SISSA
Scuola Internazionale Superiore di Studi Avanzati

Neuroscience Area - PhD course in
Cognitive Neuroscience

The impact of irregularities on the learning of words and letters

Candidate:
Olga Šolaja

Advisor:
Prof. Davide Crepaldi

Academic Year 2022-23

MA PER SEGUIR VIRTUTE E CANOSCENZA · SISSA ·

Abstract

Nature surrounds us with an intricate web of patterns and regularities. As natural systems evolve, so do the regularities inherent to them, all in order to achieve maximum efficiency. Humans' perceptual system is exceptionally good at picking up regularities across different domains, and seamlessly integrating them with previously assimilated knowledge. In the present thesis, I explore the impact of regularities on various aspects of novel word learning and the development of orthographic representations. We learn new words every day as literate adults; however, it is still unclear how this occurs. To explore this question, the first two studies I carried out examined the advantage morphology – as a prime example of a regular, frequent, and informative letter chunk – might have in comparison to other letter chunks that may lack either frequency or informativeness. In the third study I adopt a somewhat opposite perspective regarding the potential contribution of regularity in the learning process, and explore the possibility that irregular visual input may be more effective in eliciting the development of abstract orthographic representations.

Contents

Chapter 1

General introduction

Learning to communicate in written language may seem as an incredibly laborious process, especially when we think of how much time children need to start reading or writing fluently. Looking more into the details of what language processing entails, it appears an even greater achievement. For example, in order to correctly read a written word, a learner has to engage in visual processing, successfully identify all the letters, and access its meaning. Learning an unknown word requires an even more complex set of processes beyond this. However, once the language system has been established and consolidated, our language skills allow us to use language effortlessly: adult skilled readers are able to recognize words in terms of milliseconds. Even learning novel words typically occurs with ease, often without much attentional engagement. This can be linked to the incredible sensitivity of the human perceptual system to different regularities. Throughout life, we continuously pick up regularities from the environment that surrounds us and seamlessly integrate them into the existing web of patterns that we carry.

In this thesis, I focus on two domains of language learning – word learning and orthographic learning – and how are these processes affected by the different types of regularities (or lack thereof) inherent to them. In the domain of word learning, the focus is firstly on morphology, which represents a prime example of regularity thanks to its consistent relationship between form and meaning. Secondly, I explored to which extent the contribution of such language regularities (i.e., affixes) in the word learning process can be reduced to more basic statistical learning principles (e.g., frequency of letter co-occurrence), which seem to be a part of many aspects of language learning. Finally, I assume an opposite view in the third study, by exploring how irregular input contributes to development of abstract orthographic representations.

1.1 Visual word recognition

1.1.1 Orthographic processing

Orthographic processing refers to the cognitive mechanism involved in recognizing the written form of language. It is a fundamental aspect of reading and language comprehension, involving the recognition and understanding of the visual representations of words and their spelling patterns. Research in orthographic processing has mainly focused on letter and word level processing, in the attempt to connect low-level visual processes with higher level cognitive processes. Within this framework, a written word is perceived as both a visual and linguistic entity. Orthographic processing operates precisely at this crossroads, enabling the integration of fundamental visual processing with the specialized linguistic processing, which is specific to word stimuli, as opposed to other visual object forms. For orthographic processing to be successful, it is essential that the orthographic representation contains information about letter identity and letter position. Letter identity allows for the differentiation between words such as `KISS` and `MISS`, while letter position is crucial for distinguishing words like `DOG` from `GOD` (see Grainger, 2018, for a review).

Mature orthographic representations are abstract, meaning that they don't rely on letter-to-letter reading, as evidence from above also suggests, but also that they are stable and tolerate noise in the visual input (Grainger, 2018). As a result, proficient readers instantly recognize symbols **A**, **a**, 𝒶 and $\mathscr{A}$, as referring to the same phoneme-grapheme combination. Indeed, Dehaene et al. (2005) in their Local Combination Detector (LCD) neuronal model of reading present their now widely accepted claims that specialized neuron layers, which respond to the letter A, also get activated by various versions of the same letter, supporting the idea of abstract letter representations in the brain. According to the LCD model, the process of learning to read involves developing a sophisticated neural system for visual word recognition. This system comprises neurons that are each specialized to encode different components of written language, ranging in complexity from single letters to bigrams, and ultimately to entire words, as detailed by Vinckier et al. (2007). Initially, these components are encoded in a manner akin to any other visual stimuli. However, through repeated exposure to written language, neurons adapt to preferentially recognize these components, leading to the formation of what are known as abstract orthographic representations.

One of the key characteristics of abstract orthographic representations is their stability and resistance to noise (Grainger, 2018). Indeed, readers demonstrate remarkable proficiency in interpreting highly distorted text, such as CAPTCHAs (Completely Automated Public Turing test to tell Computers and Humans Apart (Ahn et al., 2003)). For example, Hannagan et al. (2012) assessed the impact of identity priming (using either identical or unrelated primes) on printed target words in a masked priming lexical decision task. The primes were either presented in their original printed form or altered to resemble CAPTCHA distortions. Although the priming effect was not as pronounced with distorted primes as with printed ones, a significant identity priming effect was still observed. This suggests that due to our extensive exposure to words with varied visual forms, our letter detectors have adapted to tolerate distortions by recognizing a multitude of letter shapes. Consequently, this allows for the identification of a wide array of potential visual representations for any given letter.

A very common source of noise in the written input is the handwritten text. However, since both sensorimotor activity and variability are integral to handwriting, it is challenging to distinguish their contributions in establishing orthographic representations (see Araujo et al., 2022; Fernandes & Araújo, 2021, for recent reviews). With the aim to disentangle these effects, Li and James, 2016) investigated the impact of different training materials and methods on learning Greek letters in children. They contrasted different types of training (handwriting, typing, tracing) and noise in input (handwritten, single font, multiple fonts). They found that learning with variable input, irrespective of training type, enhanced the ability to categorize letters correctly in subsequent tests. This suggests that recognizing various instances of a letter requires forming a perceptual category that encompasses different versions of that letter, a concept dating back to Gibson et al. (1962), who proposed that children learn to identify defining features of specific symbol categories early in development.

There are two markers specific of orthographic processing, which sets it apart from visual recognition of other visual forms (such as numbers, punctuation marks, and other elements). One of these markers is location-invariant processing. This process refers to the mechanism that codes for relative positions of letters within a word and is most commonly explored using the same-different matching task (Ratcliff, 1981; Krueger, 1978), which yields a so-called transposed-letter effect (Perea & Lupker, 2003, 2004; see Grainger, 2018 for a review). In this paradigm, the participant's task is to decide if the two strings that appear one after the other on the screen are the same or different. The crucial manipulation is in the different condition, where two letters can be either transposed (JUDGE – JUGDE) or replaced (JUDGE – JUDPE). A proficient reader is insensitive to the position of the letters, which results in slower and more error prone responses to the transposed compared to replaced condition. Even though this effect is present also with symbols (e.g., # $ % ^ - # % $ ^ induces more mistakes than # $ % ^ - # $ * ^) (Gomez et al., 2008; Norris & Kinoshita, 2012), it is notably smaller (Duñabeitia et al., 2012; Massol et al., 2013).

The second marker of orthographic processing is location-specific processing, which entails parallel processing of letters within a word. To study this mechanism, target-in-string task identification task based on Reicher (1969) and Wheeler (1970) task is used. Applied to orthographic processing (Tydgat & Grainger, 2009), in this task participants are first shown a string of five characters and are instructed to concentrate on the central character. After the string is displayed, a pattern mask appears, with each mask covering one character. A letter is then shown above and below each mask position. Results show that when the strings are letters, participants are more accurate in identifying the first, last, and middle letters, forming a W-shaped pattern of accuracy. However, with non-letter strings, the accuracy is highest in the middle position, creating a Λ-shaped pattern.

Recently, Fernández-López et al. (2020) made an attempt to recreate the process of development of orthographic representations in a controlled lab environment, i.e., they explored whether acquiring a new script affected location-invariant and location-specific processing. Participants underwent training in one of two unfamiliar scripts, each consisting of 11 printed BACS characters (C. Vidal et al., 2017), spread over six sessions. In the trained (experimental) alphabet, each written character was associated with a phonological sound. The participants were taught to proficiently read and write, with the training encompassing a comprehensive mix of handwriting, listening, and reading exercises. The untrained script served as a visual control, with the training focused solely on recognizing the visual forms of the letters. To assess the effects of the training, participants administered same different matching task and target-in-string task, in order to test for location-invariant processing and location-specific processing, respectively. Results did not show any evidence for location-invariant or location-specific processing, as the performance was quite similar before and after the training, thus suggesting that the orthographic representations were not very robust.

Thus, it seems that the specific coding of letters and words develops more thorough experience and exposure. Considering that reading is a relatively recent invention in human culture, its core representations and processing are likely rooted in general visual perception. This interpretation aligns also with the above mentioned findings of transposed-letter effect in other visual symbols (Duñabeitia et al., 2012; Gomez et al., 2008; Massol et al., 2013; Norris & Kinoshita, 2012). Recent work by Y. Vidal et al. (2021) attempted to reconcile these apparently different mechanisms in a study which compared sensitivity to co-occurrence of features in three types of visual forms: orthographic-like stimuli, made up 3D objects, and sinusoidal gratings. They found that that participants were sensitive to the patterns of co-occurrence across all three types of visual objects. This indicates that the mechanisms involved in visual recognition of novel letter strings support processing of other visual objects, further supporting the idea that letter and word-specific coding evolves with increased familiarity with orthographic material.

1.1.2 Morphological processing

Morphological processing refers to the cognitive and neural mechanisms involved in recognizing and understanding the structure and meaning of morphemes, the smallest units of meaning in language. A distinctive characteristic of morphemes is that they bring together the form and meaning of word units as regularities in language. While assignment of meaning in language is largely arbitrary (e.g., there is no concrete reason why the word *mug* was assigned to a recipient for a drink while *plate* is a recipient for food, and not vice versa), morphological structure breaks this arbitrariness by establishing one-to-one relationship between a morpheme form and its meaning. For example, the suffix *-ist* always creates a noun that describes a person who is associated with a certain profession or activity (e.g., *pianist*, *activist*, *dentist*). Thus, it is clear that suffix itself already carries some meaning which is then associated with the meaning of the stem to create a final product: a novel morphologically complex word, with clear morphological structure.

Morphologically complex words have repeatedly been shown to have important influence on visual word recognition (e.g., Beyersmann et al., 2020, 2016; Grainger & Beyersmann, 2021; Leminen et al., 2016; for review see Amenta & Crepaldi, 2012; Leminen et al., 2019). In a seminal study, Taft and Forster (1975) offered first evidence for the so-called affix-stripping accounts.

In a masked priming lexical decision paradigm, they contrasted facilitation obtained from morphologically complex nonwords with an existing stem such as *dejuvenate* to nonwords that contain a nonword stem, such as *depertoire*. They found that items with an existing stem are more difficult to reject as nonwords than those that contain a nonexisting stem. This suggests that the word has undergone decomposition into a stem and affix, followed by the activation of their meanings. Since Taft and Forster's seminal work, there has been a proliferation of studies employing morpheme interference effect paradigm in nonwords to study nature of morphological decomposition, as it offers a rather intuitive way to observe the power of morphological decomposition (e.g., Burani, Marcolini, De Luca, & Zoccolotti, 2008; Burani et al., 2002; Caramazza et al., 1988; Crepaldi et al., 2010; Dawson et al., 2018). For example, Burani et al. (2002) demonstrated that nonwords composed of stems and suffixes (e.g., *woman-ist*) were more frequently classified as possible words in a lexical decision task, and they were also named with greater speed and accuracy compared to corresponding nonwords lacking suffixes (e.g., *woman-ost*) in the naming task. While the sensitivity to morphemes can depend on reading proficiency (Beyersmann, Casalis, et al., 2015; Beyersmann, Grainger, et al., 2015), it is clear that morphemes have an important role in processing unknown words. Another finding in favor of morphological decomposition is that the stem frequency has a role in the process: in a lexical decision task, New et al. (2004) observed faster response times for the for the items that contain a stem of higher frequency than those with lower frequency.

The aforementioned literature strongly indicates that morphemes have a specific role in word processing. However, many studies have now shown that the observed effects might not be due only to morphological content of the affixes, but also to the frequency of the letter clusters. Evidence supporting this idea comes from a series of studies (e.g., Beyersmann et al., 2016; Longtin et al., 2003; Rastle et al., 2004; for a review see Rastle & Davis, 2008). These studies demonstrate that words such as *corner*, where *-er* does not act as a genuine suffix, facilitate processing of the stem corn in masked primed lexical decision task to a similar extent as genuinely suffixed primes (e.g., *farmer* – *farm*). Importantly, it has been shown that morphological decomposition effects are dependent on the position within the nonword: suffixes are processed as morphemes only when they appear in their typical position (e.g. *gasful*) but not in an incorrect position (e.g. *fulgas*) (Crepaldi et al., 2010).

This body of research has led to the development of several key models in complex word processing, among which the full-parsing and dual-route models are particularly prominent. Full-parsing models are based on the idea that morphologically complex words have to be decomposed into stem and affix, which are both contained in the lexicon (Taft et al., 1986; Taft & Forster, 1975; Taft & Nguyen-Hoan, 2010). On the other hand, dual-route models, as proposed by Grainger and Ziegler (2011), present two routes for the input to be processed. According to this perspective, one route involves processing and storing the complex word in its entirety, while the other involves decomposing the word into its morphological components, much like the full-parsing model. The activation of the route depends on a number of factors, including word frequency, familiarity, transparency, and the intermediate lemma level (Crepaldi, Rastle, Coltheart, et al., 2010; Diependaele et al., 2009; Taft, 1994). Recently, Beyersmann and Grainger (2023) introduced a novel perspective that integrates the latest findings in this field. Their model focuses on the full-decomposition of pseudo-suffixed words, like *corner*, underscoring the significant role that stems play in complex word processing (Grainger & Beyersmann, 2017). They propose that the activation of the embedded word, such as *corn* in *corner*, can be explained by the parallel activation of affix and edge-aligned embedded word. It is important to note that the process of embedded word activation is non-morphological, as stems often exist as independent words and do not require morphological decomposition in order to process their meaning. A key divergence of this recent model from its predecessors is its emphasis on the spatial positioning of word parts: the space on the edge of a word serves as anchor point for the encoding of letter position. This is evidenced by the observed priming for edge-aligned (*pimebook-BOOK*), but not for mid-embedded (*pibookme-BOOK*) or outer-embedded constituents (*bopimeok-BOOK*) (Beyersmann et al., 2018).

1.2 Word learning

Word learning is something that we do on a daily basis, throughout life, even as proficient speakers of a language (Keuleers et al., 2015; Leach & Samuel, 2007; Saffran et al., 1997). Thus, the mental lexicon is constantly changing. In order to become a part of it, a new word has to undergo the process of lexicalization. Once the word has been lexicalized, it will interact with the other words in the lexicon (Leach & Samuel, 2007). For example, activation of the representation for "rat" will lead to activation and possibly subsequent inhibition of the representation of "cat" according to most theoretical models of lexical processing (e.g., Cohort (Marslen-Wilson, 1990), TRACE, (McClelland & Elman, 1986), Shortlist, (Norris, 1994), Neighborhood Activation Model (Luce & Pisoni, 1998), Interactive Activation (McClelland & Rumelhart, 1981), Dual Route Cascaded Model (Coltheart et al., 2001)). In other words, the word has to compete for activation with other neighboring lexical items.

In their seminal study in the domain of novel word learning, Gaskell and Dumay (2003) make use of the lexical competition phenomenon to explore the formation of new lexical representations. Their participants learnt new words such as *cathedruke*, derived from the base word *cathedral*. The idea was that, if the new word has developed a lexical representation, it should enter in competition with the lexical units within its lexical neighborhood, including *cathedral*. In the lexical decision task this would be evident from a delay in the processing of the base word *cathedral* due to activation of the novel neighbor/competitor. An important point to emphasize here is that the new words did not have any meaning associated with them, since the training was a simple phoneme-monitoring task, which deviates from what is usually considered to be a "lexical item" (Gaskell & Dumay, 2003). The results showed that lexicalization does not occur immediately after training, but only after three days after the initial exposure. This study paved the way to the rich body of research that explored different aspects of learning and lexicalization, including studies on how the new words get integrated into the memory system, as well as the impact of the consolidation period and sleep on this process (e.g., Bakker et al., 2015; Davis & Gaskell, 2009; Gaskell et al., 2019; Gaskell & Ellis, 2009; Henderson et al., 2015; James et al., 2017, 2018; Mirkovic et al., 2019; Sobczak & Gaskell, 2019; Takashima et al., 2017; Tamminen et al., 2012; Walker et al., 2020; Weighall et al., 2016).

Importantly, there has been a move towards exploring the role of semantics (e.g., Bakker et al., 2015; Gaskell et al., 2019; Takashima et al., 2017). This aligns with the primary purpose of language acquisition in real-life situations, which is to allow communication. For example, Bakker et al. (2015) investigated whether lexical and/or semantic N400 and later positive component (LPC – typically observed with following N400 response to word stimuli) change with offline consolidation. Participants learnt two sets of novel words and their definitions during two consecutive days. The novel words were preceded by semantically related or unrelated existing word primes. In the testing phase, participants' EEG responses were recorded during a semantic categorization task, in which the target novel word could be from the first or the second learning set – thus assessing the effect of consolidation period. In addition, the target was preceded either by a related or unrelated prime. The results showed that recently learned novel words behaved like pseudowords, exhibiting a N400 effect, whereas consolidated novel words patterned more closely with existing words. However, some semantic effects were observed when preceded by semantically related primes, eliciting a more positive LPC response (a semantic-priming effect), both before and after consolidation. This suggests that certain semantic effects can be observed even when words have not been fully lexicalized.

Another noteworthy shift from the Gaskell and Dumay's initial paradigm involves a transition toward more naturalistic learning contexts. In these studies, participants learned novel words from sentence or text contexts, mirroring real-life scenarios, instead of artificial paradigms such as phoneme monitoring (e.g., Ginestet, Shadbolt, et al., 2020; Ginestet, Valdois, et al., 2020; Henderson et al., 2015; Henderson & James, 2017; Joseph et al., 2014; Joseph & Nation, 2018; Pagán & Nation, 2019). Furthermore, adults predominantly acquire new vocabulary through reading (Nagy et al., 1985), thus eye-tracking paradigms offer a particular advantage here. Eye tracking allows exploring learning as it happens, on line, eliminating the need to postpone assessment until after the learning process. For example, in an eye-tracking study Pagán and Nation (2019) explored how word learning depends on contextual diversity (among other factors, e.g., retrieval practice). By comparing eye-movements before, during and after learning, together with behavioral results they showed that learning from diverse contexts is more effortful, but also more successful in an immediate posttest.

Yet another line of research went to explore the role of morphology in word learning. When confronted with a new word, a reader has two sources of information based on which they uncover the meaning of the unknown word: the context in which the word appears, and the structure of the word itself (i.e., its morphological structure). As it has been illustrated in section 1.1.2., people are very sensitive to morphology; thus, there is a reason to believe it may play a role in word learning. In this context, Tamminen et al. (2015) explored how does suffix learning occur. In a series of word learning experiments, participants were introduced to novel words that consisted of an existing word stem and a new suffix (e.g., *crabafe*). After an extensive training where they read the definitions of the novel words (e.g. if *-afe* designates a place, *crabafe* is a zoo building with exotic crab species and *gunafe* is the section of an armory where one can find a gun), participants were tested on different aspects of learning. The results showed that when speeded online language processing is required, such as in a semantic priming task, generalization was achieved only after a period of memory consolidation. In addition, this required a sufficient number of unique exemplars in the training. Knowledge was not generalized unless one affix meaning was consolidated before introducing inconsistencies. Conversely, in tasks that required slow, deliberate reasoning, generalization largely did not depend on the above constraints. (Havas et al., 2015) focused on the same problem, the learning of novel suffixes, but they used word-picture pair paradigm. In this experiment, each picture represented an animal marked for gender (e.g., a monkey wearing a skirt is female, and a monkey wearing pants is male). Under the picture, a novel complex word was written, consisting of a nonword stem and an affix that marks for gender (e.g., *elu-ro* – male, *elu-mo* – female). When tested, participants were able to recognize the items from their training even when presented alongside distractors. Additionally, they were able to apply the novel gender-marking system to unfamiliar word stems in the generalization task. Importantly, in this study there was no consolidation period, which suggests that participants were ready to use the newly acquired morphological knowledge immediately after learning.

It's important to note that in the studies above, training was extensive and explicit (although no information was given about suffixes), which stands in sharp contrast with the implicit nature of real-life word learning from the context. A recent study by (Ginestet, Shadbolt, et al., 2020) explored novel complex word learning in a more naturalistic setting, using eye-tracking. The training comprised of reading short stories that had novel words embedded in them, while the eye-movements were tracked. Each story consisted of five sentences. As example would be: "Mike has to go back outside during the storm and his hat gets wet. Mike decides to use the *relurber* to fix his problem. The handy *relurber* will make his hat dry again. Mike is glad the *relurber* can dry his hat. Mike makes sure to put the *relurber* away when he is finished.”. Novel words were composed of a nonword stem, an existing prefix and an existing affix (e.g. *re-lurb-er*), or nonmorphological chunks instead of the affixes (e.g. *pe-lurb-le*). Overall, the results showed that nonmorphological words were fixated more often than complex words. Moreover, nonmorphological words exhibited longer gaze durations at the beginning of the training, but this difference was dramatically reduced after four encounters. Finally, on the spelling test after the training, complex words were spelled more accurately compared to their nonmorphological counterparts. Thus, it seems that morphology affects reading behavior during the learning of novel morphologically complex words.

1.3 Statistical learning

In the studies described in the previous section on word learning (1.2.), the focus has always been on chunks of letters, either as whole words (Gaskell & Dumay, 2003) or parts of words (e.g., Ginestet et al., 2020; Havas et al., 2015; Tamminen et al., 2015). However, research in the domain of statistical learning (SL) has shown that the units typically studied in language, such as words or morphemes, can be described also in terms of statistical regularities within the unit itself, as well as in the environment that surrounds it. This is not unique for language: humans are extremely sensitive to such patterns in all domains of perception, they pick them up rapidly and they are able to use them very shortly after exposure.

In the context of language, statistical learning began to gain prominence first in the field of language acquisition in infants. In their seminal study, Saffran et al. (1996) exposed 8-month infants to a 2-minute continuous speech stream of concatenated nonsense words. Each word consisted of three syllables (e.g., *bi/da/ku/pa/do/ti/ go/la/bu/bi/da/ku*), and the only way to uncover the structure (i.e., to identify word boundaries) was to rely on the statistical structure of the stream. This is because although the words were repeated in a random order, syllables within them followed a fixed structure. Therefore, the only cue to the word boundaries were the transitional probabilities (TP): they were higher between the syllable pairs (TP = 1) than between words (TP = 0.33). The results showed that infants are indeed able to segment the continuous stream into words, as shown by their ability to differentiate between the syllable orderings heard in the stream and the novel ones encountered in the test. Moreover, they were able to recognize items consisting of the last syllable and the first syllable of the words (thus containing the word boundary, for example as in *prettybaby* → *tyba*), as incorrect.

This study led to the proliferation of statistical learning research in different aspects of language and perception, and their intersection. For example, to translate Saffan et al.'s (1996) design from the auditory domain into vision, researchers explored the human capacity to compute the probabilities of co-occurrence of simple abstract shapes. With this aim, Fiser and Aslin (2001) exposed their participants to configurations consisting of multiple such shapes (see 1.1). After the exposure phase, participants were shown the pairs that co-occurred during the exposure and distractor pairs (i.e., pairs of shapes that were never presented together). Similarly to Saffran et al. (1996), they found that participants extracted the shape co-occurrence probabilities since they were able to distinguish between the trained items and distractors. This result has been replicated in many studies (for a recent review see Frost et al., 2019), and also in different modes of presentation. For example, instead of having a grid with shapes like in the Fiser and Aslin (2001) study, some studies (e.g. Kirkham et al., 2002; Siegelman et al., 2017; Siegelman & Frost, 2015; Turk-Browne et al., 2005) used a continuous stream of shapes. In this design, participants are required to categorize shapes into groups based on whether one shape can follow another. Therefore, it is essential to keep track of temporal sequencies.

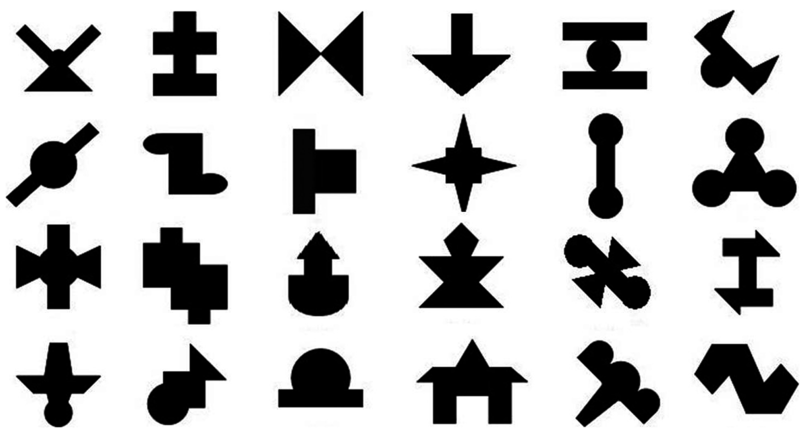
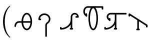
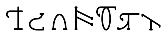
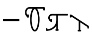


Figure 1.1: Typical shapes used to create sequences in visual statistical learning paradigms.

Reading represents an intriguing phenomenon in the context of statistical learning since it incorporates visual object processing, statistical learning, and language: in order to decode a linguistic message, a reader has to recognize letter forms which embed both linguistic patterns and visual regularities. There is a rich literature showing readers' ability to extract all kinds of statistical regularities from the written input, including frequency of letters and words (Brysbaert et al., 2018; New & Grainger, 2011), letter co-occurrence (Chetail et al., 2015), position of letters within a word (Crepaldi et al., 2010), the relationship between letters and semantics in morphology (Marelli et al., 2015; Ulicheva et al., 2020), and predictability of words in a sentence (Ashby et al., 2005; Smith & Levy, 2013).

Moreover, it seems that some of the effects that are typically found with linguistic elements, such as morphemes, can largely be replicated by using only statistical information. For example, Lelonkiewicz et al. (2020) trained participants using pseudoletter strings (, ), where each string contained an affix-like chunk (). These chunks were affix-like only in the sense that they were embedded in several items and their position was consistent across items. In the test phase, participants were presented with novel items, some of which included the trained affixes. Results showed that participants categorized the test items as belonging to the training set more often if they contained a trained affix. In this experiment, learning was necessarily based on visual statistical cues alone, as no orthographic, phonological, or semantic cues were involved. Building on this, Lelonkiewicz et al. (2023) found that the inclusion of actual linguistic elements, such as using real letters and associating letter strings with objects, led to a stronger learning effect. In a similar vein, Vidal et al. (2021) tested whether bigram frequency effect, typical for orthographic processing, can be replicated with nonlinguistic materials. As described in section 1.1.1., pg. 9, they found that this effect can be observed even in visual processing of novel visual objects and Gabor patches. These findings suggest that statistical learning is inherent to structuring the linguistic input, probably because it is inherent in the visual mechanisms that get recruited for reading when we become literate.

Since statistical learning seems to permeate multiple sensory perception systems, it is plausible to conceive it as a general-purpose mechanism that guides learning across different domains of perception. Based on this, one might expect that the correlation of the statistical learning skills of an individual in one domain would highly correlate with their SL skills in another domain. There is also evidence of correlations between SL and other cognitive skills, such as reading or musical skill (Arciuli & Simpson, 2012; Qi et al., 2018). Compatible with this idea, it was found that SL skills measured both in the auditory (Rao Mandikal Vasuki et al., 2017) and visual modality (Arciuli & Simpson, 2012) predicted reading skills. Similarly, Siegelman et al. (2018) found significant correlations between SL visual and auditory tasks, when the prior linguistic knowledge was not involved. Such correlations are reflected also in the neural activity: the same subset of brain regions is activated during SL tasks across modalities and stimuli (see Frost et al., 2015 for a review). However, as argued in Bogaerts et al. (2022), these correlations could be a consequence of high similarity in the statistical patterns they employ, such as pairs or triplets within a continuous sensory stream. Thus, the correlations between tasks that involve learning pairs and triplets of any kind of stimuli might reflect the uniform design of the tasks used to test the SL skill, not how SL works in the natural environment. Another contributing factor could be the fact that most studies used a two-alternative forced-choice task as a test, which involves a similar meta-cognitive decision processing independently of the type of stimuli that have been learned (Christiansen, 2019). As for the neural evidence goes, the observed activation of the medial temporal lobe memory system (Schapiro et al., 2012, 2017) may primarily reflect hippocampal activation inherent to learning a limited set of patterns. Such a learning process does not necessarily generalize to continuous tracking and assimilation of statistical information in the natural environment. Moreover, correlations with other cognitive measures are weak, with many reports of null results and are potentially at least in part driven by other factors such as attention and decision making strategies (see Bogaerts et al., 2022).

The question of domain generality vs. specific processing is tied to the issue of individual differences within the statistical learning research framework. In the context of language, one may hypothesize that good statistical learners would also be good language learners. While there is some evidence supporting such relationship (e.g., Siegelman et al., 2020), the strength of this relationship is variable. Furthermore, there is clear evidence that it is influenced by the task, as well as by the measures used to assess proficiency (e.g., Schmalz et al., 2019; Siegelman et al., 2017). One of the challenges in assessing statistical learning ability is that while the tests typically used are reliable at the group level, they lack reliability at the individual level (Siegelman et al., 2017; Siegelman & Frost, 2015). However, recent efforts to improve the reliability of such tests, such as improved standard visual statistical learning tasks (e.g., Siegelman et al., 2017) and the Statistically Induced Chunking Recall (SICR) task (Isbilen et al., 2020), have offered encouraging results. Finally, learning statistical regularities in a statistical learning task differs from learning regularities in real life setting. Thus, statistical learning research would benefit from developing more nuanced paradigms that reflect more accurately learning in real world, which implies regularities that vary in their reliability, that are presented concurrently, and assimilated over long time-scales (Siegelman et al., 2020).

1.4 Research questions

Experiment 1:

The literature on visual word recognition and morphological processing seems to agree that affixes represent a rather specific unit of language, in terms of both form and meaning regularity. In the word learning framework, this might represent an advantage compared to other word parts that do not carry any meaning. Importantly, previous literature has not addressed an important feature of affixes – the fact that they are often highly frequent letter chunks. The question of how word structure impacts on word learning is highly relevant given that vocabulary grows throughout the lifetime, and that as adults we mostly learn morphologically complex words. To fill this gap, Experiment 1 aims to disentangle effects due to letter chunk frequency and those due to genuine morphology in a semantically rich context. I hypothesize that if suffixes are more helpful in learning, this will be reflected in a sharper reduction of fixation times upon repeated encounters with items that contain suffixes (*flibness*), compared to those that have non-meaningful endings, independently whether these endings are frequent (*fliban*) or rare (*flibov*). In addition, I expected this advantage to emerge also in post-learning tests. Finally, if learning the stem of a word (e.g., as in ”teach” within ”teacher”) strongly depends on morphological decomposition, it is predicted that affixed words specifically trigger this phenomenon.

Experiment 2:

Humans are exceptional in picking up regularities in their environment. Studies focused on statistical learning have shown that assimilating statistical regularities is an important part of language learning in different domains. As mentioned above, affixes represent a highly regular unit; however, their facilitatory role might be due both to the semantics carried by the suffix or by the high frequency of the letter combination that constitutes it. Thus, this information could be at least partially statistical in nature, which led me to the second study of this thesis. In this experiment, I propose that if statistical learning underlies the acquisition of novel words, participants who are relatively stronger on the former should also be relatively stronger on the latter. If this correlation is mostly due to meaning, it should show up specifically with suffixed words (*flibness*). If, instead, it also depends on frequency, it should be apparent with items with non-meaningful, but frequent endings (*fliban*).

Experiment 3:

Abstract orthographic representations are typical for skilled adult reading, and are marked by two fundamental components of orthographic processing: the encoding of letter identity and letter position. In terms of letter identity, this implies that the letters are resistant to noise that may introduce various distortions of a letter shape. On the other hand, encoding of letter position refers to location-invariant processing, whereby transposed letters within a string (e.g. JUDGE – JU<u>GD</u>E) evoke more errors and are slower to process compared to replaced letters (JUDGE – JU<u>DP</u>E). Critically, this effect is larger in orthographic material compared to digits or other visual symbols (e.g., $, &, #, +, !). In this experiment, I propose that training with variable, handwritten input may lead to enhanced development of orthographic representations. More specifically, I hypothesize that if novel characters have become resilient to distortion, they would exhibit identity priming effects in a masked priming, same-different task even when primes are presented in distorted format. As for the transposed-letter effects, if the novel characters have established abstract orthographic representations, they are expected to emerge, or at least become stronger, after training.

Chapter 2

Experiment 1: The role of morphology in novel word learning: A registered report

Manuscript under review as a Stage 2 Registered Report at Royal Society Open Science journal. All the materials related to this paper (stimuli, data, and code for data analysis and power analysis and Stage 1 approved manuscript) are publicly available at the Open Science Framework: https://osf.io/x7ctg/).

2.1 Abstract

The majority of the new words that we learn everyday as adults are morphologically complex; yet, we don't know much about the role of morphology in novel word learning. In this study, we tackle this issue by comparing the learning of (i) suffixed novel words (e.g., *flibness*), (ii) novel words that end in non-morphological, but frequent letter chunks (e.g., *fliban*), and (iii) novel words with non-morphological, low-frequency endings (e.g., *flibov*). Words are learned incidentally through sentence reading, while the participants' eye movements are monitored. We show that morphology has a facilitatory role compared to the other two types of novel words, both during learning and in a post-learning recognition memory task. We also showed that participants attributed meaning to word parts (if *flibness* is a state of happiness, then *flib* must mean happy), but this process was not specifically triggered by the pres-

ence of a suffix (*flib* must also mean happy in *fliban* and *flibov*), thus suggesting that the brain tends to assume similar meanings for similar words and word parts.

2.2 Introduction

People encounter new, previously unknown words on a daily basis. In order to preserve successful communication, these new words must be interpreted quickly, essentially on-line. This is one of the mechanisms whereby adults expand their vocabulary throughout their lifetime (Hartshorne & Germine, 2015; Keuleers et al., 2015; Share, 1995). People encounter most new words while reading, which implies that learning happens mostly implicitly, in the absence of any instruction or explanation. This suggests that people are able to compute meaning for novel words online during reading, by relying only on the information provided in the text (Swanborn & de Glopper, 1999).

Understanding an unknown word is mostly supported by the context in which the novel word appears, i.e., the surrounding words (Firth, 1957). For example, in a sentence like "John was incredibly hungry, so he reached for the kitchen and had all the wugs that were left over from dinner", it is not difficult to gather that *wugs* are some kind of food. Upon multiple encounters, especially across different conceptual domains (Mak et al., 2021; Pagán & Nation, 2019), people reliably attribute meaning to the novel lexical item.

Some information about word meaning can also come from the word's form itself, however. As symbolic systems, human lexicons are largely arbitrary; so, generally speaking, we would typically not be able to guess the meaning of a word based on how it looks, or sounds. However, regularities in the mapping between form and meaning do exist (Marelli et al., 2015), and they seem to affect lexical processing (Amenta et al., 2016, 2020; Marelli & Amenta, 2018; Siegelman et al., 2022).

Form and meaning maximally correlate through morphology. The words *gardener*, *seller*, *influencer*, and *driver* all indicate a profession, and they seem to do so by virtue of their ending, *-er*. Similarly, the words *grasp*, *graspable*, *grasping*, and *ungrasp* share a common core, which is related to their stem, grasp-. Even though the form-meaning mapping brought about by morphology is not always perfectly straightforward (e.g., *corner* is not someone who corns, and *irony* is not made of iron), morphology does establish some regularity in the way that words' form is connected to words' meaning. In this paper, we focus on this source of information in the novel words and investigate the effect of suffixes on the acquisition of new lexical items.

Morphology is widespread in human lexicons (Aronoff & Anshen, 2017; Leminen et al., 2016; Roelcke, 1997; Talamo & Celata, 2011) and polymorphemic words account for most novel words that enter the lexicon (Algeo & Algeo, 1993). As such, it is a primary player in word identification, particularly in the visual modality (Leminen et al., 2016; Beyersmann et al., 2012, 2016, 2020; Grainger et al., 2021; for a review, see New et al., 2004). For example, response times in lexical decision are proportional to the frequency of the stem of the target word (New et al., 2004), and the visual word identification of a stem (e.g., *depart*) is speeded by the previous presentation of a morphological relative (e.g., *departure*) in a way that cannot be traced back to the semantic and orthographic similarity between the two words (Drews, 1995; Rastle et al., 2000). It is now completely undisputed that words' morphological structure is engaged during lexical processing.

More critically for the present work, these morphological effects are not limited to well known, familiar lexical material. In fact, morphology has made its way to the psycholinguistic stage when Taft and Forster (1975) discovered that nonwords embedding existing stems (e.g., *de-juvenate*) are more difficult to reject in a lexical decision task than non-morphological controls with nonexisting stems (e.g., *de-pertoire*). Since this seminal work, there have been many reports of morpheme interference effects in nonwords (Burani et al., 2002, 2008; Crepaldi et al., 2010; Dawson et al., 2018) and nonword morphological priming (Beyersmann et al., 2013; Crepaldi et al., 2016; Grainger & Beyersmann, 2021; Hasenäcker et al., 2016). This clearly shows that morphemes are addressed in unfamiliar letter strings, which is of course a necessary condition for word learning to be affected by the morphological structure of words. However, it is not entirely clear whether this information – which this work shows to be available to the readers' cognitive system – is effectively used during word learning; in none of these experiments, in fact, the unfamiliar stimuli were learned as potentially meaningful novel lexical items.

This was the case, instead, in Tamminen et al. (2015). In a series of word learning experiments, participants were familiarised with novel words made up of an existing stem and a new suffix (e.g., *crabafe*). For each of these novel words, a definition was created by using the meaning of each novel affix consistently, to modify the meaning of each familiar stem. For example, *crabafe* would be the zoo building where you can see exotic crab species and *gunafe* the section of an armoury where one can find a gun; in these examples, -*afe* refers to a place, similarly to -*ery* in *bakery* and *nunnery*. Participants were presented with the novel words and their definitions, and were then asked to type the word back. In a separate task, they were asked to recollect the word upon hearing its definition. Based on this training, participants were able to extract the suffix meaning, and to generalize the newly acquired knowledge to untrained novel words after a memory consolidation period of 7 days. For example, they read aloud more quickly *sailafe*, an item to which they were never exposed during training, when a preceding sentence context was consistent with the locative meaning of -*afe*. Interestingly, learning came up considerably more quickly (e.g., right after training) and with lesser constraints (e.g., without a need for high contextual diversity) in tasks that required deliberate reasoning, thus showing a dissociation between implicit and explicit memory. We will address this issue more in depth below. Again on the explicit vs. implicit dichotomy, but this time with regards to the training routine, participants were explicitly instructed to learn the novel words, but no explicit information was given about the suffixes, whose existence was left for the participants to figure out. The morphological training was thus entirely implicit.

In Havas et al. (2015), participants were Finnish and Spanish speakers who were exposed to novel word-picture pairs. The novel words consisted of a non-word stem (*elu-*) and a novel suffix that indicated gender (*elu-ri*). The novel words were paired with pictures that depicted animals wearing typical male or female clothing. Participants were asked to learn the novel word-picture correspondences, but were not informed about the morphological structure – similarly to Tamminen et al. (2015), there was no explicit morphological training. The recognition memory and rule generalization tasks showed, respectively, that participants were able to successfully recognize the items from the training among the distractors, and also to generalize the novel gender-marking system to new stems. In the generalisation task, participants were presented with a new picture of an animal in male or female clothes, paired with two letter strings: both contained a stem that wasn't seen in the training, paired with either the feminine or the masculine suffix. In this study, there was no period of consolidation, suggesting that participants were able to use the newly acquired morphological knowledge straight away.

More recently, Dawson et al., (2021) examined whether developing readers learn novel words better when these words contain suffixes. More specifically, they compared novel words whose meaning was congruent with the dominant meaning of a suffix in the language (e.g., *brint-ise*, to make an object clean again), against novel words that were still suffixed, but with a meaning that was not congruent with the meaning of the suffix in the real language (e.g., *drict-ful*, to put something in fancy dress). They found that congruency facilitated the learning of the meaning of the novel words, but this advantage didn't extend to the phonological, orthographic, or lexical level.

These findings suggest that humans extract morphological information while learning the meaning of unknown words. Moreover, they do so in the absence of any explicit instruction regarding morphology—in none of the studies above participants were cued to the presence of suffixes, or to their contribution to the form and meaning of the novel words. On the other hand, though, an important part of morphology-based learning in real life is based on familiar suffixes attached to unfamiliar stems, e.g., one might learn what *glare* means based on hearing the word *glaring* in an informative context. Tamminen et al., (2015) focused exclusively on the opposite case, that of an unfamiliar suffix attached to a familiar stem. Havas et al., (2015) did use unfamiliar stems, but the focus was on the generalization of the morphological rule based on the novel suffixes; the learning of the novel stem was not tested. In a sense, these studies focus on how we learn morphemes, while the present work will focus instead on how morphemes affect learning. In this sense, we're closer to Dawson et al. (2021), which, however, did not test whether there was any meaning attribution to the novel stems, that is, they did not assess stem learning. Moreover, and perhaps more importantly, participants in these studies received extensive and explicit training on the novel words themselves, which differs substantially from how word learning mostly happens in real life—implicitly, without instructions or explicit feedback.

A step toward a more ecological training regime was taken by Ginestet et al. (2020), who followed previous studies with eye tracking (Joseph et al., 2014; Pagán & Nation, 2019) and trained their participants by embedding novel words into short stories. Of relevance for this paper, the novel words could contain an existing prefix and an existing suffix (e.g., *re-lurb-er*) or non-morphological chunks (e.g., *pe-lurb-le*). Similarly to previous studies (Joseph et al., 2014), Ginestet et al. tracked the learning pattern during repeated exposures via eye tracking metrics. In general, they found that non-complex words attracted more fixations than complex words. It was not clear, however, whether there was a difference in looking times; there was a significant effect of word type on gaze duration, but not on the duration of single fixations, first-of-two fixations and second-of-two fixations. The learning pathway through repeated exposures did not change according to the morphological status of the novel words, when this was measured via the number of fixations or the duration of first-of-two fixations; however, it did change when considering gaze duration, or total duration. The overall picture was a bit unclear across different eye tracking measures. There was, however, an interesting pattern in gaze duration: while nonmorphological words required longer looking times at the beginning of the training, the difference disappeared, or at least shrunk, after four encounters with the novel words. In terms of post-training accuracy, complex words were spelled more accurately than orthographic controls.

In the present work, we build on and extend Ginestet et al. (2020) in two main ways. Firstly, we will connect with a large branch of the recent morphological literature, which focuses on the role of meaning-bearing per se vs. the role of letter co-occurrence statistics. This was triggered by the finding that a pseudo-suffixed word like corner speeds up the processing of its pseudo-stem corn in primed lexical decision to a similar extent as a genuinely suffixed prime (e.g., *farmer-farm*) (Rastle et al., 2004; Rastle & Davis, 2008). More recently, it has been shown that chunks of pseudo-letters, with no connection with phonology or semantics, can mimic some classic morphological effects (Lelonkiewicz et al., 2020). This further underlines the strong effect that the mere frequency of a letter cluster can yield. In the context of the present work, one might imagine that novel words with existing suffixes are learned more easily because they contain meaningful elements, or because they feature frequent letter chunks. Thus, we will contrast: (i) novel words that contain a suffix to items that contain an ending of the same frequency, but do not have any meaning; and (ii) novel words with high-frequency endings to items with endings that are lower in frequency. Such a design will rather cleanly separate meaning-based morphological effects from frequency-based effects.

Secondly, in addition to focusing on orthographic learning, we will also investigate the extraction of the meaning of the novel stem—we will ask whether participants learn that, if *flibness* is a state of happiness, then *flib* must mean happy.

A further important feature of the experiment we propose is that we will investigate the outcome of learning both implicitly and explicitly. We mentioned above evidence that explicit representations of word meanings can develop rapidly, but implicit representations take more time and/or exposure. This notion is supported by several other studies. In an experiment by Batterink and Neville (2011), participants were exposed to pseudowords embedded in a narrative. Afterwards, they were administered behavioural tasks while ERPs were recorded. The results showed that in a primed lexical decision task, which measured learning implicitly, there was no effect of priming, neither in behaviour nor in electrophysiology. On the other hand, in the recognition task (explicit measure), a robust N400 effect was found for the correctly learned word. Exploring the timeline of the development of implicit word representations, Qiao and Forster (Qiao & Forster, 2013) taught participants novel words that were designed as neighbours of existing words (e.g., *bontract*, a neighbour to *contract*). They were tested in a primed lexical decision task. The prediction was that if the novel word is successfully learned, it will yield inhibition, as it is typically found for pairs like *contract-CONTRAST*) (Davis & Lupker, 2006; Forster & Veres, 1998; Qiao et al., 2009). The results showed that inhibition only emerged after four sessions of training spread over four weeks. This evidence sits nicely with general theories of learning and memory that postulate slow integration of novel information into a highly interconnected system of overlapping memories (Alvarez & Squire, 1994; Kumaran & McClelland, 2012; McClelland et al., 1995; Reilly & Norman, 2002). In such a system, new information must be acquired and integrated at a slow pace, in order to avoid interference with pre-existing knowledge. This slowness of learning also allows the system to capture structural, persistent aspects of the input (e.g., a new word occurring across a set of different sentences), rather than volatile details or irrelevant information (e.g., whether the same word is pronounced by a male or a female speaker). Within this framework, explicit and implicit memory subserve fundamentally different goals, and are structurally different; the former is fast, doesn't lead to integration with existing knowledge and decays relatively quickly; the former is slow, but embeds into a highly sophisticated system of existing memories and is more resistant to decay. Notice that, as general statements about learning and memory, these considerations do not speak directly on word learning, and even less on the role of morphology in the process. Thus, there are no direct predictions that emerge here and that this study wants to test. Rather, these general theories provide the broader landscape for the investigation of word learning, and the psycholinguistic data and theorizing described above beg the question of how morphology plays out here, given the deep role that it plays in word processing and the scant evidence we have collected thus far.

In the present experiment, eye tracking will provide an implicit measure of learning, where we will monitor the reduction of looking times across progressive encounters with the novel words. We will use recognition memory task as an explicit measure. In addition, we will also investigate both the implicit and the explicit extraction of the stem meaning–via a sentence congruency task and a definition selection task, respectively.

In what follows, we first describe the experimental design and paradigm; then we will present a pilot study that we have conducted with 14 participants; and finally, we will illustrate how the pilot helped us fine-tune some of the experiment parameters and adapt the design to the outcome of a power analysis.

2.3 Methods

Novel words

Eighteen novel stems were created as readable combinations of letters that do not exist as stems or words in Italian (e.g., *pobed-*, *cribot-*). They were 5 or 6 characters in length; their mean log bigram frequency was 5.77 (SD = 0.31); their mean average Levenshtein distance to the 20 closest lexical neighbours (OLD20; (Yarkoni et al., 2008) was 2.13 (SD = 0.27); and they had no immediate lexical neighbour (their edit distance to the closest word was >1). These 18 stems were sorted in 6 triplets, so that the members of each triplet would be rotated through the three main experimental conditions (see next paragraph). Within each triplet, stems were of the same length, and matched as closely as possible for log bigram frequency (5.72 ± 0.25 vs. 5.66 ± 0.41 vs. 5.91 ± 0.25) and OLD20 (2.25 ± 0.34 vs. 2.13 ± 0.31 vs. 2±0).

To generate the novel words to be learned in the experiment, the stems were paired with 3 types of endings: (i) suffixes (e.g., *-enza*; a corresponding example in English would be *-ness*); (ii) non-morphological endings matched in frequency to suffixes (5.01 ± .24 vs. 4.73 ± .36; e.g., *-ondo*; a corresponding example in English would be *-an*); (iii) non-morphological ending lower in frequency than suffixes (2.37 ± .20 vs. 4.73 ± .36; e.g., *-espa*; a corresponding example in English would be *-ov*). Frequency was specific to word final position (Crepaldi et al., 2010). The stimuli were selected from De Rosa and Crepaldi (2022). As mentioned above, we rotated the 18 stems across the three types of endings in a classic Latin Square design; therefore, the overall stimulus set was composed of 18x3=54 novel words, but each participant only learned 18 (that is, for each participant, each stem was paired with only one given ending, with six endings per condition). Overall, the novel words had a mean log bigram frequency of 6.01 (SD = 0.19) and a mean OLD20 of 3.78 (SD = 0.36). None of the words had any immediate lexical neighbour.

Learning

Each target word was embedded in 10 different sentences (e.g., "Fare il *cribotista* non porta tanti soldi, però sai che fai bene all'ambiente e agli animali", "You don't make much money working as a *cribotista*, but you help the environment and the animals"). Thus, each participant will read 180 sentences in total, and will be exposed to each novel word 10 times. Sentences were constructed to convey the meaning of the novel word; for a couple of illustrative examples, see Table 1 (the full stimulus set is available at https://osf.io/x7ctg/; see Data Availability below). We tried to use fairly diverse contexts (see, for example, Bolger et al., 2008; Mak et al., 2021; Pagán & Nation, 2019) and to give sentences similar syntactic structures. The novel words never appear in the first or last position. The training sentences contain 16 words on average (SD=1.57; range:13-19) and will be presented one by one in the center of the screen. Participants will be asked to press the spacebar when they are ready to read a new sentence. They will be instructed to read sentences and try to understand them even if there will be some unknown words. The order with which the sentences will be presented will be randomised across participants.

Target Word	**Training Sentence**
rugobenza	Marco non ha *rugobenza*, quindi quando la madre gli ha sgridato, si è messo a urlare anche lui. Marco doesn't have any *rugobenza* so when his mother scolded him, he also started to yell. Quando ho rotto la bottiglia dell'olio, ho dovuto ascoltare con *rugobenza* il rimproverò di papà. When I broke a bottle with oil, I had to listen to dad's scolding with *rugobenza*. Trent'anni fa i bambini dovevano sopportare con *rugobenza* e senza rispondere i rimproveri della maestra. Thirty years ago, children had to endure teacher's criticism with *rugobenza* and in quiet.
cribotista	Alessandro lavora da tanto come un *cribotista*, ormai è famoso per le sue capanne sotto terra. Alessandro has been working as a *cribotista* for a long time now, he's famous for his underground huts. Nel neolitico, il *cribotista* era molto importante per la tribù perché costruiva i ripari per tutti. *Cribotista* was very important for a tribe in the Neolithic because they built shelter for everybody. Per fare il *cribotista*, bisogna intendersi della natura e soprattutto del terreno e acque sotterranee. To work as a *cribotista*, you have to know a lot about nature, and terrain and groundwater above all.

Table 1: Example of two novel words each used in three different sentences. Original followed by a translation into English.

Testing

Recognition memory

Participants will be asked to identify each trained novel word in a recognition memory task. Similarly to the approach of (Tamminen et al., 2015), there will be three types of distractors: (i) untrained stem + trained ending (e.g., *bepolenza*); (ii) trained stem + untrained ending (e.g., *rugobiera*); (iii) trained stem + trained ending, but in new, unseen combinations (e.g., *rugobondo*). Each item will be presented once, for a total of 60 trials; 18 trained items, 18 recombinant items, and 12 items for each of the other two distractor types. Clearly, the trained and recombinant items depend on the rotation a participant was assigned to in the learning phase; thus, we created three matching rotations in this task as well. Items will appear one at a time at the center of the screen, and participants will be asked to decide whether they remember the item from the training via a button press on the keyboard. Items will stay on the screen until participants respond.

Sentence congruency

The sentence congruency task was designed to assess whether participants assign meaning to the stems of the novel words they learned. For example, the novel noun *rugobenza* contains the familiar, existing suffix -*enza*, which implies that *rugob*- is a novel stem. The training sentences will provide meaning to the novel words, and therefore to their novel stems (e.g., *rugobenza* refers to "being able to stand someone yelling at you without overreacting", and thus the stem *rugob*- must have something to do with the ability to accept scolding without much complaining). Free stems do not exist in Italian[1] , at least for content words, so we have to use what we call the *base words*, i.e., words where the novel stems are attached to the unmarked morphological suffix (e.g., *rugob-are*, an infinitive verb; *cribot-ista*, a singular noun). Clearly, it is not obvious that readers would infer the existence of a novel stem when the trained word does *not* contain a suffix, i.e., in the non-morphological ending conditions. In fact, if the learning of the stem is triggered by the presence of the familiar suffix, there should be no stem learning in these conditions. For each base word, two sentences were constructed, which are either consistent or inconsistent with the meaning of the base word itself (e.g., congruent meaning: "Quando

[1]Except for a few loan words (e.g., brioche, chef, élite) and a small set of accented items (e.g., caffé).

la mamma mi ha sgridato, abbiamo litigato di brutto, non ce l'ho fatta a rugobare"; "When my mother yelled at me, we had a big row, I was incapable of *rugobare*"; incongruent meaning: "Quando la mamma mi ha sgridato, era troppo tardi, non ce l'ho fatta a rugobare."; "When my mother yelled at me, it was too late, I was incapable of *rugobare*"). The critical base word is always at the end of the sentence in both sets. Although there is some debate as to whether the sentence final position has a special role in handling inconsistencies that were not resolved within a sentence (Stowe et al., 2018), our primary goal here was being able to focus on an area that is most likely to reflect integration processes. Sentences in the two sets have very similar structures. They were divided into two rotations, which we will counterbalance across participants, so that no single participant will see corresponding sentences in both congruent and incongruent conditions. Sentences will appear one at a time on the screen while the eye movements will be recorded. Each sentence will be followed by the question "Does this sentence make sense?", to which participants will reply with a button press on a keyboard. There will be no timeout, neither for sentence reading nor for the comprehension questions.

Definition selection

Participants will be presented with the 18 base words corresponding to the 18 novel words that they had learned, and will have to choose among 4 definitions. Beside the correct option (e.g., "standing someone yelling at you without overreacting" for *rugobenza*), the alternatives are: (i) related, but underspecified (e.g., "to be patient"); (ii) unrelated (e.g., "to listen to someone's stupid comments"); and (iii) the definition of another base word (e.g., "distrust in gossip", which is the definition for another novel word in the experiment, *zudulare*). This latter option allows us to make sure that participants will not simply choose some semantic content that they were exposed to during the experiment, but specifically remember the correct link between form and meaning. Trial order will be randomized across participants. Each item will appear in red in the top part of the screen. The four possible definitions are numbered and displayed one under the other. Participants will select their choice by pressing the appropriate number on the keyboard (1, 2, 3 or 4). There will be no timeout.

Sensitivity to morphemes

Prior to the experiment, we will probe participants' sensitivity to the morphological structure of nonwords using a Morpheme Interference Task—a lexical decision paradigm where the critical comparison is between the rejection time for morphologically structured nonword (e.g., *fruitness*) and orthographic controls (e.g., *fruitnuss*). This task has been extensively used in the morphological literature (Beyersmann et al., 2020; Burani et al., 2006; Crepaldi et al., 2010; Taft & Forster, 1975; Yablonski & Ben-Shachar, 2016), although not as a measure for individual variability. Recently, De Rosa and Crepaldi (2022) developed a version of this task with Italian materials; we will use that task here. The critical conditions (morphological and control, orthographic nonwords) contain 60 items each, which differ by one letter (e.g., *lesionaggio* vs. *lesioneggio*). Stimuli in these conditions are matched for the number of orthographic neighbours ($0.25 \pm .44$ vs. $0.05 \pm .22$), OLD20 ($2.66 \pm .43$ vs. $3.02 \pm .48$) and length (8.97 ± 1.19 vs. 8.97 ± 1.19). Of course, the task also includes 48 word stimuli as fillers (i.e., YES trials), 24 of which are complex ($N = 1.83 \pm 1.09$, OLD20 $= 2.40 \pm .49$, length: 8.71 ± 1.04) and 24 of which are not ($N = 1.58 \pm 1.14$, OLD20 $= 2.48 \pm .45$, length: 8.79 ± 1.14). To ensure that no single participant will be exposed to both the complex and noncomplex version of the same critical item (e.g., *fruitness* and *fruitnuss*), the task includes two rotations that are counterbalanced across participants; so, each participant will be presented with 30 complex nonwords, 30 control, orthographic nonwords, 24 complex words and 24 simple words. Item presentation will be randomized across participants. Items will be presented one at a time at the center of the screen until keyboard response.

Apparatus and software

The experiment was programmed in Python (Van Rossum & Drake, 2009) (v. 3.6.6), using Psychopy (v. 1.11.0.0; (Peirce et al., 2019)) and Pylink (v. 2020.2.3; (SR Research Ltd.). Stimuli will be presented in white against a black background, using the Courier New font (size 25). They will be presented on a 27-inch monitor, at a viewing distance of 62cm. The refresh rate and resolution of the monitor will be 144 Hz and 1920x1080 pixels, respectively. To record eye movements, we use an Eyelink 1000 Plus (tower mount; (SR Research Ltd.). We will record from the right eye, with a sampling rate of 1000 Hz. In order to minimise head movements, participants will be asked to use a chin rest. Responses will be collected via a keyboard.

Procedure

Participants will first complete the Morpheme Interference task, which lasts around 10 minutes. They will then move on with the Learning task (~30 minutes), Recognition Memory task (~10 minutes), Sentence Congruency task (~20 minutes) and Definition Selection task (~10 minutes). Participants will be encouraged to take a break between the tasks. The whole procedure will last around 1 hour 30 minutes.

Data modelling and statistical analysis

Data will be modelled and analysed in R (RCoreTeam, 2021). For the computation of eye tracking measures, we will use the Eyekit package for Python, version 0.3.10 (Carr, 2021). To model the data, we will fit (generalized) linear mixed-effects models as implemented in the *lme4* package (Bates et al., 2015). We will implement treatment coding in all models. For visual inspection of the data distribution, we will use the Box-Cox plot as implemented in the MASS package (Venables & Ripley, 2002); data will be either logarithmically or inverse transformed, whatever will make the relevant distribution more Gaussian. All models will include random intercepts for subject and stimulus, and the maximal random slopes structure as appropriate for the experimental design. In case of convergence issues, we will gradually simplify the random slope structure by removing (i) the interactions; (ii) then the random slope correlations; and finally (iii) the random slope for the main effects. After modelling the data, we will assess statistical significance using *Anova* from the car package (Fox & Weisberg, 2019). To interpret the statistical patterns, we will use both the estimated Beta parameters in the model and the reconstruction of the expected response times/looking times/accuracies per condition, as computed by functions in the package *Effects* (Fox & Weisberg, 2019)[2].

Exclusion criteria

Individual datapoints will be discarded when they are affected by technical errors (e.g., we displayed the wrong word, the program deviated from the standard trial timeline). If the prevalence of these technical errors is high (> 20%) for a given participant, or for a given item, we will reject the entire set of data referring to that participant or item. The eye tracking data will

[2]We did not carry out null hypothesis significance testing with the pilot data described below, given that it would obviously be heavily underpowered. Even more importantly, assessing statistical significance is not the goal of the pilot experiment.

be cleaned following a standard approach in the field. Entire trials will be rejected when they're heavily affected by head movement, blinking, or drift. Individual fixations under 80ms will also be discarded, along with excessively long fixations; the precise threshold here will be defined after looking at the distribution of the datapoints. In the sentence congruency and the definition selection tasks (that is, those tasks assessing the generalisation of the meaning to the stem), we will also exclude data from participants whose d-prime in the recognition memory task will be below 1; such a low d-prime would indicate that these participants did not learn the novel words in the first place, so it doesn't make sense to ask whether they learned any meaning for the stems.

Data availability statement

The full set of materials, data and analysis scripts will be available on OSF. The following link, https://osf.io/x7ctg/), now contains the stimuli, the pilot data and the analysis that we conducted on them; it will of course be updated with the final set of data and analysis, once data collection will be carried out (along with any eventual exploratory analysis that is not pre-registered in this report).

Ethical approval

The experiment obtained clearance from the SISSA Ethical Committee. All participants gave informed consent prior to the experiment.

2.3.1 Pilot study

We carried out a small-scale pilot study with the goal of putting our paradigm and experimental design to the test. We wanted to check whether people do indeed learn the novel words, and whether the way we operationalised our theoretical questions holds the promise to answer those questions, once a larger sample of data will be collected in the main study. In line with these goals, we limited ourselves to explore the pilot data and model them to estimate effect sizes that might inform a power analysis (of course, with consideration of the fact that effect sizes might deviate from the real, population effect size in a small-scale pilot).

Participants

Fourteen adults (3 female; mean age = 28.6 years, SD = 3.7 years) took part in the pilot experiment. They were all native speakers of Italian, with normal or corrected-to-normal vision and no reading disabilities. Participants were paid 20 Euros for their participation.

2.3.2 Results

Sensitivity to morphemes

The average accuracy and response time in the Morpheme Interference task are .93 and 1390ms, respectively. This latter figure is quite slow, which suggests that the participants in the pilot study might have missed our emphasis on speed; to fix this, in the main experiment we will implement a time-out for response time in the practice phase of this task. The analysis of the data aggregated by item showed that two stimuli, *urtevole* and *flauteria*, were considered to be existing words by most participants, i.e., their accuracy was below 50%. They will be substituted with two new items (*divaneria* and *untevole*) in a way that doesn't affect the matching between the complex nonwords and their orthographic control.

As expected, morphologically structured nonwords were responded to less accurately (.88) and more slowly (1734 ms) than their control, non-morphological controls (.99 and 1557 ms, respectively). The effect size (calculated as the Beta coefficient in the model) was estimated to be 55.17 (95% CI= 23.91 – 86.43) for the response times, and -2.61 (95% CI= -3.65 – -1.58) for accuracy [3]. These figures nicely replicate previous literature (Burani et al., 2006; Crepaldi et al., 2010; Taft & Forster, 1975), and thus attest to the reliability of the task.

Because we're primarily interested in individual variability here, an index of sensitivity to the morphological structure of nonwords was computed for each participant according to the following procedure. We subtracted the average on simple nonwords from the average on complex nonwords, separately for accuracy and response time. Since these two measures did not exhibit strong correlation (-0.18), we standardized them and summed them up. In the pilot sample, this index varies from -2.8 to 1.89 across participants (median = -0.23). In an exploratory analysis in the main study, we will correlate these indices with the accuracy and reading times measures in the tasks that we

[3]Because accuracy models were estimated using a logistic link function, these numbers are expressed in log odds.

describe below. In this way, we aim to explore whether individual sensitivity to morphology is associated with how one uses morphological information in the word learning process.

Learning

The novel words were fixated 5.29 times on average. The proportion of single fixations is relatively low, 24%, which is not surprising given that these were novel, unfamiliar items. The overall mean first fixation and gaze durations are 255 ms and 732 ms, respectively, and the correlation between these two metrics is 0.14 (which nicely reflects the low proportion of single fixations). Moreover, the target novel words were refixated quite often, 39% of the times.

Based on this initial overview of the data and on the relevant literature (Chaffin et al., 2001; Joseph & Nation, 2018; Pagán & Nation, 2019) we considered the following dependent variables: (i) the duration of the first fixation into the novel word; (ii) the sum of the duration of all fixations on the novel word during first pass (i.e., until the gaze moves away from the word for the first time; *gaze duration*); (iii) the sum of the duration of all fixations on the novel word from the first fixation until the gaze moves past to the right (i.e., including any regression to preceding words on the left; *go-past time*); (iv) the sum of the duration of all fixations on the novel word (*total duration*); (v) the probability of regressing into the target, after having gone past it to the right. This set of variables encompasses early to late processing.

Figure 2.1 shows how the eye tracking metrics change from the first to the last encounter with the novel words. Gaze duration, go-past time, total fixation time and probability of regression into the target word all decrease with more encounters with the novel words. First fixations remain relatively flat, or even become longer. As to the comparison between the complexity conditions, it is not easy to observe patterns with such large confidence intervals (which is to be expected in a pilot experiment). Suffixed, high-frequency ending and low-frequency ending novel words very much overlap in the first fixation duration. Overall, the suffixed items seem to elicit shorter gaze durations, go-past times and total durations. The pattern is quite less clear as far as probability of regression is concerned. Overall, it is difficult to see any strong sign of an interaction between condition and the way eye tracking metrics evolve over successive encounters with the novel words.

To explore these results further, we fitted a linear mixed model to each of the eye tracking measures, with low frequency ending items as a baseline.

In line with the considerations offered above, number of encounters seems to have an effect on all our variables: first fixation duration (b = 0.01, CI = [0.00 – 0.02]), gaze duration: b = -0.02, CI = [-0.03 – 0.00]), go-past time (b = -0.05, CI = [-0.07 – -0.04]), total durations (b = -0.05, CI = [-0.06 – -0.03]), and regressions back into target (b = -0.07, CI = [-0.12 – -0.01]). As for the complexity effects, modeling confirms what was observed from the raw data: this effect is virtually non-existent in the first fixation duration (HF: b = 0.00, CI = [-0.07 – 0.08]; SUFF: b = 0.01, CI = [-0.06 – 0.09], only a little more visible in gaze duration (HF: b = 0.04, CI = [-0.11 – 0.19]; SUFF: b = -0.02, CI = [-0.17 – 0.13]), and a bit more apparent in go-past time (HF b = -0.18, CI = [-0.31 – -0.05]; SUFF b = -0.24, CI = [-0.36 – -0.11]), total duration (HF b = -0.09, CI = [-0.24 – 0.05]; SUFF b = -0.17, CI = [-0.32 – -0.03]) and regressions (HF b = 0.27, CI = [-0.24 – 0.77]; SUFF b = 0.27, CI = [-0.23 – 0.77]).

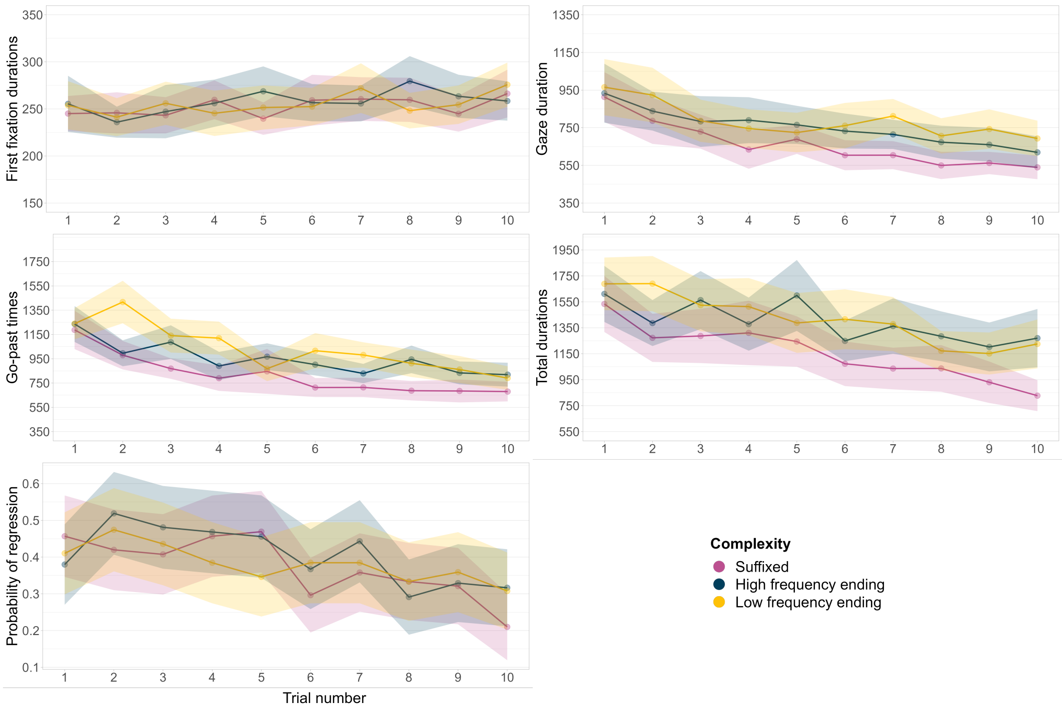


Figure 2.1: Eye tracking measures as they changed during the experiment, across the three main conditions. The shadowed areas represent the standard error of the mean.

Overall, these pilot data suggest that fixation durations change as the number of encounters with the novel words increases, in a way that is compatible with easier processing as a consequence of learning. Reduced number of regressions back into the novel words supports this suggestion. We find no apparent sign of an interaction between number of encounters and condition.

Recognition memory

The mean overall accuracy in this task was .81, suggesting that participants learned the novel words quite successfully. Recognition is much higher with suffixed words (.90) than high-frequency (.76) and low-frequency (.76) items, although variance is quite high across participants (.30, .43 and .43 in the three conditions, respectively) and therefore it's difficult to understand precisely how solid this difference might be. The mixed-effects model estimates the difference between suffixed and low-frequency items to be 1.16 in the log odds space, and the 95% CI lies entirely above zero (0.04 – 2.28).

As in any YES-NO recognition memory task, a good performance in the identification of familiar items can be achieved with a bias towards YES responses; at the limit, one might always respond YES, and this would guarantee a 100% performance in the identification of known items (hits), at the cost of a very high rate of false alarms (i.e., identifying as familiar items that are new instead). In addition, with the current design there are more NO than YES expected responses (following the path set by previous studies, e.g., (Tamminen et al., 2015)), which might invite respondents to develop a bias towards NO answers. To control for this potential confound, we computed a d-prime score for each participant, which corrects the performance on familiar items based on the rate of false alarms (d' = z(hit rate) - z(false alarm rate)). The mean d-prime value across participants was 2.18, with a range of 1.33 – 3.91, showing that participants were indeed able to recognize novel words among distractors, both at the group level and all of them individually (there is no clear-cut threshold for reliable discrimination on the d-prime scale, but 1 is traditionally taken as indicating some ability to tease apart familiar from novel items; (Bottini & Crepaldi, 2016; Macmillan & Creelman, 2005). Looking at the d-prime separately across conditions, we see that subjects were more sensitive to distinguish trained items from foils in the suffixed condition (d' = 3.13), compared to the high-frequency ending (d' = 2.15) and low-frequency ending conditions (d' = 2.3). In addition, we checked the distribution of the c index, which tracks bias (Stanislaw & Todorov, 1999). The median value was -0.10 (min: -0.60; max:1.24). Thus, there is no sign that the higher number of expected NO responses elicited a widespread NO bias.

Interim considerations

Overall, the eye tracking data in the learning task and the participants' performance in recognition memory clearly suggest that the paradigm is effective in inducing some learning of the novel words. This learning also seems to interact nicely with our experimental manipulation. Of course, we cannot say whether the differences between conditions illustrated above are statistically reliable at this point, but they do suggest that the paradigm has the potential to reveal how word learning is affected by the presence of suffixes or high-frequency, non-morphological clusters.

One important note, however, is that the data described thus far only indicate that participants gain familiarity with the novel words; they don't say, however, whether the readers also learned the meaning of these words, or whether they attributed meaning to the stems. This was the goal of the tasks that we're going to consider next.

Sentence congruency

The overall accuracy in the behavioural part of this task, i.e., whether participants are able to distinguish congruent vs. incongruent sentences, was quite low, .57 on average. Breaking down by participants (see Figure 2.2), only two out of 13 are 5% or less likely to be guessing randomly. Interestingly, the accuracy rate seems to change quite little across the complexity conditions (SUFF: mean = 0.55, SD = 0.50; HF: mean = 0.60, SD = 0.49; LF: mean = 0.56, SD = 0.50), which is quite revealing. If any stem meaning attribution might happen, this should show up specifically in the suffix condition, if it's primarily driven by the presence of a meaningful suffix; or in the suffix and the high-frequency ending conditions, if it's primarily driven by frequency. These data, instead, seem to suggest that there is little difference between the three conditions, and that stem meaning extraction didn't happen – which is in line with the overall performance in this task, as illustrated at the beginning of this paragraph.

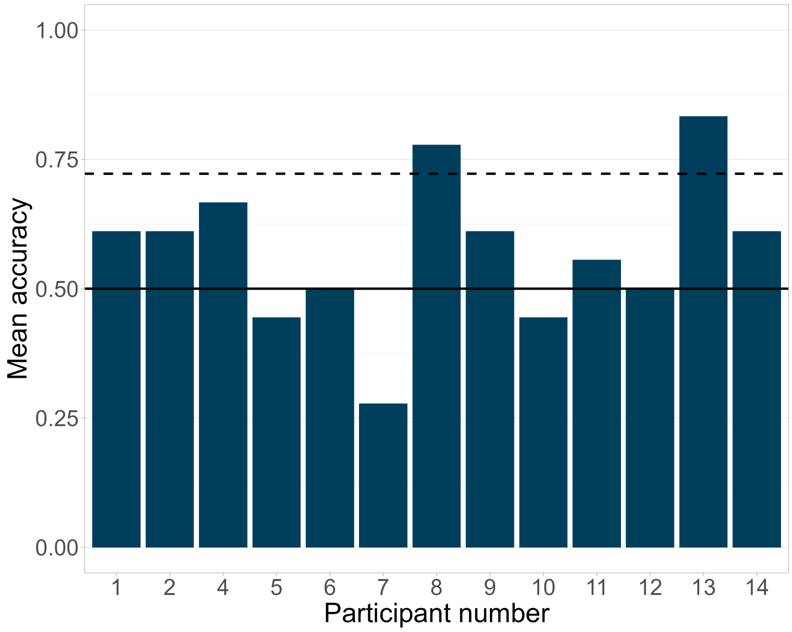


Figure 2.2: Accuracy across participants. The full line represents the chance level, while the dashed line represents the 95th percentile of a binomial distribution with p(correct)=.5 (chance level) and n=18 (the number of trials that our participants undertook); that is, participants above this threshold are 5% or less likely to be responding randomly.

What do the eye tracking data suggest? Overall, the novel words were fixated 5.91 times on average. They were also refixated after rereading the previous part of the sentences on 67% of the trials. These figures suggest that participants were making an effort to assign a meaning to the novel base word, or at least integrate it somewhat into the broader semantic context provided by the sentence.

In addition to the metrics that we considered in the learning task (first fixation duration, gaze duration and total looking time), we examined here the summed durations of second pass fixations and the probability of making a second pass. The critical word was always at the end of the sentence, and we reasoned that if participants are unsure about its meaning, they are likely to reexamine the sentence and then refixate the critical item. This behaviour would be captured most effectively by sum-of-second-pass and probability-of-second-pass.

The congruency effects on the different eye tracking metrics are represented in Figure 2.3. First fixation and gaze duration show similar patterns; the congruency effect is stronger with base words that come from novel words with high-frequency endings, as compared to low frequency endings (Beta = -0.30, CI = [-0.57 – -0.03] and Beta = -0.14, CI = [-0.51 – 0.25], for first fixation and gaze duration, respectively) and suffixes (Beta = -0.15, CI = [-0.42 – 0.11] and Beta = -0.02, CI = [-0.42 – -0.37]). The probability of refixation is quite similar with base words from the three conditions (the model Betas for the comparison across conditions are all very close to zero), which indicates that, even if first fixation and gaze duration might suggest better stem learning with high-frequency endings, all base words equally required extra processing. That is, whatever difference in learning across conditions is revealed by the first-pass metrics, it is probably not entirely solid. On the summed duration of second-pass fixations, the congruency effect is stronger with base words from the non-morphological conditions (Beta = -0.51, CI = [-1.06 – 0.04] and Beta = -0.41, CI = [-0.97 – 0.15] for low frequency and high frequency ending words, respectively, both contrasted with suffixed words). On total looking time, the pattern is reversed compared to first fixation and gaze duration: the congruency effect seems larger with base words from the suffix and low-frequency ending conditions, and smaller with base words from the high-frequency ending condition. However, the model Betas for the cross-condition comparisons are close to zero, and their confidence intervals are quite symmetrical around zero (low frequency vs. high frequency: Beta = 0.01, CI = [-0.33 – 0.34]; suffixes vs. high frequency:congruent: Beta = -0.05, CI = [-0.39 – 0.28]).

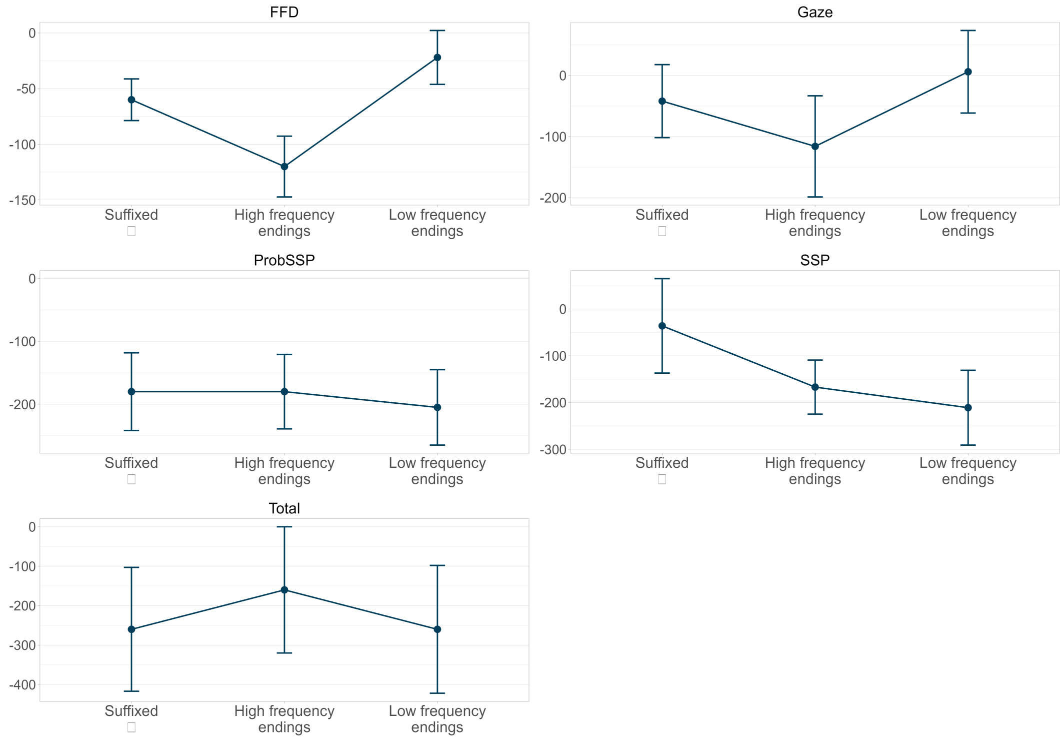


Figure 2.3: Congruency effect in the Sentence Congruency task. Error bars represent the standard error of the mean. FFD, first fixation duration; Gaze, gaze duration; SSP, sum of second pass; ProbSSP, probability of second pass; Total, total durations.

Debriefing

At the end of the whole testing session, we asked the participants whether they could guess the aim of the experiment, if they employed any specific strategies to complete the tasks, and how did they find the experiment overall. The aim was to gather information that might be useful to adjust the design of the experiment.

The participants' responses indicate that they understood the task correctly. Unsurprisingly, several participants mentioned the morphological content in some items in the learning task; however, none of them guessed that this was a crucial aspect of the experiment. Moreover, participants did not report to have developed any specific strategy that would invalidate the outcome of the experiments. Most of them explained that they were simply trying to remember the words, while some reported that they tried to create a mental image of the concept. Finally, some participants reported that they looked for existing words in Italian that refer to similar concepts, to help memorize the novel words.

Final considerations

The eye tracking data in the learning task and the recognition memory performance show that the paradigm works—participants seem to learn the novel words. There is also a fairly clear effect of word ending on this learning, which is particularly prominent in the recognition memory task.

At the same time, however, the pilot data seem to indicate that this learning doesn't lead to any solid extraction of the meaning of the stems. The behavioural data in the sentence congruency task suggest that participants don't distinguish very clearly congruent from incongruent sentences containing the novel stems. The eye tracking data are a bit less clear, and possibly leave some room for the existence of a congruency effect showing up implicitly in eye movements. Yet, the congruency effects that timidly emerged in the different eye tracking metrics do not seem to form a coherent pattern.

The definition selection task offers evidence that participants were able to single out the correct definition of the base words; should this finding emerge in the main study, it would describe an interesting dissociation between the ability to process the base words while reading a sentence and the ability to figure out a definition for them, at least among the competing alternatives that we used.

Finally, there seems to be very little evidence overall that whatever is learned about the meaning of the stems, it comes from a morphological analysis triggered by the presence of a familiar suffix, or from an analysis of the internal structure of the words that is informed by letter chunk frequency. The mechanism that seems to be in place is rather one whereby a more general correspondence between form and meaning is assumed on the part of the reader, so that whatever novel word begins with the same letter chunk, it must have a similar meaning—independently of morphology or frequency. Should this conclusion emerge in the main study, it would resonate nicely with theoretical accounts (Baayen et al., 2011; Rueckl et al., 1997) and experimental data (Amenta et al., 2020; Marelli et al., 2015; Siegelman et al., 2022) that place morphology in the context of a more general attempt of the brain to find probabilistic ties between form and meaning of whatever nature. Under this view, morphology might not be a special domain of linguistic analysis that is qualitatively unique, and requires specific cognitive mechanisms and representations; rather, morphological sets would be "lexical islands" where a reliable correspondence between form and meaning just emerges more clearly, but not in a way that clearly sets it apart from other types of more fuzzy, more probabilistic regularities in the correspondence between form and meaning.

Power analysis

Methods

Since we plan to analyze our data via mixed-effect models, and there is no analytical treatment of power in this context as far as we are aware, we resorted to data simulation. We followed the approach described in (DeBruine & Barr, 2021), and generated 1000 datasets using the relevant effect sizes, and noise levels informed by previous experience in the lab and the pilot data described above. The code for this analysis is reported in the OSF repository for the project, of course.

More specifically, for the Learning and Recognition Memory tasks, we targeted   of the effect size that we observed in the pilot experiment, to account for a possible overinflation of these effects (which is likely to happen in small-scale, pilot studies). It was more complex to understand which is the most proper approach for the Sentence Congruency and Definition Selection tasks. In fact, the effects that emerged there were very small, which would require a practically unfeasible sample size, if we are to reach a sufficient power level. In addition, if the effects are truly as small as the pilot data would suggest, it is not even clear they would be interesting at all from a theoretical standpoint; so, we believe that aiming at statistical significance there is not fitting. An alternative approach was to settle on an effect size that would be large enough to be of any value. But this is another very difficult issue to tackle: how do we decide what is big enough to be theoretically relevant? Therefore, we eventually decided to reverse our reasoning; we computed the sample size that would be required to reach 90% power in the Learning and Recognition memory tasks, where we believe that the effects of interest are very likely to be large enough. Using that sample size, we then back-computed what effect size we would be able to address in the Congruency and Definition Selection tasks with a 90% power.

For the learning task, the questions that we want to address are (i) whether the novel words are learned in the first place, and (ii) whether their learning depends on the type of ending. This translates into a main effect of number of encounters, and an interaction between number of encounters and type of ending. Based on the literature (Ginestet et al., 2020; Joseph et al., 2014; Joseph & Nation, 2018; Pagán & Nation, 2019), the results of the pilot experiment and the results of the power analysis itself, and in an effort to keep the design and the analysis as simple as possible, we chose to consider gaze duration and total duration. The former would specifically track relatively early processing, while the latter would encompass both early and late processing stages. The correlation between the two dependent variables is 0.43 in the pilot data, which indicates that they do indeed capture different cognitive processes.

For the recognition memory task, the behavioural part of the sentence congruency task and the definition selection task, we want to ask whether there's an influence of ending type in their performance, i.e., we are looking for an effect of complexity.

Finally, for the eye movements analyses in the sentence congruency task, we considered gaze duration and total duration, again with the aim of simplifying the design and the analysis as much possible, while still addressing the relevant processing stages.

Results

The results of the power analysis for the learning task and recognition memory task are illustrated in Figure 2.4 and Figure 2.5, respectively. The effect that requires the most participants is, predictably, the interaction between number of encounters and type of ending in the gaze duration analysis; in order to reach a 90% power here, we need 84 subjects. This sample size would guarantee an even higher power on all other effects of interest, for all other tasks and dependent variables.

Learning task

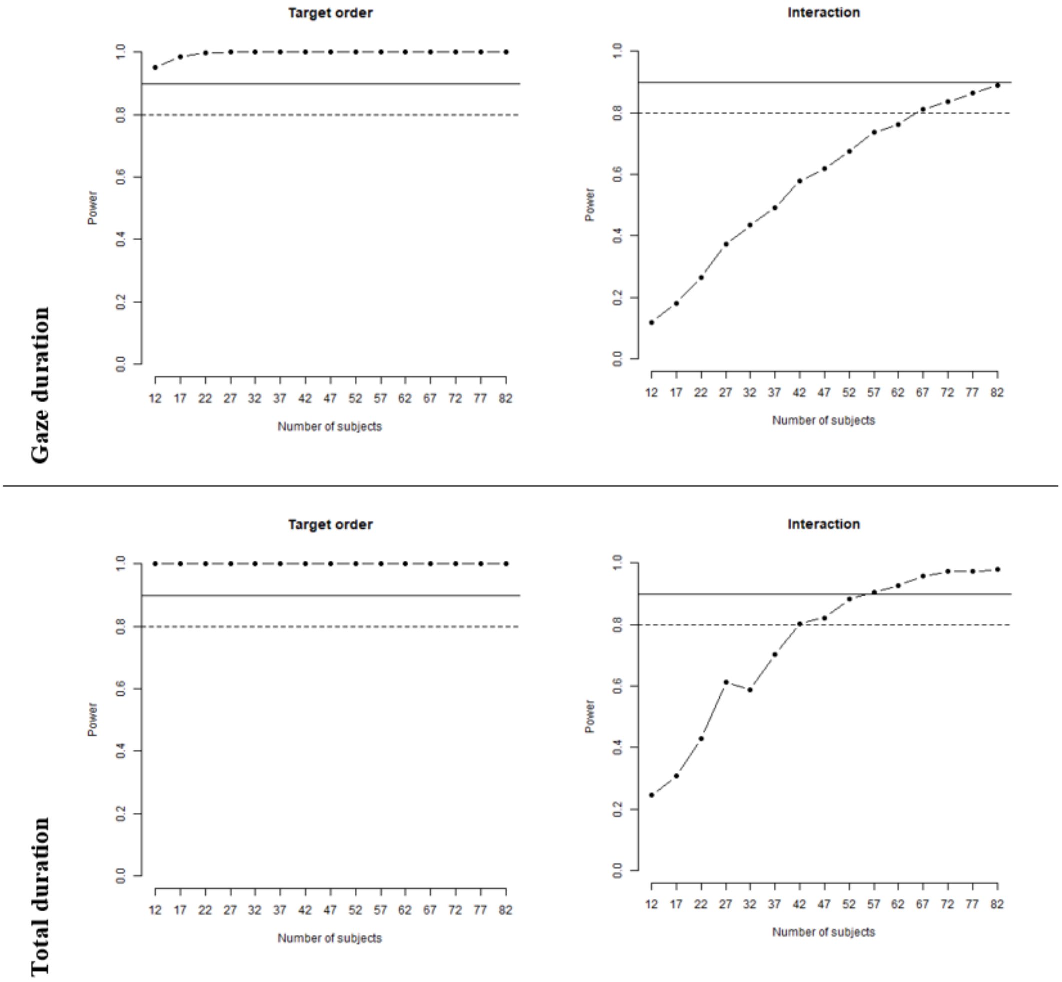


Figure 2.4: Power analysis for the learning task. Effect of target order and interaction between target order and complexity.

Recognition memory

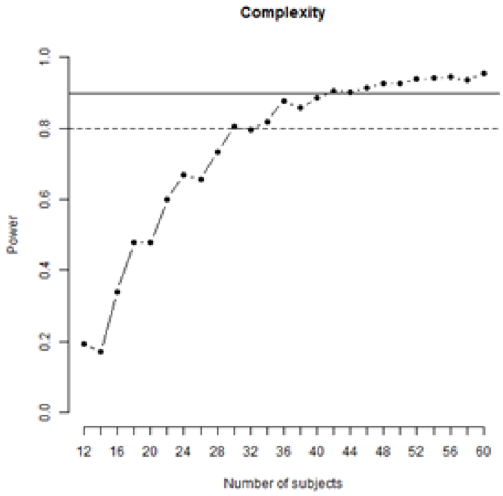


Figure 2.5: Power analysis for the effect of complexity in the recognition memory task.

Considering a sample size of 84 participants and a desired power of 90%, we computed the effect size that would identify as significant in the sentence congruency and the definition selection tasks. For the behavioral part of the sentence congruency task (Figure 2.6), the power analysis showed that we can target an effect size of Beta=0.28. For the eye-tracking part, we are able to target an effect of Beta=0.19 and Beta=0.15 for gaze and total duration, respectively (Figure 2.7). Finally, in the definition selection task we can detect an effect size of Beta=0.31 (Figure 2.8).

Sentence congruency - behavioral

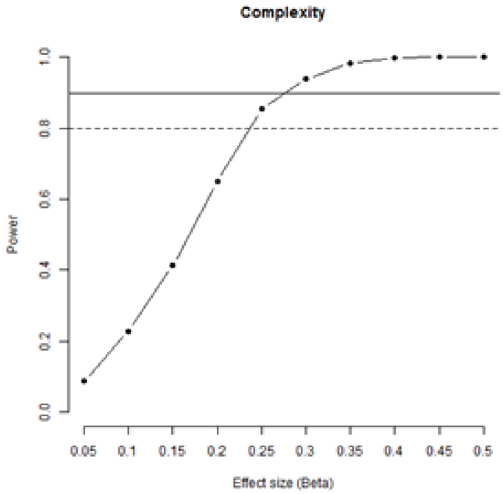


Figure 2.6: Power analysis for the effect of complexity in the sentence congruency task, as tracked by the explicit judgment.

Sentence congruency - eye-tracking

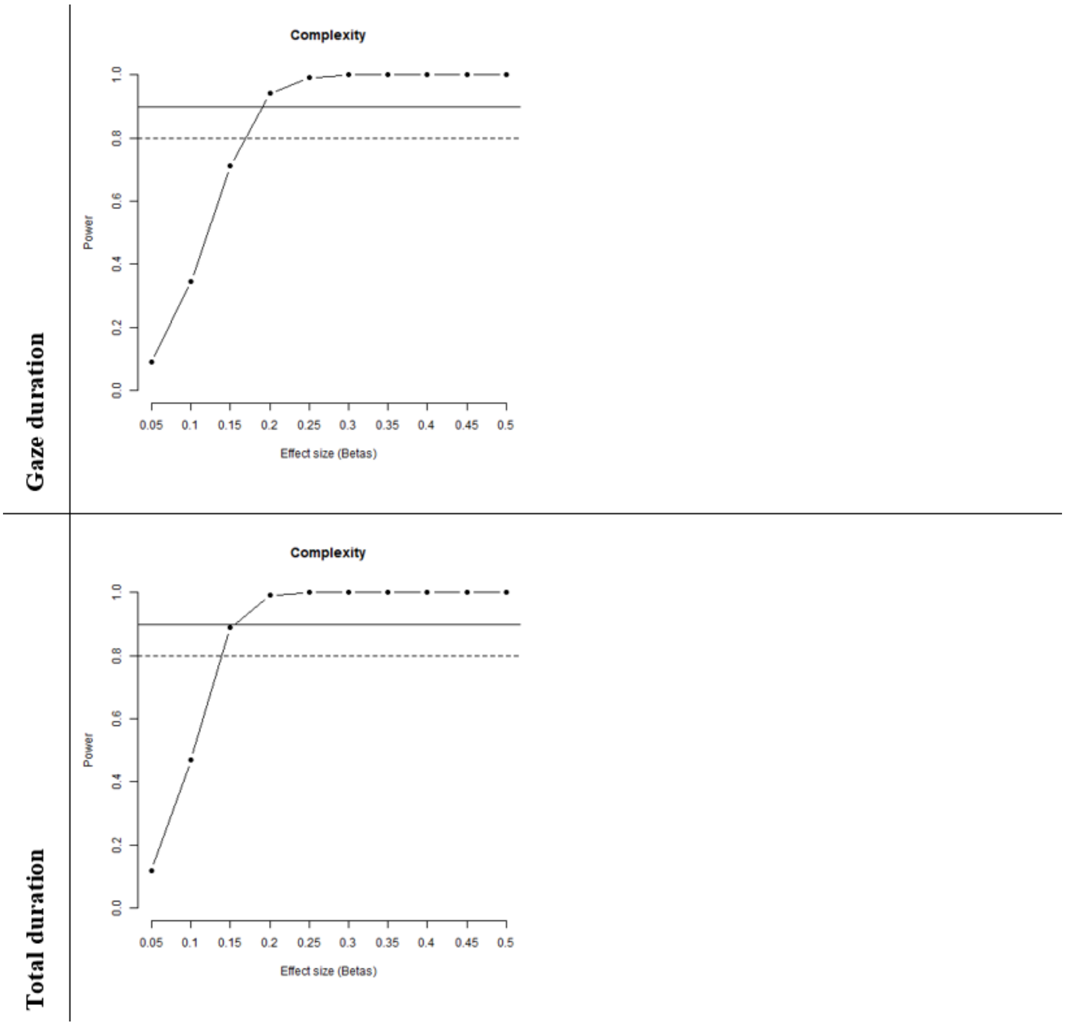


Figure 2.7: Power analysis for the effect of complexity in the sentence congruency task, as tracked by the gaze duration and total looking time.

Definition selection

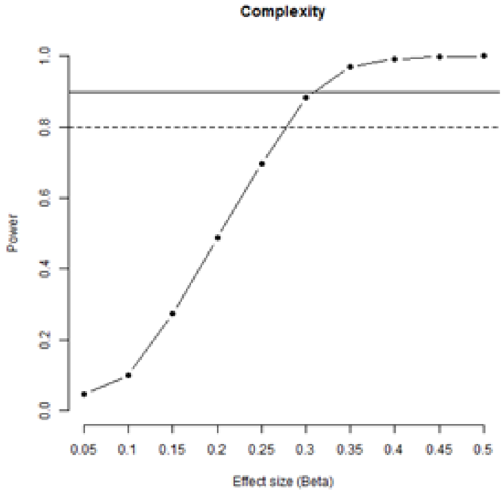


Figure 2.8: Power analysis for the effect of complexity in the definition selection task.

Revised analysis plan

To sum up, in consideration of the pilot experiment and the power analyses illustrated above, this is the revised data analysis plan. Unless stated otherwise, everything remains as described in the Methods section above, and the reader can refer to the analysis script of the pilot data for reference.

For the learning routine, we will analyse gaze duration and total looking times as a function of complexity (suffixed novel words vs. high-frequency ending novel words vs. low-frequency ending novel words), number of encounters and their interaction. We expect the eye tracking metrics to shrink with more encounters with the novel words; this will be interpreted as a sign of word learning. If morphology has a role in this process, we expect the eye tracking metrics to reduce more with number of encounters in suffixed condition compared to the other two. If learning is driven more by letter frequency instead, the reduction in looking times should emerge to a similar extent with suffixed and high-frequency items, but should be stronger in these conditions than in the low frequency condition. This task provides an implicit measure of learning and word representation.

In the Recognition Memory task, we will model accuracy as a function of complexity. Again, if the presence of a suffix enhances learning, we expect accuracy to be higher in the suffix condition, as compared to high-frequency and low-frequency ending items. If cluster frequency is what makes learning stronger, then we'd expected suffixed novel words to pattern with high-frequency items, and both conditions should yield higher accuracy than low-frequency ending novel words. This task provides an explicit measure of learning and word representation. To control for potential biases towards a YES or a NO response, particularly given their different number in the current design, we will also conduct a d-prime analysis, as per the pilot data above. That is, we will compute the average d-prime by condition, to check that whatever pattern will have emerged in the accuracy analysis above is confirmed with a measure of accuracy that is independent of response bias. This analysis will be used descriptively, in support of the main accuracy analysis (hence the lack of a power analysis here).

For the sentence congruency task, we will model again gaze duration and total looking time as a function of complexity, congruency with the carrier sentence, and their interaction. If the participants extract the meaning of the stem from the novel words, they should exhibit longer reading times in the incongruent condition. If the stem is more successfully extracted in the suffixed novel words, the effect of congruency is expected to be stronger for these items. If, however, it is due to frequency of occurrence, we expect to see a difference between suffixed and high-frequency items on one side, and low frequency items on the other. This task will provide an implicit measure of learning of the stem's meaning/representation.

In the sentence congruency task, participants will also provide a congruency judgment – whether the sentence makes sense or not. This provides a more explicit index of stem learning. The model will be the same as above; the accuracy of the congruency judgment will be modeled as a function of complexity and congruency itself. We will only focus on the complexity effect here, and the possible interpretation of the results is the same as above; if stem learning is triggered by morphology, then accuracy should be higher with the suffixed novel words than in the two non-suffixed conditions. If the learning is triggered instead by the frequency of letter chunks, we expect the high-frequency ending items to pattern with the suffixed novel words, and both conditions being better than low-frequency endings.

For the definition selection task, we will model accuracy as a function of complexity and, again, we expect suffixed items to be better than the non-suffixed conditions, if stem learning is triggered by morphology. We expect instead the suffixed and high-frequency ending conditions to be better than the low-frequency ending condition if the learning of the stem is triggered by the frequency of the final cluster.

Of course, it is also logically possible that some stem learning will happen in all three conditions, without significant differences. In fact, the pilot data suggest that this scenario is not so unlikely. Such pattern would indicate that readers tend to assign meaning to sub-lexical chunks independently of the structure of the novel words – or at least, independently of the factors we manipulated here, that is, letter chunk frequency and meaning.

As an exploratory analysis, we will also investigate whether the eventual morphological effects that might emerge in the analyses described above correlate with each individual participant's morpheme interference effect. We will perform this analysis by using the morpheme interference score as a further fixed effect in the models described above.

Modifications of the materials upon revision

Upon a reviewer's suggestion, we decided to validate the sentence congruency manipulation through a questionnaire. We created two surveys, one for each rotation, so that the raters would see each novel word only once, either in the congruent or the incongruent sentence (see Methods, Sentence congruency section). We asked the respondents to assign ratings from 1 (completely implausible) to 10 (completely plausible). A definition of each novel word was provided with each sentence. We collected data from 41 native Italian speakers, 21 for rotation 1 and 20 for rotation 2. The standardized scores are illustrated in Figure 9, and are overall very encouraging; what we designed as congruent sentences were rated much higher in plausibility than what we designed as incongruent sentences (1Q: 0.48 vs. -0.79; median: 0.63 vs. -0.67; 3Q: 0.75 vs. -0.45; mean: 0.59 ± 0.26 vs. -0.59 ± 0.28). However, the ratings were quite close in the congruent and incongruent condition for some novel words. As a cut-off metric, we took the difference between the medians in the congruent and incongruent sentences. There were four items for which this metric was below 1: *cettobo*, -0.97, *ceveco*, -0.43, *clivuno*, -0.82 and *criboto*, -0.92. Since the data were standardized within subjects, this cut-off corresponds to a distance of one standard deviation in each participant's response. We decided to change these sentences, and collected further ratings with the revised items with a new group of 22 native speakers. The relevant medians differed enough to satisfy our criterion: *cettobo*, -2.15, *ceveco*, -1.53, *clivuno*, -1.84, *criboto*, -1.04). We will therefore administer the revised sentences for these items in the main experiment.

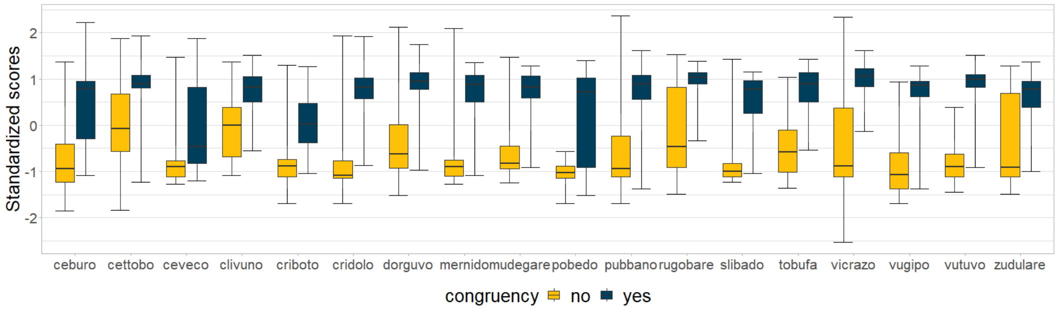


Figure 2.9: Rating scores across congruency for each novel word, as represented by boxplots. Scores were standardized within subjects, so that we got rid of any effect related to a different use of the scale (some participants tended to give higher scores than others, as one might expect).

Study design table

To sum up, the design table below illustrates the general questions we'll be addressing in the Registered Report, and how they translate into specific hypotheses and analysis plans. This table does incorporate the insight we obtained from the pilot study.

Question	Hypothesis	Sampling plan	Analysis Plan	Sensitivity of the test	Interpretation given different outcomes	Theory that could be shown wrong
Does the presence of a suffix or a high-frequency word ending (vs. a low-frequency word ending) modulate the (implicit and explicit) learning of a novel word?	If morphology facilitates word learning, suffixed items will have higher accuracy in the Recognition memory task compared to the high-frequency and low-frequency endings (explicit learning). This might also translate into an effect on implicit learning; if so, gaze duration and total looking time for suffixed items will shorten more across subsequent encounters with the novel words as compared to the other conditions.	We will use Null Hypothesis Significance Testing. We determined sample size via power analysis and obtained an estimate of effect size in a pilot study.	We will use (generalized) linear mixed effect models. See the Methods and the Power analysis section for details. The analysis pipeline is completely specified in the OSF repository.	Pilot data and power analysis.	See above in the table. At the most general level, if suffixed words contrast with high- and low-frequency ending words, we will attribute the experimental effects to a genuine effect of morphology. If instead suffixed and high-frequency ending words will pattern and contrast with low-frequency ending words, we will draw the conclusion that letter chunk frequency is the main driver of the effects.	There is no general theory of word learning that we're trying to specifically address here. Rather, we want to start establishing the role of morphemes in the processing, with a particular focus on meaningfulness vs. letter chunk frequency.

Continued on the next page.

Continued table:

Question	Hypothesis	Sampling plan	Analysis Plan	Sensitivity of the test	Interpretation given different outcomes	Theory that could be shown wrong
Does the presence of a suffix or a high-frequency word ending (vs. a low-frequency word ending) trigger meaning attribution to a novel stem?	Stem meaning extraction should happen specifically for suffixed words, if it is driven by morphology, or for suffixed and high-frequency ending words, if it is driven instead by letter frequency. This will be measured through: (i) accuracy in the sentence congruency task; (ii) gaze duration in the sentence congruency task; (iii) total looking times in the sentence congruency task; and (iv) accuracy in the definition selection task.	We will use Null Hypothesis Significance Testing. We determined sample size via power analysis and obtained an estimate of effect size in a pilot study.	We will use (generalized) linear mixed effect models. See the Methods and the Power analysis section for details. The analysis pipeline is completely specified in the OSF repository.	Pilot data and power analysis.	See above in the table. At the most general level, if suffixed words contrast with high- and low-frequency ending words, we will attribute the experimental effects to a genuine effect of morphology. If instead suffixed and high-frequency ending words will pattern and contrast with low-frequency ending words, we will draw the conclusion that letter chunk frequency is the main driver of the effects.	There is no general theory of word learning that we're trying to specifically address here. Rather, we want to start establishing the role of morphemes in the processing, with a particular focus on meaningfulness vs. letter chunk frequency.

2.4 Main experiment

Participants

Eighty-five adults (12 male; mean age = 24.36 years, SD = 3.05 years) took part in the experiment, one more than required by the power analyses. They were all native speakers of Italian, with normal or corrected-to-normal vision and no reading disabilities. Participants were paid 15 Euros for their participation.

Exclusion criteria

As described in the registered methods, participants for which the d-prime was below 1 in the Recognition memory task were excluded from any further analysis; they were considered unsuccessful learners. This procedure led to the exclusion of 9 participants.

Methods

Testing procedures, apparatus and software were identical to the pilot study, except that we used a desk-mount instead of a tower mount EyeLink 1000 eye tracker.

2.4.1 Results - Registered analyses

Learning task

The novel words were fixated 5.61 times on average, with a relatively small proportion of single fixations (27%). This mirrors the results of the pilot study.

As per the approved protocol, we ran linear mixed models separately for gaze durations and total durations, with low frequency endings items set as a baseline. We found a strong influence of number of encounters in both measures (gaze duration: $b = -0.06, t = -22.06$, CI = [-0.07 – -0.06], $p < 0.001$; total duration: $b = -0.11, t = -41.04$, CI = [-0.12 – -0.11], $p < 0.001$). Moreover, both measures showed that complexity influenced reading, with an advantage for suffixed items compared to both high frequency and low frequency endings (gaze durations: $b = -0.15, t = -3.70$, CI = [-0.22 – -0.07], $p < 0.001$; total durations: $b = -0.10, t = -2.63$, CI = [-0.18 – -0.03], $p = 0.01$). For total duration, a significant interaction emerged between suffixed items and order of presentation ($b = -0.01, t = -2.28$, CI = [-0.02 – 0.00], $p = 0.02$); suffixed items had a larger advantage of order of presentation compared to the other conditions. No such effect surfaced in gaze durations ($b = 0.00, t = 0.71$, CI = [-0.01 – 0.01], $p = 0.48$). There was no indication of any difference between high and low frequency endings. Figure 10 represents this set of analyses.

Recognition memory task

The mean d-prime across participants was 2.34 [-0.66 – 5.7], showing that the participants as a group performed the task very well. However, nine individuals had a d-prime lower than 1; as per the approved protocol, we removed them from any further analysis. The mean overall accuracy was .86, confirming that participants learned the novel words successfully. Accuracy was influenced by complexity, with better recognition for suffixed items (mean = .88, SD = .33) as compared to high frequency (mean = .80, SD = .41) and low frequency endings items (mean = 0.79, SD = .40). Generalized linear mixed models confirmed the descriptive results, showing a significant advantage for suffixed items ($b = 0.61, z = 2.34$, CI = [0.10 – 1.12], $p = 0.01$).

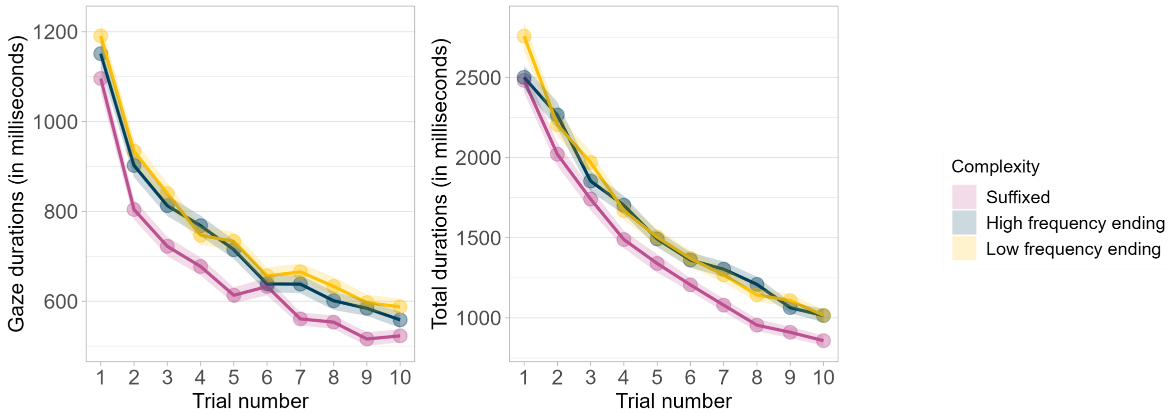


Figure 2.10: Gaze durations and total durations as they changed during the experiment, across the three complexity conditions. The shadowed areas represent the standard error of the mean.

Sentence congruency task

Behavioral

The overall accuracy in the behavioural task was .64. This represents a fairly good performance, which is significantly different from chance[4]. However, it is also considerably worse than the Recognition Memory task, which indicates that wordforms were learned much better than the meaning of their stems. Despite the mean is somewhat different in the three conditions (SF: mean = .69, SD = .46; HF: mean = .61, SD = .49; LF: mean = .61, SD = .49), generalized mixed effects model failed to show any significant difference (all $ps > 0.1$).

Eye tracking

The average number of fixations was 5.33, with a mean number of refixations 0.7, similarly to the pilot study, probably reflecting an effort to assign a meaning to the novel word. Results showed no significant effect for gaze durations (all $ps > 0.1$). However, an effect of complexity emerged as shorter total duration times for the suffixed items ($b = -0.12, t = -2.32$, CI = [-0.22 – -0.02], $p < 0.02$), as compared to high and low frequency endings. More importantly, congruent sentences required less reading time ($b = -0.14, t = -2.76$, CI = [-0.24 – -0.04], $p < 0.01$). However, congruency did not interact with

[4]The Beta for the intercept in the model was 0.4, with a Standard Error of 0.2, which corresponds to a z value of 2.02 and p value of 0.04. Because the intercept represents performance in the baseline condition, which if low frequency endings in this case, and because the logistic link function implies that 0 represents chance level, these statistics attests that participants were significantly above chance in the low frequency ending condition — and therefore, in the other conditions, too.

complexity ($b = 0.06, t = 0.90$, CI = [-0.08 – 0.21], $p = 0.4$).

Definition selection task

Overall accuracy was .75, indicating that participants as a group were able to assign the correct definition to the base word. Replicating the pattern from the pilot study, the complexity of the novel word in the learning phase doesn't seem to have affected performance: participants picked the correct definition 74% of the times (SD = 0.44) for base words in the suffix condition, and 76% of the times for base words in the high-frequency ending and low-frequency ending condition (SD = 0.43 and SD = 0.43, respectively). The linear mixed-effects model confirms this pattern (all $ps > 0.1$).

2.4.2 Results - Exploratory analysis – Sensitivity to morphemes

As indicated in the *Revised analysis plan* section, we ran exploratory analyses to check for the influence of sensitivity to morphemes in the tasks where effects of complexity were found. This is achieved by incorporating the morpheme interference index (MIF) in the relevant models, as a measure of participants' sensitivity to the presence of morphemes in a nonword. *Learning task* The morpheme interference index (MIF) did not influence either gaze or total durations as a fixed effect. However, MIF interacted significantly with trial order (gaze durations: $b = -0.01, t = -2.18$, CI = [-0.01 – 0.00], $p < 0.03$; total durations: $b = -0.01, t = -2.33$, CI = [-0.01 – 0.00], $p < 0.02$); a larger gain in reading time across multiple encounters with the novel word was associated with stronger sensitivity to morphology. Most importantly, suffixed items specifically interacted with MIF in gaze durations ($b = -0.05, t = -2.56$, CI = [-0.09 – 0.01], $p < 0.01$), with more advantage in the suffixed condition for participants who are more sensitive to morphology. Additionally, suffixed items were involved in a triple interaction, involving the order of presentation and MIF ($b = 0.01, t = 2.90$, CI = [0.00 – 0.02], $p < 0.004$). Participants with higher sensitivity to morphemes displayed lesser learning across multiple encounters with the novel suffixed words, although this is likely a side effect of these participants starting off with shorter gaze durations upon their first encounter (see Figure 11, left panel).

Recognition memory task

The generalized linear mixed-effect model applied to accuracy data revealed

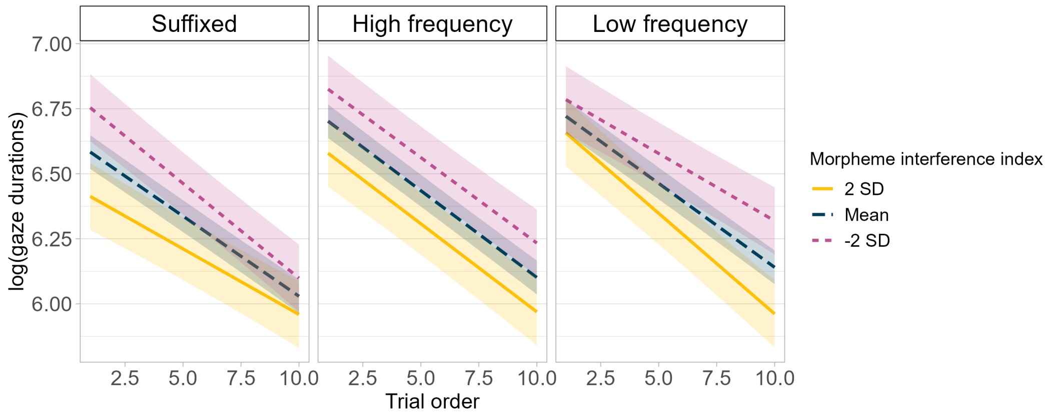


Figure 2.11: Predicted gaze duration as a function of the interaction of the order of presentation and morpheme interference index. The grey band is the 95% confidence interval.

no interactions between sensitivity to morphology and complexity (all $ps > 0.1$).]

Sentence congruency task

Eye tracking

The generalized linear mixed models ran on total durations revealed no interactions between morphology and neither complexity nor congruency (all $ps > 0.1$).

2.5 Discussion

The goal of the present experiment was to explore the role of affixes in novel word learning during sentence reading. To this aim, we presented participants with novel words composed of a novel stem and a familiar affix (e.g., *flib-er*). These items were contrasted with novel words featuring non-meaningful endings that were either matched in frequency with the suffixes (e.g., *flib-an*) or much less frequent (e.g., *flib-ov*). In this way, we were able to separate the role of morphemes as meaning-bearing units vs. chunks of letters that frequently co-occur (i.e., that are statistically associated). We assessed both participants' memory for the novel word itself and their ability to assign meaning to the novel stem, both implicitly (e.g., via eye tracking) and explicitly (e.g., via recognition memory).

In the first task of our experimental paradigm (the Learning task), participants encountered each new word in 10 different sentences, while their eye movements were monitored. We found that both gaze durations and total looking times decrease with every subsequent encounter with the novel word. This indicates that processing became increasingly less effortful, suggesting that participants were more familiar with the items to be learned. This is in line with the results of the pilot study and confirms that the paradigm was successful and did trigger word learning (see also Pagán & Nation, 2019; Ginestet et al., 2020; Joseph et al., 2014). Suffixed items were consistently read more quickly throughout the entire task, both in gaze durations and total looking time, probably reflecting a relatively higher familiarity with the encounter of a novel stem attached to a familiar affix, as compared to non-affixal word endings. These findings replicate the pattern found by Ginestet and colleagues (2020), who also found reduced gaze durations and total durations for complex pseudowords compared to their orthographic counterparts. The frequency of the endings doesn't seem to play any relevant role here.

Most importantly, word type interacted with number of encounters, reflecting a more substantial gain in processing speed for suffixed compared to non-suffixed items. In other words, not only did suffixed words lead to shorter fixations overall, but they also yielded a stronger learning effect. Again, the frequency of the novel word ending doesn't seem to play any particular role here. This interaction only emerged in total looking times, not in gaze duration. This indicates that the effect is probably driven by refixations: apparently, readers have less of a need to revisit unfamiliar words when they are suffixed. This could be explained both by orthographic familiarity – the novel item looks more natural – or by semantics – readers were able to attribute meaning to the novel item, at least to some extent, already during first pass. We will expand upon this in the subsequent sections when we discuss the results of the novel stem tasks.

The results of the Recognition Memory task confirm the main findings illustrated above: suffixed items were easier to remember and recognize as compared to non-suffixed words, and the frequency of the word ending did not appear to have a significant impact. Therefore, the learning pattern appears to be the same, independently of whether it was tested implicitly (that is, via eye tracking) or explicitly (via recognition memory). Note that the distractors contained also recombinant items, i.e., combinations of stems and endings that both appeared in the Learning task, but not together; this made it impossible for the participants to perform the task based on the identification of single word parts. Importantly, the present results replicate Tamminen al.'s (2015) results, whose recognition memory task served as a model for ours. In their experiment, Tamminen and colleagues found that novel complex words (which in their experiment consisted of an existing stem and a novel affix) were successfully recognized after training. However, their study showed no immediate post-training effects in the measures of implicit learning. One important difference here is that we tested for implicit learning *during* training, while Tamminen and colleagues administered their implicit learning task *after* training, as is the case in other word learning studies tapping into explicit/implicit representations (e.g., Batterink & Neville, 2011; Qiao & Forster, 2013). We do not see any obvious reason why this might have caused inconsistent results; yet, it is an important procedural difference, which we thought it was important to mention.

Our findings are congruent with the general consensus that affixes are easily identifiable within a word and constitute a critical processing unit (Amenta & Crepaldi, 2012; Beyersmann et al., 2012, 2016, 2020; Grainger et al., 2021; Leminen et al., 2016). These results are also in line with the abundant evidence showing that affixes play a critical role even in non-words (Beyersmann et al., 2013; Burani et al., 2008; Crepaldi et al., 2016; Dawson et al., 2018; Hasenäcker et al., 2016; Rastle et al., 2000; Taft & Forster, 1975). We go beyond these findings by showing that suffixes play a prominent role also during the word learning process.

We also showed that the morphological advantage we uncovered here is due to the fact that affixes are meaning-bearing units; frequency of occurrence doesn't seem to play any role. On the one hand, this is consistent with previous findings that nonwords with suffixes are more easily assigned meaning compared to simple nonwords (e.g., Dawson et al., 2021; McCutchen & Logan, 2011). On the other hand, however, the finding that high-frequency endings didn't exhibit any similarity with the suffixed items is at odds with the general pattern found in masked priming studies, where pseudo-complex words (e.g., corner) still display morphological effects (e.g., Rastle et al., 2004; Rastle & Davis, 2008; Longtin et al., 2003). Moreover, Lelokiewicz et al. (2020) demonstrated that readers automatically identify chunks of unfamiliar pseudoletters (that is, with no connection to orthography, phonology or meaning) that resemble suffixes in their statistical features, indicating the significance of regularities in letters co-occurrence.

A fairly straightforward explanation for these discrepancies lies in the nature of the tasks employed. Our study focused on the learning of the meaning of the novel words (e.g., they were presented in the context of otherwise familiar sentences), while masked priming studies primarily address early processing stages that rely heavily on the visual characteristics of words rather than their semantics. Along the same lines, Lelokiewicz et al. (2020) deprived the input of any linguistic information: it's only natural that learners would rely on the only information available – in this case, the statistics of the visual symbols.

In addition to learning the novel words themselves, our participants used the contextual information provided by the carrier words and sentences to infer the meaning of the novel stem. This is the main novelty of our paradigm; in contrast to previous work that focused predominantly on the learning of novel words with a morphological structure (Ginestet et al., 2020), our study also evaluates whether participants actually attribute meaning to the novel stems. Here we show that this is the case in three ways. Firstly, at the explicit level, readers were able to (i) distinguish sentences that were congruent with the meaning of the stem from sentences that were not, and (ii) to identify the correct definition of the stem among distractors. These effects appear to be smaller in size than for the whole-word learning (although it is quite difficult to make a direct comparison due to different nature of the tasks and measures); nonetheless, they are clearly statistically significant. Secondly, at the implicit level, novel stems embedded in sentences whose meaning was incongruent with that of the stem tended to elicit longer fixations. Thus, it appears that the process of meaning assignment to the novel stems was deep enough to emerge both when participants' intuitions were probed directly with an explicit question (e.g., does this sentence make sense?) and in a kind of behaviour that leaves little access to conscious, strategic control (eye movements during reading).

Interestingly, the eye movement effect only emerged in total looking times; during first pass reading (i.e., in gaze duration), congruent and incongruent sentences didn't trigger any different visual exploration of the text. This suggests that during first pass, visual/orthographic familiarity with the stem was the main driver of readers' behaviour, while semantic integration with the sentence context only happened later in time, when participants' eyes had already moved forward to the text following the critical word. Of course, this doesn't necessarily generalize to everyday reading. This result might be fully dependent on the fact that our participants were only starting to acquire the novel words, which they encountered only ten times in total during the learning phase of the experiment While this is a solid number for a controlled, lab-based experiment such as ours (for a similar design, see Joseph et al., 2014;

Joseph & Nation, 2018; Pagán & Nation, 2019), it is still far from any realistic exposure in everyday life. This might be the reason why meaning integration was slow in the present experiment – quite possibly slower than during the reading of entirely familiar sentences.

Another critical aspect of these stem meaning effects is that they don't seem to be specifically connected to morphology. In fact, we should probably not speak of stem meaning entirely, but rather, of some more general attribution of meaning to word parts. This is shown by the lack of statistically significant difference between suffixed, high-frequency and low-frequency words in the eye tracking measures, as well as in the explicit judgments on sentence congruency and in the definition selection task. Thus, quite interestingly, there was no need for the presence of an affix in order for the readers to assign meaning to word parts. As we discussed in the Registered Protocol, before any data were collected, this finding is very relevant theoretically. It reveals that readers would assume a general correspondence between orthography and semantics, such that if, e.g., *fliber*, *fliban* or *flibov* have a meaning related to food, then *flib* must too – independently of the fact that *-er* is an affix, while *-an* and *-ov* aren't, and also independently of the fact that -er and -an are very frequent word endings, while *-ov* is not. It appears that when we encounter a novel word, the process of semantic generalization is not specifically driven by the identification of sub-lexical chunks with a clear meaning on their own. Instead, it is more likely that readers assume a broader, though possibly less precise, correspondence between a word's form and its meaning, one that doesn't require a discretized analysis of lexical items or the rule-based combinatorial computations often implied by morphological theory. These findings align more closely with work and theories placing morphology in the wider context of linguistic regularities (e.g., Baayen et al., 2011; Marelli et al., 2015; Amenta et al., 2020; Siegelman et al., 2022). In this context, the brain actively searches for any probabilistic relationships between form and meaning, and then leverages this information to interpret novel input.

The individual variability data seem to support this view, at least in part. In the exploratory section of the analyses, we examined whether readers' sensitivity to morphology, as assessed using a morpheme interference task, could explain their learning patterns – either of the whole words or of the stems' meaning. We didn't observe any effect for what concerns the latter: the size of each participant's morpheme interference effect doesn't seem to correlate with their visual exploration of the novel stems in congruent compared to incongruent sentences. Similarly, it did not correlate with their ability to explicitly discern if a sentence is meaningful or nonsensical, or to identify the definition of the novel stems.

On the other hand, we did obtain a morpheme interference effect in the learning task. We would tend to discard the highest level, three-way interaction as quite irrelevant theoretically. In the suffixed condition, there seems to be a lesser learning effect for those individuals with higher sensitivity to morphemes (see Figure 2.11). This is clearly theoretically implausible. Indeed, the effect might be explained by the fact that suffixed words attracted relatively quick fixations already at the beginning of the learning routine in individuals with strong morpheme interference effects. As a consequence, there would simply be less room for improvement compared to people less sensitive to morphemes.

We tend to trust more the two-way interactions, because their statistical power was certainly higher and because they depict a more coherent theoretical message. People with a stronger morpheme interference effect generally displayed stronger learning across conditions; and also, they showed a larger advantage for suffixed items overall, across successive encounters with the novel words. This latter result suggests that the morpheme interference effect might capture some specific morphological sensitivity independent of the learning effect shown in the present study. Conversely, the former finding, points again to some general learning ability, which might be reflected in the morpheme sensitivity score, but, again, cuts across conditions in the present experiment – that is, does not seem to be specific for affixes.

In summary, the present study revealed that suffixes significantly contribute to novel word learning. This influence is driven by their semantic component, not by the statistical associations between letters that morphology implies. Furthermore, this effect emerged in both explicit and implicit tasks. Importantly, readers seem to generalize the meaning of the whole word to its constituent parts, but this process is not specifically triggered by the presence of an affix. Instead, it seems to reflect a more general attempt of the brain to identify regularities in the mapping between form and meaning.

Chapter 3

Experiment 2: Does morphology support novel word learning through statistical learning?

Manuscript submitted to the *Memory & Cognition* journal. All the materials related to this paper (stimuli, data, and code for data analysis and power analysis) are publicly available at the Open Science Framework: https://osf.io/9kt74/ The study is preregistered at https://aspredicted.org/WBL_-QSF.

3.1 Abstract

Most novel words that speakers learn are morphologically complex (e.g., *columnist*, *whistleblower*). But far from acting as an impediment, such morphological complexity might actually facilitate word learning. In particular, affixes (e.g., *pre-*, *-ness*) could do so in at least two ways: because they are associated with meaning or because they are frequent clusters of letters. This information is statistical in nature and might thus be captured via statistical learning. To investigate this, we conducted a preregistered novel word learning experiment with Italian native speakers, who learned words that have (i) existing suffixes (*rugob-enza*, akin to *spoot-er* in English), (ii) non-meaningful endings matched in frequency (*rugob-ondo*, *spoot-an*), and (iii) non-meaningful, low-frequency endings (*rugob-allo*, *spoot-ov*). Participants also completed a visual statistical learning task. Results show that items with suffixes and low-frequency endings

were learned best. Moreover, these two types of words exhibited the strongest correlation with general statistical learning ability. We discuss these results in the context of ongoing debates about the role of statistical learning in reading.

3.2 Introduction

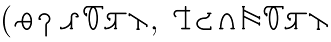
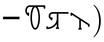
Compositionality is a core feature of language that allows for the meaning of a word to be derived from the meaning of its parts (e.g., *kind-ness*, *like-able*, *buy-er*). For example, the suffix -able indicates the capacity to be a certain way (e.g.*usable*, *likeable*). This is a highly efficient way to construct new lexical items because understanding the meaning of the verb and the suffix provides a complete understanding of the word, even if the specific combination has not been encountered before. It comes as no surprise that morphology is pervasive in the lexicons of the world's languages. In English, for instance, approximately 80% of words consist of more than one morpheme (Castles et al., 2018).

Morphology has been studied thoroughly within the field of visual word recognition where research has established morpheme recognition as a crucial step in the reading process. Most of this evidence comes from lexical decision studies. For example, it has been consistently demonstrated that the time taken to recognize a complex word is directly related to the frequency of occurrence of its stem (e.g., New et al., 2004). In skilled readers, morpheme identification is automatized to such an extent that people decompose complex words into their constituent morphemes even outside of awareness, as shown in masked priming studies (e.g., Longtin et al., 2003; Rastle et al., 2004). Due to their semantic role, morphemes occur frequently across different words; that is, affixes are recurring clusters of letters (e.g., *-able* is a much more frequent letter cluster than *-oble* or *-afle*). Their twofold nature as meaning-bearing and statistically associated units presents an intriguing contrast with other letter clusters that are equally prevalent in language, but which do not carry meaning (e.g., *con-*, as in *convert*, *contain*, *confetti*, etc.).

Differentiating the causal influence of frequency and meaning in morphological effects is challenging since these aspects are closely intertwined in natural language. Nonetheless, research has started to examine the specific role of orthographic factors in morphological effects. Longtin et al. (2003) and Rastle et al. (2004) found that words that only have an apparent morphological structure facilitate the recognition of their "stems" in the initial stages of processing (i.e., *corner* primes *corn*, even though corner is not somebody who *corns*). This behavior mimics the effects observed with genuinely morphologically complex words (*farmer* primes *farm*). Such evidence strongly suggests that segmentation into morphemes can occur independently of semantic factors (Amenta et al., 2016; Hasenäcker et al., 2016; Marelli et al., 2015). Notably, these effects depend on the morpheme position within the word. Crepaldi, Rastle, and Davis (2010) demonstrated that suffixes are processed as morphemes only when they appear in their typical position. For example, the nonword *gasful* produces interference in a lexical decision task, while *fulgas* does not. This body of evidence has fostered several competing models of complex word processing (Beyersmann & Grainger, 2023; Crepaldi et al., 2010; Grainger & Ziegler, 2011; Taft & Nguyen-Hoan, 2010). While these models disagree on several important aspects (e.g., the role of bound morphemes) and are certainly underspecified in important ways (e.g., how are positional constraints implemented?), they build a fairly clear picture of the role of morphemes in the identification of printed, familiar words.

Importantly, sensitivity to morphemes extends to nonwords as well. For example, in Burani et al. (2002), nonwords made up of stems and suffixes (e.g., *woman-ist*) were categorized more frequently as possible words and were named more quickly and accurately than matching nonwords without suffixes (e.g., *woman-ost*). Similar findings emerged in several other languages (e.g., Spanish: Duñabeitia et al., 2008; English: Taft & Forster, 1975). However, it is still unclear whether this information is utilized during word learning; none of these experiments specifically explored the learning of unfamiliar stimuli as novel, meaningful lexical items.

A step in this direction has been made by Tamminen et al. (2015). In several word learning experiments, they explored how participants acquire novel words composed of familiar stems and new suffixes (e.g., *crab-afe*). Definitions for the novel words were created such that the meaning of each novel affix modified the stem in a consistent way. The results indicated that participants were able to extract the meaning of the suffix and generalize this knowledge to untrained novel words after a period of memory consolidation. Other related studies (e.g., Dawson et al., 2021; Havas et al., 2015) confirmed that humans naturally acquire morphological information when learning unfamiliar words, even without explicit instruction about morphology. Importantly, none of the above-mentioned studies focused on a situation that happens very often in real life—coming across a new lexical item formed of a familiar suffix attached to an unfamiliar stem. With that in mind, Ginestet et al. (2020) used nonwords composed of nonword stems and existing prefixes and suffixes (e.g. *re-lerb-er*) in an eye-tracking experiment. The looking time patterns were somewhat unclear (e.g., morphological structure played a role in gaze durations, but not in several other metrics, such as number of fixations and first-of-two fixation durations), but the post-training behavioral findings showed that complex words were spelled more accurately than orthographic controls.

A recent perspective emerging from the field of statistical learning offers a fresh viewpoint. As mentioned above, morphology establishes regularities not only in the mapping between form and meaning, but also in the way letters co-occur within words. Lelonkiewicz et al. (2020) trained participants using pseudoletter strings (), each containing an affix-like chunk () that occurred across several items. These affix-like chunks were positioned consistently within the string. During the testing phase, participants were presented with newly composed items, some of which included the trained affixes. Results showed that participants were more inclined to categorize newly composed testing items as belonging to the training set if they contained a trained affix. Because the experiment did not involve any orthographic, phonological, or semantic information, these data suggest that the extraction of morpheme-like chunks was based solely on visual statistical information, suggesting that there could be a role for statistical learning in structuring linguistic input. In a further study by the same group (Lelonkiewicz et al., 2023), linguistic information was progressively factored in (e.g., by using real letters instead of pseudocharacters, or associating strings with objects), and the learning effect increased in size.

Interestingly, the connection between reading, visual word identification, and our ability to discern the numerous regularities present in (written) languages has been explored beyond the domain of morphology. After all, letters might be statistically associated more generally, independently of whether they are part of morphemes. For instance, in English, upon encountering the letter C, we might naturally expect the letter A to follow, while the letter S is less likely. Chetail (2017) also used an alphabet unfamiliar to participants and embedded regularities (e.g., bigrams and their specific position within a letter string) in a set of novel words that participants were asked to learn. Participants were found to judge novel combinations of letters containing frequently occurring bigrams as more word-like compared to random letter combinations. Similar to Lelonkiewicz et al. (2020), this research removed any semantic, orthographic, or phonological information, yet the observed effects align with those typically associated with orthographic processing.

There have also been efforts to study more directly the contribution of statistical learning to reading skills. One promising avenue in this respect is the study of individual differences (e.g., Siegelman, 2020). This approach offers clear predictions at the individual level: if statistical learning is a relevant mechanism in language acquisition, individuals who excel in statistical learning should also excel in language learning. For instance, Siegelman et al. (2020) predicted that early readers who are more adept at learning statistical regularities would exhibit stronger reading abilities compared to those with weaker statistical learning (SL) abilities. The results indeed showed that readers who were more sensitive to orthography–phonology correspondences in a word naming task were better readers compared to those who relied more on orthography–meaning relationships. It is important to note that the strength of the correlation between statistical learning and reading skills varies across different studies and could be influenced by the specific task or measure used to assess reading abilities (e.g., Schmalz et al., 2019; Siegelman et al., 2017). Recent work (Siegelman et al., 2017; Siegelman & Frost, 2015) has pointed to the fact that the tests used in a large part of the published literature yield reliable effects at the group level, but not at the individual level. In order to tackle this problem, Siegelman et al. (2017) developed a visual statistical learning task that gave encouraging results in this respect, with a rather wide distribution of participants' scores and, most importantly, good reliability.

Overall, morphology clearly plays a very important role in visual word identification and has been shown to provide a valuable anchor when we process unfamiliar, novel words. It is not clear, however, what specific role morphemes play in word learning, and what the specific contribution of letter chunk frequency is in comparison to the specific associations of these chunks with a given piece of meaning. Because these phenomena are all related to linguistic regularities, statistical learning might be a powerful contributing factor. With this picture in mind, we investigated the learning of novel, artificial lexical items formed by a nonexisting stem (e.g., *spoot-*) and: (i) an existing suffix (*spoot-er*); (ii) a letter cluster of similar frequency, but with no association with meaning (*spoot-an*); or (iii) a letter cluster with a lower frequency of occurrence (*spoot-ov*). With such a design, we can tease apart the contributions of meaning and frequency; any difference between items (i) and (ii) can only be due to meaning, while any difference between items (ii) and (iii) can only be due to frequency. To obtain evidence on the potential connection between the appreciation of statistical regularities and the role of morphology in word learning, we complemented this design with a statistical learning task aimed at assessing the general statistical learning skills of our participants (Siegelman et al., 2017), which we correlated with their performance in the three conditions in the main learning task.

The study was preregistered at https://aspredicted.org/WBL_QSF, and included the following hypotheses. First, we hypothesized that in the novel word learning task, high-frequency endings words would be learned better compared to the low-frequency endings words. Second, we hypothesized that suffixes would provide additional advantage; thus, suffixed words were expected to exhibit better performance than high-frequency endings words. Finally, we hypothesized that a correlation might emerge between statistical learning skills and novel word learning, particularly with suffixed items and, perhaps to a lesser extent, with high-frequency endings items.

3.3 Methods

The experiment was programmed using jsPsych (de Leeuw, 2015; version 7.1.2) and was hosted on the Cognition.run platform (www.cognition.run).

Participants

Two hundred and six participants were recruited through Prolific (www.prolific.co). The sample size was computed based on data simulations (see below). All participants were native speakers of Italian with normal or corrected-to-normal vision and no reading disabilities. Prolific filtering system was used to limit the study to this population. They were paid for their participation a base amount of 5 GBP and received an additional bonus of up to 1.14 GBP according to performance. The experiment was approved by the SISSA Ethics committee.

Procedure

The experiment consisted of two main parts. The first part tested novel word learning and the second assessed general statistical learning skills.

Novel word learning

Stimuli

Participants were exposed to novel words that consisted of a stem and an ending. All stems were nonexisting words in Italian without any lexical neighbor (e.g., *rugob-*; an English example would be *spoot-*). Specifically, they were either five or six characters long; their mean OLD20 (Yarkoni et al., 2008) and log bigram frequency were 2.09 (SD = 0.19) and 5.96 (SD = 0.30) respectively. These measures were calculated based on the SUBTLEX-IT word frequency database (Crepaldi et al., 2016). The stems were paired with three types of endings: (i) suffixes (e.g., *-enza*, token frequency mean = 4.77, SD = 0.49), (ii) nonmorphological endings matched to suffixes in token frequency in word-final position (e.g., *-ondo*, mean = 4.96, SD = 0.22), or (iii) nonmorphological endings with low frequency in word-final position (e.g., *-espa*, mean = 2.49, SD = 0.29). The endings were either three or four characters long. We used 72 stems and 24 endings, eight of each type. To create the novel words, we pseudorandomly paired each ending with three stems. Thus, the final set of words to be learned consisted of 72 items, eight to ten letters in length (mean = 9.44) and with no lexical neighbors in Italian. We also made sure that none of the items contained an embedded existing Italian word. All participants were exposed to all the stimuli (i.e., there were no rotations). The complete list of stimuli is reported in Appendix 1.

Procedure

During the learning phase of the task, a novel word appeared at the center of the screen and participants were instructed to read it and remember it. In order to facilitate learning, they had to type it back into a textbox that appeared after a button press. Each word was repeated in a pseudorandom order three times during the learning, so each participant was exposed to 216 learning trials in total. The entire task was self-paced. The learning phase was followed by a two-alternative forced choice task in which a trained item and its foil appeared together at the center of the screen. The foils were created by substituting one of the letters in the stem (e.g. *rugobenza* vs. *rufobenza*) and under the same constraints described above. Two rotations took care of counterbalancing for position (i.e., whether the foil was on the left or the right side of the screen).

Visual statistical learning

This task followed closely the visual statistical learning task described by Siegelman et al. (2017). The only difference was the language of the instructions and the fact that our experiment was run online rather than in a laboratory setting.

Stimuli

The materials consisted of 16 shapes (Figure 3.1), which were used to create eight triplets, further subdivided into four triplets with transitional probabilities (TP) between the shapes equal to 1, and four triplets with TP = 0.33. The combinations of shapes that constitute the triplets were randomized for each participant.

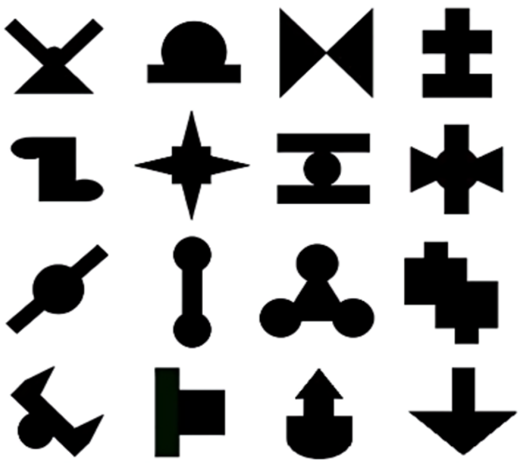


Figure 3.1: Shapes from the exposure phase.

Procedure

In the exposure phase, the shapes were presented individually on the screen for 800 ms, with a 200 ms pause between stimulus presentations. Their sequence was determined by the triplet structure described above. Each of the eight triplets appeared 24 times during this phase, with the constraint that the same triplet could not appear two times in a row. The order was randomized in each run. Overall, the only cue to the structure of the stream were the TPs between the shapes. Participants were instructed to look at the sequence of shapes that appeared on the screen and were told that they would be tested on them. The testing phase included two different tasks. In one, participants were required to recognize a pattern from the training among two, three, or four distractors (34 trials; see Figure 3.2). In the second task, participants were presented with either one or two shapes from a triplet and had to select the appropriate third shape from three alternatives (8 trials; see Figure 3.3). The statistical learning ability of each participant was computed as the total number of correct trials in these two tasks (total N=42).

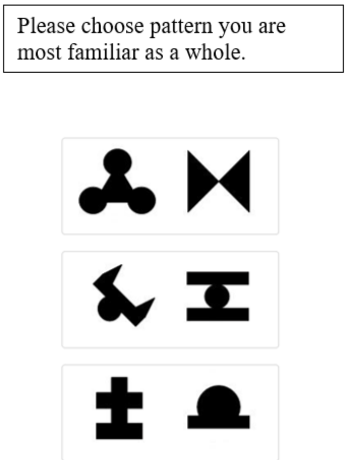


Figure 3.2: . Example of a trial in Pattern recognition task with two distractors.

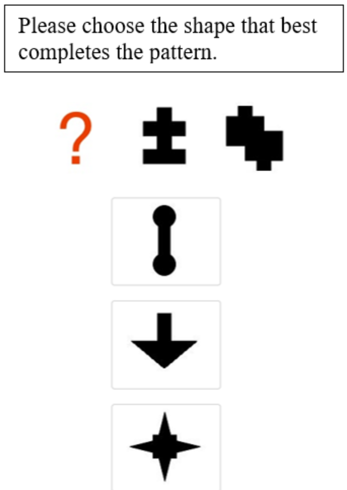


Figure 3.3: Example of a trial in Pattern completion task with two distractors.

Power simulation

A power simulation was conducted to determine the minimum justified sample size. The methodology described in the study by DeBruine and Barr (2021) was employed, which involved running data simulations. In these simulations, we generated 1000 datasets where the effect sizes in the learning task were set to a 6% accuracy advantage for the high-frequency condition compared to the low-frequency condition. Additionally, there was an extra 6% advantage for the suffixed condition compared to the high-frequency condition, resulting in a total 12% accuracy boost for the suffixed condition compared to the low-frequency condition. The correlation between visual statistical learning (VSL) and the suffixed condition, as well as VSL and the high-frequency condition, was set to 0.4. In the absence of substantial literature to leverage, all the values above were informed by considerations on the minimal effect that we deemed of theoretical interest. The noise parameters are outlined in Table 2. Using this model, we achieved 100% power for the learning task and 79.9% power for the correlations with a sample size of 160 participants.

Table 2: Parameters defining noise in the simulation model.

Overall	**Item**	**Subject intercept learning task**	**Subject frequency slope**	**Subject meaning slope**	**Subject intercept VSL task**
0.06	0.03	0.06	0.03	0.03	0.8

3.4 Results

The data were analyzed using R (R Core Team, 2020), version 2021.9.0.351.

Participants' exclusion

As per the preregistration, we excluded participants based on their performance in the typing task during the training phase. We calculated the total Levenshtein edit distance between the target words and the words that participants typed back; participants were excluded from the analysis when this index was higher than 30 (i.e., participants who made more than 30 typos during the learning phase). As a result, we excluded 46 participants, which left 160 participants for the final analyses (male: 86, mean age: 28.81, range: 19–60 years).

Novel word learning task – 2AFC

The overall mean accuracy in the novel word learning task was 0.78 (SD = 0.41). Figure 3.4 breaks down test accuracy by condition, while Figure 3.5 shows mean difference between conditions. As predicted, and as pre preregistered analysis, suffixed words were learned better than high-frequency ending words (0.81 ± 0.12 vs. 0.75 ± 0.12; $t[159] = 6.21$, $p < 0.001$). This contrast unveils the effect of the ending having a meaning, while holding frequency constant. We also expected that high frequency ending items would be learned better than their low-frequency ending counterparts. This prediction was not confirmed (0.75 ± 0.12 vs. 0.80 ± 0.11; $t[159] = -5.21$, $p > .999$). In fact, when we ran a two-tailed t-test on the same comparison, which was not preregistered, the low-frequency ending items turned out to be learned significantly better than the high-frequency ending items ($t = -5.21$, $df = 159$, $p < 0.001$), contrary to our hypothesis.

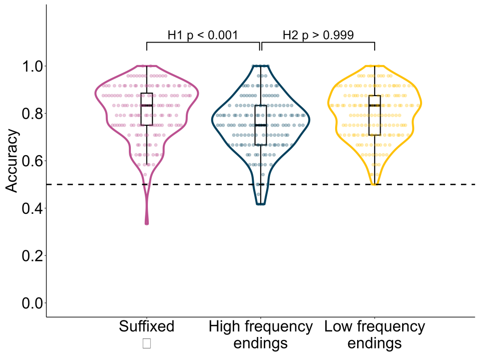


Figure 3.4: Accuracy in Novel word learning task across the complexity conditions.

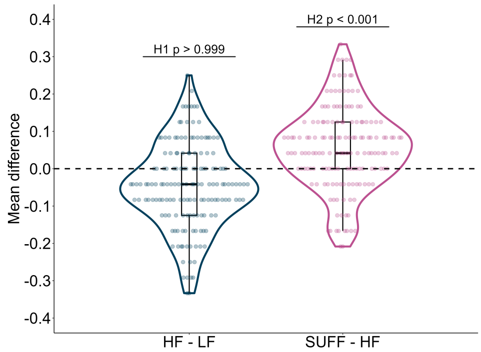


Figure 3.5: Mean difference between high-frequency vs. low-frequency endings, and suffixes vs. high frequency endings.

Visual statistical learning task

The overall mean accuracy in the VSL test was 0.53 (SD = 0.50). The average number of correct responses by subject was 22.32 (SD = 5.54). The distribution of the scores can be seen in (Figure 3.6).

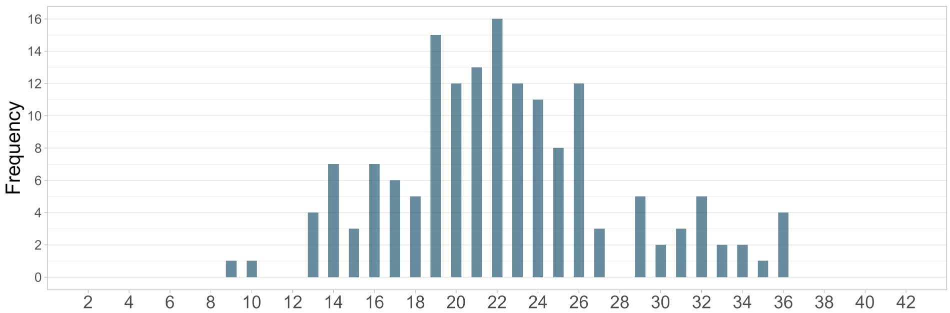


Figure 3.6: Visual statistical learning task. Distribution of scores

Following our preregistration, we correlated the accuracy scores on the VSL task with the accuracy scores in each of the novel word learning conditions. As can be seen in Figure 3.7, the suffixed and low frequency conditions showed a similar correlation with the VSL performance. With suffixed items, the correlation was 0.15, sitting just outside the conventional significance threshold ($t = 1.94$, $df = 158$, $p = 0.06$), while with low-frequency endings, the correlation was 0.18, below the significance threshold ($t = 2.32$, $df = 158$, $p = 0.02$). On the other hand, the correlation with accuracy in the high frequency ending condition was smaller ($r = 0.05$) and not significant ($t = 0.64, df = 158, p = 0.53$).

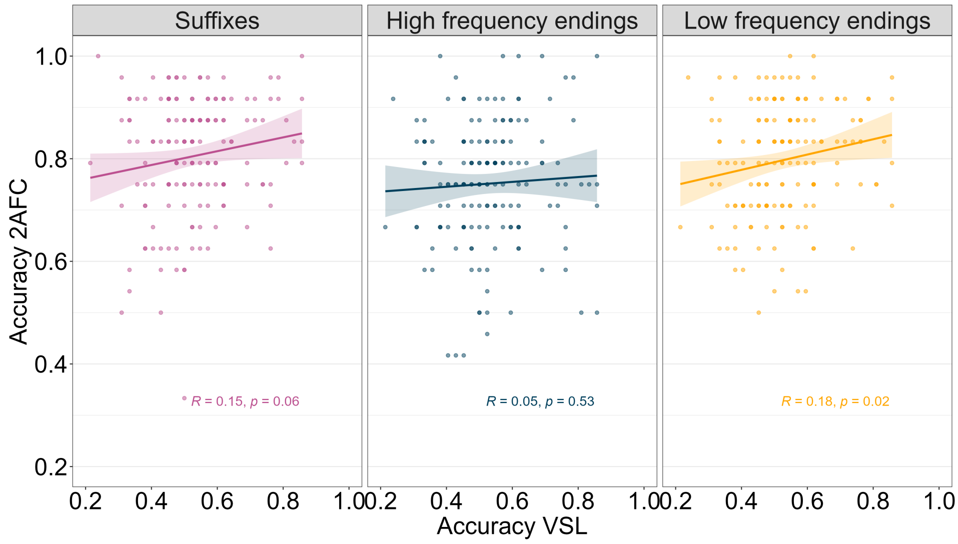


Figure 3.7: Correlation of Novel word learning and Visual statistical learning accuracy across three complexity conditions.

3.5 Discussion

This work sits at the intersection between two lines of research. On the one hand, it is unclear what the specific role of morphology is in the acquisition of novel words, particularly words that feature a known suffix and novel stem. On the other hand, recent studies (Chetail, 2017; Lelonkiewicz et al., 2020, 2023) have suggested a connection between certain morphological effects and statistical learning, specifically the frequency of letter chunks and letter co-occurrence. This leads to the prediction that individuals with good statistical learning abilities might be more capable of learning novel words comprised of frequent letter clusters. In addition to that, they might also be more capable of using morphology in this process, when available, as an additional source of information that is ultimately based on statistical patterns.

To address these questions, we administered a novel word learning task where the items consisted of a nonword stem and a final chunk that could be a suffix (e.g., -er), an equally frequent, but non-meaningful ending in Italian (e.g., -an) or an infrequent ending (e.g., -ow). This design enabled us to specifically single out the effect of the frequency of the final chunk (comparing words with low frequency vs. high frequency endings), and the effect of meaning (comparing high-frequency endings vs. suffixed words). To investigate whether the possible frequency and meaning effects were connected to statistical learning, we tested our participants for their SL skills using the task developed by Siegelman et al. (2017) which was designed to capture individual variability.

Three main results emerged. First, we confirmed our predictions regarding the semantic effect: novel words with suffixes were learned more effectively compared to their counterparts with meaningless, high-frequency endings. Second, a frequency effect emerged, but it contradicted our initial expectations: items with low-frequency endings were better remembered than novel words with high-frequency endings. Third, the correlations between statistical learning skills and the word learning performance in the different conditions were not particularly strong, but lay around the significance threshold in the conditions where the best learning outcomes were observed (i.e., suffixed and low-frequency ending words). For the high-frequency ending condition, the correlation was not significant.

The semantic effect confirms that morphology plays a role in the learning of novel words (e.g., Dawson et al., 2021), and that this suffix advantage relates specifically to the fact that suffixes carry meaning, not to their frequency as letter chunks. These results extend the previous literature highlighting the role of morphology in the processing of novel words (e.g., Burani et al., 2002, 2008; Crepaldi, et al., 2010; Taft & Forster, 1975; Yablonski & Ben-Shachar, 2016; see Amenta & Crepaldi, 2012 for a review). Apparently, not only do we detect the morphological structure of unfamiliar letter strings, but we also actively use that information to store novel words in memory.

There are two ways in which we can account for these results. One explanation is that novel words are learned more effectively due to the dynamics of the lexical system. According to several models of complex word identification (Crepaldi et al., 2010; Grainger & Ziegler, 2011; Marelli & Baroni, 2015; Taft & Nguyen-Hoan, 2010), suffixes have distinct representations within this system. When encountering a novel word, these representations would be activated, thereby facilitating the formation of the novel word's representation itself. Conversely, this is less likely to occur with letter clusters lacking such specific representations. Importantly, this account refers to the same lexical dynamics that are generally believed to underlie the processing of familiar words (Coltheart et al., 2001; McClelland & Rumelhart, 1981). These dynamics are automatic and implicit, that is, they do not require any strategic control, nor do they typically reach the reader's awareness. Another possibility is that familiar suffixes contribute to the improvement of episodic memory for the novel words, resulting in better retention and recall of these words. This might again be related to the structure of the underlying lexical system, of course.

However, in this account, this phenomenon per se would be mostly related to the workings of explicit memory, rather than to the implicit, non-conscious dynamics of the lexical system. In other words, while demonstrating the role of suffixes in novel word learning, the present data open a new question: are we constructing a new implicit representation of the novel words specifically linked to the representation of the suffixes within them? Or rather, are the suffixes merely making us remember better that we have encountered a novel word in the recent past? Of course, the two possibilities are not necessarily mutually exclusive; we might have a better episodic memory for suffixed novel words, but perhaps this is precisely because those novel words activate a suffix representation in the lexical system. Future work might address this issue.

These theoretical considerations relate to the vast literature on the consolidation of novel lexical entries – the process by which newly encountered words are integrated into a person's long-term memory, and stored as representations that interact with previously existing ones (e.g., Davis & Gaskell, 2009; Plaut et al., 1996; Ullman, 2001). The morphological effect observed here unveils the possibility that suffix representations are involved in the process of lexical consolidation, which in turn opens a new set of interesting questions. How does this happen? Does this create new ties in the lexical system that, for example, might be amenable to investigation via completely implicit phenomena like priming (see, e.g., Viviani & Crepaldi, 2022)? And under which conditions would a suffix-mediated consolidation emerge? Does it depend on the previous status of the morphological representations, so that it could bring more benefit to one individual than another (such as in "semantic" vs. "orthographic" profile in Andrews & Lo (2013) study)? Does it require several encounters with the novel word, or a sufficiently varied experience (e.g., Mak et al., 2021; Tamminen et al., 2015)?

One clear finding from our present results is that, if a connection with suffix representations is established during word learning, this process certainly involves semantic information. This conclusion is supported by the fact that non-meaningful letter chunks, despite being equally frequent, did not provide the same learning advantage. Different models of morphological processing have different ways of addressing the contrast between more visual, frequency-based stages of processing on the one hand, and more linguistic, meaning-rich morphological processing (Beyersmann & Grainger, 2023; Crepaldi et al., 2010; Grainger & Beyersmann, 2017; Grainger & Ziegler, 2011; Taft & Nguyen-Hoan, 2010). However, all these models agree that both types of representations/processing exist. Our findings suggest that the learning advantage provided by suffixes during the acquisition of novel words is due to the predominant involvement of more linguistic and meaning-oriented morphological representations. This aligns with previous research that suggests that establishing links between form and meaning in new lexical items is key to learning such items (Dawson et al., 2021). Dawson et al. (2021) found an advantage of suffixes only in a semantic recall task, which relies heavily on semantics. However, they failed to observe similar findings in less semantics-dependent tasks (phonological shadowing, lexical decision, and spelling). Interestingly, our results align with the findings from Dawson et al. (2021), even though our task did not particularly emphasize semantic processing.

The second important finding of this experiment is that novel words with low-frequency endings were learned significantly better than their high-frequency counterparts. This "inverse" frequency effect was perhaps somewhat surprising. Frequency effects are very well established in the processing of familiar words (e.g. Longtin et al., 2003; New et al., 2004; Rastle et al., 2004), but from our results, it appears that frequency of occurrence yields opposite effects in novel word learning vs. familiar word identification. Thinking more closely, however, this is not so surprising after all: novel words with low-frequency patterns are more striking, and therefore, they might perhaps be more memorable. This sits nicely with the literature on word and sentence processing that revolves around concepts such as surprisal (e.g., Amenta et al., 2023; Frank & Bod, 2011; Hale, 2001; Levy, 2008). Amenta et al. (2023) showed that there is a tradeoff between sentence-level and word-level cues, and that reader behavior is effectively described based on the amount of information carried by a given word in the context of a given sentence. Somewhat related, error-driven learning postulates that larger mismatches between expected and observed states generate stronger memories (e.g., Henson & Gagnepain, 2010). From this perspective, it is conceivable that low-frequency endings determined a violation of the general word-ending pattern in the language, which might have triggered better learning.

Overall, it seems that we then have two separate factors at play here. On the one hand, suffixed items benefited from the inclusion of a meaning-bearing unit, which is likely to have an established, dedicated representation within the lexical system. On the other hand, novel words with unusual endings might have broken the readers' expectations about word forms, thus triggering a surprisal effect that led to better learning. Novel words with high-frequency endings would not benefit from either the meaning or the surprisal effect, and therefore yielded the worst performance overall.

The individual variability results support this interpretation. Statistical learning skills exhibited stronger correlation with the learning of suffixed and low-frequency-ending items, as compared to novel words with high-frequency endings (these correlations were not very strong overall, however; more on this below). This suggests a link between the reader's ability to capture regularities in the visual input and morphology on the one hand, and surprisal on the other. It is not difficult to see why statistical learning might be specifically related to morphology. As a symbolic system, the human language is characterized by a fundamental arbitrariness in the way form conveys meaning; there is generally nothing in the way a word sounds, or looks, that is suggestive of its meaning (e.g., Hockett, 1963). Morphology, however, establishes some regularity in this mapping and thus introduces predictability. Although the association between, for example, the suffix -er and its meaning remains arbitrary, the fact that, for example, driver, settler, buyer, dealer, all denote someone who performs a specific action increases the predictability of lexical mappings. Therefore, when we encounter a novel word such as *libber*, while its exact meaning might be unknown, we can confidently assume that a *libber* is someone who *libs*. Substantial evidence supports our sensitivity to these regularities (e.g., Marelli et al., 2015). Moreover, recent studies (Lelonkiewicz et al., 2020, 2023) have provided direct causal evidence that morphology-like statistics can induce morphology-like effects, even when the stimuli are entirely unfamiliar to the readers and lack any reference to semantics and/or phonology.

The connection between statistical learning and the surprisal effect is even more intuitive; there can be no surprise without a good representation of how words "normally" look, that is, of the statistics that characterize the functioning of the lexical system. Interestingly, this correlation highlights that statistical learning does not only capture the correspondence between form and meaning, but also the statistics of the orthography per se. In fact, the only difference between high-frequency and low-frequency ending items was frequency itself – neither type of word ending is connected to any specific piece of meaning in the language. Therefore, the frequency of letter chunks seems to be another relevant piece of information that statistical learning captures and that affects the acquisition of novel lexical material. This observation is again in line with previous work investigating the role of letter statistics in visual word identification and processing (Chetail, 2017; Lelonkiewicz et al., 2020; Seidenberg, 1987; Treiman et al., 2017). Moreover, it directs our attention to a critical question in the statistical learning literature: since the information that is used is different in letter statistics vs. meaning, is the cognitive phenomenon that underlies the two correlations one and the same? Or rather, are some neural circuits/cognitive processes at work at the interface between form and meaning, while other, separate neural circuits/cognitive processes capture letter statistics? Our data certainly cannot settle this question, although the fact that the same individual SL skill correlates with both the performance on suffixed words and the performance on low-frequency endings might be taken to suggest that the same fundamental statistical learning process may support word learning in both conditions (for a broader discussion of this issue, see Bogaerts et al. (2022).

Finally, it is important to stress that although the relevant correlations above are removed enough from zero to be statistically significant – and, we believe, of theoretical interest – they are still quite small. One reason for this could be the different training methods in the statistical learning and word learning tasks. While in the latter participants were exposed to static stimuli without any time constraints, in the visual statistical learning training, they encountered fast-paced, dynamic stimuli. Furthermore, during word learning, participants were actively engaged (i.e., they had to type back the words they were learning), whereas the SL task simply required passive viewing. These differences are certainly non-negligible; however, they make the correlations that we observe here even more compelling and provide further support to the idea that statistical learning contributes to word learning, especially for what concerns form and form-to-meaning regularities. At the same time, the fact that these correlations are not big in size also suggests that the role of statistical learning in these mechanisms might be fairly limited, and effective word learning is likely to depend on a host of cognitive processes (e.g., attention, working memory, vocabulary size; Gathercole & Baddeley, 1993; I. Nation, 2006; K. Nation, 2008; Ouellette, 2006; Smith et al., 2010) where statistical learning is only one of many factors.

Chapter 4

Experiment 3: Catching a CAPTCHA: The impact of variable input on the processing of emerging orthographic representations

Manuscript submitted to the *British Journal of Psychology*. All the materials related to this paper are publicly available at the Open Science Framework. Data and analysis code are available at https://osf.io/85dmp/. Training materials are available at https://osf.io/um6rw/. The study is preregistered at https://aspredicted.org/PLZ_FNB.

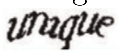
4.1 Abstract

Variability inherent to handwriting has been suggested to help establish more robust letter representations than other methods (e.g., typing or tracing). The present study tests whether encoding letter strings from a novel alphabet becomes more resistant to distortion when trained with variable input. Over five days, participants learned an 11-character artificial alphabet in a variable handwritten format involving reading, listening, and handwriting practice. Another set of 11 artificial characters served as a visual control. Before and after the training, participants completed a masked priming same-different task with the novel alphabet letters. The key manipulation was in the primes: the identity/unrelated primes could be presented in a printed or

distorted format. Results showed identity priming in both conditions, with a stronger effect for the printed primes. These effects increased post-training for experimental and control scripts, indicating that exposure to variable input enhances distortion resistance even without explicit training. Another experiment assessed the transposed-letter effect—another marker of orthographic processing—in the novel script with a same-different task, revealing an increase in the post-training phase for both scripts. These findings demonstrate improved identification of novel letter forms through training, while the development of abstract orthographic representations is still in progress.

4.2 Introduction

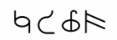
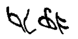
Reading and writing are revolutionary inventions of human civilization and are essential communication tools in modern society. During reading, the eyes typically focus on each word in the text, often fixating only once, thereby providing a brief foveal glimpse. In this fleeting moment, skilled readers process the word, transforming the visual input into an increasingly abstract orthographic code. This code is crucial for retrieving the word's phonological, morphological, and semantic properties from the mental lexicon. The present paper focuses on the emergence of orthographic processing, the critical interface between visual input and the mental lexicon. Orthographic processing serves as a vital bridge, linking the initial stages of visual perception to the more complex processing of words. It encompasses the encoding of abstract letter identities and the serial order of the letters, playing a crucial role in guiding the selection of the appropriate entries in the mental lexicon. This process is key in distinguishing orthographically similar words such as `KISS` from `HISS` or `GOD` from `DOG` (see Grainger, 2018, for review).

The process of encoding letter identities in the brain, as agreed upon broadly in the research community, involves specialized neuron layers attuned to abstract letter representations during visual word identification. A notable example is the hierarchical model of visual word recognition proposed by Dehaene et al. (2005), where specific neuron layers exhibit similar responses to different forms of the same letter (d, *d*, D, or D). As reviewed by Grainger (2018), developing this level of abstraction is necessary for proficient reading, and these abstract orthographic representations are both stable and resilient to visual noise—critical factors for effective reading. Notably, this ability is thought to develop relatively early after learning to read (Jackson & Coltheart, 2001). Empirical evidence supporting this view comes from masked priming experiments showing sizeable repetition priming effects with handwritten words (Gil-López et al., 2011; Qiao et al., 2010). An even more striking example of the brain's proficiency in handling distorted visual input occurs with CAPTCHAs (Completely Automated Public Turing test to tell Computers and Humans Apart; Ahn et al., 2003). In a masked priming lexical decision task, Hannagan et al. (2012) found significant repetition priming effects with printed target words when primes were distorted in a CAPTCHA-like manner (e.g., ). Although the repetition priming effects were less pronounced than with printed primes, they were still sizeable, suggesting that the letter detectors, weathered by exposure to a wide variety of visual inputs, are quite adaptable to distortion. This adaptability allows them to respond to a broad spectrum of potential visual inputs for any given letter. In our study, we utilize the identity priming effect of CAPTCHA primes to investigate the tolerance of letter detectors to noise in the processing of letter strings for a novel—recently learned—script.

The main aim of this paper was to examine whether stable orthographic representations, resilient to distortion, can be developed in the initial stages of literacy. To explore this, we trained adult participants in a novel, artificial script over five sessions, using varied visual inputs. We opted for adults learning an artificial script instead of preliterate children to have better control over participants' prior letter knowledge and use more standardized tasks—note that in experiments with preliterate children, the tasks and procedures must be simplified enormously (Perea, Jiménez, et al., 2016). This approach has proven effective, akin to how children learn to read (Taylor et al., 2011; C. Vidal et al., 2017).

Employing a similar methodology, Fernández-López et al. (2020) examined the emergence of orthographic representations focusing on the encoding of letter order. In their study, participants were trained in two unfamiliar scripts, each comprising 11 BACS characters (C. Vidal et al., 2017), throughout six sessions, with each letter being assigned a phonological value. The training for one script encompassed extensive handwriting, listening, and reading exercises, while the other script served as a control, concentrating solely on superficial letter recognition. The participants were tested on two aspects of orthographic processing before and after the training. Firstly, a same-different task was administered to assess letter transposition effects, thereby evaluating the emergence of flexibility in letter position within the new script, a phenomenon known as location-invariance processing. This concept posits that letter strings evoke greater transposition effects than strings of symbols or artificial letters (Duñabeitia et al., 2012; Massol et al., 2013). Secondly, a target-in-string task examined the parallel processing of letter positions in a string. Prior research has shown a divergence in the accuracy functions of symbols versus letters in this task: Participants tend to be more accurate when identifying centrally fixated characters in symbols, while an advantage is also observed for the exterior letters, especially the initial letter, in letter strings (Tydgat & Grainger, 2009). Fernández-López et al. (2020) observed that the findings in both tasks were strikingly similar before and after training for both the script learned by participants and the one with which they were merely visually familiarized. They suggested that more robust orthographic representations are required to distinguish orthographic processing from the processing of other visual symbols—particularly regarding location invariance and position-specific processing. Nonetheless, their study primarily examined the emergence of location-invariance and location-specific processing in letter strings of the newly trained script without directly examining the development of the encoding of letter representations. Our study seeks to address this gap in the literature.

One effective strategy to develop more stable letter representations involves increasing the variability of the visual input. This approach was examined in a study by Li and James (2016), which focused on five-year-old children learning four Greek letters previously unfamiliar to them. The children were trained using either variable or invariable input coupled with either visual-motor or visual-auditory training methods. Li and James found that variable input, irrespective of the training type, enhanced the children's ability to categorize a letter correctly in the subsequent testing phase. They argued that exposure to only a singular form of a letter (e.g., a) would make it challenging for a learner to recognize that (a and a) belong to the same abstract letter unit. However, through repeated exposure to various letter forms within the same context, learners can develop a more robust sense of letter invariance.

In the present paper, we examined whether the development of orthographic representations could be accelerated by introducing variability in the visual input during the reading learning process. Specifically, we primarily focused on how such variability affects the encoding of letter identities. Our methodology was based on the training protocol used by Fernández-López et al. (2020) but with two significant modifications. Firstly, to increase the variability of the training input, we presented the learning materials in four different handwritten fonts, as opposed to the printed BACS2serif font used previously (refer to Table 1 for details). Secondly, to assess whether the newly formed orthographic representations could withstand distortion, we conducted the masked priming same-different task introduced by Norris and Kinoshita (2008; Kinoshita & Norris, 2009). In this task, primes were either identical or unrelated to the target and were presented in either a regular, printed format (e.g., ) or a distorted, CAPTCHA-like format (e.g., ). Notably, the repetition priming effects potentially observed in this task—even for familiar alphabetic stimuli—are considered to be prelexical, indicating that any observed effects would primarily reflect bottom-up activation from the visual input to the letter detectors (Kinoshita et al., 2018; Perea, Marcet, et al., 2016).

If increasing the variability in training materials indeed bolsters the emergence of a greater tolerance to noise in the evolving detectors for letter identities, we predict an increased masked repetition priming effects post-training, particularly for the printed format, but, crucially, also for the CAPTCHA-like primes, only for the alphabet that participants learned to read. Conversely, the lack of differences in repetition priming effects between the trained alphabet and the visual control would imply that the obtained priming effects are not uniquely orthographic but rather stem from greater visual familiarity with the script. Additionally, as a secondary objective, we explored the encoding of letter order in this new setup using the same task employed by Fernández-López et al. (2020)—the specific details will be discussed in the context of Experiment 2.

4.3 Experiment 1. The emergence of abstract letter representations

4.3.1 Method

The analysis, exclusion criteria, and sample size justification were preregistered at https://aspredicted.org/PLZ_FNB. The study was approved by the local Ethical Committee.

4.3.2 Participants

Participants were 28 native Spanish speakers (mean age = 20.69 years, SD = 1.75) with normal or corrected-to-normal vision and reported no language-related or learning disorders. All participants gave informed consent and were given monetary compensation upon completing the experiment.

4.3.3 Materials and Design

In the experiment, we used two novel scripts from previous research (Fernández-López et al., 2020), available at https://osf.io/um6rw/. Each script is a subgroup of the BACS alphabet (C. Vidal et al., 2017), composed of 9 consonants and two vowels. Each participant learned one script via print-to-sound training to establish grapheme-phoneme associations (i.e., experimental script). In contrast, learning the other script referred to the visual familiarization with the characters (i.e., control script). The use of the script across participants was counterbalanced. Four different hand-written fonts were created using the Calligraphr online app (www.calligraphr.com) to create the variability in the input. The complete scripts and their handwritten versions can be seen in Figure 4.1.

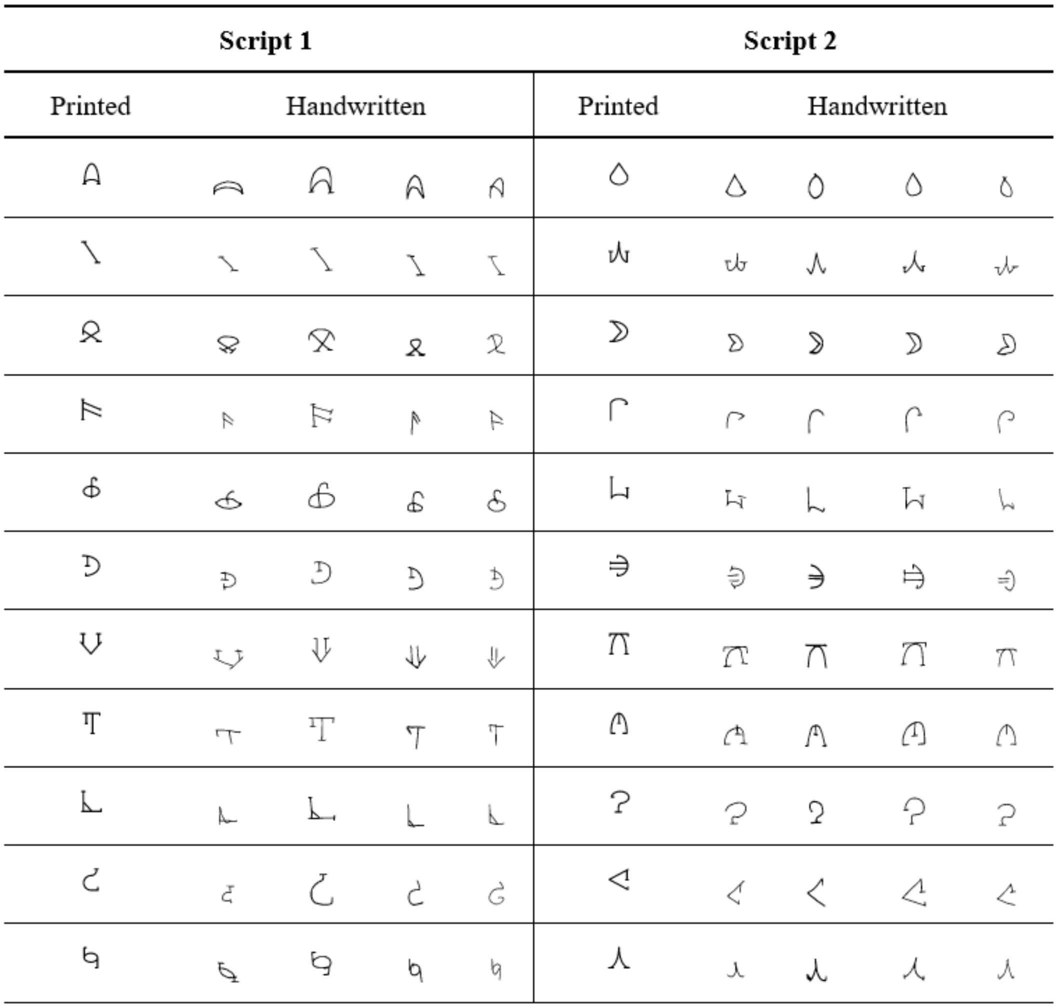


Figure 4.1: Letters from Scripts 1 and 2 in printed BACS2serif and handwritten versions.

4.3.4 Training

Learning to read

As mentioned at the beginning of this subsection, the learning materials were identical to those of Fernández-López et al. (2020) with one crucial difference: instead of the capital letters of BACS2serif font (C. Vidal et al., 2017), they were presented in a handwritten font. Within each character string, only one font was used. The use of different fonts was pseudorandomized across items.

The training in the experimental script was done as follows. On the first day, participants familiarized themselves with the grapheme-phoneme correspondences of the experimental script. They were presented with graphemes of the novel script and their corresponding phonemes. They were asked to read, listen, and hand-copy them on a piece of paper until they felt confident in remembering the associations. On day 2, they briefly reviewed the associations and completed three read-aloud tasks involving 12 sequences of 4 and 5 characters. They also completed three write-down tasks, which consisted of listening and writing down another 4 and 5-character long sequences. On day 3, they completed the same tasks with six or 7-character long sequences, and on day 4, they completed the same set of tasks with 8-character long sequences and then with 6, 7, and 8-character long sequences.

Visual familiarization

To familiarize the participants with the control script's visual form, they were presented with the list of all control script characters on the first day. They were asked to try to remember them. To practice the control script, we administered a character count and detection tasks on days 2, 3, and 4.

Character count task. In the character count task, a fixation point appeared on the screen for 500 ms, substituted by a character string for a maximum of 2000 ms or until response. Participants were asked to press "yes" only if the character string that appeared on the screen contained three or more nonalphabetical symbols.

Character detection task. In the character detection task, a character from the untrained script, which acted as a probe, appeared on the screen for 1000 ms, followed by a pattern mask (####) for 500ms. The mask was then substituted by a target, a string of characters from the untrained script that had an equal length as the mask. It stayed on the screen until response or until a timeout of 2000 ms. The participants were instructed to respond "yes" if the probe appeared in the target or "no" if it did not. The length of the character string in both tasks corresponded to the length of the string in the training on that day (i.e., if the training was on four and five-character strings, so were the visual familiarization tasks). In 27 out of 36 trials, the BACS characters were presented in a hand-written font. The use of hand-written fonts was pseudorandomized across trials.

Testing task – masked priming same-different task

Materials. The probes, primes, and targets were 4-consonant strings composed of characters belonging to one of the scripts, thus creating two versions of the task, one for each script. Characters were never repeated within a single string. Three hundred twenty probe-target pairs were created, 160 belonging to the “same” condition (probe and target were identical) and 160 to the “different” condition (probe and target consisted of entirely different characters). Half of the primes were distorted similarly to CAPTCHAs (Completely Automated Public Turing test to tell Computers and Humans Apart; (Ahn et al., 2003)) while the other half was presented in a regular printed format. CAPTCHA items were generated using Python script (Python version 3.6.6; packages: pandas (version 1.1.5.), PIL (version 8.0.1.)). This yielded a 2 x 2 x 2 design (same-different strings x related-unrelated probe x printed-distorted prime). Four lists, each comprising 320 trials, were created following the Latin square design. In addition, a practice list with the same criteria, containing 24 trials, was generated.

Procedure. The task was programmed using PsychoPy3 Builder version 2020.2.10 (Peirce et al., 2019). Like the standard masked same-different paradigm, the trial sequence began with a 500ms presentation of a fixation cross at the center of the screen. It was followed by a 1000ms probe, after which a mask consisting of four hashtags appeared lower on the screen for 1000 ms. Subsequently, the prime was presented in the location of the mask for 50 ms, which was then replaced by the target until a button press or a 2000 ms timeout (see Figure 4.2)—to prevent perceptual continuity, the prime was in a smaller size than the target. Participants were instructed to respond as accurately and as quickly as possible whether the two-character strings were the same or not. The session lasted approximately 18 min.

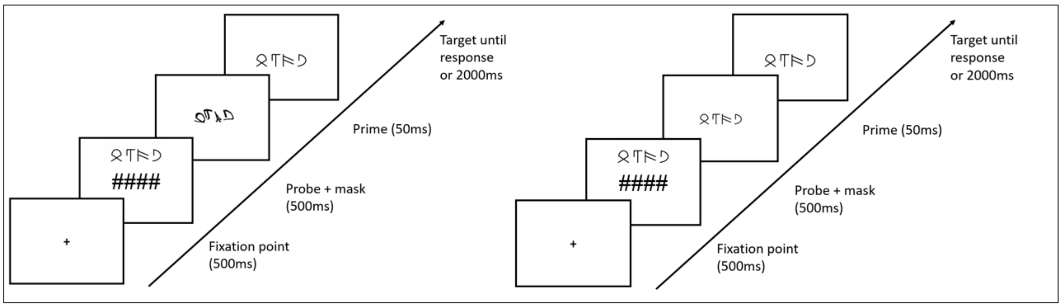


Figure 4.2: Illustration of the masked priming same-different task with the distorted (CAPTCHA) prime on the left and the printed prime on the right for “same” trials.

4.3.5 Overall procedure

The experiment took place in a quiet laboratory setting over five consecutive workdays. It consisted of the pre-training, training, and post-training phases. On the first day, participants completed the pre-training phase: They were administered masked priming same-different task and the regular same-different task in both script one and script 2. The order of the tasks was counterbalanced. Then, they were familiarized with the experimental and control scripts. Half of the participants learned script one, and half learned script two. On days 2, 3, and 4, they were trained on letter strings of increasing length (4 to 8 characters) for the experimental script, and they completed the visual familiarization tasks for the control script (detailed description below). On the last day, they were first administered a test consisting of one read-aloud exercise and one listening exercise. After passing the test with a minimum of 84% correct responses (20 correct responses out of 24 in reading and writing), they completed the post-training phase, which consisted of masked priming same-different task and the regular same-different task in the script they were trained on. For a graphic depiction of the training, see Figure 4.3. For more details on the procedure, see (Fernández-López et al., 2020).

Figure 4.3: Scheme of the experiment procedure over five days.

4.3.6 Data analysis

Table 3 shows average response times and accuracy. In the preregistered statistical analysis, the critical dependent variable was response time. All response times shorter than 250ms and incorrect responses were removed from the analyses (8.53% of data points were removed). The analysis focused on the "same" trials because that is where the priming effect can be observed. We also conducted a parallel analysis of the accuracy data – this analysis was not preregistered (see Appendix). All data and data analysis scripts are available at: https://osf.io/um6rw/.

		Pre-training		Post-training	
		Trained	Untrained	Trained	Untrained
Related	Captcha	613 (7.7%)	625 (7.7%)	538 (6.6%)	547 (8%)
	Printed	570 (4.3%)	581 (4.4%)	500 (3.9%)	500 (5%)
Unrelated	Captcha	638 (9.8%)	650 (11%)	576 (11.3%)	575 (11.1%)
	Printed	607 (8.5%)	628 (6.4%)	552 (9.7%)	552 (11%)
Priming effect	Captcha	25 (2.1%)	25 (3.3%)	38 (4.7%)	28 (3.1%)
	Printed	37 (4.2%)	47 (2%)	52 (5.8%)	52 (6%)

Table 3: Masked priming same-different task: mean correct reaction times (in milliseconds) and accuracy (in parenthesis) across conditions.

We ran Bayesian linear mixed models to analyze the data using the brms package (Bürkner, 2017, 2018) in R (R CoreTeam, 2023). Phase, script, prime relatedness, and prime distortion, and their 4-way interaction were contrast-coded as fixed effects – these effects were zero-centered: *related* vs. *unrelated* [-0.5 and as 0.5], *pre-training* vs. *post-training* [-0.5 and as 0.5], *trained* vs. *untrained* [-0.5 and as 0.5], and *captcha* set vs. *printed* set [-0.5 and as 0.5]. We used the maximal random structure both for participants and items[1]. We used a shifted log-normal distribution for the reaction time data. The priors for the RT data were weakly informative: Normal ($\mu = 0$, $\sigma = 10$) for the intercept and Normal (0, 1) for each of the fixed effects/interactions and standard deviation parameters[2]. For the covariance matrix of random effects, we

[1] Brms_captcha_rt_model <- brm(data = captcha_data_rt, rt~pre_post_c * trained_c * prime_relatedness_c * prime_distortion_c + (1 + prime_distortion_c * pre_post_c * trained_c * prime_relatedness_c| participant) + (1 + prime_distortion_c * pre_post_c *prime_relatedness_c | item), warmup = 1000, iter = 5000, chains = 4, family=shifted_lognormal(), sample_prior = T, prior = priors, save_all_pars = T, control = list(adapt_delta = 0.95), cores = 4)

[2] As a further check that the present findings were not affected by the choice of priors done in the preregistration, we also conducted the analyses using the default prior from

had a regularization of 2. The model was fitted using four chains with 5,000 iterations (1,000 as warmup). We consider an effect credible if the 95% credible interval (CrI) estimated from the posterior distribution does not contain zero. Simple tests effects in case of evidence for interactions were made using the *emmeans* package (Lenth, 2021).

4.3.7 Results

The results of the reaction time data showed evidence of a main effect of phase ($b = -0.22$, Estimation Error = 0.05, 95% CrI [-0.31, -0.13]) where response times were faster after training (542 ms) than before (614 ms). We also found evidence of an effect of prime-target relatedness ($b = 0.13$, Estimation Error = 0.01, 95% CrI [0.10, 0.15]) with faster responses for related targets (559 ms) compared to unrelated targets (597 ms), and an effect of prime distortion ($b = -0.12$, Estimation Error = 0.01, 95% CrI [-0.14, -0.09]) with advantage for targets preceded by printed primes (561 ms) than distorted primes (595 ms). Prime relatedness interacted with phase ($b = 0.04$, Estimation Error = 0.01, 95% CrI [0.02, 0.07]). Unpacking this interaction showed a larger masked repetition priming effect in the post-training phase (42 ms; $b = -0.15$, 95% CrI [-0.17, -0.12]) than in pre-training (33 ms; $b = -0.10$, CrI [-0.13, -0.08]). Prime relatedness also interacted with prime distortion ($b = 0.08$, SD = 0.02, 95% CrI [0.04, 0.11]). This interaction revealed greater repetition priming effects for targets preceded by printed (47 ms; $b = -0.17$, 95% CrI [-0.19, -0.13]) than distorted primes (29 ms; $b = -0.09$, 95% CrI [-0.11, -0.06]). Figure 4.4 depicts posterior distributions from the estimates in the model. As shown in Figure 4.4, we found no evidence of the effects related to the script (either trained vs. untrained) or its interactions with the other factors.

brms package. They produced virtually the same estimates.

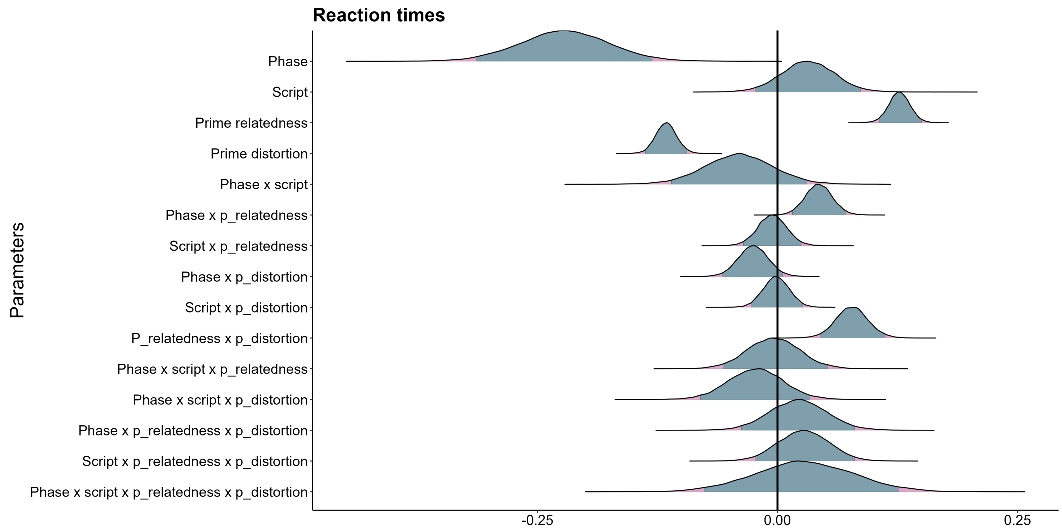


Figure 4.4: 95% and 100% Highest density intervals from the Bayesian linear mixed effects model for the reaction times in the masked priming same-different task.

Thus, we found a stronger repetition priming effect after training than before training. Critically, this increase in the priming effect cannot be attributed to orthographic processing, as a similar pattern was observed for both the script that participants learned to read and the script with which they were merely visually familiarized. In addition, we found a stronger repetition priming for the intact primes than for the distorted primes; again, this effect was similar for the trained and untrained scripts.

4.4 Experiment 2. The emergence of location-invariant processing with variable visual input

A secondary goal of the present paper was to test whether variability in input could enhance the flexibility of letter position encoding in orthographic representations. The underlying rationale was that increased variability in letter forms might influence how letter order is encoded in the novel-trained script. To this end, we implemented a same-different task parallel to the one used by Fernández-López et al. (2020, Experiment 1). As discussed in the Introduction, they found a consistent pattern of transposed-letter effects in the new script, both before and after training. This pattern was parallel for trained and visual control scripts. Our aim was to reassess these results under conditions where the visual input included variability in letter forms to understand the impact of these variations on the transposed-letter effect. Should variability in visual input facilitate the emergence of orthographic processing, we anticipated an increase in the transposed-letter effect in the post-training test compared to the pre-training test. However, this increase was expected only for participants trained to read the script. Conversely, if this added variability does not impact the development of orthographic processing regarding letter-position encoding, we would expect a similar pattern of transposed-letter effects for both the trained and untrained scripts similar to that reported by Fernández-López et al. (2020).

4.4.1 Method

Participants, overall training procedure, and materials for the training were identical to Experiment 1. The testing task differed, as described below.

Testing task – same-different task

Materials. A separate set of items was created for each of the two scripts. Each set consisted of 240 probe-target 5-character consonant string pairs, displayed in 15pt BACS2serif font (C. Vidal et al., 2017). All character strings were composed of non-repeated letters. One hundred twenty items belonged to the "same" condition, and another 120 belonged to the "different" condition. In the "different" condition, 60 pairs were created by transposing two adjacent letters (e.g., 1-2-3-4-5 →1-3-2-4-5; ᐊ⋺ᒷᕱᒥ →ᐊᒷ⋺ᕱᒥ), and 60 pairs were created by replacing two adjacent letters (e.g., 1-2-3-4-5 →1-6-7-4-5; ᐊ⋺ᒷᕱᒥ →ᐊ♢ʔᕱᒥ). The proportion of transpositions/replacements was the same in all letter locations. To counterbalance the probe-target pairs, we created two lists for each script. For the practice phase, we created eight five-consonant string pairs for each script with the same criteria. Participants were instructed to press "yes" if the two strings were the same or "no" if they did not. They were encouraged to be as quick and as accurate as possible. The session lasted approximately 18 min. The task was programmed using DmDX software (Forster & Forster, 2003).

Procedure. The task followed the same design as in Fernández-López et al. (2020). In each trial, a fixation point appeared in the center of the screen for 500ms. Then, it was substituted by a probe positioned 3mm above the center of the screen for 300ms. Next, the target appeared 3mm below the center of the screen and remained at display until response or until a timeout of 2000ms.

4.4.2 Data analysis

Table 4 shows average accuracy and response times only for the correct items. Following the preregistered statistical analysis, the only dependent variable was accuracy. We also analyzed the reaction times in a non-preregistered analysis (see Appendix). We analyzed only the “different” trials because that is where the critical manipulation lies (i.e., transposed-letter vs. replacement-letter pairs). All data and analysis scripts are available at: https://osf.io/85dmp.

		Pre-training		Post-training	
		Trained	Untrained	Trained	Untrained
Different	Transposed	659.96 (45%)	704.10 (46%)	620.77 (41%)	613.62 (43%)
	Replaced	630.08 (27%)	675.36 (27%)	569.62 (24%)	570.24 (24%)
Same		584.44 (8%)	617.93 (10%)	539.74 (8%)	530.25 (9%)

Table 4: Same-different task: mean correct reaction times (in milliseconds) and accuracy (in parenthesis) across conditions.

We analyzed the accuracy data using Bayesian generalized mixed models. The fixed effects were a phase (pre- vs. post-training), training (trained vs. untrained [visually familiarized]), and probe-target relationship (transposed vs. replaced). We used the maximal random factor structure for participants and items. They were contrast-coded as zero-centered fixed effects: *pre-training* vs. *post-training* [-0.5 and as 0.5], *trained* vs. *untrained* [-0.5 and as 0.5], and *transposed* vs. *replaced* set [-0.5 and as 0.5]. For the fits, and due to the binary nature of the responses (1 denoting a correct response and 0 an incorrect response), we used the Bernoulli distribution with a logit link. The priors and model fitting were identical to the masked priming same-different task. Again, we considered an effect as credible where the 95% credible interval (CrI) estimated from the posterior distribution did not contain zero. The *emmeans* package (Lenth, 2021) was used to unpack significant interactions.

4.4.3 Results

The results of the accuracy data showed evidence of an effect of probe-target relationship (b = 1.00, 95% CrI [0.84, 1.16]) where the error rate in the replaced condition was lower compared to the transposed condition (43.75%

vs. 25.5%). We did not find any signs of the other effects or interactions (Figure 4.5 depicts the posterior effects estimates from the model).

While not preregistered, the analysis of the response times (RTs) replicated the transposed-letter effect. Moreover, phase also affected the reaction times, with faster RTs after the training, and the interaction of the probe-target relationship and phase showed a stronger effect after the training (see Appendix for details).

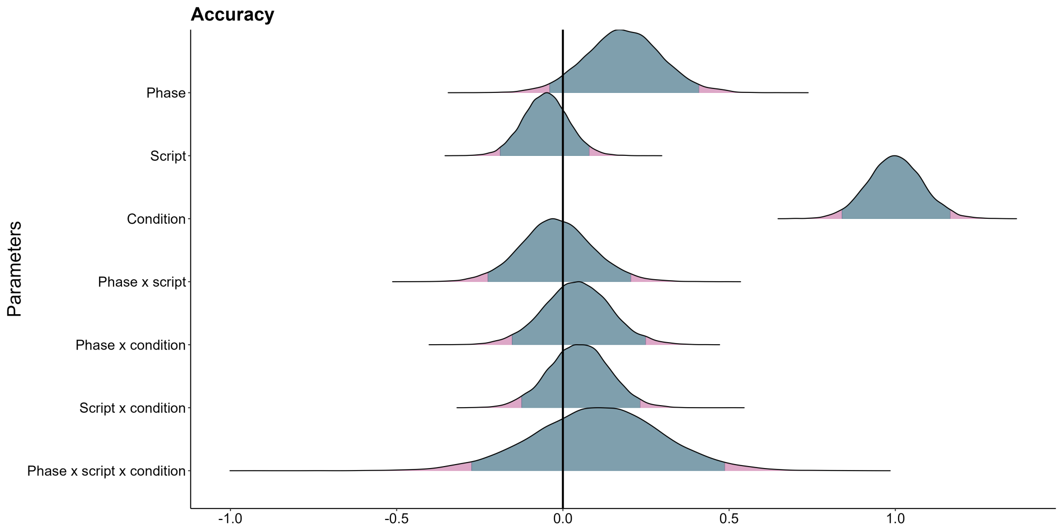


Figure 4.5: 95% and 100% highest density intervals from the Bayesian generalized mixed effects model for accuracy in same-different task.

Thus, we found the typical transposed-letter effect: participants' responses were more accurate for replacement-letter pairs than transposed-letter pairs. Critically, as also occurred in the Fernández-López et al. (2020) experiment, the magnitude of this effect was similar before and after the training. Also, it was similar for both the trained script and the visual control script. In other words, adding variability to the letter shapes in the training phase did not modulate how participants encoded the position of the characters in the letter strings.

4.5 General Discussion

In the present experiments, we examined whether the variability in visual input influences the development of two fundamental components of orthographic processing: the encoding of letter identity and letter position. The underlying premise was that exposure to varied visual inputs would facilitate the formation of more robust letter representations (e.g., see Li & James, 2016). To test this hypothesis, we trained participants in an artificial script, using handwritten versions, until they attained proficiency in reading, listening, and handwriting. Additionally, we introduced a second set of novel letters as a visual control to better assess the specific impact of reading and handwriting training. For the control script, participants were only familiarized with their visual forms without the associated reading and handwriting training. In Experiment 1, participants completed a masked priming same-different task both before and after training. This task involved identity versus unrelated priming conditions, employing both printed and distorted (CAPTCHA-like) primes. The aim was to assess the tolerance of the newly learned letters against visual distortion. Specifically, the emergence of a greater repetition priming effect with CAPTCHA-like items after training in the novel script would suggest that these representations are resilient to distortion. In Experiment 2, we shifted our focus to determine whether the novel characters had been internalized as abstract orthographic representations by examining whether training induced a more flexible encoding of letter order—an index of orthographic processing (Grainger, 2018). To assess this, we employed a same-different task comparing transposed-letter versus replacement-letter pairs. Here, an increased transposed-letter effect observed post-training would indicate effective orthographic processing. This is supported by previous findings, which suggest that letter strings typically exhibit stronger transposition effects than strings of non-letter symbols (Duñabeitia et al., 2012; Massol et al., 2013).

The results of the masked priming same-different task (Experiment 1) revealed several key findings. First, we found that even before any training, distorted primes (such as CAPTCHA) were capable of producing repetition priming effects. This suggests that the cognitive system of adult readers can handle a certain degree of variability in the letters of a novel script, even when they are still entirely unfamiliar. This finding aligns with and extends previous research indicating that visually variable primes, including handwritten and CAPTCHA primes, aid in word processing (Gil-López et al., 2011; Hannagan et al., 2012; Qiao et al., 2010). However, it is important to note that these earlier studies employed tasks like lexical decision (Gil-López et al., 2011; Hannagan et al., 2012) and semantic categorization (Qiao et al., 2010), which might involve top-down processes aiding in the integration of primes and targets (see Vergara-Martínez et al., 2015). In contrast, the masked priming same-different task primarily targets prelexical processing, minimizing the influence of top-down information (Norris & Kinoshita, 2008; Perea, Marcet, et al., 2016). Therefore, our results demonstrate that distorted primes can facilitate the processing of letter strings without lexical feedback.

Secondly, we found that the repetition priming effect was greater in the post-training phase, indicating that participants' familiarity with the novel letters increased over the training sessions, yielding a sizable processing advantage. Notably, this boost in the priming effect was observed both for the experimental and visual control scripts. Thus, the enlarged repetition priming effect in the post-training phase should be attributed more to heightened visual familiarity with the input rather than the development of orthographic representations specific to the script participants learned to read.

Therefore, Experiment 1 uniquely demonstrated that distorted CAPTCHA primes can be effectively normalized even in the absence of top-down influences. This finding suggests that after training, participants had become more attuned to the visual forms of the characters. In addition, participants could quickly develop stable visual representations of the characters, which remained tolerant to input variability. Crucially, the absence of differences in the results between the script that participants learned to read and write, and the script with which they were only visually familiarized implies two key points: (1) participants were able to construct resilient representations of the novel characters even when there were not connected to phonological information, and (2) these representations are not orthographic in nature. The results from Experiment 2 further corroborate our initial conclusion. This experiment aimed to assess the development of location-invariance processing. We employed a same-different task to compare participants' accuracy in responding to transposed-letter versus replaced-letter pairs, measuring the transposed-letter effect. The findings revealed similar transposed-letter effects for the trained and control scripts, consistent in both the pre- and post-training phases. This extends the observations of Fernández-López et al. (2020) to a context with high variability in visual input. Consequently, our study suggests that learning to read and write in a new script does not necessarily lead to a rapid emergence of location-invariant processing.

All in all, the results of our experiments provide valuable insights for understanding the process of learning to read in a novel script. Despite achieving fluency in reading and writing the new script over five training sessions, participants did not sufficiently develop orthographic representations in terms of both letter identities and letter order. A possible explanation for this pattern might be that the intensity of training in a novel script cannot replicate the extensive exposure and experience with letters and letter strings typically received by children when learning to read and write. The development of children's reading skills often involves substantial visual familiarization with letters, even before formal reading instruction begins, and the association of graphemes and phonemes with words. Thus, the emergence of orthographic processing likely occurs progressively through print exposure (Gomez et al., 2021; Mano & Kloos, 2018). Crucially, the absence of specific patterns typical for language in our experimental letter strings (e.g., frequent letter co-occurrences acquired via statistical learning) may have impacted our results. Such patterns are crucial in orthographic processing and visual word recognition (e.g., Chetail, 2017; Fernández-López & Perea, 2023; Lelonkiewicz et al., 2020, 2023; Y. Vidal et al., 2021). For instance, Chetail (2017) found that adults developed sensitivity to the frequency of bigrams in artificial character streams after brief exposure, regardless of learning grapheme-phoneme associations. Fernández-López and Perea (2023) extended this to the encoding of character order, demonstrating that participants also became attuned to the order of characters in frequently occurring sequences. This evidence suggests that a certain level of regularity, such as letter co-occurrences, is necessary for the cognitive system to effectively modulate letter position encoding.

Our findings can also be interpreted as revealing an intermediary phase in the progression from the general processing of visual objects by the visual system to the specialized processing of orthography. Since reading is a relatively recent development in human history, its foundational representations and processes likely originate from basic visual perception mechanisms. This notion aligns well with the Neuronal Recycling Hypothesis (Dehaene & Cohen, 2007), which posits that the brain repurposes existing neural pathways for new tasks, such as reading. Recent research by Y. Vidal et al. (2021) builds upon this hypothesis, investigating if the bigram frequency effect, typically associated with orthographic material (e.g., Binder et al., 2006; Chetail, 2015; Lochy et al., 2018; Vinckier et al., 2007), could also be observed in non-orthographic stimuli. They discovered that participants were sensitive to co-occurrence patterns across various visual objects, suggesting that mechanisms used in visual word recognition might apply more broadly. This finding supports the idea that letter and word-specific processing evolves from pre-existing visual processing systems as familiarity with orthographic material increases.

Central to this discussion is the role of the Visual Word Form Area (VWFA) in the left ventral occipitotemporal cortex, which is crucial for rapid word recognition in skilled reading (Cohen et al., 2000; Cohen & Dehaene, 2004; Lochy et al., 2018; Vinckier et al., 2007). Neurons in this area become tuned to recognize orthographic regularities, showing increased activation when processing letter sequences resembling words (Binder et al., 2006; Cohen et al., 2002; Vinckier et al., 2007; but see Brem et al., 2006; Tagamets et al., 2000, for alternative views). Developmental studies indicate that VWFA specialization is influenced by early reading experiences (Brem et al., 2010; Dehaene-Lambertz et al., 2018; Eberhard-Moscicka et al., 2015; Lochy et al., 2016; Maurer et al., 2006; Schlaggar & McCandliss, 2007). In this line, recent work underscores the importance of teaching methods, particularly those that automate grapheme-phoneme connections rather than relying on the visual memorization of whole words, in developing advanced reading skills and preventing reading disabilities (Castles et al., 2018; van de Walle de Ghelcke et al., 2020). In light of these findings, models of reading development should incorporate this transitional stage where letters evolve from mere visual objects to recognized orthographic entities. Examining this incipient phase may have significant implications for instructional approaches in early literacy education, thus emphasizing the need for strategies that support this fundamental aspect of learning to read.

4.6 Conclusion

In sum, our experiments examined if orthographic processing, defined as the encoding of letter identities and their positions (Grainger, 2018), could rapidly emerge when learning to read a novel script with visually variable input. Although participants achieved fluency in reading and writing the new script, the evidence from our study does not support the hypothesis of rapid development of orthographic processing under these conditions— a similar pattern of results emerged with a visual control script. Critically, our findings also revealed that exposure to variable visual input did foster the formation of resilient character representations, demonstrating high resistance to distortion. This resilience, however, appears to be rooted in visual cognitive mechanisms rather than the development of orthographic representations.

Chapter 5

General Discussion

The aim of the research presented in this thesis is to enhance our understanding of the mechanisms that underlie learning from the written text. As reviewed in the Introduction, it is clear that humans are remarkably fast learners. This is often observed in connection to the high sensitivity of the human perceptual system to different kinds of regularities in the environment, and the ability to assimilate these regularities rapidly. This exceptional learning capacity has been shown to be particularly useful in order to achieve the primary objective of language: comprehension. In the domain of visual word recognition, it is clear that depending on the context, the system relies on different sources of information, or at least it gives more weight to different sources of information depending on the context. Moreover, it is also clear that sets of rules coalesce into higher-order units when they co-occur within the same context, which generally enables advantage in processing. That is the case of suffixes, which combine meaning and phoneme/grapheme co-occurrence. Yet, these units retain the flexibility to accommodate alternative options, since the matching is rarely complete and exclusive, as in the case of word endings that in some words act as suffixes and in others they do not (e.g. corner – corn effect). This flexibility is essential for the development of a language system, which is in continuous mutation. In this context, suffixes can be conceived as shortcuts to meaning, thus particularly useful for word learning. In the first experiment, I explored the impact of genuine morphology compared to the mere frequency of the letter clusters, with the aim to disentangle potential overlap of the effects due to each of these variables. In the second experiment, I tested whether the success in learning such words could be explained by a more general sensitivity to statistical regularities, bearing in mind that visual object processing is at the root of decoding any written input. Finally, in the third experiment, I look at regularity from a somewhat different perspective and asked if readers can develop stable orthographic representations in the initial stages of literacy, and how this process changes when people are presented with more varying/irregular input.

Summary of the findings

Chapter 2.

In this study, I examined if suffixes have a different role in word learning compared to other type of letter clusters that do not contain any semantics, with the aim of disentangling the genuine morphological effects and those related only to letter frequency. With this aim, I monitored the process of learning and then tested the outcome in three types of words: (i) suffixed novel words (e.g., *flibness*), (ii) novel words that end in non-morphological, but frequent letter chunks (e.g., *fliban*), and (iii) novel words with non-morphological, low-frequency endings (e.g., *flibov*). Each novel word was assigned a meaning, as a consequence of being embedded in a set of sentences that were otherwise familiar (e.g., "Marco doesn't have any *rugobenza* so when his mother scolded him, he also started to yell."). It was found that genuine morphology has a supportive role relative to the other two categories of novel words throughout the learning process, and it also facilitates immediate recognition of the novel words. Additionally, the findings indicate that participants tend to ascribe meaning to constituent word parts (i.e., they establish a connection as in *happy* – *happiness*). Notably, this meaning attribution process is not solely triggered by the presence of a suffix, as similar inferences were made for the novel words without morphology. This suggests that the semantic generalization generated by the learning of a novel word does not depend on the identification of the sub-lexical chunks of the word. Rather, it seems to be driven by a broader correspondence between form and meaning.

Chapter 3.

In Experiment 2, I further explore the behavior of suffixes and their equally frequent non-meaningful counterparts, but now in relation to statistical learning. Since the frequency of letter clusters, and consequently, also suffixes, is statistical in nature, the aim was to examine how much the learning patterns correlate with the statistical learning skill of an individual. With this goal, participants learnt the same type of novel words as in the previous experiment, but this time, they were stripped of any semantic content except for the meaning carried by the existing suffixes (that is, they were presented in isolation and no definition was provided). The outcomes of the learning task were correlated with the statistical learning skill, as measured by a visual statistical learning task. The results of the learning task showed that the items with suffixes and low-frequency endings were learned best, and equally well. Moreover, these two types of words exhibited the strongest correlation with general statistical learning skill. Although the correlations were quite small in size, their robustness and the fact that they follow the pattern of the learning outcome at the group level might suggest that statistical learning has a role in word learning.

Chapter 4.

In the final experiment of this thesis, I focused on the emergence of novel orthographic representations and their resistance to distortion. In this experiment, participants went through a 5-day training to learn a novel alphabet of 11 BACS characters in handwritten format. The training routine included reading, listening, and handwriting. Another set of non-trained 11 BACS characters was used during the sessions as a visual control. Before and after the training, participants completed a masked priming same-different task with the novel alphabet letters, and a standard same-different task, in order to measure fundamental components of orthographic processing: the encoding of letter identity and letter positions. The resistance to distortion was measured as a difference in the priming effect between printed and distorted (captcha-like) primes. It was found that distorted letters serve as effective primes, and more so after the training, suggesting efficient development of the visual representations tolerant to the variability of the input. However, the lack of difference between the learned alphabet and the visual control suggests that the nature of these representation is not specifically orthographic. The results of the same-different task corroborated this interpretation, which similarly revealed no distinction between the two scripts.

Is word learning guided by statistics or by semantics?

The first two experiments of this thesis explored the learning of morphologically complex words in two different contexts: in the first experiment, the emphasis was on semantics, thus an attempt was made to provide a setting close to the environment in which word learning typically happens in real life. On the other hand, in the second experiment, semantic content was completely removed (except for the suffixes themselves), thus arguably encouraging a learning strategy that strongly relies on statistical regularities. In both experiments, suffixes determined a clear advantage in the recognition of the novel words as a whole. In addition, Experiment 1 shows that the advantage of suffixes begins already during the learning process, since the reading times dropped more quickly for the suffixed items compared to words with non-meaningful endings. This is not surprising given previous findings indicating that morphology facilitates guessing meanings of unknown words (McCutchen & Logan, 2011), and highlighting the importance of morphology in visual word recognition (see Amenta & Crepaldi, 2012; Beyersmann & Grainger, 2023; Rastle & Davis, 2008 for reviews). In addition, the present findings replicate those of Ginestet et al. (2020), indicating that complex pseudowords exhibit an advantage compared to their nonmorphological counterparts. Importantly, the present study also extends those findings by controlling for the potential effects of letter cluster frequency. As reviewed in the Introduction, recent studies have shown that even chunks of pseudoletters can yield suffix-like effects. These effects include sensitivity to the position of the letter chunk, in addition to a more general sensitivity to the presence of the letter chunk (Lelonkiewicz et al., 2020).

Further work by Lelonkiewicz and colleagues showed that these effects are increased when letter strings are associated to semantics (Lelonkiewicz et al., 2023). Taking a reverse approach, the second experiment in this thesis stripped the novel words of any meaning, examining the extent to which such effects could be attributed to statistical learning. Notably, suffixes exhibited an advantage even in this context; suffixed items were more easily recognized after the training compared to the other items. In addition, the participants who learned these items better were those with the best performance in the SL task. Interestingly, the same pattern of results was obtained with the low-frequency endings items, in both word learning and SL tasks. Here it is important to consider the issue of regularity versus irregularity in learning. One of the advantages of regular information is that it is reliable. This reliability is well exemplified by suffixes, both at the level of semantics and frequency of letter co-occurrence (e.g., suffix – able is a common word ending and denotes a capacity of the action described by the verb, as in *readable* or *comfortable*). These factors collectively contribute to the establishment of a form-to-meaning relationship, characteristic of suffixes. This relationship is consistent and highly informative (e.g., based only on the suffix *-able*, we know that the word that contains it designates a capacity), which makes it highly reliable. Consequently, due to such reliability, suffixes are a robust source of information within the context of word learning. On the other hand, low-frequency endings represent the opposite extreme in the regularity vs. irregularity opposition: they are deprived of any regularity, in fact they *violate* regularities. As we have observed, they are learned equally well as suffixed items, and the nearly equal correlation with statistical learning suggests that it is the same mechanism that drives the learning success: tracking regularities in the input. Thus, it appears that the cognitive system is engaged in attempting to uncover patterns in the input that do not conform to one would typically experience, and that this leads to better memorization. Clearly, this involves more effort than uncovering an easily identifiable pattern, as in the case of high-frequency endings. Consistent with this interpretation, high-frequency endings exhibited the worst performance (although they were also learnt successfully, with ~75% accuracy); they did not contain either highly regular and reliable suffixes, or highly irregular and engaging low-frequency endings.

This interpretation could be seen as generally aligned with the main postulate of error-driven learning, which states that larger mismatches between expected and observed states generate stronger memories (e.g., Henson & Gagnepain, 2010). While evidence supporting the role of prediction errors in word learning is scarce, some recent findings suggest their potential contribution. For example, Gambi et al. (2021) in Experiments 4 and 5 presented participants with novel words embedded in sentences such as “Now, Peppa will eat the *cheem*.”. Each sentence was accompanied by an image of a pair of objects where one was representing the novel word (e.g. exotic fruit - *cheem*) and another one that could be either plausible direct object of the verb (e.g., eat *an apple*), or implausible as an object of that verb (e.g., eat *a car*). The retention test showed better performance for the words that have been learnt in implausible conditions (i.e., which violated the expectations – when cheem was presented together with a car). In the context of the present experiment, one may conceive low-frequency endings as violating the word-ending pattern, especially in comparison to the two frequent types of endings, while suffixed items may represent the expected, plausible meanings. It is important to note that the existing body of literature on prediction error and error-driven learning has predominantly focused on semantically rich contexts, whereas semantics remains absent in this experiment. Crucially, the studies on error-driven learning were designed to explore the effect of violation of predictions based on the immediate context of the sentence. The mechanisms that are in use in such cases might differ significantly from the those that are at play when the only context is the word itself and its constituent parts. Importantly, in Experiment 1, where novel words were embedded in semantically rich context, no advantage was observed for the low-frequency endings items. Thus, looking at the results of the two experiments together, it appears that error-driven learning might be influenced by the information accessible during the learning process. Therefore, when a rich semantic context is present, learners tend to seek information within that context. Conversely, when the only available context is the structure of the word itself, learners are compelled to search for information within that structure. In the framework of the present experiments, it is likely that sentences of Experiment 1, designed to be consistent and aid in the understanding of a novel word as a whole, did not prompt participants to form strong expectations regarding word endings. Future work could explore further how the learning strategies change depending on the informativeness of the available input. Notably, Experiment 2 has been designed to be comparable to the standard statistical learning task, which might have constrained participants to pay more attention to the visual features of the novel words compared to a task where they were actively searching to comprehend a word in sentence.

Finally, when people read for comprehension, the findings suggest that the type of word ending does not modulate the learning of the word stem, as evidenced by the general lack of an effect of word ending on understanding the stem of the novel word. In all three cases, participants were able to assign appropriate meaning to the stem.

In summary, the results of the Experiment 1 and 2 suggest that whether the cognitive system will rely more heavily on regularities or irregularities during word learning depends on the context of the learning process, the available sources of information, and the specific learning goals.

Does irregular input impact processing of emerging orthographic representations?

The findings reported in Chapter 3 showed that the human visual system can successfully process certain level of distortion. As one might expect, this ability improves with practice, as shown by the stronger priming effects after the training. This confirms previous findings on priming with degraded stimuli (Gil-López et al., 2011; Hannagan et al., 2012). Moreover, the present finding confirms that the effects are prelexical, since other studies used masked lexical decision or semantic categorization task paradigms, which might involve top-down processing (Vergara-Martínez et al., 2015). Furthermore, it appears that participants have become more attuned to the visual forms of the characters, as evidenced by an increase in priming effects post-training.

Nevertheless, the data suggest that orthographic representations have not emerged. The results of the second experiment support this interpretation: no evidence was found for location-invariant processing, an effect widely considered as evidence of abstract orthographic processing (Grainger, 2018). One reason for this pattern could be the difference between a laboratory setting and the much richer experience through which children go during literacy acquisition. However, within the framework of this thesis, another intriguing aspect of the stimuli deserves attention. Unlike real language structures, the letter strings presented to participants lacked typical linguistic patterns, and importantly, connection with the meaning of the word and broader context. Recent research convincingly shows that patterns, such as letter co-occurrence, are essential for orthographic processing and visual word recognition (Chetail, 2017; Fernández-López & Perea, 2023; Lelonkiewicz et al., 2020, 2023; Vidal et al., 2021). In this context, it is important to highlight that letters are primarily visual objects. As individuals gain experience and engage with regularities within them and in the environment surrounding them, they develop into orthographic representations. Thus, it is possible that our results reveal a stage in development that sits halfway between unknown visual objects and language-specific, abstract orthographic representations. Given that reading is a relatively recent invention in human history, its foundational representations and processes likely originate from basic visual perception mechanisms. This notion aligns with the Neuronal Recycling Hypothesis (Dehaene & Cohen, 2007), which posits that the brain repurposes existing neural pathways for new tasks, such as reading. Recent research by Vidal et al. (2021) supports this hypothesis by showing that coding schemes typically associated with orthographic material (e.g., Binder et al., 2006; Chetail, 2015; Lochy et al., 2018; Vinckier et al., 2007) can also be observed in non-orthographic stimuli (e.g., 3D novel visual objects and Gabor patches). Similarly to Lelonkiewicz et al. (2020), who demonstrated that effects typically associated with affixes can be replicated by relying only on statistical patterns, Vidal et al. (2021) found that mechanisms characteristic for visual word recognition might apply more widely to visual objects. This finding gives support to the notion that the development of processing systems specific to letters and words stems from pre-existing visual processing mechanisms. In accordance with this view, developmental studies show that the specialization of the visual word form area in the left occipitotemporal cortex, a brain area crucial for skilled reading, is affected by early reading experiences (Brem et al., 2010; Dehaene-Lambertz et al., 2018; Eberhard-Moscicka et al., 2015; Lochy et al., 2016; Maurer et al., 2006; Schlaggar & McCandliss, 2007). Considering these findings, it would be important for the models of reading development to consider this transitional phase wherein letters transform from visual objects into abstract orthographic representations.

5.1 Conclusion

The work in this thesis explored how regularities influence different aspects of learning in the context of language. In the domain of word learning, I have explored the impact of morphology. The kind of form-to-meaning mapping embodied in affixes is one of the most reliable sources of information when one attempts to uncover the meaning of a novel word. The first two studies showed that indeed, suffixes facilitate word learning. However, where the cognitive system searches for information, it seems to depend on the context of learning: when the learning environment is impoverished semantically, and the acquisition of the novel word is entirely based on rather superficial visual information, basic statistical regularities seem to be a fruitful source of information. Thus, whether the system will favor regular or irregular sources of information might depend on the context in which the learning mechanism is activated. Aligned with the idea that irregular input provides a valuable contribution in forming novel linguistic representations, the third experiment showed that the developing reading system is highly resistant to distortion. As this resistance becomes more robust with exposure to the variable stimuli, it is likely that the cognitive system is able to extract regularities in a very irregular input and learn from it. Overall, the present thesis offers a glimpse into the versatility and adaptability of our language learning and visual word recognition systems.

Chapter 6

Appendices

6.1 Appendix 1
Stimuli from Experiment 2 - (Chapter 3)

Table 5: Description of Stems

ID	Stem	Length	N	OLD20	Bigram Frequency
1	rugob	5	0	2	5.35
2	zudul	5	0	2	5.13
3	mudeg	5	0	2	5.69
4	cebur	5	0	2	5.73
5	vutuv	5	0	2	5.45
6	tobuf	5	0	2	5.60
7	clivun	6	0	2.6	5.84
8	vicraz	6	0	2	6.12
9	cettob	6	0	2	6.24
10	cribot	6	0	2.15	5.76
11	mernid	6	0	2	6.18
12	pubban	6	0	2	5.85
13	dorguv	6	0	2.75	5.55
14	slibad	6	0	2.75	5.56
15	cridol	6	0	2	6.13

Continued on next page

Table 5 – continued from previous page

ID	Stem	Length	N	OLD20	Bigram Frequency
16	cevec	5	0	2	6.09
17	vugip	5	0	2	5.53
18	pobed	5	0	2	5.96
19	netonc	6	0	2	6.45
20	minfur	6	0	2	5.88
21	cotlec	6	0	2	5.90
22	guerud	6	0	2.15	5.83
23	balcus	6	0	2	5.92
24	ghipam	6	0	2.5	5.80
25	morpug	6	0	2.5	5.72
26	fertuc	6	0	2	6.08
27	pummut	6	0	2.6	5.59
28	laccet	6	0	2	6.27
29	relism	6	0	2	6.15
30	gualep	6	0	2	5.95
31	fiolic	6	0	2	6.30
32	funels	6	0	2	5.93
33	suffol	6	0	2	5.71
34	bolset	6	0	2	5.84
35	rastuc	6	0	2.3	6.25
36	crodap	6	0	2.2	5.95
37	lerteb	6	0	2	6.31
38	peglev	6	0	2	6.18
39	smoric	6	0	2	6.14
40	gropal	6	0	2	6.12
41	lonsif	6	0	2	6.16
42	ciamed	6	0	2	6.37
43	vulasc	6	0	2.4	6.04
44	ceghit	6	0	2.05	5.88
45	sapit	5	0	2	6.18
46	raclat	6	0	2	6.17
47	mavub	5	0	2	5.87

Continued on next page

Table 5 – continued from previous page

ID	Stem	Length	N	OLD20	Bigram Frequency
48	palcar	6	0	2	6.25
49	prodan	6	0	2	6.22
50	pelten	6	0	2	6.43
51	vutrac	6	0	2.4	6.15
52	speluf	6	0	2.25	5.86
53	nortub	6	0	2.4	6.21
54	porzor	6	0	2	5.99
55	tuvub	5	0	2	5.23
56	trodab	6	0	2	6.05
57	miscet	6	0	2	6.25
58	chicem	6	0	2	6.24
59	sepuf	5	0	2	5.50
60	futef	5	0	2	5.81
61	tefug	5	0	2	5.54
62	gralad	6	0	2	6.26
63	bolcod	6	0	2	5.87
64	filrec	6	0	2	5.66
65	scicov	6	0	2.3	6.32
66	bonart	6	0	2	6.25
67	bletic	6	0	2.4	6.13
68	cafum	5	0	2	5.56
69	trecut	6	0	2	6.18
70	spulan	6	0	2	6.11
71	sotag	5	0	2	6.27
72	sotiz	5	0	2	6.14
Mean		5.75	0.00	2.09	5.96
SD		0.44	0.00	0.19	0.30
Min		5.00	0.00	2.00	5.13
Max		6.00	0.00	2.75	6.45

Table 6: Description of Endings

ID	Ending	Type	Token frequency	Length	Bigram frequency
			Suffixes		
1	enza	morph	5.20	4	6.04
2	izia	morph	4.76	4	6.06
3	iere	morph	4.66	4	6.62
4	ista	morph	4.90	4	6.48
5	ario	morph	4.74	4	6.57
6	ismo	morph	4.12	4	5.84
7	iero	morph	4.17	4	6.51
8	ale	morph	5.60	3	6.47
Mean			4.77	3.88	6.32
SD			0.49	0.35	0.30
Min			4.12	3.00	5.84
Max			5.60	4.00	6.62
			High frequency endings		
9	ondo	hf	5.00	4	6.52
10	enso	hf	4.93	4	6.31
11	ento	hf	5.42	4	6.65
12	allo	hf	4.69	4	6.44
13	esto	hf	5.05	4	6.61
14	ogno	hf	4.97	4	6.18
15	ordo	hf	4.89	4	6.26
16	oco	hf	4.78	3	6.25
Mean			4.96	3.88	6.40
SD			0.22	0.35	0.18
Min			4.69	3.00	6.18
Max			5.42	4.00	6.65
			Low frequency endings		
17	espa	lf	2.31	4	6.22
18	olmo	lf	2.60	4	6.01

Continued on next page

Table 6 – continued from previous page

ID	Ending	Type	Token frequency	Length	Bigram frequency
19	iaba	lf	2.16	4	6.04
20	asio	lf	2.16	4	6.38
21	inna	lf	2.60	4	6.27
22	epre	lf	2.37	4	6.03
23	iplo	lf	2.70	4	5.64
24	upe	lf	2.99	3	5.91
Mean			2.49	3.88	6.06
SD			0.29	0.35	0.23
Min			2.16	3.00	5.64
Max			2.99	4.00	6.38

6.2 Appendix 2 Exploratory analyses from Experiment 3 - (Chapter 4)

Supplementary non-preregistered analyses.

6.2.1 Experiment 1. The emergence of abstract letter representations

Data analysis - accuracy

We ran Bayesian generalized linear mixed models to analyze the data using the brms package (Bürkner, 2017, 2018) in R (R CoreTeam, 2023). Phase, script, prime relatedness, and prime distortion, and their 4-way interaction were contrast-coded as fixed effects—these effects were zero-centered: *related* vs. *unrelated* [-0.5 and as 0.5], *pre-training* vs. *post-training* [-0.5 and as 0.5], *trained* vs. *untrained* [-0.5 and as 0.5], and *captcha* set vs. *printed* set [-0.5 and as 0.5]. We used the maximal random structure both for participants and items. We used the Bernoulli distribution with a logit link. The priors for the RT data were weakly informative: Normal ($\mu = 0$, $\sigma = 10$) for the intercept and Normal (0, 1) for each of the fixed effects/interactions and standard deviation parameters. For the covariance matrix of random effects, we had a regularization of 2.

The model was fitted using four chains with 5,000 iterations (1,000 as warmup). We consider an effect credible if the 95% credible interval (CrI) estimated from the posterior distribution does not contain zero. Simple tests effects in case of evidence for interactions were made using the *emmeans* package (Lenth, 2021).

Results

The results of the accuracy data mimicked the same pattern of priming effects as the latency data (see Figure 6.1). We found evidence for an effect of prime-target relatedness ($b = -0.71$, Estimation Error = 0.12, 95% CrI [-0.94, -0.48], with target words being responded more accurately when preceded by an identity prime than an unrelated prime (5.95% vs. 9.85% error rate), and also an effect of prime distortion ($b = 0.44$, Estimation Error = 0.10, 95% CrI [0.24, 0.64], where printed primes yielded more accurate responses than distorted primes (6.65% vs. 9.15% error rate). Moreover, prime-target relatedness interacted with phase ($b = -0.54$, Estimation Error = 0.21, 95% CrI [-0.96, -0.16]: the identity priming effect was smaller after the training (4.9% error rate; b = 0.44, 95% CrI [0.15, 0.73]) than before training (2.9% error rate; $b = 0.98$, 95% CrI [0.67, 1.31]) (see Table 2). Prime-target relatedness also interacted with distortion ($b = -0.38$, Estimation Error = 0.15, 95% CrI [-0.68, -0.08]), where identity priming was larger when the prime was in printed format (4.5% error rate reduction; $b = 0.90$, 95% CrI [0.61, 1.19]) than in distorted format (3.3% error rate reduction; $b = 0.52$, 95% CrI [0.27, 0.78]). Finally, we found no evidence of the effect of phase or the other interactions (see Figure 6.1).

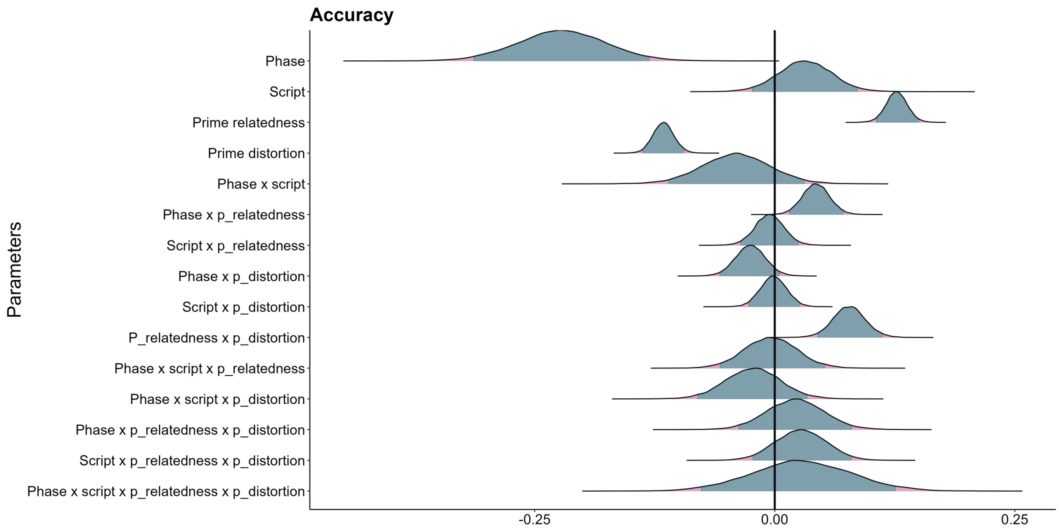


Figure 6.1: 95% and 100% highest density intervals from the Bayesian generalized mixed effects model for the accuracy in the masked priming same-different task.

6.2.2 Experiment 2. The emergence of location-invariant processing with variable visual input

Data analysis - reaction times

We analyzed the data using Bayesian linear mixed model. The fixed effects were phase, training, transposed/replaced letters, and their interaction, with the maximal random structure for participants and items. We used shifted log-normal distribution. The priors and model fitting were identical to the masked priming same-different task. Again, we consider an effect as credible where the 95% credible interval (CrI) estimated from the posterior distribution did not contain zero. The *emmeans* package (Lenth, 2021) was used to unpack significant interactions.

Results

We found evidence for an effect of phase (see Figure 6.2), with an advantage in post-training compared to pre-training (667 ms vs. 593 ms), ($b = -0.11$, Estimation Error = 0.03, 95% CrI [-0.16, -0.06]). Moreover, transposed letter effect emerged independently ($b = -0.05$, Estimation Error = 0.01, 95% CrI [-0.07, -0.04]), with transposed-letter condition being slower compared to replaced (649 ms vs. 611 ms). The effect emerged also in interaction with phase ($b = -0.03$, Estimation Error = 0.01, 95% CrI [-0.05, -0.00]), with stronger effect emerging after the training (pre-training: 29.31 ms; $b = 0.04$, 95% CrI [0.02, 0.06] vs. post-training: 47 ms; $b = 0.07$, 95% CrI [0.05, 0.09]). Phase also interacted with script ($b = -0.07$, Estimation Error = 0.03, 95% CrI [-0.12, -0.01]), where facilitation was strong in pre-training but was lost in post-training (44 ms; $b = -0.06$, 95% CrI [-0.11, -0.02] vs. -3 ms $b = 0.01$, 95% CrI [-0.03, 0.04]).

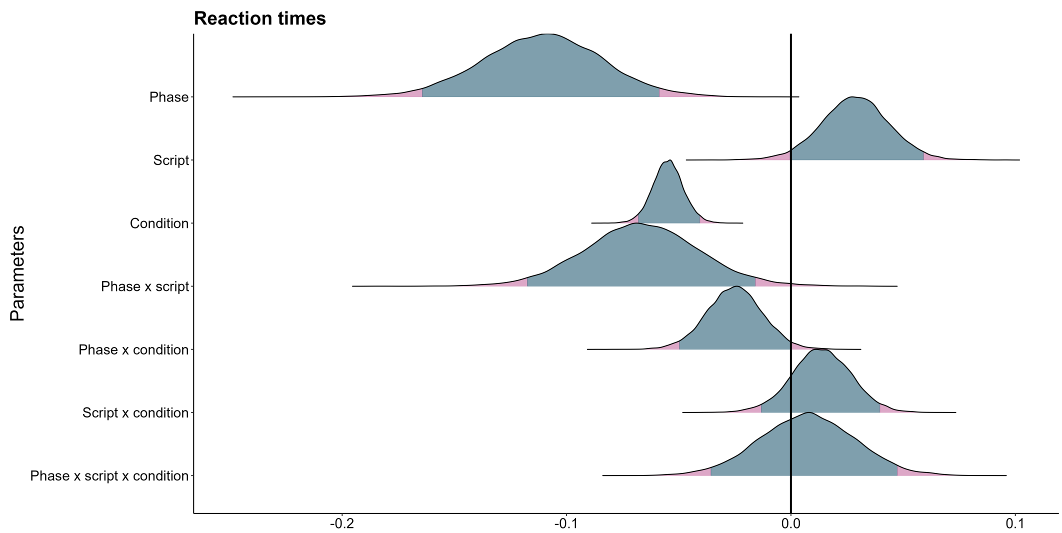


Figure 6.2: 95% and 100% highest density intervals from the Bayesian linear mixed effects model for reaction times in same-different task.

Bibliography

Ahn, L. Von, Blum, M., Hopper, N. J., & Langford, J. (2003). CAPTCHA : Using Hard AI Problems for Security. *Lecture Notes in Computer Science, 2656*, 294–311. https://doi.org/10.1007/3-540-39200-9_18

Algeo, J., & Algeo, A. (1993). Fifty Years among the New Words: A Dictionary of Neologisms 1941–1991 (1st ed.). Cambridge University Press.

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe : A simple network model. *Proceedings of the National Academy of Sciences, 91(July)*, 7041–7045. https://doi.org/10.1073/pnas.91.15.7041

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(JUL)*, 1–12. https://doi.org/10.3389/fpsyg.2012.00232

Amenta, S., Crepaldi, D., & Marelli, M. (2020). Consistency measures individuate dissociating semantic modulations in priming paradigms: A new look on semantics in the processing of (complex) words. *Quarterly Journal of Experimental Psychology, 73(10)*, 1546–1563.

Amenta, S., Hasenäcker, J., Crepaldi, D., & Marelli, M. (2023). Prediction at the intersection of sentence context and word form: Evidence from eye-movements and self-paced reading. *Psychonomic Bulletin and Review, 30(3)*, 1081–1092. https://doi.org/10.3758/s13423-022-02223-9

Amenta, S., Marelli, M., & Sulpizio, S. (2016). From sound to meaning : Phonology-to-Semantics mapping in visual word recognition. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-016-1152-0

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(3)*, 279–296. https://doi.org/10.1016/j.jml.2012.12.001

Araujo, S., Domingues, M., & Fernandes, T. (2022). From Hand to Eye: a Meta-Analysis of the Benefit from Handwriting Training in Visual Graph Recognition. September. https://doi.org/10.1007/s10648-021-09651-4

Arciuli, J., & Simpson, I. C. (2012). Statistical Learning Is Related to Reading Ability in Children and Adults. *Cognitive Science, 36(2)*, 286–304. https://doi.org/10.1111/j.1551-6709.2011.01200.x

Aronoff, M., & Anshen, F. (2017). Morphology and the Lexicon: Lexicalization and Productivity. In A. Spencer & A. M. Zwicky (Eds.), *The Handbook of Morphology (First, pp. 237–247)*. John Wiley & Sons, Inc. https://doi.org/https://doi.org/10.1002/9781405166348.ch11

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 58(6)*, 1065–1086. https://doi.org/10.1080/02724980443000476

Baayen, R. H., Milin, P., Filipović Đurđević, D., Hendrix, P., & Marelli, M. (2011). An Amorphous Model for Morphological Processing in Visual Comprehension Based on Naive Discriminative Learning. *Psychological Review, 118(3)*, 438–481. https://doi.org/10.1037/a0023851

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia, 79*, 33–41. https://doi.org/10.1016/j.neuropsychologia.2015.10.020

Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67(1)*, 1–48. https://doi.org/10.18637/jss.v067.i01

Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: an event-related potential study. *Journal of Cognitive Neuroscience, 23(11)*, 3181–3196. https://doi.org/10.1162/jocn

Beyersmann, E., Casalis, S., Ziegler, J. C., & Grainger, J. (2015). Language proficiency and morpho-orthographic segmentation. *Psychonomic Bulletin & Review, 22*, 1054–1061. https://doi.org/10.3758/s13423-014-0752-9

Beyersmann, E., Castles, A., & Coltheart, M. (2012). Morphological processing during visual word recognition in developing readers: Evidence from masked priming. *Quarterly Journal of Experimental Psychology, 65(7)*, 1306–1326. https://doi.org/10.1080/17470218.2012.656661

Beyersmann, E., Duñabeitia, J. A., Carreiras, M., Coltheart, M., & Castles, A. (2013). Early morphological decomposition of suffixed words : Masked priming evidence with transposed-letter nonword primes. *Applied Psycholinguistics, 34*, 869–892.

Beyersmann, E., & Grainger, J. (2023). The Role of Embedded Words and Morphemes in Reading. In D. Crepaldi (Ed.), *Linguistic Morphology in the Mind and Brain (1st ed., pp. 26–49)*. Routledge. https://doi.org/10.4324/9781003159759-3

Beyersmann, E., Grainger, J., Casalis, S., & Ziegler, J. C. (2015). Effects of reading proficiency on embedded stem priming in primary school children. *Journal of Experimental Psychology, 139*, 115–126. https://doi.org/10.1016/j.jecp.2015.06.001

Beyersmann, E., Kezilas, Y., Coltheart, M., Castles, A., Ziegler, J. C., Taft, M., & Grainger, J. (2018). Taking the Book from the Bookshelf: Masked Constituent Priming Effects from Compound Words and Nonwords. *Journal of Cognition, 1(1)*, 1–13. https://doi.org/https://doi.org/10.5334/joc.11

Beyersmann, E., Mousikou, P., Javourey-Drevet, L., Schroeder, S., Ziegler, J. C., & Grainger, J. (2020). Morphological Processing across Modalities and Languages. *Scientific Studies of Reading, 24(6)*, 500–519. https://doi.org/10.1080/10888438.2020.1730847

Beyersmann, E., Wegener, S., Nation, K., Prokupzcuk, A., Wang, H. C., & Castles, A. (2020). Learning morphologically complex spoken words: Orthographic expectations of embedded stems are formed prior to print exposure. *Journal of Experimental Psychology. Learning, Memory, and Cognition.*

Beyersmann, E., Ziegler, J. C., Castles, A., Coltheart, M., Kezilas, Y., & Grainger, J. (2016). Morpho-orthographic segmentation without semantics. *Psychonomic Bulletin and Review, 23(2)*, 533–539. https://doi.org/10.3758/s13423-015-0927-z

Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E., & Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. *NeuroImage, 33(2)*, 739–748. https://doi.org/10.1016/j.neuroimage.2006.06.053

Bogaerts, L., Siegelman, N., Christiansen, M. H., & Frost, R. (2022). Is there such a thing as a 'good statistical learner'? In *Trends in Cognitive Sciences (Vol. 26, Issue 1, pp. 25–37)*. Elsevier Ltd. https://doi.org/10.1016/j.tics.2021.10.012

Bolger, D. J., Balass, M., & Landen, E. (2008). Context Variation and Definitions in Learning the Meanings of Words : An Instance-Based Learning Approach. *Discourse Processes, 45*, 122–159.

Bottini, R., & Crepaldi, D. (2016). The Nature of Semantic Priming by Subliminal Spatial Words : Embodied or Disembodied ? *Journal of Experimental Psychology: General, 145(9)*, 1160–1176.

Brem, S., Bach, S., Kucian, K., Guttorm, T. K., Martin, E., Lyytinen, H., Brandeis, D., & Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proceedings of the National Academy of Sciences of the United States of America, 107(17)*, 7939–7944. https://doi.org/10.1073/pnas.0904402107

Brem, S., Bucher, K., Halder, P., Summers, P., Dietrich, T., Martin, E., & Brandeis, D. (2006). Evidence for developmental changes in the visual word processing network beyond adolescence. *NeuroImage, 29(3)*, 822–837. https://doi.org/10.1016/j.neuroimage.2005.09.023

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science, 27(1)*, 45–50. https://doi.org/10.1177/0963721417727521

Burani, C., Arduino, L. S., & Marcolini, S. (2006). Naming morphologically complex pseudowords : A headstart for the root ? *2*, 299–328.

Burani, C., Marcolini, S., De Luca, M., & Zoccolotti, P. (2008). Morpheme-based reading aloud: Evidence from dyslexic and skilled Italian readers. *Cognition, 108(1)*, 243–262. https://doi.org/10.1016/j.cognition.2007.12.010

Burani, C., Marcolini, S., & Stella, G. (2002). How early does morpholexical reading develop in readers of a shallow orthography? *Brain and Language, 81(1–3)*, 568–586. https://doi.org/10.1006/brln.2001.2548

Bürkner, P. (2017). brms : An R Package for Bayesian Multilevel Models. *Journal of Statistical Software, 80(1)*. https://doi.org/10.18637/jss.v080.i01

Bürkner, P. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal, 10(1)*, 395–411. https://doi.org/https://doi.org/10.32614/RJ-2018-017

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morrphology. *Cognition, 28*, 297–332.

Carr, J. W. (2021). *Eyekit*. https://jwcarr.github.io/eyekit/

Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest, 19(1)*, 5–51. https://doi.org/10.1177/1529100618772271

Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning New Word Meanings from Context: A Study of Eye Movements. *Journal of Experimental Psychology: Learning Memory and Cognition, 27(1)*, 225–235. https://doi.org/10.1037/0278-7393.27.1.225

Chetail, F. (2015). Reconsidering the role of orthographic redundancy in visual word recognition. *Frontiers in Psychology, 6(645)*, 1–10. https://doi.org/10.3389/fpsyg.2015.00645

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

Chetail, F., Balota, D., Treiman, R., & Content, A. (2015). What can megastudies tell us about the orthographic structure of English words? *Quarterly Journal of Experimental Psychology, 68(8)*, 1519–1540. https://doi.org/10.1080/17470218.2014.963628

Christiansen, M. H. (2019). Implicit Statistical Learning: A Tale of Two Literatures. *Topics in Cognitive Science, 11(3)*, 468–481. https://doi.org/10.1111/tops.12332

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *NeuroImage, 22*, 466–476. https://doi.org/10.1016/j.neuroimage.2003.12.049

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area. Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain, 123(2)*, 291–307. https://doi.org/10.1093/brain/123.2.291

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain, 125(5)*, 1054–1069. https://doi.org/10.1093/brain/awf094

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC : A Dual Route Cascaded Model of Visual Word Recognition and Reading

Aloud. *108(I)*, 204–256.

Crepaldi, D., Amenta, S., Mandera, P., Keuleers, E., & Brysbaert, M. (2016). Frequency estimates from different registers explain different aspects of visual word recognition. *International Meeting of the Psychonomic Society, Granada, 5-8 May.*

Crepaldi, D., Hemsworth, L., Davis, C. J., & Rastle, K. (2016). Masked suffix priming and morpheme positional constraints. *Quarterly Journal of Experimental Psychology, 69(1)*, 113–128. https://doi.org/10.1080/17470218.2015.1027713

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. https://doi.org/10.1016/j.jml.2010.03.002

Crepaldi, D., Rastle, K., & Davis, C. J. (2010). Morphemes in their place: Evidence for position-specific identification of suffixes. *Memory and Cognition, 38(3)*, 312–321. https://doi.org/10.3758/MC.38.3.312

Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance, 32(3)*, 668–687. https://doi.org/10.1037/0096-1523.32.3.668

Davis, M. H., & Gaskell, G. M. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1536)*, 3773–3800. https://doi.org/10.1098/rstb.2009.0111

Dawson, N., Rastle, K., & Ricketts, J. (2018). Morphological effects in visual word recognition: Children, adolescents, and adults. *Journal of Experimental Psychology: Learning Memory and Cognition, 44(4)*, 645–654. https://doi.org/10.1037/xlm0000485

Dawson, N., Rastle, K., & Ricketts, J. (2021). Bridging form and meaning: support from derivational suffixes in word learning. *Journal of Research in Reading, 44(1)*, 27–50. https://doi.org/10.1111/1467-9817.12338

De Rosa, M., & Crepaldi, D. (2022). Letter chunk frequency does not explain morphological masked priming. *Psychonomic Bulletin & Review, 29*, 589–599.

DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through Data Simulation. *Advances in Methods and Practices in Psychological*

Science, 4(1), 1–15. https://doi.org/10.1177/2515245920965119

Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS Biology, 16(3)*, 1–34. https://doi.org/10.1371/journal.pbio.2004103

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences, 9(7)*, 335–341. https://doi.org/10.1016/j.tics.2005.05.004

Diependaele, K., Sandra, D., & Grainger, J. (2009). Semantic transparency and masked morphological priming: The case of prefixed words. *Memory and Cognition, 37(6)*, 895–908. https://doi.org/10.3758/MC.37.6.895

Drews, E. (1995). Morphological and Orthographic Similarity in Visual Word Recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21(5)*, 1098–1116.

Duñabeitia, J. A., Dimitropoulou, M., Grainger, J., Hernández, J. A., & Carreiras, M. (2012). Differential sensitivity of letters, numbers and symbols to character transpositions. *Journal of Cognition, 24(7)*, 1610–1624. https://doi.org/10.1162/jocn

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2008). Does darkness lead to happiness? Masked suffix priming effects. *Language and Cognitive Processes, 23(7–8)*, 1002–1020. https://doi.org/10.1080/01690960802164242

Eberhard-Moscicka, A. K., Jost, L. B., Raith, M., & Maurer, U. (2015). Neurocognitive mechanisms of learning to read: Print tuning in beginning readers related to word-reading.

Feldman, L. B., Milin, P., Cho, K. W., Moscoso del Prado Martín, F., & O'Connor, P. A. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience, 9(March)*, 1–19. https://doi.org/10.3389/fnhum.2015.00111

Feldman, L. B., O'Connor, P. A., & Moscoso del Prado Martín, F. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin and Review.* https://doi.org/10.3758/PBR.16.4.684

Fernandes, T., & Araújo, S. (2021). From Hand to Eye With the Devil In-Between : Which Cognitive Mechanisms Underpin the Benefit From Handwriting Training When Learning Visual Graphs? *Frontiers in Psychology, 12*, 736507. https://doi.org/10.3389/fpsyg.2021.736507

Fernández-López, M., Marcet, A., & Perea, M. (2020). Does orthographic processing emerge rapidly after learning a new script? *British Journal of Psychology, 112(1)*, 52–91. https://doi.org/10.1111/bjop.12469

Fernández-López, M., & Perea, M. (2023). A Letter is a Letter and its Co-Occurrences: Cracking the Emergence of Position-Invariance Processing. *Psychonomic Bulletin and Review, 30(6)*, 2328–2337. https://doi.org/10.3758/s13423-023-02265-7

Firth, J. (1957). A synopsis of linguistic theory, 1930-55. *Studies in Linguistic Analysis*, 1–31.

Fiser, J., & Aslin, R. N. (2001). UNSUPERVISED STATISTICAL LEARNING OF HIGHER-ORDER SPATIAL STRUCTURES FROM VISUAL SCENES. *Psychological Science, 12(6)*, 499–504.

Forster, K. I., & Forster, J. C. (2003). DMDX : A Windows display program with millisecond accuracy. *Behavior Research Methods, 35(1)*, 116–124. https://doi.org/10.3758/bf03195503

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(2)*, 498–514. https://doi.org/10.1037/0278-7393.24.2.498

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression (3rd ed.)*. Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Frank, S. L., & Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science, 22(6)*, 829–834. https://doi.org/10.1177/0956797611409589

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin, 1–87*. https://doi.org/10.1037/bul0000210

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. In *Trends in Cognitive Sciences (Vol. 19, Issue 3, pp. 117–125)*. Elsevier Ltd. https://doi.org/10.1016/j.tics.2014.12.010

Gambi, C., Pickering, M. J., & Rabagliati, H. (2021). Prediction error boosts retention of novel words in adults but not in children. *Cognition, 211(January)*, 104650. https://doi.org/10.1016/j.cognition.2021.104650

Gaskell, G. M., Cairney, S. A., & Rodd, J. M. (2019). Contextual priming of word meanings is stabilized over sleep. *Cognition, 182(May 2018)*, 109–126. https://doi.org/10.1016/j.cognition.2018.09.007

Gaskell, G. M., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89(2)*, 105–132. https://doi.org/10.1016/S0010-0277(03)00070-2

Gaskell, G. M., & Ellis, A. W. (2009). Word learning and lexical development across the lifespan. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1536)*, 3607–3615. https://doi.org/10.1098/rstb.2009.0213

Gibson, E. J., Gibson, J. J., Pick, A. D., & Osser, H. (1962). A developmental study of the discrimination of letter-like forms. *Journal of Comparative and Physiological Psychology, 55(6)*, 897–906.

Gil-López, C., Perea, M., Moret-Tatay, C., & Carreiras, M. (2011). Can masked priming effects be obtained with handwritten words? *Attention, Perception, and Psychophysics, 73(6)*, 1643–1649. https://doi.org/10.3758/s13414-011-0174-y

Ginestet, E., Shadbolt, J., Tucker, R., Bosse, M. L., & Deacon, S. H. (2020). Orthographic learning and transfer of complex words: Insights from eye tracking during reading and learning tasks. *Journal of Research in Reading, 00(00)*, 2–20. https://doi.org/10.1111/1467-9817.12341

Ginestet, E., Valdois, S., Diard, J., & Bosse, M. L. (2020). Orthographic learning of novel words in adults: effects of exposure and visual attention on eye movements. *Journal of Cognitive Psychology, 32(8)*, 785–804. https://doi.org/10.1080/20445911.2020.1823987

Gomez, P., Marcet, A., & Perea, M. (2021). Are better young readers more likely to confuse their mother with their mohter? *Quarterly Journal of Experimental Psychology, 74(9)*, 1542–1552. https://doi.org/10.1177/17470218211012960

Gomez, P., Ratcliff, R., & Perea, M. (2008). The Overlap Model: A Model of Letter Position Coding. *Psychological Review, 115(3)*, 577–600. https://doi.org/10.1037/a0012667

Grainger, J. (2018). Orthographic processing: A 'mid-level' vision of reading : The 44th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology, 71(2)*, 335–359. https://doi.org/10.1080/17470218.2017.1314515

Grainger, J., & Beyersmann, E. (2017). Edge-Aligned Embedded Word Activation Initiates Morpho-orthographic Segmentation. *Progress in Brain Research*. https://doi.org/10.1016/bs.plm.2017.03.009

Grainger, J., & Beyersmann, E. (2021). Effects of Lexicality and Pseudo-Morphological Complexity on Embedded Word Priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47(3)*, 518–531. https://doi.org/http://dx.doi.org/10.1037/xlm0000878

Grainger, J., Snell, J., & Beyersmann, E. (2021). Morphological processing in the flankers task. *Language, Cognition and Neuroscience, 36(3)*, 288–295. https://doi.org/10.1080/23273798.2020.1810292

Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00054

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *2nd Meeting of the North American Chapter of the Association for Computational Linguistics, NAACL 2001*. https://doi.org/10.3115/1073336.1073357

Hannagan, T., Ktori, M., Chanceaux, M., & Grainger, J. (2012). Deciphering CAPTCHAs : What a Turing Test Reveals about Human Cognition. *PLOS ONE, 7(3)*, 5–7. https://doi.org/10.1371/journal.pone.0032121

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the lifespan. *Psychological Science, 26(4)*, 433–443. https://doi.org/10.1177/0956797614567339.When

Hasenäcker, J., Beyersmann, E., & Schroeder, S. (2016). Masked morphological priming in German-speaking adults and children: Evidence from response time distributions. *Frontiers in Psychology, 7*, 1–11. https://doi.org/10.3389/fpsyg.2016.00929

Havas, V., Waris, O., Vaquero, L., Rodríguez-Fornells, A., & Laine, M. (2015). Morphological learning in a novel language: A cross-language comparison. *Quarterly Journal of Experimental Psychology, 68(7)*, 1426–1441. https://doi.org/10.1080/17470218.2014.983531

Henderson, L., Devine, K., Weighall, A. R., & Gaskell, G. M. (2015). When the daffodat flew to the intergalactic zoo: Off-line consolidation is critical for word learning from stories. *Developmental Psychology, 51(3)*, 406–417. https://doi.org/10.1037/a0038786

Henderson, L., & James, E. (2017). Consolidating new words from repetitive versus multiple stories: The role of prior knowledge. *Journal of Experimental Child Psychology, 1–41*. https://doi.org/https://doi.org/10.1016/j.jecp.2017.09.017

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus, 20(11)*, 1315–1326. https://doi.org/10.1002/hipo.20857

Isbilen, E. S., Mccauley, S. M., & Kidd, E. (2020). Statistically Induced Chunking Recall : A Memory-Based Approach to Statistical Learning. *44*. https://doi.org/10.1111/cogs.12848

Jackson, N. E., & Coltheart, M. (2001). Routes To Reading Success and Failure: Toward an Integrated Cognitive Psychology of Atypical Reading (1st ed.). *Psychology Press*. https://doi.org/https://doi.org/10.4324/9780203759905

James, E., Gaskell, G. M., & Henderson, L. (2018). Offline consolidation supersedes prior knowledge benefits in children's (but not adults') word learning. *Developmental Science*. https://doi.org/10.1111/desc.12776

James, E., Gaskell, G. M., Weighall, A. R., & Henderson, L. (2017). Consolidation of vocabulary during sleep: The rich get richer? *Neuroscience and Biobehavioral Reviews, 77*, 1–13. https://doi.org/10.1016/j.neubiorev.2017.01.054

Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology, 166*, 190–211. https://doi.org/10.1016/j.jecp.2017.08.010

Joseph, H., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition, 133(1)*, 238–248. https://doi.org/10.1016/j.cognition.2014.06.015

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. *Quarterly Journal of Experimental Psychology, 68(8)*, 1665–1692. https://doi.org/10.1080/17470218.2015.1022560

Kinoshita, S., Gayed, M., & Norris, D. (2018). Orthographic and phonological priming effects in the same-different task. *Journal of Experimental Psychology: Human Perception and Performance, 44(11)*, 1661–1671. https://doi.org/10.1037/xhp0000548

Kinoshita, S., & Norris, D. (2009). Transposed-Letter Priming of Prelexical Orthographic Representations. *Journal of Experimental Psychology: Learning Memory and Cognition, 35(1)*, 1–18. https://doi.org/10.1037/a0014277

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy : evidence for a domain general learning mechanism. *Cognition, 83*, 4–5. https://doi.org/10.1016/S0010-0277(02)00004-5

Kumaran, D., & McClelland, J. L. (2012). Generalization Through the Recurrent Interaction of Episodic Memories : A Model of the Hippocampal System. *Psychological Review, 119(3)*, 573–616. https://doi.org/10.1037/a0028681

Leach, L., & Samuel, A. G. (2007). Lexical Configuration and Lexical Engagement: When Adults Learn New Words. *Cognitive Psychology, 55(4)*, 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001.Lexical

Lelonkiewicz, J. R., Ktori, M., & Crepaldi, D. (2020). Morphemes as letter chunks : Discovering affixes through visual regularities. *Journal of Memory and Language, 115*, 104152. https://doi.org/10.1016/j.jml.2020.104152

Lelonkiewicz, J. R., Ktori, M., & Crepaldi, D. (2023). Morphemes as letter chunks: Linguistic information enhances the learning of visual regularities. *Journal of Memory and Language, 130(104411)*. https://doi.org/10.1016/j.jml.2023.104411

Leminen, A., Lehtonen, M., Bozic, M., & Clahsen, H. (2016). Morphologically complex words in the mind/brain. *Frontiers in Human Neuroscience, 10*, 1–3. https://doi.org/10.3389/fnhum.2016.00047

Leminen, A., Smolka, E., Duñabeitia, J. A., & Pliatsikas, C. (2019). Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding). *Cortex, 116*, 4–44. https://doi.org/10.1016/j.cortex.2018.08.016

Lenth, R. V. (2021). emmeans: Estimated Marginal Means, aka Least-Squares Means (1.6.3). https://doi.org/10.1080/00031305.1980.10483031

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106(3)*, 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Li, J. X., & James, K. H. (2016). Handwriting generates variable visual input to facilitate symbol learning. *Journal of Experimental Psychology General, 145(3)*, 298–313. https://doi.org/10.1037/xge0000134

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. *Proceedings of the National Academy of Sciences of the United States of America, 115(32)*, E7595–E7604. https://doi.org/10.1073/pnas.1718987115

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 of the United States of America, 113(30)*, 8544–8549. https://doi.org/10.1073/pnas.1520366113

Longtin, C. M., Segui, J., & Hallé, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes, 18(3)*, 313–334. https://doi.org/10.1080/01690960244000036

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19(1)*, 1–36. https://doi.org/10.1097/00003446-199802000-00001

Macmillan, N. A., & Creelman, D. C. (2005). *Detection Theory: A User's Guide (2nd ed.)*. Psychology Press.

Mak, M. H. C., Hsiao, Y., & Nation, K. (2021). Anchoring and contextual variation in the early stages of incidental word learning during reading. *Journal of Memory and Language, 118(104203)*. https://doi.org/https://doi.org/10.1016/j.jml.2020.104203

Mano, Q. R., & Kloos, H. (2018). Sensitivity to the Regularity of Letter Patterns Within Print Among Preschoolers: Implications for Emerging Literacy. *Journal of Research in Childhood Education, 32(4)*, 379–391. https://doi.org/10.1080/02568543.2018.1497736

Marelli, M., & Amenta, S. (2018). A database of orthography-semantics consistency (OSC) estimates for 15,017 English words. *Behavior Research Methods, 50*, 1482–1495. https://doi.org/https://doi.org/10.3758/s13428-018-1017-8A

Marelli, M., Amenta, S., & Crepaldi, D. (2015). Semantic transparency in free stems: The effect of Orthography-Semantics Consistency on word recognition. *Quarterly Journal of Experimental Psychology, 68(8)*, 1571–1583. https://doi.org/10.1080/17470218.2014.959709

Marelli, M., & Baroni, M. (2015). Affixation in Semantic Space: Modeling Morpheme Meanings With Compositional Distributional Semantics. *Psychological Review, 122(3)*, 485–515. https://doi.org/10.1037/a0039267

Marslen-Wilson, W. D. (1990). Activation, Competition, and Frequency in Lexical Access. In G. T. M. Altmann (Ed.), *Cognitive Models of Speech Processing* (pp. 148–172). MIT Press. https://doi.org/10.7551/mitpress/1889.003.0008

Massol, S., Duñabeitia, J. A., Carreiras, M., & Grainger, J. (2013). Evidence for Letter-Specific Position Coding Mechanisms. *PLOS ONE, 8(7)*, 1–9. https://doi.org/10.1371/journal.pone.0068460

Maurer, U., Brem, S., Kranz, F., Bucher, K., Benz, R., Halder, P., Steinhausen, H. C., & Brandeis, D. (2006). Coarse neural tuning for print peaks when children learn to read. *NeuroImage, 33(2)*, 749–758. https://doi.org/10.1016/j.neuroimage.2006.06.025

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18(1)*, 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). McClelland PsycRev 1995 HPC-Neocortex.pdf. *Psychological Review, 102(3)*, 419–457. https://doi.org/10.1037/0033-295X.102.3.419

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(5)*, 375–407. https://doi.org/10.1016/S0022-0728(02)01421-3

McCutchen, D., & Logan, B. (2011). Inside incidental word learning: Children's strategic use of morphological information to infer word meanings. *Reading Research Quarterly, 46(4)*, 334–349. https://doi.org/10.1002/RRQ.003

Mirkovic, J., Vinals, L., & Gaskell, G. M. (2019). The role of complementary learning systems in learning and consolidation in a quasi-regular domain. *Cortex, 6(116)*, 228–249. https://doi.org/10.1016/j.cortex.2018.07.015

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning Words from Context. *Reading Research Quarterly, 20(2)*, 233–253.

Nation, I. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63(1)*, 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, K. (2008). Learning to read words. *Quarterly Journal of Experimental Psychology, 61(8)*, 1121–1133. https://doi.org/10.1080/17470210802034603

New, B., Brysbaert, M., Segui, J., Ferrand, L., & Rastle, K. (2004). The processing of singular and plural nouns in French and English. *Journal of Memory and Language, 51*, 568–585. https://doi.org/10.1016/j.jml.2004.06.010

New, B., & Grainger, J. (2011). On letter frequency effects. *Acta Psychologica, 138(2)*, 322–328. https://doi.org/10.1016/j.actpsy.2011.07.001

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition, 52(3)*, 189–234. https://doi.org/10.1016/0010-0277(94)90043-4

Norris, D., & Kinoshita, S. (2008). Perception as Evidence Accumulation and Bayesian Inference: Insights From Masked Priming. *Journal of Experimental Psychology: General, 137(3)*, 434–455. https://doi.org/10.1037/a0012799

Norris, D., & Kinoshita, S. (2012). Reading Through a Noisy Channel: Why There ' s Nothing Special About the Perception of Orthography. *Psychological Review, 119(3)*, 517–545. https://doi.org/10.1037/a0028450

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98(3)*, 554–566. https://doi.org/10.1037/0022-0663.98.3.554

Pagán, A., & Nation, K. (2019). Learning Words Via Reading: Contextual Diversity, Spacing, and Retrieval Effects in Adults. *Cognitive Science, 43(1)*, 1–24. https://doi.org/10.1111/cogs.12705

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51(1)*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perea, M., Jiménez, M., & Gomez, P. (2016). Does Location Uncertainty in Letter Position Coding Emerge Because of Literacy Training? *Journal of*

Experimental Psychology: Learning Memory and Cognition, 42(6), 996–1001. https://doi.org/10.1037/xlm0000208

Perea, M., & Lupker, S. J. (2003). Does jugde activate COURT? Transposed-letter similarity effects in masked associative priming. *Memory and Cognition, 31(6)*, 829–841.

Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language, 51*, 231–246. https://doi.org/10.1016/j.jml.2004.05.005

Perea, M., Marcet, A., & Vergara-Martínez, M. (2016). Does top-down feedback modulate the encoding of orthographic representations during visual-word recognition? *Experimental Psychology, 63(5)*, 278–286. https://doi.org/10.1027/1618-3169/a000327

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(1)*, 56–115. https://doi.org/10.1037/0033-295x.103.1.56

Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D. E., & Arciuli, J. (2018). Hearing Matters More Than Seeing: A Cross-Modality Study of Statistical Learning and Reading Ability. *Scientific Studies of Reading, 23(1)*, 101–115. https://doi.org/10.1080/10888438.2018.1485680

Qiao, E., Vinckier, F., Szwed, M., Naccache, L., Valabrègue, R., Dehaene, S., & Cohen, L. (2010). Unconsciously deciphering handwriting: Subliminal invariance for handwritten words in the visual word form area. *NeuroImage, 49(2)*, 1786–1799. https://doi.org/10.1016/j.neuroimage.2009.09.034

Qiao, X., & Forster, K. I. (2013). Novel Word Lexicalization and the Prime Lexicality Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39(4)*, 1064–1074. https://doi.org/10.1037/a0030528

Qiao, X., Forster, K. I., & Witzel, N. (2009). Is banara really a word? *Cognition, 113(2)*, 254–257. https://doi.org/10.1016/j.cognition.2009.08.006

Rao Mandikal Vasuki, P., Sharma, M., Ibrahim, R., & Arciuli, J. (2017). Statistical learning and auditory processing in children with music training: An ERP study. *Clinical Neurophysiology, 128(7)*, 1270–1281. https://doi.org/10.1016/j.clinph.2017.04.010

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes, 23(7/8)*, 942–971. https://doi.org/10.1080/01690960802069730

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(5)*, 507–537. http://www.tandf.co.uk/journals/pp/01690965.html

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. In *Psychonomic Bulletin & Review (Vol. 11, Issue 6)*.

Ratcliff, R. (1981). A Theory of Order Relations in Perceptual Matching. *Psychological Review, 88(6)*, 552–572. https://doi.org/10.1037/0033-295X.88.6.552

R Core Team. (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing.* https://www.r-project.org/

Reilly, R. C. O., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends in Cognitive Sciences, 6(12)*, 505–510. https://doi.org/https://doi.org/10.1016/S1364-6613(02)02005-3

Roelcke, T. (1997). *Sprachtypologie des Deutschen: Historische, regionale und funktionale Variation (1st ed.).* De Gruyter.

Rueckl, J. G., Mikolinski, M., Raveh, M., Miner, C. S., & Mars, F. (1997). Morphological Priming, Fragment Completion, and Connectionist Networks. *Journal of Memory and Language, 36(3)*, 382–405. https://doi.org/https://doi.org/10.1006/jmla.1996.2489

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science, 274(5294)*, 1926–1928.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and Learning) out of the Corner of Your Ear. *Psychological Science, 8(2)*, 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology, 22(17)*, 1622–1627. https://doi.org/10.1016/j.cub.2012.06.056

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1711)*. https://doi.org/10.1098/rstb.2016.0049

Schlaggar, B. L., & McCandliss, B. D. (2007). Development of neural systems for reading. *Annual Review of Neuroscience, 30*, 475–503. https://doi.org/10.1146/annurev.neuro.28.061604.135645

Schmalz, X., Moll, K., Mulatti, C., & Schulte-Körne, G. (2019). Is Statistical Learning Ability Related to Reading Ability, and If So, Why? *Scientific Studies of Reading, 23(1)*, 64–76. https://doi.org/10.1080/10888438.2018.1482304

Seidenberg, M. S. (1987). Sublexical Structures in Visual Word Recognition: Access Units or Orthographic Redundancy? In M. Coltheart (Ed.), *Attention & performance XII: The psychology of reading (pp. 245–263)*. Erlbaum.

Share, D. L. (1995). Phonological recoding and self-teaching: sine non of reading acquisition. *55*, 151–218.

Siegelman, N. (2019). Statistical learning abilities and their relation to language. *Language and Linguistics Compass, 14(May)*, 1–19. https://doi.org/10.1111/lnc3.12365

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition, 177(August 2017)*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods, 49(2)*, 418–432. https://doi.org/10.3758/s13428-016-0719-z

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language, 81*, 105–120. https://doi.org/10.1016/j.jml.2015.02.001

Siegelman, N., Rueckl, J. G., Lo, J. C. M., Kearns, D. M., Morris, R. D., & Compton, D. L. (2022). Quantifying the regularities between orthography and semantics and their impact on group- and individual-level behavior. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 48(6), 839–855. https://doi.org/https://doi.org/10.1037/xlm0001109

Siegelman, N., Rueckl, J. G., Steacy, L. M., Frost, S. J., van den Bunt, M., Zevin, J. D., Seidenberg, M. S., Pugh, K. R., Compton, D. L., & Morris, R. D. (2020). Individual differences in learning the regularities between orthography, phonology and semantics predict early reading skills. *Journal of Memory and Language, 114*. https://doi.org/10.1016/j.jml.2020.104145

Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as Process: Contextually Cued Attention and Early Word Learning. *Cognitive Science, 34(7)*, 1287–1314. https://doi.org/10.1111/j.1551-6709.2010.01130.x

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128(3)*, 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

Sobczak, J. M., & Gaskell, G. M. (2019). Implicit versus explicit mechanisms of vocabulary learning and consolidation. In *Journal of Memory and Language (Vol. 106)*. https://doi.org/10.1016/j.jml.2019.01.003

SR Research Ltd. (2020). *Pylink (2020.2.3)*. SR Research Ltd.

SR Research Ltd. (n.d.). *Eyelink 1000 Plus*. SR Research Ltd.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31(1)*, 137–149. https://doi.org/https://doi.org/10.3758/BF03207704

Stowe, L. A., Kaan, E., Sabourin, L., & Taylor, R. C. (2018). The Sentence Wrap-up Dogma. *Cognition, 176*, 232–247.

Swanborn, M. S. L., & de Glopper, K. (1999). Incidental Word Learning While Reading: A Meta-Analysis. *Review of Educational Research, 69(3)*, 261–285. https://doi.org/10.3102/00346543069003261

Taft, M. (1994). Interactive-activation as a Framework for Understanding Morphological Processing. *Language and Cognitive Processes, 9(3)*, 271–294. https://doi.org/10.1080/01690969408402120

Taft, M., & Forster, K. I. (1975). Lexical Storage and Retrieval of Prefixed Words. In *JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR (Vol. 14)*.

Taft, M., Hambly, G., & Kinoshita, S. (1986). Visual and Auditory Recognition of Prefixed Words. *The Quarterly Journal of Experimental Psychology Section A, 38(3)*, 351–365. https://doi.org/10.1080/14640748608401603

Taft, M., & Nguyen-Hoan, M. (2010). A sticky stick? The locus of morphological representation in the lexicon. *Language and Cognitive Processes, 25(2)*, 277–296. https://doi.org/10.1080/01690960903043261

Tagamets, M. A., Novick, J. M., Chalmers, M. L., & Friedman, R. B. (2000). A parametric approach to orthographic processing in the brain: An fMRI study. *Journal of Cognitive Neuroscience, 12(2)*, 281–297. https://doi.org/10.1162/089892900562101

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2017). Interaction between episodic and semantic memory networks in the acquisition and consolidation of novel spoken words. *Brain and Language, 167*, 44–60. https://doi.org/10.1016/j.bandl.2016.05.009

Talamo, L., & Celata, C. (2011). Toward a morphological analysis of the Italian lexicon: developing tools for a corpus-based approach. *Quaderni Del Laboratorio Di Linguistica, 10(1)*.

Tamminen, J., Davis, M. H., Merkx, M., & Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition, 125(1)*, 107–112. https://doi.org/10.1016/j.cognition.2012.06.014

Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology, 79*, 1–39. https://doi.org/10.1016/j.cogpsych.2015.03.003

Taylor, J. S. H., Plunkett, K., & Nation, K. (2011). The Influence of Consistency, Frequency, and Semantics on Learning to Read: An Artificial Orthography Paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition, 37(1)*, 60–76. https://doi.org/10.1037/a0020126

Treiman, R., Kessler, B., Boland, K., Clocksin, H., & Chen, Z. (2017). Statistical Learning and Spelling: Older Prephonological Spellers Produce More Wordlike Spellings Than Younger Prephonological Spellers. *00(0)*, 1–13. https://doi.org/10.1111/cdev.12893

Turk-Browne, N. B., Junge, J. A., & Scholl, B. J. (2005). The Automaticity of Visual Statistical Learning. *Journal of Experimental Psychology: General, 134(4)*, 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Tydgat, I., & Grainger, J. (2009). Serial Position Effects in the Identification of Letters, Digits, and Symbols. *35(2)*, 480–498. https://doi.org/10.1037/a0013027

Ulicheva, A., Harvey, H., Aronoff, M., & Rastle, K. (2020). Skilled readers' sensitivity to meaningful regularities in English writing. *Cognition, 195(September 2018)*, 103810. https://doi.org/10.1016/j.cognition.2018.09.013

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research, 30(1)*, 37–69. https://doi.org/10.1023/A:1005204207369

van de Walle de Ghelcke, A., Rossion, B., Schiltz, C., & Lochy, A. (2020). Impact of Learning to Read in a Mixed Approach on Neural Tuning to Words in Beginning Readers. *Frontiers in Psychology, 10(January)*, 1–15. https://doi.org/10.3389/fpsyg.2019.03043

Van Rossum, G., & Drake, F. (2009). Python 3 Reference Manual. *CreateSpace.*

Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S (J. Chambers, W. Eddy, W. Härdle, S. Sheather, & L. Tierney (eds.); 4th ed.). *Springer.*

Vergara-Martínez, M., Gómez, P., Jiménez, M., & Perea, M. (2015). Lexical enhancement during prime–target integration: ERP evidence from matched-case identity priming. *Cognitive, Affective and Behavioral Neuroscience, 15(2)*, 492–504. https://doi.org/10.3758/s13415-014-0330-7

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. https://doi.org/10.3758/s13428-016-0844-8

Vidal, Y., Viviani, E., Zoccolan, D., & Crepaldi, D. (2021). A general-purpose mechanism of visual feature association in visual word identification and beyond. *Current Biology, 31(6)*, 1261-1267.e3. https://doi.org/10.1016/j.cub.2020.12.017

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(1)*, 143–156. https://doi.org/10.1016/j.neuron.2007.05.031

Viviani, E., & Crepaldi, D. (2022). Masked Morphological Priming and Sensitivity to the Statistical Structure of Form–to–Meaning Mapping in L2. *Journal of Cognition, 5(1)*. https://doi.org/10.5334/joc.221

Walker, S., Gaskell, G. M., Knowland, V. C. P., Fletcher, F. E., Cairney, S. A., & Henderson, L. (2020). Growing up with interfering neighbours: The influence of time of learning and vocabulary knowledge on written word learning in children. *Royal Society Open Science, 7(3).* https://doi.org/10.1098/rsos.191597

Weighall, A. R., Henderson, L., Barr, D. J., Cairney, S. A., & Gaskell, G. M. (2016). Eye-tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and Language.* https://doi.org/10.1016/j.bandl.2016.07.010

Yablonski, M., & Ben-Shachar, M. (2016). The morpheme interference effect in Hebrew A generalization across the verbal and nominal domains. *2*, 277–307. https://doi.org/10.1075/ml.11.2.05yab

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15(5)*, 971–979. https://doi.org/10.3758/PBR.15.5.971