On Coding Non-Contiguous Letter Combinations

Frédéric Dandurand, Jonathan Grainger, Jon Andoni Duñabeitia and Jean-Pierre Granier

On Coding Non-Contiguous Letter Combinations

Frédéric Dandurand [1*], Jonathan Grainger [2], Jon Andoni Duñabeitia [3], Jean-Pierre Granier [2]

[1] Department of Psychology, Université de Montréal, Montréal, Canada.
[2] Laboratoire de Psychologie Cognitive, CNRS & Aix-Marseille University, Marseille, France.
[3] Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain.

Correspondence:
Dr. Frédéric Dandurand
Department of Psychology
Université de Montréal
Pavillon Marie-Victorin
90, ave. Vincent-d'Indy
Montréal, Québec
H2V 2S9, Canada
frederic.dandurand@gmail.com

Abstract

Starting from the hypothesis that printed word identification initially involves the parallel mapping of visual features onto location-specific letter identities, we analyze the type of information that would be involved in optimally mapping this location-specific orthographic code onto a location-invariant lexical code. We assume that some intermediate level of coding exists between individual letters and whole words, and that this involves the representation of letter combinations. We then investigate the nature of this intermediate level of coding given the constraints of optimality. This intermediate level of coding is expected to compress data while retaining as much information as possible about word identity. Information conveyed by letters is a function of how much they constrain word identity and how visible they are. Optimization of this coding is a combination of minimizing resources (using the most compact representations) and maximizing information. We show that in a large proportion of cases, non-contiguous letter sequences contain more information than contiguous sequences, while at the same time requiring less precise coding. Moreover, we found that the best predictor of human performance in orthographic priming experiments was within-word ranking of conditional probabilities, rather than average conditional probabilities. We conclude that from an optimality perspective, readers learn to select certain contiguous and non-contiguous letter combinations as information that provides the best cue to word identity.

1. Introduction

When trying to understand the functioning of cognitive architectures, important insights can be obtained by studying characteristics of the task and the constraints they place on the processing system. Rational Analysis (Anderson 1990; Oaksford and Chater 1998a, Oaksford and Chater 1998b) aims at studying such characteristics and constraints, taking the standpoint that the brain optimizes processing to perform near optimality given limited cognitive resources. A number of mechanisms, such as heuristics and biases, can be interpreted as optimizing performance under minimal resources. For instance, the inferotemporal neurons of monkeys trained to recognize wire-frame objects generalized their responses to stimuli rotated by 180 degrees around the vertical axis (Logothetis and Pauls 1995). Similar biases and tendencies to spontaneously generalize to symmetrical or mirror stimuli can be found across the cognitive system (e.g., Chater and Brown 2008; Chater and Vitányi 2003; Feldman 2003; Freyd and Tversky 1984). For instance, children learning letter identities typically go through a mirror stage in which they can write indifferently in both directions (Walsh and Butler 1996). In this line, Dehaene et al. (2010) have recently presented behavioral and fMRI data showing that mirror-invariance is present for pictures in a masked mirror priming paradigm, while this was not the case for words, under the assumption that once a certain level of reading experience is attained, mirror writing and/or reading is no longer an automatic spontaneous process.

An important function of the process of learning to perform a particular cognitive activity is to build compact and general representations of stimuli that condense important information while discarding their irrelevant and idiosyncratic details (generalizing to symmetrical or mirror stimuli is often, but not always, useful in this respect). In the domain of reading, because letter size, font, and case are usually irrelevant, the cognitive system learns to devise abstract representations of letters devoid of size, font, and case information (Chauncey, Holcomb, and Grainger 2008; Grainger 2008; Petit et al. 2006; see also Carreiras, Dunabeitia, and Perea 2007). However, a key question driving current research on printed word perception concerns the nature of the orthographic code that enables location-invariant recognition of printed words. This is the focus of the present study.

In the present analysis, we need to make three assumptions in order to get started. The first is that visual word recognition in languages that use an alphabetical script is essentially letter-based. That is, we assume that the printed word is recognized via its constituent letters as opposed to some more holistic word-shape representation (e.g., Pelli, Farell, and Moore 2003; see Grainger 2008 for a summary of the arguments). The second assumption is that the simple visual features that make up the different letters of the alphabet are coded retinotopically in visual cortex (see Tootell et al. 1998, for a review of retinotopic coding in human visual cortex). This suggests that the mapping of such features onto a string of letters must retain a certain amount of retinotopicity (Dehaene 2007; Dehaene et al. 2005). On the other hand, although there is a processing cost associated with shifts in word position relative to eye fixation (e.g., Stevens and Grainger 2003), printed words can be recognized at different locations. Finally, we assume that letter information is input to the cognitive system in a parallel fashion during an eye fixation (Slowiaczek and Rayner 1987). With letter data available in parallel, the cognitive system can easily extract or code for letter combinations. So, what is the nature of the orthographic code that enables location-invariant recognition of printed words?

Today, there is a general consensus that the literate brain implements some form of word-centered, location-independent orthographic coding such that letter identities are coded for their position in the word independently of their position on the retina, at least for words

that require a single fixation for processing. This consensus also extends to the principle that such within-word position coding of letter identities is flexible and approximate – in other words, letter identities are not rigidly allocated to a given position (Davis 2010; Gomez, Ratcliff, and Perea 2008; Grainger and van Heuven 2003; Whitney 2001). Key evidence in favor of such approximate, flexible coding has been obtained using the masked priming paradigm. For a given proportion of letters shared by prime and target, priming effects are not affected by small changes of letter order (transposed-letter priming: Perea and Lupker 2004; Schoonbaert and Grainger 2004), and length-dependent letter position (relative-position priming: Grainger et al. 2006; Peressotti and Grainger 1999). Furthermore, using a combination of masked priming and EEG recordings, Dufau, Grainger, and Holcomb (2008), Carreiras, Duñabeitia and Molinaro (2009) and Grainger and Holcomb (2009) have studied the time course of such flexible orthographic processing. They found that, by the time the ERP components were found to be sensitive to flexible orthographic priming, they were no longer sensitive to whether or not prime and target stimuli were spatially aligned.

In the present work we start from the further assumption that flexible orthographic processing is achieved by coding for letter combinations (Dehaene et al. 2005; Grainger and van Heuven 2003; Whitney 2001; Whitney and Berndt 1999). Our study does not aim to pit this particular type of coding scheme against alternative coding schemes that do not use letter combinations, such as the coding schemes implemented in the SOLAR model (Davis 2010) or the overlap model (Gomez, Ratcliff, and Perea 2008), although this is indeed an important goal for present and future research (see Davis 2010, for a recent comparison of these different theoretical approaches). Given the popularity of the general notion of n-gram coding (especially bigrams and trigrams), and the almost exclusive use of contiguous n-grams before the pioneering work of Mozer (1987) and Whitney and Berndt (1999), it seemed important to analyze the arguments behind this theoretical shift to including non-contiguous letter combinations in n-gram coding. We present and analyze these arguments within the specific framework of Grainger and van Heuven's (2003) parallel open-bigram model, and although some of the arguments are shared with Whitney's (2001) serial open-bigram model (the SERIOL model), a detailed analysis of this particular approach is beyond the scope of the present work (see Whitney and Cornelissen 2008; Whitney 2008, for more information about this specific approach). Instead, we focus on two distinct perspectives within Grainger and van Heuven's parallel open-bigram approach, which we now describe.

The Grainger and van Heuven (2003) proposal is illustrated in Figure 1. In this model, location-invariant orthographic coding occurs at what is referred to as the "relative position map". The relative position map abstracts away from absolute letter position and, instead, focuses on relationships between letters. The first stage of processing in this model is the parallel mapping of visual features onto letter identities at a specific horizontal location with respect to eye fixation (the alphabetic array). This location-specific coding of letter identities is then transformed into a location-invariant prelexical orthographic code (the relative position map) before matching this information with whole-word orthographic representations in long-term memory. The key distinguishing feature of this specific approach is that flexible orthographic coding is achieved by coding for ordered combinations of contiguous and non-contiguous letters (see Mozer 1987; Whitney 2001 for earlier proposals along the same lines).

O-words   silence   science   slice

Relative

position

map

S-I	S-L	S-E	I-L	I-E	I-N
L-E	L-N	L-C	E-N	E-C	E-E
	N-C	N-E	C-E		

Alphabetic
array

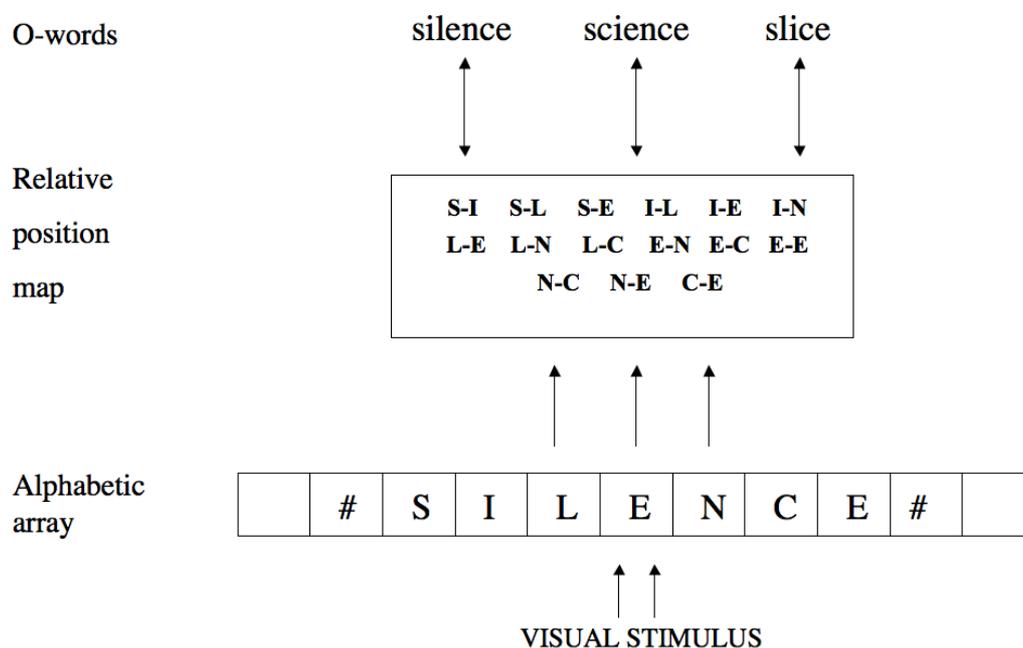| | # | S | I | L | E | N | C | E | # | |

VISUAL STIMULUS

Figure 1 – Grainger and van Heuven's model of orthographic processing (2003). Visual features activate location-specific alphabetic character detectors (the alphabetic array) aligned along the horizontal meridian. Next, information from the alphabetic array is mapped onto ordered pairs of contiguous and non-contiguous letters (not all possible combinations are shown). Finally, relative-position coded letter identities activate whole-word orthographic representations (O-words).

Grainger and van Heuven (2003) proposed open-bigram coding as one means of achieving the gain in flexibility that is necessary in order to capture empirical findings that standard bigram coding schemes cannot account for. As already pointed out by several authors (see Grainger 2008, for a review), standard bigram and trigram coding schemes cannot capture some basic empirical phenomena, such as the transposed-letter and relative-position priming effects mentioned above. In the present theoretical note we provide an exploration of Grainger and van Heuven's open-bigram coding scheme using the general framework of rational analyses of human information processing. More specifically, we examine the utility of coding for non-contiguous as well as contiguous letter combinations within this general approach to orthographic processing, over and above the fact that this confers the flexibility that is necessary to account for key empirical results.

When we begin to think about how coding for non-contiguous letters might emerge in a system that starts with parallel location-specific letter encoding, then two different perspectives emerge. On the one hand, it has been proposed that the coding of non-contiguous letter combinations arises because of noisy position coding in location-specific (retinotopic) letter detectors (e.g., Dehaene et al. 2005). Within this particular perspective, the system aims to code for contiguous letters combinations, and ends up coding for non-contiguous elements as well, because of positional errors (see Grainger et al. 2006, for a

specific implementation of this mechanism referred to as the "overlap open-bigram model"). Therefore, within this perspective, there is no need to justify the coding of non-contiguous letter combinations, they simply arise by accident, and accidentally confer the additional flexibility that allows the coding scheme to capture key data patterns. Here we take a very different perspective, proposing on the contrary that the coding of non-contiguous letter combinations is <u>deliberate</u>, and not the result of inaccurate location-specific letter coding. In other words, non-contiguous letter combinations are coded because they are beneficial with respect to the overall goal of mapping letters onto meaning, and not because the system is not accurate enough to determine the precise location of letters. Within this particular perspective it is therefore important to ask why the coding of non-contiguous letter combinations might have developed during reading acquisition in order to optimize orthographic processing. To answer this question, we first perform analyses of the mathematical dependencies and regularities in real corpuses of words, and the constraints they place on the process of lexical identification. We then apply these analyses to explain some experimental results. These analyses also take into consideration visual constraints on the quality of the information being processed. Before presenting these analyses, we first describe the overall methodology employed in the present work.

1.1. Optimal letter combinations for word recognition

Here we apply a rational analysis to Grainger and van Heuven's (2003) model of orthographic processing, in order to understand why a biological system learning to read a language with an alphabetical orthography would code for non-contiguous letter combinations. Expressed in terms of Anderson and Milson's (1989) methodology for deriving rational analyses of cognitive processes, we define one key goal of the word recognition system as *the mapping of location-specific letter representations onto location-invariant whole-word orthographic representations*. This, of course, represents one (critical) sub-goal of the overall task of mapping visual information onto meaning. A key characteristic of the environment that constrains the optimization of this mapping is letter-in-string visibility (see section below). Also, inputs available to the reading system vary in informativeness or diagnosticity. The probability of encountering a unit (letter or bigram) is a function of its frequency. Rare (low probability) events are more informative than frequent (i.e., predictable) events (c.f., information theory: Shannon 1948)[1].

We therefore propose three reasons why a biological system learning to read a language with an alphabetical orthography would deliberately code for non-contiguous letter combinations. The first has to do with letter visibility: the most visible letters – the letter at fixation, the first and the last letter of a word – are non-contiguous (Stevens and Grainger 2003), see section below for details. These letters are more easily and reliably identified. The second concerns the positions of letters in a word that are the most informative with respect to lexical identity. We investigate whether a system that tries to minimize letter-level processing while maximizing information would be well advised to code for non-contiguous letter combinations. The third possible reason is that coding for non-contiguous letter combinations is easier (i.e., less complex) in that less precise information about letter positions is required compared with a system that codes for contiguous letter combinations. Put differently, it is clear that relative position information (A is before B) is more reliable when A and B are farther apart. In other words, non-contiguous letter combinations might provide a more robust relative-position code than contiguous letter combinations.

---

[1] It should be noted that the use of diagnostic information is thought to be only one part of the complete set of processes involved in skilled word reading (Grainger and Ziegler 2011).

In order to test this proposal, in what follows we first describe the corpuses of words on which our analyses were based. Second, we provide empirical data concerning letter-in-string visibility. Third, we describe how conditional probabilities are computed under this visibility constraint. We then provide mathematical analyses of the information carried by different types of letter combinations, providing a measure of a bigram's informativeness with respect to a given word, and a ranking of bigrams within a word in terms of their relative informativeness. Finally, we provide empirical tests of these measures by examining to what extent they can capture key patterns of data obtained in orthographic priming experiments.

2. Materials and methods

2.1. Corpuses and word lengths

The current study covers three languages: French, English and Spanish. In all three languages, analyses are based on lemma forms of words composed of letters only (i.e., lemmas that included hyphens, spaces, etc. were excluded) and of lengths between 5 and 7 letters. Words between 5 and 7 letters were selected because visibility data was available for such lengths (see section below), or could be interpolated[2]. As a French lexicon, we used the Lexique (version 3.55) database (New et al. 2004). It contains 125,653 words (including 46,942 lemmas) along with information such as phonology, lexical category and lexical neighborhood. After selection based on the criteria above, 11811 words were chosen. As an English lexicon, we used the Celex database (Baayen, Piepenbrock, and Van Rijn 1993). It contains 50591 lemma forms. After selection based on the criteria above, 18,275 words were selected. As a Spanish lexicon, we used the BuscaPalabras database (Davis and Perea 2005) which is a subset of the Spanish lexicon called LEXESP (Sebastián-Gallés et al. 2000). It contains 31,491 lemmas. After selection based on the criteria above, 11394 words were selected.

2.2. Letter-in-string visibility

An important aspect of our analyses is the use of realistic letter-in-string visibility constraints in the computation of conditional probabilities, along with a comparison with a control condition in which all letters are equally visible. In fact, empirical evidence suggests that not all letters are equally visible when an eye fixation is made on a word. It is possible that two constraints play a role. First, visual acuity is best at the centre of the fovea, and degrades as eccentricity increases. Thus, the letter at fixation is more visible. Furthermore, outer letters of words are more visible, which may be explained by the fact that crowding is reduced for outer letters compared to inner letters (e.g., Grainger, Tydgat, and Isselé 2010; Tydgat and Grainger 2009).

Empirical data on letter-in-string visibility used here are from Stevens and Grainger (2003). For the present theoretical note, we assume that fixations occur on the central position (e.g., letter position number 4 in a string of seven letters), which is a useful first approximation of actual distributions of fixation positions (Rayner 1979). Stevens and Grainger presented uppercase letter targets embedded in strings of uppercase Xs (e.g., XXTXXXX) forming strings of either 5 or 7 letters. These stimuli were presented briefly and were preceded and followed by a pattern masking stimulus. Participants first focused on a central fixation point and then had to identify the letter in the string that was not an X. Figure 2 presents the visibility graph for seven-letter strings. Data also exists for five-letter strings, and, for our analysis, visibility values for 6 letters were interpolated from empirical values for

---

[2] We considered other word lengths to be unwarranted because they would have required an extrapolation of the available visibility data.

5 and 7 letters. It should be noted that similar patterns of letter visibility have been found when strings are composed of a random series of consonants, and post-cueing is used to indicate the location for letter identification (e.g., Tydgat and Grainger 2009).
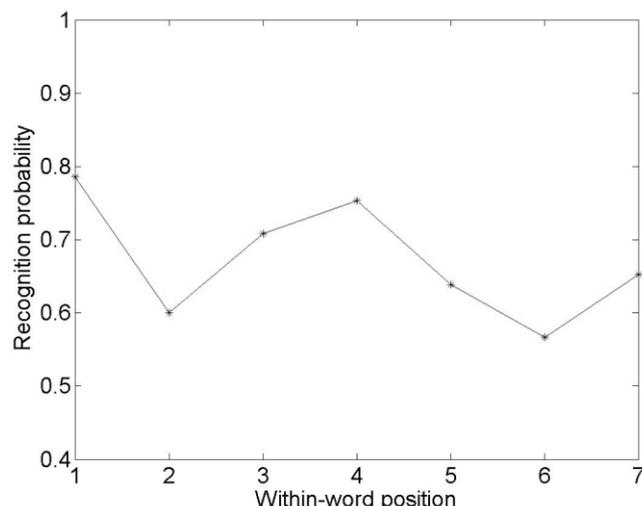

Figure 2 – Letter visibility (recognition probability) as a function of within-string position, for a seven-letter string with fixation on the central position (i.e., the fourth letter) from (Stevens and Grainger 2003). Note that this recognition curve is based on data collected in French.

2.3. Computation of conditional probabilities

Analyses 2 and 3 examine the information carried by different ordered pairs of letters (i.e., bigrams) that can be contiguous or not in the target string. For example, for the word TABLE, bigrams TA, AB, BL and LE are contiguous, whereas TB, TL, TE, AL, AE and BE are non-contiguous. Unless specified otherwise, the analyses consisted of comparing contiguous and non-contiguous bigrams. Note that contiguous and non-contiguous bigrams are collectively called open-bigrams. For the intermediate level of representation between individual letters and whole words described earlier, open-bigrams represent a candidate coding scheme because they abstract away from absolute letter position while preserving information about the relative position of letters. Furthermore, open-bigrams have the advantage of using minimal information about position - only the left-right relationship of the two constituent letters. In contrast, other coding schemes (e.g., Overlap model: Gomez, Ratcliff, and Perea 2008; Spatial coding: Davis 2010) often need some continuous and transitive measure of position (e.g., position of letter 1 < pos 2 < pos 3).

Extending Stevens and Grainger's (2003) method applied to single letter frequencies, we approximated the conditional probability of a word given a bigram. We compare two visibility conditions: perfect visibility and realistic visibility.

Under *perfect visibility*, both letters are optimally visible and the equation is as follows:

p (word | bigram) = freq(word) / ∑ freq (words that contain bigram x number of times bigram occurs in the word)   Equation 1

Under *realistic visibility*, the visibility constraints are taken into account, as follows:

p (word | bigram) = vis_1 * vis_2 * freq(word) / ∑ freq (words that contain bigram x number of times bigram occurs in the word)   Equation 2

where vis_1 and vis_2 are the visibility of letters constituting the bigram, as they appear in the target word. For example, for the word SILENCE, bigram SL contains letter S in position 1

which has a visibility vis_1 of 0.786, and letter L in position 3 has a visibility vis_2 equal to 0.708 (see Figure 2).

3. Results

3.1. Analysis 1: Information values of letter combinations - Minimal substrings to identify words

For a reading system aiming at the compression of data while maximizing the information retained, is it beneficial to code for non-contiguous letter combinations? To answer this question, we identified the shortest ordered (but not necessarily contiguous) letter subset that uniquely identifies words. These ordered letter subsets can be seen as compressing input data by dropping letters that are not critical to the identification of the target word, defining in a sense the optimal abbreviation of a word. For example, the word "fatigue" can be uniquely identified by ordered letter substrings "ftge" and "atge". The analysis consists in counting how many contiguous letter pairs of the substring are formed of contiguous and non-contiguous letters in the baseword. For example, the sequence "atge" from the baseword "fatigue" contains one contiguous letter pair ("at"), and two non-contiguous pairs ("tg" and "ge"). When multiple ordered letter subsets of the same length existed (for example, there were 2 sequences of 4 letters for word "fatigue"), all these sequences were considered in the analysis.

Note that, for analysis 1 only, we further selected words that were not strict letter subsets of other words (ex. "table" is a strict letter subset of "tableau") because, for such words, any substring (including the word itself) also identifies words composed of supersets of letters[3]. Of the total words reported in the "Corpuses and word lengths" section, 85.5% of the French words (N=10129) were selected for analysis under this criterion, as well as 74.9% of the English words (N=13696) and 87.9 % of the Spanish words (N=10012).

We found that the mean percentage of non-contiguous letter pairs in minimal substrings were 48.4%, 47.9% and 47.8% for the words selected from the French, English and Spanish corpuses, respectively. While consonants constitute 55.2%, 62.8% and 53.0% of letters in the French, English and Spanish corpuses, they formed respectively 59.2%, 65.4% and 57.4% of letters in the minimal substrings. Given that consonants occur less frequently than vowels, they carry more information, and so their over-representation is indeed reflected in greater information carried by these elements. Analysis 3a will further explore informativeness of consonants and vowels.

Thus, a system that a) computes letter combinations and b) tries to extract the minimal set of letters that uniquely identifies a given word, should pay attention to non-contiguous letters. In other words, the system would be performing sub-optimally if only contiguous letter combinations were taken into consideration.

3.2. Analysis 2: Conditional probabilities

The second analysis involved comparing how informative non-contiguous bigrams are compared with contiguous bigrams.[4] The higher the conditional probability, the more evidence a bigram is for the word, and therefore the more informative that bigram is. We

---

[3] The subset-superset relationship described is only for selected words of 5 to 7 letters, and not for all words in the language.

[4] We acknowledge that this analysis is biased by the fact that there are more non-contiguous bigrams than contiguous bigrams in a given string. However, from the perspective of demonstrating that non-contiguous bigrams do carry information that should not be ignored, the proportion of contiguous vs. non-contiguous bigrams is irrelevant.

wanted to determine if the most informative bigram for a word tended to be contiguous or not.

Table 1 – Proportion of letter positions in the most informative bigrams under perfect visibility (* indicate non-contiguous bigrams).

Bigram	French			English			Spanish		
	5 letters	6 letters	7 letters	5 letters	6 letters	7 letters	5 letters	6 letters	7 letters
12	10.9%	5.2%	4.1%	9.7%	4.2%	2.6%	7.6%	6.4%	3.1%
13	22.8%	16.2%	11.9%	22.3%	15.8%	10.1%	23.2%	15.5%	11.8%
14	19.4%	18.4%	14.5%	22.7%	20.1%	19.8%	33.7%	12.7%	11.2%
15	10.1%	9.0%	9.6%	14.8%	8.4%	9.0%	4.5%	17.0%	6.2%
16	N/A	6.3%	5.8%	N/A	11.9%	4.3%	N/A	2.3%	9.0%
17	N/A	N/A	3.6%	N/A	N/A	10.1%	N/A	N/A	1.2%
23	8.5%	6.5%	3.7%	4.9%	4.5%	3.3%	5.3%	8.5%	7.8%
24	8.0%	7.3%	5.3%	6.5%	5.4%	4.8%	8.7%	6.9%	7.4%
25	4.9%	5.0%	6.5%	6.5%	4.3%	3.9%	1.9%	6.4%	4.0%
26	N/A	3.0%	3.0%	N/A	3.4%	1.9%	N/A	0.7%	4.6%
27	N/A	N/A	1.6%	N/A	N/A	2.7%	N/A	N/A	0.3%
34	6.4%	5.7%	3.7%	3.4%	3.7%	3.5%	9.5%	3.3%	5.3%
35	6.4%	5.1%	5.7%	5.9%	4.0%	3.6%	4.2%	11.1%	5.0%
36	N/A	4.5%	4.4%	N/A	4.2%	2.2%	N/A	1.5%	9.8%
37	N/A	N/A	3.5%	N/A	N/A	2.5%	N/A	N/A	1.1%
45	2.6%	3.0%	2.0%	3.4%	2.1%	3.1%	1.4%	4.3%	0.9%
46	N/A	3.5%	4.0%	N/A	6.7%	3.1%	N/A	2.6%	6.0%
47	N/A	N/A	3.8%	N/A	N/A	4.6%	N/A	N/A	0.8%
56	N/A	1.3%	1.0%	N/A	1.3%	0.9%	N/A	0.9%	3.0%
57	N/A	N/A	1.8%	N/A	N/A	3.4%	N/A	N/A	1.4%
67	N/A	N/A	0.5%	N/A	N/A	0.5%	N/A	N/A	0.1%
Proportion of non-contiguous bigrams									
	71.6%	78.3%	85.0%	78.7%	84.2%	86.1%	76.3%	76.7%	79.9%

We computed the conditional probability of words given bigrams, p (word | bigram) and we examined the distribution of the positions occupied by the letters that composed the most informative bigrams. For this analysis, we computed conditional probabilities based on all words, but we selected only the words without any repeated letters as test words. In this way we could unequivocally assign a letter in a bigram to a precise position in the word, and therefore know whether the bigram in question is contiguous or not. The number of words included in the test set were 5838 in French, 8412 in English, and 4750 in Spanish. Proportion of letters at each position in words of a given length (5-7 letters) that appear in the most informative bigram of words are provided for French, English, and Spanish (see Table 1 for the perfect visibility condition, and Table 2 for the realistic visibility condition). The results show that the 1st letter of words occurs more often in the most informative bigram than any other letter position, while the last letter occurs the least often. More important for the present purposes is that, after initial letters, it is letters located near the centre or the left-of-centre of words that occur most often in the most informative bigram, such that the non-contiguous bigrams 13 and 14 are most often the most informative bigram in words of lengths 5-7 letters. This result follows logically from the fact that letter frequency typically shows a serial position function with the lowest value for initial letters, followed by positions 3 and 4 (e.g., Grainger and Jacobs 1993), such that it is a combination of the positions with

the lowest letter frequencies that carry the most information. Overall, this analysis shows that the most informative bigrams in words are non-contiguous bigrams (on average 78% in French, 83% in English, and 78% in Spanish), and adding a visibility weighting does not change this pattern much.

Table 2- Proportion of letter positions in the most informative bigrams under realistic visibility (* indicate non-contiguous bigrams).

Bigram	French			English			Spanish		
	5 letters	6 letters	7 letters	5 letters	6 letters	7 letters	5 letters	6 letters	7 letters
12	11.7%	7.0%	4.4%	10.5%	6.7%	2.3%	7.8%	8.2%	3.2%
13*	25.8%	20.2%	13.8%	25.7%	20.8%	12.0%	25.5%	19.9%	15.1%
14*	19.8%	19.0%	18.1%	23.5%	21.0%	23.7%	34.5%	12.9%	13.9%
15*	10.0%	9.1%	10.1%	14.8%	8.2%	9.5%	4.7%	15.9%	6.6%
16*	N/A	6.2%	5.8%	N/A	11.0%	3.9%	N/A	2.5%	8.4%
17*	N/A	N/A	4.1%	N/A	N/A	10.9%	N/A	N/A	1.4%
23	8.1%	6.4%	2.8%	4.5%	4.8%	2.8%	4.3%	9.0%	6.5%
24*	6.8%	6.0%	4.7%	5.3%	4.7%	4.1%	7.6%	5.9%	7.0%
25*	4.5%	4.0%	4.9%	5.7%	3.5%	2.8%	1.8%	5.5%	2.5%
26*	N/A	2.8%	2.0%	N/A	2.8%	1.2%	N/A	0.6%	3.7%
27*	N/A	N/A	1.2%	N/A	N/A	1.8%	N/A	N/A	0.3%
34	5.9%	5.2%	4.0%	3.0%	3.3%	3.9%	9.0%	3.0%	7.1%
35*	5.3%	4.5%	5.5%	4.5%	3.2%	3.3%	3.8%	9.7%	4.2%
36*	N/A	4.0%	3.6%	N/A	3.1%	1.6%	N/A	1.6%	8.0%
37*	N/A	N/A	3.3%	N/A	N/A	2.2%	N/A	N/A	1.0%
45	2.1%	2.3%	2.3%	2.4%	1.4%	3.1%	1.0%	3.1%	1.6%
46*	N/A	2.4%	3.7%	N/A	4.5%	2.3%	N/A	1.7%	5.1%
47*	N/A	N/A	3.8%	N/A	N/A	4.6%	N/A	N/A	0.9%
56	N/A	0.8%	0.6%	N/A	0.8%	0.4%	N/A	0.5%	2.2%
57*	N/A	N/A	1.3%	N/A	N/A	3.3%	N/A	N/A	1.1%
67	N/A	N/A	0.2%	N/A	N/A	0.3%	N/A	N/A	0.1%
Proportion of non-contiguous bigrams									
	72.3%	78.2%	85.6%	79.6%	83.0%	87.2%	77.9%	76.2%	79.3%

3.3. Analysis 3: Orthographic priming

In the following analyses, we investigated how conditional probabilities of letter combinations could explain some key patterns of orthographic priming effects, and in particular (a) effects of consonant vs. vowel status of letters shared by primes and targets, and (b) effects of position of letters shared by prime and target. Note that these effects arise over and above the effects driven by amount of orthographic overlap between primes and targets. The hypothesis guiding our investigation is that the more informative the primes are with respect to target identity (as measured by conditional probabilities), the larger the priming effects will be. We performed two analyses based on bigram conditional probabilities: the first one using the average of conditional probabilities and the second one using the within-word ranking of these probabilities. The conditional probabilities are calculated for primes and targets tested in experiments with human participants, and the values compared with the priming effects found in the experiments.

As a step common to both analyses, we first computed the conditional probability of each word in some realistic corpus given their constituent bigrams (as in Analysis 2), and then ranked bigrams by decreasing conditional probability (see example below taken from an experiment in Spanish). Table 3 presents an example for the target word DOPAJE.

Table 3 – Example of ranked conditional probabilities for prime DPJ and target word DOPAJE. *** identify bigrams that the target and the prime share.

Rank	Bigram	Conditional Probability
1	PJ	0.00666***
2	DJ	0.00530***
3	OJ	0.00408
4	DP	0.00350***
5	OP	0.00135
6	AJ	0.000941
7	JE	0.000920
8	DA	0.000192
9	OE	0.000166
10	OA	0.000158
11	PE	0.000147
12	DE	0.000146
13	DO	0.000129
14	PA	0.000127
15	AE	0.000126

Next, for each word in the testing material (e.g., prime = DPJ & target =DOPAJE), we identified the bigrams from the prime (e.g., made of letters 1-2, 2-3 and 1-3 for prime DPJ, identified using *** in Table 3), some of which could be repeated (e.g., 1 x TT & 2 x TL for prime TTL, and target word TÍTULO).

For the first analysis, we simply took the average of the conditional probabilities of the target word given the bigrams in the prime. In the example above, we obtained conditional probabilities of 0.0067 for PJ, 0.0053 for DJ, and 0.0035 for DP, and thus an average of $(0.0067 + 0.0053 + 0.0035) / 3 = 0.0052$.

For the second analysis based on within-word ranking, we sought what ranks these prime bigrams had in the target word, and computed informativeness as the inverse of rank (1/rank). We computed an overall informativeness measure for the prime by averaging the three individual values of inverse ranks. The inverse of ranks measure is indicative of priming: the lower the rank, the more informative the prime. In the example above, PJ has a rank of 1, DJ has a rank of 2, and DP has a rank of 4, which yields an informativeness measure of $1/3 * (1/1 + 1/2 + 1/4) = 0.58$.

3.3.1. Analysis 3a: Letter type (consonants vs. vowels)

Consonants and vowels are known to play different roles in language processing, as shown for example in the processing of continuous speech (Bonatti et al. 2005). Differences between consonants and vowels might stem from the different lexical constraints they impose; for instance, vowel and consonants may activate different numbers of lexical candidates (Carreiras, Duñabeitia, and Molinaro 2009). Consonants tend to be more indicative of word identity than vowels are, as we have seen in Analysis 1. This advantage for consonants can be attributed, at least in part, to differences in frequencies. Vowels being

more frequent, they are thus less informative. However, Duñabeitia and Carreiras (2011 submitted) have recently shown that even when vowels and consonants are matched for frequency and various measures of lexical constraint, consonant primes are still more effective than vowel primes. They concluded that consonants and vowels have distinct properties over and above frequency.

More specifically, Duñabeitia and Carreiras (2011 submitted) tested relative-position priming using the masked priming paradigm with a lexical decision task. Relative-position priming constitutes a specific form of subset priming in which a masked prime is made of some of the letters of the target word, preserving their relative ordering within the string (e.g., CSN from CASINO; see Grainger et al. 2006). In one of their experiments, targets were 6-letter Spanish words, while the 3-letter primes were either related to the target (e.g., prime = DPJ & target = DOPAJE) or unrelated (e.g., prime = MBZ & target = DOPAJE). The latter condition acted as a control to establish baseline response times for the targets (see Figure 4). Critically, the three letters of the related primes could be formed uniquely of vowels, high-frequency consonants, or low-frequency consonants (according to the number of appearances of each letter in the Spanish lexicon). The three critical letters that constituted the related primes were in all cases the first, third and fifth letters of the targets. In further detail, the letters selected for the low-frequency consonant set were [b,d,j,m,v,z,g,f,p], which altogether summed 17.16% of letter occurrences in Spanish (mean frequency in percentage of occurrences: 1.91%). The letters that formed the high-frequency consonant subset were [t,n,c,r,s,l], summing 35.68% of letter occurrences (mean frequency in percentage of occurrences: 5.95%). The vowels subset accounted for 45.63% of the letter-frequency distribution (mean frequency in percentage of occurrences: 9.13%). These figures are reflected in the conditional probabilities computed on the basis of letters (see Figure 3): conditional probabilities of high-frequency consonants and of vowels are relatively well-matched when calculated at the level of individual letters. In their study, Duñabeitia and Carreiras showed that masked relative-position priming is found for primes consisting of consonants of high or low frequency, but that vowel relative-position masked primes did not represent any processing benefit as compared to the corresponding controls. According to what was also shown in other experiments reported in that study (see also Carreiras, Duñabeitia, and Molinaro 2009), vowels did not yield significant relative-position priming effects, while both high and low-frequency consonants led to reliable priming effects that were comparable in magnitude. Hence, the authors concluded that frequency of the letters is not the underlying factor that determines whether relative-position priming is found. Rather, Duñabeitia and Carreiras hypothesized that the consonant-vowel processing difference relies on the main role of each of these letters. While the role of consonants is related to lexical access by constraining the lexicon, the main role of vowels is related to the identification of properties of the syntactic structure and the rhythmic class (see also Bonatti et al. 2005; Mehler et al. 2006; Pons and Toro 2010). Across many languages, consonants are more numerous than vowels, and consequently consonant combinations are by default less frequent than vowel combinations, leading to a higher lexical constraint imposed by consonants as compared to vowels.
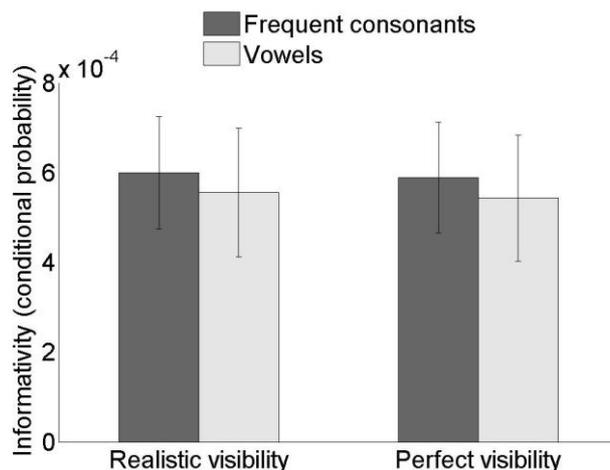
Figure 3 - Conditional probabilities of the frequent consonant and vowel stimuli used in Duñabeitia and Carreiras (2011 submitted) calculated using the method described by Stevens and Grainger (2003) for individual position-coded letters.

In the present theoretical note, we focused on the key contrast between high-frequency consonant and vowel primes. Empirical results in human participants are presented in Figure 4. We tested the hypothesis that, despite the fact that the materials used by Duñabeitia and Carreiras (2011 submitted) were well-matched in terms of individual letter frequencies (as well as for lexical frequency and number of orthographic neighbors of the target words), the two critical prime conditions (high-frequency consonants and vowels) were not equivalent when taking into account the information carried by letter combinations. Our hypothesis was that observed results could be explained, at least in part, by the fact that informativeness of high-frequency consonants was significantly greater than informativeness of vowels when calculated at the level of letter combinations.
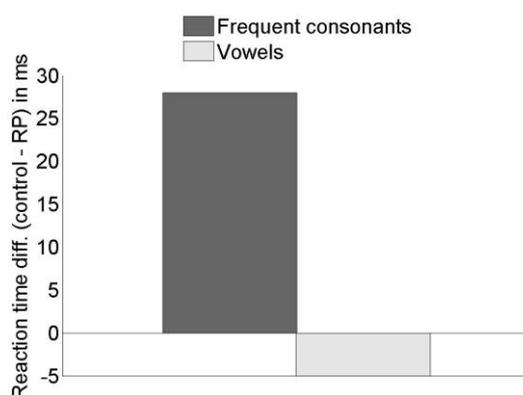

Figure 4 - Empirical results in human participants for the effect of consonants and vowels (Duñabeitia and Carreiras, (2011 submitted).

Using the primes and targets of the Duñabeitia and Carreiras (2011 submitted) study, we computed average conditional probabilities of the target words given the open-bigrams (contiguous and non-contiguous) of their respective 3-letter primes. These informativeness values are shown in Figure 5. As we can see, the pattern of informativeness measures is

14

consistent with human response times: more information in high-frequency consonant primes results in greater priming effects.
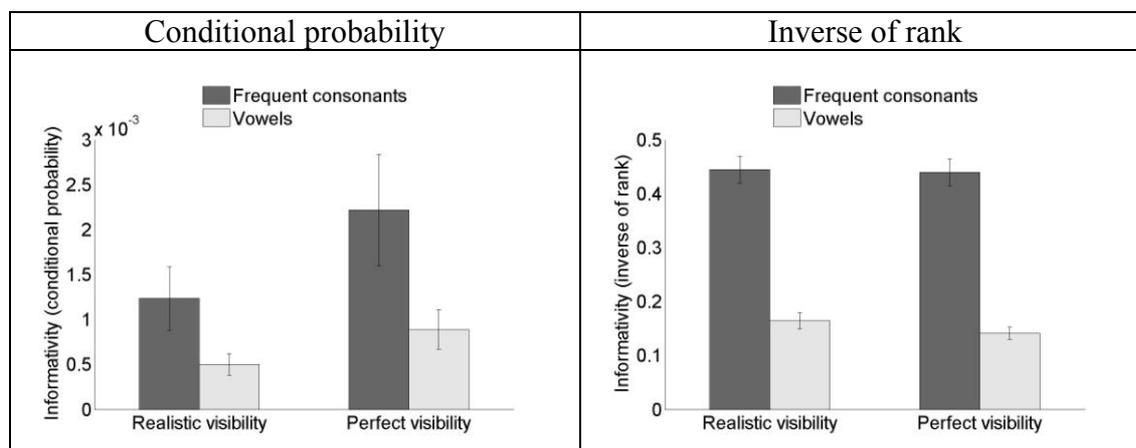

Figure 5 - Effect of consonants and vowels in bigram-based analyses. Left: informativeness calculated using an average conditional probability measure. Right: informativeness calculated using the inverse of within-word ranking of conditional probabilities.

We next tested the results of the two informativeness measures (inverse of rank measure and average conditional probability) using two-way mixed ANOVAs with visibility as a repeated factor with 2 levels (realistic and perfect) and letter type as an independent factor with 2 levels (vowels and frequent consonants). Both measures of informativeness were log-transformed prior to the analyses.

For the ranks-based measure of informativeness, we found a main effect of letter type ($F(1,94) = 111$, $p < 0.001$), a main effect of visibility ($F(1,94) = 29$, $p < 0.001$), and an interaction ($F(1,94) = 16$, $p < 0.001$). A similar pattern was observed for the average conditional probability measure of informativeness: a main effect of letter type ($F(1,94) = 16$, $p < 0.001$), a main effect of visibility ($F(1,94) = 5348$, $p < 0.001$), and an interaction ($F(1,94) = 27$, $p < 0.001$). The pattern of results was consistent for both realistic and perfect visibilities: informativeness was higher for consonant primes than vowel primes when calculated using both the rank-based measure and the average conditional probability measure.

The explanation of Duñabeitia et al.'s data, which was based on the lexical constraint imposed by relative-position consonantal masked primes as compared to the constraint imposed by vowel primes, can be effectively modeled in terms of the informativeness of the bigrams that formed the primes. Although frequencies of letters were effectively matched as measured by their frequency of occurrence in the Spanish corpus, high-frequency consonants still had an advantage over vowels when considered at the level of letter combinations. This effect may be partly due to the within-word ranks of letter positions of the letters that form the bigrams, as evidenced by the control condition. Recall that the primes, composed of letters 1, 3 and 5, corresponded to two categories (letter 1 is the first letter of the word, whereas letters 3 and 5 are inner letters). Since most of the letters of the word consist of inner letters, it is possible that the first letter of the word drives the difference. Namely, the fact that consonants tend to be more often among the most informative letters for the word than vowels at the first position. Thus, this analysis based on combinatorial probabilities of open-bigrams provides computational confirmation of the idea suggested by Bonatti, Peña, Nespor, and Mehler (2005) and Duñabeitia and Carreiras (2011 submitted), among others, who

claimed for a higher lexical information value of consonant combinations as opposed to vowel combinations.

3.3.2. Analysis 3b: Effect of the central letter on priming

We next investigated if bigram conditional probabilities could explain some experimental results on relative-position priming. In an experiment involving masked priming of 7-letter French words using 5-letter primes, Granier and Grainger (unpublished data) investigated the effect of primes on response times in a lexical decision task. Primes were selected such that they either included the central letter of the target word or did not. Granier and Grainger found better priming for primes that contained the central letter compared to those that did not. Our hypothesis is that better priming can be explained by larger informativeness when primes include the central letter. More specifically, Granier and Grainger (unpublished data) tested sixty 7-letter target words in two sub-experiments each containing 4 prime conditions. For one sub-experiment the four priming conditions were: 13457, 12457, 13467, ddddd; and for the other sub-experiment: 13457, 12367, 12567, ddddd. Numbers indicate the positions (out of 7 letters) in targets of the letters selected to form the primes, and "d" stands for a letter that is not in the target. For example, for the word SILENCE, the prime 12367 would be formed by the letters SILCE. Prime-target pairings were counterbalanced across four lists associated with four independent groups of participants in each sub-experiment. 52 participants were tested in the first sub-experiment, and 56 in the other sub-experiment. Standard masked priming procedure was used with the lexical decision task and 50 ms prime durations (see Grainger et al. 2006, Experiment 1, for a description of the procedure that was used). The results revealed a dissociation between priming effects obtained with primes containing the central letter of the target (primes 13457, 12457, 13467) compared with primes that did not contain the central letter of the target (primes 12367, 12567). Priming effects measured against the unrelated prime condition (ddddd) were significant for the former and non-significant for the latter.

In order to test our account of this central letter effect, we computed average conditional probabilities of target words given the open-bigrams (contiguous and non-contiguous) of their respective 5-letter primes. Because the present study focuses on the presence of the central letter, data were aggregated over conditions containing the central letter (that is, primes 13457, 12457, 13467) and conditions that did not (primes 12367, 12567). Empirical results in human participants are shown in Figure 6, and results of the informativeness analysis in Figure 7.
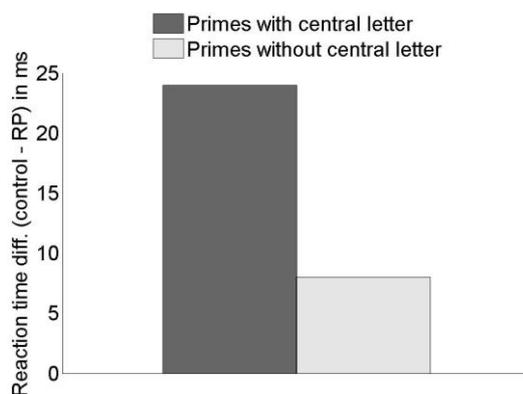

Figure 6 - Empirical results (priming effect sizes measured as RT differences) in human participants for the effect of central letters on priming.
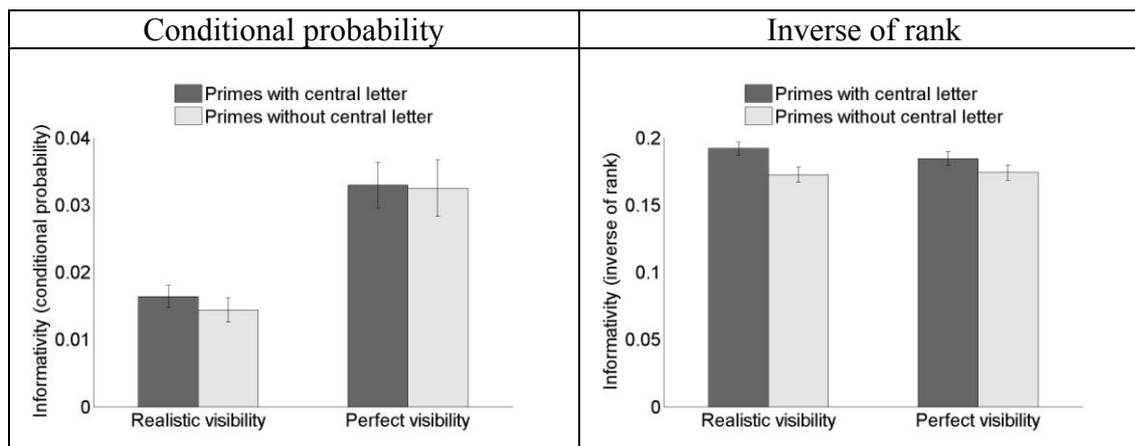
Figure 7 - Effect of central letter on priming in bigram-based analyses. Left: informativeness calculated using an average conditional probability measure. Right: informativeness calculated using the inverse of within-word ranking.

Next, for the bigram-based measure, we performed two-way mixed ANOVAs of the two measures of informativeness (inverse of within-word ranking, and global conditional probability) with visibility as a repeated factor with 2 levels (realistic and perfect) and the presence of the central letter as an independent factor with 2 levels (with and without). Both measures of informativeness were log-transformed prior to the analyses.

The analysis of the inverse ranking measure revealed a trending main effect of the presence of the central letter ($F(1,298) = 3.8$, $p = 0.052$), a strong effect of visibility ($F(1,298) = 21$, $p < 0.001$), and an interaction ($F(1,298) = 50$, $p < 0.001$). Rank-based informativeness was higher for primes that included the central letters compare to those that did not, and the interaction shows that the advantage of primes with the central letter was larger in the realistic visibility than the perfect visibility condition. Guided by the interaction, we verified that presence of the central letter had indeed a significant impact in the perfect visibility condition only ($F(1,299) = 6.6$, $p = 0.011$). Thus, results with realistic visibility are consistent with experimental results on priming in humans: given an equal number of letters as primes, those that contain the central letter are more effective.

For the informativeness measure based on average conditional probabilities, the ANOVA revealed a strong effect of visibility ($F(1,298) > 10000$, $p < 0.001$), an interaction ($F(1,298) = 180$, $p < 0.001$), but no effect of the presence of the central letter ($F(1,298) < 1$). In sum, average conditional probabilities of bigrams do not explain the effect of letter centrality, but the inverse ranking measure does. The within-word ranking measure of informativeness appears to be more sensitive than the conditional probability measure, an effect that was already observed in the size of the effects reported in Figure 5. This suggests that the relative informativeness of each bigram in the prime stimulus with respect to the other bigrams of the target word might be a better measure than the average measure of informativeness of letter combinations in the entire corpus (average conditional probability). In other words, the stimulus-specific relative information is perhaps more critical than the absolute measure of informativeness.

4. Discussion

In this theoretical note, we investigated the nature of the orthographic code that subserves the parallel mapping of location-specific letter identities onto location-invariant

lexical representations. We hypothesized that some form of intermediate orthographic code is necessary, and that this code involves subsets of letter identities and information about their relative positions. We then examined the kind of intermediate orthographic code that would enable skilled readers to map letters onto words as efficiently as possible by optimizing utilization of the available information while minimizing resources. More precisely, we asked whether the constraints of optimality would lead to the coding of non-contiguous letter combinations as well as contiguous ones.

We examined two kinds of constraints that an optimal reader should take into consideration when processing orthographic information. The first involved variations in letter visibility across the different letters of a word during a single fixation. The second kind of constraint concerned the varying amount of information carried by the different letters in the word. We hypothesized that a system that aims to optimize orthographic processing, that is, to map letter identities onto whole-word orthographic representations as efficiently as possible, will adapt to such variations in visibility and informativeness. More specifically, we hypothesized that this optimization would involve coding of non-contiguous letter combinations, under two assumptions: i) that letter combinations need to be computed, and ii) that optimal letter combinations are not necessarily contiguous. Furthermore, we pointed out that in principle it should be easier to code for non-contiguous letter combinations, since order information is more reliable (at the cost of being less precise) in this case than when letters are contiguous.[5]

Our corpus analyses suggest that, indeed, optimal readers should code for non-contiguous letters. First, when selecting an ordered subset of letters which are critical to the identification of a word (that is, dropping non-essential letters that bear little information), about half of the letters in the resulting subset are non-contiguous. Second, the most informative pair of letters in a word is a non-contiguous combination in 83% of 5-7 letter words (having no letter repetition) in English, and 78% in French and Spanish. In the second part of the paper, we saw that the superiority of the central letter in relative-position priming can be explained by rational analyses based on maximizing conditional probabilities of words given bigrams. Also, we reanalyzed proposed evidence that consonants and vowels play a different role during word recognition. We re-examined the finding that consonants form better primes in masked relative-position priming even when matched for letter frequencies with vowels. When considering consonants and vowels not individually, but as letter combinations (bigrams in the present analyses), we can account for the superior priming effect of consonants over vowels using a parsimonious explanation based on informativeness, without having to appeal to qualitatively distinct representations for these two types of letter. This is in line with the proposal of Mehler and colleagues (Bonatti et al. 2005, Bonatti et al. 2007; Mehler et al. 2006; Nespor, Peña, and Mehler 2003; Pons and Toro 2010; Toro et al. 2008) that consonants are lexically more constraining than vowels, and therefore more important for word identification (see also Carreiras, Duñabeitia, and Molinaro 2009).

These results suggest that an optimal or rational agent learning to read corpuses of real words should deliberately code for non-contiguous letters based on informational content, and given visibility constraints. Thus, the optimality or rationality argument is

---

[5] In the more general framework for visual word recognition developed by Grainger and Holcomb (2009) and Grainger and Ziegler (2011; see also Grainger and Dufau 2011), the coding of non-contiguous letter combinations is used in one of the two routes of a dual-route approach to orthographic processing. As argued in the present work, this particular route is thought to optimize processing by selecting the most informative letter combinations in order to get from print to meaning as efficiently as possible. The other route is thought to optimize processing by performing data compression via the chunking of frequently co-occurring contiguous letters such as multi-letter graphemes and affixes. This route is therefore thought to be involved in getting from print to meaning via intermediate phonological and morphological representations.

sufficient to explain the coding of non-contiguous letters. However, this does not preclude some contribution of biological constraints in readers, such as noise and inaccuracies, in such encoding.

In the present work we compared two measures of bigram informativeness with respect to their ability to account for orthographic priming data. Both measures used the conditional probability of bigrams in a given prime stimulus, with one measure taking the average of the conditional probabilities for a given prime, and the other taking the average rank value of each bigram in the prime stimulus among the complete set of bigrams in the target stimulus. Only the rank values successfully predicted the complete set of empirical findings, and rank values were generally more sensitive than conditional probabilities. So, given a prime stimulus TBL for the target TABLE, it would appear that priming effects depend more on how constraining the letters shared by prime and target are with respect to the other bigrams in the target word (TA, TE, AB, AL, AE, BE, LE) than on how the prime's bigrams (TB, TL, BL) constrain target word identity (i.e., their average conditional probabilities). This important finding suggests that during the process of learning to read, the most informative bigrams are "selected" or at least given priority over less informative bigrams, and the manner in which this priority is accorded has a long-lasting influence on the speed with which this information is processed. Furthermore, this finding explains why, in prior unpublished research from our laboratory, manipulations of lexical constraint failed to modulate the size of orthographic priming effects, since the manipulation of lexical constraint was based on average conditional probabilities.

The finding that the informativeness ranking of open-bigrams was a superior predictor of word identity than the average informativeness of open-bigrams can be linked to the work on diagnosticity and visual object recognition (F Gosselin and Schyns 2001). In this work, participants are typically asked to classify visual images as belonging to one of a possible set of object identities, and one examines how classification performance varies as a function of the nature of the information that is made available using the so-called "bubbles" technique. For example, Fiset et al. (2008), using this technique, showed that on average only 32% of the printed area of uppercase and 24% of lowercase letters was used by observers to identify letters. Furthermore, their analysis revealed that terminations were by far the most diagnostic piece of information for letter identification, with intersections and horizontal lines providing further significant sources of information for uppercase letters. For example, the letter W was mainly distinguished from other letters by the presence of two terminations, one in the upper left corner and the other in the upper right corner, an example of a highly informative non-contiguous feature combination.

Finally, taking into consideration estimates of letter visibility had a significant impact on our ability to capture the pattern of priming effects when letter position varied (outer and central letter manipulations), but not when letter type varied (consonant vs. vowel manipulation). This is a logical result in that only changes in letter position cause a change in letter visibility, and the position of letters was matched for the consonant and vowel primes (across different targets) in the work of Duñabeitia and Carreiras. However, we acknowledge here that our letter visibility values were obtained with French native speakers, and there is evidence that such letter visibility functions can differ as a function of language. More precisely, there is evidence that languages with a more transparent relation between spelling and sound show a less pronounced advantage for the final position in the string (e.g., Ktori and Pitchford 2008, for Greek). Therefore future work will need to examine whether such variations in letter visibility across different languages, combined with a measure of informativeness such as applied in the present work, could be usefully applied to examine possible differences in orthographic priming effects as a function of the language under study.

Summing up, our analyses of the information carried by orthographic prime stimuli in different orthographic priming conditions (matched in terms of the number of letters shared by prime and target) revealed one measure that captured the empirical data patterns. This is the rank of the conditional probabilities of the bigrams shared by prime and target among the complete set of target bigrams. This key finding suggests that further exploration of the role of letter combinations, and in particular non-contiguous letter combinations, in optimizing the extraction of meaning from print during skilled reading, is in order.

Acknowledgments

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Anderson, J. R. 1990. *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
Anderson, J. R., and R. Milson. 1989. "Human memory: an adaptive perspective." *Psychological Review* 96: 703–719.
Baayen, H. R, R. Piepenbrock, and H. Van Rijn. 1993. "The CELEX lexical database—Dutch, English, German." *Philadelphia, PA: Linguistics Data Consortium*.
Bonatti, L. L., M. Peña, M. Nespor, and J. Mehler. 2005. "Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing." *Psychological Science* 16 (6): 451-459.
———. 2007. "On consonants, vowels, chickens, and eggs." *Psychological Science* 18 (10): 924-925.
Carreiras, M., J. A. Duñabeitia, and N. Molinaro. 2009. "Consonants and vowels contribute differently to visual word recognition: ERPs of relative position priming." *Cerebral Cortex* 19: 2659-2670.
Carreiras, M., J. A. Dunabeitia, and M. Perea. 2007. "READING WORDS, NUMB3R5 and \$YMβOL\$." *Trends in Cognitive Sciences* 11: 454–455.
Chater, N., and Gordon D. A. Brown. 2008. "From universal laws of cognition to specific cognitive models." *Cognitive Science* 32 (1): 36-67.
Chater, N., and Paul Vitányi. 2003. "Simplicity: a unifying principle in cognitive science?" *Trends in Cognitive Sciences* 7 (1): 19-22.
Chauncey, K., P. J. Holcomb, and J. Grainger. 2008. "Effects of stimulus font and size on masked repetition priming: An ERP investigation." *Language and Cognitive Processes* 23 (1): 183-200.

Davis, C. J. 2010. "The spatial coding model of visual word identification." *Psychological Review* 117 (3): 713–758.

Davis, C. J., and M. Perea. 2005. "BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish." *Behavior Research Methods* 37 (4): 665-671.

Dehaene, S. 2007. *Les neurones de la lecture*. Odile Jacob.

Dehaene, S., L. Cohen, M. Sigman, and F. Vinckier. 2005. "The neural code for written words: a proposal." *Trends in Cognitive Sciences* 9 (7): 335-341.

Dehaene, S., K. Nakamura, A. Jobert, C. Kuroki, S. Ogawa, and L. Cohen. 2010. "Why do children make mirror errors in reading? Neural correlates of mirror invariance in the visual word form area." *Neuroimage* 49 (2): 1837-1848.

Dufau, S., J. Grainger, and P. J. Holcomb. 2008. "An ERP investigation of location invariance in masked repetition priming." *Cognitive Affective and Behavioral Neuroscience* 8: 222-228.

Duñabeitia, J. A., and M. Carreiras. 2011. "(accepted) The relative position priming effect depends on whether letters are vowels or consonants." *Journal of Experimental Psychology: Learning, Memory and Cognition*.

Feldman, J. 2003. "The simplicity principle in human concept learning." *Current directions in psychological science* 12 (6): 227-239.

Fiset, Daniel, Caroline Blais, Catherine Ethier-Majcher, Martin Arguin, Daniel Bub, and Frédéric Gosselin. 2008. "Features for identification of uppercase and lowercase letters." *Psychological Science* 19 (11): 1161-1168. doi:10.1111/j.1467-9280.2008.02218.x.

Freyd, J., and B. Tversky. 1984. "Force of symmetry in form perception." *American Journal of Psychology* 97 (1): 109-126.

Gomez, P., R. Ratcliff, and M. Perea. 2008. "The overlap model: A model of letter position coding." *Psychological Review* 115: 577-601.

Gosselin, F, and P G Schyns. 2001. "Bubbles: a technique to reveal the use of information in recognition tasks." *Vision Research* 41 (17) (August): 2261-2271.

Grainger, J. 2008. "Cracking the orthographic code: An introduction." *Language and Cognitive Processes* 23 (1): 1-35.

Grainger, J., and S. Dufau. 2011. "(in press) The front end of visual word recognition." Ed. J. S. Adelman. *Visual Word Recognition* 2: 54.

Grainger, J., J.P. Granier, F. Farioli, E. Van Assche, and W. J. B. van Heuven. 2006. "Letter position information and printed word perception: The relative-position priming constraint." *Journal of Experimental Psychology: Human Perception and Performance* 32 (4): 865-884.

Grainger, J., and W. J. B. van Heuven. 2003. Modeling letter position coding in printed word perception. In *The Mental lexicon*, ed. P. Bonin, 1-23. New York: Nova Science Publishers.

Grainger, J., and P. J Holcomb. 2009. "An ERP investigation of orthographic priming with relative-position and absolute-position primes." *Brain research* 1270: 45–53.

Grainger, J., and A. M. Jacobs. 1993. "Masked partial-word priming in visual word recognition: Effects of positional letter frequency." *Journal of Experimental Psychology: Human Perception & Performance* 19: 951-964.

Grainger, J., I. Tydgat, and J. Isselé. 2010. "Crowding affects letters and symbols differently." *Journal of Experimental Psychology: Human Perception and Performance* 36 (3): 673-688.

Grainger, J., and J. Ziegler. 2011. "A dual-route approach to orthographic processing." *Frontiers in Psychology* 2: 54. doi:10.3389/fpsyg.2011.00054.

Ktori, M., and N. J. Pitchford. 2008. "Effect of orthographic transparency on letter position encoding: A comparison of Greek and English monoscriptal and biscriptal readers." *Language and Cognitive Processes* 23 (2): 258–281.

Logothetis, N. K., and J. Pauls. 1995. "Psychophysical and physiological evidence for viewer-centered object representations in the primate." *Cerebral Cortex* 5 (3): 270-288.

Mehler, J., M. Peña, M. Nespor, and L. L. Bonatti. 2006. "The 'soul' of language does not use statistics: reflections on vowels and consonants." *Cortex* 42 (6): 846-854.

Mozer, M. C. 1987. Early parallel processing in reading: A connectionist approach. In *Attention and Performance XII*, ed. M. Coltheart. Hillsdale, NJ: Erlbaum.

Nespor, M., M. Peña, and J. Mehler. 2003. "On the different roles of vowels and consonants in speech processing and language acquisition." *Lingue e Linguaggio* 2: 221–247.

New, B., C. Pallier, M. Brysbaert, and L. Ferrand. 2004. "Lexique 2: A new French lexical database." *Behavior Research Methods, Instruments, and Computers* 36 (3): 516-524.

Oaksford, M., and N. Chater. 1998a. A revised rational analysis of the selection task: Exceptions and sequential sampling. In *Rational models of cognition*, 372-398. Oxford: Oxford University Press.

———. 1998b. *Rational models of cognition*. New York: Oxford University Press.

Pelli, Denis G., Bart Farell, and Deborah C. Moore. 2003. "The remarkable inefficiency of word recognition." *Nature* 423 (6941) (June 12): 752-756. doi:10.1038/nature01516.

Perea, M., and S. J. Lupker. 2004. "Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions." *Journal of Memory & Language* 51 (2): 231-246.

Peressotti, F., and J. Grainger. 1999. "The role of letter identity and letter position in orthographic priming." *Perception & Psychophysics* 61: 691-706.

Petit, J. P., K. J. Midgley, P. J. Holcomb, and J. Grainger. 2006. "On the time course of letter perception: A masked priming ERP investigation." *Psychonomic Bulletin and Review* 13 (4): 674-681.

Pons, F., and J. M Toro. 2010. "Structural generalizations over consonants and vowels in 11-month-old infants." *Cognition* 116: 361-367.

Rayner, K. 1979. "Eye guidance in reading: Fixation locations within words." *Perception* 8: 21-30.

Schoonbaert, S., and J. Grainger. 2004. "Letter position coding in printed word perception: Effects of repeated and transposed letters." *Language and Cognitive Processes* 19: 333-367.

Sebastián-Gallés, N., M. A. Martí, F. Cuetos, and M. Carreiras. 2000. *LEXESP: Léxico informatizado del español [LEXESP: A computerized word-pool in Spanish]*. Barcelona: Edicions de la Universitat de Barcelona.

Shannon, C. E. 1948. "A mathematical theory of communication." *The Bell System Technical Journal* 27: 379-423.

Slowiaczek, M. L, and K. Rayner. 1987. "Sequential masking during eye fixations in reading." *Bulletin of the Psychonomic Society* 25 (3): 175–178.

Stevens, M., and J. Grainger. 2003. "Letter visibility and the viewing position effect in visual word recognition." *Perception & Psychophysics* 65 (1): 133-151.

Tootell, R. B. H., N. K. Hadjikhani, J. D. Mendola, S. Marrett, and A. M. Dale. 1998. "From retinotopy to recognition: fMRI in human visual cortex." *Trends in Cognitive Sciences* 1 (2): 174-183.

Toro, J. M., M. Nespor, J. Mehler, and L. L. Bonatti. 2008. "Finding words and rules in a speech stream: functional differences between vowels and consonants." *Psychological Science* 19: 137-144.

Tydgat, I., and J. Grainger. 2009. "Serial position effects in the identification of letters, digits, and symbols." *Journal of Experimental Psychology: Human Perception and Performance* 35 (2): 480–498.

Walsh, V., and S. R Butler. 1996. "The effects of visual cortex lesions on the perception of rotated shapes." *Behavioural brain research* 76 (1-2): 127–142.

Whitney, C. 2001. "How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review." *Psychonomic Bulletin & Review* 8: 221-243.

———. 2008. "Supporting the serial in the SERIOL model." *Language and Cognitive Processes* 23 (6).

Whitney, C., and R. S. Berndt. 1999. A new model of letter string encoding: Simulating right neglect dyslexia. In *Progress in brain research*, ed. J.A. Reggia, E. Ruppin, and D. Glanzman, 143-163. Elsevier.

Whitney, C., and P. Cornelissen. 2008. "SERIOL reading." *Language and Cognitive Processes* 23: 143-164.