Computer Aided Hyphenation for Italian and Modern Latin

Claudio Beccari

Abstract

After an essential historical sketch of the evolution of latin into italian and modern latin, the peculiarities of both languages are described so as to understand the philosophy of the hyphenation patterns. The latter are one of the few cases where the same set serves two different languages.

Sommario

Dopo aver delineato brevemente l'evoluzione del latino verso l'italiano e il latino moderno, vengono descritte le caratteristiche delle due lingue in modo da capire la filosofia dei *pattern* di divisione in sillabe. Questi *pattern* costituiscono uno dei pochi esempi applicabile a due lingue differenti.

Summarium

Latini sermonis evolutione ad italianum et latinum modernum breviter exposita, utrius sermonis specietates descriptae sunt ut philosophia de *pattern* ad syllabas dividendi intelligatur. Isti *pattern* duobus differentibus sermonibus applicabile exemplum sunt.

1 Outline of historical evolution

Classical latin as we study it in schools and universities is the language that was used, especially in written form, by the authors of the republican period and of the very beginning of the empire. Common people spoke a similar language that was open to the contribution of new words from other countries, to new constructs and to a general simplification of the inflection of nouns, adjectives and verbs.

Cicero himself was complaining about the fact that common people (the *vulgus*) used to shorten the desinences leaving out the final consonants, and used to palatalize the 'c' and 'g' followed by the front vowels 'e' and 'i'. Those were the first signals of the autoctonous evolution of latin towards the modern language; in the other parts of the Roman empire similar evolutions were going on with a stronger influence of the native languages over which latin had superimposed itself; the invasions of the "barbarians" brought in peculiar pronunciations and a lot of lexical additions.

Latin decline was very slow because it was the scholar's, the chancellor's, the notary public's language for many centuries, and it was and still is the official language of the Roman Catholic Church; latin, in its modern form, is the official language of the Vatican State, and the daily Vatican newspaper, *L'Osservatore Romano,* is published mainly in italian, but with frequent contributions in latin, even commercial adds! Modern latin is used even for comics books; I suggest Snoopy [**?**], Mickey Mouse [**?**], Asterix [**?**][1].

Nowadays latin is studied in many countries as a regular subject both in high school and in universities; in Italy it is not classified as a "foreign" language and is a compulsory subject both in classical and scientific *licei* (high schools). In the past, latin was even more important in the education of young people; forty years ago I started latin in sixth grade and had eight years of it through junior high and high schools[2].

From the common people's language of the first century several regional and local dialects evolved; in 960 A.D. there is the first document explicitly written in what we might already call italian [**?**]; several documents, mostly including poems, were produced in the following centuries, and by the end of the thirteenth century the masterpiece of Dante Alighieri, the *Divina Commedia*, is considered the main landmark of the new language, that was already so mature as to be used in a poetic treatise of history, philosophy and theology.

The modernization of Dante's language took place during the past seven centuries, but compared to modern italian there is not such a great difference as between the language used by Chaucer in his *Canterbury Tales* and modern english; today's italian high school students can read Dante's poem and other even older texts with no more difficulty than that required by any other conceptual text.

2 Alphabet

Italian and modern latin use the 26 letter alphabet that everybody knows with the name of *latin alphabet*; actually there are some fine points to consider with due attention.

Italian. The letters J, K, X, Y, and W are used only in technical terms and symbols, foreign names and some very specialized words, such as the international word *taxi*. J, K and Y survive in toponyms, family names, and english style nick names, such as Stefy for Stefania (Stephanie). The letter J (see

---

[1] The former two books are intended as didactic aids for teaching latin, and are fully accented with both prosodic and rythmic marks.

[2] I frequented the *liceo classico* and had also five years of classical greek; now I have an engineering degree and I am a professor of electric circuit theory. I am very glad I had the opportunity of completing my education by studying humanities for so long, and I wish the new generation could have the same.

also below) used to be employed in the past as a graphic device to distinguish the semivowel role of the letter I, so that you have *Ajmone* (family name) and you may write *Iugoslavia* (modern spelling), *Jugoslavia* (old fashioned spelling), and *Yugoslavia* (international spelling) according to your preference; in italian all three are correct and are pronounced exactly the same way.

Besides the above mentioned letters, there are five vowels, none of which is mute: *a, e, i, o, u,* fifteen consonants: *b, c, d, f, g, l, m, n, p, q, r, s, t, v, z,* and one diacritical letter: *h.* The latter does not correspond to any sound and is used only to mark half a dozen words in order to distinguish them from similar ones that sound the same but have a different meaning, to mark some interjections, and to mark the velar pronunciation of 'c' and 'g' when otherwise they would be palatalized.

Except for a dozen among articles, prepositions and adverbs (that nevertheless are used quite often), all common words in italian end with a vowel; of course this statement does not apply to trade marks, not assimilated foreign words, technical terms, and the like.

Another peculiarity is that every consonant may ▮ occur in its doubled form, and this corresponds to its reinforcement when the double consonant is pronounced. There are rare instances of double vowels, but in this case, contrary to what happens in english, they form different syllables instead of a diphthong; for example, *zoologico* can be divided in *zo-o-lo-gi-co.*

Latin. Classical latin missed J, U, and W, while V was used throughout wherever now U or V are used. Since the very beginning this anomaly was passed by the scholars on into the spelling and printing of all languages; capital V was used in all circumstances, while 'v' was used in printing at the beginning of words and 'u' in the middle or at the end. This confusing habit was common to all western languages but fortunately it was abandoned starting in Holland during the sixteenth century; it lasted a little more in Italy because of the wide use of latin, but was eventually done away by the end of the seventeenth century. When Knuth [**?**, reference 106] cites Pacioli's *Diuine Proportione*, published in Venice in 1509, he reports that title with the spelling of the original printing, but the pronunciation at that time already implied the consonant V instead of the vowel U.

In the middle ages and in the early times of printing there was the habit of using 'j' instead of 'i' in those cases where the letter 'i' formed a diphthong with the following vowel; it was just a graphic trick to distinguish the two roles of the letter 'i', and it was so successful that it was adopted also in other languages; this is the reason why even today we spell *junior* instead of *iunior*, although the latter is the formal latin spelling.

Modern latin uses both U and V in the proper positions, while J and W are used only in foreign names and toponyms.

There are six vowels: *a, e, i, o, u, y* and eighteen consonants: *b, c, d, f, g, h, k, l, m, n, p, q, r, s, t, v, x, z.* The ligatures *æ, œ* do not belong to latin; they were introduced in the sixteenth century in France and in England, and after that they enjoyed a certain popularity also in latin, but in modern usage, as well as in classical latin, these two diphthongs are spelled with separate letters.

3   Accents

Italian. In italian accents are used very sparingly; it is compulsory to mark with a suitable accent the last vowel of polysyllabic oxitone words (those that receive the stress on the last syllable), and to mark some well known and specified monosyllabic words that contain a diphthong. This is standardized by the Regulation UNI 6015 [**?**].

Contrary to spanish and portuguese, in italian there is no necessity to mark proparoxitone words with an accent, although the best grammars recommend to do so. In practice, if you exclude oxitone words (where accents are compulsory) and paroxitone words (where accents are not required), the other ones *may* be marked with an accent only when a different position of the stress might change the meaning; for example *làvati* means 'wash yourself' while *lavàti* is the masculine plural of 'washed'; in this circumstance it is advisable to mark the first case unless the meaning of the rest of the sentence does not make clear which case is implied. Although the 'Sommario' of this article contains five proparoxitone words, no accents were used.

The accent can be used also for denoting the open or closed nature of a vowel (only for tonic 'e' and 'o'), but this use is found only in dictionaries and grammars; a good grammar will certainly point out that *còlto* (picked up) is different from *cólto* (educated), but in practice the meaning is determined by the context while the actual pronunciation very strongly depends on the regional origin of the speaker.

The grave (`) accent is used on any vowel, while the acute (´) accent may be used only on the vowel 'e' (and on the vowel 'o', but only in optional situations) when it has a closed sound. Most Italians are not even aware of this choice; when they hand write,

they just put any kind of small surd on the vowel to be accented, and by so doing they intend to mark only the stress; the tonic value of the accent is used only in dictionaries and grammars, while in printing the difference is maintained only for the letter 'e' in oxitone words more as a tribute to the tradition than for an actual semantic necessity.

When the accent is compulsory and upper case letters are used, if the character set does not contain accented vowels, it is accepted to use an apostrophe: UNITA' (unity) in place of UNITÀ; this practice is considered bad style in typesetting, but is used quite often in advertising.

The diaeresis (¨) and the circumflex (ˆ) are not used anymore; in the past the diaeresis was used in poetry to split a diphthong, and the circumflex had several meanings such as, for example, to mark the contraction of two 'i' into one sign in those plurals that centuries ago were spelled with a double 'i': *studii* (studies, two centuries ago), *studî* (one century ago), *studi* (modern).

Latin. In latin no accents are used; the breve (˘) and the long (¯) accents are used only in dictionaries, grammars and where prosody is dealt with. The diaeresis is sometimes used in grammars and in prosody to mark the splitting of a diphthong: *aër* (air), *poëta* (poet).

4   Apocope and aphaeresis

Italian. In italian the dropping of one or more initial letters in a word (aphaeresis) takes place only in poetry and is marked with an apostrophe preceded by a white space.

The loss of one or more terminal letters in a word (apocope) either is not marked at all (see in the 'Sommario' *aver* in place of *avere*) or it is marked with an apostrophe when it corresponds to a vocalic elision (see above *l'evoluzione* in place of *la evoluzione*) or to a complete syllabic apocope. The latter case is very unusual, while the vocalic elision is very frequent, so that this case must be taken care properly when dealing with hyphenation; the rules stated in the Regulation UNI 6461 [**?**] require that when a line ends with an apostrophe, this *must not* be replaced back with the vowel it originally replaced. In the past, not too long ago, for example when I was in elementary school, the opposite rule was in use, so that there are occasional discussions between the old styled generation and the new one. Nevertheless even to day it is considered bad style to end a line with an apostrophe, and in typography this practice is tolerated only when the line width is quite small, as in the daily newspapers narrow columns.

Latin. I do not know of any case where apocope or aphaeresis are marked in any visible way; actually I am almost sure that these two spelling behaviours are not legal in latin.

5   Diphthongs

Italian. In italian a diphthong is formed by any vowel preceded or followed by an *unstressed* closed vowel ('i' or 'u'); so we have:

$$ia, ie, io, ai, ei, oi$$
$$ua, ue, uo, au, eu, ou$$
$$iu, ui$$

Italian diphthongs are always pronounced maintaining the sounds of the individual vowels, and the closed vowel plays the role of a semivowel or a glide.

There are also groups of three vowels that contain two semivowels or a semivowel and a glide:

$$iuo, uie$$
$$ieu, uoi, iei$$

An 'i' (possibly also an 'u', but I can't find examples) surrounded by two open vowels behaves always as a semivowel, so it always starts a new syllable.

Latin. In latin there are more or less the same diphthongs as in italian with the addition of

$$ae, oe$$

that one or two centuries ago were written with the corresponding ligatures æ, œ; in modern latin the pronunciation of both these diphthongs is given by a single open 'e'[3]. Furthermore in some words of greek origin, latin may have the diphthong *yi*, for example *Harpyia* [**?**][4].

The main difference between italian and latin common diphthongs is that *ia, ie, io, iu* behave as such in latin only when they are at the beginning of a word or are preceded by another vowel; in any other case they are part of two different syllables; in italian they are always diphthongs provided the 'i' is unstressed.

6   Di- and trigraphs

Italian. In italian there are groups of two or three letters that imply a sound that is not implied by any other single letter of the alphabet; besides 'c'

---

[3] I have seen a reproduction of an italian book printed in Venice in the sixteen century where both these diphthongs where replaced by their sound given by the letter 'e'.

[4] One might think that it would be the same to consider the vowel 'y' and the diphthong 'ia', since the pronunciation would be practically the same; but if you look at it from the prosody point of view, the situation becomes completely reversed; a diphthong is always long while 'y' is always short, so that in prosody Har-pyi-a becomes ¯¯˘, while Har-py-ia becomes ¯˘˘.

and 'g' modified with the diacritical 'h', and 'c' and 'g' modified with a diacritical 'i'[5] there are

$$gn, \; gli, \; sc$$

where *gn* is pronounced as in french, or as the spanish *ñ* or the portuguese *nh*; *sc* is pronounced as the english *sh* when is followed by a front vowel 'e' or 'i', and *gli* is pronounced as the portuguese *lh* when it is not preceded by 'n' and is followed by another vowel or it is at the end of a word. These digraphs and trigraphs must not be split by the hyphenation process.

Latin. In latin by itself there are no indivisible digraphs or trigraphs, but since the classical times the transliteration of greek words required *th* in place of *θ*, *rh* in place of *ρ* (but *rrh* in place of *ρρ*), *ph* in place of *φ*, and *ch* in place of *χ*; therefore these digraphs can not be split by the hyphenation process.

7 Hyphenation

Italian. The italian hyphenation rules are stated very simply as follows:

1. every syllable contains at least one vowel[6]

2. diphthongs and 'triphthongs' behave as one vowel■

3. a consonant followed by a vowel belongs to the same syllable as the vowel

4. one or more consonants not followed by a vowel (at the end o a word, possibly because of an apocope, or in technical terms, trade marks and the like) belong to the same syllable as the preceding vowel

5. when a group of consonants is found, the hyphen position is the *leftmost* one (even at the left of the whole group) such that the consonants that remain on the right of the hyphen can be found also at the beginning of an italian word;

6. prefixes and suffixes can be ignored and the compound word may be divided as if it were a single word; in any case the division according to the etymology is accepted; in practice this happens only with the technical prefixes *dis-, post-, sub-, trans-,* which are not very common.

Once it is clear what is a consonant, a vowel, a diphthong and a 'triphthong', the only difficult

---

[5] In this case the letter 'i' does not form a diphthong with the following vowel but is used just to palatalize the two consonants; under the hyphenation point of view this subtle difference may be ignored.

[6] This rule applies to all languages, although in every language the notion of vowel is different; for example in several slavic languages 'r' is considered a vowel. If TeX contained a provision for this, the bad line break (compara-nds) that occurred in *TUGboat*, vol.12, n.2, June 1991 at page 239, first column, 6-7 lines from bottom, would not have taken place.

b-d	b-n	b-s	c-m	c-n
c-s	c-t	c-z	d-g	d-m
d-v	f-t	g-m	p-n	p-s
p-t	p-z	t-m	t-n	z-t
g-fr	ld-m	ld-sp	l-st	mb-d
mp-s	nc-n	ng-st	n-scr	n-st
n-str	r-st	r-str	st-m	

Table 1: Groups of consonants that can be split across syllables

rule to apply is the rule number **??**; but with the help of a school dictionary one can always find if there exists an italian word starting whith a certain group of consonants.

The point is that if you use a dictionary of too high a quality, you will find words starting with almost any possible group of consonants: *bdelio*[7], *cnidio, ctenidio, ftalato, gmelinite, pneumatico, psicosi, pteridina, tmesi.* But many of these words, mostly of greek origin, do not find their way into school dictionaries (except *pneumatico* and *psicosi*), so that a diligent person will not be misled by too many technicalities and will find the correct division.

The Italian Standards Institute, in order to avoid■ confusion in this matter established the Regulation UNI 6461 [**?**] that lists the group of consonants that must be divided, table **??**. This table does not list the normal consonant divisions, that is

- digraphs and trigraphs can *never* be divided, except *gn* when it appears in a foreign word or in a word that derives from a foreign one and where the two letters are pronounced individually, such as *Wagner, wagneriano,...*

- geminated (double) consonants and *cq* must *always* be split

- a liquid ('l', 'r') or a nasal ('m','n') is *always* separated from a following consonant except for the cases shown in table **??**

- any consonant is *never* separated from the following liquid except for the cases shown in table **??**

- the letter 's' is *never* separated from any following consonant (unless it is another 's')

The Regulation UNI 6461 states also the rules for the apostrophe, i.e. it behaves as the vowel it replaces; line breaking (without hyphen) is allowed

---

[7] Due to the extremely specialized nature of these words, I do not give the translation in english, because I did not find a suitable italian-english dictionary that reported them; I believe, though, that their scholarly nature is such that with minor modifications they exist also in english and many other languages.

after it when the line is very short, but it is bad style to do it, so that line breaking is eliminated if no interword space is left between the apostrophe and the following word.

Italian hyphenation for TEX was already explained by Désarménien [**?**], but, although I wish I knew french as well as he knows italian, the 88 patterns that he created for italian were good only for consonants while completely ignored diphthongs and 'triphthongs'; in a previous version I prepared, 150 patterns were needed to perform italian hyphenation correctly.

For the rest the regulation is already made in such a way as to synthesize the hyphenation patterns TEX requires, without the need of running `patgen`; of course some care must be exercised in order to avoid strange situations and in order to replace TEX inability to distinguish vowels from consonants.

With the advent of Version 3.xx of TEX it is better to set `\righthyphenmin` to the value 2, because there is no need to protect the hyphenation algorithm from the mute vowels ('e') that are so frequent in english; of course it is not good style to go on a new line with just two letters, but this is so rare that it is much better to give TEX more chances to find suitable line break points than to protect it from situations that in italian never take place.

Another reason for choosing this reduced value for `\righthyphenmin` is due to the accents; it was pointed out that in practice italian has accents, if any, only on the last ending vowel of a word. It is known that TEX does not hyphenate a word after an accent control sequence, but this drawback has a negligible influence on italian since after the accent control sequence the word may have just one letter; when accented letters will find their way into the 256 symbol character sets, this simple drawback will be eliminated, but even with the actual limitations (unless virtual fonts are used) TEX peculiarity is of no influence; I admit that *virtù* (virtue) cannot be hyphenated because is too short (it could be hyphenated as *vir-tù*), while there are no problems with longer words, for example *qualità* (quality) is hyphenated by TEX as *qua-lità*, the full possibility being *qua-li-tà*. But TEX gives correctly *per-ché* (because), *af-fin-ché* (so that), and so on.

There are no known problems with the synthesized patterns listed at the end; the only point that leaves me partially unsatisfied but is grammatically perfectly correct, is the fact that technical prefixes such as *dis-, post-, sub-, trans-* must be explicitly separated with `\-` if one wants to stress their specific prefix nature. See below the solution for the same problem in latin.

Latin. The patterns that are listed at the end include a subset that was originally designed just for italian; with a little thought and few additions the pattern set was extended so as to include also modern latin.

For what concerns diphthongs, italian and latin diphthongs were merged together under the assumption that TEX is not supposed to find every possible break point but only legal break points, so that if two vowels are treated as a diphthong even if they belong to two different syllables, the only drawback is that you miss a legal break point but you do not make any wrong division. More over most Italian readers feel uncomfortable when a break point is taken such that the new line starts with a vowel (this is certainly not the case with anglophone readers) so that the extension of the set of diphthongs of either language does not bother neither italian readers, nor latin ones. The declaration of *æ* and *œ* as letters with their `\lccode` allows the hyphenation of words containing such ligatures, although their use is discouraged.

For what concerns consonant groups there is no regulation as for italian; my grammar [**?**] claims that latin hyphenation is done as in italian (except for what concerns prefixes and suffixes that must be divided etymologically) but in latin there are consonant groups that in italian never occur.

In order to find out how unusual consonant groups are treated in latin I examined an important scholar's book [**?**], the bilingual New Testament in greek and latin "apparato critico instructum", reprinted as a "reeditio photomechanica ex typographia ..., Romae" and for which "omnia iura reservantur"; clearly this is modern latin, although the book's contents, the latin part, contains the well known text that was translated from greek and aramaic by several authors across several centuries and copied by different copyists in many codices that are preserved all over the world. This critical edition is intended as a study material and is particularly cured in the language and the spelling for the very purpose of the book.

By examining the hyphenations of this book I could list a series of consonant groups, and I could realize that the digraph *gn* (which is such in italian but it is not supposed to be one in latin) was treated not uniformly so as to have both *reg-num* and *re-gnum*. I decided to chose the second form of hyphenation for two reasons: a) it does not conflict with the italian rule, and b) the pronunciation recommended to the clergy and that is being used in the catholic universities, seminaries, monasteries, etc., corresponds to the italian one.

Also the letter 's' is not treated uniformly; it is generally treated as in italian, but there are cases where it is treated as in english; for example *blasphemia* (blasphemy) is hyphenated as *blas-phe-mia*. Since this does not conflict with the italian rule (in this language the group 'sph' is missing) a suitable pattern was generated in order to cope with such situations.

Some attention was given to the prefixes and suffixes in order to find a way to separate them correctly according to their etymology; for what concerns prefixes, these must be separated regardless of the groups of letters that get split away, provided that the prefix did not loose its final vowel by elision with the initial vowel of the compound word second element. For example the prefix *paene-* (almost) looses the last 'e' in *paeninsula* and therefore the whole word is treated as a single word and is hyphenated *pae-nin-su-la*.

It was possible to find suitable patterns for certain instances of *ab-, ad-, ob-, trans-*, for the prefixes *abs-, dis-, circum-, sub-*, and for the suffixes *-dem, -que* but the problem remains, although it shows up not so often.

The solution can be found in a macro (already described by J. Braams [**?**]) that has been in use by the German TEX users, which have to cope all the time with compound words that need a little help for their correct hyphenation:

```
\def\allowhyphens{\penalty\@M\hskip\z@}
\def~#1{\if\string#1-
        \allowhyphens\-\allowhyphens
    \else
        \penalty\@M\ #1%
    \fi
}
```

Here this macro appears in a modified form; in the german version the character " (instead of ~) was made active and was given a complex definition so as to treat the umlaut in the proper way and to cover several other situations that occur in german. This implies several changes to be made here and there in the definitions of `plain`, in particular the double quote must be added to the list of special characters so as to deal with them in a consistent way when typesetting in verbatim mode. I preferred to give a new definition to the tie character ~, that is already listed among the special characters; this new definition performs the usual tie function except when is followed by the hyphen character; in the latter case the sequence ~- inserts a special discretionary break that has the property that normal hyphenation takes place in the rest of the word; remember,

in facts, that the standard sequence \- inserts a discretionary break but inhibits hyphenation in the rest of the word.

Therefore, if wrong prefix or suffix hyphenations ▇ are found in the drafts, it is possible to correct (or to write it that way since the beginning) `con~-iungo,  ob~-iurgo`▇ so that the possible hyphenation points are *con-iun-go, ob-iur-go*.

8   Generation of the format file

In the appendix the file `italat.tex` is listed and the patterns may be checked against the rules that have been stated in the previous sections. Special attention was given to the groups *ps* and *pn*, because the table **??** states that they must be separated, but the compound words with *psic-* (example *parapsicologia*) and *pneum-* (example *pseudopneumococco*) must not be hyphenated after the 'p'.

The ligatures 'æ' and 'œ' have been included with the `^^` notation, because the patterns can not contain control sequences; this poses no problems to the final user, because the hyphenation algorithm is applied after all macro expansions have been reduced to non expandable tokens.

The pattern list is preceded by some definitions:

- the category, lower case and upper case code definitions for the ligatures 'æ' and 'œ' so that they can be used in latin text; I stress again that these ligatures should not be used, except when quoting verbatim some text where they have been used.

- the definition of the special control sequence ~-;

- the definition of the new language "italian" with▇ the command (\italiano) that invokes all the auxiliary definitions; the apostrophe character must be given its \lccode=39 so as to treat it as a normal letter and as the vowel it replaces.

- the command for latin (\latino, ablative and short for "latino sermone") is simply \let to be identical with \italiano.

The patterns are enclosed within a group so that the \lccode of the apostrophe and the codes for the ligatures 'æ' and 'œ' remain local and do not mix things up with the default language and/or with the previously defined languages.

Adding these hyphenation patterns to the format that has one or more languages already defined is not a heavy overhead; if you add italian and latin to the default language 'english' you do not need a large version of TEX; the statistics, after running `initex`, say that the hyphenation trie is of size 6336 with 220 ops, 181 of which are used for english and

La lingua italiana e le lingue cosiddette romanze o neolatine, cioè lingue derivate anch'esse dal latino (francese, spagnolo, portoghese, rumeno ed altre minori), si fanno risalire all'idioma, che al tempo dell'impero romano era parlato nella penisola italiana, nelle regioni del Mediterraneo occidentale e nella Dacia, l'odierna Romania.

Tracce evidentissime si osservano ancor oggi non soltanto nel lessico e nella morfologia del gruppo linguistico neolatino, ma anche in altre lingue europee, quelle del gruppo anglo-sassone, come conseguenza dell'influsso diretto o indiretto esercitato dalla lingua di Roma sugli idiomi particolari dei popoli nordici.

Per quel che riguarda la lingua italiana, essa si collega direttamente al *sermo vulgaris latinus,* cioè al latino parlato comunemente dalle famiglie e in pubblico nei quotidiani rapporti di commercio e di affari.

Figure 1: Example of italian text typeset in narrow columns (from [**?**])

Et sicut Moyses exaltavit serpentem in deserto, ita exaltari oportet Filium hominis, ut omnis, qui credit in ipsum, non pereat, sed habeat vitam aeternam. Sic enim Deus dilexit mundum, ut Filium suum unigenitum daret, ut omnis qui credit in eum non pereat, sed habeat vitam aeternam. Non enim misit Deus Filium suum in mundum, ut iudicet mundum, sed ut salvetur mundus per ipsum. Qui credit in eum, non iudicatur; qui autem non credit, iam iudicatus est, quia non credit in nomine unigeniti Filii Dei. Hoc est autem iudicium, quia lux venit in mundum, et dilexerunt homines magis tenebras quam lucem; erant enim eorum mala opera. Omnis enim, qui male agit, odit lucem et non venit ad lucem, ut non arguantur opera eius; qui autem facit veritatem, venit ad lucem, ut manifestentur opera eius, quia in Deo sunt facta.

Figure 2: Example of latin text typeset in narrow columns (J 3,14-21)

39 for italian and latin; italo-latin hyphenation requires just 202 patterns (some of which probably never occur in practice) against the 4447 needed in english.

9 Conclusion

The hyphenation patterns valid for both italian and latin have been generated directly from the grammar hyphenation rules; for what concerns italian the set of patterns (a subset of that shown in the file `italat.tex` reported in the appendix) has been in use for two years in the Institution where I work, and after a short period of careful observation and debugging it performed absolutely without errors of any kind. Although the italian rules allow to hyphenate a compound word as if it were a simple one, some prefixes that are mainly used in technical terms may be explicitly hyphenated with the help of the special discretionary hyphen macro `~-`.

For what concerns latin the there is less experience but the impression is that also in this language there are no hyphenation errors; any how the author is grateful to anyone that might report suggestions and corrections. The special discretionary hyphen macro `~-` is very useful for prefixes and suffixes and must be used whenever unusual consonant clusters are generated by the apposition of a prefix or a suffix.

In Figures 1 and 2 two examples show the performance of the hyphenation algorithm in italian and in latin when the line width is very small; the line breaking tolerance is the default one (200) and in each example there are a just couple of underfull hboxes.

A   The `italat.tex` file

This file must be input after the last line of the file `plain.tex` (or `lplain.tex` for LATEX); the definitions given before the list of patterns are better located in the format file, so they are valid for any style and there is no possibility to forget them out.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%                       F I L E     I T A L A T . T E X
%
%               Hyphenation patterns for Italian and Latin
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%         Prepared by Claudio Beccari, Politecnico di Torino, Italy
%                       e-mail beccari@polito.it
%
% Version date  27 august 1991
%
% Useful definitions
%
\def\catcodeAE{\catcode 26=11 \catcode 29=11 \lccode 29=26    % Ligature ae,AE
               \uccode 29=29  \lccode 26=26  \uccode 26=29
               \catcode 27=11 \catcode 30=11 \lccode 30=27    % Ligature oe,OE
               \uccode 30=30  \lccode 27=27  \uccode 27=30}
\makeatletter %                       Because when this file gets read @ is "other"
\def\allowhyphens{\penalty\@M\hskip\z@}
\gdef~#1{\if\string#1-\allowhyphens\-\allowhyphens
          \else \penalty\@M\ #1\fi}
\makeatother%                                           Restore @ to "other"
%
% A number is given to italian/latin hyphenation
%
\newlanguage\italian
%
% The commands \italiano and \latino are defined
%
\def\italiano{\language=\italian \righthyphenmin=2 \lccode`\'=39 \catcodeAE}
\let\latino\italiano
%
% The patterns are defined within a group so that the \lccode of the apostrophe
% remains local and does not interfere with other languages
%
{\language\italian \catcodeAE \lccode`\'=39
%
\patterns{
.a2b2s3  .a2b3l
.o2b3l   .o2b3m .o2b3r      .o2b3s
.an1ti3  .a2p3n .di2s3ci3ne .cir1cu2m3 .wa2g3n
.ca4p5s  .pre3i .pro3i
.ri3a    .ri3e .re3i        .ri3o        .ri3u
.su4b3lu .su4b3r 2s3que.    2s3dem.
3p4si3c4 3p4neu1
^^Z1     ^^[1                                     %   Ligatures ae and oe
a1a   a2e   a2i   a2j   a1o   a2u  a2y             %   Diphthongs
a2y3o a3i2a a3i2e a3i2o a3i2u ae3u
e1a   e1e   e2i   e2j   e1o   e2u  e2y e3iu
i2a   i2e   i1i   i2o   i2u   io3i
o1a   o2e   o2i   o2j   o1o   o2u  o2y
o3i2a o3i2e o3i2o o3i2u
u2a   u2e   u2i   u2o   u1u   uo3u
1b2   2b3b   4b3d   2b3n   2b3t                    %   Consonant groups
```

```
        2b3s4a 2b3s4e 2b3s4i 2b3s4o 2b3s4u  2b3s4t    u2b3s4c
1c2     2c3c    2c3m    2c3n    2c3q    2c3s  2c3t  2c3z  2ch3h
1d2     2d3d    2d3g    2d3m    2d3s    2d3v  4d3w
1f2     2f3f    2f3t
1g2     2g3g    2g3d    2g3f    2g3m    2g3s
1h2     1j2     2j3j    1k2     2k3k
1l2a    1l2e    1l2i    1l2j    1l2o    1l2u
        1l2l3l  l3f4t   1l'     2l4l3m        1l2^^Z 1l2^^[
1m2     2m3m    2m3b    2m3p    2m3l    2m3n  2m3r    2m4p3s 2m4p3t 4m3w
1n2a    1n2e    1n2i    1n2j    1n2o    1n2u  2n3n    n2c1n  2n1l
        n2g3n   2n1r    n2s3m   n2s3f   2n'   1n2^^Z 1n2^^[
1p2     2p3p    2p3s    2p3n    2p3t    2p3z  2ph3p  2ph3t  2s3p2h
1q2     2q3q
1r2a    1r2e    1r2i    1r2j    1r2o    1r2u  1r2h    1r2^^Z 1r2^^[
1s2     2s3s    2st3m
1t2     2t3t    4t3m    2t3n    1t'     4t3w
1v2     2v3v    1w2     2w3w    wa4r
1x2a    1x2e    1x2i    1x2o    1x2u    2x3x  1x2^^Z  1x2^^[
y2a     y2e     y2i     y2o     y2u
1z2     2z3z    2z3t    1z'     }}
```

References

[1] Schulz C.M., *Insuperabilis Snupius*, translated into latin by G. Angelino, European Language Institute, Recanati, Italy, 1984

[2] Walt Disney, *Michael Musculus et Regina Africae*, translated into latin by C. Egger, European Language Institute, Recanati, Italy, 1986

[3] Goshinny and Uderzo, *Asterix gladiator*, translated into latin by K.H.G. von Rothenburg (*Rubricastellanus*), Delta, Stuttgart, 1978

[4] Migliorini B.M., *Storia della lingua italiana*, (History of the italian language), Sansoni, Firenze 1963

[5] Graham R.L., Knuth D.E., Patashnik O., *Concrete mathematics*, Addison-Wesley Publ. Co., Reading, Mass., 1989 (3rd printing)

[6] *Divisione delle parole in fin di linea* (Word hyphenation at the end of a line), published by UNI, Ente Nazionale Italiano di Unificazione, Milano, 1969

[7] *Segnaccento obbligatorio nell'ortografia della lingua italiana* (Obbligatory accent marks for the correct spelling of the italian language), bublished by UNI, Ente Nazionale Italiano di Unificazione, Milano, 1967

[8] Désarménien J., "The use of TeX in French: hyphenation and typography" in *TeX for scientific documentation*, D. Lucarella ed., Addison-Wesley Publ. Co., Reading, Mass., 1985

[9] Manna F., *Il latino ieri e oggi* (Latin yesterday and today), Signorelli, Milano, 1985

[10] *Novum Testamentum Graece et Latine* (The New Testament In Greek and Latin), A. Merk S.J. ed., Istituto Biblico Pontificio, Roma, 1984

[11] Braams J.,*Babel, a multilingual style option system for use with LaTeX's standard document styles*, *TUGboat* vol.12, n.2, June 1991, pp. 291-301

⋄ Claudio Beccari
Dipartimento di Elettronica
Politecnico di Torino
Turin, Italy
`beccari@polito.it`