A corpus of French texts with non-standard orthography

Yves Charles Morin  
Université de Montréal

The phonetic reconstruction of earlier stages of languages essentially relies on four techniques, unequally applicable to various languages: (1) comparative reconstruction, (2) internal reconstruction, (3) phonetic interpretation of early graphic systems, and (4) phonetic interpretation of poetic conventions such as rhymes and meter during its history. These techniques all have limitations and the best results are obtained when these can be combined and completed with one another, as is often possible with French.

Graphic systems may sometimes be relatively phonetic, i.e. may give more or less reliable indications of the pronunciation of a language, usually for relatively formal registers. It is rarely the case, however, that such phonetic correspondence remains stable in the course of time. Graphic systems are notoriously conservative and tend to be retained after sound changes have disrupted the phonographic correspondences that might have existed at earlier stages. Non-standard orthographies — whether they are intentionally devised as substitutes for current conventional spelling systems or elaborated by semi-literate scriptors after they acquired some rudiments of conventional orthography — tend to follow some kind of phonographic principle (not necessarily based on some prior phonemic awareness, as is sometimes claimed, however). Early documents using such orthographies thus offer a glimpse of the sound patterns of a language at the time when they were written, and have constituted some of the most important evidence used in scholarly works on the history of French.

The present survey is meant to be a reasoned, albeit limited, catalog of such documents, including those that have been entered in the computer databases1 that I have developed over the last twenty years as a tool for my work on the reconstruction of the pronunciation of French at different stages of its development. One essential characteristic for inclusion in a base was the richness of the expected information on the pronunciation and the relatively large size of documents devised by the same grammarian or semi-literate scriptor to allow for some reliability of the results.

The documents included in the databases can be divided into six categories ranging from transcriptions of French in non-Latin graphic systems (section 1) and documents written by

---

1 The documents signaled with the sign + under the heading db in the tables below have been included in the databases; those with the sign ° are currently being entered.
semi-literate individuals (section 2), to texts for which the authors or printers deliberately chose to use a reformed spelling, either fully articulated (section 5) or succinctly sketched (section 6). Less radical are documents that used a conventional orthography enriched with additional diacritics for easier reading (section 3) and documents that adapted the conventional spelling to note non-standard pronunciations (section 4). In the last section, we give a short list of texts written in relatively standard orthographies, which could nonetheless cast some light on the pronunciation (section 7), none of which, however, have been included in the databases.

1 Texts with non-Latin conventional graphic systems

One occasionally finds documents that are written with a conventional graphic system other than Latin. They usually are too short to allow for in-depth analyses. The only exceptions are documents with French words written in Hebrew characters: glossaries, dictionaries, treatises and — rarely — poems (cf. Blondheime 1927). An inventory of known glossaries can be found in Banitt (1972:xiv–xv) [G1 to G6], three of which have been edited, and a fourth only partially, together with various fragments [F1 to F9] and dictionaries [D1 and D2].

These glossaries written with Hebrew characters must not be confused with modern editions of French glosses in Rashi’s commentaries (Darmesteter 1909, Darmesteter and Blondheim 1937) that try to recover the original forms of the Talmudic commentator from often quite corrupted copies found in later manuscripts.

![Figure 1. Glossaire de Bâle, 6 first lines of fol. 27r° (from Banitt 1972)](image)

Glossaries contain series of Hebrew expressions, either from the Bible or teachers’ commentaries, followed by their French equivalents, or le’azim, as appears in the excerpt of the Glossaire de Bâle given in Figure 1. In this excerpt, one can read several glosses in Latin characters later added by an early analyst; in particular chant son is a Latin transliteration of the la’az immediately below, which explains a Hebrew term found before on the same line.
The *Traité des fièvres* on the other hand is basically a French text entirely written with Hebrew characters, with only occasional Hebrew and Latin comments and prescriptions.

Great hopes were entertained by the first scholars who examined these texts, in particular Darmesteter, that an alternate script would exhibit less graphic conservatism and add significantly to our knowledge on the phonetic characteristics of Old French. It appeared soon, however, that the use of the Hebrew script for French and for Latin belonged to a long tradition, which could also be quite conservative (cf. Banitt 1972:58). Different Hebrew characters could thus be used for Latin and French in the same medieval texts for sounds that medieval speakers certainly did not distinguish, such as *samekh* ס for Latin ‹s› and *sin/shin* ש for French ‹s› (besides ‹ts› for [s] < [ts]), probably because they belonged to independent scriptural traditions that developed at different times (cf. Banitt 1972:61, Kiwitt 2001:33).

<table>
<thead>
<tr>
<th>Conventional name</th>
<th>date</th>
<th>Editions and characteristics</th>
<th>db</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glossaire de Paris (G4)</td>
<td>13th c.</td>
<td><em>edition</em>: Lambert and Brandin 1905; the Hebrew characters of the French <em>le’azim</em> have not been edited, they are simply transliterated according to the authors’ own interpretation of their phonetic values, barring the possibility of a systematic reconstruction of the original Hebrew script. <em>dialect</em>: Lorrain (prob.).</td>
<td>+</td>
</tr>
<tr>
<td>Glossaire de Parme (G5)</td>
<td>14th c.</td>
<td><em>edition</em>: (partial); Siskin 1981.</td>
<td></td>
</tr>
<tr>
<td>Traité des fièvres</td>
<td>early 13th c.</td>
<td><em>editions</em>: Katzenellenbogen 1933, Kiwitt 2001; in Kiwitt’s edition, the Hebrew characters are transliterated according to a protocol which preserves most of the information found in the original script. <em>dialect</em>: Bourguignon (prob.).</td>
<td></td>
</tr>
</tbody>
</table>

**References**
2 Texts with deviant orthographies

There is a large body of research on deviant orthographies, usually of semi-literate scribes from various social backgrounds (cf. Ernst 1999). A survey of the relevant documents is well beyond the scope of this paper. Many of them can be found in some of the raw materials used by historians of private life: private letters, intimate and travel diaries, “livres de raison” (family diaries/registers), memoirs (cf. Beaurepaire and Taurisson 2003), which more often than not, however, have been adapted to modern spelling conventions and need to be re-edited to be useful as linguistic documents.

The very concept of deviant orthography may be ill suited for texts written during or before the fourteenth century, as their specific spelling may sometimes simply reflect a scriptural tradition of partial phonetic adjustment to a regional norm. One can probably find a complete continuum between the spelling innovations found in the Annales de Laval by Le Doyen (Figure 2) and those in the Haynin’s Mémoires (Figure 3).

Figure 2. Annales de Laval by Le Doyen, f° 1r° (cf. Godberg–Beauluère, p. 6)

lesquelles ne peulte toutes estre venues a ma connoissanse, plusieurs du Roiaume se conplaindoit et doleoit. Et tant que de fet a la remonstranse et requeste du duc de Berry, seul frerre du roi et le plus prochain de la couronne, et de plusieurs autres prinche et seigneurs de Roiaime, estant desplaisant de che quil veoit le Roiaime ensi gouverné et eux mime pareillement, desirant di pourvoir et mettre remede de tout leur pooir come tenus, iestoite. Et afin qu’ensi en avenist, il proume l’eun a l’autre d’estre aidant et confortant et d’eus trouver et coumuniquer ensembel ...

Figure 3. Haynin’s Mémoires, f° 9v° (cf. Brouwers, vol. 1, p. 6)

Le Doyen is a royal notary in Laval, well accustomed to writing. He abides by the current orthographical system with its regional phonetic interpretation, e.g., «oi» for a [e] or [ɛ] sound in some words, such as froit ‘froid’. His only significant departure from the dominant normative patterns is the use of «z» in word-final position for «r» as in recitez or ferchez in the
excerpt given above (which Godberg–Beauluère correct as \(r\)). This is a feature found in many manuscripts of the same period and probably reflects a phonetic development \(r > \theta \sim \emptyset\). On the other hand, Haynin is a member of the gentry, whose education was certainly more concerned with warfare than writing skills. His spelling reflects the current Picard scripta, e.g. \(\text{ch}\) for regional \(ʃ\) or \(ʧ\) in *prinches* and *che* and \(\text{ei}\) for \(i\) in *pourveir* ‘pourvoir’. One quite original feature, however, is the generalization of mute \(e\) in word final-position in 3sg endings, as in *peulte* (normally written *peut*, sometimes *peult*) and 3pl endings, as in *complaindoite* (for regional *complaindoient*) or *proumerte* (for pro(u)mirent), which signals the retention of word-final \(t\), a normal feature of Picard, still found in modern dialects with the 3pl ending.

Lodge (2004:143, 166–167) lists, together with large excerpts, some classic examples of deviant orthographies relevant to the history of Parisian French. A larger corpus, including other regions has been undertaken and is described by Ernst and Wolf (2002). Another large corpus for varieties of French spoken between the seventeenth and nineteenth centuries in North America and Northwestern France is under construction and described in Martineau (to appear) and Bénétéau and Martineau (2006).

<table>
<thead>
<tr>
<th>Conventional name</th>
<th>date</th>
<th>Editions and characteristics</th>
<th>db</th>
</tr>
</thead>
<tbody>
<tr>
<td>Œuvres poétiques de Ménétra</td>
<td>?–1802</td>
<td>edition: none – found in the same ms. as the Mémoires dialect: Parisian French.</td>
<td>+</td>
</tr>
</tbody>
</table>

References


Le Doyen, Guillaume. 1859. *Annales et chroniques du pais de Laval et parties circonvoisines, depuis l’an de Nostre Seigneur Ihesu-Crist 1480 jusqu’a l’année 1537,*
3 Annotated conventional orthography

Many grammarians felt the necessity to remedy the phonetic opacity of the conventional orthography by adding some diacritical marks. Sylvius (1531) even championed a two-tiered orthography for French: one tier for the etymological representation of words and a superposed tier for their pronunciation, e.g. <g> for [ɡ], <œ> for [gw], <œ> for [z] in <lions> for (nous) lisons, <œ> for [s] in <poisier> for poisier, etc., including a diacritic to mark mute letters <ś>, <ť>, etc.

Accented vowel-letters such as <é, è, ë> and <ç> with a cedilla are the only annotations still found in the standard orthography that can be traced back to Sylvius’ proposal. The use of diacritics to mark mute letters also found its way into some manuals for the learning of French as a second language. Sainliens, alias Holliband, (1566–1580) and Milleran (1692) made a regular use of such diacritics, which can now be most usefully applied to the study of liaison and elision (cf. Crevier 1993, 1994). For instance, the following excerpt from Sainliens shows that word-final <t> and <z> where silent before a consonant (ou eft le, preftez moy) and articulated before a vowel (quant à, eftiez en vn) and at the pause (aifément, que):

quant à voftre haquenée, elle va lès
ambles auflî aifément, que fi vous eftiez
en vn bateau: preftez moy voftre e-
fcharpe de Taffetas, à caufe de la poujdrê

eft du foleil : courage, ie voy la ville
Logeronç nous aux faulç àbourôs, ou
en la ville ? où eft le meôlleur logis?

Figure 4. Sainliens, Littleton (1566 f° A iiiij r°)
Some early printers may also have put this system to a more general use. Estiene Caveiller in his edition of Burrier (1542) used “un système orthographique complet et tout à fait nouveau” (Beaulieux 1927:43). The facsimile of the printer’s notice given by Catach (1968:276) shows that mute letters are marked with a subscript ↦, e.g. chaſcun, endroict, cognofftre, queſ; a subscript dot indicates the consonant value of ⧺ and ⟨a/v⟩, e.g. ſulque, vous, oeuvre; ⧺ notes palatal [ʎ], e.g. fueiſe; ⧺ has the same function as ⧺, e.g. conſeration. An internal dot may also appear inside the body of the following letters ⟨e e h m n⟩, whose distinctive function is not altogether obvious.

<table>
<thead>
<tr>
<th>Authors</th>
<th>date</th>
<th>Characteristics</th>
<th>db</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burrier, Antoine</td>
<td>1542</td>
<td>Mute letters are marked with a subscript ↦, values of ⟨i, v/u, e⟩ distinguished with subscript or superscript dots, ⧺ for palatal [ʎ], ⟨c, e, h, m, n⟩ marked with internal dots.</td>
<td></td>
</tr>
<tr>
<td>Sainliens, Claude</td>
<td>1566</td>
<td>Mute letters are marked with a subscript ↦ or superscript ↦.</td>
<td>*</td>
</tr>
<tr>
<td>Milleran, René</td>
<td>1692</td>
<td>Mute letters are in italics (or in roman in italicized words).</td>
<td>+</td>
</tr>
</tbody>
</table>

References


4 Phonetic use of conventional orthography

Conventional orthography may also be adapted to render pronunciations that do not conform to the prevalent norm, or at least what the author considered it to be. Such “phonetic” transcriptions of what could be considered genuine speech are relatively rare. One may mention Jean Héroard’s Journal (Ernst 1985, Foisil 1989) in which the author, a physician, recorded verbatim the conversations between the future King of France and his entourage, adapting the orthography to reflect the pronunciation of the young king (at least until he considered it to be no longer divergent from that of the adult norm). One also occasionally finds small passages in literary written works, where the authors tried to give an indication of some actual pronunciation (cf. Saint-Gérand, 2004). More often than not, however, this technique is used with definite parodic intentions, where the stereotypical features associated with the variety of speech they portray are often grossly multiplied (cf. Lodge 2004:136, 137, 154–158, 173–175, 210–214, 224–225 for Parisian French; some of the documents he presents are available from http://www.ahds.ac.uk/collections/ as Paris speech in the past lll-2423-I [as of November 30, 2006]).

Divergent pronunciations are usually signaled through modifications to the current orthographic usage, in ways that are sometimes difficult to interpret.

\begin{verbatim}
Lamant defpourueu de fon esperit efcri-
      uant a famie, oulant parler
      le courtifian.

MA Dame ie vourayme tan

May ne le dite pa pourtan
Les mufaille on derozille:
Celuy que fet les gran merueille
Nou doin bien to couché ensemble,
\end{verbatim}

**Figure 5. Le beau fils de Paris** (1549:270)

The omission of word-final letters in *Le beau fils de Paris*, as in *tan* (= *tant*) or *couché* (= *coucher*), not only indicates that they were not sounded in the speech portrayed, but are a tell tale sign that final [-t] and [-r] were retained in the “polite” pronunciation. The presence of final graphic -s in *les*, however, does not indicate that an [-s] or [-z] was pronounced before *fran* in *les fran merueille*, as this spelling conforms to the regular spelling convention allowing -s to be mute in such context. It is not clear, on the other hand, why the author (or the printer) used *Nou* (= *Nous*) in *Nou doin*, as the regular orthographic *Nous* would have signaled the same pronunciation. (The retention of -s after -e in *les*, as in *les fran merueille*, may actually have been prompted by the desire to keep a mark of the plural.)
## References


---

5 Spelling reformers

The Renaissance renewed interest in the pronunciation of Latin and its relation to spelling soon extended to vernacular language. Meigret and Peletier were the forerunner, in the middle of the sixteenth century, of a never-ending debate on a never fully completed reform of French orthography. In often-passionate discussions, the protagonists gave us invaluable information on the pronunciation of French, and a key to understanding the phonetic regularities embodied in the various reformed spellings put forward. The concrete proposals made by spelling reformers are useful only insofar as there survived a large enough body of texts embodying a specific reform with relatively infrequent inconsistencies — reformers indeed often blame the carelessness of printers for many of the blemishes in their work. It must also be emphasized that, contrary to optimistic views, spelling reformers are not necessarily aware of the phonemic distinctions in their own usage of French and, if they are,
do not necessarily put forward a new orthography that allows such distinctions to be
represented. They are usually convinced that their own usage — which they embody in their
spelling — is representative of some well-established norm, which accounts for the specific
regional characteristics one often finds in their texts.

5.1 Teachers

One may distinguish two groups of reformers. The first comprises schoolteachers, tutors, or
religious instructors, whose long experience teaching children or ordinary persons how to
read convinced them that the reformation of spelling was a social, if not moral, necessity.

Rambaud’s spelling system (1578) was somewhat of an *abugida* in which the majority of
the symbols do not represent simple indivisible sounds, but complete syllables. There were no
obvious resemblances between the shape of these symbols and those of the Latin alphabet (cf.
figure 6). Rambaud’s contribution to the debate on French orthography was probably nil.

![Figure 6. Rambaud script (pp. 62–63)](image)

Le Gaygnard’s influence was not any greater. His book contains long lists of words, but no
connected text. It is difficult to have a precise idea of the ways he intended specific words to
be spelled, not only because he often applied his reform only to such parts of words as were
relevant to some topic under discussion at that moment, but mostly because the printer did an
extremely poor work, making it difficult sometimes to even guess which were the words the
author actually intended to give as examples (cf. Morin, *to appear*).
la a la drète en disant: In nomine Patris et Filii, et Spiritus Sancti.

**Figure 7.** Vaudelin script (1715:137)

Although Vaudelin’s (cf. Figure 7) and Féline’s scripts were less radically departing from the conventional Latin script, they did not leave any trace either in the later adjustments made to the French orthography. Linguists, however, often used them as reliable transcriptions of earlier stages of the pronunciation of Standard Parisian French (cf. Cohen’s 1946 and Martinet’s 1946 analysis of Vaudelin’s work). Thurot (1881: LXXXVII) hailed Féline’s work as “le guide le plus sûr que je connaisse pour la prononciation de notre temps” (the most reliable guide I know for the current pronunciation [of French]).

<table>
<thead>
<tr>
<th>Conventional name</th>
<th>date</th>
<th>Editions and characteristics</th>
<th>db</th>
</tr>
</thead>
</table>
| Rambaud           | 1578       | *edition: J. de Tournes, Lyon.*  
observations: syllabic abugida not abiding to the orthographic principle, not readable without complete retraining, also used for Latin and Provençal.  
dialect: Marseilles regional French. | +  |
observations: regular alphabet without new characters.  
dialect: Poitou regional French. |    |
observations: very large number of new characters, reduced e not represented, not readable without specific training.  
dialect: no obvious regional features. |    |
observations: true “phonetic alphabet”; ë, łò for [œ, ø], ë̃ for [u] and ø̃ for [y], ą̃ for [ŋ] and ū̃ for [ʒ], ą̃̃ for [į] and ę̃ for [ǐ].  
dialect: minimal Parisian system. | +  |

**References**


5.2 Humanists

Humanists were less concerned with social issues — although there are notable exceptions, e.g., Lesclache — than internal consistency with an orthographic principle requiring spelling to be a faithful “mirror” of the pronunciation of the language, as they found argued in some of the earlier Latin grammarians’ works. The pronunciation on which they based their orthography strongly reflected that of their native regional variety, as contemporary witnesses did not fail to highlight. Pasquier thus commented in a letter to Ramus the regional features he noted in Meigret’s, Peletier’s, Baïf’s and Ramus’ works:

> Ceux qui mettent la main à la plume prennent leur origine de divers païs de la France, et est mal-aisé qu’en nostre prononciation il ne demeure tousjours en nous je ne sçay quoy du ramage de nostre païs. ‘Those who put their hand to the pen originate from various parts of France and it is uneasy for one not to retain in his pronunciation some ineffable part of his birthplace’s ways of sounding’.

(quoted by Firmin-Didot 1868:195, Pasquier 1723, tome 2, page 65)
As a rule, humanists’ reformed spellings retained much of the familiar aspects of traditional spelling and mostly used diacritics to further distinguish between the different phonetic values of ambiguous letters, in particular accented vowels and cedilla 〈ç〉, which eventually found their way into the modern orthography.

Meigret’s and Peletier’s works make up the two largest corpora of homogeneous documents. Their usefulness is nonetheless tempered by the typographical canons of that time, which limited to one the number of accents a word could receive and which did not find fit to use diacritic marks to forms that would not otherwise be ambiguous.

With the exception of Peletier’s work, documents written in a specific reformed spelling were produced within a relatively short period of time. Notwithstanding the restrictions mentioned earlier, Peletier’s work is extremely rich and presents a variegated sample of texts bearing on issues as diverse as grammar, mathematics, poetic rhetoric and actual poetical creations, written over a period of thirty years during which he adjusted his transcriptions to the changing mores in the recognized norms of pronunciation.

<table>
<thead>
<tr>
<th>Reformer</th>
<th>date</th>
<th>Editions and characteristics</th>
<th>db</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meigret</td>
<td>1548–1551</td>
<td>editions: Chrestien, Paris: 1548, 1550a, 1550b, 1550c, 1551. observations: few new characters – sandhis indicated between host and clitics. dialect: Lyon regional French.</td>
<td>+</td>
</tr>
<tr>
<td>Ramus</td>
<td>1562</td>
<td>edition: A. Wechel, Paris: 1562. observations: few new characters – sandhis indicated between host and clitics. 1572 editions: A. Wechel, Paris: 1572; Du Val, Paris: 1587. observations: relatively more new characters than previously – sandhis (except regular elision) are no longer indicated; the characters and the inventory of vowels are different from those of the previous editions. dialect: Picard regional French.</td>
<td>+</td>
</tr>
</tbody>
</table>
observations: regular alphabet without new characters.
dialect: Picard regional French.

Mercure de Trévoux 1708 appeared in a periodic journal.
observations: regular alphabet without new characters, radical simplification and neutralization of vowels distinctions.
dialect: unidentified.

observations: regular alphabet without new characters.
dialect: unidentified.

References
Dobert, le père Antoine. 1650. Récréations literales et mysterieuses, où sont curieusement estalez les principes et l’importance de la nouvelle orthographe, avec un acheminement à la connoissance de la poësie, et des anagrammes. Lyon: F. de Masso.


6 Texts with reformed spellings

A large number of other works have been published for which the printer — either on his own or at the author’s request — used non-traditional spellings, for which little or no justification is presented. One of the first systematic study of such innovative spellings is that of Firmin-Didot (1868), who used labels such as “personal orthography”, “intelligent personal orthography”, “independent orthography”, “containing instructive details” to describe them. This was later completed by Beaulieux (1927) and recent systematic studies of sixteenth and seventeenth-century spelling systems (Catach 1968 and Biedermann-Pasques 1992). On the basis of these analyses and of an earlier and quite succinct “survey” by Le Gaygnard (1609), I have selected a tentative list of potentially informative documents. They are listed in this section when the printer or the author used what may be considered a reformed spelling,
whether original or adopted from someone else; they appear in the next section when the spelling — without being as innovative — appears to embody distributional regularities that may reveal some of the phonetic properties of the variety of French they use.

The criteria used for this selection do not bear on the quality of the spelling systems as is normally examined in these works on orthography, but rather on prospective usefulness for the discovery of new information on the pronunciation; thus we include Domergue’s work in spite of Firmin-Didot’s severe critique (1868:306). Conversely, some editions have been omitted which are often singled out for the originality of their spelling reforms, such as Fouquelin (1555) or Tahureau (1555, 1565). The editions that I have consulted (Fouquelin 1557, Tahureau 1555, 1567), for instance, do not appear to warrant the large investment required to enter such documents into a computer database, as they would probably not add to our knowledge of the evolution of French pronunciation. (This conclusion, however, does not necessarily apply to the editions of Fouquelin 1555 and Tahureau 1565, which I have not yet consulted.)

Printers, as already mentioned, found it very difficult to follow the authors’ instructions, and more often than not departed from their plans. When the principles governing the new orthography are clearly presented, it is possible to identify as mistakes forms that blatantly contravene them. With texts written in a reformed orthography with little indication on its nature, however, printer’s mistakes are more difficult to notice, in particular with relatively short texts that do not allow for statistical control of the results.

The list below also contains the work of Jean-Antoine de Baïf written with his reformed spelling, most of which has not been printed during his lifetime and only survived in one autograph manuscript. The Pléiade poet announced a forthcoming grammar in which he would expose the characteristics of his script, which however was never published or has since been definitely lost. A comparison of his manuscript with his Étènes de poezie fransoëz an vërs mezurés (1574), the only one of his printed texts using the reformed script, shows that printing imposed heavy constraints on the liberty of spelling reformers. In his manuscript, Baïf made a rich use of diacritics above vowel-symbols: breve vs. macron for poetic meter, grave vs. circumflex accent for phonological length and acute accents for ‘adverbs’ as well as for some still unidentified usage (cf. Morin 1999b). The use of accents, however, is quite limited in the printed work; in particular his new characters never appear in print combined with a circumflex accent.

<table>
<thead>
<tr>
<th>Author</th>
<th>date</th>
<th>Editions and characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baïf, Jean-Antoine de</td>
<td>1569-??</td>
<td>editions: D. du Val 1574, Paris; Groth 1888; Bird 1964. observations: new characters (the 1574 printed material used the same fonts as Ramus 1572 with different phonetic values). dialect: no obvious regional features.</td>
</tr>
</tbody>
</table>
**French texts with non-standard orthography**

<table>
<thead>
<tr>
<th>Author</th>
<th>Year(s)</th>
<th>Editions</th>
<th>Observations</th>
<th>Dialect</th>
</tr>
</thead>
</table>

**References**


Domergue, Urbain. 1794 [an V]. *La prononciation française, déterminée par des signes invariables, avec application à divers morceaux, en prose et en vers, contenant tout ce qu’il faut pour lire avec correction et avec goût; suivie de notions orthographiques et de la nomenclature des mots à difficulté*. Paris: F. Barret.

Domergue, Urbain. 1805. *Manuel des étrangers amateurs de la langue française, ouvrage... contenant tout ce qui a rapport aux genres et à la prononciation, et dans lequel l’auteur*
a prosodié, avec des caractères dont il est l’inventeur, la traduction qu’il a faite en vers français de cent cinquante distiques latins, des dix élogues de Virgile, de deux odes d’Horace, et quelques morceaux en prose de sa composition. Paris: Librairie économique.

Domergue, Urbain. 1806. La prononciation française, où l’auteur a prosodié, avec des caractères dont il est l’inventeur, sa traduction en vers des dix élogues de Virgile et quelques autres morceaux de sa composition; augmentée d’un tableau des désinences françaises, pour faciliter l’étude des genres: manuel indispensable pour les étrangers, amateurs de cette langue, infiniment utile aux Français eux-mêmes, 2e édition. Paris: Librairie économique.


Taillemon, Claude de. 1556. La Tricarite, plus qelques chants an faveur de plusieurs damoézelles. Lyon: J. Temporal.
7 Texts with personal spellings

The systematic study of distributional orthographical regularities may eventually throw some unexpected light on some aspects of pronunciation in any text, if it is sufficiently long, be it written in the most traditional conservative orthography. The odds, however, are against it. Firmin-Didot’s comments suggest that the three following authors may be more promising than most others. A forthcoming computer edition in *Epistemon* of Matthieu’s work (http://www.cesr.univ-tours.fr/Epistemon/cornucopie/Cornuc.asp) will soon allow one to put the conjecture to the test.

<table>
<thead>
<tr>
<th>Author</th>
<th>date</th>
<th>Editions and characteristics</th>
<th>db</th>
</tr>
</thead>
</table>

References


8 Concluding remarks

The documents marked with the sign “+” in the tables above have been entered in different databases at different periods, with largely different protocols. Each was built to solve a specific problem. Some of them have been completely lemmatized; some are little more than a raw concordance. Little effort has yet been made to use a common approach, and it appears unlikely that a uniform protocol may even be profitably considered for printed and written documents to be included in a common database.

Their authors referred to a common norm for the pronunciation of French that was escaping them. Each had largely different conceptions of the organization of sounds in language and of the ways a spelling system had to incorporate sound distinctions. On the other hand they shared a common vocabulary to describe the sounds of the language that often masked their divergences and could underlie widely different realities.

The experience gathered these last twenty years has shown that each document must be analyzed on its own. There does not appear much to be gained by cross-referencing the data between different authors. It appeared most profitable to examine each spelling system individually and relate it to the forms that the French, the high language of the privileged classes, had taken in the different regions where it eventually became the official language.

Acknowledgments

Most of the databases presented here have been created over a period of twenty years and benefited from several grants from the Government of Québec (FCAC, FCAR) and Social Sciences and Humanities Research Council of Canada (SSHRC). I would like to acknowledge with thanks the help of Michèle Bonin, Carole Bouchard, Catherine Courchesne, Isabelle Crevier, Jocelyne Cyr, Sophie Daoust, Danielle Frêchette, Jocelyn Guilbault, Martine Ouellet, Julie Richard and Sandra Thibault. These bases are now being maintained and expanded within the framework of the MCRI project Modeling Change: The Paths of French, directed by F. Martineau, also funded in part by the SSHRC. I would like to thank for their contribution Stéphanie Brazeau and Jaïme Dubé. The databases for the texts with deviant orthographies (section 2) have been created for the realization of a project also funded in part by the SSHRC, Évolution et variation dans le français du Québec du XVIIe au XIXe siècles, under the direction of France Martineau, with the collaboration of Alain Desrochers. We would like to acknowledge with thanks the help of Marianne Chevrier, Jennifer Dionne, Maxime Glandon, Amélie Hamel, Philippe Leblond, Ashley Lemieux and Mylène Pérault for their participation to this project.