Running head: MORPHOLOGY IN SPELLING-TO-SOUND RELATIONS

The Contribution of Morphology to the Consistency of Spelling-to-Sound Relations: A Quantitative Analysis Based on French Elementary School Readers

Ronald Peereman[a] , Liliane Sprenger-Charolles[b-c] , Souhila Messaoud-Galusi[d]

a. CNRS and Pierre Mendes-France University, LPNC (Laboratoire de Psychologie & NeuroCognition, UMR 5105), BP 47, 38040 Grenoble Cedex 9. France

b. CNRS and Aix-Marseille University, LPC (Laboratoire de Psychologie Cognitive, UMR 7290, Bat 9, Case D, 3, place Victor Hugo, 13331 Marseille Cedex 1, France

c. Associated CNRS and ParisDescartes University (LPP, Laboratoire Psychologie de la perception, UMR 8158, PRES Sorbonne-Paris-Cité, Labex EFL), France

d. Research Triangle Institute (RTI International), Education Policy and Systems, 701 13th St NW, Suite 750 * Washington, DC 20005. USA

Abstract

In alphabetic writing systems, the consistency of grapheme-phoneme (GP) and phoneme-grapheme (PG) relations is a key factor in learning to read and write. The difficulty of the tasks facing beginning readers/spellers in different orthographies is typically estimated using quantitative data on the consistency of GP or PG relations in the written words to which children are exposed. However, until now, these relations have been examined independently of morphological analysis. The present study aims to examine the contribution of morphological information (mainly inflectional morphology) to the consistency of GP and PG correspondences in a set of nearly 10,000 words found in French primary school readers (Lété et al., 2004). The database generated during this study, *Manulex_morpho*, has been made available via the internet.

Keywords

Grapheme-phoneme consistency; phoneme-grapheme consistency; morphological cues; learning to read; learning to spell; linguistic database

The Contribution of Morphology to the Consistency of Spelling-to-Sound Relations: A Quantitative Analysis Based on French Elementary School Readers

Alphabetical writing systems differ greatly in the consistency of their mappings between the minimal units of the written language (graphemes) and the minimal units of the spoken language (phonemes). Grapheme-phoneme (GP) and phoneme-grapheme (PG) inconsistencies arise from the assignment of different pronunciations to the same graphemic unit or the transcription of the same phonemic sequence using different spellings. For example, the vowel *oa* has different pronunciations in the words *road* and *broad*, and *deep* and *heap* are pronounced similarly despite different spellings. Like English, the French language is characterized by the presence of ambiguous GP and PG mappings. For example, the sounds /k/ and /ɛ̃/ have several spellings: c, qu, k, ch; in, ain, ein, ym, aim…. Conversely, spelling units such as *ille*, *en* or *er* have different phonological counterparts: /i/ vs. /j/ in ***v*ille** (town) and ***fille*** (girl); /ɑ̃/ vs. /ɛ̃/ or /ɛn/ in ***v*en*t*** (wind), ***bi*en** (good), or ***dolm*en** (dolmen); and /e/ vs. /ɛʀ/ in ***dang*er** (danger) and ***fer*** (iron).

In a cross-linguistic study of letter-phoneme associations at the initial position of words, Borgwaldt, Hellwig and de Groot (2005) showed large variations in seven different orthographies. Finnish displayed the strongest consistency, English the weakest, with French occupying an intermediate position. Estimates not restricted to word-initial graphemes indicate greater consistency in French than in English in the orthography-to-phonology direction but reduced consistencies in both languages in the phonology-to-orthography direction (Peereman & Content, 1997, 1998). Particularly in French, many inconsistencies are located at word endings. Hence, several final consonants are pronounced in some words but silent in others. For instance, the consonants are pronounced in the first word of each of the following pairs, not in the second: *brut* (raw) vs. *chat* (cat); *choc* (crash) vs. *croc* (fang)*; dix* (ten) vs. *prix* (price); *hamac* (hammock) vs. *tabac* (tobacco); *métis* (mixed-race) vs. *tapis*

(carpet); *ours* (bear) vs. *cours* (lesson). In addition, many phonological word endings have multiple orthographic renderings due to the presence of silent consonants: ‘ar’, ‘art’ and ‘ard’ for /aʀ/, as in *bar* (bar), *part* (slice) and *tard* (late); ‘ord’, ‘ort’, ‘ors’ and ‘orc’ for /ɔʀ/ as in *bord* (edge), *fort* (strong), *alors* (then), and *porc* (pig). However, in French, a large number of these silent letters correspond to morphological markers. This is the case for the *-s* and *-ent* inflections specifying number for nouns and verbs respectively: *ami* - *amis* (friend-friends); *il mange* - *ils mangent* (he eats, they eat). This is also the case for ending consonants that mark an inflectional or derivational relation between words: e.g., the silent *d* in *grand* (tall, masculine) vs. *grande* (tall, feminine) or *grandeur* (size); and the silent *g* in *long* (long, masculine) vs. *longue* (long, feminine) and *longueur* (length). Thus, previous characterizations of GP and PG mappings in French and other languages might well have underestimated consistency because possible relevant morphological cues were not considered in the analyses.

The aim of the present study was to determine the quantitative impact of morphological cues, mainly inflectional, on the GP and PG consistency of the French language that children encounter during the early years of learning to read. As we discuss below, several psycholinguistic observations suggest that such analyses stand to provide a better picture of the learning difficulties that children have to face when starting to read and write. Many studies have shown that the consistency of GP and PG relations influences reading and spelling acquisition in English (Backman, Bruck, Hebert, & Seidenberg, 1984; Caravolas, Kessler, Hulme, Snowling, & 2005; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Treiman, Mullenix, Bijeljac-Babic, & Richmond-Welty, 1995; Waters, Seidenberg, & Bruck, 1984; Weekes, Castles, & Davies, 2006) and in French (Alegria & Mousty, 1996; Lété Peereman, & Fayol, 2008; Sprenger-Charolles, Siegel, Bechennec, & Serniclaes, 2003; Sprenger-Charolles, Siegel, & Bonnet, 1998). Reading and/or spelling acquisition also

happens faster in transparent orthographies such as Greek or Italian than in less consistent orthographies such as English (Bruck, Genesee, & Caravolas, 1997; Ellis et al., 2004; Ellis & Hooper, 2001; Frith, Wimmer, & Landerl 1998; Goswami, Gombert, & Barrera, 1998; Goswami, Ziegler, Dalton, & Schneider, 2003; Seymour, Aro, & Erskine, 2003). Variations in the degree of consistency between orthographies also lead to differences in reading strategies (Frost, Katz, & Bentin, 1987; Katz & Frost, 1992; Goswami et al., 1998; Goswami, Ziegler, Dalton, & Schneider, 2001, 2003; Paulesu et al., 2000; Wimmer & Goswami, 1994, for reviews: Sprenger-Charolles, 2004, Ziegler & Goswami, 2005). For instance, while transparent orthographies encourage the use of GP conversion processes, the existence of numerous spelling inconsistencies leads the reader confronted to a deep orthography to rely more frequently on stored lexical knowledge to read words (but see Seidenberg, 2011, for an alternative hypothesis). Differences in consistency between orthographies also appear to modulate the size of the unit used to segment spoken and written words, with English readers relying on large processing units, such as the rimes of words (Goswami et al, 2003; Ziegler et al., 2010), but not Spanish or German readers (for reviews: Ziegler & Goswami, 2005, 2006). English readers rely on large processing units because the high level of inconsistency in the pronunciation of vowels drops when the surrounding consonants are taken into account (Treiman & Kessler, 2006; Treiman, Kessler, & Bick, 2003; Sprenger-Charolles, 2004). Not surprisingly, the critical role of GP and PG conversion processes in literacy acquisition has led to quantitative analyses of the orthography-phonology mappings of the words that beginning readers are exposed to in different orthographies (Garabik et al., 2007; Masterson Stuart, Dixon, & Lovejoy, 2003; Peereman, Lété, & Sprenger-Charolles, 2007).

Children's reading and writing performance seems, however, to be related not only to phonological skills: several studies have found that morphological knowledge too has an impact on learning to read and write (e.g., Casalis & Louis-Alexandre, 2000; Clin, Wade-

Woolley, & Heggie, 2009; Colé, Bouton, Leuwers, Casalis, & Sprenger-Charolles, 2012; Deacon & Kirby, 2004; Mahony, Singson, & Mann, 2000; Shankweiler et al., 1995). A particular source of difficulty in learning the French morphological system lies in the prevalence of inflectional morphemes that remain silent in speech (Catach, 1980). For instance, the markers for gender (final *-e*) and number (final *-s*, *-x* for nouns, final *-ent* for verbs) are not pronounced in normal speech, and phonological word forms are often homophonic. This is the case for verbs: il *mange*-ils *mangent* (he eats, they eat); for nouns: *ami-amie-amis* (friend: masc. sing. form, friend - fem. form, friends); *jeu-jeux* (game, games); for pronouns: *il-ils* (he, they - masc. form); *elle-elles* (she, they - fem. form); and also for adjectives (French adjectives agree in gender and number with the nouns they qualify): *joli-jolie-jolis* (nice: masc. form, fem. form, plural). This presumably explains why children often omit number agreement when spelling words at the onset of reading acquisition (Largy, Cousin, Bryant, & Fayol, 2007). Moreover, the use of the *-s* plural inflection seems to be more closely related to the individual frequencies of occurrence of the singular and plural forms of each word in the written language that the child encounters (Largy et al., 2007). Acquisition of the morphology of number takes place during the second year of schooling, and is faster for nouns than for adjectives and verbs. An overgeneralization of the use of final-s to mark plurality is observed for verbs (*ils *manges* instead of *ils mangent*, they eat). In addition, erroneous substitution in number agreement for adjectives with the verbal inflection (*-ent*) is seen until the sixth year of schooling (Thévenin, Totereau, Fayol, & Jarousse 1999; Totereau, Barrouillet, & Fayol, 1998; Totereau, Thevenin, & Fayol, 1997). Such overgeneralization errors are also observed for past participles when used as an epithet with the auxiliary *avoir* (to have; e.g., *les garçons ont *coiffés…* instead of *les garçons ont coiffé,* the boys have combed…: Fayol & Pacton, 2006) and for the verbal inflection /e/ (spelled *-er* and not *-é* in the infinitive form: Brissaud, Chevrot & Lefrancois, 2006).

Although it is acquired slowly, morphological knowledge should help children in spelling words. This is particularly obvious with respect to the numerous silent letters that occur at the end of words. Indeed, in such cases, inflectional or derivational knowledge can remove uncertainty about the identity of the silent letters. As an example, the -e marking the feminine gender can lead to a phonological transformation of the stem, denasalizing the vowel (e.g., *brun* /bRœ̃/ - *brun**e*** /bRyn/, brown) or altering the pronunciation of a consonant that is silent in the masculine form (e.g., *gris* /gRi/ - *gris**e*** /gRiz/, grey). Knowing the pronunciation of the feminine form of a word should thus help to determine the identity of the silent letter. Similarly, derivational knowledge can be useful where silent final consonants are phonologically realized in derived forms (e.g., *chant* /ʃɑ̃/ - *chanter* /ʃɑ̃te/ - *chanteur* /ʃɑ̃tœʀ/, song - to sing - singer). This may explain why children more often correctly produce silent consonants occurring at the end of words when they are pronounced in derivatives than when no derivation exists: e.g., the silent -d in the word *buvard* (blotter; Sénéchal, 2000; Sénéchal, Basque, & Leclaire, 2006). Interestingly, in less frequent cases, sensitivity to such derivational knowledge in spelling can yield errors consisting in the addition of an extra consonant to words ending with a vowel (e.g., *numéro* /nymeʀɔ/, number, spelled **numérot* because a /t/ is heard in the derivative *numéroter* /nymeʀɔte/, to number; Pacton & Casalis, 2006).

Compared to studies on the acquisition of French morphology in spelling, very little research has focused on children's use of morphological information during reading. Several studies, however, have found evidence of a relationship between reading performance and morphological awareness, defined as the ability to reflect on and manipulate morphological relations between words (Carlisle, 2000; Wink et al., 2009; Mahonny et al., 2000). Casalis and Louis-Alexandre (2000) found that kindergartners' inflectional and derivational

knowledge partially predicts their reading performance in Grade 2. In addition, Plaza and Cohen (2003) observed a relationship between syntactic/morphemic skills and reading and spelling in Grade 1. Children who manifest deficits in decoding words are also delayed in morphological analysis. For instance, St. Pierre and Béland (2010) showed that 9- to 12-year-old French-speaking Canadian children with delayed reading acquisition were also delayed in the acquisition of inflectional morphology. In these children, reading performance was more impaired than in control children for words ending with a silent consonant that is only heard in case of inflection for gender (e.g., adjectives: *petit* /pəti/- *petite* /pətit/, little) or number agreement (e.g., verbs: *il sort* /sɔʀ/ - *ils sortent* /sɔʀt/, he goes out - they go out). Unlike studies in adults, few studies with children have investigated the role of morphemic knowledge during on-line processing in reading acquisition. However, the data available from some tasks (completion and reading aloud tasks) suggest an early contribution of morphological knowledge in reading. Two distinct phases characterize the completion task. In the first phase the child reads a list of words, including critical items such as *turned* or *turnip*. In a second phase, the child must complete spelling patterns (e.g., T_ _ N). The results indicated that the percentage of correct completions (TURN) was higher when the initial phase included a morphologically related word (turned) than when it did not (turnip). In addition, the effect was similar from Grade 1 to Grade 8 (Deacon, Campbell, Tamminga, & Kirby, 2010; Rabin & Deacon, 2008). Consistent observations that suggest morphological processing during reading in children have been described in the reading aloud task (Colé et al., 2012; Laxon, Rickard, & Coltheart, 1992; Marec Breton, Gombert, & Colé, 2005). For example, data reported in Colé et al. (2012) highlight the fact that Grade 2 and Grade 3 children read pseudowords made up of a root and an existing French suffix (like -*ure* in *chature*) better and faster than pseudowords composed of a pseudo-root and a real suffix (*choture*, *chot* is not a French word), the latter being read better and faster than pseudowords

composed of a pseudo-root and a pseudo-suffix (*chotore*, *-ore* is not a suffix in French). In a second experiment, real words were presented in four different formats: normal (e.g., *dentiste*, dentist), syllabically segmented (den-tiste), morphologically segmented (dent-iste), or segmented without regard to syllables and morphemes (denti-ste). Only the last condition slowed down reading. These observations suggest that young French readers make use of morphological knowledge in reading words.

Acquisition of morphological knowledge and its use in reading and spelling might be useful to circumvent grapheme-to-phoneme (GP) and phoneme-to-grapheme (PG) inconsistencies in deep orthographies. As mentioned above, flexional and derivational knowledge can be used to reduce uncertainty with regard to the spelling of words that end with silent letters. In other cases, the identification of inflectional morphemes reduces the number of possible orthographic alternatives. For example, numerous orthographic counterparts of the /ɔ̃/ phoneme exist at the end of words (*-ond*, *-ong*, *-ont*, *-onc*, *-ons*, *-on*, *-omb*, *-om* …), but only two remain when it corresponds to a verbal inflection (-ons, -ont). Similarly, there are several allographs for the /ɛt/ ending of words (*-et*, *-aite*, *-ète*, *-ette*, *-ept*, *ête*…), but only one spelling when it corresponds to a diminutive suffix (*-ette*; e.g., *maisonnette*, small house). In reading aloud, identifying the word finals *-er* or *-ent* as verbal inflections should also facilitate GP conversion, since in this case *-er* is always pronounced /e/ as in *marcher* (to walk, and never /ɛʀ/ as in the adjective *cher*, expensive), and -ent is always silent as in *les poules couvent* (the hens are brooding) and never pronounced /ɑ̃/ as in *le couvent* (the convent).

Until now, quantitative analyses of GP and PG mappings performed on broad word corpora have not incorporated morphological coding. For instance, the French studies described by Ziegler, Jacobs, and Stone (1996) and Peereman and Content (1999) focused on

grapho-phonemic relations in monosyllabic words at the level of the rhyme (vowel and optional following consonants) without morphological cueing. Moreover, the word corpus (taken from the *Brulex* database: Content, Mousty, & Radeau 1990) included few inflected forms, no plural forms, and no inflected verbs. Similarly, the Dubois-Buyse word corpus (Ters, Mayer, & Reichenbach, 1977) analyzed by Veronis (1986, 1988) included only non-inflected forms. The more recent estimates provided by the *Manulex_infra* database (Peereman et al., 2007) included inflected and derived forms but no morphological cueing, so that consistency estimates were established without regard to the morphological status of the included orthographic and phonological strings. Thus, GP and PG consistency for words ending with an -s grapheme did not distinguish words in which the final -s corresponded to a plural marker (e.g., *jours*, days), or to a final pivot-letter for inflection or derivation (e.g., *gris*, grey), from the final -s of *alors* (then), *concours* (competition) or *mais* (but). In light of findings suggesting that children make use of morphological information in early reading and writing, the current available quantitative estimates may therefore provide an incomplete picture of the GP and PG consistency of the words included in the books used in elementary schools in France.

The primary aim of the present work was to assess the contribution of morphological information to the consistency of GP and PG associations. Because derivational affixes are generally supra-graphemic units, the morphological cues taken into account were mainly inflectional, and related especially to word endings, which previous studies have shown to include strong PG inconsistencies (Lété et al., 2008; Peereman & Content, 1997, 1998; Peereman et al., 2007). The analyses we present should thus provide quantitative estimates of the impact of morphological information on GP and PG mappings in the French language. In order to better capture the reading and spelling difficulties encountered by beginning readers, the study was based on a corpus of nearly 10,000 words drawn from school readers (Lété,

Sprenger-Charolles, & Colé, 2004). To evaluate the contribution of morphological information, the study contrasts quantitative analyses carried out either with or without morphological cues.

Method

Word corpus. The words were taken from Manulex (Lété et al., 2004), a database that provides the frequencies of occurrence of nearly 49,000 orthographic forms encountered in fifty schoolbooks in use from Grade 1 to Grade 5. Forms corresponding to abbreviations or interjections were excluded from the analyses (0.3%). The analyses were also restricted to words whose frequency of textual occurrence estimated by the U index was higher than 2.99 (according to Lété et al., 2004). Introduced by Carroll et al. (1971; see also Zeno et al., 1995), the U index has the advantage over the F index (frequency per million words) of being weighted by the dispersion of the words across the different books. Indeed, the computation of the U index takes into account the total frequency of occurrence of the words in the schoolbooks as well as the number of different books in which the words are encountered. The U-value is higher when the word appears in many books. Thus, the U index makes it possible to distinguish words of similar F frequency when one of them appears frequently in a few books, and the other in many books. Previous analyses have indicated that the U frequency index is a significant predictor of spelling scores in children from Grade 1 to Grade 5 (Lété et al., 2008). The selected word corpus (hereinafter *Manulex_morpho)* corresponded to approximately 20% of the lexical entries occurring in Lété et al. (2004) but to 98% of the words encountered by children in their schoolbooks (textual occurrences). It included 9,949 word forms corresponding to 5,253 lemmas. The proportions of the different grammatical categories are reported in Table 1, and the corresponding values observed in Lété et al. (2004) are shown for comparison. As can be seen, relative to Manulex, our selection leads to a slight

reduction in the percentages of proper names, adjectives, and verbal forms, and an increase in the percentage of nouns.

TABLE 1

Phonetic notations, graphemic segmentation and GP and PG associations were derived from Peereman et al. (2007). Phonetic notations and segmentations were checked by two of us (LSC and SMG) and corrected: several modifications were introduced, particularly in relation to the coding of the vowels and the coding of the silent E (schwa), for which we introduced the distinction between the obligatory schwa (e.g., *tabl**e***, table or *r**e**nard*, fox) and the optional one (e.g., *rout**e***, road, or *boul**e**vard*, boulevard). Details on these changes are provided in Appendix A. On the whole, approximately 12% of the phonetic symbols were modified. The phonological representations were based on 17 vowels, 3 semi-vowels, 18 consonants, and an additional mark to code silent graphemes: e.g., the letter "t" in *petit* (little) or the letter "d" in *grand* (tall).

Morphological marking (set-up by LSC with the help of SMG). Four categories of morphological cues were used to mark graphemes and phonemes. The vast majority related to word finals for which previous estimates (Peereman & Content, 1998; Peereman et al., 2007) indicated low consistency. Altogether, 7,456 grapheme-phoneme associations were morphologically marked, distributed over 6,377 words (64% of the lexical corpus). The morphological cues considered in the present work were selected following Brunet (1981) and Catach (1980, 1995). For the verbs, Dubois (1967) and Le Goffic (1997) were also consulted.

1. Gender and number. Graphemes and phonemes corresponding to gender and number inflections were marked. This was used to distinguish cases where the grapheme S or X corresponded to the plural (e.g., *vélos,* bikes; *fourmis,* ants; *cheveux,* hair; *choux,* cabbages) from cases where they did not correspond to inflections (e.g., *mais,* but; *houx,* holly). Gender marking (grapheme -*e*) was noted when the masculine and feminine

forms of the word differ (e.g., *ami-amie*, friend -masc., fem. form). Gender was also noted for words that are intrinsically marked as feminine words such as *colonie* (colony), *fusée* (rocket), *joue* (cheek), *avenue* (avenue), etc. That choice allows to differentiate these words ending with a marker for the feminine from masculine words ending with -e, such as *lycée* (secondary school), *musée* (museum), *génie* (genius), *incendie* (fire), etc. Graphemes and phonemes of verbs at the past participle were also marked for gender and number (e.g., *elle est venue*, she came; *ils sont prêts*, they are ready). Overall, 2,136 words were marked for gender and/or number, some of them being marked twice (e.g., *Mesdames*, *lesquels*).

2. Verbal inflections. The grapheme *-er* occurring at the end of verbs in the infinitive form (*marcher*, to walk), and the *-ant* ending for present participles (*marchant*, walking) were marked as verbal inflections. The other graphemes and phonemes corresponding to verbal inflections were coded following Dubois (1967) and Le Goffic (1997). This was the case for the morphemes of the past participle: *-é* for the verbs of the first group, as in *il a mangé* (he ate); *-i* for the second group as in *il a fini* (he finished); *-é*, *-i*, *-u*, *-is*, and *-t* for the third group as in *il est allé*, *il a dormi*, *il a vu*, *il est assis*, *il a peint* (he went, he slept, he saw, he is sitting, he painted). Desinences of other verb tenses were also marked even when the grapheme was silent in speech. In the present tense, this was the case for the following inflections: *-e*, *-es*, *-ons*, *-ez* , *-ent* (verbs of the first group); *-is*, *-it*, *-ons*, *-ez*, *-ent* (verbs of the second group); and *-s*, *-t*, *-x*, *-d*, *-ds*, *-ts* (additional inflections for the third group: *je dois, il peut, je peux, il prend, je prends, tu bats*). In the future tense, the following graphemes were coded as verbal inflections: *ai*, *as*, *a*, *ons*, *ez* and *ont*. Overall, 3,166 words were marked with verbal inflections.

3. Adverbs. The final grapheme *-ent* and its corresponding pronunciation (/ɑ̃/) was marked where it occurs in adverbs ending with the derivation *-ment* pronounced /mɑ̃/,

such as *rarement* (rarely), *suffisamment* (enough), and *vraiment* (really). This marking applied to 134 adverbs.

4. Graphemes allowing inflection or derivation. The fourth category of morphological tagging pertained to final consonants that can be silent in the word, but heard in inflected or derived words: e.g., -d in *grand-grande-grandeur* (tall - masc. gender, fem. gender, size) and -s in *anglais-anglaise* (English - masc. gender, fem. gender). This marking was also made when the inflection or the derivation involves an orthographic transformation: e.g., *-g* vs. *-gu-* in *long-longue* (long - masc., fem. gender), *-c* vs. *-ch-* in *blanc-blanche* (white - masc., fem gender), *-x* vs. *-s-* in *époux-épouse* (husband, wife), *-f* vs. *-v-* in *vif-vive* (bright - masc., fem. Gender). It was also included when the inflection or derivation involve a denasalization of the vowel, as in *brun-brune* /bʀœ̃/-/bʀyn/ (brown - masc., fem gender), *bon-bonne* /bɔ̃/- /bɔn/ (good - masc., fem. gender), *marin-marine* /maʀɛ̃/- /maʀin/ (sailor, navy). Cases with irregular spelling transformations (e.g., *-p* vs. *-v-* in *loup-louve*, wolf - masc., fem. gender) were not considered, nor were morphemes that did not correspond to single graphemes (e.g., *-eur*, *-ice*, and *-euse* in *lecteur-lectrice* (reader masc., fem. gender), *serveur-serveuse* (waiter, waitress). The number of morphological markings in this category amounted to 1,865.

The various graphemes tagged with morphological cues are listed in Table 2.

TABLE 2

Grapho-phonemic (GP) and phono-graphemic (PG) consistency. Consistency indices measure the ambiguity of GP and PG mappings, taking into account both the frequency with which a particular grapheme is associated with a particular pronunciation and the overall frequency of the grapheme whatever its pronunciation. For example, the consistency of the relationship between the CH grapheme and its pronunciation /ʃ/ (e.g., *cheval*, horse) is

determined by the ratio between the number of times the grapheme CH is associated to /ʃ/ in the word corpus and the number of times the grapheme CH is encountered, regardless of its pronunciation (/ʃ/ or /k/ as in *chorale,* or /ʃt/ as in *sandwich*). When the grapheme is always associated with the same pronunciation, the ratio is equal to 1 (maximum consistency), as for the grapheme PH which systematically maps to /f/ (e.g., *photo, éléphant*). When multiple associations exist, the ratio is less than 1. Consistency values vary between 0 and 1. Overall consistency values can be determined for each individual grapheme and phoneme. As an example, there are 588 occurrences of the grapheme CH in the corpus; among them, CH is associated to /ʃ/ in 574 cases, to /k/ in 13 cases, and to /tʃ/ in one case. Thus, the consistency of the CH-/ʃ/ association equals .976 (i.e. 574/588), the consistency of CH-/k/ is .022, and the consistency of CH-/tʃ/ is .002. The overall consistency of the grapheme is obtained by weighting the consistency of each GP mapping by their frequencies of occurrence. For CH, overall consistency equals .953, i.e. ([.976*574]+[.022*13]+[.002*1])/588.

Consistency is usually estimated taking into account either lexical frequency (i.e. by-type count) or textual frequency (i.e., by-token count) of the GP and PG associations (Kessler & Treiman, 2001; Peereman & Content, 1999; Peereman & al., 2007; Ziegler et al., 1996). In the first case, the frequency of each association is determined by the number of words in the lexical database that include the association. In the second case, the frequency of each mapping is weighted by the textual frequency of the words that include the mapping. Therefore consistency measures are sensitive to the frequency of words in texts. For instance, “eu” for /y/ is a very rare GP mapping; however, this mapping occurs in frequent words such as *eu* (had); therefore, its lexical frequency is very low compared to its textual frequency. The word frequency (U) values used in the present study were extracted from Lété et al. (2004). Finally, consistency estimations were performed bi-directionally, that is for both GP and PG

mappings. GP consistency is a measure of the predictability of pronunciation (as in reading aloud), while PG consistency measures the predictability of spelling (as in writing to dictation).

Consistency according to morphological markers. To evaluate the contribution of morphological information to GP and PG consistencies, computations were performed with and without regard to morphological markers. A way to conceptualize the difference between these two analyses is to consider consistency computations according to morphological markers as conditional consistencies, with morphological information being the condition under which the estimates are derived. For example, in final position of words, the /ɑ̃/ phoneme has multiple allographs, such as *-an*, *-ant*, *-aon*, *-aen*, *-ean*, or *-ent*. Consequently, the overall consistency of the /ɑ̃/ phoneme is low (.429). However, considering morphological information improves this situation. Indeed, the consistency of /ɑ̃/ if the word is an adverb or a present participle (conditional consistency) is 1 (always *-ent* in the first case and always *-ant* in the second case). Similarly, in word-final position, the *-er* grapheme can be pronounced /e/ as in *chanter* (to sing), *léger* (light), *boulanger* (baker), /ɛʀ/ as in *cher* (expensive), *fer* (iron), *hier* (yesterday), or /œʀ/ as in *Peter*. However, only /e/ is possible when *-er* corresponds to a verbal inflection. Although the pattern is not perfectly regular, the *-er* grapheme in nouns or adjectives is also frequently pronounced /e/ when the final letter *-r* is heard in inflections and derivations (e.g., *premier-première*, *léger-légèreté*, *boulanger-boulangerie*); but it can be also pronounced /ɛʀ/ (e.g., *cher-chère*, *fer-ferrure*). Hence, taking morphological information into account can be expected to increase GP and PG consistency.

Results and Discussion

Three distinct analyses were carried out to estimate GP and PG consistency. The first takes into account all the morphemic markers listed above (hereinafter *with-M analysis.)* The second analysis was performed in the absence of any morphological cues (hereinafter *without-M analysis*). These first two analyses were based on strictly identical graphemic and phonemic segmentations, the only difference lying in that morphology was marked for the GP and PG mappings in the first analysis, but not the second. For example, in the with-M analysis, GP the consistency of *-s* when occurring at the end of words was determined separately for (1) nouns and adjectives in plural, (2) verbal inflections (third verb group, *tu viens,* you come; *tu crois*, you believe), (3) cases where the final -s is heard in inflected or derived forms (*gris* /gʀi/ - *grise* /gʀiz/, grey - masc. gender, fem. Gender; *trois* /tʀwa/ - *troisième* /tʀwazjɛm/, three - third), (4) cases of final -s not marked morphologically. Conversely, in the without-M analysis, no distinction was made between these cases: GP consistency was computed without regard to morphological information. In both analyses, word segmentations were based on the same 111 graphemic units and 64 phonemic units. Without morphological marking (without-M analysis), the number of different GP and PG associations was 212. Because morphological marking led to differentiated associations according to the morphological status of graphemes and phonemes, a larger number of associations was obtained in the with-M analysis (306). The comparison between these first two analyses should make it possible to assess the contribution of morphological information to the GP and PG consistency of the French words that children are exposed to in elementary school books.

Relative to the previous study on GP and PG mappings in French words described in Manulex-infra (Peereman et al., 2007), our morphemic analysis of orthography resulted in slightly different graphemic segmentations. This is reflected in the larger number of

graphemic units considered here (111 vs. 97). In particular, 14 additional graphemes were identified in the present study as potential verbal inflections *(-ent, -ant, -as, -es, -est, -ons, -ont, -ais, -ait, -is, -it, -ds, -ts, -pt*). Because the contribution of morphology was not the focus of the previous work, graphemic segmentations did not preserve morphemic units and orthographic strings such as *-ent*, *-ant*, *-ons*, and *-ont* were split into *-en/t*, *-an/t*, *-on/s*, *-on/*. Therefore, to connect our findings to those of this previous study, we also briefly report the results of a third analysis (hereinafter *Manulex-infra analysis*) based on the graphemic and phonemic segmentations initially adopted in Manulex_infra, except for the phonetic corrections described in Appendix A. This third analysis, also performed without taking morphological information into account, differed from the without-M analysis mentioned above exclusively in the choice of graphemic segmentations.

The results of the analyses for graphemic or phonemic units at any position in the word are presented first. Results according to position are then described: initial GP and PG associations, internal ones, and final ones (the grapheme or the phoneme in the initial or final positions is respectively the first and the last in the word, e.g., in the words “ami”, “amis”, “amie”, “amies”, the last graphemes are respectively “i”, “s”, “e”, and “s”). Finally, analyses for the different word-ending morphological markers are reported for GP and PG mappings. For each analysis, the consistencies of the graphemic or phonemic units are estimated taking into account either lexical frequency (by-type values) or textual frequency (i.e., by-token values) of the word including the unit considered.

Consistency of GP and PG mappings

Figure 1 displays the average consistencies for GP and PG associations obtained by the three different analyses. As can be seen, both GP and PG consistency values are higher when graphemes and phonemes are morphologically marked (with-M analysis). In comparison with the without-M analysis, morphological information causes an increase of 4.5% for GP

mappings, and of 3.5% for by-type PG mappings. For the analyses based on the textual frequency of the word (by-token mapping) the increase for GP and PG mapping is 5.3% and 3.6% respectively. The results of the analyses not considering morphological markers (without-M analysis, Manulex-infra analysis) are quite similar for PG associations. A weak increase in consistency appeared for GP mappings in the without-M analysis (2% per type and 2.4% per token), probably because of the use of slightly different graphemic segmentations.

FIGURE 1

The benefit of including morphological information can be more precisely determined by looking at consistency for the 63 graphemic units (out of 111) that were morphologically marked in the with-M analysis, but not in the without-M analysis. This is the case, for example, for the grapheme *–s*, which could correspond to a plural inflection, a verbal inflection, or a consonant heard in inflected/derived forms. Average consistency (per type) for these 63 graphemes was .902 (.895 per token) in the with-M analysis and .865 (.864 per token) in the without-M analysis. The benefit of drawing on morphology can be determined similarly for the 33 phonemic units (out of 64) that were associated to morphological markers. For these 33 phonemic units, average consistency was higher in the with-M analysis than in the without-M analysis (per type: .788 vs. .758; per token: .808 vs. .769). An advantage of the with-M analysis over the without-M analysis was also observed for the whole set of graphemic units (n = 111; per type: .883 vs. .862; per token: .892 vs. .875) and the whole set of phonemic units (n = 64; per type: .840 vs. .824; per token: .853 vs. .833).

A finding that deserves further attention is that while GP consistency is higher than PG consistency in the with-M analysis, this is not the case in the Manulex-infra analysis (see Figure 1). This last observation contrasts with the greater consistency of print-to-sound associations compared to sound-to-print associations usually reported for French (Lété et al., 2008; Peereman & Content, 1998; Peereman et al., 2007; Ziegler et al., 1996). However, a

critical difference between the data displayed in Figure 1 and previous estimates is that our initial analyses were performed without regard to the position of the graphemes and phonemes within the words. The weakness of GP consistency values in the Manulex-infra analysis likely results from the fact that consistency was determined independently of position.

Consistency as a function of position in the with-M and without-M analyses

The weakening of consistency when estimated independently of relative position in the word can be illustrated with the example of the phoneme /ɑ̃/, which has multiple orthographic renderings. In initial position, the grapheme *en-* is the most frequent (*encore, endroit, enfant, enfin, ensuite, entrer…*) whereas in final position *-an* (*maman, ocean, plan, ruban, toboggan.…*) predominates. Taking account of the relative position of the grapheme thus leads to a higher degree of consistency. The with-M analysis was not affected by the lack of positional cue, probably because morphological markers are strongly correlated with position (e.g., the grapheme *-s* is necessarily in final position if it is marked as a plural inflection). Additional analyses (by type) indicate that not considering position is more detrimental for GP than for PG consistency. When, in the Manulex-infra analysis, consistency was determined as a function of grapheme and phoneme position (initial, internal, final), GP consistency increased from .696 to .777 (8.1% increase) but PG consistency increased only from .701 to .728 (2.7%). These data also show that the higher consistency for GP than for PG mappings re-emerges once position is taken into account.

Hence, the increased consistency in the with-M analysis over the without-M analysis might have resulted from the fact that morphological markers were correlated with position in the word. We therefore compared the two analyses while taking grapheme and phoneme position into account. Average consistencies for GP and PG mappings are displayed in Figure

2a (counts per type) and 2b (counts per token).

FIGURE 2

Overall, consistencies were higher for initial positions than for internal and final positions, an observation already well documented for the grapho-phonemic system of French (Peereman et al., 2007). GP consistencies were higher than PG consistencies for all positions. These data strengthen the idea mentioned above that consistency is higher for GP than for PG mappings when estimated as a function of position in the word.[1] Figures 2a and 2b also show an increase in consistency for GP and PG units when grapho-phonemic associations were marked morphologically. The location of the effect on the final graphemes and phonemes is not surprising in view of the fact that most of the morphological cues occurred at the end of words (n = 6,362 in final position, n = 1,078 in internal position such as in the words *ami**e**s, pointu**e**s, bleu**e**s, vrai**e**s, au**x**quels…*). Finally, this increase appears to be larger for PG than for GP associations.

Among the 79 graphemic units encountered in word-final position, 58 were morphologically marked in the with-M analysis. Average consistency for these 58 graphemes was higher than in the without-M analysis (per type: .926 vs. .900; per token: .944 vs. .920). Among the 43 phonemic units ending words, 32 had morphological cues in the with-M analysis. Similarly to graphemic units, consistency of phonemic units was higher in the with-M analysis than in the without-M analysis (per type: .867 vs. .817; per token: .907 vs. .842).

1. Note that previous consistency estimations have suggested a larger difference between GP and PG consistency (e.g., e.g., Peereman et al., 2007). A first possible explanation for these differential outcomes is the present work's introduction of several changes to phonetic codes and graphemic segmentations (see Appendix A). A second possibility relates to the use of different methods to estimate average consistency. The statistical descriptions in the present study are derived by averaging the mean consistency of the 97 graphemic and 64 phonemic units. This procedure allocates exactly the same weight to each grapheme (or phoneme), thus preventing some very frequent graphemes (or phonemes) from inflating or on the contrary attenuating average consistencies. On the contrary, consistency as estimated in previous works corresponds to the mean consistency of all words included in the lexical corpus. This procedure does not prevent some graphemes or phonemes occurring in many words from disproportionately influencing consistency estimates.

The benefit of morphology was also observed for the whole set of graphemic and phonemic units (graphemic: n = 79; .939 vs. .919 per type and .956 vs. .939 per token; phonemic: n = 43; .876 vs. .839 per type and .913 vs. .865 per token).

Morphological contribution as a function of morphological cues

The consistency gains produced by the four different categories of morphological markers were examined for graphemes and phonemes occurring in word-final position. It is, however, worth noting that several graphemes and phonemes were attached to morphological cues belonging to more than one category. For example, the grapheme -*s* could be a marker for the plural or for a verbal inflection, or a final pivot-letter for inflection or derivation. In such cases, the grapheme or phoneme plays a role in consistency estimation for each of the relevant categories of morphological cues. Thus, the grapheme -*s* contributes to the evaluation of morphological consistency gains for number/gender cues, verbal cues, and for final pivot letters for inflection or derivation. The results, expressed in percentage consistency increase (by type), are shown in Figure 3.

FIGURE 3

The consistency benefits provided by the different morphological markers appear similar for graphemes and phonemes. Marking the ending grapheme -*ent* and its corresponding /ɑ̃/ phoneme as part of the adverbial derivation -*ment* caused the largest consistency benefit. The GP and PG mappings are indeed perfectly consistent when the grapheme -*ent* and the phoneme /ɑ̃/ occur in adverbs ending in -*ment*. Because the frequency of the *ent*–/ɑ̃/ mapping is essentially due to adverbs, the consistency of the other associations involving either the grapheme -*ent* or the phoneme /ɑ̃/ increases indirectly when occurring in a non-adverbial word. Gender and number markings had very limited effects on consistency. This results from similar distributions of the phonological counterparts of the final -*s* (silent or

less frequently /s/) in words where it corresponds to a morphological marker and in words where it does not. Hence, morphological cues do not increase consistency. The situation is slightly different for PG mappings because the silent grapheme is systematically *-e* for feminine gender words and almost always *-s* for plural forms of words. Finally, the marking of verbal inflections and the marking of final consonants heard in inflected or derived forms had varied effects depending on the grapheme or phoneme. For example, the final grapheme *-t* reaches maximum consistency when marked as a verbal inflection. Conversely, marking the final grapheme *-ez* as a verbal inflection did not increase its consistency because it is pronounced similarly in verbs and other words (as in *assez*, *chez*, and *nez*).

General Discussion

The use of morphological segmentation during reading is generally discussed within the framework of models postulating a morphemic decomposition of the words before their identification (Taft, 1979, 1994) or the existence of parallel pathways involving morphemic decomposition on the one hand and whole-word recognition processes on the other (Caramazza, Laudanna, & Romani, 1988; Schreuder & Baayen, 1995). In skilled readers, Meunier et al. (2008) showed that deciding on the masculine/feminine gender of morphologically complex words (gender decision task) is faster when the word and its base have identical gender (e.g., *la maisonnette* and *la maison*: maisonette, house) than when they are of opposite gender (e.g., *la camionette* and *le camion*, van, truck). This observation suggests a morphemic decomposition leading to a gender conflict when the complex word and its base are of different gender.[2]

2. Several studies based on the masked priming paradigm have shown that morphologically complex words (e.g., *gaufrette*, wafer) facilitate recognition of their bases (*gaufre*, waffle), even when the semantic relations between the prime and the target are opaque (*fauvette-fauve*, warbler-wild animal; *lunettes-lune*, glasses-moon; see Lavric, Clapp, & Rastle, 2007;

Most studies on morphological processing during reading have focused on derivational morphology, and empirical works on French inflectional morphology are scarce. Interestingly, however, the use of a morphemic decomposition procedure is also suggested by the finding that lexical decision latencies to words presented in the singular are a function of the frequency of *both* singular and plural forms (New, Brysbaert, Segui, Ferrand, & Rastle, 2004). This influence of the frequency of the plural form seems to indicate that words are segmented into base + desinence during reading, with the processing of words in the singular inheriting the frequency of the forms in the plural. Thus, a morphemic decomposition of the words into base + desinence would facilitate the recognition of some low-frequency words.

Our results indicate that a second advantage of the morphological analysis of words is to reduce the inconsistency of GP and PG mappings. At word-final position, consistency benefits of 3% and 11% (by type; 2% and 9% by token) were observed for GP and PG associations, respectively. The stronger inconsistencies previously observed for word-final than for internal graphemes and phonemes (Lété et al., 2008; Peereman & Content, 1997, 1998; Peereman et al. 2007) vanished when morphological markers were considered. Thus, taking words' morphemic composition into account reduces GP and PG ambiguities in reading and spelling.

To date, models of print-to-sound conversion (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2007) have not integrated morphological decomposition processes. However, as mentioned above, previous observations have shown

---

Longtin, Segui, & Hallé, 2003; McCormick, Rastle, Davis, 2008; Rastle, Davis, Tyler, & Marslen-Wilson, 2000; but see Feldman, O'Connor, & Moscoso Del Prado Martin, 2009, for divergent observations). Moreover, priming effetcs are still observed when the prime and the target are not etymologically related (e.g., *baguette-bague*, stick-ring) and when the prime corresponds to a pseudoword that can be analyzed as the combination of a root and a suffix. These observations contrast with the absence of facilitation when the prime word (e.g., *abricot*, apricot) or pseudoword (*rapiduit*) do not correspond to a base + affix (e.g., *-cot* and *-uit* are not suffixes in French; Longtin & Meunier, 2005 ; Rastle, Davis, & New, 2004).

morphological decomposition in children when naming words and pseudowords. In English speaking children, naming performance is better for suffixed words (e.g., dancer) than for pseudo-suffixed words (e.g., dinner: Carlisle & Stone, 2005; Laxon et al., 1992), and Marec-Breton et al. (2004) observed an advantage of pseudowords composed of a stem + affix (e.g., *bougeur*) over pseudowords composed of a pseudostem + affix (*sanneur*) in French children as early as Grade 2. A number of findings suggest that at an early stage of processing, morphemic parsing operates independently of semantic factors, on a purely orthographic basis (Beyersmann, Coltheart, & Castles, 2012; Longtin et al., 2003; Rastle et al., 2004). Hence masked priming effects are observed for prime-target pairs such as *darkness-DARK* and *corner-CORN* in spite of the fact that *CORN* does not correspond to a stem in *corner* (i.e. pseudo-suffixed word). Morphological segmentation thus seems to be guided by the presence of orthographic units corresponding to identifiable morphemes (such as *-ER* in *farmer*, *singer*). A second stage of morphemic analysis relying on syntactic-semantic properties is thought to take place later, accounting for masked priming effects with irregular inflections such as *fell-FALL* (Crepaldi et al., 2010). Interestingly, Quémart, Casalis, and Colé (2011) reported evidence of early morpho-orthographic segmentation in developing readers (Grades 3-5-7). Briefly exposed primes (60 ms) facilitated target identification both for morphologically related prime-target pairs (*tablette* - *TABLE*) and pseudo-derivation pairs (*baguette* - *BAGUE*). No facilitation was observed when the prime was orthographically similar to the target but did not include a possible French suffix (*abricot* - *ABRI*). A fast-acting morphological segmentation based on the orthographic structure of the letter string might therefore influence print-to-sound conversion. Word constituents corresponding to identifiable morphemes such as CHAT and S in the French word CHATS (cats) might therefore be converted taking into account the morphemic status of the final -S. This procedure would not work when segmentation does not yield identifiable morphemes (e.g.,

ALOR + S for the word ALORS, then). Although this remains purely speculative, it is likely that the phonological hypotheses derived through the print-to-sound mechanisms are weighted by syntactic-semantic information retrieved from the word-stem or from the sentence context. An argument in support of this view comes from Campbell and Besner's (1981) finding that the way in which the TH grapheme is pronounced when occurring at the initial position in pseudowords (e.g., THED, THIT) depends on the pseudowords' position within the sentence: generally voiced in function word position, and generally unvoiced in content word position.

Because most of the morphological tags used in the present study related to inflectional morphemes, the real contribution of morphology to GP and PG encoding might have been underestimated. Nevertheless, grammatical inflections are present in a very large proportion of the words seen by French children in school readers (around 50% according to Manulex: Lété et al., 2004). It should be noted that we have taken into account final consonants that support inflection or derivation ("d" in *grand* vs. *grande* or *grandeur; 's'* in *gros* vs. *grosse* or *grosseur, 't'* in *haut* vs. *haute* or *hauteur*), which also caused consistency to increase, especially PG consistency. Indeed, three letters (D, S, T) account for 83% of all possible silent final consonants when marked as a final pivot-letter for inflection or derivation. Finally, our study looked at GP and PG associations, without considering morphemic units larger than the phoneme (e.g., *eur, -ette, -able, -ess, -age*). Such estimates await further advances in the morphological analysis of the French orthographic lexicon (see the pioneering work of Brunet, 1981).

The increase of GP and PG consistency when morphology is taken into account has a cost, however: a larger number of associations between graphemes and phonemes. For example, distinguishing between the different final types of *-s* grapheme according to whether or not they correspond to a morphological cue requires the consideration of not one, but four distinct grapho-phonemic associations (number inflection, verbal inflection, support of

inflection or derivation, other cases). The same applies, for example, to the grapheme *–ent*, which can occur at the end of adverbs, verbs or nouns. From the point of view of the acquisition of reading and spelling, the advantage of morphemic analysis could thus be attenuated by the requirement of learning a more significant number of grapho-phonemic associations. This problem is similar to that generated by the use of supra-graphemic print-to-sound associations, such as the rhyme, in English readers. As Ziegler and Goswami (2005) pointed out, while the use of larger orthographic units can reduce a portion of the grapho-phonemic ambiguities of English, it also leads to a need to learn more associations.

Analysis of the consistency advantages of morphological markers reveals, however, that the benefits are not homogeneous. For instance, 8 graphemes are responsible for 96% of the increases in GP consistency (archigrapheme “E” (“e-é-è-ê-ë”, as well as “et-êt”), “ent”, “om”, “um”, “c”, “b”, “g”). Similarly, 9 phonemes produce 97% of the increases in PG consistency (/e/-/ɛ/, /et/-/ɛt/, /i/, /ɑ̃/, /ɔ̃/, /t/, /k/, the optional schwa, and the silent graphemes). When consistency values are weighted by graphemes’ and phonemes’ frequencies of occurrence in the lexical corpus, it emerges that just 9 graphemes are responsible for 97% of the increases in GP consistency: the archigrapheme “E” (“e-é-è-ê-ë”, “er”, “et-êt”) and the graphemes, “ent”, “om”, “um”, “d”, “g”, “s”. Similarly, just 7 phonemes produce 99% of the increases in PG consistency: /e/-/ɛ/, /i/, /ɑ̃/, /ɔ̃/, the optional schwa (as in *boule*), the obligatory schwa (as in *amie*), and the other silent graphemes (such as ‘d’ in *grand*, ‘t’ in *haut* and ‘s’ in *gris*). Our results thus suggest that morphological markers are the most productive for these GP and PG associations

The developments here in the quantification of the contribution of morphological information to grapho-phonemic encoding also provide a valuable tool to investigate reading and spelling in children, and a basis for the establishment of teaching programs. In this context, the present work represents an extension of the Manulex (Lété et al., 2004) and

Manulex_infra (Peereman et al., 2007) databases on the written words that children encounter in elementary school. As mentioned above, the morphological coding of words is currently restricted to morpho-graphemes and focuses mainly on inflectional markers. Developments currently in progress should extend morphemic analyses to derivational morphemes and supra-graphemic units. Lété et al. (2008) have indeed recently described observations with French-speaking children suggesting that phono-graphemic conversion involves processing units of increasing size as a function of reading pratice. Despite the limits that are intrinsic to any development of linguistic databases, the estimations of grapho-phonemic consistency generated during this study are, to our knowledge, the only ones to consider the contribution of morphological information. The new Manulex_morpho database is available on the internet.

Based on Manulex-morpho, the Research-Triangle-Institute (RTI International) is developing teaching programs designed for low-income Francophone countries. These programs consist of scripted lessons that require minimal resources and that teachers can easily replicate in the classroom. The progression takes into account the frequency and consistency of spelling-to-sound units for GP (including morpho-phonograms), as well as word frequency (see Dehaene, Dehaene-Lambertz, Huron, Gentaz & Sprenger-Charolles, 2011). This progression begins with the most consistent and most frequent GP mapping. For each lesson, a corpus of words containing the GP correspondences introduced in the new lesson and those already learned is established. This corpus is used to develop appropriate reading materials to allow students to practice reading based on utterances that are at a level corresponding to their decoding skills. In addition, the most frequent words that are needed to read small texts are introduced quickly, whether or not they are consistent in terms of GP

correspondences.[3] Since in these countries French is generally not the children's mother tongue, the differences between their spoken language and French are taken into account, especially phonological differences such as the presence of "u" vs. "ou", the neutral "e", or nasal vowels. The other main aims of the teaching programs are improvements in oral and written vocabulary (taking morphology into account) and in oral and written comprehension.

3. Determinants (*le/la/les*, the; *un/une*, a; *mon/ma*, *ton/ta*, *son/sa*, my, your, his/her…); pronouns (*je, tu, il/elle*, I, you, he/she; *qui/que*, who, which or that…); prepositions (*a*, to; *de*, from; *par*, by; *pour*, for…); conjunctions (*et*, and; *ou*, or; *car*, *parce que*, *puisque*, because...), nouns and adjectives (*fille*, daughter; *garçon*, boy; *monsieur*, sir*; nez*, nose; *texte*, text; *histoire*, history; *maison*, house; *beau/belle*, beautiful for a boy versus a girl ...) and frequent verbs (*être*, to be; *avoir*, to have; *faire*, to do; *savoir*, to know; *aller*, to go; *dire*, to say; *lire*, to read...) and their most frequent conjugated forms (present, past tenses, future and imperative).

Files available on the Internet

The *Manulex_morpho* database is available on the internet (http://webu2.upmf-grenoble.fr/LPNC/resources/ronald_peereman/Manulex_morpho/). The database provides the consistency values of grapheme-phoneme (GP) and phoneme-grapheme (PG) mappings (*with_M analysis*) for each of the 9,949 words considered in the present study. A list of the different GP and PG associations in the word corpus is given in a separate file, together with their corresponding frequency and consistency values. This list can be useful to estimate consistency of grapho-phonemic associations in pseudowords or in words not included in the *Manulex_morpho* corpus. Additional information is presented in the user manual.

Several variants of the database are available. For the results of the present paper, consistency of GP and PG mappings is weighted by the U textual frequency values reported in Lété et al. (2004). For various reasons (e.g., cross-linguistic comparisons), the user may prefer to use consistency values weighted by the textual frequency per million words. This variant of *Manulex_morpho* is also available, based on the F frequency data (Lété et al.) and reported per million words. Finally, for each of these two variants, two different estimates were calculated according to how vowels were distinguished. Indeed, in the present study, consistency values were derived after removal of several distinctions between vowels (i.e. removal of the distinctions /o/ vs. /ɔ/, /e/ vs. /ɛ/ and /ø/ vs. /œ/; see Appendix A). However, depending on their objective, users may prefer to use consistency values which preserve these distinctions. Both types of estimates are available.

REFERENCES

Alegria, J., & Mousty, P. (1996). The development of spelling procedures in French-speaking, normal and reading-disabled children: Effects of frequency and lexicality. *Journal of Experimental Child Psychology, 63*, 312-338.

Anderson, S. R. (1982). The analysis of French shwa: Or, how to get something for nothing. *Language, 58*, 534-573.Backman, J., Bruck, M, Hebert, M., & Seidenberg, M. S. (1984). Acquisition and use of spelling sound correspondences in reading. *Journal of Experimental Child Psychology, 38*, 114-133

Beyersmann, E., Coltheart, M., & Castles, A. (2012). Parallel processing of whole words and morphemes in visual word recognition. *The Quarterly Journal of Experimental Psychology, 65*, 1798-1819.

Borgwaldt, S.R., Hellwig, F. M., & de Groot, A. M. B. (2005). Onset entropy matters – Letter-to-phoneme mappings in seven languages. *Reading and Writing, 18*, 211-229.

Brissaud, C., Chevrot, J.-P., & Lefrançois, P. (2006). Les formes verbales homophones en /E/ entre 8 et 15 ans: contraintes et conflits dans la construction des savoirs sur une difficulté orthographique majeure du français. *Langue Française, 151*, 74-93.

Bruck, M., Genesee, F., & Caravolas, M. (1997). A cross linguistic study of early literacy acquisition. In B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 145-162). Mahwah (NJ), Lawrence Erlbaum associates.

Brunet, E. (1981). *Le Vocabulaire français de 1789 à nos jours d'après les données du Trésor de la langue française*. Genève: Slatkine; Paris: Champion (Travaux de linguistique quantitative)

Bürki, A., Alario, F.-X., & Frauenfelder, U.H. (2011). Lexical representation of phonological variants: evidence from pseudohomophone effects. *Journal of Memory and Language, 64,* 424-442.

Bürki, A., Ernestus, M., Gendrot, C., Fougeron, C., & Frauenfelder, U. H. (2011). What affects the presence versus absence of schwa and its duration: A corpus analysis of French connected speech. *Journal of the Acoustical Society of America, 130,* 3980-3991.

Campbell, R., & Besner, D. (1981). This and THAP—constraints on the pronunciation of new, written words. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 33(4),* 375-396.

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

Caravolas, M., Kessler, B., Hulme, C., & Snowling, M. (2005). Effects of orthographic consistency, word frequency, and letter knowledge on children's vowel spelling development. *Journal of Experimental Child Psychology, 92,* 307–321.

Carlisle, J. F. (2000). Awareness of the structure and meaning or morphologically complex words: Impact on reading. *Reading and Writing, 12,* 169 –190.

Carlisle, J. F., & Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly, 40,* 428–449.

Carroll, J. B., Davies, P., & Richman, B. (Eds.). (1971). *The American heritage word frequency book.* Boston, MA: Houghton Mifflin.

Casalis, S., & Louis-Alexandre, M.-F. (2000). Morphological analysis, phonological analysis and learning to read French: A longitudinal study. *Reading and Writing: An Interdisciplinary Journal, 12*, 303–335.

Catach, N. (1980). *L'orthographe française, Traité théorique et pratique.* Nathan

Catach, N. (Dir.), Golfand, J., Mettas, O., Biedermann-Pasques, L., Dobovrie-Sorin, C, & Baddeley, S. (1995). *Dictionnaire historique de l'orthographe française*. Paris: Larousse.

Clin, E., Wade-Woolley, L., & Heggie, L. (2009). Prosodic sensitivity and morphological awareness in children's reading. *Journal of Experimental Child Psychology, 104*, 197-213.

Colé, P., Bouton, S., Leuwers, C., Casalis, S., & Sprenger-Charolles, L**.** (2012). Stem and Derivational-Suffix Processing During Reading by French Second and Third Graders. *Applied Psycholinguistics, 33,* 97-120.

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. *Psychological Review, 108,* 204–256.

Connine, C. M., Ranbom, L.J., & Patterson, D. J. (2008). Processing variants forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics, 70*, 403-411.

Content, A., Mousty, P., & Radeau, M. (1990). Brulex. Une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique, 90,* 551-566.

Côté, M.-H. (2007) Rhythmic constraints on the distribution of schwa in French. In J. Camacho, N. Flores-Ferrán, L. Sánchez, V. Déprez & M.J. Cabrera (Eds), *Romance Linguistics 2006* (pp. 79-92). Amsterdam: John Benjamins.

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.

Deacon, S.H., Campbell, E., Tamminga, M., & Kirby, J. R. (2010). Seeing the *harm* in *harmed* and *harmful*: Morphological processing in Grades 4, 6 and 8. *Applied Psycholinguistics, 31,* 759-775.

Deacon, S.H., & Kirby, J. R. (2004). Morphological awareness: Just "more phonological"? The roles of morphological and phonological awareness in reading development. *Applied Psycholinguistics, 25,* 223–238.

Dehaene, S. (Dir.), Dehaene-Lambertz, G., Gentaz, E., Huron, C., & Sprenger-Charolles, L. (2011). *Apprendre à lire: des sciences cognitives à la salle de classe*, Paris: Odile Jacob

Delattre, P. (1951). Le jeu de l'e instable intérieur en français. *The French Review*, *XXIV*, 4, 341-351.

Delattre, P. (1957). La question des deux "a" en français. *The French Review*, *XXXI*, 2, 141-148.

Dubois, J. (1967). Grammaire structurale du français: le verbe. Paris: Larousse (Collection langue et langage).

Durand, J., & Eychenne, J. (2004). Le schwa en français, pourquoi des corpus ? *Corpus 3,* 311-356.

Ellis, N. C., & Hooper, A. M. (2001). Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency-matched tests. *Applied Psycholinguistics, 22,* 571-599.

Ellis, N.C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Van Daal, V.H.P., Polyzoe, N., Tsipa, M.-L., & Patalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly, 39*, 438-468.

Fayol, M., & Pacton, S (2006). L'accord du participe passé: entre compétition de procédures et récupération en mémoire. *Langue Française, 151,* 59-73.

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 & Review, 16*, 684-691.

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading, 2*, 31-54.

Frost, R., Katz, L. & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multi-lingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 104-115.

Garabík, R. et al. (2007). A Cross-linguistic Database of Children's Printed Words in Three Slavic Languages. In Levická, J. and Garabík, R., editors, Slovko 2007. *Fourth International Seminar*. Bratislava, Slovakia, 25–27 October 2007, pages 51–64, Bratislava. Tribun.

Goswami, U., Gombert, J. E., & de Barrera, L. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics, 19*, 19–52.

Goswami, U., Ziegler, J.C., Dalton, L., & Schneider, W. (2001). Pseudohomophone effects and phonological recoding procedures in reading development in English and German. *Journal of Memory and Language, 45* , 648-664

Goswami, U., Ziegler, J.C., Dalton, L., & Schneider, W. (2003). Nonword reading across orthographies: how flexible is the choice of reading units? *Applied Psycholinguistics, 24*, 235-247.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: the orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning: Advances in psychology*, Vol. 94 (pp. 67–84). Oxford: North-Holland.

Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language, 44*, 592-617.

Largy, P., Cousin, M.-P., Bryant, P., & Fayol, M. (2007). When memorized instances compete with rules: The case of number-noun agreement in written French. *Journal of Child Language, 34*, 425-437.

Lavric, A., Clapp, A., & Rastle, K. (2007). ERP evidence of morphological analysis from orthography: A masked priming study. *Journal of Cognitive Neuroscience, 19*, 866-877.

Laxon V., Rickard M., Coltheart V. (1992). Children read affixed words and non-words. *British Journal of Psychology, 83*, 407-423.

Le Goffic, P. (1997). *Les formes conjuguées du verbe français: oral et écrit*. Ophrys.

Lété, B., Peereman, R. & Fayol, M. (2008). Consistency and word-frequency effects on spelling among first- to fifth-grade French children: A regression-based study. *Journal of Memory and Language, 58*, 952-977.

Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, & Computers, 36*, 156-166.

Longtin, C.-M., & Meunier, F. (2005). Morphological decomposition in early visual word processing. *Journal of Memory and Language, 53*, 26–41.

Longtin, C.-M., Segui, J., & Halle´, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes, 18*, 313–334.

Mahony, D., Singson, M., & Mann, V. (2000). Reading ability and sensitivity to morphological relations. *Reading and Writing, 12*, 191-218.

Marec-Breton, N., Gombert, J.-E., & Colé, P. (2005). Traitements morphologiques lors de la reconnaissance des mots écrits chez des apprentis lecteurs. *L'Année Psychologique, 105*, 9-45.

Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2003). *Children's Printed Word Database*. ESRC research project.

McCormick, S. F., Rastle, K., & Davis, M. H. (2008). Is there a "fete" in "fetish"? Effects of orthographic opacity on morpho-orthographic segmentation in visual word recognition. *Journal of Memory & Language, 58*, 307-326.

Meunier, F., Seigneuric, A., & Spinelli, E. (2008). The morpheme gender effect. *Journal of Memory and Language, 58*, 88-99.

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.

Pacton, S. & Casalis, S. (2006). L'utilisation d'informations morphologiques en production écrite rend-elle le cauchemar(d ?) des lettres muettes moins cauchemardesque ? *Rééducation Orthophonique, 225*, 129-144.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S.F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., Pesenti, S., Gallagher, A., Perani, D., Price, C., Frith,

C.D., & Frith, U. (2000). A cultural effect on brain function. *Nature Neuroscience, 3,* 91-96.

Peereman, R., & Content, A. (1997). Orthographic and phonological neighborhoods in naming: Not all neighbors are equally influential in orthographic space. *Journal of Memory and Language, 37*, 382–421.

Peereman, R., Lété, B., & Sprenger-Charolles, L. (2007). Manulex-infra: Distributional characteristics of grapheme–phoneme mappings, infra-lexical and lexical units in child-directed written material. *Behavior Research Methods, 39*, 593–603.

Peereman, R., & Content, A. (1998). *Quantitative analyses of orthography to phonology mapping in English and French*. Available from: <http://homepages.ulb.ac.be/~acontent/OPMapping.html>.

Peereman, R., & Content, A. (1999). LexOP. A lexical database with orthography–phonology statistics for French monosyllabic words. *Behavior Research Methods, Instruments, & Computers, 31*, 376–379.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*, 273–315.

Plaza, M., & Cohen, H. (2003). The interaction between phonological processing, syntactic awareness, and naming speed in the reading and spelling performance of first-grade children. *Brain and Cognition, 53*, 287–292.

Quémart, P., Casalis, S., & Colé, P. (2011). The role of form and meaning in the processing of written morphology: A priming study in French developing readers. *Journal of Experimental Child Psychology, 109*, 478-496.

Rabin, J., & Deacon, S. H. (2008). The representation of morphologically complex words in the developing lexicon. *Journal of Child Language, 35*, 453–465.

Racine, I., & Grosjean, F. (2005). Le coût de l'effacement du schwa lors de la reconnaissance des mots en français. *Revue Canadienne de Psychologie Expérimentale, 59*, 240-254.

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

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

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In Laurie B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Seidenberg, M. S. (2011). Reading in different writing systems: One architecture, multiple solutions. In P. McCardle, J. Ren, & O. Tzeng (Eds.), *Dyslexia across languages: Orthography and the Gene-Brain-Behavior link* (pp. 151-174). Paul Brooke Publishing.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23,* 383-404.

Sénéchal, M. (2000). Morphological effects in children's spelling of French words. *Canadian Journal of Experimental Psychology, 54*, 76-85.

Sénéchal, M., Basque, M.T., & Leclaire, T. (2006). Morphological knowledge as revealed in children's spelling accuracy and reports of spelling strategies. *Journal of Experimental Child Psychology, 95*, 231-254.

Seymour, P.H.K., Aro, M., & Erskine, J.M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143-174.

Shankweiler, D., Crain, S., Katz, L., Fowler, A. E., Liberman, A. E., Brady, S. A., Thornton, R., Lundquist, E., Breyer, L., Fletcher, J. M., Stuebing, K. K., Shaywitz, S. E., & Shaywitz, B. A. (1995). Cognitive profiles of reading-disabled children: Comparison of language skills in phonology, morphology and syntax. *Psychological Science, 6*, 149–156.

Sprenger-Charolles, L. (2004). Linguistic processes in reading and spelling. The case of alphabetic writing systems: English, French, German and Spanish. In T. Nunes, & P. Bryant (Eds) *Handbook of children's literacy*. (pp. 43-65). Dordrecht: Kluwer Academic Publishers

Sprenger-Charolles, L., Siegel, L. S., Béchennec, D., & Serniclaes, W. (2003). Development of phonological and orthographic processing in reading aloud, in silent reading and in spelling: A four year longitudinal study. *Journal of Experimental Child Psychology, 84*, 194-217.

Sprenger-Charolles, L., Siegel, L.S., & Bonnet, P. (1998). Phonological mediation and orthographic factors in reading and spelling. *Journal of Experimental Child Psychology*, 68, 134-155.

St-Pierre, M.-C., & Béland, R. (2010). Reproduction of inflectional markers in French-speaking children with reading impairment. *Journal of Speech, Language, and Hearing Research, 53*, 469–489.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition, 7*, 263–272.

Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes, 9*, 271-294.

Ters, F., Mayer, F., & Reichenbach, D. (1977). *L'échelle Dubois-Buyse d'orthographe usuelle française*. Paris: OCDL (5e édition).

Thévenin, M.-G., Totereau, C., Fayol, M., & Jarousse, J.-P. (1999). L'apprentissage/enseignement de la morphologie écrite du nombre en français. *Revue Française de Pédagogie, 126*, 39-52.

Totereau, C., Barouillet, P. & Fayol, M. (1998). Overgeneralizations of number inflections in the learning of written French: the case of noun and verb. *British Journal of Developmental Psychology, 16*, 447–64.

Totereau, C., Thévenin, M. G. & Fayol, M. (1997). Acquisition de la morphologie du nombre à l'écrit en français. In L Rieben, M. Fayol & C. Perfetti (eds), *Des orthographes et leur acquisition*, (pp. 147–165). Lausanne: Delachaux & Niestlé.

Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology, 98*, 642-652

Treiman, R., Kessler, B., & Bick, S. (2003). Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models. *Cognition, 88*, 49-78.

Treiman, R., Mullenix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124*, 107-136.

Véronis, J. (1986). Etude quantitative sur le système graphique et phono-graphique du français. *Cahiers de Psychologie Cognitive, 6*, 501-831.

Véronis, J. (1988). From sound to spelling in French: Simulation on a computer. *Cahiers de Psychologie Cognitive, 8*, 315-334.

Waters, G. S., Seidenberg, G. M. S., & Bruck, M. (1984). Children's and adults' use of spelling sound information in three reading tasks. *Memory & Cognition, 12*, 293-305.

Weekes, B. S., Castles, A. E., & Davies, R. A. (2006). Effects of consistency and age of acquisition on reading and spelling in developing readers. *Reading and Writing: An Interdisciplinary Journal, 19*, 132–169.

Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: word recognition in English and German children. *Cognition, 51*, 91-103.

Zeno, S. M., Ivenz, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.

Ziegler, J. C., & Goswami, U. (2006). Becoming literate in different languages: similar problems, different solutions. *Developmental Science, 9*, 429-436.

Ziegler, J. C., & Goswami, U. C. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3-29.

Ziegler**,** J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., et al. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science, 21*, 551–559

Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analyses of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, & Computers, 28*, 504-515.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What is the pronunciation for –ough and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers, 29*, 600-618.

APPENDIX A

Major changes to the phonetic notations of Manulex_infra

A first series of modifications were implemented to standardize the coding of vowels associated with several pronunciations.

- The opposition between / a / and / ɑ /: As already noted by Delattre (1957), large variations exist for the pronunciation of A among French speakers, with a trend towards a reduction of the posteriorization of the vowel. However, the /ɑ/ pronunciation remains generally attached to the spelling “â”. This relationship has been systematized in the corpus. Cases of “a” are coded /a/, and “â” is coded /ɑ/ (65 forms). Instances of “oi” are coded /wa/.

- The distinctions between /o/ and /ɔ/, /e/ and /ɛ/, and /ø/ and /œ/ are mainly related to the opening/closing of the syllable. The spellings “eau” and “ô”, as well as (with some exceptions) the spellings “au” and “o” in open syllables, correspond to the closed vowel /o/. The spelling “o” in closed syllables is usually coded with the open vowel /ɔ/. The grapheme “é” is coded /e/ (with a few exceptions such as *événement,* event), and “è”, “ê”, “ais”, “ait”, and “aît” are pronounced /ɛ/. The graphemes “ë”, “ai”, “aî”, “ay”, “ei”, and “ey” are (with some exceptions) coded /e/ in open syllables, and /ɛ/ in closed syllables. With a few exceptions (e.g., *peuple, œuvre,* people, work), the graphemes “eu” and “oeu” are coded /ø/ in open syllables, and /œ/ in closed syllables.

A second series of modifications were introduced to distinguish between the different cases of silent “e” (or “mute e”): the obligatory schwa (always pronounced), the deleted

schwa (never pronounced), and the optional schwa (both options co-exist). Many linguistic studies have been conducted to define a set of rules governing the pronunciation vs. deletion of the schwa (e.g., Anderson, 1982; Delattre, 1951), as well as the fluctuations related to regional variations (Durand & Eychenne, 2004) and prosodic constraints (Côté, 2007). In the case of words with optional schwa, some recent psycholinguistic studies have led to the suggestion that both variants (dominant and subdominant) of the words with optional schwa are lexically represented (Bürki, Alario, & Frauenfelder, 2011; Connine, Ranbom, & Patterson, 2008; Racine & Grosjean, 2005). The coding adopted in the corpus was according to the following principles:

- Obligatory schwa. It occurs at the end of words when preceded by a consonant cluster of increasing sonority, like occlusive + liquid or fricative + liquid (e.g., *souple*, *chèvre*, soft, goat). Mid-word, the schwa is obligatory when preceded by two consonants and followed by another consonant (law of the three consonants; e.g., *marguerite*, *vendredi*, daisy, Friday). Finally, at the beginning of words, deletion of the schwa would lead to an illegal consonant cluster (e.g., /Rn/ in *renard*, fox; /ls/ in *leçon*, lesson) or to an impossible cluster, such as /ss/ in *ceci* (this).
- Deleted schwa. When the “e” at the end of a word is preceded by a vowel that is followed by an “e” which either is a morphological inflection (e.g., *amie*, friend, fem. gender) or is not (*foie, roue, copie, laitue*, liver, wheel, copy, lettuce). In few cases in an internal syllable (*gaiement, asseoir*, gaily, to sit).
- Optional schwa. In the absence of behavioural data including a large number of tokens that can be used to evaluate the contributions of phonological and phonotactic constraints to the relative weights of the variants (with or without schwa; but see Bürki, Ernestus, Gendrot, Fougeron, & Frauenfelder, in press, for a recent study in this direction), we coded all cases not meeting the above criteria as optional schwa.

Authors' note

Corresponding author: Ronald Peereman, LPNC-UMR 5105, Université Pierre Mendes-France, BP 47, 38040 Grenoble Cedex 9, France. This study was partially funded by a grant to LSC from the labex (laboratoire d'excellence) EFL (Empirical Foundations of Linguistics), University Paris Descartes, Laboratoire Psychologie de la Perception, UMR 8158 (PRES Sorbonne-Paris-Cité), France. We would like to thank two anonymous reviewers for their comments and suggestions. The authors are grateful to Paul Reeve for his assistance with the English.

Table 1. Percentages of the different grammatical categories in Manulex (Lété et al., 2004) and Manulex_morpho

	Verb	Noun	Proper name	Adjective	Adverb	Pronoun	Deter-minant	Preposi-tion	Conjunc-tion
Manulex	39.7	32.4	9.1	16.8	1.5	0.3	0.1	0.1	0.1
Manulex-morpho	34.5	44	3.8	13.5	2.6	0.8	0.3	0.4	0.2

Table 2. Graphemes morphologically marked in Manulex_morpho

Markers	Morphographemes
Gender and number	-e, -s, -x
Verb inflections	*Infinitive form*: -er
	Present participle: -ant
	Past participle: -é, -i, -u, -is, -t
	Other conjugated forms:
	-e, es, ez, -ent, -est, -eu
	-a, -as, -ai, -ais, -ait
	-i, -is, -it, -ons, -ont
	-u, -û, -ût
	-d, -ds, -pt, -s, -t, -ts, x
Adverb ending in *–ment*	-(m)ent
Ends of words heard in inflected/derived forms	-er, -et, -êt
	-an, -aen, -ean, -en, -enn
	-ain, -aim, -in, -ein, (-en)
	-on, -om, -oin
	-un, -um
	-il
	-b, -c, -ch, -ck, -d, -f, -g, -gu, -l, -ll, -m,
	-n, -nn, -p, -q, -r, -s, -ss, -t, -tt, -v, -w, -x, -z

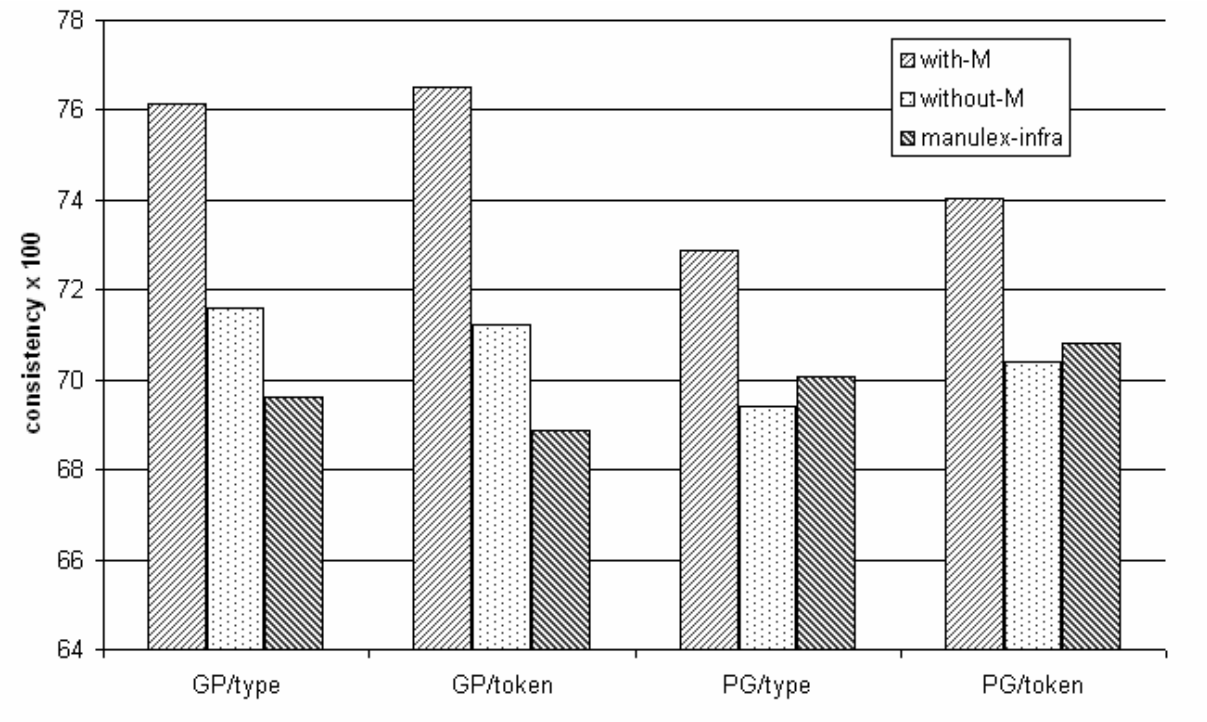
FIGURE CAPTIONS

Figure 1. Mean consistency (multiplied by 100) observed in the three analyses *(with-M, without-M, manulex-infra)* for GP and PG mappings (estimated by type and token).

Figures 2a and 2b. Mean consistency (multiplied by 100) observed in the with-M and without-M analyses for GP and PG mappings, as a function of grapheme and phoneme position (init: initial, int: internal, end: end) in words. Consistency values are computed by type (Figure 2a) and by token (Figure 2b).

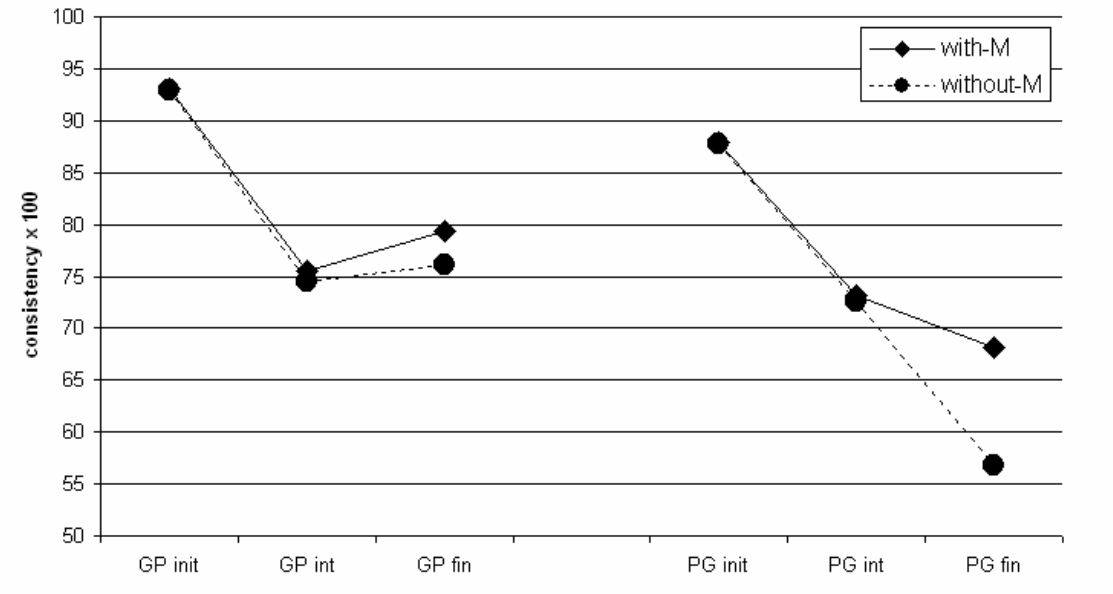
Figure 3. Increases in mean consistency (multiplied by 100) observed in the with-M analysis (in comparison to the without-M analysis) for final graphemes and phonemes, as a function of morphological markers (by-type estimations).

FIGURE 1


78
76
74
72
70
68
66
64
consistency x 100
with-M
without-M
manulex-infra
GP/type
GP/token
PG/type
PG/token

FIGURES 2a and 2b

2a



2b

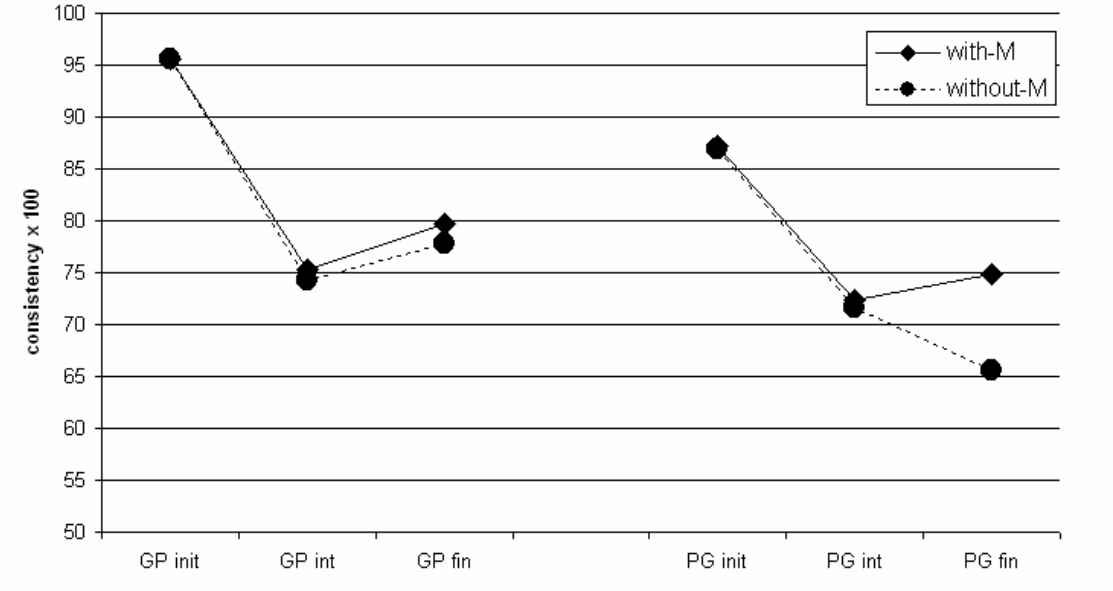
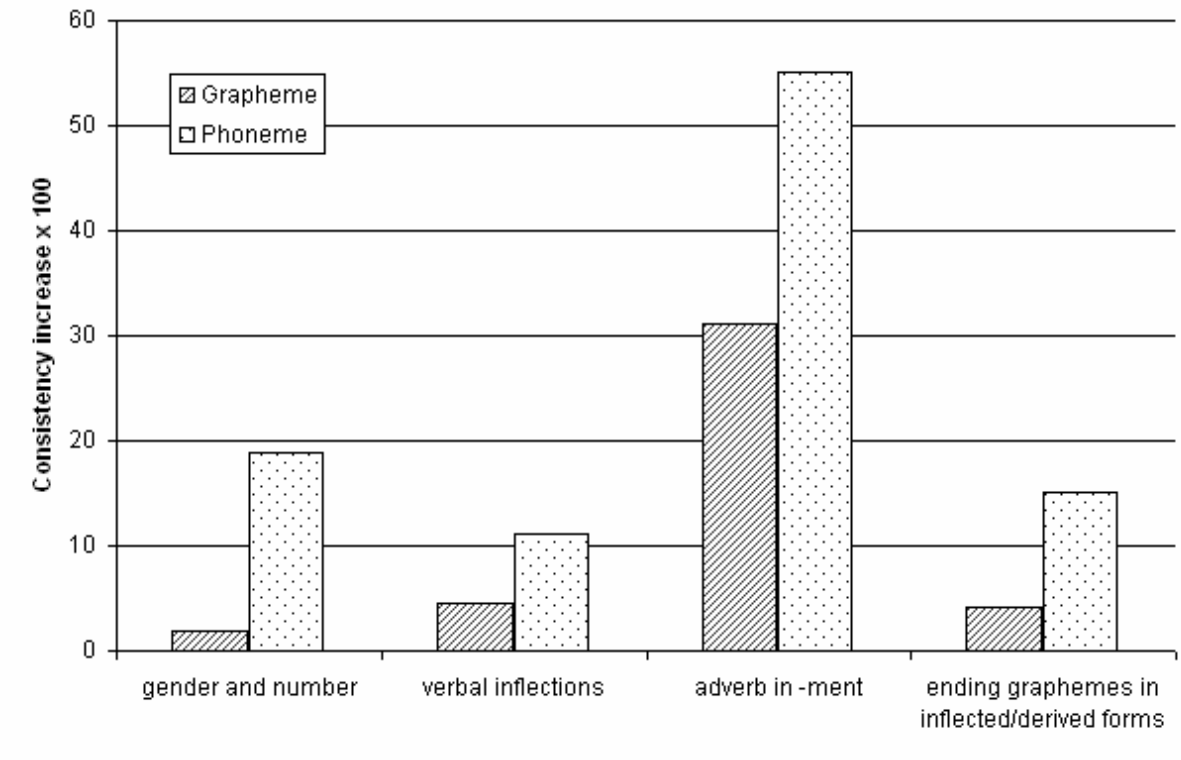


FIGURE 3


60
50
40
30
20
10
0
Consistency increase x 100
Grapheme
Phoneme
gender and number
verbal inflections
adverb in -ment
ending graphemes in inflected/derived forms