



Quina metodologia per despolhar l'enquèsta Bourciez ?

Alexandre Genadot

► To cite this version:

Alexandre Genadot. Quina metodologia per despolhar l'enquèsta Bourciez ?. Ièr, deman : diga-m'o dins la lenga, Nov 2021, Montpellier, France. hal-03471470

HAL Id: hal-03471470

<https://inria.hal.science/hal-03471470>

Submitted on 9 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Quanta metodologia per despolhar l'enquèsta Bourciez ?

Alexandre Génadot
pel grop La Dialectometria

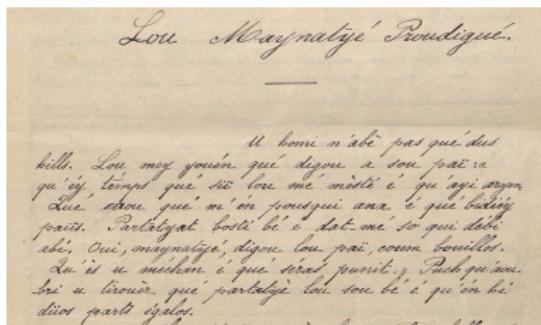
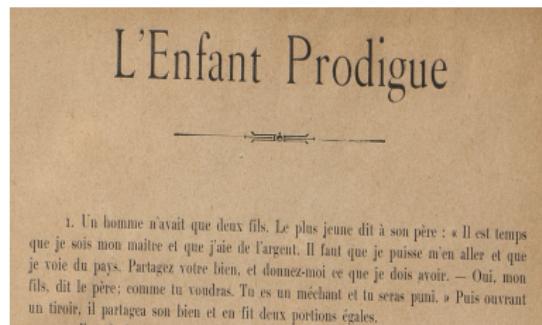
Institut de Mathématiques de Bordeaux, equipa Optim'AI
Inria Bordeaux Sud-Ouest, equipa Astral

26/11/2021

L'Enquèsta Bourciez : 1894-1895

Enquèsta concebuda per Édouard Bourciez (professor de l'universitat de Bordèu) e menada per correspondéncia mercés als regents dels rectorats de Bordèu e Tolosa.

→ revirar, dinc lo parlar de la comuna del chabiscòl, una version de la parabòla del filh prodigue (se faire ajudar se n'es besonh).



Mai de 4400 responsas !

Enquèsta jamai despolhada d'un biais sistematic levat per la partida en lenga Basca mercés a C. Videgain e G. Aurrekoetxea.
Es mens de 4 del cent del corpus → a nosautres lo demai !

Metodologia

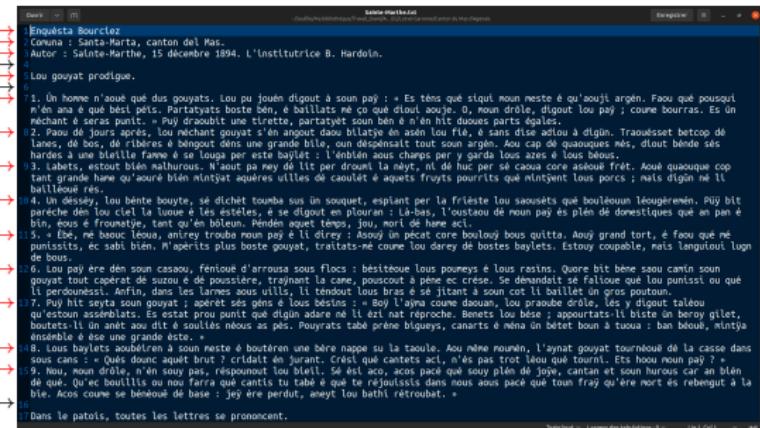
Dinc lo meteis temps :

- ▶ Encodatge informatic de las parabòlas ;
- ▶ Creacion de la basa de donadas dels mots (tablèu...) ;
- ▶ Analisa estatistica descriptiva (mapas, frequéncias, ajustaments...) ;
- ▶ Creacion d'un site web (per ajudar lo travalh e saique lo presentar).

Encodatge informatic de las parabòlas

- ▶ OCR (reconeissença optica de caractèrs) ? Intelligéncia artificiala ?
→ a mon vejaire, tot comptat e rebatut : pas encara pro bòn per nos faire ganhar de temps emb aquel corpus.
- ▶ alara... picar las parabòlas a la man !

Al format .txt emb una estructura normalizada (mai de 1200 parabòlas copiadas).



Lo nòstre grop What's App
La Dialectometria



Picaires : David Escarpit e A. G., mas tanben Julien Buziol, Gérard Judet de la Comba e Javier Martínez.

Creacion de la basa de donadas

Lo descopatge, automatizat, exemple del paragrap T1 :

1. U òmi n'abè pas qué dus hilhs. Lou méy youén qué digou aü sou pay : « Quéy téms que sioy lou mé mèsté et qu'àyi aryén . Qué cäü qué m'en pousqui ana é qué béyi drin dé pëys . Partat-yàt lou bòsté bée et dàmmé ço qui déby abé. - Oui , lou mé hilh, sé digou lou sou pay, coum bouilhos. Quès u méchan é qué séras punit . Puch, oubrin u tirouèr, qué partat-yà lou sou bée é qué hé duos pars égalos. ■

Embe :

- ▶ Los paragrafs ;
- ▶ Chausida de mots claus , reconeisseguts per expressions regularas après transliteracion del tèxt :

```
grep("^\p{ae}[\iy]",T1)
```


→ ...païe, païré, paey, payré, päy, paÿ, pay, paï
- ▶ descopatge dels paragrafs (4, 5 o 6 tròces per paragrap).

Tot aquò es fach (e çò que ven tanben) embe l'ajuda del logicial R.

Creacion de la basa de donadas

Lo tròç T11 del paragraf 1 :

U òmi n' abè pas qué dus hilhs . Lou méy youén qué digou
aü sou pay

Reconeisséncia automatica dels mòts per expressions regularas e creacion d'un tablèu .csv :

A	B	C	D	E	F	G	H	I	J	K
1	Departament	Canton	comuna	x	y	T11				
2	Gironde	Canton d'Audenge	Andrèdes les Bains	-1.0925184	44.7476897	Un homme n'asøt que dus hilhs lou mey jène dichot à son pøl	un	art. ind.	homme	subst. m.
3	Gironde	Canton d'Audenge	Aïts	-1.1379928	44.767644	Un homme n'asøt que dus hilhs lou mey jène dichot à son pøl	un	art. ind.	homme	subst. m.
4	Gironde	Canton d'Audenge	Angles	-1.1379928	44.767644	Un homme n'asøt que dus hilhs lou mey jène dichot à son pøl	un	art. ind.	homme	subst. m.
5	Gironde	Canton d'Audenge	Audenge	-1.0196969	44.6936848	Un homme n'asøt que dus hilhs lou mey jène dichot à son pøl	un	art. ind.	homme	subst. m.
6	Gironde	Canton d'Audenge	Bagnos	-0.9996271	44.6409008	Un homme n'asøt pas que dus hilhs lou mey jøuvøen dichot à son pay	un	art. ind.	homme	subst. m.
7	Gironde	Canton d'Audenge	Lacanau-de-Mios	-0.859044	44.6555763	Un homme n'asøt pas que dus hilhs lou mey jøuvøen dichot à son pay	un	art. ind.	homme	subst. m.
8	Gironde	Canton d'Audenge	Lamont	-1.0322014	44.7161218	Un homme n'asøt pas que dus hilhs lou mey jøuvøen dichot à son pay	un	art. ind.	homme	subst. m.
9	Gironde	Canton d'Audenge	Lège	-1.1472098	44.7961052	Un homme n'asøt que dus hilhs lou pe jene dichot à son pøl	un	art. ind.	homme	subst. m.
10	Gironde	Canton d'Audenge	Mios	-0.9349464	44.6045208	Un omme n'asøt que dus hilhs lou mey jøuvøen dichot à son pøl	un	art. ind.	omme	subst. m.

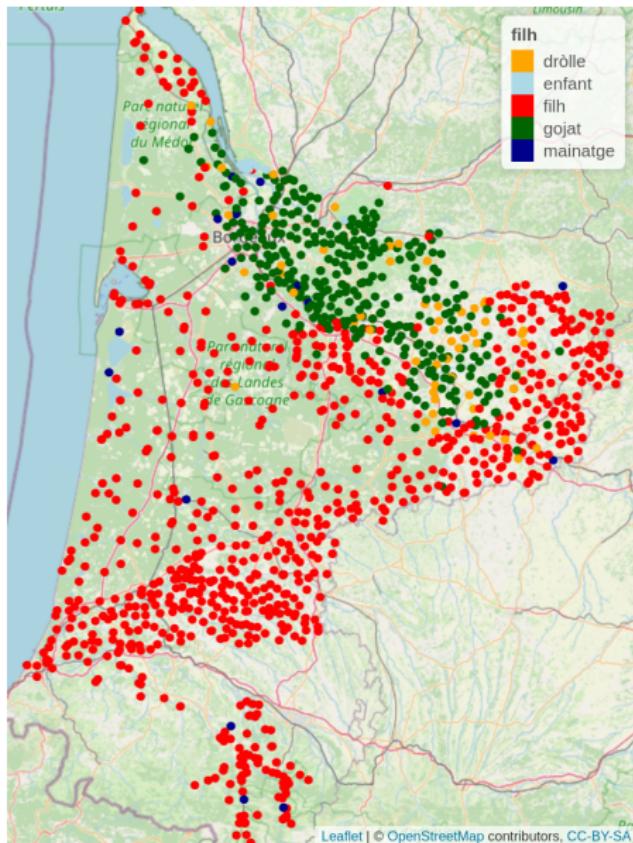
Gaireben 200 colònias per paragraf, tant de linhas coma de comunas.

Remarca : l'emplenament del tablèu se pòt tanben faire embe l'ajuda d'una distància lexicografica...

Dificultats :

- ▶ l'imaginacion descabestrada de quauques autors per escriure certans mots (mai de 200 biais d'escriure *minjar* per exemple) ; Sovent deguda a la dificultat dels autors de transcriure un son qu'existís pas en francés.
- ▶ los tròces de manca ;
- ▶ las traduccions embe reformulacions excessivas.

Showcase : Vocabulari

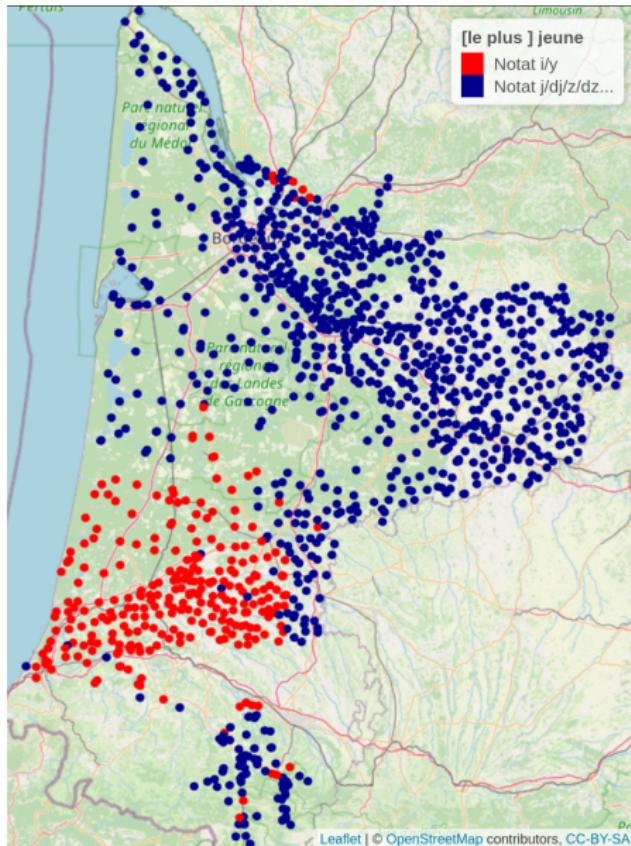


Revirada de *filh* pel tròc T11
: *Un homme n'avait que deux fils.*

- ▶ **enfant** : del latin *infans-antis*, que non parla encara ;
- ▶ **filh** : del latin *filius* ;
- ▶ **gojat** : de l'ebrèu *goy* (?), servent ;
- ▶ **mainatge** : per *mens atge* ;
- ▶ **drôle** : del neerlandés *drolle*, farfadet.

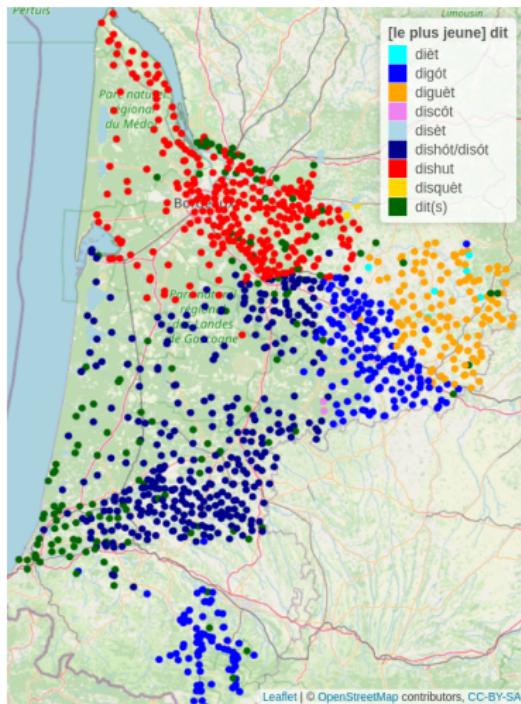
Remarca : libraria *leaflet* de R.

Showcase : Fonetica, Fonologia



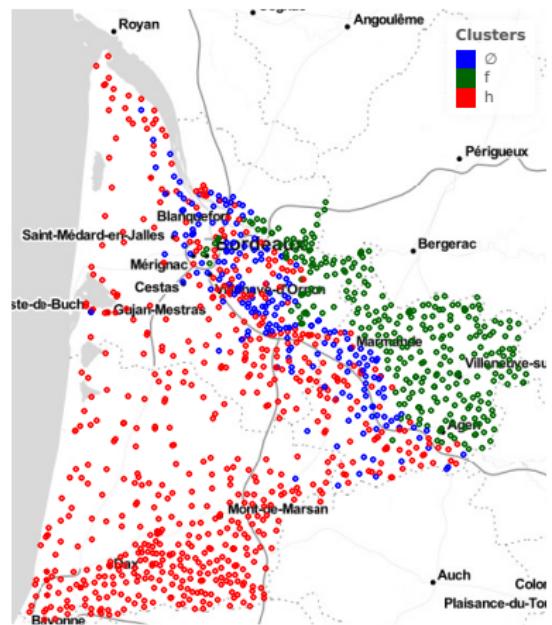
Un exemplo de pronunciación de la iòd [j] (per inferencia de la notacion).

Showcase : Morfología

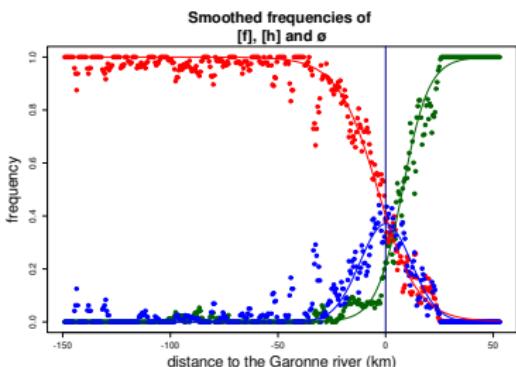


Morfología del preterit de
díser/dire.

Showcase : Clustering, Ajustament



Leaflet | © OpenStreetMap contributors, CC-BY-SA, Map tiles by Stamen Design, CC BY 3.0 — Map data © OpenStreetMap contributors



Showcase : Lo site

-/Seafile/Ma bibliothèque/Travail_David/Site_Web_Bourciez/_site/index.html

index.html | Open in Browser | Find | Publish | C

Enquèsta Bourciez

Ne saupre mai sobre Edouard Bourciez e l'enquèsta via lo site [occitanica](#).

Chada ponch blau de la mapa es una comuna de l'enquèsta. Se pòt clicar dessobre per legir la parabòla de la comuna, se pòt tanben zoomar.

Bilbao
Euskadi
Pamplona/Iruña
Donostia/San Sebastián
Bordeaux
Nouvelle-Aquitaine
Toulouse
Occitanie

Vos grandmerceji per vòstra escota !