

# What can a dictation corpus teach us about orthographic variation and the mastery of context-sensitive spellings?

Anne ~~Dister~~ and Marie-Louise Moreau  
UCLouvain | UMon

*La dictée du Balfroid* is a very popular dictation contest in French-speaking Belgium. Since its first edition in 1988, thousands of children in the 6th year of primary school have taken part in this event, which meets a large media audience.

In this article, we analyze a corpus of some 11,000 dictations produced during this championship. We focus on orthographic variation and phoneme-grapheme correspondences, particularly for context-sensitive graphemes such as *c* or *g*.

**Keywords:** *Acquisition of French spelling by children, phonographemic matches, context-sensitive graphemes*

## Introduction

*La dictée du Balfroid* is an institution in French-speaking Belgium. Since its first edition in 1988, thousands of children in the 6th year of primary school have taken part in this famous spelling competition. Yet, despite its success and the recurrent speeches about the spelling crisis made by the competition's creator herself, no study has ever been made of the results obtained. Indeed, while we know that every year, some school children obtain a “no-fault” mark, and that others, despite being motivated and trained by their teachers, accumulate more than a dozen errors, we do not know, for example, the average number of errors per pupil, let alone what kind they are. Are they errors of agreement? If so, what type of agreement? Stumbling over double consonants or accents? Problems with silent, added or omitted letters? Confusion between several possible spellings for the same sound? Other questions also arise: when a pupil makes a particular type of error, does he or she do so systematically? Are certain errors correlated? Our Ortho-cor research project aims at answering these and other questions. If a reform of

French spelling were to be considered, and if it were to address the main points on which users most often stumble, it could be based on studies of this kind.

In this article, we present some results on overall performance. We then analyze pupils' mastery of the relationship between spelling and phonics, and finally examine cases where this relationship is subject to context-sensitive rules.

## 1. A corpus of dictations

*La dictée du Balfroid* is a spelling competition open to all children in the 6th year of primary school in French-speaking Belgium. Pupils who have not repeated a year or moved up one or more grades are 11 or 12 years old at the time of the test. The competition, which was first held in 1988, is sponsored by a number of official institutions, and is supported by a press group that reports on each of the meetings. Each year, six semi-finals (one for each Walloon province and one for Brussels) are held to select the school children who will advance to the final. They are only allowed to make a maximum of 5 mistakes in the eliminatory round, for which no further selection is made.

As part of our research, we have collected dictations that took place between 2012 and 2016. In all, there were 35 different dictation texts (6 semi-finals and 1 final per year, for 5 successive years), averaging 132 words,<sup>1</sup> with a total of 11,298 completed dictations (see Table 1). We then set up a database, built according to the procedures described in [Dister](#) and Moreau (2017).

Unlike other dictation analyses (e.g. Chervel and Manesse 1989<sup>2</sup> or Manesse and Cogis 2007), we do not have any sociolinguistic information, either on the pupils' profiles (gender, educational trajectory – are they “on time” or not, etc.) or on the schools they attend. The answer sheets of the dictations are strictly anonymous. We are therefore unable to examine the data on the basis of such independent variables.

The school children who took part in these dictations cannot be assimilated to the general school population of their level: they are participants in a spelling competition, who are either designated by their school, with a maximum of 4 pupils per class, or registered individually. We can assume that the teachers responsible for the initiative are those who pay particular attention to teaching

---

1. The texts vary in length, from 101 to 182 words.

2. Chervel and Manesse not only have the dictations, but also the inspection reports of Gaspard Beauvain d'Altenheim, the Inspector General of Primary Education who introduced the dictations from 1873 onwards. The dictations can thus be related to categories of schools and categories of pupils.

spelling, and that they did not choose the least successful of their pupils. Nevertheless, performance is far from good, as we shall see in the next section.

2. Overall poor results

Although the competition is very popular, and the winners are praised in the press, no presentation of the overall results is ever given. In the following table, we answer the question “What proportion of completed dictations contain between 0 and 4 mistakes, between 5 and 9, etc.?” We can see that there are sometimes significant differences from one dictation to the next, but also that answer sheets with few errors, between 0 and 4, are always in the minority. The averages, like the shades of gray, show that the majority of pupils (62.3%) made between 5 and 14 mistakes, and one in two (50.2%) made 10 or more.

Table 1. Percentages of pupils by number of errors

|                   | N answer<br>sheets | N<br>words | % of pupils   |               |                 |                 |                      |
|-------------------|--------------------|------------|---------------|---------------|-----------------|-----------------|----------------------|
|                   |                    |            | 0-4<br>errors | 5-9<br>errors | 10-14<br>errors | 15-19<br>errors | 20 errors<br>or more |
| 12D-Avril         | 325                | 133        | 2.8           | 23.1          | 34.2            | 20.9            | 19.1                 |
| 12D-Bonbons       | 320                | 121        | 13.1          | 36.6          | 28.4            | 13.8            | 8.1                  |
| 12D-Ecureuil      | 349                | 159        | 11.2          | 26.9          | 25.8            | 15.2            | 20.9                 |
| 12D-Electroniques | 368                | 124        | 5.4           | 28.0          | 31.8            | 17.9            | 16.8                 |
| 12D-Puie          | 303                | 138        | 12.5          | 24.1          | 26.1            | 19.5            | 17.8                 |
| 12D-Lumières      | 315                | 131        | 8.6           | 26.7          | 33.3            | 18.1            | 13.3                 |
| 12F-Anniversaire  | 312                | 171        | 15.1          | 45.2          | 26.0            | 11.9            | 1.9                  |
| 13D-Bonjour       | 324                | 125        | 9.3           | 31.8          | 26.2            | 16.7            | 16.0                 |
| 13D-Oiseaux       | 395                | 120        | 15.4          | 32.9          | 27.1            | 12.7            | 11.9                 |
| 13D-Paraît        | 385                | 127        | 27.3          | 42.1          | 20.8            | 6.2             | 3.6                  |
| 13D-Retour        | 307                | 141        | 15.0          | 36.5          | 30.3            | 11.7            | 6.5                  |
| 13D-Téléphone     | 437                | 138        | 18.8          | 32.3          | 29.5            | 12.4            | 7.1                  |
| 13D-Voisinage     | 238                | 131        | 13.9          | 33.2          | 20.6            | 14.7            | 17.6                 |
| 13F-Tunnel        | 618                | 162        | 17.3          | 49.8          | 24.6            | 6.0             | 2.3                  |
| 14D-Beautemps     | 237                | 124        | 23.6          | 34.6          | 25.3            | 10.5            | 5.9                  |
| 14D-Cinéma        | 276                | 126        | 12.3          | 29.3          | 30.1            | 14.9            | 13.4                 |
| 14D-Génération    | 313                | 139        | 2.6           | 31.9          | 34.5            | 17.6            | 13.4                 |

Table 1. (continued)

|                  | N answer<br>sheets | N<br>words | % of pupils   |               |                 |                 |                      |
|------------------|--------------------|------------|---------------|---------------|-----------------|-----------------|----------------------|
|                  |                    |            | 0–4<br>errors | 5–9<br>errors | 10–14<br>errors | 15–19<br>errors | 20 errors<br>or more |
| 14D-Hiver        | 315                | 122        | 16.2          | 34.0          | 25.4            | 15.2            | 9.2                  |
| 14D-Olympique    | 253                | 102        | 19.0          | 45.8          | 23.3            | 9.1             | 2.8                  |
| 14D-Panne        | 370                | 133        | 17.8          | 39.5          | 23.0            | 11.6            | 8.1                  |
| 14F-Ailes        | 448                | 151        | 10.3          | 41.1          | 33.5            | 12.5            | 2.7                  |
| 15D-Arbres       | 416                | 130        | 9.6           | 27.6          | 26.7            | 17.8            | 18.3                 |
| 15D-Espoir       | 237                | 112        | 21.9          | 46.8          | 22.8            | 5.9             | 2.5                  |
| 15D-Lever        | 359                | 139        | 13.6          | 30.9          | 26.2            | 13.6            | 15.6                 |
| 15D-Mars         | 259                | 101        | 14.3          | 52.1          | 22.4            | 7.3             | 3.9                  |
| 15D-Merveilles   | 192                | 126        | 27.1          | 27.6          | 23.4            | 12.5            | 9.4                  |
| 15D-Perd         | 248                | 120        | 13.3          | 46.0          | 27.4            | 10.9            | 2.4                  |
| 15F-Fais         | 425                | 182        | 11.1          | 34.1          | 32.9            | 15.1            | 6.8                  |
| 16D-Alimentation | 283                | 134        | 12.4          | 37.5          | 28.3            | 12.0            | 9.9                  |
| 16D-Fenêtres     | 279                | 143        | 7.2           | 30.5          | 30.8            | 15.8            | 15.8                 |
| 16D-Gelé         | 312                | 118        | 10.3          | 38.1          | 23.1            | 18.3            | 10.3                 |
| 16D-Ordinateur   | 234                | 107        | 16.7          | 32.9          | 28.2            | 12.8            | 9.4                  |
| 16D-Perspective  | 265                | 111        | 20.0          | 44.2          | 21.9            | 9.1             | 4.9                  |
| 16D-Tempête      | 211                | 122        | 32.7          | 27.0          | 17.5            | 13.7            | 9.0                  |
| 16F-Fourche      | 370                | 157        | 8.1           | 35.1          | 34.3            | 17.6            | 4.9                  |
| Total/Average    | 11298              |            | 14.4          | 35.3          | 27.0            | 13.5            | 9.8                  |

The texts are not particularly difficult. Admittedly, the vocabulary used sometimes deviates from the *Vocabulaire orthographique de base* (Ters et al. 1968; henceforth VOB), but they do not contain those rare words (e.g. *amphitryon*, *arrhes*, *chrysanthème*) that are so popular in championships of this type. Similarly, while the pupils have to agree past participles on every occasion, there are no pronominal verb forms or special cases. What the results do show, however, is that the majority of primary school leavers still encounter substantial spelling difficulties. This is in line with data collected in other studies. To take only the case of agreement of epithets with the noun, many children are still experiencing difficulties at the junction of primary and secondary school, or even later.<sup>3</sup>

3. For an overview of studies on this subject and their results, see Dister and Moreau (2023, 76 et sq.).

Most of the words used in our dictations are also subject to a wide range of spelling and grammatical variations. By way of illustration, we have selected the noun *perce-neige*, used in the plural in 4 dictations, i.e. in 1422 answer sheets. This compound noun is affected by the 1990 spelling rectifications. The new rule proposes the plural *perce-neiges*, the traditional spelling being the same in the singular and plural: *perce-neige* (Dister and Moreau 2019). In addition to the correct spellings *perce-neiges* (new spelling) and *perce-neige* (traditional spelling), for the 4 dictations concerned, there are 71 different forms, with accents and the presence or absence of a hyphen accounting for a significant proportion of the variation: *pairsonnaiges*, *perce neige*, *pêrce neiges*, *perce neige*, etc. While a significant proportion of the variation observed here is linked to the plural and the hyphen, the same cannot be said for the adjective *insouciant*, for which the 279 pupils who took part in the 16D-Fenêtres dictation used 52 different ways: *asociente*, *assouciante*, *assouciencie*, *assoussiante*, etc. (The full list of variants is provided in the Appendix.)

### 3. The phonics-graphics relationship

If overall results are far from good, and school children fail to meet the institutionally-validated spelling standard, does this mean that they have not mastered the link between phonics and spelling?

While *perse-neige* and *persenèges* do not correspond to valid spellings, in the sense that they are faulty from the point of view of the evaluative standard, these two transcriptions are nonetheless possible according to the orthoepic rules of French, and their oralisation does indeed result in the starting form /pɛʁsənɛʒ/; the same goes for *insousciant* and *insousiante*, which are oralised /ɛ̃susjɑ̃t/.

It is this relationship between phonics and grapheme that we will now examine. To do this, we have categorized the words of the 6 dictations proposed for the 2016 semi-finals, i.e. 1584 answer sheets, into different sets depending on whether or not the oralization of the graphemes adopted by the school children coincided with the oral target:

- We have considered as deviations productions such as *pénètre* > *penaitre*, *renouveau* > *renouvant*, *valorisé* > *falorisé*, *insignifiant* > *insinifiant*. The grapheme-phoneme correspondences do not allow us to arrive, when we read them, at the same phonic sequence as the target. We focused on this category.
- On the other hand, forms such as *accentuaient* > *axentuet*, *certain* > *serteint*, *éphémère* > *éfémaire*, *insouciant* > *inssoussiante*, *foisonnant* > *foizonants*, *inconsciemment* > *inconciament*, *pénètre* > *pénaitre* were treated as conform-

ing to the oral target: they may deviate from the norm, but reading them aloud would not distinguish them from the target.

The percentage of deviations, calculated on the number of phonemes, is always less than 1% in the 6 dictations, ranging from 0.3% to 0.8%. The principle of alphabetical transcription is therefore well integrated. Children are familiar with phoneme-grapheme correspondences, and know what possibilities are available for transcribing the various phonemes (Sprenger-Charolles et al. 2003). If French orthography were like that of Finnish, Italian, Spanish, Dutch, German and many other European and non-European languages, essentially phonological, its mastery would be assured without any problem, and certainly well before primary 6. This is indeed what Seymour et al (2003) have already indicated in a study of the reading skills of children speaking ten European languages.

In strictly phonological scripts, one phoneme and one only is associated with one grapheme and one only, and vice versa, as in the International Phonetic Alphabet. In French, however, this is far from the case: a single phoneme is often associated with several graphemes, and vice versa,<sup>4</sup> although the range of solutions is smaller in the latter case. The result is an asymmetry between reading and writing: the oral meaning of a written word is more predictable than the written meaning of an oral word. What our data indicate is that the children whose output we are analyzing are unlikely to have difficulty in reading.<sup>5</sup> But when they write, they have to determine which of the possible graphemes pertains to the words they have to transcribe, and this is one of the stumbling blocks of our orthographic system.

There remains that tiny proportion of cases where children use a grapheme that does not have the desired phonological value. Which correspondences seem to present the most difficulties? The breakdown of the various deviations into categories (according to whether, for example, the /s/ phoneme in *poisson* is transcribed incorrectly by a single *s*, or by *c*, or by *t*..., or whether a *u* takes the place of an *ou*, etc.) quickly shows that the figures are closely dependent on the particular words used in the dictated text: we can obviously observe no difficulty with /aj/ endings, if the text contains neither *caille*, nor *épouvantail*, etc. As Table 2 shows, the discrepancies cluster around certain words: 10 of the 100 to 140 words in these dictations account for 57% to 89.1% of the discrepancies. If we take into account the 15 forms most frequently at fault, we find 60.3% to 93.4% of discrepancies. Let

4. The phoneme /s/, for example, is transcribed by *s* in *soulier*, *ss* in *passer*, *c* in *ceci*, *ç* in *lançait*, *t* in *potion*, *sc* in *scier*, *x* in *dix*; the grapheme *s* is read /s/ in *soulier*, /z/ in *poison*, or it is silent (*pas*), the choice between the possibilities generally being dictated by the context.

5. We are talking here about deciphering words, not accessing their meaning. A fortiori, not access to the meaning of the texts in which they appear.

us be clear: it is not the spelling mistakes we are talking about here, but only those cases where the pupils’ output does not conform to the correspondences between phonemes and graphemes laid down in the code — in other words, cases where, when we pronounce what the child has written, we do not end up with the oral signifier of the target. The words concerned are not among the most familiar, and many of them (indicated below by an asterisk) do not appear in the VOB.

**Table 2.** Words for which school children least respected phoneme-grapheme correspondences in 2016 dictations (percentages are calculated on total phonographemic deviations)

| Dictation        | Words with the most deviations                                                                                                                                                                                                                                                 | % for the first 15 | % for the first 10 |
|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|--------------------|
| D16-Alimentation | quotidiennement*(18.2), alléchants* (17.0), terrasse* (16.8), réjouissant* (7.8), alternatives* (7.1), aménager* (6.9), souhaitée (3.9), désormais (3.0), pimpant* (1.6), adresser (1.6), lointains (1.6), sainement* (1.6), condiments* (1.1), trouverez (0.9),employés (0.9) | 90.1               | 83.9               |
| D16-Fenêtre      | insignifiant* (17.6), effaçant (13.0), insouciant* (7.9), gaieté (7.4), pénètre (7.0), aubépines* (5.0), jonquilles (4.4), témoins (4.0), bosquets (3.3), inaperçues* (3.3), dès (3.3), dépouillés (2.3), précoce* (2.9), immaculé* (1.6), bulbes* (1.5)                       | 83.6               | 72.8               |
| D16-Gelé         | éphémère* (24.7), gazouillis* (13.3), déçus (5.4), effet (3.4), inconsciemment* (2.7), accentuait* (2.3), croyaient (1.8), clarté* (1.4), gelé (1.2), était (1.1), crocus* (0.8), gel* (0.8), été (0.7), frileux* (0, 5), brillait (0.4)                                       | 60.3               | 57.2               |
| D16-Ordinateur   | intéressants (29.3), aisément* (17.7), semblaient (11.4), attellent* (5.5), soigneusement (5.2), éclaircissent (4.8), né (3.0), hebdomadaires* (2.0), désormais (2.0), est (2.0), complément (1.6), indispensable (1.6), bénéficier* (1.4), obscurs (1.4), révolue* (1.3)      | 90.4               | 81.0               |
| D16-Perspective  | impeccablement* (23.1), désespéraient* (21.6), ressortis (14.7), élançaient (11.7), réjouï* (6.5),                                                                                                                                                                             | 93.4               | 89.1               |

Table 2. (continued)

| Dictation   | Words with the most deviations                                                                                                                                                                                                                                                 | % for the first 15 | % for the first 10 |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|--------------------|
|             | perspective* (4.4), engins* (2.6), joues (2.1), représentait (1.3), loueurs* (1.3), espoirs (1.1), utilisés (0.9), attendaient (0.9), étaient (0.8), comblés (0.6)                                                                                                             |                    |                    |
| D16-Tempête | déchainés* (18.0), violemment* (10.9), annonçaient (9.6), flamboyant* (8.5), empressant* (5.5), résistaient* (4.4), pressés (4.4), dépassait (4.1), tenaient (3.8), malgré (3.0), éléments* (2.7), sensations* (2.5), craindre (2.2), retournaient (1.9), incorrigibles* (1.4) | 82.8               | 72.1               |

But are we really dealing with an incomplete mastery of the correspondences between phonemes and graphemes? In fact, we have to assume that when children write a text under dictation, they are not transposing the oral form they hear into written form. As with adults, the function of the oral form is to activate a sign integrated into their mental lexicon, and it is the signifier of this internalized sign that they transcribe. It is not a coincidence that concentrations of deviations are found in words that children are unfamiliar with, both in their written and sometimes oral form. As their mental lexicon does not contain these words, children try to recompose what they have heard, not without some distortion: our corpus thus offers us cases such as, for example, *aubépines* > *aubutines*, *bosquets* > *paussetés*, *immaculé* > *inafumé*, *insouciant* > *inssidence*, *perspective* > *perpsectise*, *perpextive*.<sup>6</sup> If 20 children, in dictation 16D-Fenêtres, write *aubépines* > *opépines*, *aupépines*, *oppépines*, etc., we cannot infer that they confound voiced and unvoiced consonants, or that they are unaware that /b/ is transcribed by the grapheme *b*, and not by *p*; such confusions are not illustrated in other words within the same dictation. Above all, these errors show how they react to unfamiliar or little-known words: by linking them to neighboring units they know (in the last example, *pépin*?),<sup>7</sup> and by relying on the intuitive knowledge they may have of

6. The phenomenon is clearly visible in the spellings used by pupils when writing some of the ten difficult words, which must be used to decide between ex-aequos. Many of these words surely have the same status for pupils as logatomes, fictitious words: *okoumé* > *hocumé*, *acoumèe*, *ocouné*, *occumer*, *ocomer*, *aucoubé*, *hopimmée*, *augoumé*, *agoumer*... all forms whose oralisation does not correspond to that of the target.

regularities in the distribution of phonemes. A dictation thus sometimes tests not only spelling, but also the lexical repertoire, in both its oral and written components.

However, another hypothesis must be reserved to explain these distortions: some children may suffer from hearing loss, or at the very least not hear well what is dictated, especially as the acoustic conditions in which the tests take place are not always optimal.

The signs available to the children sometimes differ from the norm. The corpus contains several cases of reduction of consonantal groups (*arbre* > *arbe*), or, a phonic peculiarity widespread in French-speaking Belgium, substitution of a deaf final consonant for a sound one (*perspective* > *perspectif*). Similarly, the corpus contains several cases of final *-ij/* transcribed *ie* (*jonquille* > *jonquie*).<sup>8</sup> But the phenomenon is quite marginal in the dictations we have studied.

Analysis of our data shows that by the end of primary school, pupils have a good command of phoneme-grapheme correspondences. In our dictations, the percentage of discrepancies, calculated on the number of phonemes, is always less than 1%. Even if they do not necessarily write as the orthographic norm would require, children use a plausible target grapheme to transcribe the source phoneme.

#### 4. How do pupils handle graphemes whose value is context-sensitive?

The conversion of phonemes into graphemes must comply with phonographic correspondences, but also with certain rules specific to the written word, which determine how graphemes are combined, independently of the transcribed phonemes.

Thus, nasal vowels are transcribed with an *n*, but with an *m* before the graphemes *p* and *b* (e.g. *bombe*, *pompe*), the grapheme *q* is always followed by a *u*, except in *cinq* and *coq*, and a few words used only by scrabble fans (*maqam*, *qat...*) (cf. Dister and Moreau ~~2023~~ forthcoming). That being said, whether you write *an* or *am*, you transcribe always the same nasal vowel. The same is not true of the treatment of *c* and *g*, which are associated with different phonemes depending on whether they are followed by *a*, *o*, *u* or *e*, *i*. To study how children use these graphemes, our work here covers all 35 dictations, i.e. 11,298 answer sheets.

7. Other pupils use *aubépines* > *dos d'épines*, *abondamment* > *à bonds d'amants*, *éphémères* > *effet-maire*, and so on.

8. Less frequently, we also find cases such as *carie* > *carille*.

#### 4.1 C, ç and cu

In this and the next section, we will use the capitals *A, E, I, O, U*, to refer to vowels with or without accents. For example, *E* refers to *e, é, è* or *ê*.

The grapheme *c* transcribes the phoneme /k/ before *A, O, æ*<sup>9</sup> and *U* (*car, cor, cœur, cuit*), or the phoneme /s/ before *E, I* and *y*, (*cent, cil, cygne*). However, it can also refer to /s/ before *A, O* and *U*, if given a cedilla (*façade, soupçon, déçu*), and to /k/ before *E*, if combined with *u* (*cueille*). So you need to pay attention to both the strictly phonic (/k/ or /s/?) and the strictly graphic dimensions, since you need to take into account the grapheme that follows, regardless of its phonic value (*coi, coin, con, cor, cour* VS *leçon, perçois, suçoter*). The functioning of *c* and *ç* is based on a double asymmetry: on the one hand, without a cedilla, *c* can be used in all environments, with different phonic values; on the other hand, with a cedilla, it cannot be used before *E, I* or *y*; while the phonic correspondent of *c* depends on the context, *ç* always refers to /s/. The situation is therefore complex. As they relate to the relationship between phonics and writing, and to the phonological and alphabetical principles of spelling (Bousquet et al. 1999; Fayol and Jaffré 1999 and 2008), the uses of *c* and *ç* are part of basic learning. However, their distribution is not integrated by all, even among adults (Lucci and Millet 1994, 107 ff.). What do we observe in our completed dictations?

Our scope was limited to cases where the target has a *c* before a vowel, whether it corresponds to /k/ or /s/, and where children use a *c* to transcribe this phoneme. Excluded from the survey are targets such as *action*, or *lac*, and spellings like *placer* > *plasser*, *récolte* > *réquolte*. For the 477 units<sup>10</sup> meeting these criteria, we focused on the phonological value of the *c* in the target, and on how it was written (with or without a cedilla, followed or not by a *u*). The results are shown in the table below.<sup>11</sup> Percentages are calculated on the sum of spellings with *c, ç* and *cu* (the number of which is given in the “*n*=” line).

Let us look first at the left-hand side of the table, which shows the use of *c, ç* or *cu* to transcribe /k/. The penultimate line of the table shows that almost all pupils have understood that /k/ cannot be transcribed by *ç*. Exceptionally, deviations with a cedilla only appear in the second decimal place of the percentage.

9. Only one word has an *æ* (*coeur*). We have treated it as if it were an *o*, especially since handwritten scripts make a poor distinction between *æ* and *oe*.

10. We are talking about units, not words: we have provided two entries for words with two *c* (*concours, confiance...*), and we also take into account the spelling adopted by the school children when *c* is followed by a vowel other than the target (*convaincante* > *convaincente*), which slightly increases the volume processed.

11. Shaded boxes indicate cases not accepted by the system. The table does not include a column for the sequence /k/+ *y*, which is not represented in our corpus.

Table 3. How do pupils use *c*, *ç* and *cu* (in%)?

|            | Transcription of /k/     |                        |      |                      |                      | Transcription of /s/     |                      |                      |      |       |                     |
|------------|--------------------------|------------------------|------|----------------------|----------------------|--------------------------|----------------------|----------------------|------|-------|---------------------|
|            | Before the written vowel |                        |      |                      |                      | Before the written vowel |                      |                      |      |       |                     |
|            | A                        | E                      | I    | O                    | U                    | A                        | E                    | I                    | O    | U     | Y                   |
| <i>n</i> = | 12.008                   | 614                    | 12   | 37.975               | 11.563               | 1.733                    | 49.292               | 16.450               | 100  | 1.698 | 202                 |
| <i>c</i>   | <i>car</i><br>99.98      | 26.0                   | 25.0 | <i>encore</i><br>100 | <i>cuit</i><br>99.98 | 14.3                     | <i>lancé</i><br>99.7 | <i>voici</i><br>99.5 | 9.0  | 4.4   | <i>cycle</i><br>100 |
| <i>ç</i>   | 0.02                     | 0                      | 0    | 0                    | 0.02                 | <i>façade</i><br>85.7    | 0.3                  | 0.5                  | 91.0 | 95.6  | 0                   |
| <i>cu</i>  | 0                        | <i>cueille</i><br>74.0 | 75.0 | 0                    | 0                    | 0                        | 0                    | 0                    | 0    | 0     | 0                   |

The entire column concerning the use of *c* to transcribe /k/ before the vowel *i* can be disregarded: there are only 12 cases, which is derisory compared with the other numbers. On the other hand, the solution for writing a /kœj/ sequence is far more problematic. This sequence is responsible for 158 of the 159 errors in the use of *c* before *E*. The reason lies not in any particular difficulty with the grapheme *c*, but in the transcription of the vowel. Only a few words in the French lexicon (*accueil*, *cercueil*, *cueillir*, *écueil*, *recueil* and their derivatives) feature this particular grapheme sequence.<sup>12</sup> They illustrate a conflict of phonographic rules: (a) /œ/ can be transcribed by *eu* (*heure*, *seul*...), /œj/ by *euil* (*écueil*, *fauteuil*, etc.) (b) *c* cannot have the value /k/ before *E*. To resolve this conflict, French orthography adopted the solution of putting a *u* in contact with *c*<sup>13</sup> and removing the *u* from the vowel transcription. 19.6% of the children in 12F-Anniversaire and 34.8% in 14D-Générations stumbled over this solution, adopting a grapheme sequence other than *ueil* (*euil*, *uel*, *eul*, *ueui*...), thus demonstrating, more often than not, that they are more familiar with the general rules than with the special cases.

Let us now take a look at the transcription of /s/ on the right-hand side of the table. The percentage of errors is low when *c* is followed by an *E*, an *I* or a *y*. The rate of correct spellings for *c* + *E* is partly due to the frequency of the demonstratives *ce*, *cela*, *celui*, *celle*, *ces*, *cette*, *ceux* in the dictated texts, which account for half the number of targets containing the *ce* sequence. In the 24,352 demonstratives in our data, only 3 *ç* are found, compared with 154 in the 24,940 other words in

12. This category should also include, on the *g* side, *orgueil* and its derivatives.  
13. The solution -*cueil* would have been more coherent: it would have marked the occlusive character of *c* with a *u*, as in *gué* or *gui*, and would have noted the vowel as most generally in the system.

this category. Even in the latter case, the proportions are derisory. It is conceivable that the frequency of demonstratives containing the *ce* sequence helps children to integrate the regularity that /s/ followed by an *E* is transcribed with a *c* without a cedilla.

Every time a dictation includes *-ci*, *ici* or *(re)voici*, we record a few cedilla errors, albeit in a very small minority: there are 5 times more *çi* in these words than in other targets containing /s/+ *i*, but the proportions do not exceed 0.9%. These errors could be due to hypercorrection in these individual words, judging by the much more pronounced tendency to omit the cedilla on *c* before *u*, *o*, and even more so *a*.

Among the units that are the most critical in terms of cedilla handling, our corpus features verbs whose conjugation alternates *c* and *ç*: *élançaient* is written without a cedilla by 27.8% of school children, *effaçant* by 20.7%, *annonçaient* by 14.6%, *lançant* by 7.8%, *agaçant* by 6.1%. There are also a lot of cedilla errors in the forms linked to *-cevoir* verbs: for *aperçu*, *déçus*, *inaperçues* and *perçus*, we record between 2.3% and 9.4% of spellings without a cedilla. However, our corpus does not allow us to assert that the difficulty is greater in verbal radicals whose final consonant varies orthographically than in other units. *Façade* and *insoupçonnées*, with 10.2% and 9% of incorrect spellings respectively, only suggest that the problem is of a more general nature.

## 4.2 G, ge and gu

The grapheme *g* alone corresponds to /g/ before *A*, *O*, *U* (*gare*, *gout*, *figure*), and to /ʒ/ before *E* and *I* (*gentil*, *gilet*). However, it can be used to transcribe /ʒ/ before *A*, *O*, *U* by adding an *e* (*geai*, *nageons*, *gageüre*), and to transcribe /g/ before *E* and *I*, if followed by a *u* (*gué*, *gui*). A notable exception<sup>14</sup> is verbs in *-guer*, which keep their *u* throughout their conjugation (*fatiguais*, *fatiguons...*), in accordance with a principle that favors the integrity of the root, which certainly does not contribute to clarity in the distribution of variants, especially as this principle is not respected for verbs in *-cer* (*lancions*, *lançons...*), for example.

We built up the analyzed set in the same way as for *c*, *ç* and *cu*, taking into account only target words containing a /g/, transcribed by pupils using *g*, *ge* or *gu*. In cases such as *gorge* > *gaurge*, this spelling was classified as /g/+*A*. Percentages have been calculated on the sum of spellings with *g*, *ge* and *gu*. In the table below, the dark gray boxes indicate the possibilities excluded by the system and therefore corresponding to erroneous spellings. Those in light grey can only be used for the conjugation of verbs in *-guer*, of which our dictations provide no sample.<sup>15</sup>

14. Other exceptions include *baguage* (deverbal of *baguer*) and various words derived from toponyms, such as *camarguais*.

Table 4. What use (in%) do pupils make of *g*, *ge* and *gu*?

|            | Transcription of /g/     |                      |                       |                      | Transcription of /ʒ/     |                     |                      |                       |
|------------|--------------------------|----------------------|-----------------------|----------------------|--------------------------|---------------------|----------------------|-----------------------|
|            | Before the written vowel |                      |                       |                      | Before the written vowel |                     |                      |                       |
|            | A                        | E                    | O                     | U                    | A                        | E                   | I                    | O                     |
| <i>n</i> = | 6.810                    | 1.174                | 4.414                 | 1.946                | 1.229                    | 25.461              | 6.787                | 1.665                 |
| <i>g</i>   | <i>garde</i><br>98.5     | 1.7                  | <i>gorge</i><br>96.5  | <i>légume</i><br>100 | 35.6                     | <i>luge</i><br>99.9 | <i>agile</i><br>99.7 | 4.9                   |
| <i>ge</i>  | 0.1                      | 0.0                  | 1.7                   | 0.0                  | <i>rangea</i><br>64.0    | 0.0                 | 0.2                  | <i>pigeon</i><br>95.1 |
| <i>gu</i>  | <i>liguant</i><br>1.3    | <i>digue</i><br>98.3 | <i>liguons</i><br>1.7 | 0.0                  | 0.4                      | 0.1                 | 0.1                  | 0.1                   |

On the left-hand side of the table, corresponding to /g/, the second row contains percentages ranging from 0% to 1.7%: we can deduce that only a small minority of children associate /g/ with *ge*. Symmetrically, on the right-hand side, for sequences with /ʒ/, the last row contains percentages of 0.4% at most: almost all children know that *gu* cannot be used to transcribe /ʒ/.

The distribution of *g*, *ge*, *gu* seems to be conditioned by two variables, the relative importance of which would be interesting to determine.

- a. The “grapheme alone or combined” variable. This pits *g* against *ge* and *gu*. We hypothesize that the “alone” modality produces better graphemes than the “combined” modality, and that words with simple *g* are more successful than those with *ge* or *gu*.
- b. The “stable or variable grapheme” variable. Representing the “stable” modality, we have words like *gout*, *gelé*, *givre*, *guérir*, *digue*, whose representation of /g/ or /ʒ/ is the same in all their derivatives. The other modality is actualized in words like *nagions*, which neighbors *nageons*, or *Bourgogne* with *bourguignons*. The easiest words would be those whose transcription of /g/ or /ʒ/ does not vary, within their morphological family, according to the vowel it precedes.

When words lend themselves to the use of a single *g*, the results are better than when it has to be combined with an *e* or *u*: the *g* line systematically shows high percentages of correct spellings. Moreover, it is this variant that is preferred in the

15. For the sake of clarity, we have not included a column for /ʒ/+U (in the French lexicon, of the words that transcribe /ʒ/ by *g*, only *gageüre* has this characteristic), nor for sequences involving a *y*, as our corpus contains no words containing these combinations; we have also excluded the column for /g/+i, of which we have collected only two uses (*régulièrement* > *réguillièrément* and *gazouillis* > *gizouie*).

case of incorrect spellings (this is particularly evident in the 35.6% of words that should be written with *gea*, but are written with *ga*). This validates the hypothesis that the “grapheme alone or combined” variable is effective. But we also get good results for words like *pigeon* or *digue*, which use a combined grapheme. Let us take a closer look.

*Ge + O* – The dictations in this category contain *bourgeois* (3 times), *villageois* and *gageons*. Pupils who adopt the correct spelling for /ʒ/ are 96.1% to 99.7% for *bourgeois* and *villageois*, but only 86.5% for *gageons*. Is the difference in treatment linked to the frequency of these words? The VOB knows *bourgeon*, but not *villageois* (nor any other *-geois* word), nor *gager*. We do, however, find *dégager* and *engager*. The explanation seems to lie elsewhere than in frequency: the spelling of the /ʒ/ of *bourgeon* is stable (*bourgeonner...*), *villageois* is the only derivative of *village*, but the /ʒ/ of *gager*, radical ending, is written differently depending on the vowel that follows it.

*Ge + A* – On the other hand, 24.3%, 31.9% and 42.6% of pupils made many mistakes when transcribing the /ʒ/ targets *allongeaient*, *dégageaient*, *négligeables*. These are words which, within their morphological family, transcribe the final consonant of their root differently depending on the vowel that follows (*allongions*, *allongeaient...*). On the other hand, *intelligemment* > *intelligamment*, *intelligement...* is found in 66 answer sheets of the 367 from 12D-Électricité;<sup>16</sup> /ʒ/ is written with **g** alone by 42.9% of pupils, who are no doubt referring to the way in which intelligent should be written with g alone... It is therefore the variability of transcriptions of /ʒ/ that seems to be at issue.

*Gu + E* – In our corpus, the words in this category are *digue*, *longue* (2 times), and *vagues*. Their /g/ is correctly written by *gu* by all school children in the first two words, by 98.1% in the third.<sup>17</sup> These are units whose /g/ is transcribed *gu* in all their derivatives. However, we must also keep in mind the hypothesis that the good results on these words could also be attributed to the fact that /gə/ at the end of the word is always written *gue*, just as /kə/ at the end is always written *que*.

*Gu + A*, *Gu + O* – The dictations contain no *-guer* verbs. In this last row, the percentages in the light gray boxes are those of errors on words in /g/+A or O. Very few children make use of this possibility for the transcription of /g/, which suggests major difficulties if they had to write words like *fatiguons* or *distinguaient*, not to mention the distinction between *fatigant* and *fatigant*, etc.

To sum up, we can say that both variables are operative: the best results can indeed be recorded when the transcription of /g/ and /ʒ/ uses a single g; but com-

16. 88 other pupils also opt for an *a*, but write /ʒ/ with *ge*.

17. Other occurrences of *ge* instead of *gue* are found in 4 *élégante* > *élégente*, in 11 *gaité* > *gété*, *get-tée...*, and in 1 *égosillent* > *éguissent*.

bined graphemes pose few problems when they are stable in the morphological family of the words concerned.

This whole Section 4, which reminds us of the complexity of the rules in a very small sector on the scale of the whole system, indicates that it would be difficult to reform the sector cosmetically, with the use of the grapheme *g* proving the most problematic. At the very least, we could consider putting an end to the coexistence of adjectives and present participles (*exigent* VS *exigeant*), leaving only the verbal form (Legros et al. 2009).

## 5. Conclusions

As many other studies have pointed out (Chervel and Manesse 1989, Manesse and Cogis 2007, Bourdin et al. 2011, Tortereau et al. 2013, Brissaud et al. 2014, Andreu and Steinmetz 2016, Cogis and Brissaud 2019), by the end of primary school, despite the many hours of teaching time devoted to the subject, French spelling is certainly not a widely mastered competence. This is the very reason for the existence of spelling championships. While the participants in these events are not recruited among the weakest, or not even from the general public, in the case studied here, some 50% of pupils prove incapable of writing a one-page text of moderate difficulty without making 10 or more mistakes. Yet they all have a good grasp of the relationship between phonics and writing. In fact, in less than 1% of cases, the grapheme used does not correspond to the source phoneme, and most of the discrepancies are concentrated in the lexical domain (/s/ written with *s* or with *c*?), morphology (gender, number and person inflections), and syntax (agreements in particular). If the French transcription system were bi-univocal, if a single grapheme always corresponded to a single phoneme and vice versa, children at the end of primary school would certainly have a better command of French spelling, and French speakers would have a totally dispassionate relationship with dictation, and with spelling in general.

However, phonographic correspondences also have their pockets of difficulty and resistance. This is particularly true of the graphemes *c* and *g*, whose value depends on the context, and which must sometimes be accompanied by a diacritical sign or a silent vowel. By the end of primary school, many pupils have not fully mastered how these graphemes work.

Internet forums, in particular, show that secondary education fails to fill the gaps for all citizens, and that many French-speaking adults do not in fact master the written word in their own language. Is this not an issue that a democratic society should be concerned about?

## Funding

This article is part of a wider research project aimed at analyzing pupils' major spelling problems. This work would not have been possible without funding from the Direction de la langue française, Ministry of the Wallonia-Brussels Federation.

## Acknowledgements

We would like to thank Nathalie Marchal for supporting our project. We would also like to thank Liliane Balfroid, who gave us access to the answer sheets, as well as Jacqueline Lambert, who helped us whenever a question arose about the details of the championship.

## Bibliographie

Andreu, Sandra and Claire Steinmetz. 2016. *Les performances en orthographe des élèves en fin d'école primaire (1987-2007-2015)*, Direction de l'évaluation, de la prospective et de la performance, *Note d'information* 28, novembre 2016. Bousquet, Sylvie, Cogis, Danièle, Ducard, Dominique, Massonnet, Jacqueline and Jean-Pierre Jaffré. 1999. «Acquisition de l'orthographe et mondes cognitifs», *Revue française de pédagogie* 126, 23–37.

doi

Bourdin, Béatrice, Leuwers Christel and Carole Bourbon. 2011. «Impact des contraintes linguistiques et cognitives sur l'acquisition de l'accord en genre de l'adjectif en français écrit», *Psychologie française* 56, 133–143.

Brissaud, Catherine, Cogis Danièle and Catherine Tortereau. 2014. «La performance orthographique à l'articulation école-collège : une approche qualitative du marquage du pluriel», 4<sup>e</sup> *Congrès mondial de linguistique française*, Berlin, 19–22 juillet 2014.

Chervel, André and Danièle Manesse. 1989. *La dictée. Les Français et l'orthographe, 1873–1987*. Paris: Calman-Lévy, INRP.

Closet, François de. 2009. *L'orthographe, une passion française*. Paris: Fayard.

Cogis, Danièle and Catherine Brissaud. 2019. «À la poursuite des marques de genre», MORTAMET Clara, *L'orthographe. Pratiques d'élèves, pratiques d'enseignants, représentations*. Rouen: PURH, 43–71.

Dister, Anne and Marie-Louise Moreau. 2017. «Construire une base de données à partir de copies d'élèves. La base Orthocor des dictées du Balfroid 2012–2016». SESLA, en ligne. Dister, Anne and Marie-Louise Moreau. 2019. «Les tribulations d'une réforme dans un championnat d'orthographe. Comment des élèves en fin de scolarité primaire ont intégré les rectifications de 1990.», *Le français moderne* 2019 n°2, 281–295.

Dister, Anne and Marie-Louise Moreau. 2023 (à paraître). «Des phonèmes aux graphèmes», *Les grands corpus scolaires*.

Dister, Anne and Marie-Louise Moreau. 2023. «Quels enfants pourraient intégrer facilement les règles préconisées par l'écriture « inclusive » ? Conseils à la politique linguistique», *Le français moderne* 2023 n°1, 70–98.

- doi

## doi

- [illegible]

## Address for correspondence

Anne Dister  
~~UCLouvain, FIAL~~  
~~Place Cardinal Mercier 31/L3.03.11~~  
~~1348 Louvain-la-Neuve~~  
Belgium  
anne.dister@uclouvain.be

UCLouvain - Saint-Louis Bruxelles  
Boulevard du Jardin Botanique, 43  
1000 Bruxelles

## Co-author information

Marie-Louise Moreau  
UMons  
[mlmoreau@skynet.be](mailto:mlmoreau@skynet.be)

## Publication history

Date received: 17 June 2023

Date accepted: 11 December 2023

### Author Queries

- Please provide a complete reference for the citation '(Dister and Moreau 2019)' in this article.
- Please provide a citation for the reference id "CIT004 (Closet, François de. (2009))" since citation is missing in the article.
- Please provide keywords for this article.
- Please provide a shorter running header for this article.
- Please provide contact information (at least an e-mail address) for Marie-Louise Moreau.