SISSA
Scuola Internazionale Superiore di Studi Avanzati

Neuroscience Area – PhD course in
Cognitive neuroscience

Learning to recognize novel words and novel objects

Candidate:
Eva Viviani

Advisor:
Davide Crepaldi

Academic Year 2018-19

Abstract

Reading seems as easy and natural as listening. It is still not clear how we acquire this skill, and how visual word identification mechanisms are refined through reading experience. Theoretical models of word recognition describe general principles of skilled reading behaviour. However, these models have been based on averaged data from relatively small samples of skilled readers, mainly English native speakers, and are based on the assumption that skilled reading involves a specialized system of word identification. In this thesis it is proposed that expert reading requires the development and refinement of basic visual processing mechanisms originally employed to identify everyday objects, and then adapted to reading. To test this hypothesis, I carried out three experiments investigating: (i) how L2 visual word recognition changes with growing proficiency; (ii) how novel lexical memories are integrated into the lexicon, i.e., how they interact with previously existing words; and (iii) how sensitivity to the lexicon statistics plays out in the process of learning a novel set of visual stimuli, either in the language and non–language domain.

List of Contents

List of Tables

List of Figures

Chapter 1

General introduction

At first glance, skilled reading appears to be a relatively straightforward process of visual pattern recognition. However, reading is a culturally engineered skill less than 6,000 years old (Carr, 1999)– too recent for the evolution of specialized cognitive processes devoted to reading (Dehaene et al., 2010). Humans are therefore not evolutionarily prepared to learn to read as they appear to be for comprehending and producing spoken language (Pinker, 2009), such that many researchers have considered it almost an "innatural act" (Gough and Hillinger, 1980). Indeed, reading is a very complex skill. It requires basic visual processing, letter identification, the integration of orthographic information over space and time, and access to meaning. However, despite its cultural recency and complexity, reading comes so naturally that feels easy. In my thesis, I investigated how we become skilled readers, focusing on the steps that the visual system of a proficient reader applies to the problem of identifying written words, i.e., visual word recognition.

Successful readers in fact develop an automatic word identification system that supports fast, efficient access to all components of a word's identity. Convergent evidence for the important role of early visual processing in skilled reading is provided by neuroimaging comparisons of successful and unsuccessful readers. For instance, a review by Pugh et al. (2001) suggests that successful reading is associated with higher levels of activation in the ventral visual processing areas corresponding to the Visual Word Form Area (VWFA) during the early stages of word identification, as well as with stronger activation in temporo–parietal areas of the dorsal visual system during later processing, while less skilled readers rely on activation of frontal regions (e.g., Broca's area) as a compensatory strategy. These evidences suggest that skilled readers rely on automatic word recognition pathways, while less skilled readers don't. What drives these changes in the visual system?

It has been suggested that fully automating word identification strategies depends on a transition from sensitivity to larger orthographic chunks, such as syllables, bigrams, to morphological units such as roots (e.g., -mit- from admit, submit), affixes (e.g., pre-, ad-) and suffixes (e.g., -er, -ing) (Ehri and Wilce, 1985). This characterization of skilled reading is also consistent with models of visual word recognition which assume that perceptual features (e.g., letters) are automatically mapped to lexical forms. These models provided the foundation of current knowledge about what skilled reading entails.

1.1 Models of lexical processing

The visual word recognition domain has provided considerable advances in cognitive science for the development of models for mapping visual patterns onto phonology and meaning. This section provides a major overview of the different theoretical perspectives.

In alphabetic writing systems, the building blocks of words are letters, and so the recognition of letters was central to early models of visual word processing. The interactive activation model (IA) of letter perception developed by (McClelland and Rumelhart, 1981) involves indeed three levels (features, letters and words), and two types of connections across representations, i.e., facilitatory and inhibitory. Presenting a word activates the feature, letter and word level representations consistent with the visual input. For example, when a word is presented the first step is the analysis of individual letters that are case, position, color and font invariants. Subsequently, activation passes up from features to letters to words, with activated units inhibiting competitors at the same level. As a unit increases in activation, it feeds activation back down the hierarchy so that the lower-level units whose activation has been successfully matched with the input at the higher level will be strengthened. This additional top-down influence of word-level and letter-level representations drives the word superiority effect (Reicher, 1969). From this interaction of activations throughout the system, a single word unit will eventually reach an activation threshold that allows the letter string to be recognized as a word. The IA model contains word level representations, but it was primarily developed to explain letter (e.g., word superiority effect) rather than word recognition effects. For instance, both reading aloud and lexical decision (classifying letter strings as words and nonwords via button presses), map onto processes involved in a word level representation at the phonological, orthographic and semantic level, reaching threshold to produce the appropriate response, i.e., the correct pronunciation or the correct word/nonword response.

To this end, the parallel distributed connectionist model developed by Seidenberg and McClelland (1989) identifies lexical information with the activation distributed across layers of nodes of a network that represents phonology, semantics and orthography. A a set of input units codes for the orthography of the stimulus and map onto a set of hidden units, which in turn map onto a set of phonological units that code the pronunciation of the stimulus. Initially, the pathway weights are set to random levels. Gradually, through the learning mechanism of *backpropagation*, the connections across levels are adjusted to capture the correct pronunciation when a given orthographic string is presented. The PDP approach has been considered very useful for its ability to simulate the statistical relationships between orthographic, phonological and semantic layers. It is appealing because (1) it includes a learning mechanism; (2) it does not contain any formal spelling-to-sound "rules", but instead mimics rule-like behavior based on the statistical properties of spelling-to-sound mappings.

Modular views of skilled reading instead hypothesize distinct reading pathways for words and non-words. Among these models, there is the dual route cascaded model (DRC) (Coltheart et al., 2001). DRC has two distinct pathways for pronouncing a word aloud: a direct *lexical route* that maps letter strings onto a lexical representation present in memory, and a *sub-lexical route* that maps the letters within the string onto its pronunciation based on previously known grapheme-phoneme correspondence rules. These rules are absorbed via exposure on purely statistical grounds; that is, /k/ is the phoneme most commonly associated with *k* in English. This model explains very well why words that adhere to the grapheme-phoneme mapping are pronounced faster than words that violate it (irregular words), but also general frequency effects on the basis that the lexical route is frequency modulated, but the sublexical one is based on assembly mechanisms that are insensitive to whole-word frequency.

A hybrid model of speeded pronunciation developed by Perry et al. (2007) and is called CDP+ (connectionist dual process) model. The CDP+ takes inspiration from the DRC model (Coltheart et al., 2001), except that the DRC's rule based sublexical route is replaced by a two layer connectionist network that learns the most reliable spelling-sound relationships in the language.

Finally, an alternative approach is the Bayesian Reader model developed by Norris (2006) to account for visual lexical decision data. This model assumes that readers behave like decision makers who compute the probability that the visual input would be a word rather than a nonword.

A feature common to all these models is the little emphasis on intermediate steps between the recognition of letters and whole words in the lexicon. However, written words in alphabetic writing systems consist of letters, which make up syllables, and morphemes that have dedicated entries in the lexicon. To date, researchers have extensively investigated whether a word's morphological structure has consequences for word identification in reading, and have developed specific morphological processing models in order to account for it.

1.2 Morphological processing

There is little doubt that morphological structure is represented in the mental lexicon (Amenta and Crepaldi, 2012). In linguistic terms, morphemes are the smallest unit bearing meaning. Morphology captures the mapping from form to meaning of frequent patterns within words. It introduces some amount of non-arbitrariness since it enables the reader to guess the meaning of a word by its constituents (the meaning of hunter can be derived by the meaning of *hunt* and the meaning of *er*). Indeed, researchers have identified functional sublexical units (i.e., representations smaller than a word, such as letters, morphemes, and syllables) mediating word recognition (e.g., Carreiras and Grainger, 2004) (albeit not all researchers agree on this (e.g., Plaut et al., 1996; Baayen et al., 2010)). Researchers argue that morphological decomposition is particularly useful when the word is morphologically highly complex. The more a word is complex, i.e., comprised of a combination of prefix and suffix in addition to the root (e.g., re-establish, re-establish-ment), the more the visual system is encouraged to parse it via its morphological constituents. To date, different models have proposed their view about how morphemes are decomposed during reading, and how this mechanism affects the early stages of visual word recognition.

The usual distinction is between *full parsing* and *full listing* theoretical accounts. Full-listing accounts argue that complex words are stored and retrieved as whole (Burani et al., 1984; Burani and Laudanna, 1992; Butterworth, 1983), while full-parsing accounts argue for the usefulness of storing and activating constituents of complex words. In other words, it assumes that complex words have to be decomposed in their constituents in order to be accessed in the lexicon (Taft and Forster, 1975, 1976). In between there are models that posit a dual-route account that assumes activation of one of the other mechanism depending on the frequency, familiarity, transparency of a word and an intermediate lemma level (Taft, 1994; Libben and Jarema, 2006; Diependaele et al., 2009; Crepaldi et al., 2010). The models under the dual-route account vary in their predictions about whether one or both routes are activated (e.g., Caramazza et al., 1988) and whether they work in parallel or in a horse-race fashion (Andrews et al., 2004).

Experiments investigating the role of morphological processing are usually based on a priming technique for which a lexical decision (word/nonword judgement) is made. In the case of priming research, the same target word is compared when preceded by different word primes and the impact of different relationships between the prime and target can therefore be measured. This method gives researchers a window into the study of the morphological structure that is informative with regard to lexical representation. However, strategic and episodic factors can be invoked to explain priming effects because of the overt presentation of the prime, mining any link to underlying early stages of visual word recognition. This led to the development of a priming procedure that involves also a mask (#####), and a brief presentation (around 40ms) of a word prime to overcome such possibilities. In masked priming, primes facilitate the recognition of targets in several circumstances, for example when the prime is a pseudoword which shares all the letters but one with the target (bontrast-CONTRAST) (Forster and Veres, 1998), but also when prime and target share a semantic relationship in addition to orthography (*hearty-HEART*), such that the relationship is considered morphologically *transparent*. Gonnerman et al. (2007), adopting a PDP model perspective, explain the transparency effect as resulting from hidden units able to capture the relationship between form and meaning. The more transparently related two words are in both form and meaning, the greater the overlap in their pattern of activation within those hidden units.

Intriguingly, also words that are not genuinely morphologically complex and are opaque semantically cause facilitation, though to a lesser extent than genuine derivations (Rastle et al., 2004; Longtin et al., 2003). This is the case of CORNER (*corn + er*), where the meaning cannot be derived by its constituents, but it is apparently morphologically complex. This effect has been replicated extensively, and seems most readily explained by the existence of a stage of decomposition that is blind to semantic factors, namely, a morpho-orthographic stage (Rastle and Davis, 2008). On this view morphology brings structure not only through the form-meaning mapping, but also within orthography itself. Groups of letters corresponding to morphemes occur frequently in a combinatorial way, and are used frequently in one context. We might expect also that those units develop an orthographic representation per se that is activated on the basis on the mere frequency of occurrence, rather than on semantic properties. Crucially, this means that frequent units exert their influence in visual word recognition irrespective also of their linguistic morphological status – Whatever word that is composed by frequent units, even though not genuinely morphologically complex like CORNER, is stored in the mental lexicon in a decomposed fashion. The idea that the extraction of morphological constituents is a key automatic and unconscious step in visual word recognition has been confirmed by several masked priming experiments (Rastle et al., 2004; Rastle and Davis, 2008), though not all researchers fully agree on the fact that words are decomposed in a semantic blind fashion (e.g., Feldman et al., 2012, 2009).

What role do morphological representations play in becoming skilled readers? the role of morphology in learning to read is not very well understood. Morphological awareness (the ability to recognize and manipulate morphemes) has shown to be linked to reading skills in children (Carlisle, 2003, 2004; Deacon and Kirby, 2004; McBride-Chang et al., 2003), and can be used to overcome reading difficulties over the course of reading development (Arnbak and Elbro, 2000; Casalis et al., 2004). However children differ from adults in their responses to the variability of morphology (Dawson et al., 2017; Hasenäcker et al., 2017). On the other hand, models of lexical access are of little help since they assume that all skilled readers have developed homogenous high-quality representations for all words in their vocabulary, and therefore they access it all with the same precision. Recently there has been an extensive investigation of individual differences among skilled readers (e.g., Andrews, 2012, 2008; Andrews and Hersch, 2010). As a result more and more researchers believe that investigations of individuals language competence should play an important role in future refinement of models of word recognition.

1.3 The myth of the skilled reader

As reviewed earlier, models of visual word recognition assume that all words in the lexicon have fully specified connections between letter, morpheme and words units, and a perfect balance between different sources of words knowledge (orthographic, phonological and semantic). However reading behaviours are all but homogenous. Learning new word forms and word meanings arises from a variety of experiences with words, such that links between phonological, semantic and orthographic aspects of a novel word are experience dependent. Recognizing individual differences in visual word recognition means recognizing also two things: (i) that words knowledge can be partial, and (ii) that visual word identification depends on the completeness and precision of a lexical representation.

For instance, with respect to morphological masked priming, the debate about whether morphological constituents are decomposed in a semantically blind fashion or not still rages among researchers. Rastle and Davis (2008)'s meta-analysis showed that transparent and opaque primes facilitate recognition of targets relative to orthographic control pairs, as predicted by a semantically blind morpho-orthographic stage of decomposition account. However, a reanalysis of essentially the same data by Feldman et al. (2009) concluded that priming indeed was present for both transparent and opaque pairs, but was stronger for the former pairs reflecting an early involvement of semantics. Recently Andrews and Lo (2013) advanced the possibility that between form-then-meaning and semantic influence accounts there might be a common truth, that is, not all the readers decompose morphologically complex words early on in the same way. Andrews and Lo (2013) reasoned that these data were assuming a uniform reading behaviour from their sample of participants by averaging the data, but careful consideration of their spelling and vocabulary ability would have shed light on these apparent discrepancies.

Andrews and Lo (2013) implemented the same type of masked morphological priming task. They compared priming for semantically *transparent* morphologically related pairs (e.g., dealer-DEAL), pseudomorphemic pairs that share an apparent morphological relationship but are *opaque* semantically (e.g., corner-CORN), against *orthographic* controls where the prime is not morphologically complex (e.g., brothel-BROTH). A principal component analysis was used to individuate individuals reading profile based on spelling and vocabulary measurements, in contrast to a composite measure of global proficiency. Results showed that high overall proficiency predicted significantly faster responses to transparent stems but didn't interact with priming in any condition. However, individuals with a spelling reading profile showed equivalent priming for transparent and opaque pairs, and both produced stronger priming than orthographic controls, whereas participants with a vocabulary reading profile showed priming for transparent but not opaque or control pairs. These results were interpreted by the authors as supporting both form-first and early semantic models, highlighting that when readers relied more on spelling the priming was consistent with the form-first account, while the reverse was coherent more with a semantic profile as predicted by early semantic influences.

The study of Andrews and Lo (2013) showed that individual differences due to reliance on different aspects of word's knowledge critically affects visual word recognition, challenging the common assumption that all readers read in the same way (Andrews, 2012). This uniformity assumption made possible to lay down the foundation for theories of lexical processing and reading when a focus on individual differences may have impeded the discovery of general principles of skilled reading behaviour. However, according to Sally Andrews, it is time for experimental psycholinguists to consider whether and how individual differences modulate skilled lexical retrieval in order to refine theoretical models of reading (Andrews, 2012).

Perfetti and Hart (2002) in their Lexical Quality proposal already urged researchers to investigate linguistic differences reflected in the way words are accessed into the lexicon. The construct of lexical quality was first introduced by Perfetti and McCutchen (1982) in the context of their verbal efficiency theory and subsequently defined as the "extent to which a mental representation of a word specifies its form and meaning components in a way that is both precise and flexible" (Perfetti, 2007, pag. 359). High-quality representations are said to be constituted by orthographic, phonological and semantic information tightly bound together. These strong connections lead to more reliable and precise activation of the various word constituents, enabling the reader to achieve fast and automatic word recognition (Perfetti, 2007). Rather than assuming a malfunctioning in the reading system, the Lexical Quality Hypothesis (LQH) assumes that lexical knowledge determines the level of efficiency in lexical access. In this sense, reading abilities are considered along a continuum in which novice and skilled readers are at opposite ends. The more a word's orthographic, phonological and semantic constituents are consolidated and strongly interconnected, the more skilled a reader is.

Indeed recent evidence has linked the breadth and the depth of a person's word knowledge to a range of reading behaviours along the lines of Sally Andrews, giving credit to Perfetti's LQH. Most of them based their measurements on reading comprehension and vocabulary used as predictor variables (e.g., Ashby et al., 2005; Yap et al., 2009; Andrews, 2012; Andrews et al., 2004; Andrews and Lo, 2013; Andrews, 2008; Burt and Tate, 2002). For example, Yap et al. (2009) showed that participants with high average vocabulary showed an additive relationship between high-frequency words and semantic priming, while lower vocabulary participants showed larger priming effects for low-frequency words. Indeed vocabulary has been strongly associated to variability in the speed of responses to lexical decision tasks in Yap et al. (2012) as well, suggesting that higher lexical integrity is associated with more automatic and faster lexical retrieval processes. Similar conclusions have been drawn from Ashby et al. (2005) that compared the effects of predictability and frequency on eye movements during sentence reading as a function of a composite measure of reading comprehension and vocabulary in skilled readers. The highly skilled group showed similar reading behaviour for predictable, neutral and unpredictable sentences, suggesting that they relied on automatic, context-independent word identification processes. By contrast, the less skilled group was strongly influenced by sentence context. Ashby et al. (2005) suggested that even within the right-skewed distribution of skilled readers, it is possible to see how lower levels of proficiency were associated with increased reliance on context, probably to compensate for less efficient and automatic sentence reading mechanisms.

Collectively these results suggest that vocabulary, reading comprehension and spelling strongly impacts lexical access. However it is worth noting that participants in these experiments were all skilled readers. An overlooked aspect where individual differences are extremely important is in the moment of learning a second language. In this respect, adult learners show greater individual variability than native speakers (Grosjean, 2010), yet research on L2 individual variability in morphological masked priming is still in its infancy.

1.4 Morphological processing in second language learners

Global estimates of the number of bilinguals dated back to 2010 are as high as 50% (Grosjean, 2010). Yet there has been little research investigating individual variability in morphological processing in L2. The majority of those studies investigated whether L2 adult proficients were able to recognize L2 words like in their native L1 language. Focusing on areas found to be prone to error for adult learners, such as morpho-syntax and inflectional morphology, researchers have mostly concluded that L2 word recognition reflects incomplete or unstable grammars (Johnson et al., 1996), or is the product of processing problems (Prévost and White, 2000).

According to the shallow-structure hypothesis (SSH), L2 speakers does not possess the kind of information required to process morphological information in a native-like fashion, forcing L2 learners to fall on shallow parsing strategies (Clahsen et al., 2010). However studies investigating morphological processes in non-native language are very few, and very controversial. Some studies did not find any L1/L2 differences in production latencies (Beck, 1997), or priming patterns of regularly inflected word forms (Basnight-Brown et al., 2007), but others did, particularly with respect to regular inflection (e.g., Silva and Clahsen, 2008; Neubauer and Clahsen, 2009). For example, Silva and Clahsen (2008) examined regular past tense forms processing in English as second language in a lexical decision task and obtained larger frequency effects than attested in a L1 control group. Furthermore, they observed priming for regular past tense forms in the L1 group but not in L2, suggesting that L2 learners do not decompose regular past tense forms like natives. Another study of Neubauer and Clahsen (2009) compared regular and irregular participle forms of German in two different groups of L1 and L2 speakers. Again stronger reliance on frequency of occurrence was found for the L2 group, and priming was present in both participant groups for regular participles, but for irregulars only L1 group showed significant priming. Collectively, these studies are interpreted as indicators that adult L2 learners process regular and irregular inflected word forms differently from native speakers.

What about derived word forms? Derivational morphology has only been started to be examined in second language processing. In one of the experiments of Koda (2000) L2 learners of English had to express explicit judgments related to whether a series of morphologically complex words could have been decomposed in prefix+stem. Authors used a mix of existing and novel derived word forms with one of four prefixes (con-, de-, in-, re-) and as controls free stems that shared the same initial orthographic sequence as the prefixed words (e.g., re-gime, in-fant, de-part). Results showed shorter reaction times in the prefixed condition than in the free stems, suggesting that L2 learners were sensitive to the internal structure of the derived words and were using it explicitly. However, this task doesn't tap into automatic unconscious processes of language comprehension. Other studies that do investigated automatic recognition in masked priming procedures, showed again contrasting results (e.g., Silva and Clahsen, 2008; Diependaele et al., 2009).

In the end the debate has been focused mainly on the presence or absence of differences between L1 and L2, suggesting a black-and-white scenario. However, even the SSH of Clahsen and Felser (2006) posits gradual L1/L2 differences (Clahsen and Felser, 2018), similarly to what Sally Andrews proposed for native speakers. The existence of more graded differences between L1 and L2 has been proposed also by Lindsay and Gaskell (2010) in attempt to reconcile native and non-native language processing with the Complementary Learning Systems account. The literature on L2 processing is also vastly based on averaging large samples of individuals without considering their characteristics. However, it's even more important to assess individual variability in L2 learners, since their word recognition skills are not comparable to monolinguals in their native language (Robinson, 2003; VanPatten and Benati, 2015). Theoretically, identifying the contribution of L1 knowledge and skills to L2 learning can inform theories about cognitive overlap between first and additional languages (e.g., Abutalebi, 2008). In fact, bilinguals' visual word recognition might rely on a different type of word's knowledge for each language. The majority of studies investigating individual differences focused on English (e.g. Andrews and Hersch, 2010; Andrews, 2012). One could argue that word recognition mechanisms might rely on different aspects of word knowledge based on the language's characteristic, i.e., for English word recognition spelling might be essential, but not for more transparent languages such as Italian. Practically, understanding how existing skills of learners might impact their individual learning trajectory can lead to improved teaching methods, tailored to specific learner profiles.

In sum, recent discoveries recognize that reading abilities "accumulate with age and experience" (page 380 Perfetti, 2007), and greatly vary across individuals. Perfetti and Hart (2002) and Andrews and Lo (2012) make specific claims about the nature of lexical representations by distinguishing high versus low quality lexical representations. However, high quality lexical representations do not necessarily support direct, automatic identification, particularly for words that are orthographically very similar. Crucially, the degree of phonological overlap or similarity among words is source of ambiguity, and a mechanism of lexical competition helps sort out ambiguous inputs by suppressing alternative candidate words (i.e., competitors). Such interference is common in word recognition and constitutes a key component of the real-time dynamics of skilled word recognition.

1.5 Competition and facilitation within lexical neighbourhoods

Priming studies that use behavioral, eye-tracking and electrophysiological measurements highlight the interactive and dynamic nature of lexical processing and representation in the adult lexicon. These studies suggest that word recognition routinely involves simultaneous access to other words that overlap with a spoken word on phonological (Slowiaczek et al., 1987; Marslen-Wilson, 1990), semantic (Meyer and Schvaneveldt, 1971; Rugg, 1985) or other perceptual dimensions (Dahan and Tanenhaus, 2005; Huettig and Altmann, 2007). This happens to be the case also during visual recognition: when we identify a word there is an early stage at which a number of similar spelled lexical items (neighbours) are partially activated. This collection of candidates are similar at the phonological or orthographic level with the target word (Landauer and Streeter, 1973). During visual word recognition, lexical candidates are progressively deactivated until one lexical unit remains active, i.e., the perceived target (Murray and Forster, 2004; McClelland and Rumelhart, 1981; Coltheart et al., 2001; Grainger and Jacobs, 1996; Selfridge, 1958). However this process of elimination critically depends on the density of the word's neighbourhood and by whether this is phonological or orthographic.

Indeed, phonological and orthographic neighbourhoods constantly interact during visual word recognition, specifically with respect to orthographic neighbourhood density. In the auditory modality, much prior work has shown that increasing phonological neighbourhood density systematically gives rise to inhibitory effects in spoken word recognition. Words with many phonological neighbours are harder to recognize in noise (Cluff and Luce, 1990) and they take longer to recognize in auditory lexical decision tasks (Goldinger et al., 1989; Vitevitch and Luce, 1998). However, increasing the number of phonological neighbours while holding fixed the orthographic ones has been shown to provoke a facilitatory effect (Yates et al., 2004; Grainger et al., 2005; Ziegler et al., 2003). For instance, Yates et al. (2004) reported that words with many phonological neighbors were responded faster in a (visual) lexical decision task than the words with few phonological neighbors. In a subsequent study, Grainger et al. (2005) extended these results by carefully manipulating both neighbourhood densities in French by selecting words belonging to small versus large orthographic and phonological neighbourhoods. Authors observed a facilitatory effect of phonological neighbourhood density in words with dense number of orthographic neighbours. However, exactly the opposite pattern is observed for words with few orthographic neighbours, thus generating an interaction between phonological and orthographic neighbourhoods.

That is explained by assuming a direct relationship between orthographic and phonological information like in the PDP and the bimodal interactive-activation frameworks. These models predict three different ways in which phonological neighbourhood density could affect visual word recognition. For instance, *lexical inhibition* is expected whenever the number of competitors active during word recognition is high. This mechanisms operates via mutually inhibitory connections in a lateral inhibition network. The second mechanism, termed as *global activation*, predicts that overall lexical activation is boosted by a high number of phonological neighbours provoking facilitation (Grainger and Jacobs, 1996). Finally, a third possible mechanism, *cross-code consistency*, involves levels of compatibility across orthographic and phonological representations. Specifically, auditory presentation of a target word activates automatically its corresponding orthographic representations, and viceversa. If the level of consistency between orthographic and phonological representations is high, then effects of word competitors will be facilitatory, otherwise it will be inhibitory. When the number of orthographic neighbours is maintained at a low value, then increasing the number of the phonological neighbours will generate an increase in the level of inconsistency across simultaneously activated orthographic and phonological representations, causing inhibition. Conversely, when number of orthographic neighbours is high, increasing the number of phonological neighbours will decrease the level of inconsistency, causing facilitation. Such inconsistent mappings between spelling and sound are known to hamper visual word recognition. This effect was predicted by Stone and Van Orden (1994) and by Seidenberg and McClelland (1989) in their triangle model, and successfully tested by (Grainger et al., 2005). Importantly, this is expected in the lexical decision task because it uses some measure of global lexical activity. Clearly, an important topic for future research is to examine in detail the impact of both orthographic and phonological neighbors in visual word recognition and reading across a range of languages.

So far I have reviewed evidence of word neighbours effects when a known word is seen. However, what happens when we add a novel word into the lexicon? It has been suggested that learning novel words alters the process by which words compete with each other during recognition. As we acquire more words, the lexicon becomes populated by more lexical entries resulting in both more neighbors for a given word and, likely, greater overlap between them. This is thought to impact visual and speech recognition, forcing the reader/listener to develop more precise phonological and lexical representations. However, such growth may also increase demands on lexical access, requiring learners to acquire more robust processing strategies that resolve competition more efficiently in the face of a changing lexicon.

1.6 Populating the lexical neighborhood

People learn new words throughout adulthood; the lexicon is the probably the most plastic part of the linguistic system. Indeed research show that adults are remarkably fast and efficient in creating novel lexical representations. Known words however display unique dynamic behaviours. In the literature on word learning there is a distinction between knowing the form or meaning of a word (lexical configuration), and the ability of the word to engage with other words or linguistic representations in the lexicon (lexical engagement) (Leach and Samuel, 2007). Lexical configuration refers to knowledge of the factual information about a word (e.g., spelling, phonology), and the meaning of the word. Lexical engagement refers to the dynamic behaviour of the novel words with respect to other lexical or sublexical units (word neighbours effects). In order to conclude that novel words have entered the mental lexicon, one would like to see evidences of both lexical configuration and lexical engagement, as proof that the word has formed links with other lexical item and sublexical units.

Another reason why lexical engagement is important has to do with the formation of episodic memory traces. Davis et al. (2009) proposed a two-stage account of novel word learning that incorporated the *Complementary Learning Systems* framework, a.k.a, CLS (McClelland et al., 1995). They hypothesized that a new word is initially stored as a distinct episodic trace in the hippocampus, but becomes integrated into long-term memory over time, particularly over sleep (Dumay and Gaskell, 2012). According to this account, the hippocampal route rapidly accomodates new words using sparse representations that mediate the mapping from auditory areas to lexical and semantic areas. These mediators are used during early stages of word learning, until direct cortical mappings can be gradually built via consolidation. During recognition the hippocampal route does not allow newly learned words to be activated quickly enough to compete against the much swifter activation of the known words, which are accessed more directly. However, system consolidation gradually strengthens direct cortical links of novel words leading to faster activation and more inhibition, thus heightening competition effects.

The relation between hippocampus and neocortex during memory consolidation has been supported by several evidence. For instance, Davis et al. (2009) showed hippocampal activity associated with new word learning, plus changes in activity in the superior temporal gyrus (phonological/lexical encoding) after consolidation. Tamminen et al. (2010) also found a robust correlation between sleep spindle activity measured using sleep EEG recordings and the degree of lexical engagement.

To date lexical engagement has been widely replicated and linked to neural substrates of learning, becoming undoubtedly the marker of lexical integration. Nevertheless there are some evidences that novel word's ability to influence lexical processing is immediately available (e.g., Lindsay and Gaskell, 2013), whereas in other circumstances it takes longer to emerge (e.g., Gaskell and Dumay, 2003b). It has been suggested that an intervening factor is how much overlap there is between novel and prior knowledge. Recent neuroscientific and computational work suggests that information can be integrated in neocortical areas soon after learning without interference (McClelland, 2013; Tse et al., 2007; van Kesteren et al., 2010, 2013). This is in line with the most recent account of the CLS model, which emphasises that neocortical learning is not slower per se, but dependent on prior knowledge: new information that is consistent with existing knowledge produces little interference (McKenzie et al., 2014; James et al., 2017).

Experimentally manipulating the link between previous and new knowledge can shed light on novel word learning and the mechanisms by which new information capitalise on prior knowledge. Along these lines, the literature on individual differences contributes to clarify to this issue. In a seminal paper, Andrews and Hersch (2010) found that effects of word neighbour primes depended on the spelling ability of participants. Inhibitory priming effects were found in better spellers, whereas poorer spellers showed facilitatory priming. Better spellers would have more precise whole word orthographic representations, i.e., representations that specify precisely which letters are where in the word, whereas poorer spellers, but nevertheless good readers, were able to solve the task with less well-specified whole-word orthographic representations.

Collectively, these studies show that the way in which we use words is constantly shaped by experience. Facilitatory influences will dominate processing when primes are nonwords, but as soon as primes are lexicalized, lexical inhibition will emerge. Lexical competition is meant to resolve ambiguities in the input and increase processing efficiency in a vocabulary that is constantly growing. To date, the exact locus of the plasticity underlying these changes is not known. Possible candidate are Words neighbours effects, that are driven by mechanisms of lexical inhibition that operate at the word level.

This intriguing phenomenon speaks in favor of a change in sensitivity to orthographic regularities. This sensitivity arises as a set of associations between letter co-occurrences learned via reading. Also morphology might be learnt in this way. Sensitivity to orthographic regularities might be the first step towards the extraction of larger meaningful orthographic chunks, i.e., morphemes. Sensitivity to orthographic regularities, as well as morphology, thus emerge within individuals in the course of learning novel words. This idea is coherent also with Baayen's naive discriminative learner, where associations between forms and meanings are incrementally updated through experience (Baayen et al., 2010). As a result, the task of the learner is to acquire the lexicon and to generate generalizations over the lexicon in order to create shortcuts that would enable it to rapidly recognize known words, probably developing strategies to deal with interference. While these on the surface may seem like distinct tasks, they might arise from a general statistical learning principle.

1.7 Learning and generalization of orthographic knowledge

Our visual system, originally devoted to recognize objects and faces, seems to systematically channel information about written words to the Visual Word Form Area located in the occipito-temporal left hemisphere (Dehaene et al., 2010; Pugh et al., 2001). However, it does not merely respond passively and innately to anything that resembles a letter or a word, but it seems to reflect the regularities present in the language to which it's exposed.

Much evidence supports the idea that the perceptual system becomes selectively efficient at processing inputs that are encountered frequently in reading. For example, it has been consistently shown that the brain compiles statistics about letters that belong together. This is illustrated by the fact that letters embedded in words (such as S in the English word FLASH) or in word-like letter strings (S in FRISH) are more efficiently recognized than letters embedded in unusual letter strings (S in RFHSL), or presented in isolation (Hildebrandt, 1994; Reicher, 1969). The brain also prefers frequent letter combinations, like "ER" or "ING", to rare or impossible ones such as "HW" or "XFQ" (Vinckier et al., 2007; McClelland and Rumelhart, 1981; Binder et al., 2003). Other studies have suggested an indirect link between reading ability and sensitivity to orthographic regularities. For instance, Conrad et al. (2013) showed that better readers/spellers consider more often items with frequent letter clusters more wordlike than items with low-frequency or illegal clusters when are confronted with novel stimuli. Sensitivity to orthographic regularities has been also shown to improve from kindergarten to first grade and to account for a significant amount of unique variance in their reading and spelling performance (Rothe et al., 2014). Such evidences suggest that readers use information about frequently recurring letter combinations to efficiently perceive and identify letters and letter strings. Yet, if the person tested has not learned to read, or has been exposed to a different language (e.g., Hebrew), he will fail to show any of these effects (Dehaene, 2009), suggesting that our sensitivity to frequent letters and letter combinations is not merely determined by visual stimuli, but also by the cultural history of the reader's brain.

Broadly, statistical learning refers to learning about the frequency with which features occur and co-occur. The term has been proposed by Saffran et al. (1996), corroborated by findings of sensitivity to distributional properties of the language in 8-months children. Its appeal lies in part in its ability to be applied across a diverse range of domains, such as the learning of auditory (Saffran et al., 1999; Aslin et al., 1999) visual (Fiser and Aslin, 2002; Kirkham et al., 2002), and tactile stimuli (Conway and Christiansen, 2005). It is thus indisputable that we are equipped with statistical learning mechanisms able to detect the regularities of the outside world. What remains unclear are the details of the relationship between statistical learning and reading.

In fact, whether the human reading system profits from the statistical regularities present in the written material remains a hotly debated topic in the psycholinguistic community. For instance, performance in behavioral tasks, such as letter/word detection or lexical decision, tends to be modulated by n-gram frequencies when using non-words as stimuli, but results are mixed with real words stimuli (for a systematic review on this, see: Chetail (2015) and Schmalz and Mulatti (2017)).

On the other hand, Dehaene et al. (2005) in their neural model of reading (the local combination detector or LCD model) argues that n-grams act as a bridge between letters and words, and are therefore essential in visual word recognition. The model is based on the assumption that word recognition obeys to the same principles that govern the organization of the primate ventral visual system. In fact, studies on the macaque monkey brain has led researchers to reconstruct the existence of converging neural detectors with progressively larger receptive fields tuned to increasingly complex objects (Baker et al., 2002). The LCD model argues that learning to read consists in the development of a hierarchy of neurons that encode increasingly more complex and larger word fragments, such as letters, bigrams or quadrigrams, until reaching maximum complexity in whole words, creating the visual word-form system (Vinckier et al., 2007). Crucially, the model posits that these word constituents are originally encoded as any other visual feature, and only by compiling statistics about letters through exposure to written text, neurons acquire preference to it. Sophisticated statistical learning techniques detect regularities of the visual world, that are thus extracted and stored in our cortical connections ready to be evoked for fast and efficient recognition. In this sense, according to Dehaene et al., any word is just a visual stimulus, and the ability to visually identify words can be accomplished by any primate visual cortex equipped with the basic principles to solve visual recognition problems.

A recent experiment of Baker et al. (2002) nicely illustrates this point. In this experiment monkeys learned in around 7000 trials to recognize sticks with characteristics shapes at their two extremities, for instance a star at one end or a triangle at the other. After training, several neurons started to fire to their precise combinations. Crucially, when only part of the object was shown, neurons responded feebly, suggesting that neuronal activity associated to the whole learned object was greater than the sum of the individual shapes considered individually, coherently with the Gestalt principles (Gibson, 1950). In line with this, also baboons showed to be able to distinguish words from nonwords based on co-occurrence letter statistics (Grainger et al., 2012), proving to be able to use statistics as basis to categorize new unseen stimuli. Interestingly, this result was later replicated with pigeons (Scarf et al., 2016), and a group of researchers in Trento also found preference to the object global structure of novel objects rather than its parts in chicks (Rosa-Salva et al., 2018), extending these findings also to non primates. Taken together these findings are consistent with Dehaene et al. (2005)'s proposal of a reading system that "recycles" visual processes that are domain general in nature.

If we agree that our visual system is similar enough to our fellow primates, is sensitivity to orthographic regularities part of our heritage? After all, literacy has been only invented 6,000 years ago. Research on orthographic processing in humans has repeatedly showed the special status of letters as the building blocks of single word reading. Based on the assumption that "what distinguishes letters from other kinds of visual objects is that they can be aligned in strings in order to form familiar objects" (page 17, Grainger and Hannagan, 2014), the horizontal alignment imposes significant modifications in the visual mechanisms that allows visual word recognition as opposed to object recognition (Grainger and Hannagan, 2014). In this sense, there has to be something "special" about orthographic knowledge that is unique to reading.

Taken together, the literature put forward to date has strengths and weaknesses in its ability to describe and explain whether sensitivity to orthographic regularities derive from a general domain statistical learning mechanism, or is a specialized mechanisms developed for reading.

1.8 Summary and Research questions

This section briefly summarises the experimental approaches used in this thesis, aimed at addressing a main over-arching theoretical question, that is how do we become skilled readers.

First experiment. Fully understanding how skilled reading emerges and shapes the architecture of the reading system requires, among other things, the investigation of how lexical knowledge is learned and refined through experience. Consistent with this view, a small but growing body of research has started to investigate how and whether individual differences modulate skilled reading. The literature shows that the efficiency with which readers extract the visual features of written words and map them to representations in lexical memory is highly dependent by their language competence. This appear to be strictly tied to some sort of perceptual efficiency in accessing the semantic, phonological and orthographic knowledge of a word in the early stages of visual word recognition.
However, literature so far has limited its focus of attention to relatively skilled readers, while the role of experience should be more evident for poorer readers. As a result, it remains to be seen whether such patterns can be found in adults at a low level of proficiency. An area where there is sufficient variability in visual word recognition in order to test low proficient adults is bilingual morphological processing. Bilingualism provides a tool for examining aspects of visual word recognition that are otherwise obscured by skilled lexical retrieval. Moreover, to date there is an intense debate on whether L2 readers are able to decompose morphologically complex words like in their native language, and no study assessing individual differences in this domain has been carried out yet. Hence, I propose to investigate morphological priming in native and non-native word recognition, whereby individual characteristics are qualified by a battery of tests assessing L2 English language competence. Further, in line with recent developments in the literature showing that readers' morphological processing is mediated by their sensitivity to graded, probabilistic relationships between form and meaning, I propose to investigate whether priming is modulated by sensitivity to Orthography-to-Semantics consistency (Marelli et al., 2015) as a function of language proficiency. High OSC values mean that two words (e.g.,deal, dealer) have always been associated consistently to the same semantic context, becoming both linked to one concept. Low OSC values (e.g., whisk, whisker) indicate that the two words have been found in diverse/rich semantic contexts, i.e., linked to different concepts. I hypothesize

that gaining more language competence involves growing sensitivity to these probabilistic ties between form and meaning in a language, which speaks for the hypothesis that learning a novel lexicon proceeds through an increased sensitivity to the statistical structure of the orthography–semantics mapping.

Second experiment. An efficient reader is equipped with lexical competition mechanisms to disambiguate the speech or visual input by suppressing alternative candidates. Such competition is central to theoretical accounts of visual word recognition. Research on word learning suggests that lexical competition tends to arise when novel words are added to lexicon, as a sign of complete lexical integration. Participants exhibit lexical competition whenever novel words overlap phonologically or orthographically with existing words. However phonological and orthographic neighbourhoods always interact with each other critically determining whether word neighbours facilitate or inhibit each other during recognition. This matter is particularly important in languages where there is a perfect alignment between the number of orthographic and phonological neighbours, such as Italian. Thus, in the second experiment I propose to investigate the alterations that novel word learning exerts onto written word identification in Italian native speakers. Further, to better understand the implications for lexical consolidation, two learning routine were implemented. A more explicit, instructed routine, where participants are directly pointed to the material to be learned, and an implicit uninstructed routine inspired by animal research, where participants have to figure out themselves the novel words based on feedback only. On the premise that uninstructed learning conveys more information because representative of everyday word learning processes, I hypothesize that uninstructed learning routine would bring about a learning benefit over the instructed learning routine. Particularly for what concerns the estimate of lexical integration, I propose to assess lexicalization via a lexical engagement behavioural task and via a visual oddball paradigm paired with electroencephalography (EEG).

Third experiment. Another key aspect defining skilled reading is sensitivity to orthographic regularities. To date the literature suggests that readers use information about frequently recurring letter combinations to efficiently perceive and identify words. Yet, evidence of the role for statistical learning of orthographic regularities in reading is mixed. In this experiment I propose to focus on a particular type of orthographic regularity, ngrams, in order to investigate their role in learning novel written words. According to influential researchers, the visual word identification system identifies recurrent letter clusters (n-grams) as a bridge between letters and words. Here I test this hypothesis directly, and explore the boundaries of this phenomenon: what type of visual object, if any, fails n-gram based statistical learning? In a non-linguistic version of a well established paradigm in reading, I propose to teach to adult participants a set of novel objects progressively less word-like (e.g., Y-shaped objects and gabor patches) made up of smaller parts that follow a particular statistical pattern. I propose to test their ability to disentangle old from new stimuli that comply with the statistical structure of learned stimuli. I hypothesize that if participants have a hard time discarding objects that they have never seen but that comply with the statistical pattern of the smaller parts, this means that n-gram coding is a general mechanism used by the brain to learn about the visual environment. Alternatively, if participants do not seem to show any sensitivity to the statistical regularities according to which objects were made, it means that n-grams are a mechanism unique to reading, as proposed by many influential researchers.

Chapter 2

Experiment 1 - Masked morphological priming tracks the development of a fully mature lexical system in L2

Manuscript invited for re-submission at Journal of Memory and Language. All the materials related to this paper (the stimulus list, the data collected, the analysis scripts, and the tools that we used during the analysis) are publicly available at the Open Science Framework: https://osf.io/jnrvy/

2.1 Abstract

Visual word identification is based on an early morphological analysis in one's native language; how these mechanisms apply to a second language is much less clear. We recruited L1 Italian–L2 English speakers in a masked priming task where the relationship between prime and target was morphologically transparent, e.g., *employer–EMPLOY*, morphologically opaque, e.g., *corner-CORN*, or merely orthographic, e.g., *brothel–BROTH.* Critically, participants underwent a thorough testing of their lexical, morphological, phonological, spelling and semantic proficiency in their second language. By exploring a wide spectrum of L2 proficiency, we showed that this factor critically qualifies L2 priming. Genuine morphological facilitation only arises as proficiency grows; and opaque and orthographic priming shrink as L2 competence increases. Age of acquisition was also evaluated, and did

not affect the priming pattern. Furthermore, we showed that L2 priming is modulated by sensitivity to probabilistic relationships between form and meaning. Overall, these data illustrate the trajectory towards a fully consolidated L2 lexicon, and show that masked priming is a key tracker of this process.

2.2 Introduction

Visual word identification occurs effortlessly in adult skilled readers. This process has received a considerable amount of attention, and there is now wide consensus that the recognition of printed words involves an early morphological analysis—words that are made up of meaningful sub–parts, such as 'kind-ness' or 'clean-er', are identified via their constituents (e.g., Amenta and Crepaldi, 2012). Masked priming experiments have further revealed that morpheme identification is primarily based on form, to the point that parsing does not only operate on transparent words such as 'dealer', but also on words whose meaning is unrelated to that of their morphemes, such as 'corner', which is not someone who corns, or 'irony', which has nothing to do with iron (e.g., Rastle et al., 2004; Longtin et al., 2003; Grainger et al., 1991; Rastle et al., 2000; Lavric et al., 2007). This interpretation has been proposed mostly by models assuming an early orthographic stage in morphological decomposition that is semantically blind, but other researchers are equally committed to the alternative explanation of an early semantic influence in morphological segmentation and decision processes (e.g., Feldman et al., 2009, 2012; Schmidtke et al., 2017). Though both interpretations are well grounded on evidences that could account for it (Rastle and Davis, 2008; Feldman et al., 2009), these studies have always found a graded pattern of facilitation where transparent primes (dealer-DEAL) and pseudomorphemic derivations (corner-CORN) induce stronger priming than an orthographic control (dialog-DIAL).

Whether the same mechanisms apply to visual word identification in a second language (L2) is far less clear. This has been investigated in a few experiments now, but both the data and their theoretical interpretations seem to diverge. Silva and Clahsen (2008) investigated masked morphological priming with derived words (e.g., bitterness) in a group of L1–English, and in different groups of advanced L2–English readers. They compared derivational with repetition priming (rigidity–RIGID vs. rigid–RIGID), and found that the two effects are equally strong in L1, but not in L2. Based on these results, the authors argue that L2 readers might only have partial access to the combinatorial processes that are necessary to appreciate morphology as in their L1.

In a further experiment from the same group, Kirkici and Clahsen (2013) investigated the processing of inflected and derived words in non–native speakers of Turkish with a series of masked priming experiments. The non–native speakers involved in this study had a variety of L1 backgrounds, but were all highly proficient. Priming in L1 turned out to be equivalent for inflection (*sorar–SOR*, s/he asks–ask) and derivation (*yorgunluk-YORGUN*, tiredness–tired). In L2 instead, derivational priming was larger than the inflectional effect, which did not emerge at all.

These reports of a different morphological priming profile in L2 were not confirmed in Diependaele et al. (2011). These authors tested two groups of Dutch and Spanish non–native speakers of English in a masked priming study, with a design that is identical to that typically adopted in the vast L1 literature—transparent suffixed primes (e.g., viewer–VIEW) were contrasted with opaque (pseudo–)suffixed primes (e.g., corner–CORN) and orthographically–matched, non–morphological controls (e.g., dialog–DIAL). They reported no statistical difference between morphological priming in L1 and L2, contrary to Kirkici and Clahsen (2013) and Silva and Clahsen (2008). Across their three experiments, they seem to observe a graded facilitation pattern where genuine morphological priming is larger than the morpho–orthographic effect, which in turns exceeds orthographic priming. However, when one considers each individual experiment, data are not entirely clear cut. Opaque priming, for example, did not differ statistically from the orthographic baseline in any of the three individual studies, making the genuine contribution of morphology somewhat unclear.

One aspect in which the studies described above agree is, interestingly, outside of their main scope. Namely, they report more orthographic priming (e.g., colonel–COLON) in L2 than in L1, and no priming at all in this latter. This pattern is confirmed in Heyer and Clahsen (2015), where transparent, derivational priming was only contrasted with the orthographic effect—opaque primes were not part of the design. In their masked condition, these authors report the standard pattern in L1, with significant morphological facilitation and no orthographic effect. In L2, instead, this latter was equal to morphological priming.

This L1–L2 interaction with form priming did not receive much attention as in all these studies the orthographic condition was effectively a control baseline. It was generally explained in terms of slower prime processing

(Diependaele et al., 2011). Here we would like to offer a possibly more intriguing interpretation, which relates to the literature on novel word learning and form priming in L1. It is rather established, in fact, that nonword primes (e.g., contraft–CONTRACT) yield larger orthographic facilitation than word primes (e.g., contrast–CONTRACT) in L1, a phenomenon known as Prime Lexicality Effect (PLE; e.g., Forster and Veres, 1998). PLE is typically interpreted in terms of lexical competition—with word primes, the gain that one gets from shared letters is offset by the competition in the lexicon between the prime and the target representations. Lexical competition, in turn, has been often taken as a benchmark for the consolidation of new lexical memories (Gaskell and Dumay, 2003b; Tamminen and Gaskell, 2008; Davis and Lupker, 2006a). Therefore, orthographic priming from real words in L2 could be attributed to a still incomplete consolidation process, whereby memories for novel words are perhaps present in the brain, but are not fully lexicalised yet. Essentially, they would work similarly to nonwords in L1; because they do not participate in lexical competition, they yield priming based on sub–lexical processing. This also connects to general theories of lexical and language learning (Ullman, 2001, 2005).

In addition to better assess L2 morphological priming, in this paper we try to replicate this orthographic effect; qualify it better through a proper consideration of individual variability (see below); and develop its theoretical implications more fully.

Why does morphological priming prove to be so difficult to characterize in L2? A strong candidate to explain inconsistency in the previous data is surely individual variability. Evidence is accumulating that visual word identification is heavily influenced by the individual profile of each reader (e.g., Andrews and Hersch, 2010; Andrews and Lo, 2013; Feldman et al., 2010; Coughlin et al., 2019; Beyersmann et al., 2015a,b). These effects may be even magnified in L2, where inter–subject variability is likely enhanced by the diversity of the learning experience. Factors like Age of Acquisition (AoA) or proficiency may well determine different cognitive processes to be in place when L2 readers are exposed to printed words. Along these lines, Dawson et al. (2017) have recently shown that morphologically structured nonwords (e.g., earist) are more likely to be taken as words than control stimuli (e.g., earilt) in adults and adolescents, but not in younger children. These data refer to L1, but do show that less experience with printed words may determine a different morphological processing—this may apply to L2 speakers just as well as to children. Another recent study by Veríssimo et al. (2017) investigated masked morphological priming in Turkish–German bilinguals, and found an effect of AoA on inflectional, but not derivational L2 priming. This shows that sensitivity to morphological features is constrained by the learning trajectory of a (second) language.

Previous studies on L2 morphological priming did not try to characterize their participants' profile in terms of proficiency or AoA beyond self reports, nor they investigated whether/how these individual features may modulate the priming pattern. Also, they seem to be based on highly proficient bilinguals only. In the present study, we assess our participants' L2 proficiency with a battery of tests covering seven language domains (morphological awareness, fluency, phonemic discrimination, vocabulary, spelling, oral and reading comprehension). We also assess AoA (and, more generally, the participants' learning experience) through a questionnaire. Most importantly, we explicitly tried to recruit readers with varying learning experiences and proficiency, so that we could properly assess whether L2 morphological priming is affected by these factors.

Another recent development in the literature is the discovery that readers' morphological processing is mediated by their sensitivity to graded, probabilistic relationships between form and meaning. Marelli et al. (2015) quantified these relationships in terms of what they called Orthography–to–Semantics Consistency (OSC)—a frequency–weighted average of the semantic similarity between all members of a given morpho–orthographic family and their stem. OSC quantifies the relationship between a letter string and the meanings of all the words that share that same stem in a corpus. Hence, can be considered an estimate of the semantic density of target neighbourhoods (Marelli and Amenta, 2018). The more two words tend to occur in similar semantic contexts (e.g., widow and widower), the more their OSC will be high, because their meanings will be considered similar, i.e., frequently associated to the same concept within the lexicon. Conversely, low OSC values (e.g., whisk and whisker) indicate that two words have been experienced in diverse/rich semantic contexts, i.e., linked to different concepts. Crucially, contrarily to LSA (Landauer and Dumais, 1997) that computes the degree of semantic similarities between two words only, OSC quantifies the degree of semantic relatedness between the stem and the members of the whole orthographic family, thus going beyond prime-target pairs and indirectly informing about the breadth of individuals' lexicon. Adopting a connectionist framework, indeed OSC reflects the outcome of a specific probabilistic type of learning, where a word can be learned through the way in which it co-occurs with other words in the lexicon.

In a large scale regression analysis, Marelli et al. (2015) showed that words with a higher OSC (that is, words that are part of semantically consistent families) are identified more quickly. Of course, the appreciation of these fine–grained ties between form and meaning likely requires a rather extensive experience with any lexicon. Thus, one can imagine that L2 speakers would show less sensitivity to OSC; or perhaps more intriguingly, that their sensitivity grows with proficiency. Or perhaps again, one needs early exposure to a language in order to see a probabilistic form–meaning relationship structure, so that only early–AoA participants would show an effect of OSC. More generally, OSC offers an interesting perspective on the learning of a second language, which may perhaps involve a growing sensitivity to probabilistic ties between form and meaning. We will try to throw light on this issue by checking whether morphological priming – and, more generally, word identification time – is modulated by OSC in L2.

So, to sum up, the present experiment tries to clarify how bilingual readers process word morphological structure in L2, primarily by characterizing their profile in terms of proficiency and age of acquisition. Moreover, we will check how/whether fine–grained, probabilistic relationships between form and meaning (as tracked by OSC) inform L2 visual word identification, which speaks to the hypothesis that learning a novel lexicon proceeds through an increased appreciation of the statistical structure of the orthography–semantics mapping.

2.3 Methods

Participants

81 students at the University of Trieste participated in the study. They were 73 right–handed and 8 left-handed native speakers of Italian, who provided informed written consent to take part into the experiment. Their mean age was 24.3 years (range: 18–34) and their mean education was 17 years (range: 13–22); 27 of them were male. Participants had a clean history of neurological impairment or learning disabilities, with normal or corrected-to-normal vision. They were compensated for their time with 20 Euros.

Materials

The Italian set of stimuli is composed of 150 prime–target pairs, 50 in each of three conditions. Primes and targets in the *transparent* condition entertain a genuine morphological relationship (e.g., *artista–ARTE*, artist–ART). Primes and targets in the *opaque* condition are semantically independent, but entertain an apparent morphological relationship, i.e., primes are made of a pseudo-stem, which is shared with the targets, and a pseudo-suffix (e.g., *retaggio–RETE*, legacy–net; an analogous example in English would be corner–CORN). Primes and targets in the *form* condition have a purely orthographic relationship, i.e., primes share a (pseudo-)stem with their targets, but end in a non–suffix (e.g., *corallo–CORO*, coral–CHOIR; an analogous example in English would be dialog–DIAL). Targets and primes were matched across condition for frequency (as indexed by the SUBTLEX–IT database; Crepaldi et al., 2013), length, Coltheart's N and prime–target orthographic similarity (see Table 2.1).

For each related prime, we selected a control prime that is semantically, orthographically, and morphologically unrelated to the targets (e.g., *plunder–ACRE*). Control primes were matched as closely as possible to related primes on frequency, length and Coltheart's N (see Table 2.1). In order to avoid multiple presentations of the same target word to the same participant, we rotated related and control primes over two lists, in a Latin Square design; thus, each participant saw each target, either paired with its related or control prime.

	Transparent	**Opaque**	**Orthographic**
Target frequency	3.96 (0.67)	3.63 (0.87)	3.94 (0.84)
Target length	5.16 (1.07)	5.08 (0.84)	4.94 (0.88)
Target Coltheart's N	18.1 (11.3)	20.1 (11.9)	21.5 (13.4)
Related prime frequency	2.92 (0.84)	3.15 (0.78)	3.22 (0.69)
Control prime frequency	2.91 (0.68)	3.09 (0.85)	3.19 (0.67)
Related prime length	7.70 (1.24)	7.96 (1.21)	7.52 (1.18)
Control prime length	7.70 (1.24)	7.96 (1.21)	7.52 (1.18)
Related prime Coltheart's N	3.6 (2.9)	3.5 (2.6)	4.2 (6.1)
Control prime Coltheart's N	3.8 (2.9)	3.8 (2.9)	3.5 (2.5)

Table 2.1: Stimulus statistics for the Italian L1 set; We report means and standard deviations. Frequency is reported in Zipf (Brysbaert et al., 2018).

150 nonwords targets were also selected to serve as NO trials in the lexical

decision task. They were matched with word targets on length (mean= 5.06, SD= 0.95). Each of these targets was paired with a word prime, mirroring the structure of the word target set: half of these primes were orthographically similar to their targets, and 2/3 of the primes were complex words. This served the purpose of leaving the primes devoid of any information about the lexicality of their targets. These prime words were also roughly matched with the word–target primes for frequency (mean=3.18, SD=0.87), length (mean=7.42, SD=1.32), and Coltheart's N (mean=3.02, SD=3.5).

The English set of stimuli perfectly mirrors the Italian one. It is largely based on Rastle et al. (2004), with only a few additions and replacements. The lexical statistics of these stimuli are reported in Table 2.2. Frequency values are based on SUBTLEX–UK (Van Heuven et al., 2014).

The entire list of Italian and English stimuli is offered in the Appendix A.

	Transparent	**Opaque**	**Orthographic**
Target frequency	4.09 (0.72)	3.88 (0.74)	3.72 (0.82)
Target length	4.92 (0.65)	4.80 (0.69)	4.62 (0.68)
Target Coltheart's N	6.66 (5.73)	9.08 (7.78)	11.72 (8.2)
Related prime frequency	3.32 (0.93)	3.43 (0.96)	3.50 (0.93)
Control prime frequency	3.30 (0.83)	3.46 (1.03)	3.47 (0.87)
Related prime length	7.12 (1.15)	7.09 (1.19)	7.15 (1.68)
Control prime length	7.12 (1.11)	7.09 (1.16)	7.15 (1.67)
Related prime Coltheart's N	1.98 (2.5)	2.50 (2.9)	2.06 (3.2)
Control prime Coltheart's N	3.4 (4.6)	2.64 (4.7)	3.04 (4.2)

Table 2.2: Stimulus statistics for the English L2 set; we report means and standard deviations. Frequency is reported in Zipf (Brysbaert et al., 2018).

Measures of proficiency in English

English L2 proficiency was assessed via a battery of tests that cover phonemic fluency, phonemic discrimination, spelling, vocabulary, morphological awareness, and oral and reading comprehension.

Phonemic fluency. Participants were asked to produce as many words as possible starting with the phonemes /f/ or /p/, in two separate 60–seconds sessions. Answers were recorded through a microphone for off–line scoring. Each participant's score is the total number of words produced.

Phonemic discrimination. Participants were acoustically presented with a probe pseudo–word (e.g., *kneef*), and then with three test pseudo–words (e.g., *yawk, zeep, wid*). They were asked to pick up which of the test pseudo–words shared one phoneme with the probe. The score is the number of correctly identified test pseudo–words, out of the 13 trials that made up the task. The shared phoneme could be either a consonant or a vowel.

Spelling. 20 words were recorded by a native speaker of English, and included in example sentences to clarify any lexical ambiguity. These words were then presented to the participants, who were required to write them. Words were selected from Burt and Tate (2002), among those that were correctly spelled by between 30% and 90% of a sample of Australian first–year university students. Latin derivations were excluded as Italian speakers may be able to reconstruct their spelling based on etymology. Participants' score for the test was the number of correctly spelled words.

Vocabulary. This task comes from the Test of English as a Foreign Language (TOEFL), and consists of 20 sentences presented in a written form that need to be completed by choosing a proper word among three alternative choices. The score for this test is the number of correct choices.

Morphological awareness. This test was presented in a written form, and consisted of 9 sentences that participants were asked to fill with an appropriate plausible pseudo–word, chosen among two options. Nonwords contained a suffix, which unambiguously indicated a grammatical class (e.g., *swishely, valgeful*). This made only one option a plausible sentence completion (e.g., *The Richter Scale measures the* __________ *(swishety/swishely) of earthquakes*). The score for the test is the number of correct picks.

Oral comprehension. This test also comes from the TOEFL. Participants listened to two conversations between English native speakers, and were then asked 6 comprehension questions about them. They marked the correct answer among 4 alternatives. The score for the test is the number of correct answers.

Reading comprehension. Participants were required to read a text passage of approximately one page, and answer some comprehension questions. This task was taken again from the TOEFL, and consisted of seven questions, each with 4 alternative choices. The score is again the number of correct answers.

Measures of Age of Acquisition of English

Age of Acquisition of English (henceforth, AoA) was assessed via a questionnaire, which we expanded to include items on perceived proficiency and language experience more in general. The questionnaire was composed of the following questions:

1. Which age were you exposed to English for the first time? (AoA proper)
2. Indicate how much do you use English in your daily life from one (never) to five (always)
3. In which context were you exposed to English for the first time— home or school?
4. Did you grow up in a context where multiple languages were spoken?
5. How would you rate your proficiency in English, from 1 (very bad) to 5 (very good)?
6. Do you speak any other language in addition to Italian and English?

Procedure

Participants completed the AoA questionnaire online through the Department participant recruitment system. The rest of the data collection happened in the lab, in two sessions. During the first session, which lasted around an hour, participants carried out the proficiency tests. During the second session, participants underwent the lexical decision experiment, both in Italian (L1) and English (L2). This session lasted around 40 minutes. The testing order for the two languages was counterbalanced across participants.

For the lexical decision task, participants were tested in a sound–proof, dimly lit booth. Stimuli were presented in a randomized order using Psychopy (Peirce, 2007), and responses were collected through a two–button, custom–made response box based on Arduino microcontroller boards (`https://www.arduino.cc/`). The YES button was always controlled by the dominant hand.

Each trial started with a string of hash marks, presented for 500ms, which was replaced by the prime, presented for 50ms in lowercase. The prime was immediately followed by the target, presented in uppercase until participants' response, or for 2000ms. Targets and primes were presented in the center of the screen, left-aligned with respect to the stem. Participants were not informed of the presence of the prime, and were asked to respond as quickly and as accurately as possible. Twelve practice trials preceded the experiment proper, so as to allow familiarization with the task. At the end of the session, participants were debriefed to check whether they noticed the presence of a prime.

Statistical Analysis

Response time analyses were carried out on correct trials only. Data were trimmed of outliers separately for the Italian (L1) and English (L2) datasets. For Italian, we excluded one participant who was aware of the primes; two participants whose accuracy on non–words was below 80%; three target words, which were responded correctly less than 60% of the times; and individual data points below 280ms or above 2500ms. This determined the exclusion of 526 datapoints, which amounts to 4.6% of all available data. We were then left with 11009 data points for the analysis.

In the English set, we excluded two participants who reported having seen the primes; one additional participant whose mean overall response time was under 200ms; and individual data points that were below 300ms or above 2000ms. This led to the exclusion of 281 datapoints, which is 3% of the total available data. The clean dataset was comprised of 8938 data points.

Linear mixed models were used to fit reaction times within the R environment (R Development Core Team, 2008) using the packages `lme4` (Bates et al., 2015) and `lmerTest` (Kuznetsova et al., 2016). RTs were inverse transformed to make residuals' distribution more Gaussian–like. The effects of interest were prime *relatedness* (related vs. unrelated), *morphological type* (transparent vs. opaque vs. orthographic), and their interaction. For the proficiency and AoA analyses, we added each individual predictor tracking these variables (i.e., each test score and each questionnaire item) to the main interaction, one at a time to avoid excessive collinearity. Trial position in the randomised list, target frequency, target length, target orthographic neighborhood size and rotation were also added as fixed effects, to control for spurious variance. Only those effects that determined a significant increase in goodness of fit were retained into the model. Finally, following Baayen (2013), we refitted all models after excluding data points that deviated from their corresponding predicted value by more than 2.5SD; this protects us from unduly influential outliers. Response time estimates based on the models were obtained through the package `effects` (Fox and Hong, 2009).

2.4 Results

The overall mean RT and accuracy in the task were 594ms and 95% respectively, for Italian; and 672ms and 75.8% for English. Table 2.3 shows mean standard deviation RTs by condition, for both Italian and English.

		L1 - Italian	**L2 - English**
Transparent	unrelated	589 (131)	667 (177)
(dealer – DEAL)	related	551 (126)	630 (183)
	effect	38	37
Opaque	unrelated	614 (150)	685 (196)
(corner – CORN)	related	598 (157)	666 (214)
	effect	16	19
Orthographic	unrelated	607 (145)	703 (214)
(dialog – DIAL)	related	606 (155)	688 (207)
	effect	1	15

Table 2.3: Raw response times in ms by condition, in L1 (Italian) and L2 (English). Standard deviations in parenthesis.

The model for the Italian data reveals a significant interaction between *prime relatedness* and *morphological type*, $F[2, 10590.4] = 48.62, p < .001$. The interaction is driven by significantly more priming in the transparent, $\beta = -125, t[10590] = -9.85, p < .001$, and opaque conditions, $\beta = -.66, t[10590] = -5.21, p < .001$, as contrasted with the orthographic condition, which does not seem to show any facilitation, $\beta = -.0001, t[10590] = -.02, p = .98$. Transparent priming is also significantly larger than opaque priming, $\beta = -.59, t[10590] = -4.67, p < .001$. The estimated RTs for each condition are plotted in Figure 2.1, left panel.

The model for the English data also reveals a significant interaction between *prime relatedness* and *morphological type*, $F[2, 8564.7] = 9.27, p < .001$. Similarly to Italian, transparent primes yield more facilitation than orthographic primes, $\beta = -.06, t[8556] = -4.3, p < .001$, thus confirming that L2 speakers are fully sensitive to genuine, semantically transparent morphology. Although with somewhat weaker statistics, opaque primes also provide more facilitation than orthographic controls, $\beta = -.29, t[8569] = -2.01, p =$

.043, suggesting that readers capture opaque morphology also in their second language. Contrary to Italian instead, orthographic primes clearly yield significant facilitation themselves, $\beta = -.046, t[8566] = -4.3, p < .001$. Finally, transparent primes seem to provide larger priming than opaque primes, $\beta = -.031, t[8559] = -2.27, p = .02$. Figure 2.1, right panel, reports the estimated response times per condition in the English dataset.

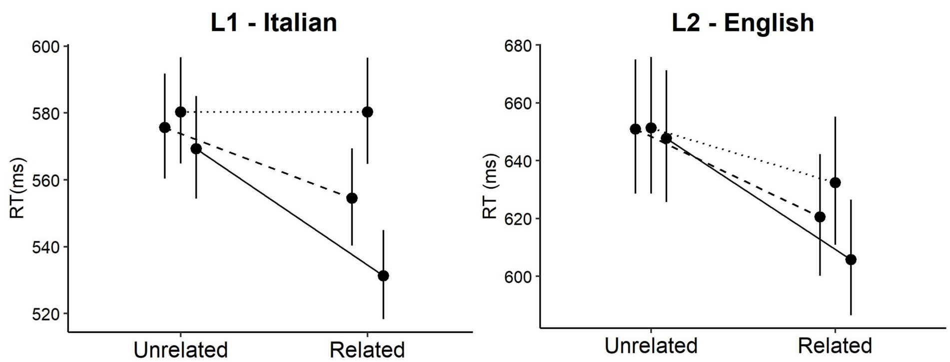

Figure 2.1: Model–based estimates of response times per condition, in L1 (left panel) and L2 (right panel). The solid, dashed and dotted lines represent the transparent, opaque and orthographic conditions, respectively. Error bars are 95% confidence intervals.

A cross–language analysis confirms that the priming pattern across conditions is different in L1 and L2, as attested by the significant interaction between *prime relatedness*, *morphological type* and *language*, $F[2, 19244.6] = 4.81, p < .008$.

Language proficiency and priming

Proficiency scores are distributed as illustrated in Figure 2.2—we were able to sample a rather wide distribution of proficiency, across different linguistic domains. Permutation–based split–half reliability was computed separately for each proficiency subtest using the `splithalf.r` function from the `multicon` package Sherman (2015) based on 5000 random splits. Average of split-half reliability (Spearman-Brown corrected) varies from 0.81 to 0.89, Cronbach's α from 0.79 to 0.84, showing good reliability for our proficiency metrics. The correlation between pairs of indexes varies between .25 and .68 (lower quartile= .43, median= .46, upper quartile= .54), which attests the effectiveness of the battery—individual scores correlate enough to be credible measure of individuals' proficiency, but also vary enough to effectively track different aspects of L2 competence.

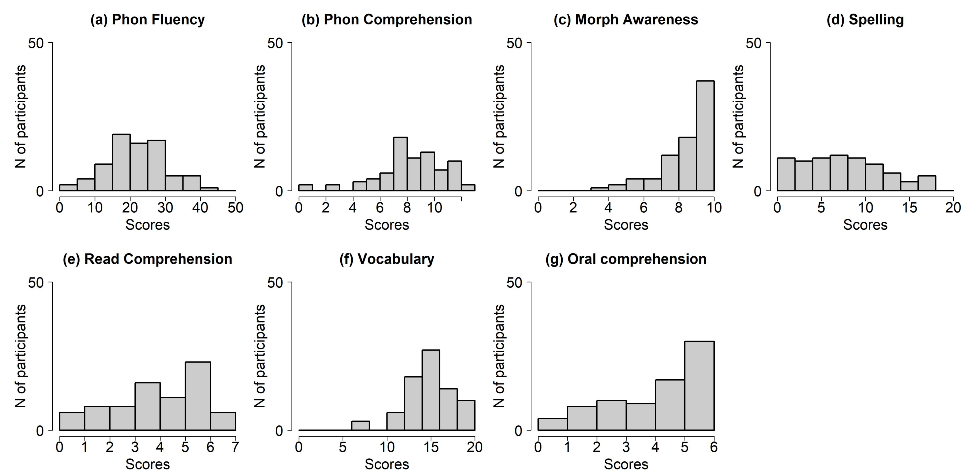

Figure 2.2: Participants' score distributions for each proficiency subtest.

As illustrated in Section 2.3 (Statistical analysis), we assessed the impact of each subtest in a separate model to avoid excessive collinearity. Every single sub-test determines an overall increase in goodness of fit, $\chi^2[6] = 15.79 - 35.02$, all p values <.014—unsurprisingly, RTs are better accounted for when participants' proficiency is taken into account. Possibly more surprising is the fact that each individual cognitive aspect of language proficiency, as tracked by the individual scores in the sub-tests, is able to guarantee this better account.

This improved goodness of fit does not necessarily comes from morphological priming modulation though; proficiency might just explain overall response speed, or general sensitivity to priming. We thus assessed which proficiency score, if any, interacted specifically with prime relatedness and morphological condition. It turned out that only phonemic fluency does so with solid statistics, $F[2, 8564] = 5.83, p < .002$, while morphological awareness is just below the significance threshold, $F[2, 8561.4] = 3.04, p < .047$. The nature of the priming modulation is illustrated through the model–based estimates in Figure 2.3; while transparent priming is solid and consistent across the whole phonemic fluency spectrum, opaque and orthographic priming shrink with growing fluency (although the former condition shows more solid statistics, $\beta = .006, t[8559] = 3.41, p = .001$ and $\beta = .003, t[8569] = 1.66, p = .09$, respectively).

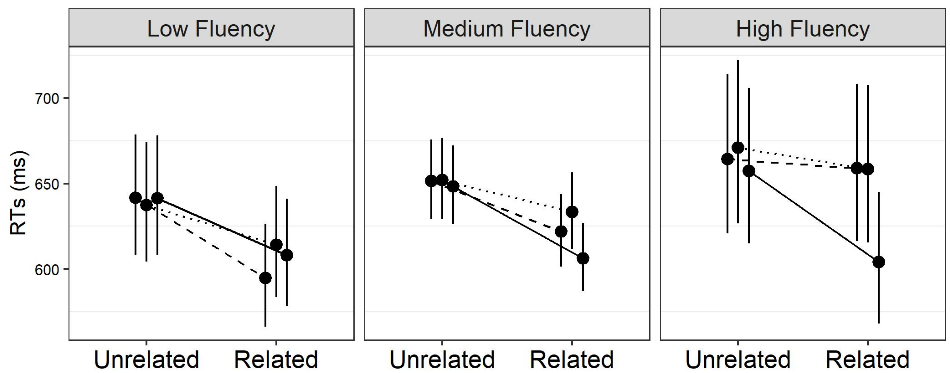

Figure 2.3: Model–based estimates of response times (RTs) relative to the interaction between prime relatedness, morphological type, and phonemic fluency in L2. The solid, dashed and dotted lines represent the transparent, opaque and orthographic conditions, respectively. Effects are estimated at the 5th, 50th (median) and 95th percentile of the phonemic fluency distribution. Error bars are 95% confidence intervals.

The pattern for morphological awareness is very similar, as illustrated in Figure 2.4. Here again, transparent priming remains strong across the board, while opaque priming shrinks with growing awareness, $\beta = .024, t[8559] = 2.36, p = .01$ (form priming doesn't seem to differ from transparent priming in this case, $\beta = .006, t[8560] = .59, p = .61$).

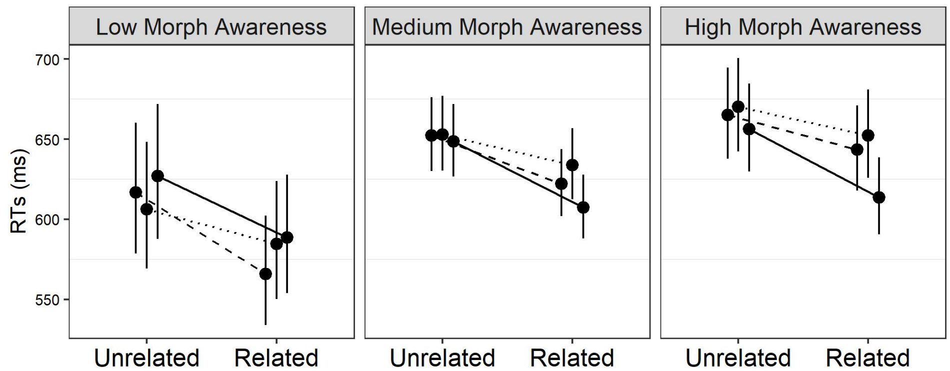

Figure 2.4: Model–based estimates of response times (RTs) relative to the interaction between prime relatedness, morphological type, and morphological awareness in L2. The solid, dashed and dotted lines represent the transparent, opaque and orthographic conditions, respectively. Effects are estimated at the 5th, 50th (median) and 95th percentile of the morphological awareness distribution. Error bars are 95% confidence intervals.

Finally, we explored whether proficiency in English also predicted priming pattern in Italian, we ran the same statistical models assessing which proficiency score interacted with morphological priming in L1. There was no effect of any proficiency subtest on L1 priming, (all $t < 1$, all $p > 1$), suggesting that proficiency in L2 does not impact L1 visual word recognition.

AoA analysis

The scores collected through the AoA questionnaire are distributed as illustrated in Figure 2.5. AoA proper, panel (a), is reasonably well distributed, with a peak around the age of 6, which is schooling age in Italy. This goes together with the fact that most of our participants learned English at school, panel (c). Interestingly, we also happened to recruit several participant with $AoA < 6$, who learned English at home. Quite notable are the nicely symmetrical distributions for daily use of English, panel (b), and self–rated proficiency, panel (e). Finally, most of our participants did not grow up in a multilingual environment, panel (d), but ended up speaking at least another language in addition to Italian and English, panel (f). Importantly, AoA proper correlates $-.15$ with daily usage, .04 with self rated proficiency, and never stronger than .23 with the objective proficiency scores; this means that we can assess the effect of AoA independently of other variables.

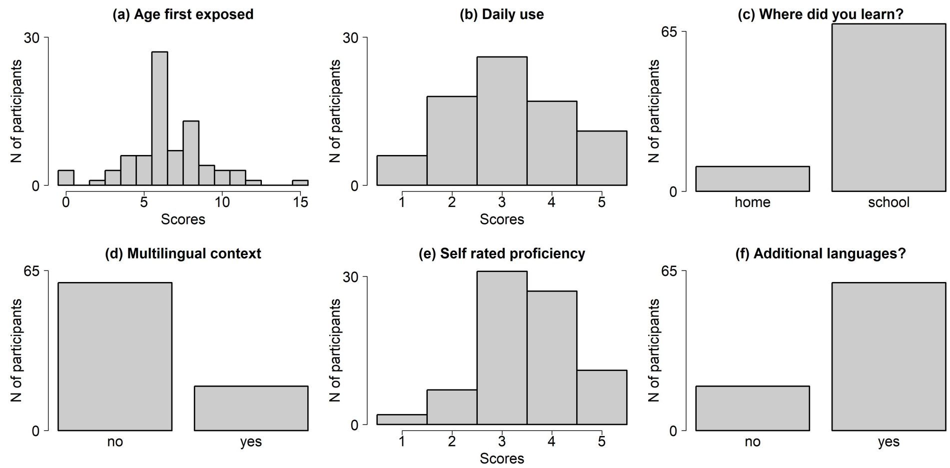

Figure 2.5: Scores distributions in the AoA questionnaire.

We followed the same modelling approach as for proficiency, that is, we first assessed whether AoA proper allows an overall better account of RTs. This does not seem to be the case, $\chi^2[6] = 9.39, p = .15$. In line with this, there is no AoA modulation of the priming patter, $F[2, 8559.1] = 1.94, p = .14$.

Among the other scores that we collected via the AoA questionnaire, only daily usage and self–rated proficiency improve the quality of the model predictions, $\chi^2[6] = 18.42, p = .005$ and $\chi^2[6] = 15.81, p = .01$, respectively; and only the latter yields a significant interaction with prime *relatedness* and *morphType*, $F[2, 8562.4] = 3.38, p = .03$, in line with the results for the objective proficiency scores. The remaining three variables (speaking a third language, learning L2 at school vs. home, and learning L2 in a multilingual environment) do not affect RTs, all $\chi^2[6] < 5.81, all p > .44$.

OSC analysis

As stated in the Introduction, we also wanted to assess the role of Orthography–to–Semantics Consistency (OSC) in L2, and particularly whether this variable affects morphological priming. Because OSC typically co–varies with morphological transparency (Marelli et al., 2015), we first checked whether this was the case in our set of stimuli too; and indeed it was (see figure 2.6;

$F[2, 144] = 21.02, p < .001$). We thus excluded *morphological type* from the modelling that involved OSC.

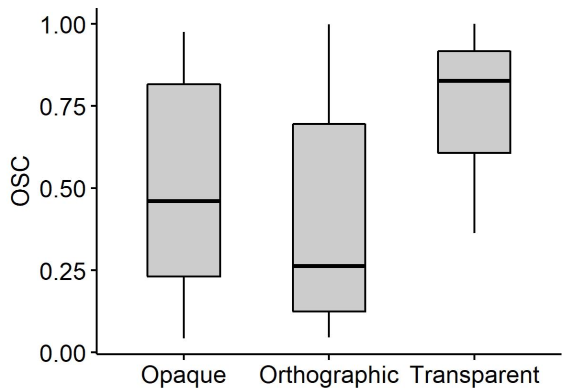

Figure 2.6: OSC distribution for the transparent, opaque and orthographic English target stems.

OSC does indeed modulate morphological priming in L2, also in interaction with phonemic fluency, $F[1, 8579.7] = 7.35, p = .006$. Figure 2.7 illustrates this interaction; priming keeps strong independently of phonemic fluency when OSC is high, but shrinks towards zero with growing phonemic fluency when OSC is low. Given that high OSC characterizes target words in the transparent condition and low OSC marks target words in the opaque and orthographic condition (see 2.6), these results essentially confirms those described above–transparent priming (i.e., priming at high OSC) is independent of phonemic fluency, while opaque and orthographic priming (i.e., priming at low OSC) decreases with increasing fluency.

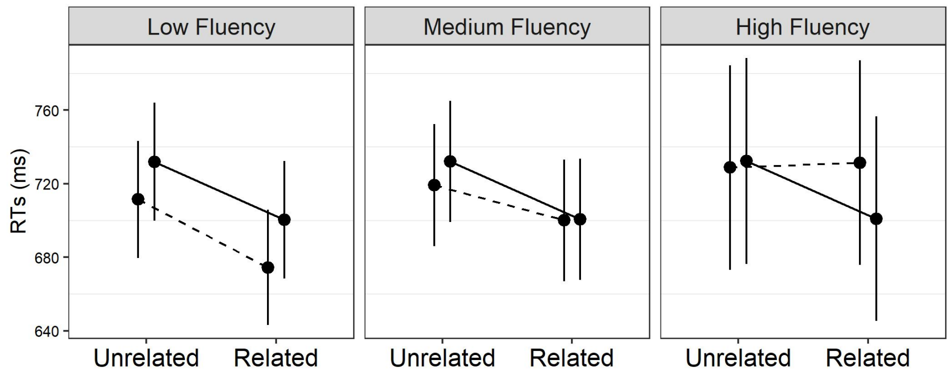

Figure 2.7: Model–based estimates of response times (RTs) relative to the interaction between prime relatedness, OSC, and phonemic fluency in L2. Effects are estimated at the 5th, 50th (median) and 95th percentile of the phonemic fluency distribution, and at the 20th (dashed line) and 80th percentile (solid line) of the OSC distribution. Error bars are 95% confidence intervals.

2.5 Discussion

In this study, we show that orthographic and morphological priming differs in L1 and L2. In L1, we replicated the widely attested pattern whereby the recognition of a target word is facilitated by the prior presentation of a semantically transparent (e.g., dealer-DEAL) or semantically opaque (e.g., corner-CORN) prime, but not by a non–morphological, orthographic prime (e.g., public-PUB). In L2, genuine and opaque derivations provide facilitation, similarly to L1. However, form primes also provide facilitation, contrary to the native language. We also found that genuine morphological derivations yield more priming than pseudo–derived primes, both in L1 and L2.

Critically, we discovered that this group–level pattern in L2 is importantly modulated by L2 proficiency in each individual reader, as tracked by phonemic fluency and, with somewhat weaker statistics, morphological awareness. While transparent priming remains strong in a second language, facilitation in the opaque (and to a lesser extent, form) condition decreases with increasing proficiency. We find no evidence for any other proficiency metric to be specifically related to morphological priming. Age of Acquisition also seems to play little or no role.

Finally, we observed that Orthography–to–Semantic Consistency (OSC) affects lexical decision in a second language, extending the growing body of evidence for OSC effects in L1 (Marelli et al., 2015; Amenta et al., 2017). Interestingly, we found that OSC also interacts with indiviudal proficiency, and is able to provide a nice account of morphological priming, similarly to the classic distinction between transparent, opaque and orthographic primes (Amenta et al., submitted).

At the group level, our results seem to support the idea that morphological processing during visual word identification differs in L1 and L2. This is in line with what suggested by, among others, Clahsen and Felser (2006), and seems to contradict Diependaele et al. (2011). However, if one looks closely to the priming pattern observed in the individual morphological conditions, inconsistency is less than it would seem. Genuine derivations provide solid facilitation in both L1 and L2, which is consistent across studies. In our experiment, opaque primes tend to yield less facilitation than transparent derivations in L2 (and also do so in L1)—again, this is comparable to what Diependaele et al. (2011) observed. Form priming is where L1 and L2 differs the most, with a clear null effect in L1, in line with a large body of literature (e.g., Rastle et al., 2004; Longtin et al., 2003; Forster et al., 1987), whereas orthographic similarity clearly brings facilitation in L2. Again, this pattern of results mirrors the findings of Diependaele et al. (2011).

The main difference between Diependaele et al. (2011) and what we observe here seems therefore to lie in the comparison between form and opaque priming in L2—form priming is somewhat smaller in our experiment, which makes it easier to differentiate from opaque priming, while these two conditions were statistically indistinguishable in Diependaele et al. (2011). So, despite a different outcome in the three–way interaction between *prime relatedness*, *morphological type* and *language*, the difference between our data and Diependaele et al. (2011) only consists in a somewhat weaker L2 form priming in the present experiment.

Even if smaller than in previous studies, L2 form priming is clearly solid here—orthographically similar words facilitate each other during lexical identification in a second language. In line with previous reports, this contrasts with a very clear null effect in L1. This pattern nicely mirrors the Prime Lexicality Effect (PLE) as observed in native speakers (e.g., Forster and Veres, 1998)—nonwords provide strong facilitation to orthographically related targets in masked priming (e.g. *contrapt–CONTRAST*), but this facilitation is reduced, and sometimes even turns into inhibition (Davis and

Lupker, 2006b), when the prime is a real word (e.g., *contract–CONTRAST*). This phenomenon is classically interpreted in terms of lexical competition—both *contrapt* and *contract* would provide the same amount of facilitation at the letter coding level, but the established lexical representation for *contract* would then compete with that of the target word, thus generating lexical inhibition that would offset the sub–lexical priming.

From this perspective, our data may suggest that L2 words behave similarly to nonwords in L1. L2 lexical representations might not be very well established (or not even present). Therefore, lexical competition would be reduced (or absent), thus providing no offset to the sub–lexical facilitation brought about by a form–related prime. This nicely connects to the growing literature on novel word learning (e.g., Gaskell and Dumay, 2003b; Tamminen and Gaskell, 2008; Tamminen et al., 2010; Walker et al., 2019; Sobczak and Gaskell, 2019; Beyersmann et al., 2015b), where lexical competition is often taken a the primary diagnostic for a fully consolidated lexical memory.

This hypothesis goes nicely with the data emerging from the proficiency analysis. Although the statistics are not entirely convincing, there is at least a trend showing that L2 form priming shrinks with growing phonemic fluency. This may suggest that L2 lexical memories become more fully established with growing proficiency, such that lexical competition progressively shows off as readers gain command over a second language. This is inline with theoretical accounts of word learning in L2 as well (Lindsay and Gaskell, 2010).

The impact of individual readers' proficiency—as tracked by phonemic fluency and, to a lesser extent, morphological awareness–extends well beyond the orthographic condition, and critically qualifies the entire pattern of form and morphological priming. The effect of transparent primes is not modulated by proficiency, while priming in the opaque condition clearly shrinks as proficiency grows. As mentioned in the previous paragraph, form priming shows the same trend. Overall, these effects paint a picture whereby visual word identification, at least as revealed by masked priming, is dominated by mere form similarity when readers do not have great command over their L2; as Figure 2.3 clearly shows, transparent, opaque and orthographic primes have hardly distinguishable effects on their (pseudo–)stems at low level of proficiency. Similarly to what we suggested above, this points to a rather weak lexical network in low–proficiency L2 readers. At this stage, the lexicon may perhaps be characterized more as a collection of unconsolidated word memories than as a network that licenses the lexical dynamics typical of L1.

There are two facets in this overall pattern. First, transparent morphological priming only sets apart from an orthographic baseline as readers gain familiarity with L2. This suggests that, even when morphological relationships are fully transparent, readers need some level of proficiency before they encode morphological ties in a way that masked priming can reveal. This conclusion contrasts with the general view that L2 processing of transparent derivations is similar to L1 (e.g., Diependaele et al., 2011); this seems to be true only at high levels of proficiency.

A second, perhaps even more interesting aspect of these results is that opaque priming clusters with form priming, rather than with the transparent effect. This would suggest that pseudo–derivations (e.g., corner) and genuine derivations (e.g., dealer) track qualitatively different morphological and lexical dynamics in L2, contrary to what happens in L1 (Davis and Rastle, 2010; Rastle et al., 2004; Diependaele et al., 2011; Feldman et al., 2009).

In other words, morpho–orthographic chunking did not emerge clearly anywhere along the proficiency spectrum that we tracked in this study. One reason for this may be that such chunking only occurs at even higher level of proficiency, which we didn't capture in our study. Some of the participants performed at ceiling in our proficiency tests, which would suggest we went as high as possible in the proficiency scale; but of course, our tasks may just not be difficult enough to truly distinguish L2 speakers/readers at the very high end of this scale. Another possibility is that morpho–orthographic chunking is related to the L1 status per se, independently of proficiency. The fact that we didn't find any effect of Age of Acquisition, however, seems to speak against this hypothesis—very early L2 speakers, which we do have in our sample, would presumably show a native–like pattern if L1 status was the driving force here.

A possible criticism to our study may be that a within-participants design doesn't allow for direct comparisons across languages. Many studies overcome such limitation by adopting a between participants design whereby L1 and L2 participants are tested on the same items. Although we purposely chose to test the same individuals in their L1 and L2 in order to directly compare patterns of morphological activation within the same person, we acknowledge that Italian and English are difficult to compare to each other because of their inherently linguistic differences. Differences in the consistency of the relationship between orthography and phonology may determine the way the reader extracts morphological information. Perhaps these differences may explain why our Italian participants did not rely on the same metalinguistic knowledge previously shown in the past (e.g., spelling, Andrews and Lo, 2013). Consistent with this interpretation, English speakers might rely on different linguistic competences in their L1 than Italian speakers reading in English because of the phonological transparency of the language (Paulesu et al., 2000). To specifically test this possibility, as also one of the reviewers of this manuscript suggested, participants' native linguistic proficiency should have been linked to priming in L1 as well. Although only anecdotally, we confirm that we have tried and repeatedly failed in the past to obtain such measures because Italians simply perform at ceiling in spelling tasks, making such measurements to all intents and purposes completely uninformative. Thus, future research is clearly needed to address this issue.

Overall, these data may suggest that morphological priming is a core metric to track the development of a fully–fledged (i.e., L1–like) morpho–lexical system (see Figure 2.8). Early on, form similarity would be the only driving force, with no morpho–lexical distinction between orthographic, opaque and transparent priming. As word representations become more and more consolidated, lexical competition arises, driving down purely orthographic priming. In this early phase, morphological detectors would only be able to capture transparent ties; therefore, while genuine morphological facilitation remains high, opaque primes behave just like mere orthographic primes, with the effect shrinking to zero. This is the stage that we would have captured here. It is only at a late stage (that we may have failed to capture here) that morphological representations become insensitive to semantics, thus yielding facilitation also for opaque pairs.

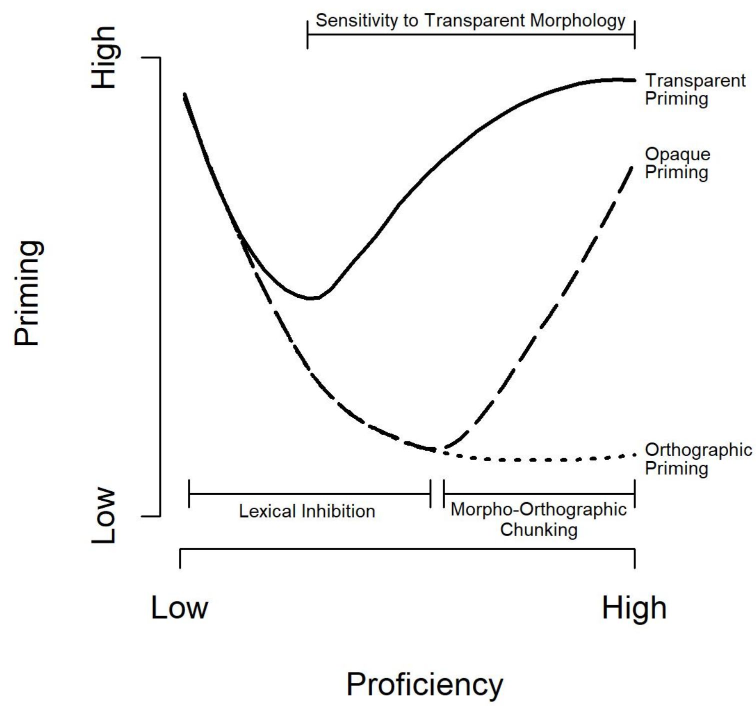

Figure 2.8: Theoretical description of the development of a fully–fledged (i.e., L1–like) morpho–lexical system in L2.

By interpreting morpho–orthographic priming as a sign of a fully matured visual word identification system, this model fits nicely with data on developing readers. Of course, acquiring literacy in one's first language isn't the same as learning to read in a second language. Yet, the evidence available on morphological priming in children does show that opaque facilitation typically emerges at a later stage of reading acquisition (Beyersmann et al., 2012), if it is even detectable at all in children (Schiff et al., 2012).

The critical importance of each reader's specific proficiency profile in these data relays to (i) the mounting evidence on the effect of individual variability in L1 (e.g., Andrews and Hersch, 2010; Burt and Tate, 2002; Andrews and Lo, 2013; Beyersmann et al., 2015a); and (ii) developmental data that also points to some changes over the course of adolescence in the way letter strings are processed (Dawson et al., 2017; Beyersmann et al., 2015b). Evidence is growing that experience with the written language (and, possibly as a consequence, better and more refined orthographic representations/processing) changes substantially the dynamics behind visual word identification. A precise characterization of the cognitive profile of each individual reader and a careful consideration of her/his experience with visual words is more and more fundamental in the field, and seems to critically qualify most the phenomena we have been studying for years as emerging in undistinguished groups of participants.

Among the many proficiency indexes we considered here, phonemic fluency and, to a lesser extent, morphological awareness turned out to be the best metrics to account for morphological priming. It is not obvious why these particular tests performed better than the others. Morphological awareness may be not too surprising. Many studies investigated the role of morphological awareness in reading acquisition. Such studies have shown that morphological awareness predicts reading comprehension in children (Carlisle, 2003, 2004; Deacon and Kirby, 2004; McBride-Chang et al., 2003), and is used to overcome reading difficulties during learning to read (Arnbak and Elbro, 2000; Casalis et al., 2004). However, this test requires explicit judgments, and thus participants had to access their morphological knowledge in full awareness. Apparently, this meta—cognitive performance builds, at least in part, on the same cognitive processes that enables identification of embedded word stimuli during visual word identification. Phonemic fluency is more general tasks, typically associated with lexical knowledge, behaviour control and attention (Ruff et al., 1997; Henry and Crawford, 2004; Whiteside et al., 2016). Despite the specific psychological construct behind these tests is not entirely clear, it is not difficult to imagine that a well structured network of lexical memories is critical to quickly retrieve a good number of words based on a phonological cue. This well structured lexical network, at least in its visual form, surely underlies morphological and orthographic priming.

Finally, we demonstrated for the first time an effect of Orthography–to–Semantic Consistency (OSC; Marelli et al., 2015) in L2, showing that readers capture fine-grain, probabilistic ties in form–to–meaning mapping also outside of their native language. This result licenses an intriguing new perspective onto the learning of a second language, which may be related (among other things, of course) to the appreciation of the statistical structure behind the relationship between form and meaning in the novel lexicon (Forster and Veres, 1998; Castles et al., 2007; Perfetti and Hart, 2002; Perfetti, 2007; Andrews and Hersch, 2010; Andrews and Lo, 2013; Hersch and Andrews, 2012). Interestingly, we found that sensitivity to OSC interacts with proficiency, which further reinforces this suggestion—the more one gains command over L2, the more sensitive it becomes to probabilistic relationships between orthography and semantics.

Although this was outside the main scope of the study, we also note that OSC provides a nice account for the priming pattern without taking into consideration the classic categorical distinction between transparent, opaque and orthographic primes Marelli et al. (2015). OSC correlates with these categories, and when we left it alone in the statistical model to account for priming, it proved able to do so. This may suggest a different interpretation of morphological priming, which would not depend on the relationship between prime and target themselves, but on transparency in form–to–meaning mapping in the lexical region where the target and the prime live (Amenta et al., submitted).

Chapter 3

Experiment 2 - Electrophysiological index of visual discrimination in the left occipito-temporal cortex shows rapid lexical integration

3.1 Abstract

The importance of consolidation in integrating new memories has received much recent attention. One important feature of integration is that new knowledge can interact with existing knowledge directly. However how and when a novel word is integrated and thus becomes functional is a critical question. The Complementary Learning Systems (CLS) framework argues that lexical competition is a symptom of lexical integration, though some studies show no competition. A possible source of clarification might involve investigation of different learning experiences. In this study we wanted to study two different learning methods, i.e., uninstructed versus instructed, while comparing the outcomes of a lexical competition task with an objective neural marker of lexical integration. We recorded scalp electrophysiological (EEG) brain responses and showed that novel words elicited a change in amplitude reflecting lexical integration without lexical competition. Cru-

cially, this response was stable after one week, suggesting that novel word memories were successfully transferred into longer–term system.

3.2 Introduction

Contrary to what one might intuitively believe, people learn novel words throughout their adulthood at a very high pace (about a thousand words per year; e.g., Nation and Waring, 1997; Keuleers et al., 2015). Most of this learning comes from reading (Nagy et al., 1985), which may appear as a simple and swift process, but in fact requires the creation of strong links between a word's orthographic, phonological and semantic information (Dehaene and Cohen, 2007; Taylor et al., 2019; Perfetti and Hart, 2002).

Learning novel words involves their integration within the mental lexicon as well. A behavioural signature of this integration is the interaction of the newly learned word with previously known memories, a process usually referred to as *lexical engagement* (Leach and Samuel, 2007). This interaction usually takes the form of a slowdown in the recognition of familiar words that are similar to the novel one. For example, in a seminal study Gaskell and Dumay (2003b) taught their participants novel words like "cathedruke" via a phoneme monitoring task. Lexical decision times on real words that were similar to the novel items (e.g., "cathedral") were monitored over five days of learning, and turned out to increase after the third day. This was interpreted in terms of lexical competition (McClelland and Rumelhart, 1981; Coltheart et al., 2001). The novel item "cathedruke" entered the complex network of representations that instantiates anyone's lexical knowledge, and therefore started to compete with similar, previously existing entries, like "cathedral".

This interpretation resonates nicely with the *Complementary Learning Systems* framework(CLS; McClelland et al., 1995; Davis and Gaskell, 2009). Davis and Gaskell (2009) argued that new words are initially encoded via a separate hippocampal route, and then later they get incorporated into existing long–term neocortical memories, mainly via hippocampal replay during sleep (Dumay and Gaskell, 2012). According to this account, the hippocampal route rapidly accommodates new words using sparse representations that mediate the mapping from auditory areas to lexical and semantic areas. These mediators are used during the early stages of word learning, until direct cortical mappings can be gradually built via consolidation. Early during learning, the hippocampal route does not allow newly learned words to be activated quickly enough to compete against the much swifter activation of the known, neocortically–coded words. At a later stage, memories consolidate via gradual strengthening of direct cortical links, leading to faster activation (and, therefore, more inhibition).

However, newly learned words do not always get to inhibit existing lexical memories. For example, Henderson and James (2018) exposed children to novel words in meaningful story contexts (e.g., "crocodol" in a story themed around "Lucy's trips to an alien zoo"), to investigate how learning interacts with contextual diversity (i.e., finding the word repeatedly within the same story vs. across different stories). Right after learning, novel words that had been encountered repeatedly in one single context yielded strong facilitation, while a null effect was found for multiple stories. The subsequent day, although non significant, a weak lexical competition effect arose for novel words belonging to the repetitive story condition, while multiple stories showed a trend towards facilitation. Sobczak and Gaskell (2019) tested the learning of novel words via repetition of syllable sequences, and showed that lexical memories can be acquired through this implicit task, but learning is generally weak and, more importantly for the issue at hand, does not yield signs of lexical competition. Hawkins (2015) in the third study of her doctoral thesis employed a learning paradigm where novel spoken words were associated to pictures referring to novel objects. She tested lexicalization via pause detection after learning and following 24 hours. Hawkins (2015) observed no lexical competition neither following 24 hours nor immediately after learning, with a trend towards facilitation after 24hr. Nonetheless, consolidation enhanced explicit recognition sensitivity of learned novel words as measured via recognition task 24 hours later. In a later follow-up of this study conducted months later, she found again absence of lexical competition, despite representations of the novel words were still retained as showed by the explicit recognition task. Taken together, these findings therefore suggest that novel words benefited from explicit memory, but in absence of the new words' entry into lexical competition. Finally, Gaskell and Dumay (2003a) found immediate competition for items learned over 60 presentations, which remained stable over time, while items presented only 12 times yielded facilitation early on, and a null effect after 7 days. These data suggest that only high frequency of exposures promotes lexical competition, and puts into the question the precise timeline of hippocampal coding and consolidation (e.g., competition emerged immediately in this study).

Importantly, these studies report examples of facilitation between known words and novel lexical material, not just null effects. The CLS framework would predict no interaction between pre–existing knowledge and unconsolidated, hippocampal memories, and therefore it is difficult to hold that novel words were simply never transferred into the neocortex in these experiments. There were signs of interaction between novel and previously established words, which does show lexical engagement and, arguably, consolidation; however, these interactions were facilitatory, instead of inhibitory.

In summary, although most studies show competition between novel words and previously established lexical knowledge under some circumstances, data are not entirely consistent as (i) these circumstances change across experiment; and (ii) there are reports of null or facilitatory effects, which are difficult to reconcile with the CLS framework.

A possible account of these facilitatory effects has emerged recently in neuroscientific and computational work (McClelland, 2013; Tse et al., 2007, 2011; van Kesteren et al., 2010, 2013). These studies suggest that information can be integrated in neocortical areas soon after learning and without any interference when the novel information is consistent with prior knowledge. For example, in Tse et al. (2007) study, rats were trained to map flavors with locations in an unfamiliar spatial environment. Crucially, when rats were trained to map a novel flavor to one of these trained, well-known location, they showed strong evidence of immediate learning, with no sign of competition between novel and old knowledge. Two days after learning, rats received extensive hippocampal lesions. After recovery from surgery, further test trials revealed that the lesioned animals retained both the original and the new flavor–place associations, supporting the conclusion that new information that were consistent with prior knowledge can be assimilated rapidly in the neocortex, without interfering with previously learned information.

It is unclear how these results may transfer into the word learning domain. In a sense, the novel material that is most typically used in word learning studies is indeed consistent with existing lexical information—words to be learned are most often neighbours of existing ones, similarly to how the original and the novel locations were close in space in Tse et al. (2007). Yet, competition emerges most often. Moreover, experiments that reported facilitation vs. inhibition between novel and pre–existing words do not seem to be based on learning material that differ in any consistent way.

Perhaps one way to conceive consistency in the context of word learning is referring to lexical networks. In fact, the very process of consolidation points to how a novel word is integrated into these networks, and therefore it is quite likely that their structure plays an important role. Fitting nicely with this perspective, a larger vocabulary was found to allow smoother integration of novel words (James et al., 2017), and it is possible that words with high consistency between phonological, orthographic and semantic networks can be added to the lexicon without interference (Davis et al., 2009).

Moreover, orthographic and phonological neighbourhoods interact constantly in visual and auditory word recognition Ziegler et al. (2003). In fact, evidence on the lexical identification of well established words shows facilitatory effects in visual word recognition for words living in a small phonological and orthographic neighborhood, but inhibitory effects for words having numerous phonological neighbours and only a few orthographic neighbours Grainger et al. (2005). The use of words from different parts of the lexical space may have thus struck a balance between facilitation and inhibition, within any individual study; or generated opposite effects, across experiments. This factor that has not received much attention in previous research. From this point of view, the language in which the present study was conducted, Italian, is rather privileged—its nearly perfect orthography–to–phonology consistency aligns the visual and acoustic lexicons, thus leaving less room for these neighbourhood interactions to cloud the effects of interest.

Another key variable in the literature described above that may have contributed to the emergence of contrasting results is the way novel lexical items were taught to participants. Learning routines were as diverse as phoneme monitoring (Gaskell and Dumay, 2003b), read aloud (Tamminen et al., 2012) or repetition tasks (Szmalec et al., 2012), word encounter within connected stories (Henderson and James, 2018), and Hebb repetition paradigm. The Hebb paradigm involves gradual learning via serial repetition of items (e.g., digits), and is considered an implicit paradigm because learning occurs without awareness (Sobczak and Gaskell, 2019). Attention has been driven only recently to this issue. For example, Szmalec et al. (2012) found that novel words learned via the Hebb repetition task are integrated into the lexicon independently of sleep, contrary to previous results based on less implicit learning routines. Sobczak and Gaskell (2019) set out to test this hypothesis by comparing Hebb repetition with phoneme monitoring within the same study, and found that the former task did not promote lexical integration at all, neither prior to nor after sleep. Results are therefore quite inconsistent.

More generally, it would be interesting to better understand the implications for lexical consolidation of a more explicit, instructed routine, where participants are directly pointed to the material to be learned (e.g., Tamminen et al., 2012; Bowers et al., 2005) vs. a more implicit, uninstructed routine, where participants have to make an active effort and figure out themselves the novel words to be learned (Grainger et al., 2012). In addition to the relevance of this question for theories of word learning, this may offer some insight onto the potential consequences of the different ways in which we learn words in daily life, e.g., via explicit study of word lists, as it often happens when we approach a second language vs. incidental, uninstructed encounters with unknown words, e.g., while reading or during conversation.

A recent development on the methodological front may also be of great help in clarifying some of the issues highlighted above. The paradigm of Fast Periodic Visual Stimulation (FPVS) has long been confined to the study of low–level visual processes and attention (see the long–standing work on Steady–State Visually Evoked Potentials, SSVEPs; Regan, 1968, 1989; Norcia et al., 2015a), but it has recently been used to measure visual discrimination responses to complex visual stimuli such as faces (Rossion and Boremanse, 2011) and words (Lochy et al., 2016). In this paradigm, participants are presented with visual stimuli in a periodic fashion, typically at a fast rate (e.g., 10 Hz). Stimuli come structured in sequences with an "oddball" design—every X instances of a given type (e.g., 4 consonant strings), an item of a different kind appears (e.g., a word). This presentation triggers a periodic EEG response at the base presentation rate (10 Hz, in the example above) and, crucially, also at the oddball presentation rate ($10/5 = 2$ Hz, in the example) if the brain distinguishes between the two kinds of stimuli. This paradigm has been used to show sensitivity to visual words in pre-literate children (Lochy et al., 2016), and in adults (Lochy et al., 2015). Critically, it is not based on facilitatory or inhibitory effects; actually, it is not based on behaviour at all, as participants are typically engaged into a totally unrelated task, like monitoring the color of a fixation cross while the stimuli change in the background. This neural marker may complement the classic behavioural paradigm, and affords an opportunity to study lexical integration by looking directly at how the brain reacts to the newly learned words, rather than exploring their interaction with pre-existing knowledge.

To sum up, novel word learning is based on the integration of the new lexical material into a complex network of pre-existing memories. We learnt a great deal about this process, but data are not entirely uncontroversial, particularly for what concerns the timing of lexical consolidation and the nature of the interaction between new and old knowledge. Also, it is not clear how these mechanisms interact with the learning routine, in particular whether teaching is instructed, explicit vs. uninstructed, based on an active effort from the learner to find out the new information to be acquired. This is the question we tackled in this study, by integrating the classic, behavioural word learning paradigm with a novel electrophysiological technique, Fast Periodic Visual Presentation (FPVS), that allow us to assess more directly the development of novel lexical representations in the brain.

3.3 Methods

Participants

Thirty Italian native speakers were recruited for the experiment (13 males, mean age = 24.4 years, range = 20-32). They had normal or corrected-to-normal vision, and a clean history of neurological impairment or learning disabilities. Participants were unaware of the goal of the study and provided their informed consent as approved by the SISSA Ethical Committee. All participants in this study were classified as right-handed after completing the Edinburgh Handedness Inventory (Oldfield, 1971). Participants recevied 50 Euros in exchange for their time.

Overall timeline of the study

Participants completed the experiment in three sessions. On Day 0, they first learned the novel words following either the instructed or the uninstructed routine described below, according to the experimental condition they were randomly assigned to. Learning was then assessed immediately after, via three tasks: Lexical engagement, explicit recognition and FPVS. Participants came back to the lab on the following day (Day 1), and underwent the same set of tasks, which they also did six days later (Day 7). We report in what follows the details on the stimuli and the procedure for these tasks.

Stimuli

The core set of novel words was comprised of 100 stimuli, which formed two perfectly parallel sets of 50 items. Each participant took up one of these sets during the learning phase, while the other was used as a control, untrained group of nonwords. The learning list was rotated over participants. Each set of 50 novel words included 30 items that were generated by changing one internal letter to an existing word (e.g., *collepio*, from *collegio*, college). We refer to these stimuli as **neighbour** novel words. Importantly, each base word had at most one substitution, transposition, deletion or addition neighbour. The other 20 items to be learned did not have any lexical neighbour, instead; their edit distance from any existing word was on average 8 or higher. We refer to this set of stimuli as **orphan** novel words. All novel words were pronounceable and fully complied to Italian phonotactic rules. Stimuli statistics are shown in table 3.1, which also shows the nice matching between the two parallel lists. The full set of base words and novel words to be learned is reported in the Appendix B.

Table 3.1: Descriptive statistics of the to-be-learned novel words

	1Q	mean	median	3Q
length in letters	6	7.17	7	8
OLD20	1.35	2.19	2.05	3
mean bigram frequency (log10)	6.35	6.79	6.80	7.12

Word Learning Routines

Novel words were learned via one of two alternative routines, each of which was administered to half of the participants. The **instructed learning routine** was very similar to what was used in previous word learning studies (Bowers et al., 2005; Tamminen et al., 2012; Walker et al., 2019; Merkx et al., 2011). Participants were shown each new word one at a time, and were asked to try to remember it and type it back on the computer. If their response was correct, the novel word turned green; if instead the participants made an error, it turned red. Each new word was presented 9 times, for a total of 450 trials presented in 6 blocks. We call this instructed learning because, similarly to what happens most often in formal teaching contexts (e.g., school), learners are presented with the items to be acquired explicitly, and do not have to figure out themselves the material to be learnt. Contrary to this, the **uninstructed learning routine** does require participants to figure out themselves what is to be learnt, i.e., what are the novel words to be acquired. The procedure is adapted from a previous study (Grainger et al., 2012), and consists of a series of lexical decisions where participants try to distinguish novel words from foils. Like the previous learning routine, participants receive the same type of feedback based on their performance.

Novel words are repeated throughout the learning session for nine times, so as to make the total number of encounters with the new items equal across learning routines, whereas foils are presented only once during the experiment. The uninstructed learning routine is organized in 15 blocks. Each block consisted of 70 trials in total: 30 novel word trials (10 items, each repeated three times) and 40 random unique foils.

Explicit Recall

Explicit memory of the novel words, both orphans and neighbours, was assessed through a forced choice task, where participants were presented with a letter string and were asked to decide whether it was one of the novel word just learned. This may be taken as a lexical decision task in the novel lexicon. The order of presentation of the items was randomized. Trial started with presentation of a cross (500ms), immediately followed by the presentation of the learned or unlearned novel word. Participants had 2s available to answer. Time between the button press (or time–out) and the presentation of the next stimulus was 1s.

Lexical Engagement

The lexicalization of the novel words was assessed behaviourally through a lexical decision task on the base words. As illustrated in the Introduction, these latter should acquire a close competitor as the novel words consolidate into the lexicon, and this should affect their identification time. Of course, this test could only be applied to the neighbour novel words, which do have a base word, contrary to the orphan items. Participants were thus shown the 60 base words from the two sets, one of which would correspond to the trained words for each participant, while the other worked as a control set. These 60 words were presented in a random order, intermixed with 60 foil pseudo–words (e.g., pelifoga) matched to the base words for length (1Q= 6, mean= 7, median= 7, 3Q= 8). Trial started with presentation of a cross (500ms), immediately followed by the presentation of the word or pseudo–word string. Participants had 2s to answer. Time between the button press (or time–out) and the presentation of the next stimulus was 1s.

EEG testing

We backed up the behavioural assessment of the novel word lexicalization with an additional EEG paradigm paired with a visual oddball paradigm (e.g., Lochy et al., 2016, 2015; Rossion, 2014; Norcia et al., 2015a). Thirty real words, unrelated to the set of stimuli employed elsewhere in the study, were embedded as oddball stimuli in a stream of pseudofont strings. The alphabetic and non-alphabetic stimuli were presented in Courier New font and BACS–2 serif font (Vidal et al., 2017), respectively. All stimuli were presented as images and at a distance of around 70cm with a screen resolution of 1920 x 1080 pixels. Stimuli were presented at a fast rate of 10Hz, i.e., every 100ms, while lexical oddballs occurred periodically every 5 items, thus with a 2Hz frequency, reaching full contrast at 50ms (see figure 3.1). The total stimulation lasted 17min with a 45s break every stimulation sequence. We know from previous work that, under these conditions, the brain shows a periodic response at the word presentation rate (2 Hz), i.e., Steady State Visual Evoked Potentials (SSVEPs, Regan, 1968, 1989; Norcia et al., 2015b), exactly by virtue of the lexicality of these items (Lochy et al., 2016, 2015). We reasoned that, if the novel words had effectively become part of the participants' lexicons, then they would elicit a similar steady–state periodic response, or at least a stronger response than the control set of items that the participants did not learn, and therefore fully maintained their nonword status. We thus arranged trained and untrained novel words within the same paradigm illustrated above, and presented them, in separate runs of course, as oddball stimuli in a stream of pseudofont strings. Neighbour and orphan novel words were tested separately, allowing a chance to assess whether the presence of an existing lexical memory similar to the word to be learned affects lexicalization—which is impossible to see in the classic behavioural tasks, where orphan words just cannot be tested. Thus, the design included 5 conditions overall: real words, trained neighbour words, untrained neighbour words, trained orphan words, and untrained orphan words, all contrasted with pseudofont strings. We followed the same testing protocol as Lochy et al. (2016). Trial runs lasted 60 seconds in the word neighbour and real word conditions, and 40 seconds in the orphan condition. Thus involving the presentation of 30 pseudofont strings and 30 oddball stimuli for real words and novel neighbour condition, and 20 pseudofont strings and 20 oddball stimuli in the orphan condition. Each individual oddball was thus presented 4 times every fifth stimuli in the word and neighbour conditions, which had a set of 30 different items, and 4 times in the orphan neighbour conditions, which had a set of 20 different items. Pseudofont strings were repeated 16 times. Stimuli were presented with a sinusoidal contrast modulation—contrast progressively increased from the background grey to full white and then back to the background grey in 100ms reaching full visibility at 50ms. Each trial consisted of a fixation period, an initial fade period, the main stimulation sequence, and an end fade period. The time course of the stimulation is represented in figure 3.1. The base rate of 10 Hz was selected because it gives the largest SSVEP to luminance changes according to previous studies (Lochy et al., 2016; Norcia et al., 2015b). Each condition was tested in three repeated trials. During the stimulation, participants continuously fixated a central cross, and were instructed to press the spacebar whenever they detected a color change (blue to red and red to blue) which occurred randomly 6 times within each stimulation sequence. No linguistic task was required on the critical stimuli. Color changes occurred independent of word-type manipulation and were included to maintain a constant level of attention throughout the entire experiment.

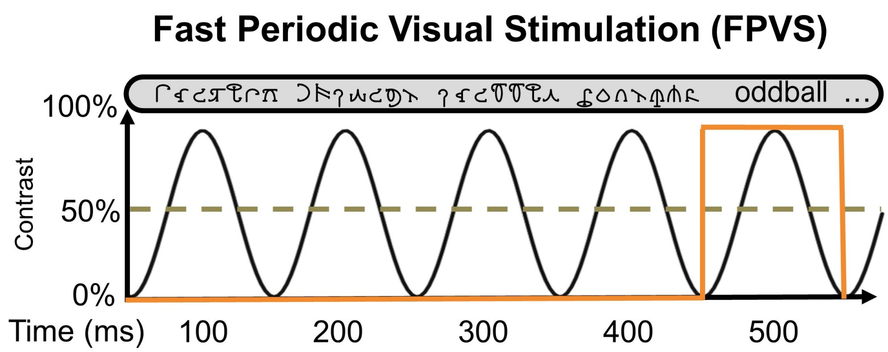

Figure 3.1: The time course of the main stimulation sequence. Stimuli were displayed with sinusoidal contrast modulation from 0% to 100% contrast at 10 Hz during the 60 or 40-sec sequence (black line). Oddball presentation occurred at 2 Hz (orange line) replacing the false fonts presentation every fifth element (orange square). Depending on the condition, oddball was either a novel trained word, or a novel untrained word, or a real word.

EEG acquisition

EEG's data were recorded in a Faraday room where participants seated comfortably at a distance of approximately 70cm from a 27" BenQ XL2720Z monitor in a dimly-lit and sound-attenuated environment. Stimuli were presented by PsychToolbox–3 (Kleiner et al., 2007) on Matlab 2015b (The Mathworks) in a Windows environment. EEG was acquired at 256 Hz using a 128 channel Biosemi Active II system (Biosemi, Amsterdam, Netherlands), with electrodes arranged according to the radial-ABC system location. Two additional electrodes (Common Mode Sense active electrode, CMS; and Driven Right Leg passive electrode, DRL) were used as reference and ground electrodes, respectively. The magnitude of the offset of all electrodes, referenced to the common mode sense, was held below 20 μV.

EEG analysis

Preprocessing: EEG data preprocessing was carried out using EEGLAB version 13 (Delorme and Makeig, 2004), and Matlab 2018a (The Mathworks). After band-pass filtering between 0.1 and 100Hz, noisy channels were manually inspected and replaced using automatic linear interpolation. Note that this approach is highly resistant to artifacts (e.g., blinks), so that we didn't remove anything else. EEG data were re–referenced to the common average and segmented into 60s and 40s trials. These trials lasted differently because reflected the different amount of items in the novel word conditions: neighbour (30 words) and orphan (20 words). Trials were grouped in bins according to the experimental condition to which they belonged.

Frequency domain analysis: We followed the same procedure of Lochy et al. (2015). The three trial repetitions of each condition were averaged in the time domain, for each word type condition (learned neighbour, unlearned neighbour, learned orphan, unlearned orphan, real words) and for each test session (day 0, day 1, day 7) within each individual participant separately to reduce EEG activity that was not phase-locked to the stimulus. A Fast Fourier Transform (FFT) was then applied to those averaged segments, and normalized amplitude spectra were extracted for all channels (square root of the sum of squares of the real and imaginary parts divided by the number of datapoints). The resulting EEG spectra had a high frequency resolution: 1/60 seconds = 0.01667 Hz for neighbour and real words condition and 1/40 seconds = 0.025 for the orphan condition, allowing thus the unambiguous identification of the brain response at the exact frequencies of interest: 10 Hz for the base stimulation rate and 2 Hz and its harmonics for the oddball stimulation rate. At this point, SNR spectra was computed for the whole frequency spectrum as the ratio of the amplitude at each frequency to the average of the 20 surrounding bins (10 on each side) (Liu-Shuang et al., 2016). This was done separately for each individual trial, then SNR spectra of individual participants were averaged within each condition resulting in 15 files per participants, i.e., 5 word type conditions x 3 test sessions.

These files were subsequently merged in two main categories, learned and unlearned, for each test session.

ROI analysis: Electrodes belonging to the pre-defined regions-of-interest (ROI) used by Lochy et al., (2015) (Lochy et al., 2015) were selected and data exported as a unique text file in order to be imported into `R` (R Development Core Team, 2008) for later statistical tests. ROI was located to the occipital area that comprised five electrodes including and around the medial occipital electrode Oz. SNR values of each electrode within the occipital ROI were averaged for each participant. For the response at the 2 Hz oddball stimulation frequency and its harmonics, the primary region of interest comprised an area of five left occipito-temporal sites including and around the electrode PO7. This left occipito-temporal (left-OT) ROI was selected to overlap with the same region that elicited a word selective response (in contrast to nonwords and pseudowords) in the study by Lochy et al. (2015). SNR spectra was initially visually inspected for the *real word* condition averaged over participants and test sessions (figure 3.2). We reasoned that if novel words were lexicalised, then they should have elicited a discrimination response at the oddball frequency and harmonics over the left occipito–temporal area resembling the one of real words inserted within the pseudofont stimulation. To this end, the significance of the response at the oddball frequency and harmonics were tested with a z-test against the baseline signal (SNR=1) in the real word condition and used it to select the same amount of harmonics over all conditions. From this condition, SNR was significantly different from the baseline (SNR=1) at the oddball, first and second harmonics, but not at the third harmonic (8Hz), as clearly evident in figure 3.2. We adopted an even more conservative threshold by selecting only the oddball and first harmonic because showed the heighest SNR, thus were more informative. Based on this result, we decided to select an identical number of harmonics for all conditions and electrodes, that is, to conduct analysis for the oddball (2Hz) and first harmonic (4Hz), averaged together. It is noteworthy that results do not change anyway even if the second harmonic (6Hz) is included in the analysis.

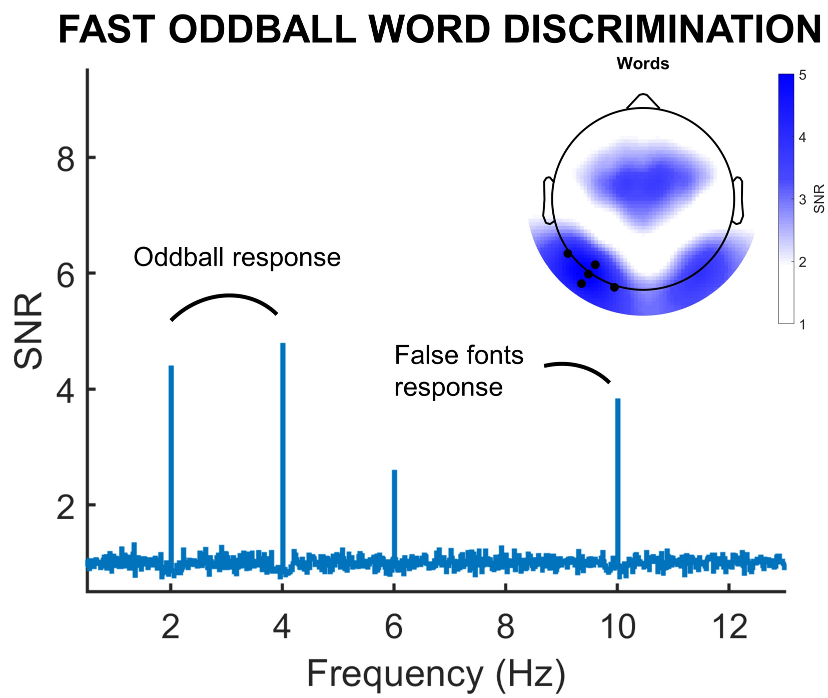

Figure 3.2: Grand-averaged (n=30) SNR EEG spectra and scalp topographies at oddball (2Hz), harmonics (4Hz, 6Hz and 8Hz) and base (10Hz) frequencies for real words condition. Only oddball (2Hz), first (4Hz) and second (6Hz) harmonics were significantly different from baseline (SNR=1).

Statistical approaches

Data were modelled inside the `R` environment (R Development Core Team, 2008). We used mixed effect models in order to account for variability at the individual and item level. Linear mixed models were implemented through the package `lme4` (Bates et al., 2015). Figures were created via `ggpubr` (Kassambara, 2019) and statistical significance was obtained via `anova` and `summary` functions updated by the `lmerTest` package (Kuznetsova et al., 2016).

3.4 Results

Learning phase. *Instructed learning.* Overall mean accuracy across blocks was 97% (SD = 1.5%). Accuracy on the last block was 98.4%, with all individual novel words being higher than 80%. A mixed–effects linear model was fitted to the accuracy scores, with *subjects* and *words* as random effects, and *block* and *word type* (neighbour vs. orphan) as fixed effects. The model showed a strong effect of *block*, $F[1, 6689.7] = 33.6, p < 0.0001$, showing that, although performance was rather good from the beginning (as expected, in a simple read–and–type–back task), it improved during the learning phase (see Figure 3.3). The effect of *word type* was around the significance threshold, $F[1, 673.6] = 2.28, p = 0.13$, with a tendency for neighbour words to be learned better, $\beta = 0.434, z = 1.83, p = .08$. No interaction was found between these variables, $F[1, 669.3] = 0.24, p = 0.62$.

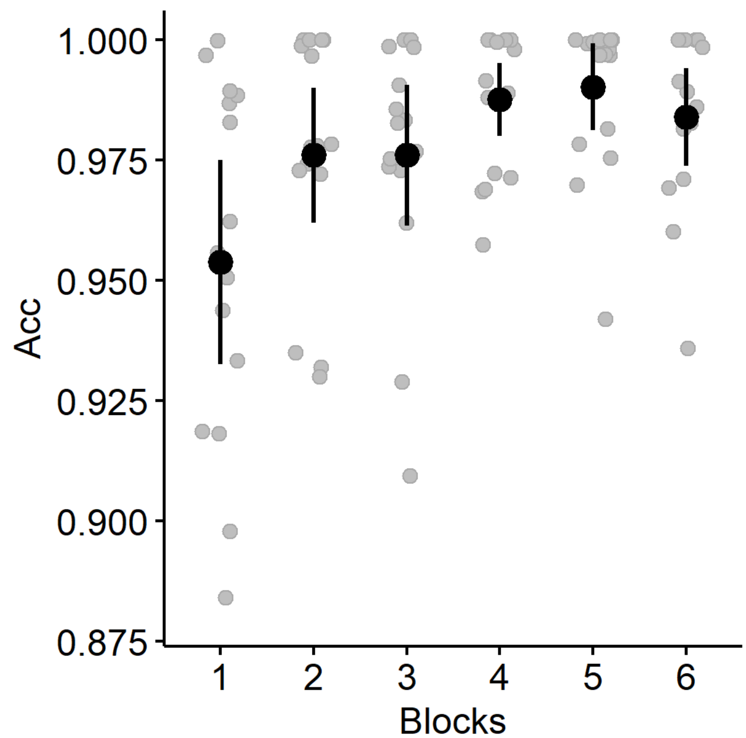

Figure 3.3: Learning trajectory of participants in the instructed learning. Error bars are 95% confidence intervals.

Uninstructed learning. Overall mean accuracy across blocks was 84% (SD= 3.6%). Accuracy on the last block was 93.4%, with all individual novel words being higher than 65%. A mixed–effects linear model was fitted to the accuracy scores, with *subjects* and *words* as random effects, and *blocks* and *word type* (neighbour vs. orphan) as fixed effects. The model showed a strong effect of *word type*, $F[1, 6732.7] = 316.55, p < 0.0001$ indicating a difference in accuracy between neighbours and orphan novel words, with the former being more accurate $\beta = 0.122, t[3725] = 5.59, p < .0001$. A main effect of *blocks* also emerged, $F[1, 6732.7] = 316.5, p < 0.0001$, indicating improvement in accuracy during the learning (see Figure 3.4).

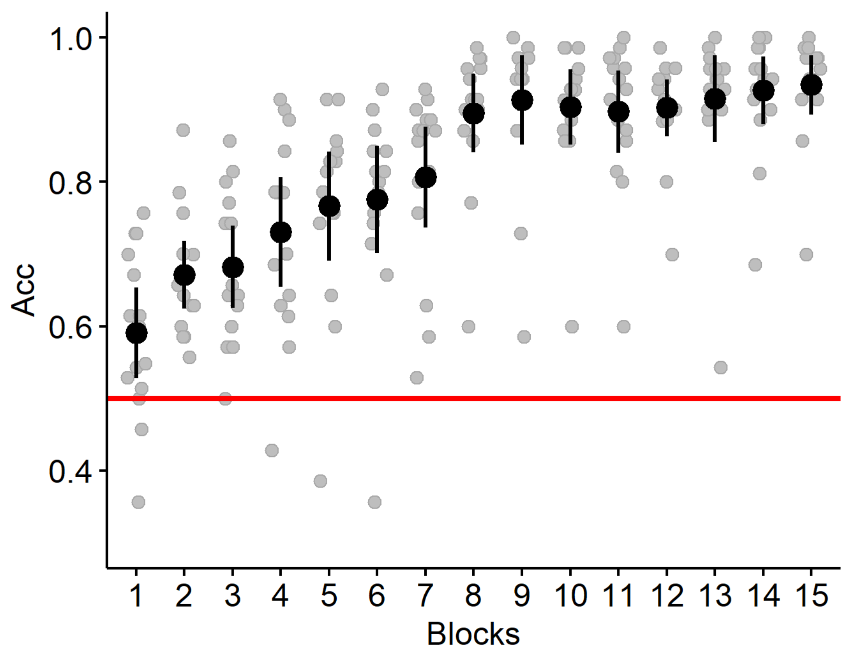

Figure 3.4: Learning trajectory of participants in the uninstructed learning. Error bars are 95% confidence intervals. Red line indicates 50% chance.

Testing phase. *Explicit Recognition.* Trials with response times shorter than 150ms or longer than 2s were removed from the analysis. Overall, accuracy was quite high ($86 \pm 3.4\%$) and did show some time–dependent modulation (D0=$92 \pm 2.7\%$, D1=$87 \pm 3.2\%$, D7=$82.9 \pm 3.7\%$). Accuracy data was transformed in d–prime measure (Macmillan and Creelman, 2005) using correct YES responses to trained novel words as "hits" and incorrect

YES responses to controls as “false alarms”. D-prime was computed separately for each participant and testing time on neighbour and orphan novel words, resulting in 180 total datapoints. A mixed-effects linear model was fitted to the dprime values, with *subjects* as random effects and *word type*, *testing time* and *learning* as fixed effects. The model reveals a main effect of *testing time*, $F[2, 140] = 41.65, p < .0001$, and a significant interaction between *testing time* and *type of learning*, $F[2, 140] = 5.4, p < .005$. D–prime decreased over testing sessions for both learning routines, but especially for instructed compared to uninstructed learning (see figure 3.5). Interestingly, on Day 7 d–primes were entirely comparable for the two learning routines. *Word type* did not affect explicit memory in any measurable way (all $F > 1$ and $p > 0.1$).

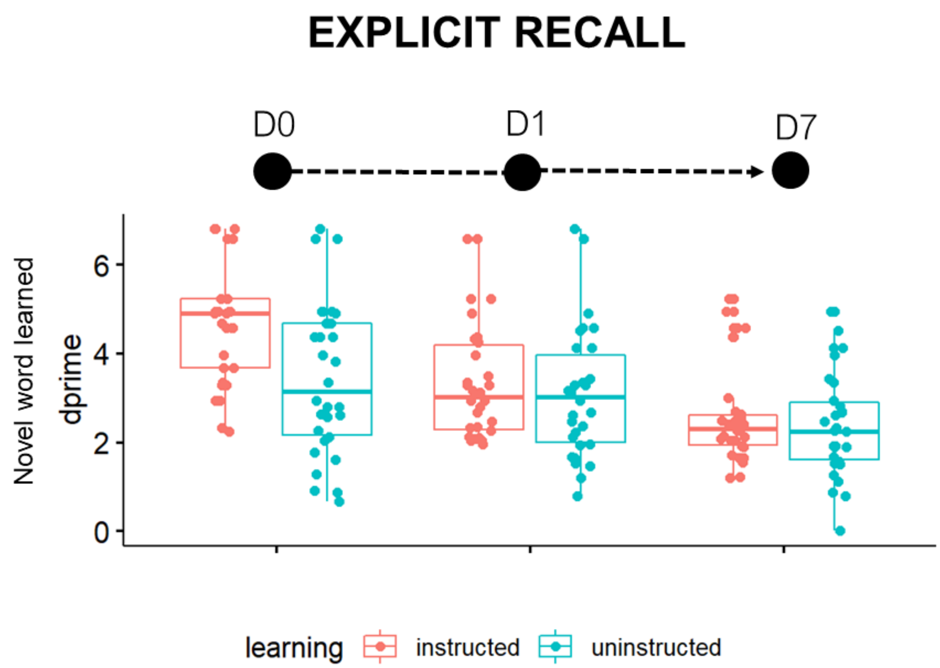

Figure 3.5: Boxplots of the d–prime scores in the explicit recognition task plotted for Day 0, Day 1 and Day 7.

Lexical Engagement. Data from two participants (subject 18 and 23) were excluded from the analysis due to an error rate higher than 20%. Response times below 200ms were also discarded. Correct responses to real words were considered for the analysis, and RTs were inverse transformed (i.e 1/RT). Table 3.2 summarizes means and SDs across conditions.

Table 3.2: Lexical engagement's RTs mean and standard deviation. Real words in the competitor condition gained a neighbour after learning, words in the control condition instead didn't.

	competitor	control
Day 0	598.79 (155)	635.49 (165)
Day 1	567.31 (124)	580.75 (137)
Day 7	586.93 (143)	596.50 (146)

A mixed-effects linear model was fitted with *subjects* and *Words* as random effects, and *testing time*, *type of learning* (instructed vs. uninstructed) and *engagement* as fixed effects. Recall that this experiment involved the base words from which the novel words were obtained (e.g., "banana" for "banara"); this latter variable contrasts those base words whose corresponding novel words were trained, thus acquiring a novel neighbour, with those base word whose corresponding novel words were not trained, thus working as a control group. A negative lexical engagement effect thus indicates faster responses to experimental compared to control base words. The model revealed a main effect of *testing time*, $F[2, 51.3] = 3.49, p < .037$, *engagement*, $F[1, 4771.9] = 44.64, p < .0001$, and a significant interaction between them, $F[2, 4771.8] = 8.71, p < .001$. Response times were faster for words with a novel competitor than for controls, but this difference decreased over testing sessions (see Figure 3.6). *Type of learning* yielded a marginally significant main effect, $F[1, 26.8] = 3.67, p = .06$, but did not interact with *engagement*, $F[1, 4792.7] = 0.34, p = 0.55$ nor with *testing time*, $F[2, 676.2] = 1.4, p = 0.24$.

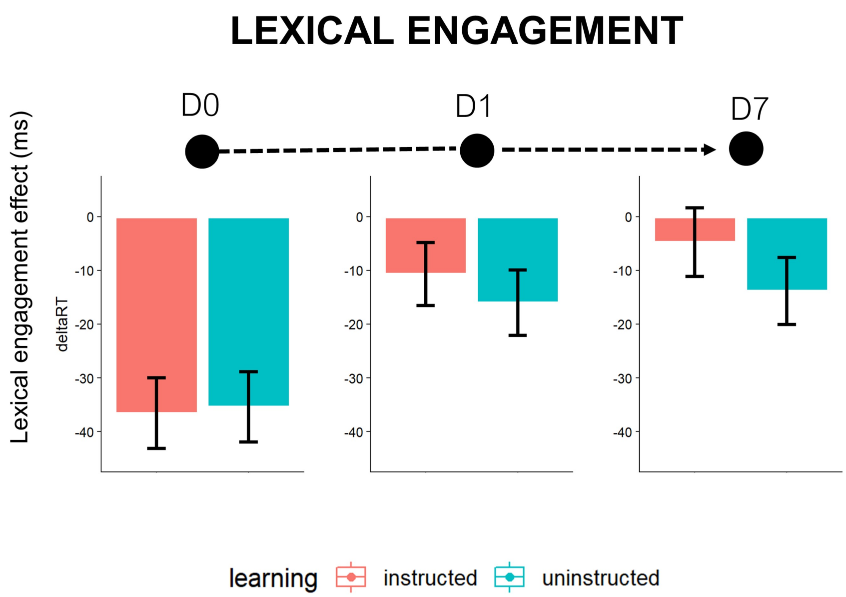

Figure 3.6: ΔRTs of lexical engagement effect in ms plotted for Day 0, Day 1 and Day 7. The lexical engagement effect is the difference between the base words which are existing neighbours of the trained novel words, and the control base words with no trained novel words. Error bars are 95% confidence intervals.

Fast Periodic Visual Stimulation (FPVS). For real words in pseudofonts, scalp topographies and EEG spectra of grand averaged data showed a clear discrimination response at 2Hz (oddball response) and 4Hz (first harmonic) over the left lateral occipital electrodes (Fig. 3.2). This nicely replicates the several recent reports showing sensitivity to words in this brain area using the FPVS paradigm (Lochy et al., 2016, 2015). Results in this condition establish a benchmark to evaluate lexicalization; therefore, data on trained and untrained novel words were assessed on the same frequency (the oddball frequency and its first harmonic) and electrodes. In these two conditions, the average of the oddball and first harmonic was analyzed with linear mixed models, with a random intercept for *subjects*, and *stimulus type* (real words, trained novel words, untrained novel words), *baseword*, *type of learning* and *testing time* as fixed effects. *word type* (neighbour vs. orphan), *type of learning* and *testing time* were added one at the time in order to avoid overly complicated models. We selected only the fixed effects that significantly explained variance in association to *stimulus type*. Analysis demonstrated a main effect of *stimulus type*, $F[2, 411.02] = 35.31, p < 0.0001$, and *testing time*, $F[2, 411.41] = 9.008, p < 0.0001$, but no interaction between the two variables, $F[4, 407.02] = 0.30, p = 0.87$. All other variables exerted no influence, all $p > 1$. Figure 3.7 shows the final analysis' results of the contrast between learned novel words and controls, while 3.8 shows the contrast between learned novel words and real Italian words.

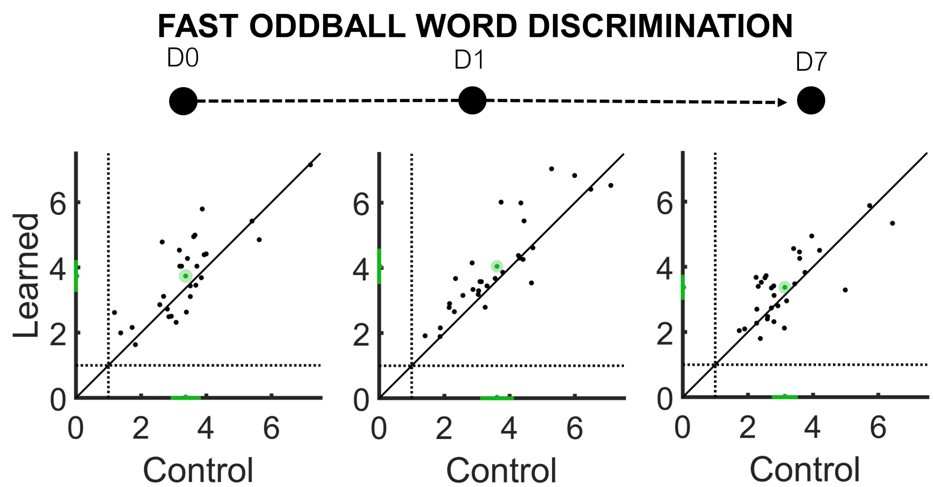

Figure 3.7: Grand-averaged (n=30) SNR EEG responses at oddball (2Hz) and first harmonic (4Hz) for the three conditions, learned novel words (hermit and non-hermit) and controls. Every black dot is a participant. Coloured dots and shaded areas indicate mean and 95% CI, respectively. Coloured bars projected on the axes indicate 95% CI of the respective conditions. Dots lying along the diagonal line indicate no difference between conditions. Dots above and below the diagonal indicate higher SNR for that specific condition.

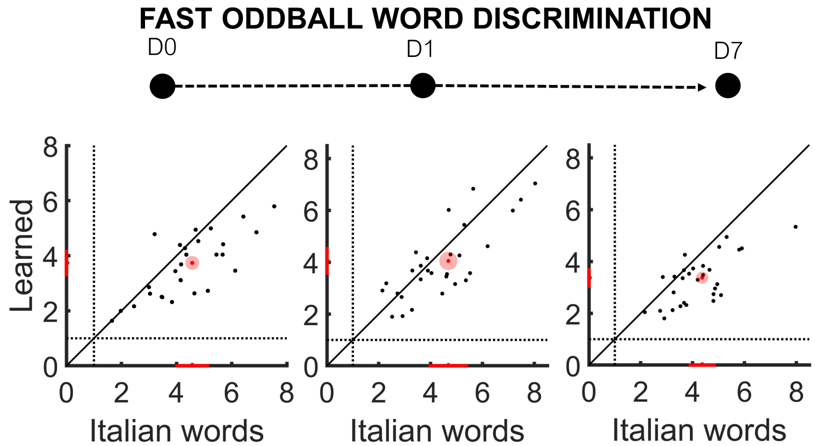

Figure 3.8: Grand-averaged (n=30) SNR EEG responses at oddball (2Hz) and first harmonic (4Hz) for the three conditions, learned novel words (hermit and non-hermit) and real words. Every black dot is a participant. Coloured dots and shaded areas indicate mean and 95% CI, respectively. Coloured bars projected on the axes indicate 95% CI of the respective conditions. Dots lying along the diagonal line indicate no difference between conditions. Dots above and below the diagonal indicate higher SNR for that specific condition.

To further validate these FPVS data, we computed the difference in SNR between trained and untrained words and assessed its correlation with the difference in SNR between trained novel words and real words. We reasoned that those participants who lexicalised better should have their novel words closer to real words, and therefore farer away from untrained novel words. Conversely, those participants who lexicalised less should have their novel words relatively far from real words, and therefore closer to untrained novel words. This predicts a negative correlation across participants between SNR(trained novel words)-SNR(real words) and SNR(trained novel words)-SNR(untrained novel words). Neighbour and orphan novel words were considered together here since there was no difference between them in the previous analysis. Results are illustrated in Figure 3.9, and show the expected negative correlation on Day 0 (Spearman ρ = -.54, p=.002), Day 1 (Spearman ρ = -.34, p=.062) and Day 7 (Spearman ρ = -.46, p=.008). All p¡0.05 except for Day 1 that is close to significance, p=0.06. Test-Retest reliability was very high: ICC(C,3) = 0.861, $F[28, 56] = 7.21, p < 0.0001$. This estimate was obtained by Intraclass Correlation Coefficient (ICC, Bartko, 1966). ICC reflects both the degree of correlation and the agreement between measurements at Day 0, Day 1 and Day 7. Reliability value ranges between 0 and 1, with values closer to 1 representing stronger reliability (McGraw and Wong, 1996). ICC estimates and their 95% confident intervals were calculated using `icc` function from the `irr` package (Gamer et al., 2019) based on a mean-rating (k = 3), absolute-consistency, two-way mixed-effects model.

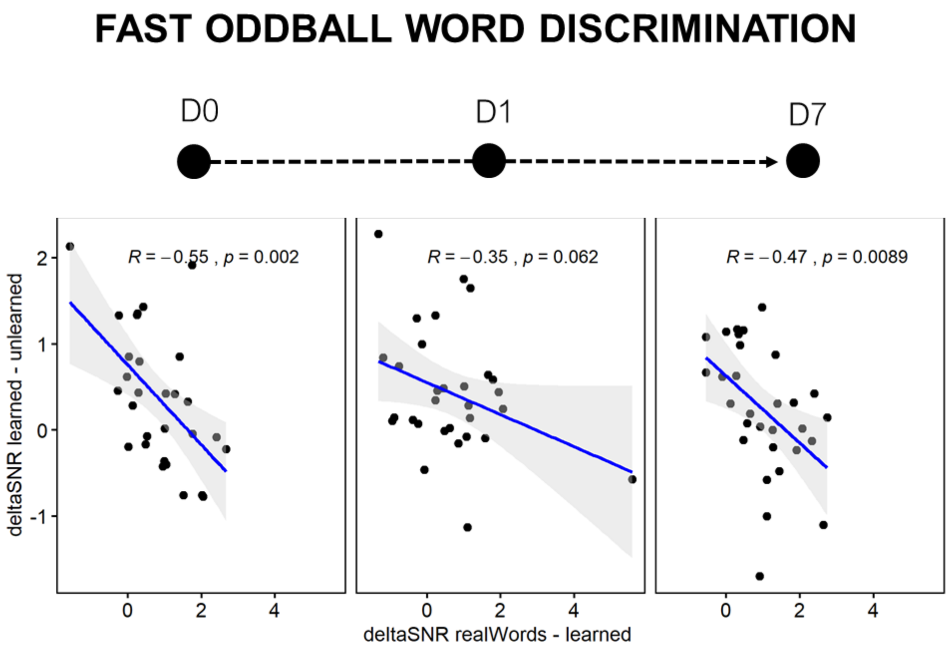

Figure 3.9: Scatter plots showing the relation between the difference in individual amplitude of the EEG discrimination responses between real words and learned novel words (on x axis) and between learned and unlearned words (on y axis).

It should be noted that in Lochy et al. (2016) study the relation between visual discrimination of words obtained via FPVS was predictive of later developing abilities in children as demonstrated by a strong correlation between SNR and two behavioral measures of prereading abilities. Hence, we tried to investigate whether our fast oddball word discrimination task was predictive of behavioural measures of explicit recall, and lexical engagement. However, correlation resulted not significant with neither lexical engagement effect (Fig. 3.10), nor with the explicit recall responses (Fig. 3.11), all $P > 0.1$.

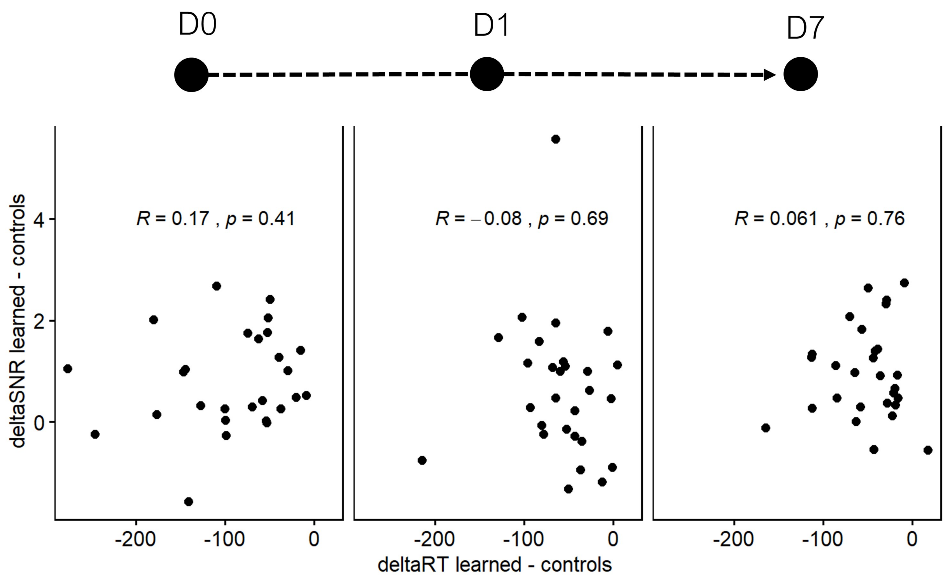

Figure 3.10: Scatter plots showing the relation between the difference in RT between learned and controls novel words in the lexical engagement task (on x axis) and between the difference in SNR of the EEG discrimination response between learned and controls novel words in the fast oddball word discrimination (on y axis).

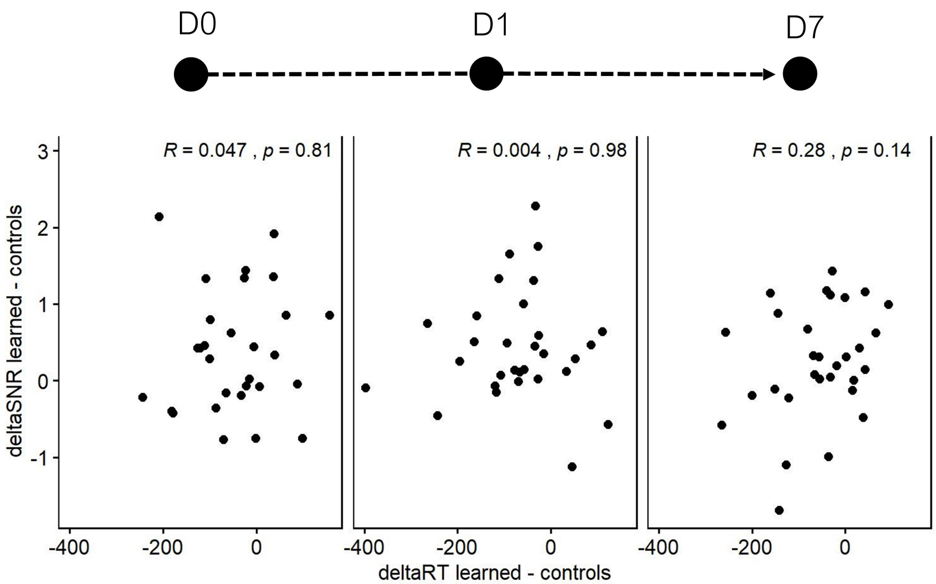

Figure 3.11: Scatter plots showing the relation between the difference in RT between learned and controls novel words in the explicit recall task (on x axis) and between the difference in SNR of the EEG discrimination response between learned and controls novel words in the fast oddball word discrimination (on y axis).

3.5 Discussion

In this study we sought to understand the learning and consolidation of novel words in adults trained via instructed and uninstructed learning routines. We tracked new word memory across three testing sessions: immediately, 24hr and one week after learning. We employed a lexical decision task in order to measure lexical integration, an old-new task to measure explicit recognition, and a Fast Periodic Visual Stimulation (FPVS) paired with an oddball design to quantify participants' ability to discriminate trained from untrained novel words, and compare their discrimination response to real words.

The performance of our participants in the explicit memory task indicates that learning was successful in both routines; by the end of the learning phase, instructed and uninstructed participants recalled a comparable and high proportion of the trained stimuli (although there were some more individual items with relatively low accuracy among the uninstructed participants).. Interestingly, in the explicit memory task, the amount of items recalled was higher immediately after learning for instructed participants. However, this difference disappeared in the other two testing sessions, 24 hours and one week later: so, although an explicit and instructed teaching strategy may apparently give better results in the short run, results are comparable in the longer term to when participants learned via an uninstructed and more implicit routine. From a slighly different perspective, while explicit memory shows a rather strong decline over time for instructed teaching, it remains more stable over time with an uninstructed approach (e.g, Dumay and Gaskell, 2007).

Do these (rather minor) differences in explicit memory translate into lexicalization and consolidation? The results from the lexical engagement task and the FPVS paradigm would suggest that this is not the case. The only hint at a difference between instructed and uninstructed learning emerged in the lexical engagement task at Day 7, but the interaction between teaching routine and engagement failed to reach significance. It is possible that studies involving more participants will eventually reveal a significant modulation, but for now, the evidence is surely too weak to claim any substantial difference in consolidation between the two learning approaches we compared in this study. Our teaching methods were of course rather different from those considered in Szmalec et al. (2012) and Sobczak and Gaskell (2019), but we do not corroborate previous results showing strong effects on consolidation from different learning routines. Apparently, having to figure out the novel words to be learned does not help their lexicalization.

As far as the quality of the interaction between newly learned and well established words, we found facilitatory rather than inhibitory lexical engagement effects. Though time exerted a detrimental effect on facilitation, this effect remained significant for the whole time-span of the experiment (except perhaps at Day 7 for one instructed participant). This result does replicate previous studies finding facilitatory effects (e.g., Henderson and James, 2018), but is somewhat in contrast with the previous literature, which predominantly reports inhibition between established words and their newly acquired competitors. These data are also inconsistent with the Complementary Learning System (CLS) account, which would predict no interaction between novel and old memories immediately after learning, and competition as soon as novel words are lexicalised, typically over a period of time including sleep. Critically, our novel words did not disobey the CLS predictions by simply not exerting *any* effect on established words. They did interact with them, showing that they have gained some form of representation able to attack the lexical network. Similarly to previous reports of facilitatory effects, then, it is difficult to hold that the new lexical material just never consolidated into the neocortex. Simply, the interaction with existing lexical memories is facilitatory, rather than inhibitory.

One possible explanation for this refers to the high consistency in phoneme–grapheme mapping in Italian spelling. Our novel (and existing) words live, pretty much unavoidably, in very consistent phonological and orthographic neighbourhoods, which may have influenced the dynamic interaction between novel and previously known words, perhaps providing enough shared ground to support rapid integration (McClelland, 2013; Tse et al., 2007, 2011; van Kesteren et al., 2010, 2013). This is consistent with evidence showing that prior knowledge enhances the ease with which new information is integrated into longer–term memory system, although this does not always result in facilitation (Tse et al., 2011, 2007; James et al., 2017). Another account, perhaps related to the one above, would refer to previous literature on words neighbours in existing words, showing facilitatory effects in visual word recognition for words living in a small phonological and orthographic neighborhoods (Grainger et al., 2005; Ziegler et al., 2003). Based on previous word learning data, we selected our base words to live in a sparse lexical space; and the perfect match between orthography and phonology in Italian guarantees that both the visual and the acoustic lexical spaces were such. Under these conditions, orthography and phonology can reinforce each other, and this reciprocal facilitation may overcome lexical competition. This notion of co–activation may also be relevant in explaining why we see facilitation effects right after learning, which is in line with previous reports of immediate lexical engagement effects (e.g., Lindsay and Gaskell, 2013; Coutanche and Thompson-Schill, 2014; Fernandes et al., 2009).

The FPVS data helps interpreting the behavioural lexical engagement results, and lend substantial support to the idea that novel words did indeed consolidate and gained some form of representation into the neo–cortex. In this task, we obtained higher discrimination response to trained novel words than untrained controls, in the very same electrodes (and therefore brain areas, arguably) that offered a strong response to existing words. This suggests that the brain developed some form of representation for the novel words, or at least showed enhanced activation in the lexical network in response to the presentation of these stimuli, as compared to untrained, unfamiliar letter strings. The FPVS paradigm has been already proven to be a successful method to measure orthographic and lexical sensitivity. For instance, it has been used to show differentiation between words and pseudowords in adult readers, in a neural pattern that was interpreted as reflecting the work of the the left occipito–temporal cortex (Lochy et al., 2015). Moreover, it has been used with 5-year-olds to show rapid discrimination of words from false fonts, with the strength of this effect predicting later developing reading ability (Lochy et al., 2016). These data are nicely in line with recent evidence showing that learning artificial words in the lab leads to the development of specific lexical–ortographic representation in the left occipito–temporal cortex (Taylor et al., 2019). In this study, we confirm and extend the potential of the FPVS paradigm by showing that the strength of the discrimination response for trained items approximates that for real words. Also, we showed that the individual difference in discrimination responses between the trained and untrained strings correlates with the difference between the former and existing words. This fits nicely with the idea that FPVS can track the lexicalization of novel words—the more they get away from non–lexical strings, the more they become similar to previously established lexical representations. Of course, the novel words investigated here never really yielded the same discrimination response of existing words. This is not particularly surprising, since our learning routines did not provide any semantic information, which real words clearly have. Even assuming that meaning is not heavily represented in this part of the ventral stream (Dehaene and Cohen, 2007; Taylor et al., 2013), the amount of exposure to novel words we gave our participants is not even remotely comparable to a lifelong exposure with existing words, which further justify to difference we observe between newly learned and existing, long–established words. Another interesting aspect of the FPVS results is that novel words elicited a strong response across all time points, thus showing that novel word memories were successfully transferred onto longer–term storage. This contrasts with the behavioural results in the lexical engagement task, which show facilitation to shrink substantially with time from learning. This comparison points to the possibility that FPVS and the lexical engagement task tap onto at least partially different cognitive mechanisms. This hypothesis gains support from the low subject–wise correlation between these metrics, and may be related to the fact that FPVS stresses the most automatic and implicit aspects of the novel word memories—not only there is no action required to the participants on the linguistic stimuli, but these are even kept unattended by the concurrent visual task of spotting a color change in the fixation cross. Lexical engagement, at least as attested here, is also an indirect measure of novel word learning, but surely requires participants to focus on lexical material, and actively and explicitly engage their lexical system. Perhaps the novel word representations are not strong enough to surface consistently over time in this more explicit task; or perhaps the memory systems captured by FPVS and lexical engagement behavioural tasks are largely different. Our data cannot speak to this issue, which is left open for future research.

As a final note, FPVS also allowed us to compare consolidation in neighbour novel words, those that were derived by substitution of a letter from real word, and therefore has a close germinate in the established lexicon; and orphan novel words, which lacked such a close neighbour. This comparison is impossible to assess behaviourally, as lexical engagement tests rely on the very presence of a neighbour, cutting off orphan novel words entirely. Interestingly, we did not find any evidence for different consolidation pathways for these types of words, which qualifies the role of prior knowledge, at least at the word form level. While a richer lexical network entirely (e.g., James et al., 2017) and/or a consistent lexical space (see above) do seem to affect the consolidation of new lexical memories, having a close neighbour may allow a slight immediate memory advantage, but does not seem to affect deep lexical learning.

Chapter 4

Experiment 3 - Domain generality and specificity of statistical learning: the case of orthographic regularities

4.1 Abstract

Individuals become sensitive to frequently recurring patterns in the environment. Yet, evidence that the human visual system profits from the statistical regularities during reading are mixed. The current series of experiments address this issue by testing whether sensitivity to orthographic regularities derive from a general domain statistical learning mechanism, or is a specialized mechanisms developed for reading only. As a test—bed we focused on ngrams frequency. We have implemented a statistical learning paradigm where stimuli are pseudofont strings made of recurrent pairs of characters. Participants have to memorize the false font strings and subsequently disentangle old from new stimuli. Our results show that participants made decisions about novel strings based on the bigram frequency of the learned stimuli. In a non-linguistic version of this paradigm, we explored the boundaries of this phenomenon: what type of visual object, if any, fails n-gram based statistical learning? Stimuli became progressively less word-like (e.g., Y-shaped objects vs. gabor patches) and we show that, similar to what happens with (pseudo)reading material, participants have a hard time discarding objects that they've never seen, but comply with the statistical pattern of the smaller parts.

4.2 Introduction

Since the publication of the seminal paper by Saffran et al. (1996), an ever growing body of research has shown that the human brain is particularly apt at exploring the statistical regularities of the environment. While the initial proposal of Saffran and collaborators was that Statistical Learning allows infants to discover novel words in the continuous speech stream, it was soon found that this ability extend beyond linguistic stimuli (Aslin et al., 1999) and is not restrained to the auditory modality (Fiser and Aslin, 2002; Kirkham et al., 2002). In fact, in recent years Statistical Learning has become a theoretical framework for information processing that can be applied to a variety of cognitive domains (Armstrong et al., 2016).

One such domain is the human ability of reading. Most writing systems rely on combinations of individual graphemes (e.g., letters) to form meaningful units (i.e. words). Grapheme combinations exhibit rich statistical regularities, to which the reading system seems to be sensitive. For example, it has been shown that letters embedded in words (such as S in the English word FLASH) or in word–like letter strings (S in FRISH) are more efficiently recognized than letters embedded in unusual letter strings (S in RFHSL), or presented in isolation (Reicher, 1969; Wheeler, 1970). Such evidence suggests that readers use information about frequently recurring letter combinations to efficiently perceive and identify letters and letter strings.

While Statistical Learning offers a plausible framework to study reading, whether the human reading system profits from the statistical regularities present in written material remains a hotly debated topic in the psycholinguistic community. In particular, performance in behavioral tasks such as letter/word detection or lexical decision tends to be modulated by orthographic regularities such as the presence of frequent letter clusters (n–gram frequencies) when using non–words as stimuli, but results become ambiguous when stimuli consist of real words.

To exemplify the current state of affairs, it is enough to compare two recent thorough reviews of the literature. While Schmalz and Mulatti (2017) points out the presence of mixed results and use Bayesian analysis to argue in favor of the null hypothesis of "absence of a bigram frequency effect in lexical decision, provided that the items are matched across lexicality and bigram frequency cannot be used as a non–lexical cue to derive a correct lexical decision response", Chetail (2015) draws the conclusion that "results are not as inconsistent as they seem, and on the contrary, sensitivity to orthographic regularities may influence visual word recognition at all levels of processing".

A less ambiguous picture emerges when considering neuroimaging evidence. For example Binder et al. (2003) used fMRI in a letter detection task where targets were embedded in letter strings of varying sublexical orthographic familiarity. Their results shown that the left fusiform gyrus is tuned to positional bigram frequency and that mean percentage signal change in this region is correlated with behavioral performance. In a further fMRI study, Vinckier et al. (2007) explored the tuning properties of the fusiform gyrus by exposing participants to character strings of varying degrees of familiarity, raging from false font to real words. Their results show a posterior to anterior gradient whereby tuning progresses from individual letters to letter pairs (bigrams) and morphemes. More recently, Lochy et al. (2018) used human intracranial EEG recordings to show evidence for populations of neurons sensitive to the statistical regularity of letter combination in the left fusiform gyrus.

Neuroimaging evidence suggests that the human brain is sensitive to the statistical regularities present in orthographic–like material. As the locus of this sensitivity is usually high–level visual areas, a question arises as to whether this coding scheme constitutes a particular adaptation of the reading system, or is instead a general processing tool that would apply to any type of visual stimuli. In this line, it has been pointed out that as writing systems are a relatively novel invention, the human brain could not have evolved neural mechanisms for visual orthographic processing ex novo (Dehaene and Cohen, 2007). Instead, the reading system might "recycle" visual processes that are domain general in nature. Supporting this view, a body of work recently emerged suggesting that during reading acquisition, the cortical surface in the left fusiform gyrus that is originally devoted to the processing of visual objects (and faces in particular) becomes tuned to orthographic material (Dehaene and Cohen, 2007; Dehaene et al., 2010; Ventura, 2014; Pegado et al., 2014, 2010; Hervais-Adelman et al., 2019).

One prediction that can be derived from the recycling hypothesis is that if reading relies on domain general visual mechanisms, then non–literate animals might also possess the ability to solve tasks that might be deemed as orthographic. In line with this, it has been shown that baboons can be trained to distinguish words from nonwords based on the orthographic regularities of letter co–occurrences (Grainger et al., 2012), and generalizing to new unseen stimuli with above chance performance. This results were later replicated with pigeons (Scarf et al., 2016). Furthermore, recent work has shown that neurons in high level visual cortex of naive macaque monkeys could support orthographic processing tasks (Rajalingham et al., 2019).

The aforementioned body of work suggests that at least part of what is usually considered as reading-specific orthographic processing is in fact the result of domain–general visual mechanisms that are also at play in non–linguistic animals. Although this cross-species evidence is very telling, more direct evidence of domain generality is needed. Therefore, the goal of this work is to test whether the effects of sensitivity to statistical regularities that are usually found with orthographic–like material can be extended to non orthographic visual stimuli in humans. This would imply that these effects are due to a domain general visual mechanism that implements statistical learning. Here we investigated this issue with a novel experimental design, which neatly isolates statistical regularities effects. The type of statistical regularity in which we focused is the frequency with which pairs of graphemes appear together. These regularities, also know as bigram frequencies, have been proposed to be an intermediate step between single graphemes and small words or word fragments (Dehaene et al., 2005).

Experiments 1 and 2 used novel words written in BACS pseudofont (Vidal et al., 2017) to show that when faced with a visual word identification task, participants' performance is modulated by mean bigram frequency. The use of pseudofont to construct novel written words allows for a tight control of confounds such as phonology, semantics and the particular history of exposure to print of the participants, allowing a complete control of stimuli statistics. This character set has already been successfully used by Chetail (2017) in a statistical learning study were the author explores how new regularities learnt by mere print exposure affect the processing of letter strings. Following this, we performed two experiments using stimuli that progressively differ from orthographic–like material in order to test whether the orthographic regularities effects persisted. Experiment 3 used exactly the same experimental design as in experiments 1 and 2, but the stimuli used consisted on images of novel Y shaped objects with distinctive shapes attached to their terminals (similar to Baker et al., 2002). These objects differed from word–like material in two ways. First, instead of being formed by adjacent but independent graphemes, the parts conforming the objects were physically connected in a unit. Second, the conforming parts followed a radial spatial arrangement, rather than the linear spatial arrangement that is typical of orthographic material. This experiment allowed us to test whether the co–occurrence effects found with orthographic material could be also found with stimuli that are not orthographic in nature. Although the objects used as stimuli in experiment 3 were not orthographic, they still shared with written words the fact of being constituted by combinations of shapes in a particular spatial arrangement. Therefore we performed experiment 4, in which stimuli were devoid of any spatial arrangement and composed by abstract features; we used circular Gabor patches in which the defining features were spatial frequency, orientation and contrast. Co–occurrence effects in this experiment would imply that the processes that help bind graphemes together into words might not depend on any spatial arrangement and might be generalized to the simple visual features that define Gabor patches.

In brief, the experiments presented here aim at capturing the effects of letter co–occurrence typically found with orthographic–like material (experiments 1 and 2). Next, they attempt to replicate such effects with stimuli that are not orthographic in nature, and that does not respect the linear arrangement of orthographic material (experiment 3). Finally, they attempt to replicate such effects with stimuli that are devoid of spatial arrangement and composed of low level visual features (experiment 4). By characterizing co–occurrence effects across different types of stimuli, we were able to test the degree of domain generality of some visual processes used in reading.

4.2.1 Experiment 1

Methods

Participants. Participants were self–reported right handed, Italian native speakers, and were recruited from the city of Trieste via on–line advertisement. They all had normal or corrected–to–normal vision and no language–related impairments. Twenty–two participants (5 male and 17 female) took part in the experiment (mean age was 23.4 ± 2.21 years). Participants signed informed consent and received a monetary compensation of 10 euro. The experiment was approved by SISSA's Ethical Committee.

Stimuli set. All stimuli where generated using the Brussels Artificial Character Set II (BACS–2, Vidal et al., 2017), whose characters have perimetric complexity, number of strokes, junctions and terminations matched to the English alphabet . We picked 23 out of the 26 available characters in BACS–2 with serifs.

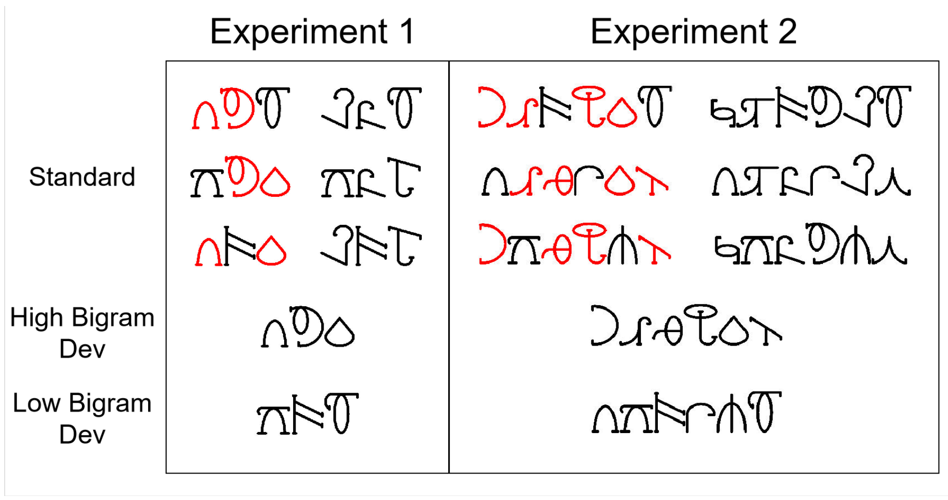

Figure 4.1: Representative stimulus sets used in Experiments 1 and 2. Note that all the characters pairs making up High Bigram Deviants are shared with Standard words (marked in red for illustration purposes)

For each participant, 9 characters were randomly selected out of the set described above, and were used to construct the stimuli for Experiment 1. An example stimuli set can be seen in Figure 4.1. First, These characters were used to construct 6 three–character combinations (e.g., XYZ), which were used as Standard words.

Next, taking these Standard words as a base, two different Deviant words were constructed. The first of such deviants was built using bigrams (pair of characters) that were present in three of the Standard words. This also included open bigrams composed by two non–adjacent characters with one other character in between. This deviant will be referred to as "High Bigram Deviant". The second deviant on the contrary, used characters present in the Standard words, but did not share any bigram with them. We will refer to it as "Low Bigram Deviant".

These stimuli were used in an oddball detection design, which, together with the way stimuli were constructed, generated a design with two independent variables, orthogonally manipulated. While word frequency was 15% for each standard and 5% for each deviant word, mean bigram frequency was instead high for standards (5.27%) and High Bigram Deviants (6.66%), but low for Low Bigram Deviant words (1.66%).

The statistical structure of the stimuli set was based on the one used in Endress and Mehler (2009), where the authors show how in the context of speech segmentation, shared pairs of syllables affect participants' word recognition. In our case, we reasoned that if participants are sensitive to character co–occurrence statistics, the rejection (i.e. detection as deviants) of High Bigram Deviants, which followed the same orthographic regularities as Standard words, should be harder than the rejection of Low Bigram Deviants, which violated these orthographic regularities. If, instead, participants are not sensitive to character co–occurrence statistics, both types of deviants should be equally easily to distinguish from standard words.

Note that characters were randomly picked for each participant. Therefore each participant was presented with a different rendering of the stimuli, but with the same statistical structure. This guarantees that differences between deviant types cannot be explained by idiosyncratic factors of a given set of characters.

Procedure. Participants sat in a sound–attenuated testing booth at around 70cm of a 27 inches computer monitor (BenQ XL2720Z). The experiment was programmed and run in MATLAB (2015b, MathWorks, Inc., Natick, MA, USA) using the Psychophysics Toolbox extensions Brainard (1997). Participants first completed a learning block in which Standard words were presented one at a time, for a total of 200 trials. Words were on screen for a time between 1.5 and 2 seconds. Participants were instructed to pay attention to the words and try to learn them; at this stage, they were not required to provide any response. After the learning phase, the experiment followed a visual oddball design. Participants completed 6 blocks of 200 trials each, where standard words were presented intermixed with deviant words. In these blocks, participants were given a maximum of 2 seconds to classify each stimulus as either "Correct" (standard) or "Mistaken" (deviant) by pressing one of two buttons on a keyboard. Participants were not informed about the existence of two different types of deviants, nor about the amount of deviants that would be presented. In these testing blocks, while standard words were presented in 90% of the trials (15%, or 30 repetitions for each of the six items), Low Bigram Deviants and High Bigram Deviants were presented in 5% of the trials each (10 repetitions each). Overall, each participant was asked to classify 1080 instances of standard words, and 60 instances of each deviant word. Each block lasted on average 7 minutes and the entire experiment had an approximated duration of 50 minutes.

Measure of performance. Trials in which participants did not provide an answer within 2 seconds were excluded from the analyses (1.79%), and participants with more than 20% of such trials for any stimulus category were also excluded (there was no such participant in Experiment 1).

To better characterize the participants' ability to detect deviant stimuli, we resorted to Signal Detection Theory and computed a d–prime score (*d'*) for each participant, for each deviant type (Stanislaw, 1999). Participants' responses were classified as "hit" (deviants classified as "mistaken") or "false alarms" (standard words classified as "mistaken"). The two types of deviants were considered separately, Next, *d'* was calculated as Z(hit rate) minus Z(false alarm rate), therefore taking into consideration the overall bias towards a "Correct" or a "Mistaken" response. As this function does not output a finite value if either the hit rate or the false alarm rate are either 0 or 1, and considering the total amount of trials of each type, hit rate was capped between 1/60 and 59/60, and false alarm rate was capped between 1/1080 and 1079/1080.

All effect sizes reported are Hedges' g Lakens (2013), which is more precise than Cohen's d, as it applies a correction for small sample sizes. Effect sizes were calculated using the Measures of Effect Size Toolbox Hentschke and Stüttgen (2011). All confidence intervals reported between square brackets are 95% CIs.

Results

The mean *d'* for High Bigram Deviants was 2.02 [1.44, 2.59], which was above 0 ($t_{(21)} = 7.33$, $p = 1.63\text{e-}7$, $g = 1.56$ [0.93, 2.18]). The mean *d'* for Low Bigram Deviants was 0.84 [0.24, 1.45], which was also above 0 ($t_{(21)} = 2.89$, $p = 0.0043$, $g = 0.62$ [0.15, 1.07]). Result are plotted in 4.3, panel A.

When comparing performance across deviant conditions, High Bigram Deviants were indeed harder to detect than Low Bigram Deviants. The difference in *d'* was 1.17 [0.77, 1.57], which was again above 0 ($t_{(21)} = 4.30$, $p = 0.00015$, $g = 0.86$ [0.36, 1.37]). Interestingly, High Bigram Deviant's *d'* was lower than Low Bigram Deviant's *d'* in 86% [65%, 97%] of the participants (19 out of 22, One side binomial test: $p = 4.27\text{e-}4$), showing impressive cross–participant consistency.

Discussion

These results show that, while participants were able to classify both deviants as "mistaken" words, the detection of High Bigram Deviants, which followed the same orthographic regularities as standard words, was harder. This implies that participants were sensitive to the orthographic regularities present in the stimuli.

The stimuli used in Experiment 1 were three–character long, which implies both advantages and disadvantages. One one hand, given the short length of the stimuli, the fact that participants were sensitive to the orthographic regularities of characters combinations, rather than encoding for whole word as a units, is even more surprising. On the other hand, three-character words are not particularly representative of the typical length of words in real languages. We therefore extended our investigation to longer words in Experiment 2, thus assessing whether the sensitivity to character co–occurrence statistics unveiled in Experiment 1 holds across a wider range of word lengths. Importantly, Experiment 2 also offers a conceptual replication of the results found in Experiment 1.

4.2.2 Experiment 2

Methods

Stimuli set and procedure. Experiment 2 followed exactly the same design of Experiment 1, with the only exception that now all words were 6 characters in length. For each participant, 18 characters out of the 23 available were selected. As in the case of Experiment 1, these characters were used to construct 6 Standard words, one High Bigram Deviant, which shared all of it's constitutive bigrams with Standard words, and one Low Bigram Deviant, which on the contrary, did not share any bigram with Standard words.

A representative stimulus set is reported in Figure 4.1. As the words presented in Experiment 2 were longer than the ones presented in Experiment 1, all bigram frequencies were exactly 1/3 of the frequencies in Experiment 1, but the ratio between frequencies remained equal.

Every other aspect of the experiment, including the arrangement of the learning and testing blocks, the overall number of trials, and the individual trial timeline, were in all identical to Experiment 1. Data were also analysed exactly as in Experiment 1. Two participants were excluded from the analysis for failing to provide an answer within 2 seconds in more than 20% of the trials in one or more conditions.

Results

In Experiment 2, the mean *d'* for High Bigram Deviants was 0.41 [-0.19, 1], which was not possible to distinguish from 0 ($t_{(22)} = 1.42$, $p = 0.084$, $g = 0.3$ [-0.12, 0.71]). The mean *d'* for Low Bigram Deviants was 1.71 [1.13, 2.29], which instead was above 0 ($t_{(22)} = 6.15$, $p =$ 1.71e-6, $g = 1.28$ [0.72, 1.83]). Results are plotted in 4.3, panel B.

As in the case of Experiment 1, High Bigram Deviants were harder to detect than Low Bigram Deviants. The difference in *d'* was 1.30 [0.88, 1.73] which was above 0 ($t_{(22)} = 4.49$, $p = 9.11\text{e-}5$, $g = 0.94$ [0.41, 1.48]). This effect was present in 87% [66%, 97%] of the participants showed an effect in the direction of our hypothesis (20 out of 23, One side binomial test: $p =$ 2.44e-4), which confirms that the effect was highly reliable.

Discussion

To sum up, Experiment 2 results show that while participants were able to detect Low Bigram deviants, this was not the case for High Bigram Deviants. Stimulus length may have taken its toll here; 6–character strings were obviously more difficult to identify than the 3–character strings used in Experiment 1, which may have made High Bigram Deviants very difficult to distinguish from Standard words in this experiment.

More importantly, the comparison across deviant types showed that the High Bigram Deviants, which contained the same orthographic regularities as Standard words, were harder to detect than the Low Bigram Deviants which violated these regularities. This confirms the core result of Experiment 1, and proves that the effect we uncovered here is resistant to change in stimulus length, and emerges also with pseudo-character strings whose length is more representative of the written language our participants were familiar with.

The results of Experiments 1 and 2 allowed us to test our novel experimental design, and capture the effect of orthographic regularities with word–like stimuli. These results show how participants' performance is biased by orthographic regularities such as mean bigram frequency when they are faced with the task of learning novel strings of letter–like symbols. One possibility is that this sensitivity to the statistics of symbol co–occurrences might have been developed ad hoc by the human visual system to assist in the task of binding graphemes into words. Instead, we consider that this sensitivity to co–occurrence regularities might be a particular case of a domain general sensitivity to statistical regularities, which would be a core feature of the whole visual system. To test this hypothesis, we performed Experiment 3, which while following exactly the same experimental design as Experiments 1 and 2, it used renderings of 3d objects as stimuli. The goal of this experiment was to replicate the effects found in Experiments 1 and 2, but using non–orthographic stimuli.

4.2.3 Experiment 3

Methods

Stimuli set and procedure. Experiment 3 followed exactly the same design as Experiments 1 and 2, with the only exception that the stimuli used were not orthographic in nature. In fact, we used images of 3D objects created using the software Blender (version 2.79b; Community, 2017). Exemplars of these objects are reported in Figure 4.2. They were all composed of a central Y–shaped body and one distinctive shape attached to each of three branches. The overall object play the role of words in Experiments 1 and 2, while the terminal shapes play the role of letters.

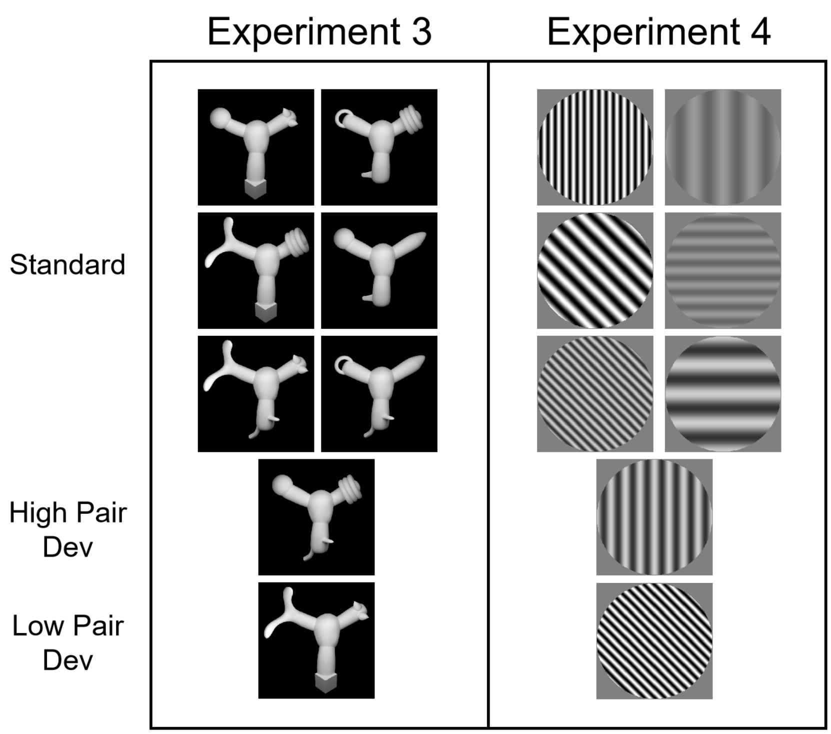

Figure 4.2: Representative stimulus sets used in Experiments 3 and 4. In the case of Experiment 3, each pair of shapes composing High Pair Deviant objects were shared with Standard objects. Whereas in Experiment 4, each pair of visual feature values defining High Pair Deviant Gabor patches were shared with a Standard Gabor patch.

As in Experiments 1, we first selected a total of 9 distinctive shapes, with which we constructed the 6 Standard objects. These objects were used as a base to construct the same two different types of deviants employed in the previous experiments. The first of such deviant objects was composed by pairs of shapes that were all present in the Standard set, therefore matching the statistical regularities of that condition. This object played a role analogous to the High Bigram Deviant in Experiments 1 and 2, and we will refer to it as "High Pair Deviant". The second deviant object was instead constructed with pairs of shapes that were not present in any Standard object. It played a role analogous to the Low Bigram Deviant in Experiments 1 and 2, and we will refer to it as "Low Pair Deviant". As each object was defined by the combinations of its 3 terminal shapes, the stimuli statistics were identical to those in Experiment 1.

Two different sets of images with the same statistical structure were created, and each participant was exposed to one of them. This had the goal of ruling out the possibility that the effects were driven by some idiosyncratic feature of a given set of shapes.

Every other aspect of the experiment, including the arrangement of the learning and testing blocks, the overall number of trials, and the individual trial timeline, were in all identical to the previous experiments. Data were also analysed exactly as in previous experiments. One participant was excluded from the analysis for failing to provide an answer within 2 seconds in more than 20% of the trials, in one or more conditions.

Results

In Experiment 3, the mean *d'* for High Pair Deviants was 0.65 [0.16, 1.15], which was above 0 ($t_{(38)} = 2.66$, $p = 0.0056$, $g = 0.43$ [0.10, 0.75]). The mean *d'* for Low Pair Deviants was 1.58 [1.14, 2.03], which once more was above 0 ($t_{(38)} = 7.22$, $p = 6.27\text{e-}9$, $g = 1.16$ [0.74, 1.56]). Results are plotted in 4.3, panel C.

As in Experiments 1 and 2, High Pair Deviants, which respected the statistical regularities of Standard objects, were harder to detect than Low Pair Deviants, which violated such regularities. The difference in *d'* between conditions was 0.93 [0.58, 1.29], which was different from 0 ($t_{(38)} = 3.81$, $p = 0.00024$, $g = 0.64$ [0.27, 1.01]). As in the previous experiments, cross–participant consistency was high, as 74% [58%, 87%] of the participants showed an effect in the direction of the hypothesis (29 out of 39, One side binomial test: $p = 0.0017$).

Discussion

Most writing systems construct words by combining adjacent graphemes in a linear arrangement. Contrary to this, the stimuli in Experiment 3 were made up of combinations of shapes that were linked into a single object. Furthermore, the shapes defining this objects were arranged radially instead of linearly. Despite these differences, the results of Experiment 3 replicate the co–occurrence effects found in Experiments 1 and 2, using material that is clearly not orthographic in nature.

On their own, these results already suggest that the sensitivity to orthographic regularities that can be found in the case of script–like material could be the result of a domain general visual mechanism that implements statistical learning. But while not orthographic in nature, the objects used as stimuli in Experiment 3 still share with written material the fact of being composed of shapes in a particular spatial arrangement.

In order to test the limits of the domain generality of this mechanism, we decided to run a final experiment in which the stimuli were entirely devoid of any spatial arrangement. In fact, they were not constructed using shapes, but consisted instead of combinations of low level visual features (spatial frequency, contrast and orientation) that eventually blended onto a given Gabor patch.

4.2.4 Experiment 4

Methods

Stimuli set and procedure. Simuli were Gabor patches defined by a particular combination of parameters in three different low level visual features (see Figure 4.2). These features, which played a role analogous to characters in Experiment 1 and 2, were spatial frequency (.4, .8 and 1.6 cycles per degree of visual angle), contrast (20%, 60% and 100%) and Orientation (0, 45 and 90 degrees). In turn, the Gabor patches played a role analogous to the words in Experiment 1 and 2.

As for previous experiments, we first defined 6 Standard Gabor patches. For example, one of such Gabor patches could have a spatial frequency of .8 cycles per degree of visual angle, 20% contrast and 90 degrees of orientation. These Standard Gabors were used as a base to construct two different types of deviants. One of them was defined by pairs of feature values that were all shared with Standard Gabor patches, therefore following the same statistical regularities of that condition. We will refer to it as "High Pair Deviant". The other type of deviant was defined using the same feature values that were use in the Standard Gabors, but in combinations that were not present in any other Gabor, therefore violating such statistical regularities. We will refer to it as “Low Pair Deviant”. Stimuli statistics were identical to previous experiments.

Results

Mean *d’* for High Pair Deviants was 1.01 [0.42, 1.60], which was above 0 ($t_{(34)} = 3.49$, $p = 0.00068$, $g = 0.59$ [0.23, 0.95]). In the case of Low Pair Deviants, mean *d’* was 2.22 [1.69, 2.74], which was also above 0 ($t_{(34)} = 8.59$, $p = 2.48\text{e-}10$, $g = 1.45$ [0.97, 1.92]). As in the previous experiments, High Pair Deviants, which respected the statistical regularities of Standard objects, were harder to detect than Low Pair Deviants, which violated such regularities. The difference in *d’* between conditions was 1.21 [0.79, 1.63], which was different from 0 ($t_{(34)} = 4.12$, $p = 0.00011$, $g = 0.74$ [0.33, 1.14]). Once more, this difference between deviants detectability was present in the majority of the participants (83% [66%, 93%], 29 out of 35. One side binomial test: $p = 5.84\text{e-}5$), which implies that the effect was highly reliable. Results are plotted in 4.3, panel D.

Discussion

The stimuli used in Experiment 4 are radically different from orthographic material. The features defining the identity of each Gabor patch are not shapes (as is the case with graphemes) and are not organized in a spatial arrangement. Instead, they consist on low level visual features which are represented in primary visual cortex (Hallum et al., 2011). Despite these differences, the results Experiment 4 neatly replicate our previous findings. Given that the visual features used to define the stimuli in this experiment constitute the building blocks of visual perception, the results of this experiment suggest that our sensitivity to orthographic regularities could be extended to any type of visual stimuli.

4.2.5 Results across experiments

We tested the hypotheses that Low Bigram Deviants and High Bigram Deviants *d’* are individually larger than zero, and that the former is higher than the latter; this was achieved via one–sample and paired–samples t tests. As all the tested hypotheses were directional, one–tailed tests were used. The results of the four experiments presented in this work are illustrated in Figure 4.3. A summary of hit and false alarm rates can be found in Table 4.1.

Furthermore, a summary of the sensitivity index (d') for each condition can be found in Table 4.3.

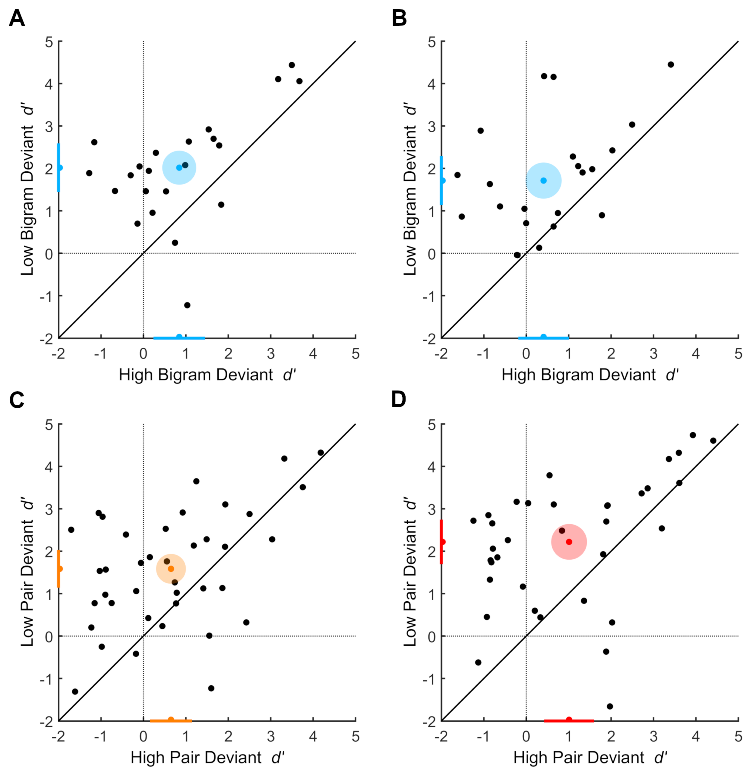

Figure 4.3: Scatter plots of participants' sensitivity indexes (d'). **A:** Experiment 1 (3 character word-like stimuli). **B:** Experiment 2 (6 character word-like stimuli). **C:** Experiment 3 (objects). **D:** Experiment 4 (Gabor patches). On each graph, the X and Y axes represent sensitivity (d') to High Bigram/Pair Deviants and Low Bigram/Pair Deviants, respectively. While each dot represents a participant, the coloured dot represents the mean performance of the group. The shaded area around the mean performance denotes the group–level 95% confidence interval. Projected on each axis, a coloured dot indicates the mean performance for the respective deviant, while the error bar reports again a 95% CI.

Comparison	BF01	*g*
3 char word-like vs Objects	4.21	0.16 [-0.36, 0.68]
3 char word-like vs Gabor patches	4.89	-0.02 [-0.56, 0.51]
6 char word-like vs Objects	3.37	0.24 [-0.27, 0.76]
6 char word-like vs Gabor patches	4.86	0.05 [-0.47, 0.58]
3 char word-like vs 3 char word-like	4.31	-0.09 [-0.68, 0.49]
Objects vs Gabor Patches	4.43	-0.16 [-0.62, 0.29]

Table 4.2: Comparison of orthographic regularities effects across experiments. All comparisons across experiments yielded BF_{01} values above 3, indicating substantial evidence in favour over the null hypothesis of no difference between experiments (relative to the hypothesis of differences between experiments). Note that all effect sizes (g) are centered around 0.

Experiment	Low Bigram/Pair Dev Hits	High Bigram/Pair Dev Hits	Standard False Alarms
1	0.78 [0.68, 0.89]	0.47 [0.33, 0.62]	0.23 [0.16, 0.30]
2	0.74 [0.66, 0.83]	0.40 [0.26, 0.53]	0.26 [0.18, 0.33]
3	0.70 [0.60, 0.81]	0.46 [0.35, 0.58]	0.25 [0.20, 0.30]
4	0.76 [0.66, 0.86]	0.45 [0.32, 0.58]	0.17 [0.11, 0.22]

Table 4.1: Hit rates and False Alarm rates for all experiments. As it can be seen, hit rates and false alarm rates are comparable across experiments.

Experiment	High Bigram/Pair Deviant *d'*	Low Bigram/Pair Deviant *d'*	*d'* difference
1	2.02 [1.44, 2.59]	0.84 [0.24, 1.45]	1.17 [0.77, 1.57]
2	1.71 [1.13, 2.29]	0.41 [-0.19, 1.00]	1.30 [0.88, 1.73]
3	1.58 [1.14, 2.03]	0.65 [0.16, 1.15]	0.93 [0.58, 1.29]
4	2.22 [1.69, 2.74]	1.01 [0.42, 1.60]	1.21 [0.79, 1.63]

Table 4.3: Sensitivity indexes (d') for all experiments

4.3 General discussion

As we stated in the introduction, the combinations of graphemes used in most writing systems to construct words, contain rich orthographic regularities (For a review, see: Chetail, 2015). Evidence suggests that the reading system might be sensitive to this orthographic regularities.

The results of Experiments 1 and 2 validate our novel experimental design, and suggest that, indeed, the visual word identification system is capable of capturing orthographic regularity. This supports theories of reading that encode some form of statistical learning, for example suggesting that one important way in which higher–level representations (e.g., words, morphemes) are learned is via letter co–occurence patterns (e.g., Davis, 2001).

Considering that reading is a recent cultural invention in mankind's history, it has been proposed that the human reading system might recycle visual processes that are domain general in nature (Dehaene and Cohen, 2007). Therefore, we asked the question of whether orthographic regularity effects observed in reading, including those we just discussed, are the result of a domain general mechanism.

We addressed this question in Experiment 3, which followed the exact same statistical structure of Experiments 1 and 2, but used renderings of 3D objects as stimuli. Contrary to orthographic–like material, the parts conforming these objects were physically connected in a unit, and arranged radially, rather than linearly. Despite these differences, the results of Experiment 3 replicate the orthographic regularity effects found in Experiments 1 and 2, using material that is clearly not orthographic in nature. The results of Experiment 3 suggest that the sensitivity to orthographic regularities that can be found in the case of script–like material extends to visual stimuli that are arguably very different. We take this to imply that the kind of statistical coding mechanisms we uncover here is the result of a domain general visual mechanism that implements statistical learning, rather than a stimulus–specific phenomenon that was carved by the unique properties of visual words.

In order to further test the generality of this mechanism, we decided to run Experiment 4, in which the stimuli were even more radically different from orthographic material. We used Gabor patches, whose defining features are not shapes, as in the case of real graphemes, the pseudofonts we used in Experiment 1 and 2, and the 3D objects we tested in Experiment 3; in fact, these defining features are not even organized into a spatial arrangement. Instead, they consist on low level visual features which are represented in primary visual cortex (Hallum et al., 2011), that is, orientation, spatial density and contrast. Despite all this, Experiment 4 results replicated the pseudofont findings once more. This suggests that the cognitive mechanisms we uncovered with this paradigm apply very widely within the visual domain, and are not related to the spatial segregation of the lower–level units.

In sum, we cannot discard the possibility that the same results we obtained in the different tasks presented here might be achieved through different processes. In a sense, this is actually quite likely in strictly mechanistic terms, given the strong stimulus specialization of the visual cortex. However, this should not take the reader's attention away from the core message of this work—whatever specific computational implementation these results are based upon, they all share the same property of being based on statistical regularities in the co–occurrence of fundamental elements, which lends support to the overall characterization of the visual identification system as a statistical learner.

In the current work, we have shown across four experiments that the human visual system is sensitive to orthographic regularities, and that this sensitivity found for co–occurrence statistics in orthographic–like material might be just one particular case of a more general sensitivity to visual statistical regularities. Furthermore, Bayesian analysis shows that not only this sensitivity is present with non–orthographic stimuli such as objects and Gabor patches, but that the magnitude of this sensitivity is actually equivalent regardless of the stimuli used.

We find the results of Experiment 4 particularly relevant. Given the primary visual cortex can be modeled as a bank of Gabor filters (Hallum et al., 2011), the visual features used to define the stimuli in Experiment 4 constitute the building blocks of visual perception. Therefore, the results of this experiment suggest that our sensitivity to orthographic regularities could be extended to any type of visual stimuli.

We consider that the evidence presented here constitutes a strong case in favor of the proposal that the human reading system recycles visual mechanisms that are domain general in nature. Rather than evolving a visual system apt for reading, we have develop writing systems that reflect the skills of our brains.

Chapter 5

General discussion

The research presented in this thesis sought to further our understanding of how do we become skilled readers. It is assumed by theoretical models of visual word recognition that the visual identification system of a successful reader supports fast and efficient access to all word components (phonology, orthography and semantics). However, these models fail to accommodate the graded nature of reading behaviours among skilled readers. In the Introduction, existing evidence was reviewed to suggest that the automaticity and efficiency of visual word recognition greatly varies within and between individuals, underscoring the importance of investigating the factors leading such differences. In this thesis, I have approached this issue by investigating how lexical knowledge is learned and refined through experience. The first two experiments presented here adopted two different approaches aiming at considering how lexical knowledge influences visual word recognition in L1 and L2. The final experiment, instead, aimed at testing whether sensitivity to orthographic regularities supports the learning of novel words, and how much these mechanisms are shared across the whole visual system.

5.0.1 Summary of the experimental findings

Chapter 2. In my first experiment I investigated morphological masked priming in L1-Italian learners of English as L2. L1 readers show true morphological priming (driver primes DRIVE) as well as morpho-orthographic priming (corner primes CORN) compared to a form condition (dialog primes DIAL). I examined this effect for Italian speakers of English to determine whether sensitivity to morphological constituents is present in their L2 and how this relates to L2 proficiency. Key findings were: 1) L2 morphological priming is qualified by L2 proficiency: genuine morphological facilitation only arises as proficiency grows; and opaque and orthographic priming shrink as L2 competence increases. 2) L2 morphological priming is modulated by sensitivity to probabilistic relationships between form and meaning as quantified by the Orthography-to-Semantic Consistency index (OSC, Marelli et al., 2015, QJEP), which interacts with proficiency as well. It was concluded that increasing command over a language is associated to qualitative changes in morphological processing and to more strongly tied orthographic and semantic representations of individual words. From this perspective, we suggest that L2 word learning might rely on the same type of mechanisms of lexical integration in L1, where pseudo–words become words, in line with Lindsay and Gaskell (2010) proposal. This brings us to the next experiment that investigate novel vocabulary learning in Italian native speakers.

Chapter 3. The experiment in this chapter investigated the learning and consolidation of novel words in adults trained via instructed vs. uninstructed learning routines. I tracked the development of novel word memory representations across three test sessions, immediately after learning at Day 0, 24hr later at Day 1, and one week after at Day 7. I employed a lexical decision task in order to measure lexical engagement, an old–new task to measure explicit recognition, and a Fast Periodic Visual Stimulation (FPVS) paradigm within an oddball design to quantify visual discrimination of newly learned words compared to untrained words and existing words. This study provided evidence of rapid word memory integration in absence of lexical competition. EEG data suggested that trained novel words (unlike untrained words) evoked neural responses similar to those for familiar words, suggesting lexical integration, and yet lexical competition effects were not observed in behaviour. Crucially, we consistently find lexical facilitation, however EEG data did not correlate with lexical decision data, suggesting that these effects tap different visual word recognition mechanisms.

Chapter 4. The last experiment tested whether readers benefit from orthographic regularities, such as letter co-occurrences, as they learn novel visual stimuli; and, more importantly, whether this is a unique feature of reading. Experiments 1 and 2, previously conducted in the lab, used pseudofont strings to show that, when faced with a visual word identification task, participants' performance is modulated by bigram frequency. This demonstrated that participants are sensitive to the frequency of co-occurrence of letters in a novel script, in the absence of semantics. Building on this evidence, two further experiments tested whether the co-occurrence effects were specific to reading. Stimuli became progressively less word-like (e.g., Y-shaped objects and Gabor patches) in order to explore the limits of the n–gram coding. Sensitivity to co-occurrence statistics was present with all stimulus types, suggesting that word recognition operates like any other visual mechanism able to capture statistical regularities in the visual environment.

5.0.2 Lexical quality and orthographic precision

The experiments within this thesis were conceived to investigate the influence of experience on visual word recognition. Collectively, the data presented in this thesis provide converging evidence that visual word recognition is qualified by lexical knowledge, probably reflecting differences in the strength of connections between lexical memories that might influence the coherence of their activation. As reviewed earlier, Perfetti and Hart (2002) identified lexical quality as precise, stable, word–specific knowledge that supports coherent activation of all components of a word's identity. On this premise, the first experiment on bilingual morphological processing confirms this theoretical contribution, and puts forward the interpretation that fully automating word identification heavily depends on a transition from an unconsolidated set of word memories, where word recognition strategies at lower L2 proficiency probably emulates novice readers (as hypothesized in the Declarative/Procedural model: Hamrick et al., 2018), to a consolidated phase where words are accessed via larger orthographic chunks, such as morphemes. Two separate proficiency measures, phonemic fluency and morphological awareness, significantly modulated readers' extraction of information in the early stages of visual word recognition—morphological masked priming appeared to be sensitive enough to show this transition. Consolidated representations also mean higher sensitivity to probabilistic ties between form and meaning as showed by OSC. These data speak in favor of the development of strongly tied lexical representations that are characterized by fully specified connections between letters and meaning, nicely paralleling the precision criterion for lexical quality of Perfetti and Hart (2002). The limitation of this experiment from this perspective is that it fails to show whether any L1 knowledge might have influenced L2 lexical access during the early stages of visual word recognition, a possibility that has been previously shown in the literature studying word cognates (e.g., Desmet and Duyck, 2007; Goldrick et al., 2016), or whether the same metalinguistic knowledge found to be critical in English (e.g., spelling; Burt and Tate, 2002; Andrews, 2012; Andrews and Lo, 2013) applies to Italian native speakers while recognizing words in their L1. Indeed, a possible interpretation is that the gradual process of refinement of lexical representations might follows a mechanism of lexical consolidation that applies to novel words in L1 as well as L2, however to what extent this relationship is based on the same cognitive mechanisms has not been addressed yet, and this thesis does not provide evidence in this respect.

5.0.3 Do linguistic differences contribute to influence orthographic precision?

The distinction between full and consolidated word representations highlights also a critical aspect of orthographic precision clearly evident from Experiment 2 results. In this experiment, novel word memories clearly show to be integrated into longer–term memory, as revealed by the FPVS electrophysiological data, but without slowing down recognition times of existing word neighbours in the lexical engagement task. This is in line with some previous studies showing that sometimes learning do not consistently produce inhibition (e.g., multiple story context in Henderson and James, 2018) (e.g., Hebb learning in Sobczak and Gaskell, 2019), but also contradicts the majority of the literature, which reported inhibition instead (e.g., Tamminen et al., 2012; Sobczak and Gaskell, 2019; Lindsay and Gaskell, 2013; Davis et al., 2009; Dumay and Gaskell, 2007; Gaskell and Dumay, 2003a). I hypothesize that the discrepancy with the literature is due to a difficult comparison between *deep vs. shallow* orthographies. It has been suggested by Andrews (2012) already that orthographic precision may not be equally important for all languages. The grain size account (Ziegler and Goswami, 2005) also assumes that the dynamics of the adult lexicon are strictly dependent on the coherence of orthography–to–phonology links. Differences in the consistency of the relationship between orthography and phonology determine the *grain size* of the smaller word units that a skilled reader needs to extract during visual word recognition (Ziegler and Goswami, 2005). This is a property that qualifies also the orthographic precision of Perfetti and Hart (2002).

Novice readers learn to specify alphabetic representations by linking consistently graphemes to phonemes (Ehri, 2005), and over time they refine this link with reading practice. In English there are many irregularities in the mapping between orthography and phonology, which are disambiguated by statistical information about the probability that a particular letter is pronounced. For example, based on whether is preceded or followed by other letters, the bigram *ea* takes many different pronunciations, as in *bread, bean.* Thus, the surrounding context reduces ambiguity in the mapping from phonology to orthography; and this relay nicely to the reader's sensitivity to statistical patterns present in the language.

This ambiguity results also in bidirectional influences between orthographic and phonological neighbours, as attested in the literature (Grainger et al., 2005; Yates et al., 2004). However, the level of ambiguity is dramatically reduced in a language with a *shallow orthography*, such as Italian, where there is high consistency between orthography and phonology, and where phonological and orthographic neighbourhood density perfectly match. Thus, from this perspective the gradual refinement of lexical representations may require different strategies of extraction of information during visual word recognition, which depends on the way the reader has learned to efficiently map orthography to phonology, which of course depends in turn onto the consistency of the orthography–to–phonology mapping. Thus, English native speakers might be motivated to rely on disambiguation mechanisms more than Italian native speakers. Consistent with the Lexical Quality Hypothesis, there might be pressure towards phonological disambiguation to achieve orthographic precision and fast recognition, but this pressure must be language dependent (e.g., English vs. Italian): words that have many similar neighbours will experience more pressure depending on the consistency of the orthographic and phonological correspondences. Languages where orthographic and phonological correspondences are inconsistently mapped might drive the development of different visual strategies than more consistent languages. As a result, I suggest the possibility that to achieve the orthographic precision required for a perfect word identification in such *deep orthography*, the English visual word recognition system is forced to stress more lexical competition mechanisms during learning novel words than Italian. From this perspective, my data provide only a hint into the lexical structures that emerge for readers of Italian vs. English—this was not the primary goal of my work. However, if we acknowledge that individuals differ as to how they process written words (e.g., Andrews, 2012; Andrews and Lo, 2013; Beyersmann et al., 2015a), comparisons of individual differences in different languages could potentially also contribute further our understanding of the skilled reading system. Thus, a more precise comparison between these two writing systems in a word learning experiment may contribute to understand how the quality of lexical memories interact with the languages characteristics.

An important limitation regarding the interpretation of lexical integration in this experiment is the absence of correlation between the FPVS and lexical engagement data. I assumed that both paradigms were measuring lexicalization from the same perspective, however there was very little evidence that this was the case. Thus, whilst both approaches inform our understanding of lexical integration and visual word recognition, they appeared to capture different aspects of this relationship, and further reasearch on this matter is clearly needed.

A second related result from this experiment is that learning induces neural changes for trained novel words that seems to be qualified by lexicality, e.g., real words vs. pseudo–words vs. pseudofonts. Coherently with the artificial nature of the learning implemented in the lab, novel word memories did not reach the same strength of activation of real words, opening the possibility that FPVS could be further used as a protocol to investigate the strength of a lexical representation from a more perceptual approach.

5.0.4 How do statistical properties contribute to novel word learning?

Clearly, skilled reading involves many steps before word identification. It has been suggested that the visual word identification system identifies recurrent letter clusters (n-grams) as a bridge between letters and words (Dehaene et al., 2005). However, the role of orthographic regularities is hotly debated. It seems that influences of orthographic regularities in visual word recognition are related to the pseudo–word domain, rather than to the word domain (for a detailed review, see: Chetail, 2015; Schmalz and Mulatti, 2017). Also, it still remains unclear whether n–gram coding could be considered a unique feature of reading, or rather derives from a binding mechanisms adopted by the visual system that groups together visual recurrent pattern independently of the visual input, like following the Gestalt principles of grouping (Pelli et al., 2009). Indeed, a quick glance into the literature on visual word recognition reveals that sensitivity to orthographic regularities is generally thought to be a *specialized* reading mechanism. This is clearly stated also by Grainger and Hannagan (2014), who hypothesizes that the visual system is a "specialized machinery that combines location-specific character detectors" (Grainger and Hannagan, 2014). This final experiment clearly shows that this is not the case, at least for what concerns letter co-occurrences. However, it also shows, intriguingly, that readers rely on orthographic regularities in tasks where the reader has to learn novel words. In contrast, the role of bigram frequency has been always contested regarding tasks within which participants had to perform lexical decisions on well known items (Chetail, 2015). Thus, I suggest that letter co-occurrences would facilitate the creation of larger orthographic chunks during learning novel words, but when a novel word is fully consolidated, other sources of information (e.g., semantics) might hinder their contribution to visual word recognition.

5.0.5 Conclusions

The work presented in this thesis investigated the overarching question of how do we become expert readers. Since reading is an acquired skill, fully understanding how skilled reading emerge requires also the investigation of the impact of a developing/not fully consolidated lexicon onto visual word recognition. This is exactly the experimental approach that I have been taken in this thesis. The experimental results clearly show that the way lexical knowledge is learned and refined through experience impacts visual word recognition. However, these experiments are not the final proof of this relationship, and more direct experimental approach can be addressed in the future.

Chapter 6

Appendices

6.1 Appendix A - Stimuli Experiment 1 (Chapter 2)

L1 – Italian

	Transparent condition	
Target	Related prime	Control prime
ARCO	arcata	melone
ARTE	artista	sottile
ASMA	asmatico	fogliame
ASTRO	astrologo	signorile
ATTO	attore	morale
BANCA	bancario	minerale
BENDA	bendaggio	minerario
CALCIO	calciatore	ventricolo
CAMPANA	campanile	variabile
CANTO	cantore	mammola
CREMA	cremoso	fertile
CUBO	cubista	tessile
DELFINO	delfinario	carotaggio
DITO	ditata	idrico
DOSE	dosaggio	camerata
ERBA	erboso	areola
FAMA	famoso	ideale
FANGO	fangoso	porcile

FARINA	farinoso	adrenale
FATO	fatale	fidata
FIENO	fienile	vettore
FORNO	fornaio	frenata
FRUSTA	frustata	ciarpame
GETTONE	gettonato	scambista
GHIACCIO	ghiacciolo	campagnolo
LEGNO	legname	puerile
MAZZA	mazzata	bombola
MITO	mitico	botola
NERVO	nervoso	turista
NOIA	noioso	sadico
OCCHIO	occhiata	naturale
ORIGINE	originario	linguaggio
ORTO	ortaggio	litorale
PAROLA	paroliere	necrotico
PENSIONE	pensionato	petroliera
POLLO	pollame	nudista
REGIA	regista	pittore
SABBIA	sabbiatura	soporifero
SANO	sanitario	alcolista
SASSO	sassata	fazione
SCHIFO	schifoso	monetario
SERVO	servile	sudista
STILE	stilista	liberale
STRADA	stradale	eleganza
TASTO	tastiera	frittata
TAVOLO	tavolata	plateale
TAXI	taxista	pontile
UGGIA	uggioso	spumame
VELLO	veliero	bravata
VETRO	vetrata	fondale
	Opaque condition	
Target	Related prime	Control prime
ABITO	abitudine	documento
ARTIGLIO	artigliere	locandiere
BALLO	ballatoio	sedimento
BILE	bilico	barile

BRIGA	brigante	revisore
CALVO	calvario	lebbroso
CARRO	carriera	fiorente
CAVIA	caviale	lunario
CERNIA	cerniera	sciabola
COLLE	collezione	parcheggio
CONO	conato	senile
COSCIA	coscienza	comunista
COSTA	costanza	pigmento
COSTO	costume	normale
DOGA	dogana	urbana
FALCO	falcata	corroso
FIRMA	firmamento	bilanciere
FORMA	formaggio	simpatico
FORZA	forziere	pompiere
FOSSO	fossile	calcolo
GARA	garante	padrone
GELO	geloso	dorato
GENERO	generoso	pazienza
GESTA	gestazione	sventurato
GOMITO	gomitolo	capienza
GRANO	granito	radioso
INDOLE	indolenza	discepolo
MAESTRA	maestranze	vivandiere
MASSO	massaggio	artistico
MATTO	mattanza	plenario
MIMO	mimosa	tisana
ORMA	ormeggio	timoroso
OSTE	ostaggio	acquario
PIETA’	pietanza	stellata
PIGNA	pignolo	festivo
QUIETE	quietanza	vituperio
RETE	retaggio	lampante
RETTA	rettile	violino
SALE	salario	formale
SERENA	serenata	volubile
SOSTA	sostanza	alleanza
STIVA	stivale	europeo
TATTO	tattico	caldaia
TEMPERA	temperanze	plafoniera

TESTA	testamento	cioccolato
TRATTO	trattore	scuderia
VANTO	vantaggio	piacevole
VENTO	ventola	pelvico
VINO	vinile	embolo
VIOLA	violenza	opinione
	Orthographic condition	
Target	Related prime	Control prime
ALBERO	albergo	istinto
AVO	avorio	patria
BANDA	bandiera	convento
BARRA	barracuda	cespuglio
BOCCA	boccia	sobria
CAMBIO	cambusa	ridosso
CAVO	cavallo	codardo
CELLA	cellula	relitto
CLAVA	clavicola	prematuro
CONGRUO	congrega	obsoleta
CORDA	cordoglio	travaglio
CORO	corallo	baruffa
CORTE	corteccia	scongiuro
FARO	faringe	omicida
GUADO	guadagno	ridicola
GUANO	guanto	stalla
LAMA	lamento	monello
LANA	lancia	radice
LENZA	lenzuola	cardiaco
LUCE	lucertola	dinosauro
LUPO	lupara	frolla
MALE	malta	riffa
MANDRIA	mandrillo	demoniaco
MANO	manto	spola
MASSA	massacro	collasso
MERCE	mercurio	castagno
META	metallo	dipinto
MUSEO	museruola	idilliaco
OBLIO	obliquo	cruento

ORDINE	ordigno	ristoro
PALLA	pallido	storico
PASSERO	passerella	salmonella
PELLE	pellicola	pagamento
PIANO	pianeta	salotto
PRODE	prodigio	prefisso
RAGGIO	raggiro	colosso
RESTO	restauro	vergogna
RISO	riserbo	ghianda
SALA	salasso	frangia
SALAME	salamandra	malaugurio
SCALO	scalogno	sonaglio
SCAMPO	scampolo	ossequio
SCIA	sciame	staffa
SOFFIO	soffitto	clausola
SPIA	spiaggia	orologio
SPINA	spinaci	litigio
SQUALO	squallido	trapianto
TRAMA	tramonto	sostegno
TRIBU'	tribuna	lattice
VELA	velcro	olezzo

L2 – English

	Transparent condition	
Target	Related prime	Control prime
ACID	acidic	yearly
ACRE	acreage	plunder
ADOPT	adopted	kingdom
AGREE	agreement	equipment
ALARM	alarming	composer
ANGEL	angelic	watcher
ARTIST	artistry	calmness
BARON	baronet	voucher
BEARD	bearded	thinker
BLOOD	bloody	active
BOMB	bomber	lessen
BULB	bulbous	leftist

CHILL	chilly	finely
CLOUD	cloudless	enactment
CREAM	creamy	watery
CRITIC	critical	tendency
DIET	dietary	wearily
DREAM	dreamer	masonry
DRUNK	drunkard	feathery
EMPLOY	employer	addition
ERUPT	eruption	vicarage
FILTH	filthy	harden
FIZZ	fizzle	touchy
FLESH	fleshy	lovers
FLOAT	floater	missive
GLOOM	gloomy	miller
GOLF	golfer	thinly
GOVERN	government	situation
GREEN	greenery	snobbish
GUILT	guilty	formal
INHIBIT	inhibitory	amateurish
LEGEND	legendary	anxiously
MARSH	marshy	thorny
MOURN	mourner	tripper
NORTH	northern	friendly
NYMPH	nymphet	acutely
OXYGEN	oxygenate	fossilise
POET	poetry	dealer
QUIET	quieten	mimicry
REACT	reaction	physical
RENEW	renewable	exemption
RISK	risky	downs
SCALD	scalding	jauntily
SOFT	soften	heroic
TEACH	teacher	finally
TOAST	toaster	wishful
TRAIN	trainee	cookery
TUFT	tufted	silken
VIEW	viewer	ranger
WIDOW	widowed	beastly
	Opaque condition	

Target	Related prime	Control prime
AMEN	amenable	palpably
AMP	ample	widen
ARCH	archer	feudal
AUDIT	audition	selfless
BOARD	boarder	factual
BRAND	brandy	safely
BRISK	brisket	foundry
BUZZ	buzzard	loyally
COAST	coaster	muffler
COUNT	country	service
COURT	courteous	developer
CRAFT	crafty	vainly
CROOK	crooked	pottery
CRYPT	cryptic	dweller
DEPART	department	production
DISC	discern	starter
EARL	early	within
FACET	facetious	distantly
FLEET	fleeting	simplify
FLICK	flicker	adviser
FRUIT	fruitless	alcoholic
GLOSS	glossary	sufferer
GLUT	gluten	bridal
GRUEL	grueling	existent
HEART	hearty	folder
HELM	helmet	brutal
INFANT	infantry	validity
INVENT	inventory	murderous
IRON	irony	sandy
LIQUID	liquidate	extremism
NUMB	number	really
ORGAN	organic	leaflet
PLAN	planet	editor
PLUCK	plucky	winger
PLUM	plumage	broiler
PUTT	putty	fishy
QUEST	question	actually

RATION	rational	steadily
SCULL	scullery	narrowly
SECRET	secretary	obviously
SIGN	signet	frosty
SNIP	sniper	hourly
SPLINT	splinter	idealism
STILT	stilted	gaseous
THICK	thicket	scruffy
TREAT	treaty	angler
TROLL	trolley	naughty
TRUMP	trumpet	chatter
UNIT	united	others
WHISK	whisker	coyness
	Orthographic condition	
Target	Related prime	Control prime
AGAIN	against	perhaps
APPEND	appendix	believer
ARSE	arsenal	timidly
BASIL	basilisk	benignly
BROTH	brothel	warfare
BUTT	button	prayer
CANDID	candidacy	epileptic
COLON	colonel	ability
COMMA	command	equally
DEMON	demonstrate	instruction
DIAL	dialog	lately
ELECT	electron	suburban
ETHER	ethereal	rumbling
EXTRA	extract	justify
FORCE	forceps	prudish
FREE	freeze	golden
FUSE	fuselage	citation
GALA	galaxy	keeper
GLAD	glade	cuffs
HEAVE	heaven	firmly
INTERN	internation	revolutionary
INVEST	investigate	anaesthetic
JERK	jerkin	twisty

NEIGH	neighbour	struggled
PARENT	parenthesis	lectureship
PHONE	phonetic	dreadful
PLAIN	plaintiff	absurdity
PLUS	plush	filmy
PUB	public	gently
PULP	pulpit	gifted
QUART	quartz	roller
RABBI	rabbit	weekly
SCRAP	scrape	ninety
SHOVE	shovel	tricky
SHUN	shunt	itchy
SIGH	sight	happy
SMUG	smuggle	twelfth
SQUAW	squawk	oddity
STAMP	stampede	defector
STIR	stirrup	buoyant
STUB	stubborn	moisture
STUN	stunt	misty
SURF	surface	medical
SURGE	surgeon	novelty
TACT	tactile	spindly
TEXT	textile	booklet
TWIN	twinkle	cheaply
TWIT	twitch	lesser
VILLA	villain	grossly
WEIR	weird	manly

6.2 Appendix B - Stimuli Experiment 2 (Chapter 3)

List of stimuli to be learned

List	Italian base word	Experimental novel word	baseWord
A	algebra	algebba	yes
A	amnesia	omnesia	yes
A	amuleto	amulero	yes
A	aneddoto	aseddoto	yes
A	anguilla	inguilla	yes

A	anguria	unguria	yes
A	antenna	antensa	yes
A	apostolo	apostomo	yes
A	astuccio	ascuccio	yes
A	biscotto	bisconto	yes
A	cellula	nellula	yes
A	ciuffo	ciaffo	yes
A	collegio	collepio	yes
A	commedia	commegia	yes
A	corazza	corazia	yes
A	dirupo	difupo	yes
A	forfora	fordora	yes
A	formica	forzica	yes
A	gorilla	gerilla	yes
A	guscio	buscio	yes
A	meringa	mefinga	yes
A	nicotina	nicopina	yes
A	ostrica	osbrica	yes
A	palude	paluge	yes
A	pioppo	pieppo	yes
A	sostegno	sosbegno	yes
A	autunno	autinno	yes
A	tartufo	tartubo	yes
A	unguento	inguento	yes
A	vassoio	rassoio	yes
A	NA	balpusta	no
A	NA	batole	no
A	NA	bomuparo	no
A	NA	buersa	no
A	NA	catalle	no
A	NA	ettodria	no
A	NA	fastero	no
A	NA	fomirto	no
A	NA	frilmodo	no
A	NA	frutirca	no
A	NA	gadodo	no
A	NA	grancile	no
A	NA	mastenza	no
A	NA	scagro	no
A	NA	scapippo	no

A	NA	sentesto	no
A	NA	spitana	no
A	NA	suomio	no
A	NA	tenopo	no
A	NA	tullordi	no
B	agguato	anguato	yes
B	analogia	enalogia	yes
B	aquilone	aquigone	yes
B	augurio	augubio	yes
B	tariffa	tariffu	yes
B	bottega	bottefa	yes
B	cannolo	gannolo	yes
B	carogna	tarogna	yes
B	chiosco	chiosto	yes
B	cicogna	ciconna	yes
B	delirio	melirio	yes
B	docile	dofile	yes
B	edicola	edicota	yes
B	equatore	equatoce	yes
B	falange	favange	yes
B	galassia	gabassia	yes
B	giraffa	gisaffa	yes
B	gregge	bregge	yes
B	iguana	igiana	yes
B	marsupio	garsupio	yes
B	mensola	menpola	yes
B	ombelico	ombelizo	yes
B	ospizio	ospigio	yes
B	pilastro	picastro	yes
B	piovra	pievra	yes
B	spiaggia	spiaglia	yes
B	tovaglia	novaglia	yes
B	trofeo	trofuo	yes
B	tulipano	tulifano	yes
B	vescovo	descovo	yes
B	NA	averuzzo	no
B	NA	bagate	no
B	NA	bandicci	no
B	NA	bosigene	no
B	NA	buensi	no

B	NA	cafugno	no
B	NA	cortario	no
B	NA	ettinita	no
B	NA	facusta	no
B	NA	forissi	no
B	NA	friborre	no
B	NA	frunicio	no
B	NA	gateri	no
B	NA	matodice	no
B	NA	scatra	no
B	NA	selgicro	no
B	NA	spriage	no
B	NA	suteni	no
B	NA	telica	no
B	NA	vimitodi	no

Bibliography

Abutalebi, J., 2008. Neural aspects of second language representation and language control. Acta psychologica 128, 466–478.

Amenta, S., Crepaldi, D., 2012. Morphological processing as we know it: an analytical review of morphological effects in visual word identification. Frontiers in psychology 3, 232.

Amenta, S., Crepaldi, D., Marelli, M., submitted. Consistency measures individuate dissociating semantic activation networks in priming paradigms: A new look on semantics in the processing of (complex) words. Quarterly Journal of Experimental Psychology .

Amenta, S., Marelli, M., Sulpizio, S., 2017. From sound to meaning: Phonology-to-semantics mapping in visual word recognition. Psychonomic bulletin & review 24, 887–893.

Andrews, S., 2008. Lexical expertise and reading skill. Psychology of learning and motivation 49, 247–281.

Andrews, S., 2012. Individual differences in skilled visual word recognition and reading: The role of lexical quality: Sally andrews, in: Visual Word Recognition Volume 2. Psychology Press, pp. 163–184.

Andrews, S., Hersch, J., 2010. Lexical precision in skilled readers: Individual differences in masked neighbor priming. Journal of Experimental Psychology: General 139, 299.

Andrews, S., Lo, S., 2012. Not all skilled readers have cracked the code: Individual differences in masked form priming. Journal of Experimental Psychology: Learning, Memory, and Cognition 38, 152.

Andrews, S., Lo, S., 2013. Is morphological priming stronger for transparent than opaque words? it depends on individual differences in spelling and vocabulary. Journal of Memory and Language 68, 279–296.

Andrews, S., Miller, B., Rayner, K., 2004. Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. European Journal of Cognitive Psychology 16, 285–311.

Armstrong, B.C., Frost, R., Christiansen, M.H., 2016. The long road of statistical learning research: past, present and future. Philosophical Transactions of the Royal Society B: Biological Sciences 372, 1–4. URL: `http://rstb.royalsocietypublishing.org/content/372/1711/20160047.abstract`, doi:`10.1098/rstb.2016.0047`.

Arnbak, E., Elbro, C., 2000. The effects of morphological awareness training on the reading and spelling skills of young dyslexics. Scandinavian Journal of Educational Research 44, 229–251.

Ashby, J., Rayner, K., Clifton, C., 2005. Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. The Quarterly Journal of Experimental Psychology Section A 58, 1065–1086.

Aslin, R.N., Saffran, J.R., Newport, E.L., 1999. Statistical learning in linguistic and nonlinguistic domains. The emergence of language , 359–380.

Baayen, R.H., 2013. languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". URL: `https://CRAN.R-project.org/package=languageR`. r package version 1.4.1.

Baayen, R.H., Milin, P., Durdevic, D.F., Hendrix, P., Marelli, M., 2010. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning .

Baker, C.I., Behrmann, M., Olson, C.R., 2002. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. Nature neuroscience 5, 1210.

Bartko, J.J., 1966. The intraclass correlation coefficient as a measure of reliability. Psychological reports 19, 3–11.

Basnight-Brown, D.M., Chen, L., Hua, S., Kostić, A., Feldman, L.B., 2007. Monolingual and bilingual recognition of regular and irregular english verbs: Sensitivity to form similarity varies with first language experience. Journal of memory and language 57, 65–80.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67, 1–48. doi:`10.18637/jss.v067.i01`.

Beck, M.L., 1997. Regular verbs, past tense and frequency: Tracking down a potential source of ns/nns competence differences. Second Language Research , 93–115.

Beyersmann, E., Casalis, S., Ziegler, J.C., Grainger, J., 2015a. Language proficiency and morpho-orthographic segmentation. Psychonomic bulletin & review 22, 1054–1061.

Beyersmann, E., Castles, A., Coltheart, M., 2012. Morphological processing during visual word recognition in developing readers: Evidence from masked priming. The Quarterly Journal of Experimental Psychology 65, 1306–1326.

Beyersmann, E., Grainger, J., Casalis, S., Ziegler, J.C., 2015b. Effects of reading proficiency on embedded stem priming in primary school children. Journal of Experimental Child Psychology 139, 115–126.

Binder, J.R., McKiernan, K., Parsons, M., Westbury, C., Possing, E.T., Kaufman, J., Buchanan, L., 2003. Neural correlates of lexical access during visual word recognition. Journal of cognitive neuroscience 15, 372–393.

Bowers, J.S., Davis, C.J., Hanley, D.A., 2005. Interfering neighbours: The impact of novel word learning on the identification of visually similar words. Cognition 97, B45–B54.

Brainard, D., 1997. The Psychophysics Toolbox. Spatial Vision 10, 433–436.

Brysbaert, M., Mandera, P., McCormick, S.F., Keuleers, E., 2018. Word prevalence norms for 62,000 english lemmas. Behavior research methods , 1–13.

Burani, C., Laudanna, A., 1992. Units of representation for derived words in the lexicon, in: Advances in psychology. Elsevier. volume 94, pp. 361–376.

Burani, C., Salmaso, D., Caramazza, A., 1984. Morphological structure and lexical access. Visible Language 18, 342–352.

Burt, J.S., Tate, H., 2002. Does a reading lexicon provide orthographic representations for spelling? Journal of Memory and Language 46, 518–543.

Butterworth, B., 1983. Lexical representation. na.

Caramazza, A., Laudanna, A., Romani, C., 1988. Lexical access and inflectional morphology. Cognition 28, 297–332.

Carlisle, J.F., 2003. Morphology matters in learning to read: A commentary. Reading Psychology 24, 291–322.

Carlisle, J.F., 2004. Morphological processes that influence learning to read. Handbook of language and literacy: Development and disorders , 318–339.

Carr, T.H., 1999. Trying to understand reading and dyslexia. mental chronometry, individual differences, cognitive neuroscience, and. Converging methods for understanding reading and dyslexia , 459.

Carreiras, M., Grainger, J., 2004. Sublexical representations and the 'front end'of visual word recognition. Language and Cognitive Processes 19, 321–331.

Casalis, S., Colé, P., Sopo, D., 2004. Morphological awareness in developmental dyslexia. Annals of dyslexia 54, 114–138.

Castles, A., Davis, C., Cavalot, P., Forster, K., 2007. Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. Journal of Experimental Child Psychology 97, 165–182.

Chetail, F., 2015. Reconsidering the role of orthographic redundancy in visual word recognition. Frontiers in psychology 6, 645.

Chetail, F., 2017. What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. Cognition 163, 103–120. URL: `http://dx.doi.org/10.1016/j.cognition.2017.02.015`, doi:`10.1016/j.cognition.2017.02.015`.

Clahsen, H., Felser, C., 2006. Grammatical processing in language learners. Applied Psycholinguistics 27, 3–42.

Clahsen, H., Felser, C., 2018. Some notes on the shallow structure hypothesis. Studies in Second Language Acquisition 40, 693–706.

Clahsen, H., Felser, C., Neubauer, K., Sato, M., Silva, R., 2010. Morphological structure in native and nonnative language processing. Language Learning 60, 21–43.

Cluff, M.S., Luce, P.A., 1990. Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. Journal of Experimental Psychology: Human Perception and Performance 16, 551.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., Ziegler, J., 2001. Drc: a dual route cascaded model of visual word recognition and reading aloud. Psychological review 108, 204.

Community, B.O., 2017. Blender - a 3D modelling and rendering package. URL: https://www.blender.org/.

Conrad, N.J., Harris, N., Williams, J., 2013. Individual differences in children's literacy development: The contribution of orthographic knowledge. Reading and Writing 26, 1223–1239.

Conway, C.M., Christiansen, M.H., 2005. Modality-constrained statistical learning of tactile, visual, and auditory sequences. Journal of Experimental Psychology: Learning, Memory, and Cognition 31, 24.

Coughlin, C.E., Fiorentino, R., Royle, P., Steinhauer, K., 2019. Sensitivity to inflectional morphology in a non-native language: evidence from erps. Frontiers in Communication 4, 21.

Coutanche, M.N., Thompson-Schill, S.L., 2014. Fast mapping rapidly integrates information into existing memory networks. Journal of Experimental Psychology: General 143, 2296.

Crepaldi, D., Mandera, P., Keuleers, E., Brysbaert, M., 2013. Subtlex-it: Aa new frequency list for italian based on movie subtitles. none .

Crepaldi, D., Rastle, K., Coltheart, M., Nickels, L., 2010. 'fell'primes 'fall', but does 'bell'prime 'ball'? masked priming with irregularly-inflected primes. Journal of memory and language 63, 83–99.

Dahan, D., Tanenhaus, M.K., 2005. Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. Psychonomic bulletin & review 12, 453–459.

Davis, C.J., 2001. The self-organising lexical acquisition and recognition (SOLAR) model of visual word recognition. Ph.D. thesis. ProQuest Information & Learning.

Davis, C.J., Lupker, S.J., 2006a. Masked inhibitory priming in english: Evidence for lexical inhibition. Journal of Experimental Psychology: Human Perception and Performance 32, 668.

Davis, C.J., Lupker, S.J., 2006b. Masked inhibitory priming in english: Evidence for lexical inhibition. Journal of Experimental Psychology: Human Perception and Performance 32, 668.

Davis, M.H., Di Betta, A.M., Macdonald, M.J., Gaskell, M.G., 2009. Learning and consolidation of novel spoken words. Journal of cognitive neuroscience 21, 803–820.

Davis, M.H., Gaskell, M.G., 2009. A complementary systems account of word learning: neural and behavioural evidence. Philosophical Transactions of the Royal Society B: Biological Sciences 364, 3773–3800.

Davis, M.H., Rastle, K., 2010. Form and meaning in early morphological processing: Comment on feldman, o'connor, and moscoso del prado martín (2009). Psychonomic bulletin & review 17, 749–755.

Dawson, N., Rastle, K., Ricketts, J., 2017. Morphological effects in visual word recognition: Children, adolescents and adults. Journal of experimental psychology. Learning, memory, and cognition .

Deacon, S.H., Kirby, J.R., 2004. Morphological awareness: Just "more phonological"? the roles of morphological and phonological awareness in reading development. Applied psycholinguistics 25, 223–238.

Dehaene, S., 2009. Reading in the brain: The new science of how we read. Penguin.

Dehaene, S., Cohen, L., 2007. Cultural recycling of cortical maps. Neuron 56, 384–398.

Dehaene, S., Cohen, L., Sigman, M., Vinckier, F., 2005. The neural code for written words: a proposal. Trends in cognitive sciences 9, 335–341.

Dehaene, S., Pegado, F., Braga, L.W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., Cohen, L., 2010. How learning to read changes the cortical networks for vision and language. science 330, 1359–1364.

Delorme, A., Makeig, S., 2004. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. Journal of neuroscience methods 134, 9–21.

Desmet, T., Duyck, W., 2007. Bilingual language processing. Language and linguistics compass 1, 168–194.

Diependaele, K., Duñabeitia, J.A., Morris, J., Keuleers, E., 2011. Fast morphological effects in first and second language word recognition. Journal of Memory and Language 64, 344–358.

Diependaele, K., Sandra, D., Grainger, J., 2009. Semantic transparency and masked morphological priming: The case of prefixed words. Memory & Cognition 37, 895–908.

Dumay, N., Gaskell, M.G., 2007. Sleep-associated changes in the mental representation of spoken words. Psychological Science 18, 35–39.

Dumay, N., Gaskell, M.G., 2012. Overnight lexical consolidation revealed by speech segmentation. Cognition 123, 119–132.

Ehri, L.C., 2005. Learning to read words: Theory, findings, and issues. Scientific Studies of reading 9, 167–188.

Ehri, L.C., Wilce, L.S., 1985. Movement into reading: Is the first stage of printed word learning visual or phonetic? Reading Research Quarterly , 163–179.

Endress, A.D., Mehler, J., 2009. The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. Journal of Memory and Language 60, 351–367.

Feldman, L.B., Kostić, A., Basnight-Brown, D.M., UREVIĆ, D.F., Pastizzo, M.J., 2010. Morphological facilitation for regular and irregular verb formations in native and non-native speakers: Little evidence for two distinct mechanisms. Bilingualism: Language and Cognition 13, 119–135.

Feldman, L.B., Kostić, A., Gvozdenović, V., O'Connor, P.A., del Prado Martín, F.M., 2012. Semantic similarity influences early morphological priming in serbian: A challenge to form-then-meaning accounts of word recognition. Psychonomic bulletin & review 19, 668–676.

Feldman, L.B., O'Connor, P.A., del Prado Martín, F.M., 2009. Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. Psychonomic Bulletin & Review 16, 684–691.

Fernandes, T., Kolinsky, R., Ventura, P., 2009. The metamorphosis of the statistical segmentation output: Lexicalization during artificial language learning. Cognition 112, 349–366.

Fiser, J., Aslin, R.N., 2002. Statistical learning of new visual feature combinations by infants. Proceedings of the National Academy of Sciences 99, 15822–15826.

Forster, K.I., Davis, C., Schoknecht, C., Carter, R., 1987. Masked priming with graphemically related forms: Repetition or partial activation? The Quarterly Journal of Experimental Psychology Section A 39, 211–251.

Forster, K.I., Veres, C., 1998. The prime lexicality effect: Form-priming as a function of prime awareness, lexical status, and discrimination difficulty. Journal of Experimental Psychology: Learning, Memory, and Cognition 24, 498.

Fox, J., Hong, J., 2009. Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. Journal of Statistical Software 32, 1–24. URL: http://www.jstatsoft.org/v32/i01/.

Gamer, M., Lemon, J., ¡puspendra.pusp22@gmail.com¿, I.F.P.S., 2019. irr: Various Coefficients of Interrater Reliability and Agreement. URL: https://CRAN.R-project.org/package=irr. r package version 0.84.1.

Gaskell, M.G., Dumay, N., 2003a. Effects of vocabulary acquisition on lexical competition in speech perception and production, in: Proceedings of the 15th ICPhS conference, Adelaide, Australia: Causal Productions.. pp. 1485–1488.

Gaskell, M.G., Dumay, N., 2003b. Lexical competition and the acquisition of novel words. Cognition 89, 105–132.

Gibson, J.J., 1950. The perception of the visual world. .

Goldinger, S.D., Luce, P.A., Pisoni, D.B., 1989. Priming lexical neighbors of spoken words: Effects of competition and inhibition. Journal of memory and language 28, 501–518.

Goldrick, M., Putnam, M., Schwarz, L., 2016. Coactivation in bilingual grammars: A computational account of code mixing. Bilingualism: Language and Cognition 19, 857–876.

Gonnerman, L.M., Seidenberg, M.S., Andersen, E.S., 2007. Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. Journal of experimental psychology: General 136, 323.

Gough, P.B., Hillinger, M.L., 1980. Learning to read: An unnatural act. Annals of Dyslexia 30, 179–196.

Grainger, J., Colé, P., Segui, J., 1991. Masked morphological priming in visual word recognition. Journal of memory and language 30, 370.

Grainger, J., Dufau, S., Montant, M., Ziegler, J.C., Fagot, J., 2012. Orthographic processing in baboons (papio papio). Science 336, 245–248.

Grainger, J., Hannagan, T., 2014. What is special about orthographic processing? Written Language & Literacy 17, 225–252.

Grainger, J., Jacobs, A.M., 1996. Orthographic processing in visual word recognition: a multiple read-out model. Psychological review 103, 518.

Grainger, J., Muneaux, M., Farioli, F., Ziegler, J.C., 2005. Effects of phonological and orthographic neighbourhood density interact in visual word recognition. The Quarterly Journal of Experimental Psychology Section A 58, 981–998.

Grosjean, F., 2010. Bilingual. Harvard university press.

Hallum, L.E., Landy, M.S., Heeger, D.J., 2011. Human primary visual cortex (V1) is selective for second-order spatial frequency. Journal of Neurophysiology 105, 2121–2131. doi:10.1152/jn.01007.2010.

Hamrick, P., Lum, J.A., Ullman, M.T., 2018. Child first language and adult second language are both tied to general-purpose learning systems. Proceedings of the National Academy of Sciences 115, 1487–1492.

Hasenäcker, J., Schröter, P., Schroeder, S., 2017. Investigating developmental trajectories of morphemes as reading units in german. Journal of Experimental Psychology: Learning, Memory, and Cognition 43, 1093.

Hawkins, E.A., 2015. The influence of meaning and memory consolidation on novel word learning .

Henderson, L.M., James, E., 2018. Consolidating new words from repetitive versus multiple stories: Prior knowledge matters. Journal of experimental child psychology 166, 465–484.

Henry, J.D., Crawford, J.R., 2004. A meta-analytic review of verbal fluency performance following focal cortical lesions. Neuropsychology 18, 284.

Hentschke, H., Stüttgen, M.C., 2011. Computation of measures of effect size for neuroscience data sets. European Journal of Neuroscience 34, 1887–1894. doi:10.1111/j.1460-9568.2011.07902.x.

Hersch, J., Andrews, S., 2012. Lexical quality and reading skill: Bottom-up and top-down contributions to sentence processing. Scientific Studies of Reading 16, 240–262.

Hervais-Adelman, A., Kumar, U., Mishra, R.K., Tripathi, V.N., Guleria, A., Singh, J.P., Eisner, F., Huettig, F., 2019. Learning to read recycles visual cortical networks without destruction. Science Advances 5, eaax0262. URL: http://advances.sciencemag.org/lookup/doi/10.1126/sciadv.aax0262, doi:10.1126/sciadv.aax0262.

Heyer, V., Clahsen, H., 2015. Late bilinguals see a scan in scanner and in scandal: dissecting formal overlap from morphological priming in the processing of derived words. Bilingualism: Language and Cognition 18, 543–550.

Hildebrandt, N., 1994. The reicher-wheeler effect and models of deep and phonological dyslexia. Journal of neurolinguistics 8, 1–18.

Huettig, F., Altmann, G.T., 2007. Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. Visual Cognition 15, 985–1018.

James, E., Gaskell, M.G., Weighall, A., Henderson, L., 2017. Consolidation of vocabulary during sleep: The rich get richer? Neuroscience & Biobehavioral Reviews 77, 1–13.

Johnson, J.S., Shenkman, K.D., Newport, E.L., Medin, D.L., 1996. Indeterminacy in the grammar of adult language learners. Journal of Memory and Language 35, 335–352.

Kassambara, A., 2019. ggpubr: 'ggplot2' Based Publication Ready Plots. URL: https://CRAN.R-project.org/package=ggpubr. r package version 0.2.1.

van Kesteren, M.T., Rijpkema, M., Ruiter, D.J., Fernández, G., 2010. Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. Journal of Neuroscience 30, 15888–15894.

van Kesteren, M.T., Rijpkema, M., Ruiter, D.J., Fernández, G., 2013. Consolidation differentially modulates schema effects on memory for items and associations. PloS one 8, e56155.

Keuleers, E., Stevens, M., Mandera, P., Brysbaert, M., 2015. Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. The Quarterly Journal of Experimental Psychology 68, 1665–1692.

Kirkham, N.Z., Slemmer, J.A., Johnson, S.P., 2002. Visual statistical learning in infancy: Evidence for a domain general learning mechanism. Cognition 83, B35–B42.

Kirkici, B., Clahsen, H., 2013. Inflection and derivation in native and non-native language processing: Masked priming experiments on turkish. Bilingualism: Language and Cognition 16, 776–791.

Kleiner, M., Brainard, D., Pelli, D., 2007. What's new in psychtoolbox-3? .

Koda, K., 2000. Cross-linguistic variations in l2 morphological awareness. Applied psycholinguistics 21, 297–320.

Kuznetsova, A., Bruun Brockhoff, P., Haubo Bojesen Christensen, R., 2016. lmerTest: Tests in Linear Mixed Effects Models. URL: `https://CRAN.R-project.org/package=lmerTest`. r package version 2.0-33.

Lakens, D., 2013. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Frontiers in Psychology 4, 1–12. doi:`10.3389/fpsyg.2013.00863`, `arXiv:9605103`.

Landauer, T.K., Dumais, S.T., 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological review 104, 211.

Landauer, T.K., Streeter, L.A., 1973. Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. Journal of Verbal Learning and Verbal Behavior 12, 119–131.

Lavric, A., Clapp, A., Rastle, K., 2007. Erp evidence of morphological analysis from orthography: A masked priming study. Journal of Cognitive Neuroscience 19, 866–877.

Leach, L., Samuel, A.G., 2007. Lexical configuration and lexical engagement: When adults learn new words. Cognitive psychology 55, 306–353.

Libben, G., Jarema, G., 2006. The representation and processing of compound words. Oxford University Press on Demand.

Lindsay, S., Gaskell, M.G., 2010. A complementary systems account of word learning in l1 and l2. Language Learning 60, 45–63.

Lindsay, S., Gaskell, M.G., 2013. Lexical integration of novel words without sleep. Journal of Experimental Psychology: Learning, Memory, and Cognition 39, 608.

Liu-Shuang, J., Torfs, K., Rossion, B., 2016. An objective electrophysiological marker of face individualisation impairment in acquired prosopagnosia with fast periodic visual stimulation. Neuropsychologia 83, 100–113.

Lochy, A., Jacques, C., Maillard, L., Colnat-Coulbois, S., Rossion, B., Jonas, J., 2018. Selective visual representation of letters and words in the left ventral occipito-temporal cortex with intracerebral recordings. Pnas 32, 1–10. doi:10.1073/pnas.1718987115.

Lochy, A., Van Belle, G., Rossion, B., 2015. A robust index of lexical representation in the left occipito-temporal cortex as evidenced by eeg responses to fast periodic visual stimulation. Neuropsychologia 66, 18–31.

Lochy, A., Van Reybroeck, M., Rossion, B., 2016. Left cortical specialization for visual letter strings predicts rudimentary knowledge of letter-sound association in preschoolers. Proceedings of the National Academy of Sciences 113, 8544–8549.

Longtin, C.M., Segui, J., Halle, P.A., 2003. Morphological priming without morphological relationship. Language and cognitive processes 18, 313–334.

Macmillan, N., Creelman, D., 2005. Detection theory: A user's guide. Detection Theory. .

Marelli, M., Amenta, S., 2018. A database of orthography-semantics consistency (osc) estimates for 15,017 english words. Behavior research methods 50, 1482–1495.

Marelli, M., Amenta, S., Crepaldi, D., 2015. Semantic transparency in free stems: The effect of orthography-semantics consistency on word recognition. The Quarterly Journal of Experimental Psychology 68, 1571–1583.

Marslen-Wilson, W., 1990. Activation, competition, and frequency in lexical access. .

McBride-Chang, C., Shu, H., Zhou, A., Wat, C.P., Wagner, R.K., 2003. Morphological awareness uniquely predicts young children's chinese character recognition. Journal of educational psychology 95, 743.

McClelland, J.L., 2013. Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. Journal of Experimental Psychology: General 142, 1190.

McClelland, J.L., McNaughton, B.L., O'Reilly, R.C., 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychological review 102, 419.

McClelland, J.L., Rumelhart, D.E., 1981. An interactive activation model of context effects in letter perception: I. an account of basic findings. Psychological review 88, 375.

McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychological methods 1, 30.

McKenzie, S., Frank, A.J., Kinsky, N.R., Porter, B., Rivière, P.D., Eichenbaum, H., 2014. Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. Neuron 83, 202–215.

Merkx, M., Rastle, K., Davis, M.H., 2011. The acquisition of morphological knowledge investigated through artificial language learning. The Quarterly Journal of Experimental Psychology 64, 1200–1220.

Meyer, D.E., Schvaneveldt, R.W., 1971. Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. Journal of experimental psychology 90, 227.

Murray, W.S., Forster, K.I., 2004. Serial mechanisms in lexical access: the rank hypothesis. Psychological Review 111, 721.

Nagy, W.E., Herman, P.A., Anderson, R.C., 1985. Learning words from context. Reading research quarterly , 233–253.

Nation, P., Waring, R., 1997. Vocabulary size, text coverage and word lists. Vocabulary: Description, acquisition and pedagogy 14, 6–19.

Neubauer, K., Clahsen, H., 2009. Decomposition of inflected words in a second language: An experimental study of german participles. Studies in Second Language Acquisition 31, 403–435.

Norcia, A.M., Appelbaum, L.G., Ales, J.M., Cottereau, B.R., Rossion, B., 2015a. The steady-state visual evoked potential in vision research: a review. Journal of vision 15, 4–4.

Norcia, A.M., Appelbaum, L.G., Ales, J.M., Cottereau, B.R., Rossion, B., 2015b. The steady-state visual evoked potential in vision research: a review. Journal of vision 15, 4–4.

Norris, D., 2006. The bayesian reader: Explaining word recognition as an optimal bayesian decision process. Psychological review 113, 327.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the edinburgh inventory. Neuropsychologia 9, 97–113.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S.F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., et al., 2000. A cultural effect on brain function. Nature neuroscience 3, 91.

Pegado, F., Bekinschtein, T., Chausson, N., Dehaene, S., Cohen, L., Naccache, L., 2010. Probing the lifetimes of auditory novelty detection processes. Neuropsychologia 48, 3145–54. URL: `http://dx.doi.org/10.1016/j.neuropsychologia.2010.06.030http://www.ncbi.nlm.nih.gov/pubmed/20600179`, doi:`10.1016/j.neuropsychologia.2010.06.030`.

Pegado, F., Comerlato, E., Ventura, F., Jobert, A., Nakamura, K., Buiatti, M., Ventura, P., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., Braga, L.W., Cohen, L., Dehaene, S., 2014. Timing the impact of literacy on visual processing. Proceedings of the National Academy of Sciences of the United States of America 111, E5233–E5242. doi:`10.1073/pnas.1417347111`.

Peirce, J.W., 2007. Psychopy—psychophysics software in python. Journal of neuroscience methods 162, 8–13.

Pelli, D.G., Majaj, N.J., Raizman, N., Christian, C.J., Kim, E., Palomares, M.C., 2009. Grouping in object recognition: The role of a gestalt law in letter identification. Cognitive neuropsychology 26, 36–49.

Perfetti, C., 2007. Reading ability: Lexical quality to comprehension. Scientific studies of reading 11, 357–383.

Perfetti, C.A., Hart, L., 2002. The lexical quality hypothesis. Precursors of functional literacy 11, 67–86.

Perfetti, C.A., McCutchen, D., 1982. Speech processes in reading, in: Speech and language. Elsevier. volume 7, pp. 237–269.

Perry, C., Ziegler, J.C., Zorzi, M., 2007. Nested incremental modeling in the development of computational theories: the cdp+ model of reading aloud. Psychological review 114, 273.

Pinker, S., 2009. How the mind works (1997/2009) .

Plaut, D.C., McClelland, J.L., Seidenberg, M.S., Patterson, K., 1996. Understanding normal and impaired word reading: computational principles in quasi-regular domains. Psychological review 103, 56.

Prévost, P., White, L., 2000. Missing surface inflection or impairment in second language acquisition? evidence from tense and agreement. Second language research 16, 103–133.

Pugh, K.R., Mencl, W.E., Jenner, A.R., Lee, J.R., Katz, L., Frost, S.J., Shaywitz, S.E., Shaywitz, B.A., 2001. Neuroimaging studies of reading development and reading disability. Learning Disabilities Research & Practice 16, 240–249.

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org. ISBN 3-900051-07-0.

Rajalingham, R., Kar, K., Sanghavi, S., Dehaene, S., DiCarlo, J.J., 2019. A potential cortical precursor of visual word form recognition in untrained monkeys. bioRxiv , 739649URL: https://www.biorxiv.org/content/10.1101/739649v2, doi:10.1101/739649.

Rastle, K., Davis, M.H., 2008. Morphological decomposition based on the analysis of orthography. Language and Cognitive Processes 23, 942–971.

Rastle, K., Davis, M.H., Marslen-Wilson, W.D., Tyler, L.K., 2000. Morphological and semantic effects in visual word recognition: A time-course study. Language and cognitive processes 15, 507–537.

Rastle, K., Davis, M.H., New, B., 2004. The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. Psychonomic Bulletin & Review 11, 1090–1098.

Regan, D., 1968. Chromatic adaptation and steady-state evoked potentials. Vision research 8, 149–158.

Regan, D., 1989. Orientation discrimination for objects defined by relative motion and objects defined by luminance contrast. Vision research 29, 1389–1400.

Reicher, G.M., 1969. Perceptual recognition as a function of meaningfulness of stimulus material. Journal of experimental psychology 81, 275.

Robinson, P., 2003. The cognitive hypothesis, task design, and adult task-based language learning. University of Hawai'I Second Langauge Studies Paper 21 (2) .

Rosa-Salva, O., Fiser, J., Versace, E., Dolci, C., Chehaimi, S., Santolin, C., Vallortigara, G., 2018. Spontaneous learning of visual structures in domestic chicks. Animals 8, 135.

Rossion, B., 2014. Understanding individual face discrimination by means of fast periodic visual stimulation. Experimental brain research 232, 1599–1621.

Rossion, B., Boremanse, A., 2011. Robust sensitivity to facial identity in the right human occipito-temporal cortex as revealed by steady-state visual-evoked potentials. Journal of Vision 11, 16–16.

Rothe, J., Schulte-Körne, G., Ise, E., 2014. Does sensitivity to orthographic regularities influence reading and spelling acquisition? a 1-year prospective study. Reading and Writing 27, 1141–1161.

Ruff, R.M., Light, R.H., Parker, S.B., Levin, H.S., 1997. The psychological construct of word fluency. Brain and language 57, 394–405.

Rugg, M.D., 1985. The effects of semantic priming and word repetition on event-related potentials. Psychophysiology 22, 642–647.

Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month-old infants. Science 274, 1926–1928.

Saffran, J.R., Johnson, E.K., Aslin, R.N., Newport, E.L., 1999. Statistical learning of tone sequences by human infants and adults. Cognition 70, 27–52.

Scarf, D., Boy, K., Reinert, A.U., Devine, J., Güntürkün, O., Colombo, M., 2016. Orthographic processing in pigeons (columba livia). Proceedings of the National Academy of Sciences 113, 11272–11276.

Schiff, R., Raveh, M., Fighel, A., 2012. The development of the hebrew mental lexicon: When morphological representations become devoid of their meaning. Scientific Studies of Reading 16, 383–403.

Schmalz, X., Mulatti, C., 2017. Busting a myth with the bayes factor. The Mental Lexicon 12, 263–282.

Schmidtke, D., Matsuki, K., Kuperman, V., 2017. Surviving blind decomposition: A distributional analysis of the time-course of complex word recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition 43, 1793.

Seidenberg, M.S., McClelland, J.L., 1989. A distributed, developmental model of word recognition and naming. Psychological review 96, 523.

Selfridge, O.G., 1958. Pandemonium: A paradigm for learning. the Mechanisation of Thought Processes, 1958 .

Sherman, R.A., 2015. multicon: Multivariate Constructs. URL: https://CRAN.R-project.org/package=multicon. r package version 1.6.

Silva, R., Clahsen, H., 2008. Morphologically complex words in l1 and l2 processing: Evidence from masked priming experiments in english. Bilingualism: Language and cognition 11, 245–260.

Slowiaczek, L.M., Nusbaum, H.C., Pisono, D.B., 1987. Phonological priming in auditory word recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition 13, 64.

Sobczak, J.M., Gaskell, M.G., 2019. Implicit versus explicit mechanisms of vocabulary learning and consolidation. Journal of Memory and Language 106, 1–17.

Stanislaw, H., 1999. Calculation of signal detection theory measures. Behavior Research Methods, Instruments, & Computers 3, 37–149. doi:10.3758/BF03207704.

Stone, G.O., Van Orden, G.C., 1994. Building a resonance framework for word recognition using design and system principles. Journal of Experimental Psychology: Human Perception and Performance 20, 1248.

Szmalec, A., Page, M.P., Duyck, W., 2012. The development of long-term lexical representations through hebb repetition learning. Journal of Memory and Language 67, 342–354.

Taft, M., 1994. Interactive-activation as a framework for understanding morphological processing. Language and cognitive processes 9, 271–294.

Taft, M., Forster, K.I., 1975. Lexical storage and retrieval of prefixed words. Journal of verbal learning and verbal behavior 14, 638–647.

Taft, M., Forster, K.I., 1976. Lexical storage and retrieval of polymorphemic and polysyllabic words. Journal of verbal learning and verbal behavior 15, 607–620.

Tamminen, J., Davis, M.H., Merkx, M., Rastle, K., 2012. The role of memory consolidation in generalisation of new linguistic information. Cognition 125, 107–112.

Tamminen, J., Gaskell, M.G., 2008. Newly learned spoken words show long-term lexical competition effects. The Quarterly Journal of Experimental Psychology 61, 361–371.

Tamminen, J., Payne, J.D., Stickgold, R., Wamsley, E.J., Gaskell, M.G., 2010. Sleep spindle activity is associated with the integration of new memories and existing knowledge. Journal of Neuroscience 30, 14356–14360.

Taylor, J., Davis, H.M., Rastle, K., 2019. Mapping visual symbols onto spoken language along the ventral visual stream. Proceedings of the National Academy of Sciences , 384–398.

Taylor, J., Rastle, K., Davis, M.H., 2013. Can cognitive models explain brain activation during word and pseudoword reading? a meta-analysis of 36 neuroimaging studies. Psychological bulletin 139, 766.

Tse, D., Langston, R.F., Kakeyama, M., Bethus, I., Spooner, P.A., Wood, E.R., Witter, M.P., Morris, R.G., 2007. Schemas and memory consolidation. Science 316, 76–82.

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H., Morris, R.G., 2011. Schema-dependent gene activation and memory encoding in neocortex. Science 333, 891–895.

Ullman, M.T., 2001. The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. Bilingualism: Language and cognition 4, 105–122.

Ullman, M.T., 2005. A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. Mind and context in adult second language acquisition , 141–178.

Van Heuven, W.J., Mandera, P., Keuleers, E., Brysbaert, M., 2014. Subtlex-uk: A new and improved word frequency database for british english. The Quarterly Journal of Experimental Psychology 67, 1176–1190.

VanPatten, B., Benati, A.G., 2015. Key terms in second language acquisition. Bloomsbury Publishing.

Ventura, P., 2014. Let's face it: Reading acquisition, face and word processing. Frontiers in Psychology 5, 10–13. doi:10.3389/fpsyg.2014.00787.

Veríssimo, J., Heyer, V., Jacob, G., Clahsen, H., 2017. Selective effects of age of acquisition on morphological priming: Evidence for a sensitive period. Language Acquisition , 1–12.

Vidal, C., Content, A., Chetail, F., 2017. Bacs: The brussels artificial character sets for studies in cognitive psychology and neuroscience. Behavior research methods 49, 2093–2112.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J.P., Sigman, M., Cohen, L., 2007. Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. Neuron 55, 143–156.

Vitevitch, M.S., Luce, P.A., 1998. When words compete: Levels of processing in perception of spoken words. Psychological science 9, 325–329.

Walker, S., Henderson, L.M., Fletcher, F.E., Knowland, V.C.P., Cairney, S.A., Gaskell, M.G., 2019. Learning to live with interfering neighbours: the influence of time of learning and level of encoding on word learning. Royal Society Open Science 6, 181842.

Wheeler, D.D., 1970. Processes in word recognition. Cognitive Psychology 1, 59–85. doi:10.1016/0010-0285(70)90005-8.

Whiteside, D.M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M.R., Roper, B., 2016. Verbal fluency: Language or executive function measure? Applied Neuropsychology: Adult 23, 29–34.

Yap, M.J., Balota, D.A., Sibley, D.E., Ratcliff, R., 2012. Individual differences in visual word recognition: Insights from the english lexicon project.

Journal of Experimental Psychology: Human Perception and Performance 38, 53.

Yap, M.J., Tse, C.S., Balota, D.A., 2009. Individual differences in the joint effects of semantic priming and word frequency revealed by rt distributional analyses: The role of lexical integrity. Journal of Memory and Language 61, 303–325.

Yates, M., Locker, L., Simpson, G.B., 2004. The influence of phonological neighborhood on visual word perception. Psychonomic Bulletin & Review 11, 452–457.

Ziegler, J.C., Goswami, U., 2005. Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. Psychological bulletin 131, 3.

Ziegler, J.C., Muneaux, M., Grainger, J., 2003. Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. Journal of Memory and Language 48, 779–793.