

# 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. In this sense, ROR can be said to encourage **trans-minority** contact and unity.

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] <â> 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).

## 2.2 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 \*I<sub>OCAT</sub> (“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*

IŪNIUM “June” → It. *giugno*

MEDIUM “middle” → Ven. [ˈmezo] \$*megio*, *mejo*, *mexo*, etc.

\*I<sub>OCAT</sub> → 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*

IŪNIUM “June” → Cat. [(d)ʒun] *juny*

MEDIUM “middle” → Ven. [ˈmezo] \$*megio*, *mejo*

\*I<sub>OCAT</sub> → Ven. [ˈzoga] \$*gioga*, *joga*

GENERUM “son-in-law” → Ven. [ˈzenero] \$*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](#)).

### 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”.

Cat. *alçar* [al'sar] (infinitive)

Cat. *alces* ['alses] “you lift”

Sp. *pez* [peθ]

Sp. *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> in front of <a, o, u> 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)l(u)m**). <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**, **ș**.

Genoese Ligurian ['køʃa] \$cheuxa

Picard (Nort-Leulinghem) ['kuiʃ] \$cuix

Bolognese Emilian ['kɔ:ʃa] \$còxa



⟨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 [ts]); 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.

## 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 attempting 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 and schwa \(\[ə\]\)](#)). 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* [plɔru] “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.

For example, in many varieties of Franco-Provençal, Latin *i* and [i] that was present at some point in the language evolved into [ə] 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 [ˈtʃəvra] “goat” as \$*chivra*; or [ˈfəʎʎə] “daughter” as \$*fiye* (from Latin *FĪLIAM*).

## **2.2.4    Suprasegmentals and vowel diacritics**

### 2.2.4.1    Stress

In languages with unpredictable stress placement, it is advisable to establish a default stress position that remains unmarked in writing. Only deviations from this default should be indicated using diacritics. For reference and inspiration, one may consult the stress-marking

conventions employed in Catalan, Spanish, and Portuguese, which provide well-established models of this approach.

Stress is best indicated through acute and grave accents (for the difference between the two see §[Vowel quality](#)), but there can be exceptions: as already stated, diacritics in unstressed syllables of languages with mobile stress are discouraged, so they can be useful to indicate stress placement and vowel quality at the same time. See the following example.

Latin CINEREM → Piedmontese \$*cënn*er ['s<sup>ː</sup>nn̩er]

In this case, I chose an umlaut on <e>—already established in Piedmontese—to indicate that the pronunciation is not [e~ɛ], but [ə]. In this case, an umlaut is allowed, as it is on a stressed vowel. In practice, <“> marks both stress and vowel quality.

#### 2.2.4.2 Length

Vowel length distinctions, present historically in some of the reference languages, have been lost in most contemporary dialects. I will present two proposals for representing vowel length within a reintegrationist orthographic framework.

**Circumflex accent.** As per the French tradition, a circumflex may be used to mark vowel length. For example, one might choose to spell Zoagli Ligurian [fu'gw<sup>aː</sup>] “hearth” as \$fogoâ.

**Doubling the following consonant.** This strategy is attested in Emilian and can be utilised if a circumflex is not a viable choice. See the example of *mānnd* in §[Economy](#).

As already mentioned, some ambiguity might be needed to avoid using diacritics on the vowels of unstressed syllables.

#### 2.2.4.3 Vowel quality

The following are proposed strategies for representing vowel qualities. It is essential that the orthographer consider all the considerations previously outlined. These suggestions are intentionally broad and may not be suitable for every RML. Etymology may offer valuable guidance in the search for an appropriate orthographic solution.

SOUND	GRAPHEME
[a]	<a>
[e]	<e, é>
[ɛ]	<e, è>
[i]	<i>
[o]	<o, ó>
[ɔ]	<o, ò>
[u]	<u, ou, o, ó>
[y]	<u, ui, iu>
[ø~œ]	<eu, oeu, œu, oe, œ, eo>
[i]	<ï, i, ë, ü>
[ə]	<e, ë>
[ʌ~ʌ]	<ö, ä>

Table 4 - Proposed sound-to-grapheme correspondence

## 3 CASE STUDIES

### 3.1 RIFONDOU WALON (UNIFIED WALLOON)

Walloon is a *langue d'oïl* spoken mainly in Wallonia. *Rifondou walon* (or *rfondou walon*; RW) is a pan-dialectal, standard variety of Walloon developed by a movement of the same name in the 1990s. Since it is pan-dialectal, it aims to unify the various Walloon dialects, primarily through harmonising their orthographic representation of sounds.

#### 3.1.1 Phonology

Although efforts have been made to assign a standard pronunciation to each grapheme, *rifondou walon* is, at its core, an orthographic endeavour; therefore, I will assign a place of articulation to graphemes just to help the reader grasp the system.

### 3.1.1.1 Consonants

Labial	Coronal	Postalveolar	Palatal	Dorsal	Glottal
m	n		gn	ng	
p	t	tch		c/k	
b	d	dj		g	
f	s/ss/ç	ch/sch/xh/sh			h
v	z/s	jh			
				r	
	l		y/i		
			u	w/ou/o	

Table 5 - Walloon consonants

### 3.1.1.2 Vowels

Front				Central	Back		
i	î	u	û		ou	oû	
é	én	eu		i*		ô	
e/ae	ê/ai	in/im	un		o	å	on/om
a	an/am						

Table 6 - Walloon vowels

\* <i> is often [ə] in unstressed positions.

Other digraphs: <ea, oe, oi, oen>.

### 3.1.2 Rifondou walon: General observations

As may already be apparent, rifondou walon (as almost all Walloon orthographies) is *phonetically* based on French; that is, the way it spells sounds is naturally grounded in how standard French spells sounds. This, of course, goes against ROR principles.

Another point to consider is that Walloon uses graphemes that are not found in the languages of reference—or elsewhere in the Romance world, for the most part—for example, <k, xh, jh, w, å>.

The orthography is also not etymologically sound, given its purely phonemic nature and its French-derived system.

Since Walloon is a *langue d'oïl*, I will mostly follow the French etymological tradition.

### 3.1.3 Consonants

#### 3.1.3.1 Latin yod (Walloon)

Latin yod—in the environments and conditions described in §[Latin yod](#)—evolves into RW <dj>.

Latin *iocum* > Wall. *djeu*

I propose <j> (*\$jeu*) for the reflex of Latin yod in this environment. This roughly follows the Catalan route, though preserving the phoneme's etymology, much like French spelling works.

#### 3.1.3.2 Modern yod (Walloon)

I propose <y> for the sound [j] in Walloon, following the French tradition. Also, [j] is written as <i> after a consonant, retaining the RW convention.

#### 3.1.3.3 The velar table (Walloon)

	NON-PALATAL ([k, g])		PALATAL ([s, dʒ])	
	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>
Latin g	<gu>	<g>	<g>	<j>

Table 7 - Proposed velar table for Walloon

This velar table follows French tradition in the *etymological* sense; in other words, it is a simplification of the French velar table. Here are some examples.

<i>kinoxheu</i> “knower”	→	<i>\$quenoxeu</i>
<i>candji</i> “to change”	→	<i>\$cangi</i>
<i>gueuye</i> “snout, mouth”		(unchanged)
<i>gozi</i> “neck”	→	<i>\$gosî</i>
<i>ceréjhe</i> “cherry”	→	<i>\$cérésje</i>
<i>çoula</i> “that”		(unchanged)
<i>djambe</i> “leg”	→	<i>\$jambe</i>

#### 3.1.3.4 Palatal [ɲ] (Walloon)

As in French, [ɲ] will be spelt <gn>.

*agnon* “onion” (unchanged)

#### 3.1.3.5 Latin cs and sc (Walloon)

Latin cs (x) generally results in <xh> in *rifondou walon*. However, since <x> is mostly avoided in RW—even though it aligns with ROR recommendations—it could be preferable to adopt it in place of the digraph <xh>.

*coxhe* “thigh” → *\$coxe*

Latin *sc* often evolves into *RW* <sch>. In my opinion, this is an acceptable grapheme.

### 3.1.3.6 The remaining consonants (Walloon)

The following is a list of graphemes that have been changed based on their etymology and correspondence with French and Latin and on ROR principles. The graphemes that don't change are not described below.

- <tch> → <ch> (etymologically the same as French)
- <ch> /ʃ/ → <ș> (based on its use in Romanian)
- <jh> → <sj> (as this sound often comes from [sj] > [zj]; if modern [zj] does not exist, <si> would also be a viable option)
- [w] → <u> before vowels (as [u] does not exist in Belgium); <u> at the end of a word, including before silent <e>. For example *\$aeue* “water” [ɛ:w].

If, at the end of a word, there is a need to represent [y] after a vowel instead of [w], a diaeresis can be used: *\*\$aeüe* (non-existent word) [ɛ:ˈy].

## 3.1.4 Vowels

### 3.1.4.1 Nasal vowels (Walloon)

- <én> is acceptable.
- <in/im> represent [ɛ̃], so it would be more faithful to the pronunciation and more etymological to spell them <en/em>.
- <an/am> is acceptable.
- <ô> has various phonetic realisations, but the “standardised” pronunciation of the *Motî d' potche walon / francês et français / wallon* is [õ]. I propose <on/om>
- <on/om> ([õ]) come from earlier [ũ], so, in order to distinguish them from [õ], they should be spelt <un/um>.
- <un> represents [œ̃], a non-native Walloon sound, which I chose to spell <eun>.
- <ai> will be changed to <ae> in order to better reflect its pronunciation [ɛ:]

Nasal vowels will not be distinguished in writing before nasal consonants.

### 3.1.4.2 Front rounded vowels (Walloon)

I will retain the French tradition of writing [ɔ, u, y] as <o, ou, u>. <eu> is acceptable.

### 3.1.4.3 Length (Walloon)

Long vowels, for the most part, will be spelt using a circumflex.

- <â>, pronounced [a:~ɔ:], corresponds to a long [a], so it will be spelt <â>.
- The rest of the long vowels are accepted in their *RW* form.

### 3.1.4.4 The rest of the vowel sounds (Walloon)

- <ae> represents a vowel that ranges from [a] to [ɛ]. Even though it uses an umlaut, I propose <ä>, since Walloon has fixed stress on the last syllable (see [§Suprasegmentals and vowel diacritics](#)).



- For the unified, pan-dialectal spellings <ea, oe, oi, oen>, I propose <ea, oi, oa, oin>. This is because RW <oe> comes from Old/Middle French <oi>, whereas the original RW <oi> does not.

### 3.1.5 Other considerations

#### 3.1.5.1 Liaison and silent <e>

As other *langue d'oïl*, Walloon features liaison. Therefore, if a phoneme typically has liaison, but in a given instance is fully pronounced, it should be written with a silent <e>. I also propose, following the French tradition, that all feminine nouns end in <e> (see \$*aeue*) and all words ending in etymological /dʒ/ (pronounced [tʃ] now at the end of words) is also written finally with an <e>—so, in practice, they should end with <ge> (see *prononçaedje* → \$*pronunçäge* “pronunciation”).

#### 3.1.5.2 Schwa [ə]

RW spells the schwa, variously realised in the dialects, as <i>. I propose <e>, as it corresponds to French <e>. When <e> is not stressed, it is assumed to be /ə/. When unstressed <e> represents [e] or [ɛ], it will be written as <é> and <è> respectively, with accents.

#### 3.1.5.3 Other silent letters

All the rules of regular, word-final silent letters in French apply to my proposal as well. Additionally, where there are vowels which are optionally removed, I have chosen to include those vowels in writing.

### 3.1.6 Example text

#### 3.1.6.1 Original *rifondou walon*

Li “prononçaedje zero-cnoxheu” ni sieve a wai d' tchoi pol moumint, veyanmint ki l' manire di lère li rfondou walon est co foirt loyî a l' accint do pays di tchaeke câzeu. Mins dins les ptits motîs, la k' on n' a waire di plaeece, i pout vni a pont po dner ene idêye d' onk des prononçaedjes possibes a des zero-cnoxheus.

#### 3.1.6.2 An example of how ROR could modify *rifondou walon*

Le “pronunçäge zèro-quenoxeu” ni siev a uai d' choa pol moument, vèyament qui l’manire de lère le refondou walun\* est co foart loyî a l’accent do pays de chèque câseu. Mens dens lès petits motîs, la qu’un n’a uaire de plâce, i pout veni a punt po dener ene idêye d’unque dès pronunçâges possibs a dès zèro-quenoxeus.

\**walun*: here the <w> is retained to maintain the identity of the language in its name.

## 3.2 STANDARD ROMANSCH (RUMANTSCH GRISCHUN) (COMING SOON!)

### 3.2.1 Phonology

#### 3.2.1.1 Consonants

LABIAL	CORONAL	(ALVEOLO-) PALATAL	VELAR	GLOTTAL
m	n	ɲ	(ŋ)	
p / b	t / d	c / ɟ	k / g	
	ts	tʃ		
f / v	s / z	ʃ / ʒ		(h)
		j		
	l	ʎ		
	r			

Note that [c, ɟ] are also transcribed elsewhere as [tʃ, dʒ]. Phonemes in parenthesis are distinctive only in some dialects.

#### 3.2.1.2 Vowels

FRONT	BACK
i (y)	u
e (ø)	o
ɛ	ɔ
	a

Phonemes in parenthesis are distinctive only in some dialects.

### 3.2.2 Consonants

#### 3.2.2.1 Latin yod

[...]

#### 3.2.2.2 Modern yod

Following the current Romansch tradition, we will preserve the grapheme <j> for modern [j].

[jau] §jau “I”

### 3.2.2.3 The velar table

[...]

## 3.3 BOLOGNESE EMILIAN

See [my YouTube video about ROR](#) for an example on Bolognese Emilian. Know that ROR was at its early stages when the video was published, so some points may be somewhat different from the present proposal.

## 4 FURTHER CONSIDERATIONS

### 4.1 SOUND VS PHONEME

It might have been noticed that the words sound, phone and phoneme have not been properly differentiated in this proposal. This is because another choice that the orthographer has to make is whether to represent phonemes (i.e. distinguish in writing only the sounds that present minimal pairs in the language) or to also mark non-phonemic distinctions. ROR leaves the orthographer or language community free choice in this case.

### 4.2 RECYCLING TRADITIONAL CONVENTIONS

The last recommendation Romance Orthographic Reintegrationism makes is to reuse graphemes or conventions that have been traditionally used in the orthography, as long as they do not violate the other principles of ROR.

For example, it has been brought to my attention that Venetian <j> (i.e. following the so-called “Catalan route”—see [\\$Latin yod](#)) is appropriate according to the prescriptions described in [\\$Latin yod](#) and following paragraphs, but it would not be intuitive to native speakers. Instead, it has been suggested to adhere to the “Italian route” and mark both the reflex of Latin yod and intervocalic GI/DI in the same way, using <z>.

While adherence to reference language conventions strengthens the reintegrationist aim, practicality and intuitive readability for native users must not be disregarded. ROR recognises that orthographic acceptance depends not only on linguistic rigour but also on community resonance; therefore, it supports adapting historically justified conventions in ways that remain accessible and meaningful within each specific linguistic context.

## 5 CONCLUSIONS

Romance Orthographic Reintegrationism (ROR) offers a principled yet flexible framework for the design of spelling systems in Romance minority languages (RMLs). By rooting choices in the etymological practices of major Romance traditions—while always keeping reader-oriented accessibility in view—ROR seeks to build bridges across the Romance continuum without imposing a one-size-fits-all solution. In addition to encouraging the reuse of familiar graphemes and empowering orthographers and communities to determine where to

designate phonemic versus non-phonemic distinctions, its principles strike a balance between historical depth and orthographic economy. In the end, ROR is not merely a list of guidelines; rather, it is an invitation for linguists, planners, and language groups to work together to develop writing systems that both honour their own heritage and speak clearly to the wider Romance world.

@Sicilanguageist