Cognition 163 (2017) 103–120

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing

Fabienne Chetail

Laboratoire Cognition Langage Développement (LCLD), Research Centre in Cognition & Neuroscience (CRCN), Université Libre de Bruxelles (ULB), Av. F. Roosevelt, 50, CP 191 - 1050 Brussels, Belgium

ARTICLE INFO

Article history:
Received 4 June 2016
Revised 18 February 2017
Accepted 27 February 2017
Available online 17 March 2017

Keywords:
Reading
Orthographic regularities
Statistical learning
Artificial script
Letter identification

ABSTRACT

Individuals rapidly become sensitive to recurrent patterns present in the environment and this occurs in many situations. However, evidence of a role for statistical learning of orthographic regularities in reading is mixed, and its role has peripheral status in current theories of visual word recognition. Additionally, exactly which regularities readers learn to be sensitive to is still unclear. To address these two issues, three experiments were conducted with artificial scripts. In Experiments 1a and 1b, participants were exposed to a flow of artificial words (five characters) for a few minutes, with either two or four bigrams occurring very frequently. In Experiment 2, exposure took place over several days while participants had to learn the orthographic and phonological forms of new words entailing or not frequent bigrams. Sensitivity to these regularities was then tested in a wordlikeness task. Finally, participants performed a letter detection task, with letters being either of high frequency or not in the exposure phase. The results of the wordlikeness task showed that after only a few minutes, readers become sensitive to the positional frequency of letter clusters and to bigram frequency beyond single letter frequency. Moreover, this new knowledge influenced the performance in the letter detection task, with high-frequency letters being detected more rapidly than low-frequency ones. We discuss the implications of such results for models of orthographic encoding and reading.

1. Introduction

We are surrounded by regularities. The full moon follows the waxing gibbous moon each month; the vast majority of flowers – whatever their appearance – have a pistil surrounded by petals; water becomes ice each time the temperature is around 0 °C. Noticing these kinds of regularities enables one to anticipate sequences of events, to detect anomalies, to make categories. In other words, the *invariants* of events over time give structure to the environment and decrease uncertainty (Gibson, 1971). Note that regularities can also lead to negative effects. The frequent occurrence of a tawny pelage with dark stains in felids helps to distinguish these mammals from those of other families (such as wolves or bears), but it makes it difficult to distinguish between different felids (e.g., cheetah vs. leopard). Accordingly, lab experiments showed that the repeated presence of a semantic feature within a set of items make them difficult to recall (e.g., Baddeley, 1966). In the present study, we examined the impact of sensitivity to regularities in the situation of visual word processing. This general aim was to test whether new regularities learnt by mere print exposure affect the processing of letter strings.

'Statistical learning' usually refers to the ability to rapidly and automatically extract regularities from the environment over time (Schapiro & Turk-Browne, 2015). Several lines of work have shown that we are able to pick up regularities from the flow of stimulation very efficiently and rapidly, be it from a flow of syllables, tones, letters, spatial positions, geometrical shapes, vibration pulses and so on (see Conway & Christiansen, 2005; Perruchet & Pacton, 2006). Investigating statistical learning processes with letter strings could therefore appear as one experimental situation among many others, with the notable exception that both the nature of the stimuli and the nature of the expected behaviour are consistent with what individuals experience out of the lab (Pacton, Perruchet, Fayol, & Cleeremans, 2001).

However, rather than seeking to understand domain-general learning mechanisms involved in statistical learning, most of the studies dealing with statistical learning in the written language aim at investigating the impact of sensitivity to regularities on reading processes per se. Nevertheless, the evidence for such an impact is mixed (Chetail, 2015), limiting its theoretical understanding. The idea that statistical learning of orthographic

E-mail address: fchetail@ulb.ac.be

http://dx.doi.org/10.1016/j.cognition.2017.02.015
0010-0277/

regularities could influence processes of visual word recognition and reading has been fed by the observation that readers very quickly pick up regularities of their language, although this has been investigated mostly for very simple regularities such as consonant doublets. Orthographic regularities refer to untaught facts about the distribution of single letters or letter sequences in print, without direct reference to higher-order levels such as phonological or morphological units. The regularity of appearance of letters is usually estimated by their relative frequency of occurrence in written texts. For example, the letters *S* and *A* co-occur more frequently in English words than the letters *J* and *A*, the letter *R* is more often doubled than the letter *D*, the trigram *CHA* is more frequent at the beginning of words than *PSA*, and the letter *T* is never followed by the letter *X*. Developmental studies have showed that after a few months of exposure to print, young readers capture some of these orthographic regularities (e.g., Cassar & Treiman, 1997; O'Brien, 2014; Pacton et al., 2001). In the wordlikeness task for example, children consider that pseudowords such as *ommera* are more wordlike than pseudowords such as *ovvera*, which is consistent with the fact that the letter M is frequently doubled in French or English whereas the letter V is almost never doubled.

The idea that sensitivity to orthographic regularities contributes to reading efficiency is all the more intriguing given that it is a priori not necessary to rely on such information for reading acquisition. Indeed, in contrast with lexicon formation which requires the *discovery* of words in fluent speech (Saffran, Aslin, & Newport, 1996), reading mastery is achieved after explicit learning of print-to-sound mapping. Given the quasi-systematic relationship between the orthographic and phonological forms of words (at least in alphabetical and syllabic writing systems, see Seidenberg & McClelland, 1989), explicit learning of print-to-sound mapping rules could be sufficient in principle to read the great majority of letter strings, and the pronunciation of exceptional words could be memorised as such. There is therefore no a priori reason to expect any relationship between reading ability and sensitivity to regularities in print. The latter would just be an epiphenomenon with no functional role for written word processing.

Some facts, however, contradict this view. Several studies have reported a link between sensitivity to orthographic regularities and reading/writing efficiency. When presented with two new letter strings, better readers/spellers consider more often that items with frequent letter clusters are more wordlike than items with low-frequency or illegal clusters (e.g., Conrad, Harris, & Williams, 2013; O'Brien, 2014; Rothe, Schulte-Körne, & Ise, 2013). In the same vein, better readers exhibit stronger effects of letter frequency than poorer readers (e.g., Butler, Jared, & Hains, 1984; Mason, 1975; Mason & Katz, 1976). However, these effects have not been consistently reported (e.g., Ise, Arnoldi, & Schulte-Körne, 2014).

To date, to what extent and how sensitivity to orthographic regularities contributes to general reading ability remain open questions (e.g., Rothe, Cornell, Ise, & Schulte-Körne, 2015). Some theoretical proposals have been put forward but not directly tested (e.g., Chetail, 2015; Mano, 2016). In the influential framework of McClelland and Rumelhart (1981), visual word recognition starts with a visual analysis of letter features, which leads to the activation of letter representations that entail the critical features, which in turn activates abstract representations of words including the critical letters. Chetail (2015) proposed that during the whole process readers employ their tacit knowledge of orthographic regularities acquired via statistical learning. Due to repeated exposure, letters or letter clusters of high frequency would have representations with higher resting levels, and their identification within strings would be facilitated. In addition, sensitivity to positional frequency of letters would help to encode letter positions (i.e., perceiving *L* in *CLUSTER* in second rather than third position). Moreover, given that visual word recognition operates on large letter chunks that are processed in parallel and that orthographic regularities naturally map onto these chunks (e.g., Adams, 1981; Seidenberg, 1987), the sensitivity to orthographic regularities would help to give structure to words. Finally, processing regularities would help to reduce the set of lexical competitors during lexical access (low-frequency clusters would be the most diagnostic for lexical selection, i.e., knowing that a word starts with the bigram *ST* is less informative than with *XY*, e.g., Grainger & Ziegler, 2011; Rice & Robinson, 1975).

Evidence supporting the influence of sensitivity to regularities in visual word processing has not been convincing to date, with effects not consistently reported (see Chetail, 2015, for a review) and many researchers in the field would hesitate to ascribe to it a functional role. To explain the mixed pattern, Chetail (2015) proposed that the impact of regularities depends on the size of the regularities manipulated (one, two, three letters) and on the task performed. Single letter frequency would have a facilitative influence in tasks tapping into early stages of word processing such as letter detection tasks, although this result has not been not systematically reported (see New & Grainger, 2011, for a review). For example, participants were more rapid and accurate in deciding that a high-frequency letter (e.g., *E, S*) was present in a letter string than a low-frequency letter (e.g., *Y, K*), especially if the letter was in the first position (e.g., Pitchford, Ledgeway & Masterson, 2008, 2009; see also New & Grainger, 2011). On the contrary, bigram frequency (i.e., frequency of two co-occurring letters) would have a detrimental impact on word processing at later stages (e.g., Chetail, Balota, Treiman, & Content, 2015; Henderson & Chard, 1980; Westbury & Buchanan, 2002 in the lexical decision task), but there is again a lack of consensus (e.g., Gernsbacher, 1984; Keuleers, Lacey, Rastle, & Brysbaert, 2012).

One reason for the weak evidence in favour of the impact of sensitivity to orthographic regularities is that such regularities naturally co-vary with other factors known to influence visual word recognition, such as pronounceability (items with low-frequency letter clusters like *pk* are less pronounceable than those with high-frequency clusters, e.g., Baron, 1975), frequency/familiarity (words with low-frequency letter clusters are usually less frequent/familiar, e.g., Chetail, 2015; Gernsbacher, 1984), or orthographic neighborhood (words with high-frequency letter clusters share letters with more words, e.g., Andrews, 1992). No or insufficient matching on these factors can explain some inconsistencies between studies, but more generally, these natural confounds have made impossible the examination of effects that are strictly associated with orthographic regularities.

One way to overcome this difficulty is to use artificial scripts, that is, sets of characters either coming from unknown scripts (e.g., Thai for monolingual French speakers) or newly devised. Artificial scripts have been used a fair number of times in the field of visual word recognition (see Vidal et al., in press, for a review), particularly as they provide a unique way to thoroughly examine the developmental course of a given orthographic process which is stable in adults. In addition, they make it possible to perfectly control the amount of exposure to the characters across participants, so that one can be sure that there is no difference of familiarity with letter clusters or words. They also enable one to independently manipulate variables that co-vary in real scripts and which are therefore hard to isolate in natural languages, while enabling one to create a large number of stimuli. These properties are particularly critical for the present study.

Some studies in the past have used artificial scripts specifically to examine the impact of sensitivity to orthographic regularities (e.g. Mason & Katz, 1976; Singer, 1980). With native English speakers, Mason and Katz (1976) found that unfamiliar characters (e.g., ξ) were detected more rapidly in artificial words (e.g., θξλΨδΩ) if they occur at a recurrent position, but there was no effect of letter