



Module Nooj pour le rromani - Contribution d'un outil informatique à R.E.D.-RRROM

Masako Watabe (INALCO)

17-18/06/2014

Les Rroms et la langue rromani

Module Nooj pour le rromani - Contribution d'un outil informatique à R.E.D.-RRROM

- Introduction : Nooj
- Module Nooj pour le rromani
 - Exemple : Lemme « çhavo »
 - Paradigmes emboîtés
 - Paradigmes en graphe
 - Concordance
 - Annotation
 - Frozen expression
- Conclusion : Utilité de Nooj pour R.E.D.-RRROM





Introduction : Nooj

- Nooj est un logiciel pour la linguistique.
- C'est un environnement pour formaliser des phénomènes linguistiques.
- Chaque langue a besoin de son module correspondant à son système linguistique.
- Le module se compose de dictionnaires et de grammaires (morphologie et syntaxe).
- Nooj permet d'annoter et d'analyser des textes d'une large couverture.



Module Nooj pour le rromani

- Le module initial remplit deux fonctions : lexicale et morphologique (flexionnelle et dérivationnelle).
- Le dictionnaire contient très peu d'entrées, pourtant chacune représente un paradigme flexionnel.
- Les paradigmes principaux de noms, de verbes et d'adjectifs ont été programmés, y compris des exceptions et des variantes dialectales.



Module Nooj pour le rromani

Exemple : Lemme « čhavo »

■ Dictionnaire

čhavo, N+hum+m+EN="Rromani boy"+FLX=čhavo+DRV=čhavorro

čhavo, N+hum+m+EN="son"+FLX=čhavo+DRV=čhavorro

■ Morphologie

1) FLX : Paradigme flexionnel « čhavo »

čhavo = <E>/sg+dr | e/pl+dr |

es/sg+ob | en/pl+ob |

e!a/sg+voc | a!len/pl+voc ;

<E>=empty, =backspace

2) DRV : Paradigme dérivationnel « čhavorro »

čhavorro = <A>rro/N+diminutive;

=> Cela fait 12 formes fléchies :

čhavo, čhave, čhaves, čhaven, čhave!a, čhava!len

čhavorro, čhavorre, čhavorres, čhavorren, čhavorre!a, čhavorra!len.



Module Nooj pour le rromani

- Paradigme flexionnel complet « çhavo »

çhavo = <E>/sg+dr | e/pl+dr |

es(<E> |:pstp2) <E>/sg+ob |

en(<E> |:pstp1) <E>/pl+ob |

e!a/sg+voc | a!len/pl+voc ;

«:pstp1», «:pstp2» = paradigmes de postpositions emboîtés

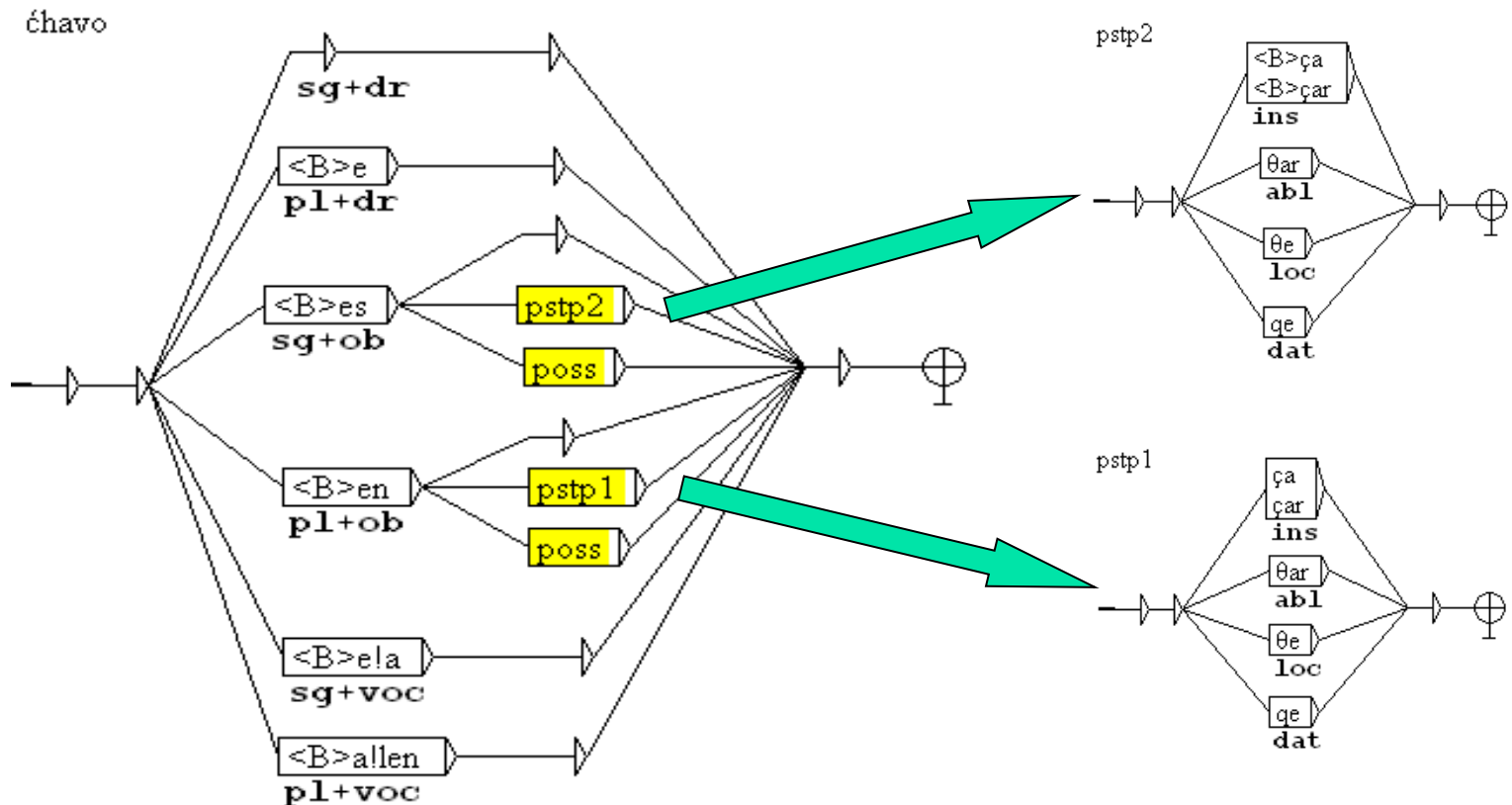
=> Cela fait 472 formes fléchies !

(Les formes obliques postposées sont considérées comme formes fléchies dans le module rromani).



Module Nooj pour le rromani

Paradigme « çhavo » en graphe-1

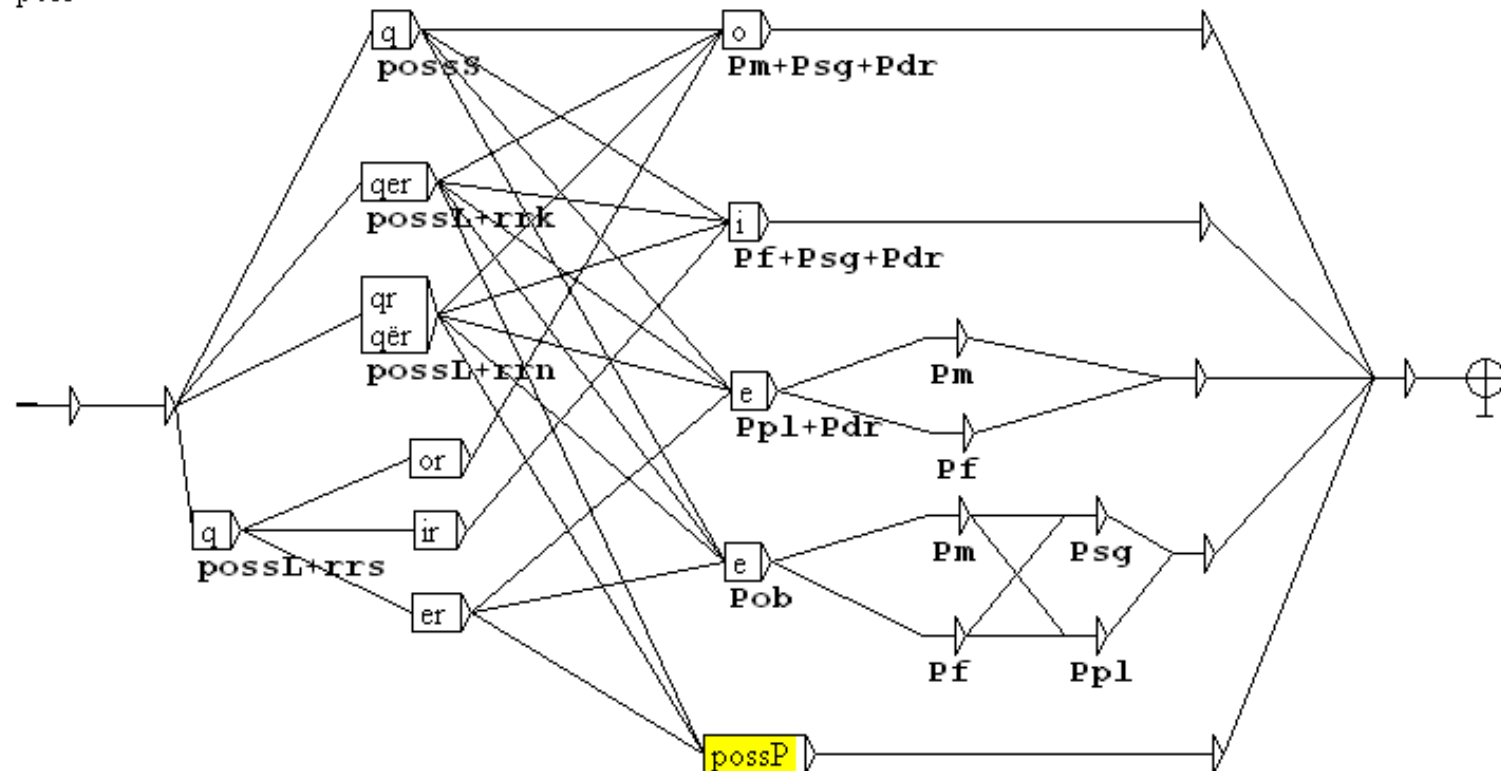




Module Nooj pour le rromani

Paradigme « çhavo » en graphe-2

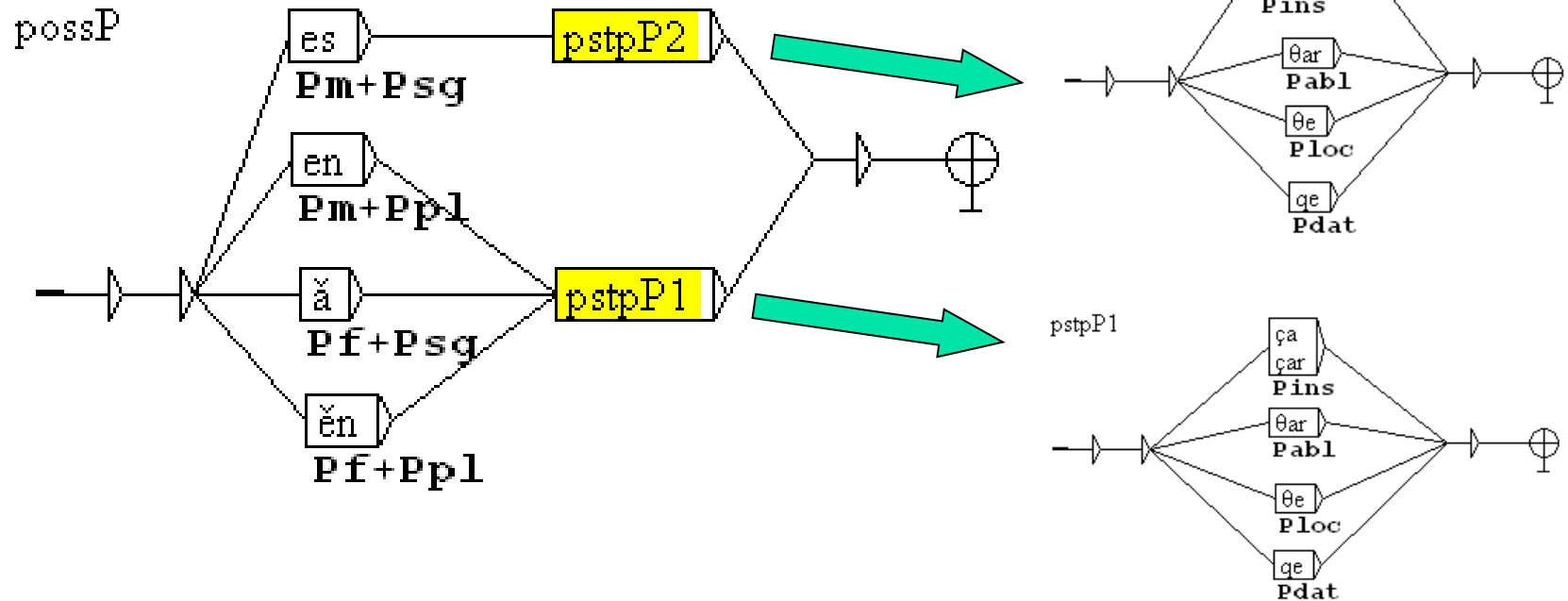
poss





Module Nooj pour le rromani

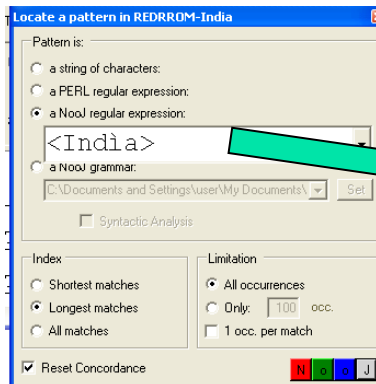
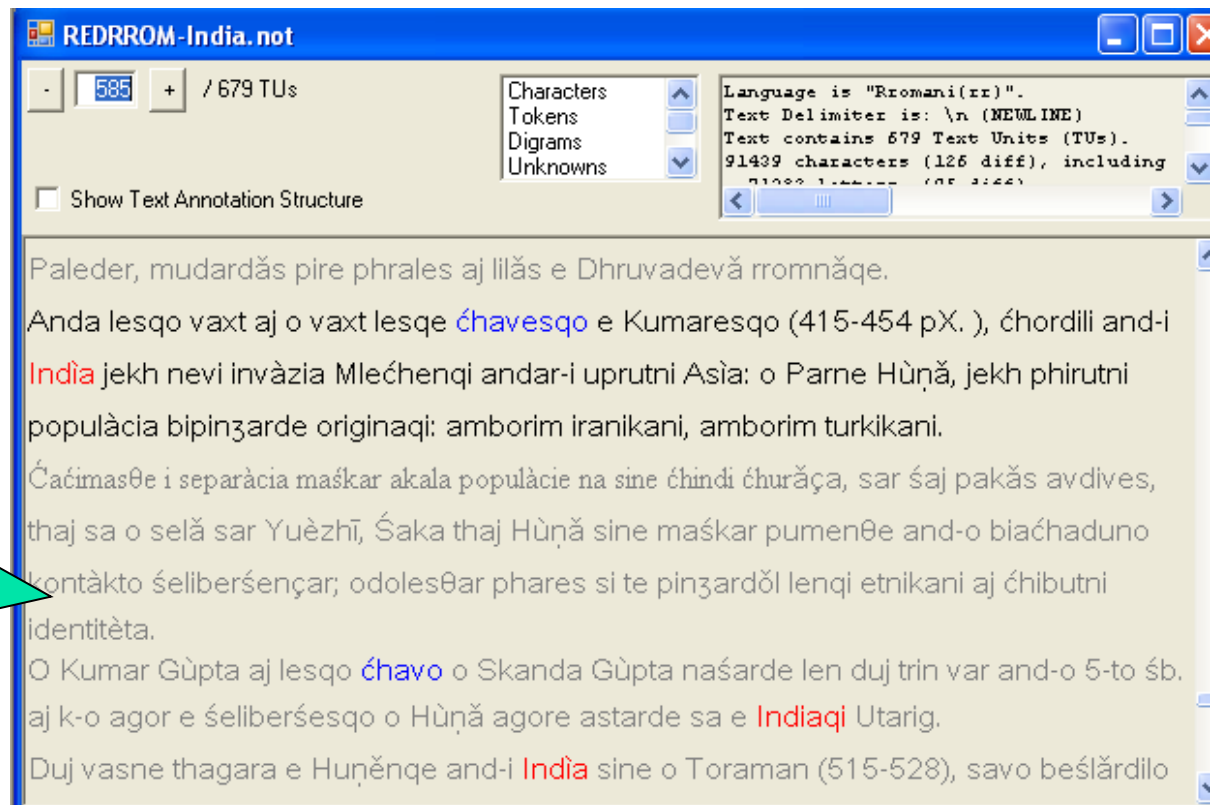
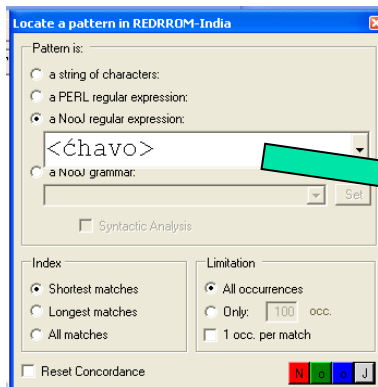
Paradigme « çhavo » en graphe-3





Module Nooj pour le rromani

Concordance





Module Nooj pour le rromani

Concordance

Concordance for Text REDRRROM-India.not [Modified]

Reset Display: 5 characters before, and 5 after. Display: Matches Outputs
 word forms

Text	Before	Seq.	After
Make donia). Karing-i Mesmerig, lesqo	čhavo		o Mahendra andäs o buddhismo
šelibers, pinzardile sar razpütä - 'rajenqe	čhave		' Akaja integrácia, sar phenel pes
vaxt aj o vaxt lesqe	čhavesqo		e Kumaresqo (415-454 pX), čhordili and
O Kumar Güpta aj lesqo	čhavo		o Skanda Güpta našarde len
-o b. 604 pX. lesqe duj	čhave		o baro o 18-beršenqo Rājja
o Prabhakar bičhalel pe duje	čhaven		: anglal and-i avangārda e
jakha, o Prabhakar phenel e	čhavesqe		e Haršasqe: 'te reses baxtagor
kurganēnqi zi k-i Utarigutni	India		Buteder desar bišupan3-trianda šelibersä
nácie, šeral and-i Evröpa,	India		aj Iräno. Si pinzarde sar
sine lil, papiro, and-i	India		thaj o manuša lekhaven sas
anglune Rajane avile and-i	India		pumare familiençar aj idovēnçar (gurumnä
širdel but paleder and-i	India		, and-o 4-to šb. pX
dikhas i čirlatuni civilizácia e	Indiaqi		sar vi rajanikani aj vi
so te aresen and-i	India		, o Rajane sine ulavde anda
čandalenqe; čacimasqe and-i purani	India		, vi o šudrāje vi o
gatēnqe. Sarkon raštra and-i	India		, si la jekh aver sistēma

Query 123/123



Module Nooj pour le rromani

Annotation

REDRRROM-India.not

585 / 679 TUs

Characters
Tokens
Digrams
Unknowns

Language is "Rromani(rr)".
Text Delimiter is: \n (NEWLINE)
Text contains 679 Text Units (TUs).
91439 characters (126 diff), including
71000 1-grams (95 diff)

Show Text Annotation Structure

Anda lesqo vaxt aj o vaxt lesqe **čhavesqo** e Kumaresqo (415-454 pX.), čhordili and-i **Indiã**
jekh nevi invàzia Mlečhenqi andar-i uprutni Asiã: o Parme Hùñã, jekh phirutni populàcia

32

čhavo,N+hum+Gd=m+EN="Rromani boy"+PGd=Pm+PNb=Psg+PCs=Pdr+Pstp=possS+Nb=sg+Cs=ob
čhavo,N+hum+Gd=m+EN="son"+PGd=Pm+PNb=Psg+PCs=Pdr+Pstp=possS+Nb=sg+Cs=ob

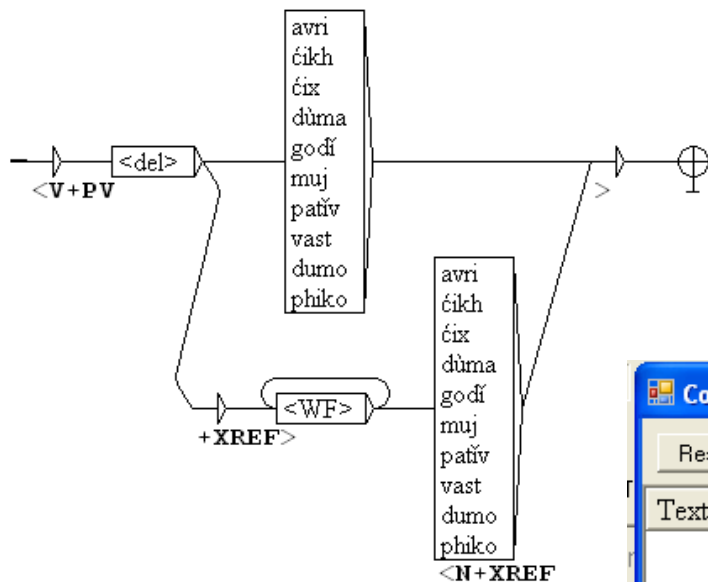
kamipe aj o bianipe.
O murš aj i zuvli **den pumen vast** jekh avres kaj te aresen o kãma, thaj odolesθar o Vaciajàna lekhlãs i butsundi pustik Kãmasutra, kaj siklõn o

18	22	28
V+PV	pumen,PRON+ref+Dict=rro+EN="themselves"+RRE="pen"	N
del,V+Tr=tr+EN="to give"+Pers=2+Nb=pl+Tense=PR		vast,N+ina+Gd=m-
del,V+Tr=tr+EN="to give"+Pers=3+Nb=pl+Tense=PR		



Module Nooj pour le rromani

Frozen expression



Concordance for Text REDRRROM-India.not

Reset Display: 5 characters before, and 5 after. Display: Matches Outputs

Text	Before	Seq.	After
	zuvlăqo, si but vazdime aj	dini patív	and-i epopèa Ramajan. Anda
	bianel lesqe çhaven aj te	del les vast	kaj te pherdârel (te realisarel
	O muré aj i guvli	den pumen vast	jekh avres kaj te aresen
7	sar manuś, thaj prandime - athòska	del phiko	e rromes. Sar nakhel o
8	baxt, pesqo karmò, bi te	del les vast	ni manuś ni devel. Maşkar
9	Mahavireça aj vi e Gautameça,	dinàs len love aj phiko	, aj vi kòrkorro lilàs lenqi
10	sa e pakăimatenqe, kherutne save	den patív	e dades aj e dia

Query 7/7

Conclusion : Utilité de Nooj pour R.E.D.-RRROM



- La morphologie Nooj sera importée dans le dictionnaire en ligne R.E.D.-RRROM, et le lexique de R.E.D.-RRROM sera importé dans Nooj pour se compléter.
- Une fois que le module Nooj est intégré dans le système R.E.D.-RRROM (i.g. Plateforme Smallcodes), on peut développer divers outils informatiques tels que correcteur d'orthographe, comparateur de dialectes etc.
- Certaines fonctions du module seraient utiles non seulement pour les apprenants mais aussi pour les auteurs et les techniciens de R.E.D.-RRROM pour qu'ils travaillent sur l'arrière-boutique du site web.

Module Nooj pour le rromani - Contribution d'un outil informatique à R.E.D.-RRROM

Références

- Marcel Courthiade. 2009. *Morri angluni rromane čhibăqi evroputni lavustik*. Cigany Haz, Budapest
- "R.E.D.-RRROM (Restoring the European Dimension of Rromani Language and Culture)"
<http://red-rrom.eu>
- Max Silberztein. 2003-. *NooJ Manual*.
<http://www.nooj4nlp.net/>





Module Nooj pour le rromani - Contribution d'un outil informatique à R.E.D.-RRROM

Te oven saste veste!

Masako Watabe (INALCO) masakowatabe@free.fr

