LaTeX e le lingue romanze alpine

Claudio Beccari

Sommario

LaTeX può gestire molte lingue, attualmente 74 più diverse varianti. Alcune delle lingue gestibili da LaTeX sono usate da poche persone, altre da milioni. Questo articolo vorrebbe essere un primo passo verso l'estensione di LaTeX alle lingue romanze usate nelle regioni alpine d'Europa.

Abstract

LaTeX is used to typeset in a variety of languages: at the time of writing, it can handle 74 different languages (plus many variants), some of them used by very few people, some by millions. At the moment, there are no facilities for typesetting documents in the various more or less official Alpine Romance languages. This paper would be a first step in order to extend the LaTeX language capabilities to the various romance languages that are being used by large communities in the European Alps.

1 Introduzione

Le Alpi sono una barriera naturale fra varie nazioni, ma non lo sono state per le popolazioni che le hanno attraversate portando di qua e di là delle Alpi le loro tradizioni, i loro costumi e specialmente le loro lingue. Gli stessi antichi romani hanno conquistato le regioni alpine annettendo i territori e le popolazioni autoctone e, successivamente, estendendo loro la cittadinanza romana. In questo modo la lingua del potente divenne anche la lingua del sottomesso, come purtroppo è successo nei secoli fino ai giorni nostri; questo ha fatto estinguere le antiche lingue autoctone, ma ha trasferito molte delle loro caratteristiche nelle nuove lingue che si sono formate.

Ora procedendo da est a ovest sull'arco delle Alpi si parlano diverse lingue romanze che hanno avuto riconoscimenti ufficiali come il romancio in Svizzera, o come il friulano nella regione autonoma Friuli – Venezia Giulia e il ladino delle regioni Trentino – Alto Adige e del Veneto. Le province autonome di Trento e di Bolzano hanno stabilito leggi particolari per il plurilinguismo che viene praticato nei loro territori; immagino che anche la regione Veneto e la provincia di Belluno abbiano provveduto nello stesso modo.

Nel canton Grigioni in Svizzera c'è un marcato plurilinguismo che comprende non solo una comunità maggioritaria di lingua tedesca, circa il 68%, e una comunità di lingua italiana di circa il 10%, ma anche una varietà di parlate romanze non italiche che rappresentano le varietà di una lingua "comune", il romancio dei Grigioni, che è stata ricostruita a tavolino e che per legge è ufficiale in tutti i rapporti fra i cittadini e il governo cantonale e con quello federale.

Esisterebbero anche le lingue ticinese, lombarda e piemontese, di cui non parlerò in questo articolo, ma arrivando alle Alpi occidentali abbiamo ancora il franco-provenzale e l'occitano.

Il franco-provenzale è poco conosciuto e spesso viene confuso con il provenzale, ma è parlato da decine di migliaia di parlanti in tre nazioni: in Italia dal Piemonte occidentale e in tutta la Valle d'Aosta; in Svizzera nel Vallese; in Francia ad est delle Alpi nella Savoia, nel Delfinato e nella Franca Contea.

L'occitano si estende su un territorio vastissimo, che comprende il sud ovest del Piemonte, tutta la Francia meridionale per arrivare fino alla Val d'Aran in Spagna. Probabilmente l'occitano, fra queste lingue romanze minoritarie è quella parlata dal maggior numero di persone.

Fra le lingue citate l'occitano e il franco-provenzale hanno avuto uno sviluppo diverso dal francese da una parte e dall'italiano dall'altra e sotto certi aspetti grammaticali assomigliano più al francese e al catalano, mentre le lingue romanze minoritarie della Svizzera sud orientale e dell'Italia nord orientale hanno avuto uno sviluppo assolutamente staccato da ogni influenza dell'italiano come, sotto certi aspetti, hanno avuto invece le lingue ticinese, lombarda e piemontese. La differenza sostanziale è che la formazione del plurale dei nomi e degli aggettivi in italiano e nelle lingue minoritarie dell'areale italiano avviene mutando la vocale finale, se detti nomi ed aggettivi finiscono con una vocale, mentre nelle lingue romanze che accomunano i grigionesi, i ladini del Trentino – Alto Adige, e i friulani formano il plurale con l'aggiunta della 's'. Tolta questa caratteristica particolare devo dire che tutte le lingue romanze citate, dal friulano all'occitano, sono comprensibili per una persona di cultura e di lingua madre italiana, e i ceppi "ladini" da una parte, e "occitani" dall'altra appaiono come varietà di due lingue principali.

Certamente come persona di formazione tecnico-scientifica non ho molto da dire su queste lingue, salvo che riesco a leggerle tutte, comprendendo il 90% delle parole e ne riconosco delle similitudini che vanno al di là degli areali dove sono sviluppate.

In Italia esiste una legge dello Stato che tutela le lingue minoritarie. In Italia sono numerosissime:

oltre a quelle citate sopra, si va dallo sloveno, al tedesco, che nella provincia di Bolzano ha lo status di lingua ufficiale accanto all'italiano, al francese, che in Valle d'Aosta ha lo status di lingua ufficiale accanto all'italiano, alle lingue dei Walser del Piemonte e della Valle d'Aosta, al catalano di Alghero, all'albanese arbaresh in molte regioni del sud, al grecanico, al croato, e probabilmente ne ho dimenticata qualcuna. In Calabria esiste persino un'enclave occitana nella zona di Guardia Piemontese, la cui lingua prende il nome di gardiol.

Come ho detto, però, solo il tedesco e il francese hanno uno status ufficiale.

A livello regionale o provinciale tutte le regioni che insistono sull'arco alpino in Italia hanno provveduto alla emanazione di leggi che hanno valorizzato le lingue locali. In Piemonte esistono leggi regionali che forniscono supporto alla conservazione delle lingue locali e in particolare la Giunta Regionale ha avanzato una petizione all'UNESCO per il riconoscimento dell'occitano come patrimonio culturale dell'umanità. In Tentino – Alto Adige esistono disposizioni delle due provincie autonome per il supporto al ladino della Val Gardena (Bz) e delle Val di Non e Val di Fassa (Tn); in Friuli – Venezia Giulia esiste sia una legge regionale, sia quella della provincia di Udine per il sostegno al friulano. Della Svizzera ho già detto e il romancio ufficiale (rumantsch grischun) è riconosciuto come lingua federale.

In questo articolo mi riferirò al romancio svizzero (chiamato ladino in alcuni dei sotto areali dei Grigioni) e al friulano (chiamato anche ladino orientale); questa scelta riduttiva dipende dal fatto che solo il rumantsch grischun e il furlan hanno per legge una grafia ufficiale e riconosciuta dalle autorità locali; ma ricordiamo che, a parte la grafia di queste lingue, i parlanti possono comprendersi reciprocamente con maggiore o minore facilità.

Le differenze di grafia sono sostanziali; il rumantsch ha molti suoni rappresentati mediante trigrafi o tetragrafi imprestati dal tedesco, come si vede anche nel nome stesso della lingua dove compaiono "tsch" e "sch" con il medesimo valore fonetico che hanno in tedesco. Il furlan invece ha una grafia più simile a quella della lingua italiana, sia pure con l'uso della 'ç' e di alcuni digrafi, che non compaiono in italiano come "cj" e "gj", oltre all'uso diffusissimo dell'accento circonflesso per distinguere le vocali lunghe che in furlan hanno una rilevanza semantica.

Riduco ulteriormente il campo di azione: mi occuperò solamente del rapporto fra il romancio e LaTeX, ma mi riprometto di affrontare questa tematica anche per altre lingue alpine. Perché questa scelta riduttiva?

Per alcuni motivi, alcuni validi, altri molto meno: in primo luogo perché il romancio è una lingua nazionale, mentre il friulano ancora ha rilevanza solo regionale, benché sia parlato da un numero molto più grande di persone; in secondo luogo perché mi sono dovuto occupare di problemi di sillabazione per le lingue che usano l'elisione degli articoli e delle preposizioni articolate quando precedono una parola che comincia per vocale; infatti per conoscenza diretta fino a poco tempo fa conoscevo solo l'italiano, il francese, il catalano e il romancio come lingue che usavano questa particolarità grafica; ora so che questo uso riguarda di fatto tutte le lingue romanze tranne il portoghese, lo spagnolo/castigliano e il rumeno. In terzo luogo la grafia del romancio non fa uso di consonanti con diacritici e per questo mio primo approccio alle lingue romanze alpine questa caratteristica mi torna molto utile. Infine ho una grande simpatia per la Svizzera e i suoi abitanti di ogni lingua e parlata, e ammiro questa loro capacità di convivere malgrado le differenze linguistiche che riescono a superare, perché quando ci si vuol capire, ci si capisce con la buona volontà; nonostante le differenze hanno un senso della cittadinanza e dello Stato dal quale noi abbiamo molto da imparare. Li ammiro perché nella loro Confederazione hanno realizzato l'*Unità nella diversità*, che sarebbe il motto dell'Unione Europea, ma che noi cittadini dell'Unione non abbiamo ancora sentito come nostro. Per questi motivi sentimentali, ho quindi dato la preferenza al romancio.

Con il romancio mi sono trovato inizialmente in difficoltà ad applicare il piccolo strumento di correzione per superare i problemi dell'elisione vocalica quando ho scoperto che LaTeX ancora non è in grado di gestire il romancio; perciò mi sono dovuto creare l'intera struttura per la gestione di questa lingua.

2 Fonemi e grafemi del romancio

Il romancio si scrive con l'alfabeto latino e l'uso degli accenti grave e acuto su alcune vocali. L'insieme di fonemi e grafemi è mostrato nella tabella 1, con i fonemi rappresentati mediante i caratteri IPA (International Phonetic Alphabet)[1].

Come si vede la grafia del romancio non è strettamente fonetica, me le differenze rispetto alla grafia puramente fonetica sono molto simili a quelle dell'italiano e riguardano principalmente la pronuncia della 'c' e della 'g' diversa a seconda che siano seguite dalle vocali 'e' e 'i' oppure dalle altre vocali.

1. Con *pdflatex* i fonemi IPA sono accessibili mediante il pacchetto tipa, (FUKUI, 2004) fornito di una discreta documentazione; sfortunatamente questi caratteri, disegnati in accordo con la famiglia dei font Computer Modern, possono essere usati solo con questi font e non sono usabili con il font Latin Modern di questo articolo. Non è un problema eseguire gli adattamenti necessari, ma certamente è una cosa che non mi sarei aspettato. Con XeLaTeX e LuaLaTeX, che usano i caratteri OpenType, i grafemi sono tutti direttamente accessibili, sempre che il font di testo scelto come "main font" ne sia dotato.

Tabella 1: Corrispondenza fra fonemi e grafemi

Fonemi IPA	Grafemi
a, ɑ, aː, ə	a,à
e, eː, ɛ, ɛː	e, è, é
i, iː	i, ì
ɔ, ɔː	o, ò
o, oː, u, uː	u
aɪ	ai
aʊ	au
iə	ie
jeʊ	ieu
wa, waː, uːɑ	ua
k	c, q
kw	qu
tʃ,	c, ch, tg
ttʃ	tsch
χ	ch
g, dʒ	g, gh
ʎ	gl, gli
ɲ	gn
h	h
s	s, ss
ts	z
z	s
ʃ	s, sch
ʒ	s, sch
ʃtʃ	stg
j	i, j
b	b
d	d
f	f
l	l
m	m
n	n
p	p
r	r
v	v

Anche il suono della ‘c’ palatale viene rappresentato in modo diverso a seconda che sia all’inizio, al centro o alla fine della parola. Ma tutto sommato le differenze rispetto all’italiano non sono moltissime; molte dipendono dal fatto che le terminazioni consonantiche delle parole sono molto frequenti.

Pertanto questa grafia può dare qualche problemino per la sillabazione (vedi più avanti), ma non dà nessun problema per scrivere un testo con una tastiera svizzera o italiana.

3 Come LaTeX gestisce le lingue

LaTeX, come anche gli altri mark-up del sistema TeX, da XeLaTeX, a ConTeXt a LuaTeX, gestisce le lingue in due modi distinti.

3.1 La divisione in sillabe

In primo luogo il file di formato per ogni mark-up deve contenere le regole di sillabazione per ogni lingua da gestire. I file di formato sono file creati apposta su ogni macchina al momento dell’installazione del sistema TeX, oppure ricreati dall’amministratore della macchina, che contengono tutto il mark-up, le configurazioni, le regole di sillabazione delle varie lingue, e molto altro, direttamente tradotti nel formato interno di ciascun motore di composizione, in modo che il grosso delle funzioni che servono per comporre con i programmi del sistema TeX siano già immediatamente utilizzabili, senza bisogno di tradurle daccapo ogni volta che si lancia uno di questi programmi.

Le regole di sillabazione sono costituite da frammenti di parole, chiamati *pattern*, che contengono una sola lettera, due lettere, tre lettere, eccetera, dove a sinistra e a destra di ogni lettera è segnata una cifra che indica con il valore pari che la cesura in quella posizione è vietata, mentre con il valore dispari è permessa; ogni cifra passa la sua informazione specifica mediante il suo valore, quindi 0 indica una blanda proibizione di dividere mentre 1 indica un blando consenso alla divisione; i corrispondenti valori 8 e 9 indicano la massima proibizione o il massimo consenso a dividere. Se in una parola divisa in pattern se ne trovano alcuni che contengono alcune lettere nella stessa sequenza e contengono cifre diverse, prevale il valore della cifra più alta: per esempio, prendiamo il caso della ‘s’ pura e impura in italiano che non può mai essere divisa da quanto segue a meno che non si tratti di un’altra ‘s’ oppure che sia la seconda di due ‘s’ seguita da un’altra consonante (per esempio mass-mediologo); ecco allora che ci sarà un pattern ‘1m’ che dice che si può dividere prima della ‘m’, e un altro pattern ‘1s2’ che dice si può dividere prima della ‘s’ ma non dopo; poi c’è ‘2s3s’ per consentire la divisione della doppia ‘s’, ma c’è anche ‘s4s3m’ per dividere correttamente questa parola italiana composta con una radice inglese. Nel nesso ‘ssm’ i pattern di una lettera consentirebbero la cesura solo prima della prima ‘s’ ma non prima della ‘m’, perché il pattern della ‘s’ vieta di dividere dopo la ‘s’ stessa; i pattern con due lettere consentirebbero la cesura solo fra le due ‘s’, mentre il pattern con tre lettere vieta la cesura fra le due ‘s’ ma la consente fra la ‘s’ e la ‘m’; in definitiva, mettendo assieme i vari pattern i codici numerici di sillabazione corrisponderebbero a 2s4s3m e questa è la regola finale che consente la cesura solo fra la seconda ‘s’ e la ‘m’.

Questo è il meccanismo, ma è facile immaginare che per dividere in sillabe una intera lingua ci vogliano moltissimi pattern; per l’italiano ne occorrono circa 330 e virtualmente non producono mai errori di sillabazione; per l’inglese americano ne occorrono un po’ meno di 5000 ma circa il 10%

delle cesure di tutte le parole americane sarebbero divise in modo scorretto[2] (parole di Liang e Knuth); per il tedesco ne occorrono un numero enorme che supera le 10 000 unità. Perché LaTeX possa fare il suo lavoro per bene con questi grandi numeri di pattern è necessario che i pattern stessi siano memorizzati nel file di formato in strutture dati molto rapide da scorrere per la loro ricerca con i rispettivi valori numerici; per questo motivo essi devono essere predigeriti una volta per tutti da una versione "vergine" del motore di composizione (oggi quasi sempre *pdftex*, oppure *xetex* o *luatex*) per assemblarli in strutture chiamate *hash* perché LaTeX le possa analizzare con la massima rapidità. Assemblati gli hash di ogni lingua, il programma associa loro un numero univoco da legare al nome della lingua che verrà usato da babel o da polyglossia per selezionare le regole giuste per ogni lingua.

3.2 La lingua nella tipografia

In secondo luogo LaTeX deve provvedere a scrivere le parole fisse nella lingua giusta, deve scegliere l'hash giusto per la lingua selezionata, deve realizzare le strutture tipografiche particolari di ogni lingua: un semplice esempio: in italiano si possono usare le virgolette alte o quelle uncinate, ma non deve essere lasciato nessuno spazio fra le virgolette e quanto esse racchiudono: "parola" o «parola»; in francese si usano solo le virgolette uncinate con uno spazio fra ognuna di esse e il loro contenuto: « mot »; in tedesco si usano sia quelle a forma di piccole virgole diritte alte o basse, ma quelle di apertura sono al piede, non sono alzate: „Wort", oppure quelle uncinate a rovescio, rispetto a quello che si fa in Italia, e senza spazio: »Wort«.

Inoltre babel e polyglossia sono in grado di distinguere alcune varietà ortografiche, come per esempio in tedesco gennaio si chiama *Januar*, mentre in austriaco si chiama *Jänner*.

Questi sono pochi esempi, ma ovviamente le tradizioni tipografiche nazionali sono molto diverse anche fra nazioni che usano la stessa lingua. I pacchetti babel e polyglossia tengono conto di tutte queste varianti e della selezione delle lingue.

Un punto importante: i pattern non sono caricati da babel o da polyglossia; questo pacchetto sceglie solo fra gli hash *già presenti nel file di formato* e se non ne trova nessuno associato alla lingua specifica, usa la sillabazione americana che, evidentemente, non ha nulla a che vedere con la sillabazione delle lingue romanze, tanto per citarne un gruppo piuttosto ampio, ma nemmeno con le lingue scandinave o germaniche o slave; non parliamo delle lingue che si compongono con alfabeti diversi da quello latino.

Oggigiorno le distribuzioni ridotte del sistema TeX spesso fanno disperare i principianti perché non riescono a ottenere le cesure in fin di riga in modo corretto anche se hanno caricato babel con l'opzione di lingua giusta. Solo con le distribuzioni complete si hanno le cesure giuste per tutte le lingue che il sistema riesce a gestire; in caso contrario bisogna essere capaci di ricreare i file di formato con i pattern delle lingue che interessa usare; questa è una operazione non difficile, ma delicata, e richiede di procedere in modi diversi a seconda della distribuzione del sistema TeX di cui si dispone, a seconda che si vogliano creare i formati per il motore *pdftex* oppure per *xetex* oppure per *luatex*; spesso le distribuzioni dispongono di opportuni programmi di configurazione con i quali l'operazione diventa più semplice; tuttavia queste operazioni esulano dall'argomento di questo articolo.

4 File di descrizione della lingua e file di pattern

La funzione descrittiva della lingua esposta nella sezione precedente è svolta da (generalmente un solo) file con estensione `.ldf`. Invece i pattern possono richiedere diversi file, diversi perché devono essere usati solo con font aventi codifiche del tipo T1, oppure con codifiche UNICODE; c'è modo di costruire file che soddisfino entrambe le esigenze, ma la predisposizione dei pattern è piuttosto complessa.

4.1 Il file di descrizione della lingua

I file `.ldf` per babel hanno il nome formato con quello della lingua e con l'estensione indicata. I file per polyglossia hanno il nome formato agglutinando il nome della lingua con il prefisso `gloss-`; il nome della lingua è solitamente il nome in inglese e in lettere minuscole; per il romancio avremo perciò i file `romansh.ldf` e `gloss-romansh.ldf`.

La funzione del file `.ldf` nei due casi è sostanzialmente la stessa, ma mentre con babel la lingua è una opzione per il pacchetto, con polyglossia questo è un file di estensione autonomo che permette di definire alcuni comandi con i quali precisare le singole lingue da usare nel documento specificandone mediante opzioni anche alcune particolarità che sono disponibili grazie all'uso dei font OpenType. Non scenderò in questi dettagli; mi limiterò a descrivere il file `romansh.ldf` che contiene le definizioni importanti e assolutamente necessarie in ogni file di questo genere.

```
\LdfInit{romansh}{captionsromansh}%
\ifx\l@romansh\@undefined
    \@nopatterns{romansh}%
```

2. Le regole di sillabazione dell'inglese americano tengono conto dell'accento tonico delle parole: il nome *rècord* e il verbo *to recòrd* hanno la stessa ortografia ma diversi accenti tonici (che in inglese non si scrivono); a causa di questo fatto le divisioni sillabiche sono rispettivamente *rec-ord* e *re-cord*; LaTeX conosce l'ortografia, ma non la pronuncia quindi se divide correttamente una delle due parole sbaglia l'altra. In inglese le parole omografe ma non omofone sono moltissime, quindi LaTeX si trova in grande difficoltà.

```
4     \adddialect\l@romansh\l@italian\fi
5  \addto\captionsromansh{%
6    \def\prefacename{Prefaziun}%
7    \def\refname{Bibliografia}%
8    \def\abstractname{Recapitulaziun}%
9    \def\bibname{Index bibliografic}%
10   \def\chaptername{Chapitel}%
11   \def\appendixname{Appendix}%
12   \def\contentsname{Tavla dal cuntegn}%
13   \def\listfigurename{Tavla da las figuras}%
14   \def\listtablename{Tavla da las tabellas}%
15   \def\indexname{Register da materias}%
16   \def\figurename{Figura}%
17   \def\tablename{Tabella}%
18   \def\partname{Part}%
19   \def\enclname{Agiunta(s)}%
20   \def\ccname{Copia a}%
21   \def\headtoname{A}%
22   \def\pagename{pagina}%
23   \def\seename{vesair }%
24   \def\alsoname{vesair era}%
25   \def\proofname{Demonstraziun}%
26   \def\glossaryname{Glossari}%
27   }%
28 \def\dateromansh{%
29   \def\today{\ifcase\day\or 1.\else
30     ils~\number\day\fi~da~\ifcase\month\or
31     schaner\or favrer\or mars\or avrigl\or matg\or
32     zercladur\or fanadur\or avust\or settember\or
33     october\or november\or december\fi\space
34     \number\year}}%
35 \providehyphenmins{\CurrentOption}{\tw@\tw@}
36 \addto\extrasromansh{%
37   \babel@savevariable\clubpenalty
38   \babel@savevariable\widowpenalty
39   \babel@savevariable\@clubpenalty
40   \clubpenalty3000\widowpenalty3000%
41   \@clubpenalty\clubpenalty}%
42 \addto\extrasromansh{%
43   \babel@savevariable\finalhyphendemerits
44   \finalhyphendemerits50000000}%
45 \addto\extrasromansh{%
46   \lccode‘’=‘’ }%
47 \addto\noextrasromansh{%
48   \lccode‘’=0 }%
49 \ldf@finish{romansh}%
50 \endinput
```

Credo che siano opportuni alcuni commenti. Innanzi tutto la prima linea di codice serve per l'inizializzazione, ma fra le altre cose definisce `\captionsromansh` che serve a babel tra l'altro per controllare se il file sia già stato caricato.

Le righe da 2 a 4 verificano se esista un hash collegato alla lingua romancia; se un tale hash non esistesse, sarebbe necessario comunque dare un significato al comando `\l@romansh` associandolo all'hash di un'altra lingua come se ne fosse una varietà. Io ho preferito associarlo all'hash dell'italiano, perché sulla mia macchina lavoro con una distribuzione completa del sistema TEX; se questo file fosse usato in una distribuzione di base, non è detto che l'hash collegato all'italiano sia stato generato, e quindi una associazione come quella che ho fatto io potrebbe dare origine ad errori; per evitarli sarebbe necessario eseguire un altro test; data la provvisorietà di questa descrizione della lingua romancia ho preferito omettere questo lievissimo perfezionamento.

Le successive righe dalla 5 alla 27 definiscono tutte le parole fisse che compaiono nei titolini o in certi ambienti; ho preso questi nomi dal vocabolario tedesco–romancio, messo a disposizione on line dalla Lia Rumantscha (la Lega Romancia, una istituzione grigionese per la promozione del rumantsch grischun); in parte ho preso queste traduzioni da libri in romancio che ho trovato in rete.

La data in romancio è definita nelle righe da 28 a 34; come si vede ci vuole l'articolo plurale prima dei numeri dei giorni superiori a 1, mentre per il primo del mese è necessario indicare l'ordinale secondo la tradizione tedesca, cioè con la cifra seguita dal punto ma senza articolo (non lo si scrive ma viene detto a voce quando si parla: *l'emprim da schaner 2012*).

Le righe da 35 a 44 impostano alcuni parametri per la sillabazione; la sillaba iniziale e quella finale non devono contenere meno di due caratteri; siccome moltissime parole romance terminano con gruppi di consonanti, tocca ai pattern provvedere a che l'ultima sillaba contenga almeno una vocale.

Inoltre sono impostate le penalità per le righe vedove e le righe orfane; il valore 3000 è abbastanza alto ma non infinito, per cui le righe orfane e vedove potranno essere presenti ma *pdflatex* cercherà di evitarle con sufficiente determinazione. Il valore enorme assegnato a `\finalhyphendemerits` serve per evitare che la penultima riga di un capoverso subisca la cesura in fin di riga; in questo modo l'ultima riga di un capoverso non comincia (quasi) mai con una frazione di parola.

Infine le righe da 45 a 50 impostano e disimpostano la condizione per la quale l'apostrofo possa essere considerato come parte di una parola in romancio e non lo sia più quando si cambia lingua. Le ultime due righe completano le impostazioni per il romancio e finiscono il file; dopo `\endinput` potrebbe esserci scritta qualunque cosa, ma questa parte del file non verrebbe nemmeno letta dal motore di composizione.

4.2 I file per la sillabazione

I file per generare gli hash di sillabazione sono tre: `loadhyph-rm.tex`, `hyph-quote-rm.tex` e `hyph-rm.tex`; `rm` è la sigla del romancio secondo le norme ISO; anche gli altri file destinati allo stesso scopo per le altre lingue hanno i tre file con i nomi strutturati nello stesso modo, salvo che `rm` è sostituito da `it` per i file destinati all'italiano, `de` per i file destinati al tedesco, eccetera[3].

3. Ci sono alcune eccezioni, ma in sostanza questa è la convenzione generale per dare un nome ai file dei pattern. Invece in greco, per esempio, esiste `loadhyph-grc.tex` per caricare i pattern per il greco classico validi se si usano i font codificati con la codifica di default `LGR`; mentre esiste un altro file che carica i corrispondenti pattern se si usano i font codificati in "BETA code"; le due codifiche si distinguono perché nella prima i segni diacritici sono premessi alle lettere a cui si riferiscono, mentre con la seconda i segni diacritici

Il primo file carica gli altri due quando si devono creare o ricreare i file di formato, al momento di generare gli hash per la sillabazione, non durante la composizione di un particolare testo. Il secondo contiene le impostazioni per la codifica dell'apostrofo (ma potrebbe contenere anche altre macro di utilità per scrivere i pattern); infine il terzo file contiene i pattern veri e propri.

Ho provato innanzi tutto a non generare il formato con i pattern per il romancio ed ho usato quelli italiani al loro posto come indicato nel paragrafo precedente. Mediante un mio programmino per scrivere tutte le parole di un testo spezzate in sillabe, ho verificato come funzionassero i pattern italiani per il romancio; evidentemente c'erano degli errori ma erano relativamente pochi; quasi tutti riguardavano le terminazioni consonantiche singolari e plurali del romancio, visto che in italiano normalmente tutte le parole terminano con una vocale e quindi l'italiano, in teoria, non dovrebbe prevedere pattern per parole terminanti in consonante[4]. Questi erano errori evidenti perché l'ultima "sillaba" trovata dal motore di composizione, non era una sillaba secondo la grammatica, in quanto non conteneva nemmeno una vocale; anche gli altri pochi errori erano tali da isolare una "sillaba" priva di vocali.

Sono inoltre riuscito a trovare in rete la *Grammatica per l'instrucziun dal rumantsch grischun*, molto completa e ben fatta e lì ho trovato le regole per la divisione grammaticale in sillabe; sono molto simili a quelle per l'italiano finché sono enunciate a parole, ma diventano specifiche per il romancio in molti dettagli che hanno richiesto molte modifiche e correzioni rispetto ai pattern italiani, come descrivo più sotto. Le regole sono le seguenti:

- Ogni sillaba contiene una vocale o un dittongo o un trittongo; in romancio le vocali possono essere accentate; i dittonghi e i trittonghi sono solamente *ai, au, ie, ia, iu, ui, uo, ieu, uai*. Con lo stesso concetto dei pattern per l'italiano ho scritto i pattern per il romancio evitando di esplicitare le vocali, cosicché, eventualmente, non vengono divisi alcuni iati, ma non vengono commessi errori dividendo gruppi vocalici dove non è consentito.

- Ogni consonante semplice o digrafo o trigrafo o tetragrafo seguito da una vocale o da un dittongo o un trittongo fa sillaba con la vocale che segue. Analogamente non si dividono altri gruppi che formano "suoni semplici" (di fatto sono quasi tutti digrafi, trigrafi e tetra grafi): *ch, tg, gl, gn, qu, sch, stg, tsch* e, nei toponimi, *s-ch, dsch*.

- Le consonanti doppie si dividono fra le due sillabe adiacenti.

- Due consonanti diverse si dividono fra le sillabe adiacenti se la coppia di consonanti non può trovarsi all'inizio di una parola romancia; più precisamente, in presenza di un gruppo di due o più consonanti la divisione va fatta lasciando a destra della cesura il massimo numero di consonanti che possano trovarsi all'inizio di una parola.

- La regola precedente implica quindi che a destra della cesura si può cominciare con *bl, cl, fl, pl, vl, br, cr, dr, fr, gr, pr, tr, vr*, ma non si può lasciare a destra della cesura *dl, sl, tl, sl, sr*; non dispongo di un *pledari* (dizionario) romancio, quindi non so se esistano delle parole che cominciano per *sl, sr*, che invece in italiano esistono; quindi i relativi pattern devono essere aggiustati a seconda della lingua; per esempio *pasler* si divide fra la 's' e la 'l'. In tutti gli altri casi, anche se una consonante è seguita da una liquida, *l, r*, la divisione avviene fra le due consonanti, al contrario dell'italiano; per esempio il nome della Svizzera, *Svizra* si divide fra la 'z' e la 'r'.

- La 's' seguita da una delle consonanti o digrafi *c, d, p, t, tg* non consente la cesura dopo la 's'.

- I prefissi si dividono etimologicamente al contrario dell'italiano dove la sillabazione fonetica è sempre lecita, anche se non è vietata la sillabazione etimologica; bisogna però tenere presente se dopo la preposizione-prefisso la parola che segue comincia con una consonante o con una vocale; nel primo caso la cesura cade fra la preposizione e la parola, mentre nel secondo caso si fa lo stesso solo se la parola esiste anche isolata. Per esempio il prefisso *ex* nella parola *examinar* non viene diviso etimologicamente perché la parola *aminar* non esiste isolata. Dunque per poter eseguire le divisioni etimologiche bisogna avere il dizionario a disposizione, cosa che LaTeX non può fare.

- Nello stesso modo occorre eseguire la divisione etimologica nelle parole composte evitando di dividere le parole componenti; questo LaTeX lo consente inserendo a mano il comando per la cesura facoltativa `\-`, comando che impedisce

sono aggiunti dopo la lettera a cui si riferiscono. Altri pattern greci invece di `gr` contengono la sigla `el`.

4. Questa leggenda metropolitana sopravvive da tempo immemorabile; certo la stragrande maggioranza delle parole termina in vocale, questo è vero, ma l'italiano accetta non solo l'elisione ma anche l'apocope che può lasciare le parole monche e quindi anche terminanti in consonante; inoltre esistono una quantità di termini scientifici e tecnici, sportivi, architettonici, medici, eccetera che terminano in consonante o perché prestiti assimilati da lingue straniere, o perché nati così come voci dotte greche o latine (straniere fino ad un certo punto, almeno quelle latine); esiste certamente almeno una parola italiana che termina con ciascuna delle 21 consonanti dell'alfabeto latino! E i pattern per l'italiano prevedono tutte queste terminazioni consonantiche.

la divisione negli altri punti della stringa che costituisce la parola composta.

- È consigliato di non dividere gli iati, ma la divisione è consentita nella divisione di parole composte, come *extra-ordinari, co-operativa*, eccetera. In questi casi i pattern per il romancio che ho predisposto prevedono la divisione al margine della separazione delle due parole componenti, ma non ne impediscono la divisione. Se necessario, si può intervenire a mano con `\-`.

- Le necessità tipografiche, non la grammatica, richiedono di non dividere dopo la prima sillaba o prima dell'ultima sillaba di una parola se questa prima o ultima sillaba è formata da una sola vocale.

- È vietato andare a capo dopo l'apostrofo, esattamente come in italiano.

Con queste regole ben esplicitate nella grammatica romancia, (Caduff *et al.*, 2009), ho creato i pattern per il romancio semplicemente prendendo i pattern per l'italiano che già facevano abbastanza bene i loro lavoro all'interno delle parole e ho aggiunto i pattern specifici del romancio, correggendo eventualmente quelli dell'italiano, quando erano in contrasto con le regole enunciate sopra. I pattern totali sono saliti a 378, una cinquantina in più di quelli per l'italiano. Ho rigenerato i file di formato e sulla mia macchina ora posso scrivere in romancio senza problemi.

Ecco ora un breve testo in romancio tratto dalla suddetta Grammatica[5]:

> En rumantsch n'èn las reglas da comma betg uschè strictas sco p.ex. en tudestg. Sco regla generala vala: nua ch'ins vul far ina pausa en la frasa, mett'ins ina comma. La comma è en rumantsch per ordinari il segn d'ina pitschna pausa e betg il segn d'ina disposiziun sintactica sco en tudestg. Las reglas da comma dependan damai pli ferm da quai ch'ins vul exprimer. Ellas han surtut la funcziun d'inditgar la melodia, il ritmus e las pausas ed èn tras quai in agid per chapir meglier il text. Tuttina n'è il diever da la comma betg arbitrar; i dat er contexts, nua ch'ella è fixa.

5. In romancio le regole della virgola non sono così stringenti come in tedesco. Come regola generale vale la seguente: dove si vuol fare una pausa nella frase mettiamo una virgola. La virgola in romancio serve per indicare il segno di una piccola pausa e non il segno di una costruzione sintattica come in tedesco. Le regole della virgola dipendono di più da ciò che si vuole esprimere; esse hanno soprattutto la funzione di indicare la melodia, il ritmo e le pause con le quali capire meglio il testo. Tuttavia l'uso della virgola non è arbitrario; ci sono dei contesti dove la virgola è fissa.

5 Conclusione

Conto di inviare i file di gestione del Romancio ai due team che si occupano di sillabazione e di descrizione delle lingue; un team si occupa di babel e l'altro di polyglossia; con il secondo team credo che non ci sano problemi di sorta; con il primo, invece, credo che ce ne saranno, perché a me appare inattivo: non so se sia solo una impressione mia o se si tratti di un fatto concreto.

Chiaramente sarebbe desiderabile che il sistema TEX sia in grado di gestire il romancio nel giro di poche settimane, invece che fra qualche anno; ma bisogna rimettersi alle tempistiche del lavoro volontario dei membri dei due team.

Per me è stato un esercizio molto interessante e mi ha dato la possibilità di arricchire le mie conoscenze linguistiche in campi che non avrei mai pensato di esplorare; la descrizione in questo articolo, benché il romancio sia solo una curiosità per la maggior parte dei lettori italiani, ha un valore didattico perché permette di rendersi meglio conto di chi fa che cosa fra i vari elementi che costituiscono il sistema TEX.

Riferimenti bibliografici

Braams, J. (2011). *Babel, a multilingual package for use with LATEX's standard document classes*. TUG. Leggibile con `texdoc babel` nella distribuzione TeX Live.

Caduff, R., Caprez, U. N. e Darms, G. (2009). *Grammatica per l'instrucziun dal rumantsch grischun*. Seminari da rumantsch da l'Univesitad da Friburg. Dipartament da lingugatgs e litteraturas romanas, Friburg, CH. URL `http://www.unifr.ch/rheto/documents/Gramminstr.pdf`.

Charette, F. (2011). *Polyglossia: a Babel replacement for XELATEX*. TUG. Leggibile con `texdoc polyglossia` nella distribuzione TeX Live. L'attuale versione è affidata alla manutenzione di Arthur Reutenauer.

Fukui, R. (2004). *TIPA manual*. Graduate School of Humanities and Sociology, Tokyo University, Tokyo. Leggibile con `texdoc tipa` nella distribuzione TeX Live.

Liang, F. M. (1983). *Word Hy-phen-a-tion by Com-puter*. Tesi di Dottorato, Department of Computer Science, Stanford, CA. URL `www.tug.org/docs/liang/liang-thesis.pdf`.

▷ Claudio Beccari
Villarbasse
`claudio dot beccari at gmail dot com`