



HAL
open science

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'ò dins la lenga, Nov 2021, Montpellier, France. hal-03471470

HAL Id: hal-03471470

<https://hal.inria.fr/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

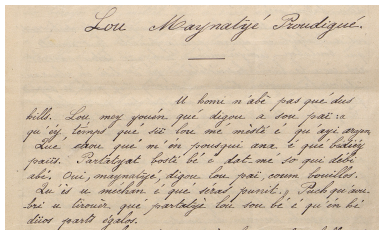
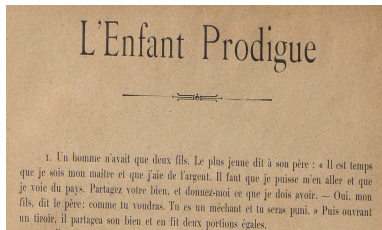
Institut de Mathématiques de Bordeaux, equipa Optim'AI
Inria Bordeaux Sud-Ouest, equipa Astral

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 :

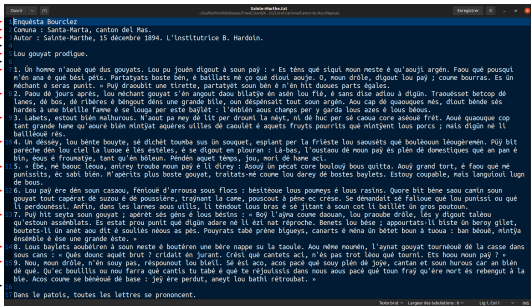
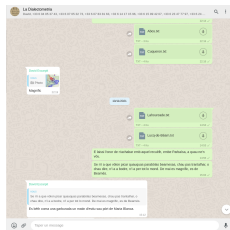
- ▶ Encodatge informatic de las parabòlas ;
- ▶ Creacion de la basa de donadas dels mots (tablèu...) ;
- ▶ Analisa estatistica descriptiva (mapas, frequéncias, ajustaments...) ;
- ▶ Creacion d'un site web (per ajudar lo trabalh e saique lo presentar).

Encodatge informatic de las parabòlas

- ▶ OCR (reconnaissance optique de caractères) ? Intelligéncia artificiaa ?
→ a mon vejaire, tot comptat e rebutat : 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 paragraf 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 càü qué m'en pousqui ana é qué béyi drin dé pèys . 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 paragraf).

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

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

	A	B	C	D	E	F	G	H	I	J	K		
1	Donatari	Larzon	carrene	x	y	T11							
2	Gasade	Carton d'Auxerge	Ardenes-les-Bains	-1.0525184	44.7476697	un homme n'asou que dus hilhs lou méy jéne dichut à son paï	un	art.	ind.	homme	subst.	m.	NA
3	Gasade	Carton d'Auxerge	Mus	-1.1379628	44.767664	un homme n'asou que dus hilhs lou méy jéne dichut à son paï	un	art.	ind.	homme	subst.	m.	NA
4	Gasade	Carton d'Auxerge	Auxerge-bis	-1.0234926269303	44.64603028110	un homme n'asou que dus hilhs lou méy jéne dichut paï	un	art.	ind.	homme	subst.	m.	NA
5	Gasade	Carton d'Auxerge	Auxerge	-1.0198268	44.6863648	un homme n'asou que dus hilhs lou méy jéne dichut à son paï	un	art.	ind.	homme	subst.	m.	NA
6	Gasade	Carton d'Auxerge	Bagnac	-0.9696371	44.6449008	un homme n'asou pas que dus hilhs lou méy jéne dichut à son paï	un	art.	ind.	homme	subst.	m.	NA
7	Gasade	Carton d'Auxerge	Lacroux-de-Mais	-0.950644	44.6555761	un homme n'asou pas que dus hilhs lou méy jéne dichut à son paï	un	art.	ind.	homme	subst.	m.	NA
8	Gasade	Carton d'Auxerge	Larzon	-1.0363114	44.702574	un homme n'asou que dus hilhs lou méy jéne dichut à son paï	un	art.	ind.	homme	subst.	m.	NA
9	Gasade	Carton d'Auxerge	Léze	-1.1472948	44.7651652	un homme n'asou que dus hilhs lou méy jéne dichut à son paï	un	art.	ind.	homme	subst.	m.	NA
10	Gasade	Carton d'Auxerge	Mus	-0.9349464	44.6545208	un homme n'asou paï que dus hilhs lou méy jéne dichut à son paï	un	art.	ind.	omme	subst.	m.	NA

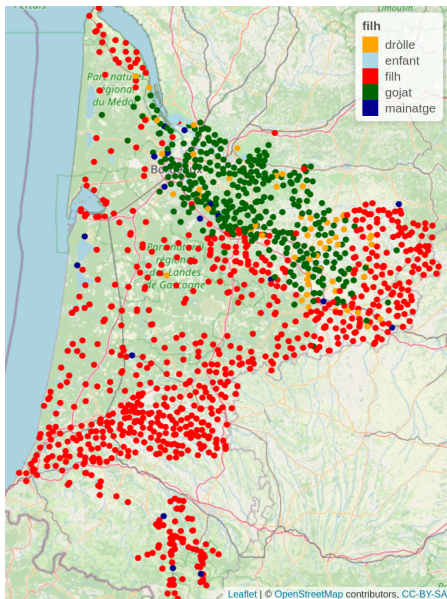
Gaireben 200 colònas per paragraf, tant de linhas coma de comunas.

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

Dificultats :

- ▶ l'imagination descabestrada de quauques autors per escriure certains 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 tròcs de manca ;
- ▶ las traduccions embe reformulacions excessivas.

Showcase : Vocabulari

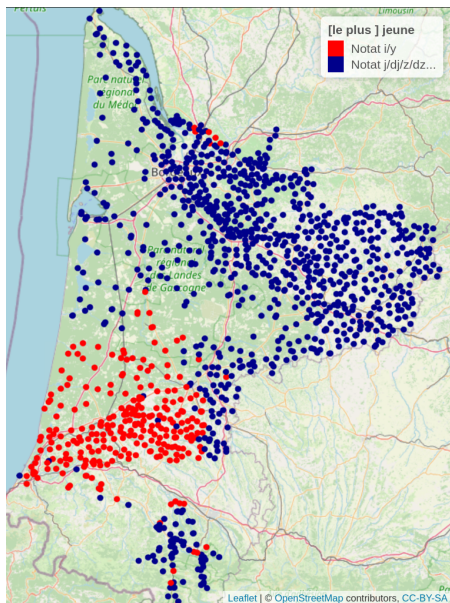


Revirada de *filh* pel tròç T11 : *Un homme n'avait que deux fils.*

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

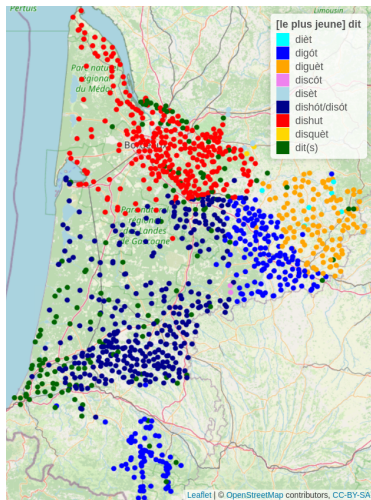
Remarca : libreria *leaflet* de R.

Showcase : Fonetica, Fonologia



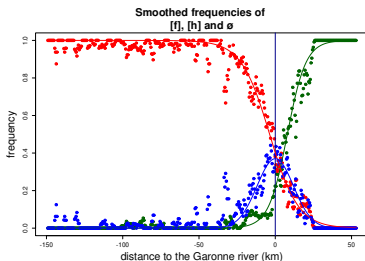
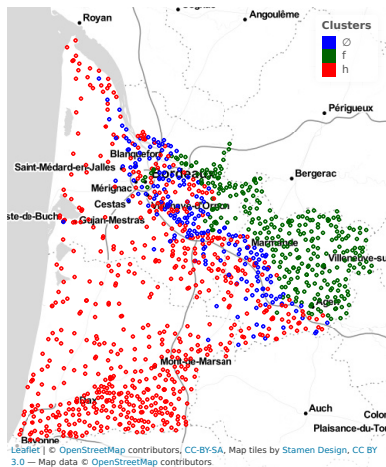
Un exemple de prononciation de la iòd [j] (per inférença de la notacion).

Showcase : Morphologia

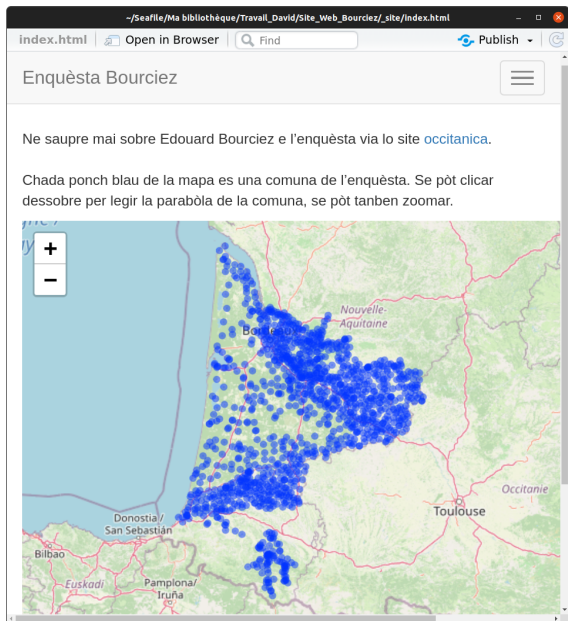


Morphologia del preterit de *díser/dire*.

Showcase : Clustering, Ajustement



Showcase : Lo site



The screenshot shows a web browser window with the following elements:

- Address bar: `~/Seafille/Ma bibliothèque/Travail_David/Site_Web_Bourciez/_site/index.html`
- Page title: `index.html`
- Navigation: `Open in Browser`, `Find` search bar, `Publish` button, and a refresh icon.
- Section header: `Enquèsta Bourciez`
- Text: `Ne saupre mai sobre Edouard Bourciez e l'enquèsta via lo site occitanica.`
- Text: `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.`
- Map: A map of Occitania, France, with numerous blue dots representing communes. Labels on the map include `Bordeaux`, `Nouvelle-Aquitaine`, `Occitanie`, `Toulouse`, `Donostia/San Sebastián`, `Bilbao`, `Euskadi`, and `Pamplona/Iruña`. A zoom control with `+` and `-` buttons is visible on the left side of the map.

Vos grandmerceji per vòstra escota !