

Romance Orthographic Reintegrationism: Orthographic Guidelines for Romance Minority Languages

by Sicilanguageist

GLOSSARY, SYMBOLS AND ABBREVIATIONS

TERM	MEANING
\$	Proposed spelling, example ROR spelling. Remember that the spellings introduced by this symbol are purely explanatory of a ROR rule or prescription and not an official proposal
§	Introduces the title of a section of this document
languages of reference	Italian, Spanish, Portuguese, Catalan, Occitan, French and Romanian; the languages used as reference for the present proposal
orthographic reintegration	The implementation and adaptation of the spelling conventions of a language or language group by the orthography of another language
palatal <c>, palatal <g>	The reflexes of Latin c and g before e and i at the beginning of word-initial, stressed syllables; for example, when they represent [s, ʒ] in French, <c> and <g> are called 'palatal' in this proposal
RML	Romance minority language

1 INTRODUCTION

1.1 PURPOSE AND SCOPE

The core rationale of Romance Orthographic Reintegrationism (ROR) is that Romance minority languages (RMLs) should be represented orthographically as part of the wider Romance linguistic continuum. Effectively, this means that the orthographies of RMLs should be grounded in the practices of the major Romance orthographies—namely Italian, French, Spanish, Portuguese, Occitan, Catalan and, to a lesser extent, Romanian. This is achieved through etymological justifications for carefully considered spelling choices aimed at achieving orthographic reintegration with the Romance language family. (For a definition of “reintegration”, see below.)

ROR does not seek to create a single orthography for all RMLs. Instead, it provides general guidelines for creating orthographies, leaving room for a diversity of outcomes.

The present proposal is potentially useful for all stakeholders involved in the study of Romance minority or regional languages, such as linguists and philologists (including orthography experts), language planners, policymakers, educators and community activists.

1.2 WHAT DOES “REINTEGRATION” MEAN?

In its only official and established sense, reintegrationism is a linguistic movement which postulates that Galician and Portuguese constitute a single language. As a result, Galician reintegrationism promotes orthographic unification with Portuguese through the implementation of Portuguese spelling conventions (see [this link](#) for more). The implementation and adaptation of the spelling conventions of a language or language group by the orthography of another language is here referred to as *orthographic reintegration*.

While Galician can look to Portuguese for reintegration on both historical and linguistic grounds, the aim of the present proposal is not necessarily to reintegrate a minority language into a dominant one—although ROR guidelines could be used toward that end if desired. In actuality, ROR advocates for orthographic similarity or convergence among minority languages belonging to the same subgroup; conversely, it advises against orthographic convergence with the majority language of the region or state, unless it is consistent with and does not contradict the objective of subgroup cohesion.

The kind of orthographic reintegration advocated for by ROR is that of RMLs towards the major Romance languages with an established and continuous orthographic tradition.

1.3 AIMS AND OBJECTIVES

Increasing mutual intelligibility with other Romance languages in writing is one of the primary reasons why a given RML should use a Romance reintegrationist orthography. This is achieved thanks to focused etymological considerations. With a more informed, limitedly historical spelling, the lexicon would be more recognisable for speakers of other Romance languages, all the while still maintaining a degree of orthographic complexity that is manageable for native speakers with enough training.

Through the enhanced mutual intelligibility that reintegrationist spellings would provide, ROR aims at breaking the linguistic isolation most minority languages face, creating interconnectedness between minority and majority languages of the Romance family.

Since the creation of these orthographies should require deep etymological and possibly philological understanding, ROR encourages and fosters etymological research and the creation of an active scholarly community.

2 GUIDELINES AND RULES

2.1 GENERAL PRINCIPLES

In this section, I will explain the overarching principles behind ROR guidelines.

2.1.1 Attested Romance orthographic practices

The central idea behind ROR is to reintegrate the orthography of Romance minority languages through the implementation of orthographic conventions used in Portuguese, Spanish, Catalan, Occitan, French, Italian and, to a lesser extent, Romanian.

This is achieved by acknowledging what these orthographic conventions are and by understanding what their *etymological* role is within the system they are borrowed from. A further recommendation is to avoid letters and multigraphs (di-, tri- and quadrigraphs) that are not used in the languages of reference. The use of <w> and <k> is especially discouraged, unless finding another solution proves unworkable.

This idea is also valid for diacritics. For example, the use of double dots above a character as a diacritic (<ä, ë, ï, ö, ü>) should be reserved for use as a diaeresis rather than an umlaut, i.e. it should not indicate a change in vowel quality, unless finding another option is difficult.

2.1.2 Etymological criterion

The details of the so-called *etymological criterion* are better explained in [§ROR prescriptions](#). The main idea is that the reflexes and evolution of certain Latin consonants and consonant clusters determine how certain sounds are represented. These newly established spelling conventions can then be extended to non-etymological contexts. Therefore, a ROR orthography ought not to be exclusively historical: etymology is just the method used to *justify* spelling choices in

certain contexts so that they can be used non-etymologically—according to the restrictions that will be explained below.

I will give an example. Old Spanish spelt [ʃ] as <x> and [ʒ] as <j> or <g>. Over the centuries, these sounds merged into the sound [x], and the distinction between them was lost. To reflect this merger, it was decided that <j> should be used for the sound [x] where <x> was used in the past (in contrast, the use of <g> remained the same, but that falls outside the scope of this example). In this case, the use of <j> was extended to non-etymological contexts—an example of how a historically motivated choice can develop into a general spelling rule, within the limits of ROR.

2.1.3 Reader-oriented approach and orthographic depth

Depending on the language's phonological history, a reintegrationist orthography may turn out to be historical to various degrees, and therefore more or less difficult to write in. For example, compare the Catalan words *cena* “dinner” (an archaic term) and *sena* “number six in a game of dice, sice”. Both words are pronounced the same, but their initial consonant is spelt differently according to etymology. This makes it more difficult for the writer to predict or recall the exact spelling of a word.

Romance reintegrationism promotes facilitating the reader rather than the writer. That is, while a single phoneme may be represented by multiple graphemes, each grapheme should ideally correspond to only one phoneme. It is possible that this principle cannot be implemented fully, and that mostly depends on how complex the phonemic inventory of the language is. Sacrificing this principle might be needed to make the orthography less crowded with diacritics (see §[Economy](#)) and less cumbersome for both the reader and the writer.

2.1.4 Economy

An important rule of ROR is the avoidance of double marking: a particular element or feature cannot be indicated more than once. As an example, Bolognese Emilian [mʌnd] “world” is written *mānnd*. The double <n> marks the preceding vowel as short, but [ʌ] <ā> is always short, so there is no need for a double <n>.

ROR also discourages the extensive use of diacritics, unless they are well established in the reference languages mentioned above. This ties nicely with the principle of avoidance of double marking: stress is sometimes overmarked in writing by RMLs. One way of preventing this is to establish rules for a default interpretation of stress placement in an accent-less word; those words which defy the established rules will be marked by an accent (acute or grave).

2.1.5 Other considerations

...

2.2 ROR PRESCRIPTIONS

Here, the word “prescription” is used to define guidelines that are not considered general principles behind ROR, and which are more practical and specific in nature. These prescriptions are rooted in Latin etymology, but—as already stated—etymology should not be overly

dominant in a Romance reintegrationist orthography. Remember that these prescriptions are suggestive in nature and allow for flexibility. **I will use the symbol ‘\$’ to indicate proposed spellings. These are merely examples used to illustrate ROR prescriptions and should not be interpreted as formal orthographic proposals.**

Some of these ideas can contradict each other, so choosing one option over another is definitely acceptable. If a prescription is essential for creating a ROR orthography, it will be stated.

2.2.1 Palatal and palatalised sounds

2.2.1.1 Latin yod

“Yod” here refers to the phone [j].

The modern reflex of word-initial Latin yod at the beginning of a stressed syllable will determine the use of <j>. For example, since Latin *Iocat (“he/she plays”) evolved into Venetian [ˈzoga], that initial Venetian [z] can be written <j> (for example \$joga).

The Italian route. The evolution of yod in this context usually coincides with the evolution of the prevocalic Latin sequences -GI- and -DI-. Therefore, the resulting sound from this merger can be spelt the same as these sequences even in contexts where it comes from yod. This is the convention adopted in Italian.

HODIĒ “today” > It. *oggi*
 IUNIUM “June” > It. *giugno*
 MEDIUM “middle” > Ven. [ˈmɛzo] \$megio, mejo, mexo, etc.
 *Iocat > Ven. [ˈzoga] \$gioga, joga, xoga, etc.

The Catalan route. The general rule for this in Catalan is different. I will explain it to the best of my abilities, but note that the following is an oversimplification. Catalan treats the [(d)ʒ] sound that came from Latin yod the same as Latin G before E and I. For most cases, [(d)ʒ] is written <g> before <e, i> and <j> before <a, o, u>. This means that, if Latin yod has the same reflex as G before E and I, then they can be treated the same as Catalan: said reflex—which is [(d)ʒ]—is written <g> before <e, i> and another choice has to be made before other vowel letters.

GENTEM “people” > Cat. [(d)ʒen(t)] gent
 IUNIUM “June” > Cat. [(d)ʒun] juny
 MEDIUM “middle” > Ven. [ˈmɛzo] \$megio, mejo
 *Iocat > Ven. [ˈzoga] \$gioga, joga
 GENERUM “son-in-law” > Ven. [ˈzɛnero] \$gènero

2.2.1.2 Modern yod: the use of <j> vs <y> vs <i>

Different Romance languages have different traditions for writing the sound [j].

Some do not distinguish it from [i] at all in writing, such as modern Italian, Catalan, Portuguese and Romanian; so that is certainly an option for a ROR orthography, if no or little ambiguity arises from such a choice. Other languages, however, do differentiate [j] in writing. Older Italian

and many Romance minority languages in Italy traditionally use <j> for [j], while French and Spanish use <y>. The preference for <i>, <j> or <y> relies mainly on the interconnected strategies adopted for handling Latin yod (see §[Latin yod](#)) and the treatment of Latin <c> and <g> before front vowels (see §[Latin c and g before front vowels and the use of <c> and <g>](#)).

2.2.1.3 Latin c and g before front vowels

Probably the most important choice to make when creating a reintegrationist orthography is how the letters <c> and <g> are handled. The following suggestions are especially recommended.

Before <a, o, u>, the sound [k] should always be written with the letter <c>, and the sound [g] should always be written with the letter <g>. [k, g] should be written <c, g> also at the end of a word—unless some kind of liaison is present, in which case it is advisable to use a system inspired by French.

The use of the sequences <ce, ci> depends on the evolution of Latin c and g before e and i at the beginning of word-initial, stressed syllables. These will be referred to as *palatal <c>* and *palatal <g>*, regardless of whether their realisations in modern languages have a palatal place of articulation. Take for example French *cent* and *sent*. Much like the Catalan examples *cena* and *sena* above, they are pronounced the same, but the initial [s] is represented according to its origin in Latin. A ROR orthography should reflect this: palatal <c> and palatal <g> before <e, i> cannot be represented with letters other than <c> and <g>.

Many RMLs do not represent this faithfully, since their orthography is based on the way sounds are represented in the majority language of the state or nation where they are spoken. See the following examples.

CURRENT RML ORTHOGRAPHY	PROPOSED SPELLING	LATIN WORD OF ORIGIN
Jèrriais Norman <i>chent</i> [ʃɑ̃] “hundred”	<i>\$cent</i>	CENTUM
Walloon <i>djins</i> [dʒɛ̃] “person”	<i>\$gens</i>	GENTĒS
Genoese Ligurian <i>Zêna</i> [ˈzeːna] “Genoa”	<i>\$Gena</i>	GENU
Rumantsch <i>tschintg</i> [ˈtʃɪnc] “five”	<i>\$cintg</i>	CINQUE < QUINQUE

Table 1 - Spellings of RMLs for palatal <c> and <g>

2.2.1.4 The <s> rule

It is recommended that a sound derived from Latin S be written with <s> or <ss>, especially when the sound is [s] or [z]. However, this is a flexible guideline that may be disregarded in cases where following it would result in an orthographic form that makes it difficult for the reader to predict the pronunciation of a word.

2.2.1.5 The *velar table*

A way to write [k] and [g] before front vowels must be found. Strategies must also be developed to represent the sounds written with palatal <c> and <g> when they occur before back vowels.

I devised the following table which helps with creating said conventions, and which should be valid for all languages of reference *and* ROR orthographies. To illustrate how the table works, I will take Catalan as an example.

	NON-PALATAL ([k, g])		PALATAL ([s, (d)ʒ])	
	before <e, i>	before <a, o, u> and word-finally	before <e, i>	before <a, o, u> and word-finally
Latin c	<qu>	<c>	<c>	<ç>/<ss>
Latin g	<gu>	<g>	<g>	<j>

Table 2 - Velar table of the Catalan language

There are two prominent strategies to write the sounds [k] and [g] before <e, i>.

The first is the use of <qu> and <gu> (e.g. as in Catalan), while the second is the use of <ch> and <gh> (as in Italian, for example). This choice rests on the language's phonotactics and sound inventory. If the language has the sequences [kw] and [gw], the orthographer might want to use <qu> and <gu> to represent them, reserving <ch> and <gh> for [k] and [g] before <e, i>. If the language does not display frequent use of sequences [kw, gw], the choice is a little freer. It must be noted that a diaeresis can be used to distinguish [kw, gw] from [k, g] if needed or desired (<qü, gü>), much like in Catalan; another option is the Spanish strategy of contrasting <qu> [k] vs <cu> [kw].

One consideration to make is whether the digraph <ch>, which is quite common among the languages of reference, is useful to represent another phoneme or sound entirely; it is most common for some kind of palatalisation before <a, o, u>—see French *chat* (< CATTUM) and Portuguese *chorar* (< PLŌRĀRE). Selecting <ch> to represent [k] rather than some kind of palatalised sound depends on the orthographer and, most importantly, to the language's sound inventory: it is necessary to evaluate what graphemes the RML needs to represent all of its phonemes.

In order to write the sounds typically associated with palatalised <c> and <g> before <a, o, u>, Catalan uses <ç, ss, s> for [s] and <j> for [(d)ʒ]. The choice between <ç> and <ss> is often etymological, while <s> is most commonly employed at word boundaries. As an example, let's take the Catalan verb *alçar* “to lift” and the Spanish noun *pez* “fish”.

alçar [al'sar] (infinitive)
alces ['alɐs] “you lift”
pez [peθ]
peces ['peθes] (plural)

In Catalan, <ç> and <c> are used to maintain the pronunciation [s] throughout the paradigm. The same goes for Spanish, where <z, c> are part of an alternating orthographic paradigm.

In this case, the orthographer has much liberty. The most common spellings for the sounds associated with palatal <c> and <g> are:

- for <c>: **ci** + vowel, **z, ç/ss/s** (also attested among RMLs: **ch, tch, tg**; other additional possibilities: **sh, sch, x**);
- for <g>: **gi** + vowel, **j, ge** + vowel (also attested among RMLs: **z, sgi** + vowel (for [ʒ]), **sg** (for [ʒ]); other additional possibilities: **x, sj**); see also [\\$Latin yod](#).

Here is an example of a hypothetical velar table for Aostan Franco-Provençal.

	NON-PALATAL ([k, g])		PALATAL ([ts, dz])	
	before <e, i>	before <a, o, u> and word-finally	before <e, i>	before <a, o, u> and word-finally
Latin c	<qu>	<c>	<c>	<ch>
Latin g	<gu>	<g>	<g>	<j>

Table 3 - Hypothetical velar table for Aostan Franco-Provençal

When planning an orthography, I strongly recommend filling in a velar table, in order to have a visual representation of a cohesive orthographic system.

2.2.1.6 [ɲ] and [ʎ]

To write the palatal nasal and the palatal lateral approximant, there are specific Romance traditions and it is advisable to stick to them. They are illustrated below.

- For [ɲ]: **gn, ñ, nh, ny**.
- For [ʎ]: **gli** + vowel, **ll, lh**, vowel + **ill** + vowel, vowel + **il** (also attested among RMLs: **gl, ly**).

<gn> is recommended only if some instances of [ɲ] come from Latin GN. <ñ> is recommended only if some instances of [ɲ] come from Latin NN. <nh, ny> can be used in any orthography.

<gli> + vowel or <gl> are recommended only if some instances of [ʎ] come from Latin GL or from the -CL- sequence often found in diminutives and some other words (see Romansh [eʎ] **egl** “eye” from OC(U) LUM). <ll> is recommended only if some instances of [ʎ] come from Latin LL. <lh, ly> can be used in any orthography. <ill, il> are not recommended because they can cause confusion and be mistaken for [jɪl(l)].

2.2.1.7 The use of <x>

It is possible to use <x> to represent a reflex of Latin cs which is different from [s]. See the following examples for “thigh”, from Latin COXA. Other possibilities exist for this sound—namely **sci** + vowel, **sch, sh, ch, ş**.

Genoese Ligurian [ˈkøʃa] \$cheuxa

Picard (Nort-Leulinghem) ['kɥij] \$cui*x*

Bolognese Emilian ['kɔːja] \$cò*xa*

<sci> + vowel and <sch> are especially recommended for a sound (usually [ʃ]) which comes, in at least some instances, from Latin sc.

2.2.2 What to do with the remaining consonants

There is no other recommendation for the representation of the remaining consonants other than using the languages of reference as a guide.

For example, let's take the affricates [ts, dz]. Assuming that these affricates are not the realisation of palatal <c> and <g> in front of <e, i> (as it happens in Franco-Provençal, for example), one could choose to adhere to the Italian convention and use <z(z)>; or to the Catalan and Sardinian convention using <tz>; or even to the Romanian convention using <ț> (for [tz]); or to the medieval Iberian convention by using <ç> (for [ts]). The orthographer might also combine these strategies to differentiate between [ts] and [dz], if needed.

gorpiy

2.2.3 Vowels

2.2.3.1 Nasal vowels

There are two main strategies for marking nasal vowels in the languages of reference.

The Portuguese route. Portuguese marks its nasal vowels with a tilde on vowels: <ã, õ>.

The French route. French uses mainly <m, n> to mark nasal vowels: <am, an, aim, em, en, eim, ein, eun, im, in, um, un> etc.

The Breton route. This strategy is not found among the Romance languages, but it can be useful if it is difficult to find another solution. Breton, a Celtic language, marks nasality writing a <ñ> after the vowel.

The choice rests on the shoulders of the orthographer. One must avoid the crowding of diacritics whilst also attempt to represent the phonemes of the language faithfully. Accessibility is also an issue: font support and keyboard layouts may influence this choice (see §[Diacritics](#)).

2.2.3.2 Front rounded vowels

It is preferable to not use diacritics to mark front rounded vowels; instead, the use of bare vowel letters or multigraphs is encouraged. Some options are as follows.

- For [ø, œ]: **eu, oeu/œu**. (Also possible: **oe/œ**; other digraphs may be useful, such as **oi, eo**.)
- For [y]: **u**. (Some useful digraphs, not found in Romance languages for [y]: **ui, iu**.)

If every other option fails to meet the goals of the orthographer, ROR suggests using the 'double dot above' diacritic on vowels, mainly in stressed syllables. This is only possible if the

orthography does not use this diacritic as a diaeresis, i.e. to divide a diphthong or to highlight other features already discussed.

It might have been noticed that ROR proposes <u> for [y]. This is found in French, where, in contrast, [u] is <ou>. Some RMLs have a distinction between [ɔ~o, u, y] and, again, this choice is left to the orthographer. Striking a balance between aesthetics and phonemic faithfulness is particularly challenging, especially when the language presents all of [ɔ~o, u, y], possibly in unstressed syllables too (see §[Unstressed and reduced vowels](#)). In stressed syllables, it might be useful to utilise <u/ú>, <o/ó>, <>ò> for [y, u, o] respectively. Alternatively, one might take the French route: <u, ou, o>. Where the RML's phonology distinguishes even more vowels, it may be reasonable to allow some degree of phonological ambiguity to minimise the use of diacritics.

2.2.3.3 Unstressed and reduced vowels and schwa ([ə])

The schwa is most commonly written with <e> in the languages of reference, although other possibilities exist. Eastern Central Catalan alternates between <a> and <e> according to etymology. Romanian uses <ă>, but this use would not be reintegrationist, as diacritics in unstressed syllables of languages with mobile stress are discouraged (see below).

As already stated, diacritics in unstressed syllables of languages with mobile stress are discouraged. This means that diacritics that are normally used for stress (namely the acute and grave accents) can be used in RMLs that have fixed stress placement with the purpose of changing vowel quality. This mainly happens in Belgium and northern France.

Another important thing to keep in mind is stress alternation and paradigmatic consistency. In Eastern Central Catalan, [o, ɔ] always reduce to [u] in unstressed syllables. This means that [u] is written <o> in those unstressed syllables where it would be [o, ɔ] somewhere else in the paradigm. See Eastern Central Catalan *plorar* [plu'ra] “to cry” and *ploro* [ˈplɔru] “I cry”.

Some RMLs make a distinction between all of [e, ε, ə]. These vowels are traditionally transcribed as <e> in the languages of reference; but, especially in unstressed syllables in languages with mobile stress, it is hard to manage both ROR guidelines and make a distinction for all of the three. As stated for [ɔ~o] vs [u] vs [y], it might be better to accept ambiguity rather than overcomplicate the orthography.

2.2.3.4 Other guidelines for vowels

It might be useful to resort to etymology to avoid the crowding of diacritics or.

For example, in many varieties of Franco-Provençal, Latin *i* and [i] that was present at some point in the language evolved into [ø] in some instances. Since this is a pretty common sound, it may be best to spell it <i>, while reserving some other solution, maybe a diacritic on <i>, for [i]. For example, one might spell Aostan Franco-Provençal [ˈtʃøvra] “goat” as *\$chivra*; or [ˈføʎʎə] “daughter” as *\$fiʎe* (from Latin *FĪLIAM*).

2.2.4 Vowel Diacritics

2.2.4.1 Stress

grave vs acute

2.2.4.2 Length

preferably through 1) circumflex or 2) single vs double/multiple consonants after the vowel

3 CASE STUDIES

3.1 STANDARD ROMANSCH

3.2 BOLOGNESE EMILIAN

3.3 RIFONDOU WALON (UNIFIED WALLOON)

4 IMPLEMENTATION AND FURTHER CONSIDERATIONS

- // vs [] “sound/phone” vs “phoneme”
- ...

5 CONCLUSIONS