



Free Software Magazine

e

GUIT

presentano

combelow: abbasso i segni diacritici di serie B

Gianluca Pignalberi

Pisa, 17 ottobre 2009

Indice

- 1 Sommario
- 2 Introduzione
- 3 L^AT_EX e cediglia
- 4 Comma below e casi speciali
- 5 Peripezie di un codice
- 6 `combelow 0.99a`
- 7 `combelow 0.99b`
- 8 `combelow 0.99c`: un fantastico passo indietro
- 9 `combelow 0.99d`
- 10 Sillabazione e crenatura
- 11 Carattere vietato in Unicode
- 12 Conclusioni, sviluppi futuri e ringraziamenti

Sommarío

Sommarío

Romeno e lettone devono essere considerate lingue di serie B nel mondo di T_EX? Se finora lo potevano essere, questo piccolo pacchetto tenta di riportarle al livello giusto con il semplice uso del segno diacritico corretto. Niente piú cediglia al posto del comma below.

Abstract

Should Romanian and Latvian be considered second choice languages in the T_EX world? Though up to now they could have been, this small package tries to put them back at the right level, just using the correct diacritic mark. No more cedilla instead of comma below.

Introduzione

Nei giornali di lingua romena si trovano spesso parole con un segno diacritico strano: una virgola centrata sotto le lettere 's' e 't'.

Non è possibile riprodurre queste lettere con L^AT_EX, visto che non c'è un comando per comporle.

Su alcuni giornali si trova, al posto della virgola sottostante, una cediglia (,). A seconda del font, questa ha una forma a uncino.

Introduzione

Dilemma

Qual è il segno corretto da usare?

Per capire dove fosse la ragione ho ritenuto di fare affidamento alle fonti “ufficiali”.

Risposta

Chiediamo alle fonti autorevoli: Korpela (2006); Unicode Consortium (2009a,c,d,b); Wikipedia (2009b,a); Typo.cz e designiq.cz (2009).

Chi meglio di loro poteva aiutarmi a dirimere la questione su quale fosse la via migliore da percorrere?

Introduzione

Cosa dice Unicode

Unicode ci dice che le lettere con il segno a uncino o quello a virgola sono sempre “con cediglia”.

La forma della cediglia varia in base alla lettera: ç vs. ç̣.

Ț e Ș sono considerate varianti tipografiche per il romeno di Ț and Ș.

Il simbolo ‘,’ ha comunque il suo nome: *comma below* (o *comma accent*).

L^AT_EX e cediglia

Possiamo usare la cediglia al posto del comma below?

Abbiamo visto che con 'c', 'S', 's', 'T' e 't' non ci sono problemi.

Il lettone ha le lettere g, k, l, n, r che possono avere il comma below. Vediamo come vengono con la cediglia:

g, k, l, n, r...

Francamente è **inaccettabile**: non si può usare la cediglia con le lettere discendenti.

Comma below e casi speciali

Il segno comma below, con le lettere discendenti, muta in *turned comma above*.

Dobbiamo quindi scrivere \grave{g} : il segno è ruotato di 180° e posto *sopra* il carattere.

Il pacchetto `combelow` compone i caratteri appropriatamente, distinguendo i casi:

$$\backslash\text{cb}\{t\} \rightarrow \text{t}$$
$$\backslash\text{cb}\{g\} \rightarrow \grave{g}$$

Comma below e casi speciali

Le versioni vecchie del pacchetto erano molto inefficienti.
Le quattro versioni nuove (da 0.99a a 0.99d) sono migliori, quale per un aspetto, quale per un altro.
Non bisogna dimenticare che tutte le versioni possono ricevere caratteri multibyte in input; l'importante è racchiudere i caratteri tra parentesi graffe.

Peripezie di un codice

La prima versione, 0.99a, è stata ampiamente giudicata da Claudio Beccari, che mi ha spronato verso risultati (e un codice) migliori, e mi ha chiarito diversi punti oscuri.

Per puro scopo didattico Claudio ha scritto un codice molto compatto per realizzare il comando.

Studiando questo codice, e il codice del comando plain T_EX `\c #1`, ho realizzato le versioni successive a 0.99a.

combelow 0.99a

```

1 \NeedsTeXFormat{LaTeX2e}
2 \ProvidesPackage{combelow}
3 [2009/08/23 v0.99a comma below
  command]
4 \RequirePackage{graphicx}
5 \RequirePackage{ifthen}
6 \newdimen\fcxpos
7 \newdimen\fcypos
8 \newdimen\ftpos
9 \newcommand{\cb}[1]{%
10 \setbox1=\hbox{#1}%
11 \setbox2=\hbox{,}%
12 \fcxpos=.5\wd1%
13 \advance\fcxpos by .5\wd2%
14 \ftpos=.5\wd1%
15 \advance\ftpos by -.5\wd2%
16 \ifthenelse{\lengthtest{\dp1>
17 .1\ht1}}%
18 {\fcypos=\ht1%
19 \advance\fcypos by .6ex%
20 \raisebox{0pt}{#1\hspace
21 {-\fcxpos}\raisebox{\fcypos}%
22 {\rotatebox[origin=c]{180}{,}
23 \hspace{\ftpos}}}%
24 {\fcypos=\dp1%
25 \advance\fcypos by .4ex%
26 \raisebox{0pt}{#1\hspace
27 {-\fcxpos}\raisebox{-\fcypos},
28 \hspace{\ftpos}}}%
29 }
30 \endinput

```

1–6 : definizioni generali;

10, 11 : box con lettera da accentare e
“accento”;

12, 13 : calcolo posizione orizzontale
dell’accento; la sua retta
verticale di simmetria sarà
allineata a quella del carattere;

14, 15 : posizione del testo dopo il
carattere accentato;

16–28 : posizione verticale del segno
diacritico:

16, 17 : verifica profondità (carattere
discendente);

18–23 : posizione turned comma
above;

24–28 : posizione comma below.

combelow 0.99a: Risultati

Vediamo come viene posizionato il segno diacritico dal programma appena visto:

Sebbene il calcolo della posizione del carattere successivo a quello accentato sia basata sempre sui box usati, il risultato sembra essere buono:

București

combelow 0.99a: Pro e contro

L'unico vantaggio è l'indipendenza dalla codifica.

Molti gli svantaggi:

- 1 Viene incluso `graphicx` per ruotare una virgola e ottenere una virgoletta singola aperta;
- 2 Errato posizionamento dell'accento nei corsivi:

g' g'

- 3 Usare `\newcommand` non permette di mettere caratteri con comma below nelle note a piè di pagina;
- 4 Nel caso di `ř` l'accento non è posto sotto l'asta, ma centrato;
- 5 Si perdono le informazioni sulla crenatura e per la sillabazione;
- 6 Alcune versioni dei font mettono un comma below piú piccolo della virgola qui usata.

combelow 0.99b

```

1 \NeedsTeXFormat{LaTeX2e}
2 \ProvidesPackage{combelow}
3   [2009/08/23 v0.99b comma below
   command]
4 \RequirePackage{graphicx}
5 \RequirePackage{ifthen}
6 \newdimen\fcxpos
7 \newdimen\fcypos
8 \newdimen\ftpos
9 \DeclareRobustCommand{\cb}[1]{%
10 \setbox1=\hbox{#1}%
11 \setbox2=\hbox{,}%
12 \fcxpos=.5\wd1%
13 \advance\fcxpos by .5\wd2%
14 \ftpos=.5\wd1%
15 \advance\ftpos by -.5\wd2%
16 \ifthenelse{\lengthtest{\dp1>
17   .1\ht1}}%
18 {\fcypos=\ht1%
19 \advance\fcypos by -.7ex%
20 \raisebox{0pt}{#1\hspace
21   {-\fcxpos}\raisebox{\fcypos}
22   {\hspace{\ftpos}}}%
23 {\fcypos=\dp1%
24 \advance\fcypos by .4ex%
25 \raisebox{0pt}{#1\hspace
26   {-\fcxpos}\raisebox{-\fcypos}
27   ,\hspace{\ftpos}}}%
28 }
29 \endinput

```

La differenza sostanziale tra questo codice e il precedente sta nelle righe 19–22.

combelow 0.99b: Soluzioni

La versione b del programma risolve i seguenti errori:

- 1 Non si carica piú graphicx ma si usa la virgoletta singola aperta;
- 2 L'accento è ora corretto anche nei corsivi:

g'

- 3 `\DeclareRobustCommand` elimina il problema introdotto da `\newcommand`.

combelow 0.99c: un fantastico passo indietro

A partire dal codice di Claudio Beccari e da quello del comando `\c #1` ho prodotto due versioni. La prima, di seguito, con il comando `\accent`:

```
1 \NeedsTeXFormat{LaTeX2e}
2 \ProvidesPackage{combelow}
3   [2009/08/23 v0.99c comma below command]
4 \DeclareRobustCommand{\cb}[1]{%
5 \setbox\z@\hbox{#1}%
6 \ifdim\dp\z@=\z@{\accent44 #1}%
7 \else{\accent96 #1}%
8 \fi}%
9 \endinput
```

combelow 0.99c: Risultati

Il codice è molto compatto, ma questo è il risultato ottenuto:

ġ k l ṅ ṙ ṡ t

Il risultato pessimo dipende dal fatto che i caratteri usati non sono disegnati per essere accenti, e `\accent` non offre controllo sui posizionamenti verticali.

Oltre alla bruttezza, questa versione introduce la dipendenza dalla codifica.

combelow 0.99d

Questa è l'altra versione; il codice è meno compatto, ma funziona tanto bene quanto la versione 0.99b:

```
1 \NeedsTeXFormat{LaTeX2e}
2 \ProvidesPackage{combelow}
3   [2009/08/23 v0.99d comma below command]
4 \DeclareRobustCommand{\cb}[1]{%
5 \setbox\z@\hbox{#1}%
6 \ifdim\dp\z@=\z@
7   \oalign{\unhbox\z@\crcr
8   \hidewidth\lower.4ex
9   \hbox{,}\hidewidth}%
10 \else\oalign{\unhbox\z@
11   \crcr\hidewidth\raise.3ex
12   \hbox{'}\hidewidth}%
13 \fi}}%
14 \endinput
```

Sillabazione e crenatura

L'algoritmo di sillabazione di T_EX non è in grado di sillabare se trova un comando `\accent`.

Anche nel caso del comando qui presentato T_EX non è in grado di sillabare la parola che lo contiene.

Usare un font con codifica adeguata, cioè che contenga il carattere voluto, elimina questo problema.

Sillabazione e crenatura

Mutu: „Amenda îmi afectează viața privată“

„În afara terenului nu sunt deloc în apele mele, nici nu pot să fiu. Acea amendă este un lucru particular care îmi afectează viața privată... Totuși, când intru pe teren este ca o ușurare su-

fletească“, a precizat Mutu la postul de televiziune italian Sky. Tribunalul de Arbitraj Sportiv (TAS) a anunțat, la 31 iulie, că a respins memoriul formulat de Adrian Mutu la decizia Camerei

pentru Rezolvarea Disputelor (DRC) din cadrul FIFA, care a hotărât, la 7 mai 2008, că jucătorul trebuie să plătească clubului Chelsea despăgubiri în valoare de 17 milioane de euro. Că-

Rezolvarea Disputelor (DRC) din cadrul FIFA a decis, la 7 mai 2008, că Mutu, acum la AC Fiorentina, trebuie să plătească fostului său club, Chelsea, despăgubiri în valoare de 17.173.990 de

încălcarea contractului, iar internaționalul român a cerut TAS să anuleze hotărârea FIFA și să stabilească faptul că nu trebuie plătită nicio compensație.

Articolo in romeno, caratteri con cediglia, codifica OT1

Sillabazione e crenatura

Mutu: „Amenda îmi afectează viața privată“

„În afara terenului nu sunt deloc în apele mele, nici nu pot să fiu. Acea amendă este un lucru particular care îmi afectează viața privată... Totuși, când intru pe teren este ca o ușurare su-

fletească“, a precizat Mutu la postul de televiziune italian Sky.

Tribunalul de Arbitraj Sportiv (TAS) a anunțat, la 31 iulie, că a respins memoriul formulat de Adrian Mutu la decizia Camerei

pentru Rezolvarea Disputelor (DRC) din cadrul FIFA, care a hotărât, la 7 mai 2008, că jucătorul trebuie să plătească clubului Chelsea despăgubiri în valoare de 17 milioane de euro. Că-

Rezolvarea Disputelor (DRC) din cadrul FIFA a decis, la 7 mai 2008, că Mutu, acum la AC Fiorentina, trebuie să plătească fostului său club, Chelsea, despăgubiri în valoare de 17.173.990 de euro, **pentru**

încălcarea contractului, iar internaționalul român a cerut TAS să anuleze hotărârea FIFA și să stabilească faptul că nu trebuie plătită nicio compensație.

Ecco i problemi che questa codifica comporta.

Sillabazione e crenatura

Mutu: „Amenda îmi afectează viața privată“

„În afara terenului nu sunt deloc în apele mele, nici nu pot să fiu. Acea amendă este un lucru particular care îmi afectează viața privată... Totuși, când intru pe teren este ca o ușurare su-

fletească“, a precizat Mutu la postul de televiziune italian Sky. Tribunalul de Arbitraj Sportiv (TAS) a anunțat, la 31 iulie, că a respins memoriul formulat de Adrian Mutu la decizia Camerei

pentru Rezolvarea Disputelor (DRC) din cadrul FIFA, care a hotărât, la 7 mai 2008, că jucătorul trebuie să plătească clubului Chelsea despăgubiri în valoare de 17 milioane de euro. Camera

Rezolvarea Disputelor (DRC) din cadrul FIFA a decis, la 7 mai 2008, că Mutu, acum la AC Fiorentina, trebuie să plătească fostului său club, Chelsea, despăgubiri în valoare de 17.173.990 de

tru încălcarea contractului, iar internaționalul român a cerut TAS să anuleze hotărârea FIFA și să stabilească faptul că nu trebuie plătită nicio compensație.

Articolo in romeno, caratteri con cediglia, codifica T1

Sillabazione e crenatura

Mutu: „Amenda îmi afectează viața privată“

„În afara fletească“, a pentru Rezol- Rezolvarea
terenului nu precizat Mutu varea Dispu- Disputelor tru încălcarea
sunt deloc în la postul de telor (DRC) Disputelor contractului,
apele mele, televiziune din cadrul FIFA (DRC) din iar internațio-
nici nu pot italian Sky. FIFA, care cadrul FIFA nalul român
să fiu. Acea Tribunalul a hotărât, la mai 2008, că să anuleze ho-
amendă este de Arbitraj Sportiv (TAS) a anunțat, la 7 mai 2008, că tărărea FIFA
un lucru par- traș Sportiv (TAS) a anunțat, la 31 iulie, că a și să stabi-
ticular care nșat, la 31 iulie, că a lească faptul
îmi afectează viața pri- respins me- moriul formu- lat de Adrian Mutu la de- cizia Camerei
vată... To- tuși, când intru pe te- ren este ca o Mușurare su- cizia Camerei
câtă, care a respins memoriul formulat de Adrian Mutu la decizia Camerei

pentru Rezolvarea Disputelor (DRC) din cadrul FIFA, care a hotărât, la 7 mai 2008, că jucătorul trebuie să plătească clubului Chelsea despăgubiri în valoare de 17 milioane de euro. Camera pentru

Rezolvarea Disputelor (DRC) din cadrul FIFA a decis, la 7 mai 2008, că Mutu, acum la AC Fiorentina, trebuie să plătească fostului său club, Chelsea, despăgubiri în valoare de 17.173.990 euro, pen-

tru încălcarea contractului, iar internaționalul român a cerut TAS să anuleze hotărârea FIFA și să stabilească faptul că nu trebuie plătită nicio compensație.

Ecco come risolve i problemi questa codifica.

Sillabazione e crenatura

Mutu: „Amenda îmi afectează viața privată“

„În afara fletească“, a pentru Rezol-
terenului nu precizat Mutu varea Rezol-
sunt deloc în la postului de varea Dispu-
apele mele, televiziune din (DRC) Disputelor
nici nu pot italian Sky. FIFA, care (DRC) din
să fiu. Acea Tribunalul a hotărât, la cadrul FIFA internaționalul
amendă este de Arbi- 7 mai 2008, că român a ce-
un lucru par- traj Sportiv (TAS) a trebuie să mut TAS să
ticular care anunțat, la plătească clu- la AC Fioren- anulaze ho-
îmi afec- 31 iulie, că a bului Chelsea să plătească și să stabi-
tează viața 31 iulie, că a bului Chelsea fostului său plătească lească faptul
privată... Totuși, când moriul formu- in valoare de despăgubiri plătită nicio
intru pe te- lat de Adrian 17 milioane în valoare de despăgubiri compensație.
ren este ca o Mutu la de- 17 milioane în valoare de despăgubiri compensație.
ușurare su- cizia Camerei mera pentru euro, pen-

Articolo in romeno, caratteri con comma below, codifica T1

Sillabazione e crenatura

Mutu: „Amenda îmi afectează viața privată“

„În afara fletească“, a pentru Rezol- Rezolvarea tru încăl-
terenului nu precizat Mutu varea Dispu- Disputelor carea con-
sunt deloc în la postul de telor (DRC) (DRC) din tractului, iar
apele mele, televiziune din cadrul FIFA internaționalul
nici nu pot italian Sky. FIFA, care a decis, la 7 român a ce-
să fiu. Acea Tribunalul a hotărât, la mai 2008, că rut TAS să
amendă este de Arbi- 7 mai 2008, Mutu, acum anuleze ho-
un lucru par- traj Spor- că jucătorul la AC Fioren- tărârea FIFA
ticular care tiv (TAS) a trebuie să plătească și să stabi-
îmi afec- anunțat, la plătească clu- să plătească lească faptul
tează viața 31 iulie, că a bului Chelsea despăgubiri club, Chelsea, plătită nicio
privată... Totuși, când moriul formu- lat de Adrian Mutu la de- cizia Camerei în valoare de în valoare de
intru pe te- Mutu la de- cizia Camerei în valoare de 17 milioane de euro. Ca- 17.173.990 de
ren este ca o Mutu la de- cizia Camerei în valoare de 17 milioane de euro. Ca- 17.173.990 de
ușurare su- cizia Camerei în valoare de 17 milioane de euro. Ca- 17.173.990 de euro, pen-

Tornano quasi tutti i problemi, nonostante la codifica.

Sillabazione e crenatura

Il problema si può risolvere in almeno due modi:

- utilizzando i font virtuali, cioè quei font “costruiti” associando quei comandi L^AT_EX che descrivano esattamente la costruzione dei caratteri;
- usando programmi di composizione *Unicode-compliant*, ad esempio LuaT_EX.

Quanto detto vale anche per le crenature, disattivate quando usiamo il comando di `combelow`.

Carattere vietato in Unicode

Vista l'assenza di un comando nativo per il comma below, e prima di decidere di scrivere il pacchetto, ho provato a inserire '‡' in un documento L^AT_EX da un editor Unicode.

La compilazione si interruppe proprio nel punto in cui L^AT_EX incontrò quel carattere.

Perché?

L^AT_EX "supporta" Unicode solo perché può leggere caratteri multibyte mettendoli in corrispondenza biunivoca ai comandi L^AT_EX.

Non c'è/era ancora un comando L^AT_EX per comporre lettere con comma below.

Conclusioni, sviluppi futuri e ringraziamenti

Il pacchetto dovrebbe finire per comprendere (magari cambiando nome in diacrit) tutti quei comandi per comporre i segni diacritici non compresi in T_EX/L^AT_EX.

Anche babel potrebbe fare in modo di usare combelow per comporre i termini noti (capitolo, bibliografia...) in base alla lingua selezionata.

Nell'immediato futuro il pacchetto dovrebbe arrivare a usare i comandi L^AT_EX per comporre le lettere accentate; bisognerebbe anche decentrare il segno diacritico in base alla forma della lettera. Grazie a Claudio Beccari: senza il suo acume questo lavoro sarebbe stato ben povero. Grazie anche a tutti quelli che, avendolo letto, mi hanno suggerito miglioramenti e correzioni.

Riferimenti bibliografici

Korpela, J. K. (2006). *Unicode explained*. O'Reilly, Sebastopol, CA.

Typo.cz e designiq.cz (2009). «Diacritics project». URL

<http://diacritics.typo.cz/>.

Unicode Consortium (2009a). «Code charts». URL

<http://www.unicode.org/charts>.

— (2009b). «Combining Diacritical Marks». URL

<http://www.unicode.org/charts/PDF/U0300.pdf>.

— (2009c). «Latin Extended A». URL

<http://www.unicode.org/charts/PDF/U0100.pdf>.

— (2009d). «Latin Extended B». URL

<http://www.unicode.org/charts/PDF/U0180.pdf>.

Wikipedia (2009a). «Cediglia». URL

<http://it.wikipedia.org/wiki/Cediglia>.

— (2009b). «Cedilla». URL

<http://en.wikipedia.org/wiki/Cedilla>.

Grazie a tutti per l'attenzione.
Potete pormi i quesiti che qui non avete potuto a
g.pignalberi@freesoftwaremagazine.com