

The orthography of machine-readable Neolatin texts: A plaidoyer for minimal intervention

J. Ramminger (version of 5 October 2006)
[originally: Considerations for the Early modern latin meeting, Tufts U, Boston, 2-4 June 2006],

Contents:

Abstract

The role of the editor

Some changes are necessary

Editors can do what they want, but shouldn't

Neolatin orthography is neither arbitrary nor uniform

'Etymological' orthography: Examples

Contemporary and ancient Greek

Neolatin and Romance languages

Some Neolatin features will successfully resist 'modernization'

'Retro-spelling' new words

Proper names

Fixed orthography, or: Modernizing Mozart ?

Automated 'correction' can create new problems: A test case

What is standard orthography?

The user's needs and expectations

Coping with orthographical diversity: Designing a search machine

Further reading

Many of the following samples are taken from CAMENA, a project which over the years has been tremendously useful for me. If the following sometimes appears to criticize it, this is merely from the wish to contribute to an enterprise on which I as well as many researchers have come to rely as an indispensable tool of reference for early modern studies. Please note that I refer to the texts as contained in the html-files posted by CAMENA. When I wrote the first draft of this article I was not aware of the fact that the introduced changes are documented in the xml-files posted by CAMENA in parallel.

Abstract:

The Latin of the early modern period (Neolatin) is an independent stage in a continuing development of Latin. It is not a failed attempt to write correct Latin, which we can help succeed by improving its orthography. Standardization of orthography was developed with and for editions printed on paper. Publications on the web (in the following equally called 'editing') offer much more flexible editorial models.

'Normalizing' a text according to a presumed classical orthography should be avoided for two reasons: 1) The orthography of Neolatin texts is hardly ever uniform, but it is not therefore arbitrary; on the contrary it often reflects either deliberate or unconscious choices of the author. 2) There is no such thing as classical orthography (modern lexica diverge, and for new words there obviously cannot be a classical orth.).

Machine-readable texts allow completely new types of statistics-based research of post-medieval Latin - if we don't destroy the evidence by imposing our own version of normalcy. Scholars can easily change the original text to suit their purpose, but cannot roll back undocumented changes to get to the original state.

For many task on the agenda of the Boston meeting we need orthographical standards. These should not be implemented in the original texts, where - once enacted - they block further evolution, but in an intermediate layer, where they can be adapted and refined to meet the needs of present (searching, parsing, tagging) and future tasks.

The role of the editor

Some changes are necessary

Every time we transcribe a text, we interfere with it. This is unavoidable (in any medium), but should be kept at a minimum. Basically, editorial interventions should make the information in a text accessible, not reduce it.

E.g., since many modern readers will not be familiar with standard abbreviations / contractions / suspensions (*sm* = *secundum*, *roe* = *ratione*, *cotur* = *communicatur*, *ptas* = *potestas*, to name but a few), these need to be expanded. If a change does not improve understanding, it need not be introduced (ex.: *ferè* changed to *ferè*).

Some arbitrary decisions are unavoidable. Contractions are often ambiguous. Should *nūquam* be transcribed as *nunquam* or *numquam*? We try to follow examples where a given word is written in full, but these may be nonexistent or not uniform.

In some texts word breaks are not hyphenated. Establishing hyphenation means costly human intervention; in many cases, where one or both parts of a split word are meaningless (e.g. *au-tem*), hyphenation could be introduced automatically. Coping with hyphenation in the course of a search itself can be left to the retrieval software (see [last paragraph](#)).

Tachygraphical signs are to be expanded: æ-Ligature to *ae*, the ampersand to 'et', etc.

Punctuation should be respected, but often needs to be translated into modern systems, since many of the same signs (e.g. the colon, even the full stop) mean something different in earlier texts. This is a thorny issue (cost), because it needs a person with sufficient knowledge of Latin.

Editors can do what they want, but shouldn't

Classical texts are 'normalized'. So why not do the same with Neolatin texts?

For most texts from antiquity the original orthography is lost, because they are only transmitted by medieval witnesses. If we could, we would surely try to respect Cicero's orthography. The same ought to be the case with the orthography of modern Latinity which, in contrast to that of Cicero, is amply documented.

Most texts on the net today are based on one witness, whose orthography therefore can be followed. 'Poor' spelling may reflect an author's lack of competence, and is thus part of the text. On the other hand, the editor needs to identify misprints that do not reflect the author's/writer's intentions (e.g. CONTZEN *pol* p.934b *Atchistratergo* for *Archistratego*).

Before Gutenberg, - less so later -, texts often are transmitted in several witnesses with equal authority. Even in texts with a more rich textual transmission an orthography akin to the author's can be established (ex.: Lorenzo Valla's letters ed. Regoliosi, Salutati's correspondence ed. Novati).

Neolatin orthography is neither arbitrary nor uniform

The impression of arbitrariness can be misleading. Uniformity of spelling was not a goal of early modern authors, but they may try to be consistent in points they consider important. Moreover, orthography may offer clues to the linguistic environment in which the author worked. We lose such messages by standardization.

'Etymological' orthography: Examples

The Italian humanist Niccolò Perotti usually spells Pliny as *Plynius*, not *Plinius*, because he connected it with the Greek *plyno*.

Humanists since Tortelli's *De orthographia* were prone to believe that *ceterum* came from Greek *kai heteros* and therefore considered *caeterum* the more correct spelling; if anything we would have to standardize in the opposite direction from the modern norm.

aequiperare in Neolatin texts is usually spelt *aequiparare*, probably from a dim impression of kinship with *comparare*. If we change this we remove a feature which may be important for the understanding of the text.

Contemporary and ancient Greek

Greek loanwords are often spelt following the CONTEMPORARY pronunciation of Greek; if they are changed into 'correct' forms (ex.: Lat. *chiromantis*, Gr. *cheiromantis*) evidence of the influence of contemporary Greek pronunciation is lost.

Greek words in Latin texts follow contemporary standards, not modern norms of classical Greek (ex.: Vida, *De dign. rep.*, 2,54,6 *pimena* for classical *poimena*).

Whether there is aspiration in Greek loanwords, depends as much on an author's knowledge of Greek as on the evolving state of Greek philology in general - since Medieval Latin is mostly unaware of the aspiration, which had already ceased in Greek before the end of the Classical period (cp. Hugutio, *Derivationes*, C 306.13 *ypocrisis*). But despite the victory of Erasmian pronunciation*, as late as 1582 there is an example of *ypo-* for *hypo-* (Bruno, *Cantus circaeus*, p.187 *ypocrisim*).

Se also *Iphicrates/Hiphicrates* below.

*) See Chrys C. Caragounis, "The Error of Erasmus And Un-Greek Pronunciations of Greek", in: *Filología Neotestamentaria* 8 (1995) 151-185, URL: www.bsw.org/?l=72081&a=Art06.html.

Neolatin and Romance languages

'Reforming' Neolatin orthography will cut it off from the Romance languages.

praesumptuosus is an established spelling AFTER antiquity and leads (through intermediate stages) to Engl. *presumptuous*. The correct classical spelling is *praesumptiosus*.

Some words do not have what even remotely resembles a standard spelling. *sonetum/sonettus/sonetium*: Clearly the spelling reflects assumptions about the etymology, the relationship to the volgare and about how the volgare evolved from earlier Latin.

cygnus (swan) is akin to Ital. *cigno* and Fr. *cigne*. Neolatin texts have both *cygnus* and *cycnus*. Introducing the uniform spelling *cycnus* (proposed e.g. by CAMENA's reglat-tool) destroys this connection. In this case the spelling in antiquity is mostly a matter of conjecture based on etymology (and is hardly ever transmitted in classical texts).

If we 'normalize' Latin translations from modern languages, we may inadvertently destroy evidence for the influence of the original (ex.: Pierre d' Avity - Johann Ludwig Gottfried, *Archontologia cosmica*; Bodin, *De republica*).

Some Neolatin features will successfully resist 'modernization'

'Retro-spelling' new words

Should we reform *poenitet* to *paenitet*, but leave the papal *poenitentiarius*, an office which antiquity did not know?

What about the famous dat./abl. plur. *hiis*? Connect it with *hic*, *haec*, *hoc*, or *is*, *ea*, *id*?

Proper names

Even proper names known from antiquity can have divergent spellings, that may reveal not only an author's linguistic competence, but also the sources he used.

Girolamo Vida, *De dignitate rei publicae*, 1,37,2 writes *Hiphicrates* for a name, which in Athenaeus, *Deipnosophistae*, is spelt *Iph*-. Clearly this cannot be an intervention of the typesetter. Should we improve Vida? The same name is earlier spelt *Hyphicrates* in Gauricus, *De sculptura*, p.257.

If we interfere with *Virgilius/Vergilius*, we eradicate a discussion which has occupied a good many humanists.

See also *Cygnea* below.

Fixed orthography, or: Modernizing Mozart ?

Parts of a Neolatin text may be determined by external factors (e.g. being set to music) and thus resist modernization. Should the text of Mozart's Requiem (and countless masses) *Pleni sunt coeli ... gloria tua* be improved to *caeli* ? Should such phenomena be left unchanged in a text otherwise normalized?

Parts of a Latin text may be encrypted (e.g. in the diplomatic correspondence of the Chancellor of Sweden, Axel Oxenstierna); if these were transcribed with 'modernized' spelling, the code would become nonsense. In such a text, preferably also the parts not encrypted should preserve the original spelling (btw, *Oxenstiernaprojektet* retains the original orthography).

If we modernize the Latin, in plurilingual texts parts in other languages (French, German) for consistency's sake should also be modernized; but of course nobody would want to interfere with French spelling of the 17th century. Better also to leave the Latin.

Automated 'correction' can create new problems: A test case

The automated correction of *cygn-* to *cycn-* has created problems in several passages in the Mannheim corpus.

Cygnea, Latin name for Schwanenstadt, Zwickau. Matthiae, *Theatr. hist.*, p.1050b *cum ... Cygneam et Torgam ... occupavit*. CAMENA changes the name to *Cycn-*; this seems to be without basis, I have not been able to find a single instance where the place name *Cygnea* is originally spelt *Cycnea*.

Confusion is introduced into the Lemma CYGNUS in Hofmann's *Lexicon universale*, where the spelling -g- is retained in the Lemma (unavoidable because of the alphabetical position between CYDRUS and CYTES), but changed to -c- in the text.

Completely mystifying is Nolte, *Antibarbarus*, p.128 "in veteribus libris legi ... *cignus*, pro ... *cycnus* l. *cycnus*", until one realizes that the first *cycnus* is not Nolte's, who wrote "*cygnus* l. *cycnus*", a reference to his previous discussion of this word (p.47). The initial statement in these pages in CAMENA refers the reader, if in doubt, to the page image. But in many cases the reader will be unsuspecting initially, or even assume a word does not occur, if he cannot find it, because its spelling has been changed.

cygnaeus/cycnaeus/cycneus: Sometimes, *cygnaeus* is 'emended' to *cycneus* (Widl, *Carmina*, p.268), but *cycnaeus* remains unchanged. A further drawback: With the change from *cygnaeus* to *cycneus* the long vowel in the penultimate syllable is no longer visible (*cycneus* can also be measured with short -e-, e.g. in Boiardo, *Pastoralia*, G 5.63).

There is the further possibility *cicn-*, e.g. Taubmann, *Postuma schediasmata*, p.217 *Cicnos* (registered as a variant by CAMENA). The spelling is rare, but not unique, e.g. Verino, *Carlias*, 6.619 *leporem aut cicnum* (Byzantine pronunciation?). A reader might expect also these (and further?) cases to be normalized.

--

Even a seemingly harmless or even apparently useful 'improvement' like the change of *u* before a vowel to *v* can have side-effects; thus in GOTTFRIED *Avity arch* p.334 *In coronatione regum suorum vel cum bellum immineret solebant Pervani ducentos pueros trucidare*, in the CAMENA-transcription the Peruvians (*Peruani*) have become 'very vain' (*pervani*).

What is standard orthography?

There is no unified 'classical' spelling (cp. Lewis&Short vs. OLD vs. ThLL). An early example of the contradictory result of standardization (unavoidable in print) is the fact, that the TLL spells some compounds with ad- not assimilated (e.g. *adfleo*), but inserts them in the alphabetical sequence as if they were assimilated (i.e. *adfleo* after *afflatus*), while others appear under ad-, even though assimilated forms prevail in the ms. tradition (e.g. *admoneo/ammoneo*, which thus appears some hundred pages before *adfleo*). Electronic publications need not make that kind of decision.

Also, surprisingly, modern conventions may be untenable etymologically.

cum: Modern editorial conventions for classical texts mostly write *cum* for both the conjunction and the preposition. Already in the Renaissance some scholars were aware that the conjunction *cum/quom* should be distinguished from the preposition and spelt it *quum*, *qum* or *quom*. Are we to normalize that away? Or introduce *quum* everywhere?

In some cases modern conventions make texts more difficult to read. Esp. the separation of enclitics (-*ve* and -*que*) favoured by some Neolatin authors can be helpful. Obviously, this feature cannot be re-introduced, but might as well be retained if already existent.

The user's needs and expectations

Large corpora of machine-readable texts make new kinds of research possible (e.g. phonetics). We should not preclude this by destroying the evidence on which such research could be based.

cygnus/cycnus: A cursory study of the available texts seems to indicate that *cygn-* prevails in earlier texts, while *cycn-* is the spelling prevalent in the later part of our period. This may reflect a growing tendency towards etymologically correct spelling. *cicnus* in Verino may be influenced by late Byzantine pronunciation. Evidence for this development should not be destroyed by standardization.

Partial changes in a text increase the confusion instead of decreasing it. How should the reader know where we changed ? And what standard we are following ?

Since uniformity may be unattainable, we should not lead the user to expect it. In many cases we will leave the user guessing what we consider the norm.

Coping with orthographical diversity: Designing a search machine

Standardization is needed for many tasks. It is the advantage of an electronic text that this need no longer be the duty of the editor. It can be delegated to a middle layer between text and user, where the question of the user is transformed into a query adequate for the text. It can be the task of the editor of an electronic text to empower the reader to make informed choices. We should not disenfranchise the user by preempting decisions about the text.

One obvious use is searching. Google, while meritorious and used by us all, is unsuited for searches in heavily inflected languages such as Latin. Neither does it index long texts completely nor - contrary to a widely held belief - support the joker (*) in search queries.

We need search algorithms which, e.g., transform a search for *cael-* into one for *coel-/cel-/cael-*, *committ-* into *conm-/comm-*, which correctly understand the ampersand, æ- and œ- ligatures (if retained), *i/j/y*, *u/v*, *gracia/gratia*, cope with hyphenation/word breaks, Some such algorithms were developed for the NLW, more broadly for the Salutati-CDROM, and are implicit in Camena's reglat-tool.

Such a search machine could even be open for texts spread over other sites, which follow wildly differing conventions and are now only retrievable by chance. Similar mechanism could be applied to other task and allow for a flexible and adaptive model of text retrieval.

Further reading

A thoughtful analysis of some of the problems discussed above and others can be found in: Thomas Stäcker, Die Lateinischen Briefe Athanasius Kirchers an Herzog August Herzog d.J. zu Braunschweig und Lüneburg (1579-1666) - eine Internetedition. Preprint 2003, <http://www.hab.de/bibliothek/wdb/preprint/2003-11-staecker.pdf>>here.