Cartes et graphiques

Quelques pistes de représentation de données linguistiques en utilisant R

Timothée Premat | Univ. Paris 8 & CNRS : UMR 7023 (SFL)
Séminaire des doctorants de SFL (Univ. Paris 8 & CNRS : UMR 7023), 26/01/2021.

Introduction

Problèmes de visualisation de données en linguistique

- La linguistique est souvent une affaire de variation
  - *Faire émerger l'ordre du magma chaotique des données*
- Pour ça, rien de mieux que de visualiser (correctement) les données !

- Présenter quelques problèmes de visualisation de données que j'ai pu rencontrer
- Présenter les solutions que j'ai pu trouver, les discuter avec vous
- Et éventuellement vous donner des pistes si vous avez rencontré des problèmes analogues

Plan

Objectif : introduction à R pour les données variationnelles

- Contexte : mes recherches
- Outil : présentation de R
- Applications
  - Variation dialectale : cartographie linguistique
  - Variation dialectale : représentation non cartographique
  - Distribution de données non dialectale
- Conclusion

Avant R

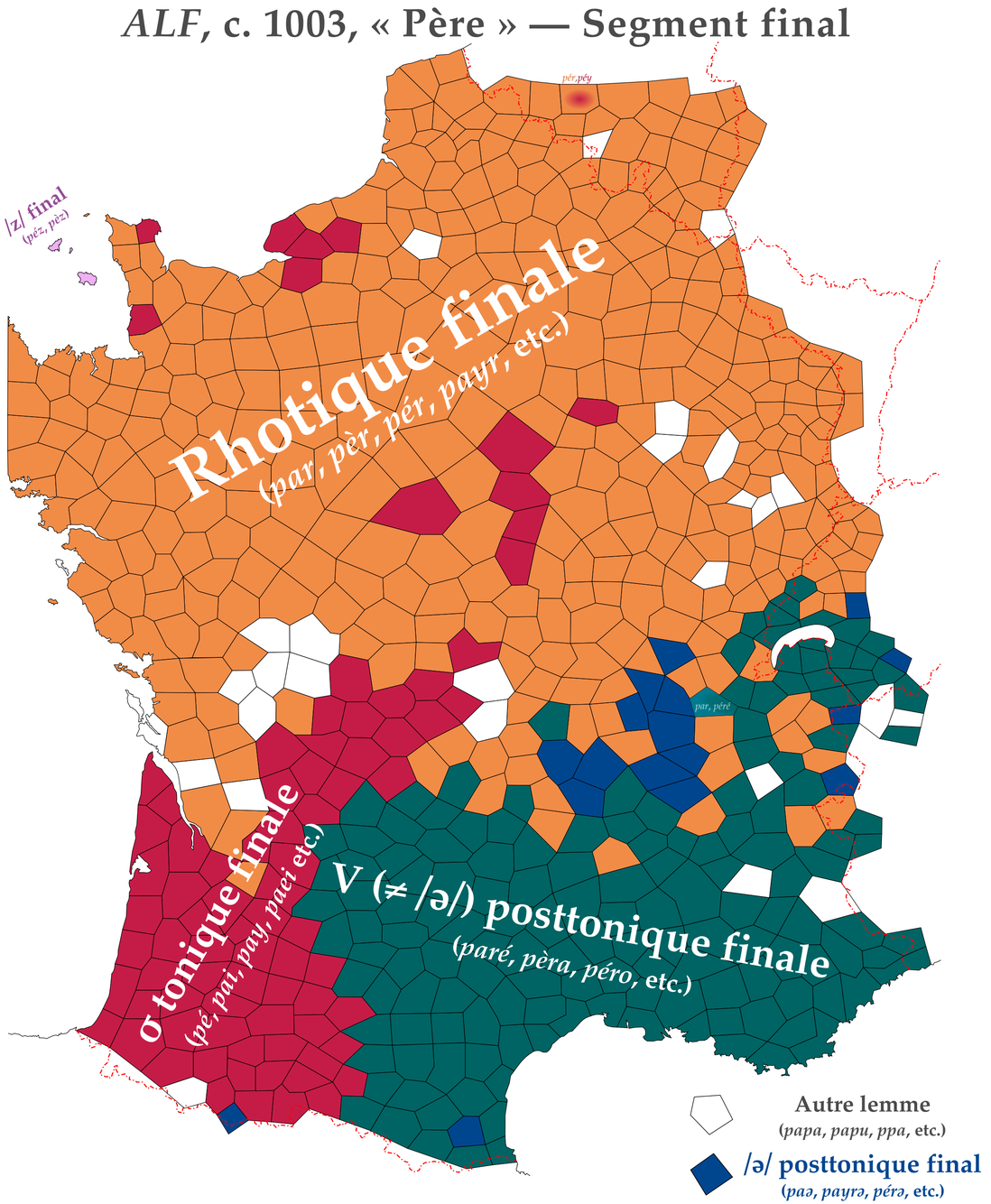
Faire des cartes sans R…

- À la main
- Avec des programmes de graphismes
  - (Inskape ci-contre)

➡ Très mauvaise idée !

  - Très long
  - Très compliqué à modifier en cas d’erreur
  - Aucun traitement *automatisé*

- Les applications SIG (*système d’information géographique*) ne sont pas nécessairement adaptées au besoin du linguistique.



Contexte

Contexte

Ma thèse

- Traitement du /ə/ final en ancien et moyen français (IX^{e}-XVI^{e} s.)
  - Dir. Michela Russo & Sophie Wauquier
- Le schwa français est un objet fondamental pour la théorie phonologique, y compris à l'échelle *cross-linguistic* (Anderson 1982)
- Français contemporain 'standard' : schwa résulte d'une insertion phonotactique (insertion de schwa là où il faut une voyelle/un noyau de syllabe et qu'il n'y a pas de voyelle disponible)
  - *belle fille* : /bɛl fiʎ/ [bɛl fiʎ]
- Ancien français : schwa final est voyelle lexicale, supprimée uniquement en hiatus (= élision)
  - *belle fille* : /bɛlə fiʎə/ [bɛlə fiʎə]
  - *belle amie* : /bɛlə amiə/ [bɛl amiə]
- Moyen français : schwa final commence à être supprimé progressivement (⇢ fr.cont.)

Contexte

Dimensions

- Phonologie représentationnelle : qu'est-ce que schwa en a.fr. et m.fr. ?
- Phonologie prosodique : quel est le domaine de l'élision ?
- Métrique : comment ça marche dans le discours versifié ?
- Musication : comment ça marche en musique ?
  - métrique et musication $\in$ linguistique
  - donnent accès à des information prosodique en l'absence d'enregistrements.
- Dialectologie :
  - a.fr. : langue non standardisée = multiplicité de normes
  - m.fr. : langue en voie de standardisation = unification progressive des normes
  - Donc : grosse variation diatopique.
- Diachronie

Variations diachronique et diatopique… = données confuses

Comment y voir clair ?

→ Représentation spatiale :

- cartes
- graphes

R

R

Langage de programmation et logiciel libre

- langage informatique : permet d'interagir avec un ordinateur pour lui faire faire des opérations
- dédié aux statistiques et à la science des données
- permet d'effectuer des opérations sur un jeu de données :
  - faire des calculs (moyenne, coefficient de corrélation (r, ρ, τ), test de probabilités χ etc.)
  - produire des représentations spatiales
    - 'graphiques'
  - produire des transformations

R

Langage de programmation et logiciel libre

- langage ‘simple’ à apprendre
  - syntaxe explicite, faite pour être écrite et lue par un humain
  - larges documentation et communauté en ligne
    - Beaucoup de tutoriels/MOOC en ligne
- permet de faire des choses *vraiment pénibles* à faire dans des programmes grand public comme Excel.
- il existe des alternatives propriétaires et payantes *a priori* plus simple à manier, et des alternatives libres plus compliquées (p.ex. on peut faire tout ça directement en Python)

R

Workflow sur R

1. Préparer les données (saisie manuelle, automatique, extraction depuis une base de données, etc.)

2. Rédaction du *script* R (ensemble d'instructions R) dans un éditeur de texte

   - Je conseille *Atom* (gratuit et libre)

3. Exécution de ces instructions dans un logiciel spécialisé

   - Je conseille *RStudio* (gratuit), qui propose une interface graphique pour un certain nombre d'actions

➡ Ce qui produit le résultat des instructions (2) sur le jeu de données (1) : graphiques, résultats de calculs, etc.

Dataset

Exemple de jeu de données cartographiques

name;item

name	Lat	Long	description	CODE_DEPT	Nom_dept	Pays	name
1;[õ,õ]	493	48.508533	-2.916611	Plouvara	22	C�tes-dArmor	France	493
3;[õ,õ]	494	48.280427	-2.84085	Uzel	22	C�tes-dArmor	France	494
4;[õ,õ]	485	48.03455	-2.765412	Cr�din	56	Morbihan	France	485
5;[ã,ã]	486	47.837868	-2.639578	Plumelec	56	Morbihan	France	486
6;[ã,ã]	399	49.448627	-2.553337	Saint-Peter-Port	0	Guernesey (Guernsey)	Normandie (Couronne
7;Ø	482	48.448	-2.486	Noyal	22	C�tes-dArmor	France	482
8;[ã,ã]	475	47.591915	-2.456789	Noyal-Muzillac	56	Morbihan	France	475
10;[ã,ã]	484	47.988586	-2.383094	Loyat	56	Morbihan	France	484
11;[ã,ã]	476	47.356155	-2.369867	Gu�rande	44	Loire-Atlantique	France	476
12;[ã,ã]	398	49.430737	-2.360516	Sark	0	Serq (Sark)	Normandie (Couronne britannique)	398
13;Ø	479	46.72451	-2.348529	L'Ile d'Yeu	85	Vend�e	France	479
14;[ã,ã]	481	48.655113	-2.331348	Pl�venon	22	C�tes-dArmor	France	481
16;[ã,ã]	478	47.024269	-2.303349	Noirmoutier-en-l'�le	85	Vend�e	France	478
17;[ã,ã]	483	48.183334	-2.233333	Loscou�t-sur-Meu	22	C�tes-dArmor	France	483
19;[õ,õ]	396	49.722038	-2.201772	Sainte-Anne	0	Aurigny (Alderney)	Normandie (Couronne britanni
20;[ã,ã]	465	47.636829	-2.126369	Saint-Jean-la-Poterie	56	Morbihan	France	465
21;[ã,ã]	466	47.396751	-2.09179	Besn@	44	Loire-Atlantique	France	466
22;[õ,õ]	397	49.235168	-2.090838	Trinity	0	Jersey	Normandie (Couronne britannique)	397
23;[ã,ã]	463	47.876583	-2.084128	Comblessac	35	Ille-et-Vilaine	France	463
24;N/A	471	48.383331	-2.066667	Tr@vron	21	C@te-dOr	France	471
25;N/A	394	49.711658	-1.932204	Auderville	50	Manche	France	394
26;N/A	467	47.128164	-1.910011	Ch@m@r@	44	Loire-Atlantique	France	467
27;Ø	470	48.604076	-1.893617	La Gouesni@re	35	Ille-et-Vilaine	France	470
28;Ø	459	46.671261	-1.883768	Givrand	85	Vend@e	France	459
30;Ø	458	46.883331	-1.833333	La Garnache	85	Vend@e	France	458
31;[a,ɑ]	462	48.12746	-1.817488	L'Hermitage	35	Ille-et-Vilaine	France	462
32;[ã,ã]	453	47.824821	-1.806856	Messac	35	Ille-et-Vilaine	France	453
33;[ã,ã]	395	49.402164	-1.780064	Les Moitiers-d'Allonne	50	Manche	France	395
35;N/A	461	48.309021	-1.668098	Montreuil-sur-Ille	35	Ille-et-Vilaine	France	461
36;[ã,ã]	540	46.466759	-1.617908	Talmont-Saint-Hilaire	85	Vend@e	France	540
38;[õ,õ]	387	49.200645	-1.562132	Cr@ances	50	Manche	France	387
40;[õ,õ]	690	43.483093	-1.558613	Biarritz	64	Pyr@n@es-Atlantiques	France	690
41;[õ,õ]	460	48.516666	-1.55	Vieuxviel	35	Ille-et-Vilaine	France	460
42;[õ,õ]	446	47.340271	-1.526231	Suc@-sur-Erdre	44	Loire-Atlantique	France	446

Script

Exemple de Script R

```r
#Timothée Premat | 25 octobre 2020, based on original script by:
#Mathieu Avanzi | 22 avril 2020

#packages that need to be installed prior to run the code
library(maps)
library(plyr)
library(dplyr)
library(rgdal)
library(rgeos)
library(ggplot2)
library(ggsn)
library(scales)
library(ggmap)

#load data
dataALF = read.table("commencent_ALF_simple.txt", header=T, sep=";", quote="", dec=".")
dataALF2 = read.table("ALF_point.txt", header=T, sep="\t", quote="", dec=".")
dataALF = merge(dataALF,dataALF2,by="name")
dataALF[!complete.cases(dataALF),]

head(dataALF)

#get map
world <- get_googlemap("Paris", zoom = 6, style = s, size=c(640,640), scale=2)

#plot it
g= ggmap(world)+
  geom_point(data=dataALF, aes(x=Long,y=Lat,shape=item), size=3) +

  scale_x_continuous(limits = c(-3.5,8), expand = c(0, 0)) +
  scale_y_continuous(limits = c(45,51.15), expand = c(0, 0))+

  scale_shape_manual(values=c(15,16,17,18,14,0,1,2,5,43,42), "",
                     limits=c("[ẽ,ɛ̃]","[ã,ɑ̃]","[õ,ɔ̃]","[ũ]","[ə̃]",
```

RStudio

Exécution du script dans RStudio

RStudio

BFM_date.txt | script_NCA_map_dial_compo.R | script_NCA_map_dial_ms.R | SCR_A_commencent.R

```
1  #Timothée Premat | 25 octobre 2020, based on original script by:
2  #Mathieu Avanzi | 22 avril 2020
3
4  #packages that need to be installed prior to run the code
5  library(maps)
6  library(plyr)
7  library(dplyr)
8  library(rgdal)
9  library(rgeos)
10 library(ggplot2)
11 library(ggsn)
12 library(scales)
13 library(ggmap)
14
15 #load data
16 dataALF = read.table("commencent_ALF_simple.txt", header=T, sep=";", quote="", dec=".")
17 dataALF2 = read.table("ALF_point.txt", header=T, sep="\t", quote="", dec=".")
18 dataALF = merge(dataALF,dataALF2,by="name")
19 dataALF[!complete.cases(dataALF),]
```

72:1 (Top Level) R Script

Console | Terminal | Markers | Jobs

~/Documents/Thèse/Atlas/ALF_commencent/

```
+         legend.justification=c(0,0),
+         axis.title.x=element_blank(),
+         axis.text.x=element_blank(),
+         axis.ticks.x=element_blank(),
+         axis.title.y=element_blank(),
+         axis.text.y=element_blank(),
+         axis.ticks.y=element_blank(),
+         legend.position=c(0.0,0.0),
+         legend.spacing=unit(1,"lines"),
+         legend.box="vertical",
+         legend.key.size=unit(1.2,"lines"),
+         legend.text.align=0,
+         legend.title.align=0,
+         legend.text=element_text(size=12, color="black"),
+         legend.key = element_rect(fill = "transparent", color="transparent"),
+         legend.title=element_text(size=1, color="transparent", face = "bold"),
+         legend.background = element_rect(fill = alpha("white",0.5), colour = "transparent")) +
+
+  annotate("rect", xmin = -3.5, xmax = 1.1, ymin = 50.53, ymax = 51.15, fill="White", alpha = 0.5)+
+  annotate("text", x = -3.3, y = 51, label = "Commencent", size=5, color = "black", hjust = 0) +
+  annotate("text", x = -3.3, y = 50.8, label = "Présence d'une voyelle posttonique", size=4, color ="b
lack",hjust = 0) +
+  annotate("text", x = -3.3, y = 50.62, label = "ALF c. 311 : (Les pommiers) commencent (à fleurir)",
 size=3, color ="black",hjust = 0) +
+  annotate("text", x = -3.3, y = 50.47, label = "© Timothée Premat", size=2.5, color ="white",hjust =
 0)
Scale for 'x' is already present. Adding another scale for 'x', which will replace the existing
scale.
Scale for 'y' is already present. Adding another scale for 'y', which will replace the existing
scale.
>
>
> #Print plot
> g
Warning messages:
1: Removed 1 rows containing missing values (geom_rect).
2: Removed 50 rows containing missing values (geom_point).
>
> #Save plot
> ggsave(g, filename = "Aacommencent_ALF.png", scale=1)
Saving 11.7 x 11.1 in image
Warning messages:
1: Removed 1 rows containing missing values (geom_rect).
2: Removed 50 rows containing missing values (geom_point).
>
```

Environment | History | Connections | Tutorial

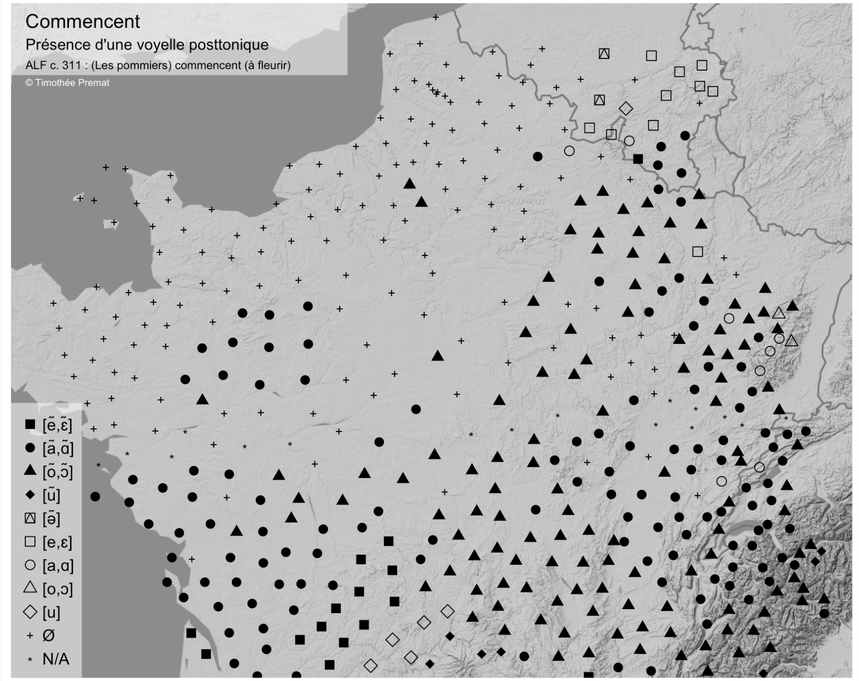
Import Dataset | List

Global Environment

HOM_R_2	153 obs. of 4 variables
HOM_R_3	222 obs. of 3 variables
HOM_R_3_Prose	119 obs. of 3 variables
HOM_R_3_Vers	91 obs. of 3 variables
HOMME	266 obs. of 5 variables
opar	List of 1
p	List of 9
palA_et_U	52 obs. of 13 variables
Scatter_accents	20 obs. of 4 variables
Scatter_voix	71 obs. of 4 variables
VIDE	26 obs. of 11 variables
Voix_accents	4 obs. of 3 variables
Values	
lev	int [1:12] 90 100 110 120 130 140 150 160 170 180 ...
n	100
name	"Forme"
pie.sales	Named num [1:6] 0.12 0.3 0.26 0.16 0.04 0.12
pin	num [1:2] 2.52 4.32
s	"https://maps.googleapis.com/maps/api/staticmap?key=AIzaSyDo7mu1ICdeIVlK…

Files | Plots | Packages | Help | Viewer

Zoom | Export | Publish



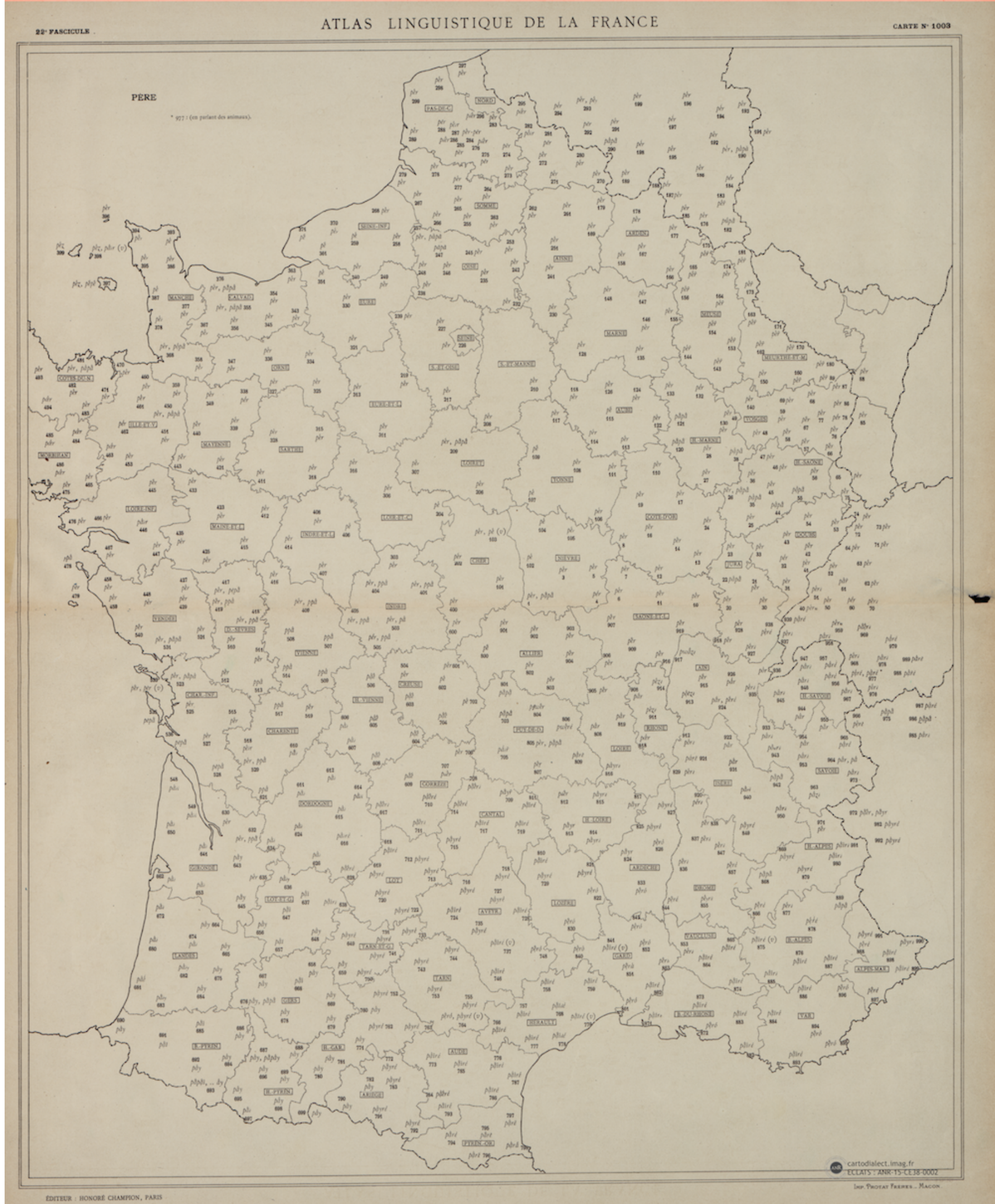
Applications

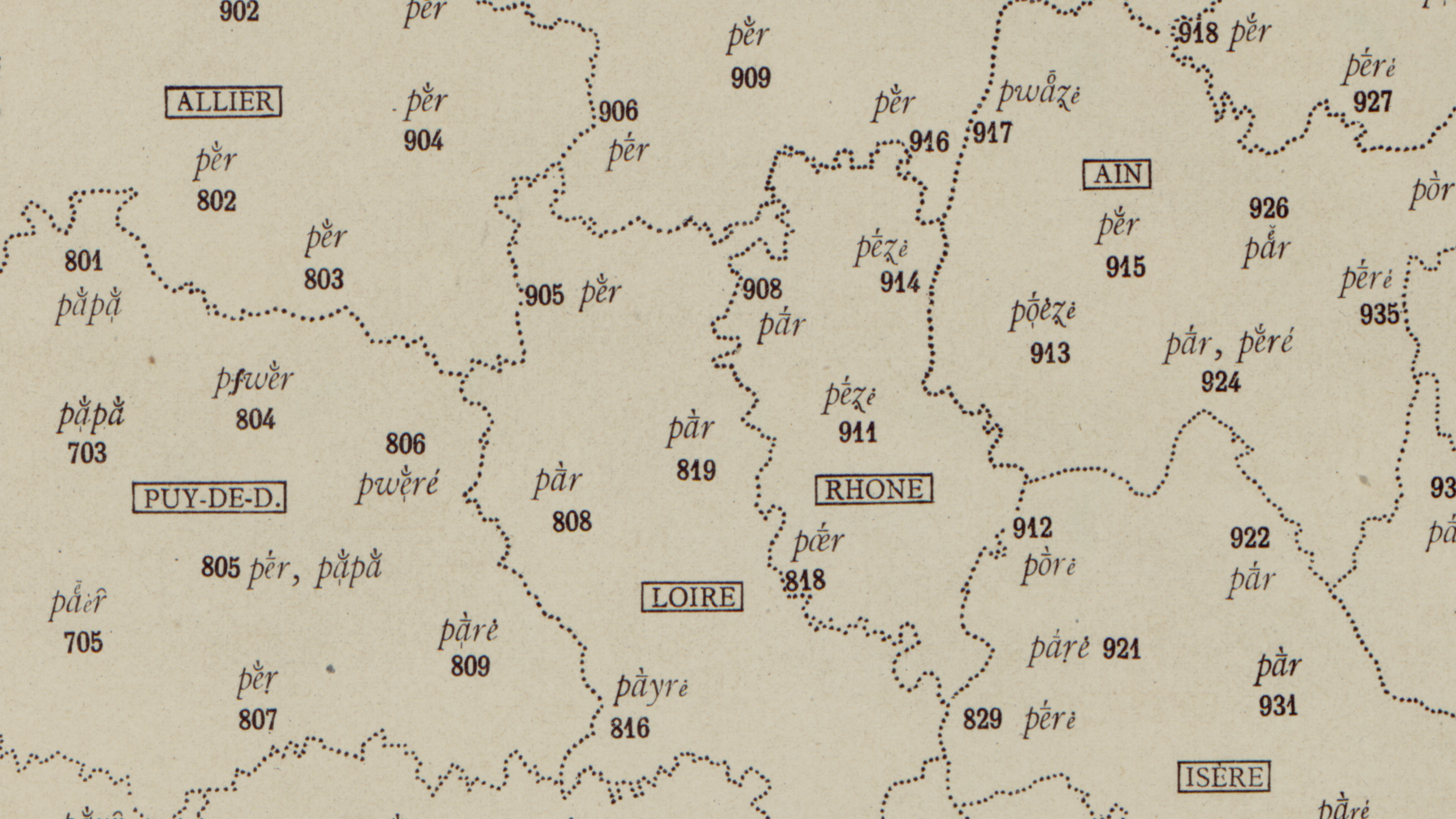
Cartographie avec R

ALF

Exemple de source de données

- *Atlas linguistique de la France*
- Gilliéron & Edmond (1902-1910)
  - http://lig-tdcge.imag.fr/cartodialect5/
- *Comment vous dites ça ?*
- Une carte par mots
- Une transcription phonétique par point d'enquête
- Carte brute (≠ carte interprétative)



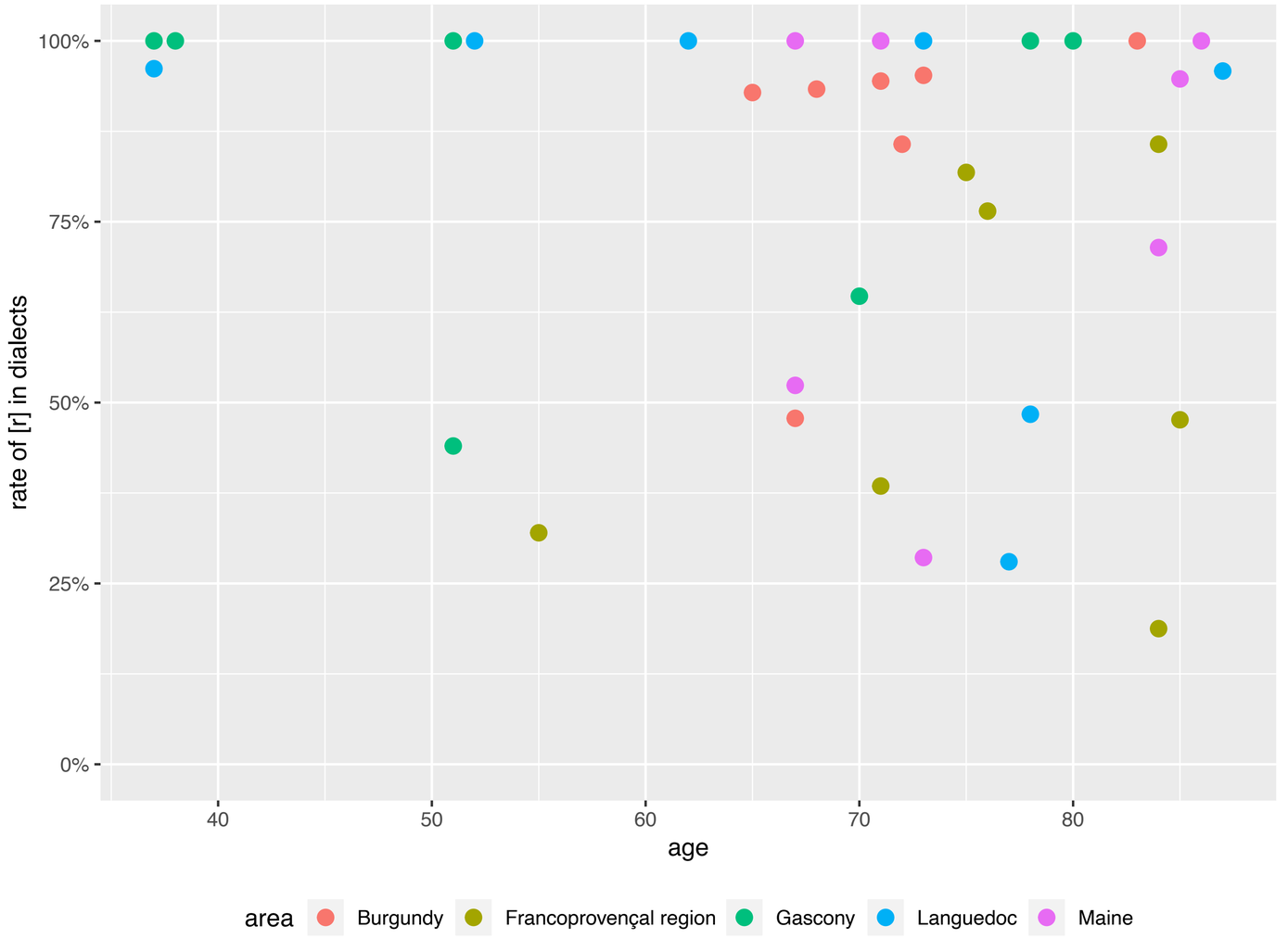

902
pĕr
ALLIER
pĕr
904
906
pĕr
909
pĕr
916
pwåzė
917
AIN
pĕr
915
926
pă̆r
927
pĕrė
801
802
pĕr
803
pă̆pă̆
905 pĕr
908
pằr
914
pẹ̄ẹ̄zė
913
pằr, pĕré
924
pĕrė
935
703
pă̆pă̆
pfwĕr
804
806
pwẹ̄ré
PUY-DE-D.
pằr
808
pằr
819
pẹ̄zė
911
RHONE
912
pœ̄r
818
pṑrė
922
pằr
931
805 pĕr, pă̆pă̆
LOIRE
pă̆ėr
705
pằrė
809
pằrė 921
pằyrė
816
pĕr
807
829 pĕrė
ISÈRE
pằrė

Comment produire des cartes interprétatives de manière automatisée ?

Variation dialectale

Des cartes dans R ?

- R est un programme qui peut produire des graphiques
- Nuage de points (*dotplot*)
  - Chaque point est défini par :
    - des coordonnées *x*, *y*
    - une esthétique
  - Exemple :
    - *x* = âge des locuteurs
    - *y* = taux de [r] sur l'ensemble des /R/
    - couleur : région du locuteur

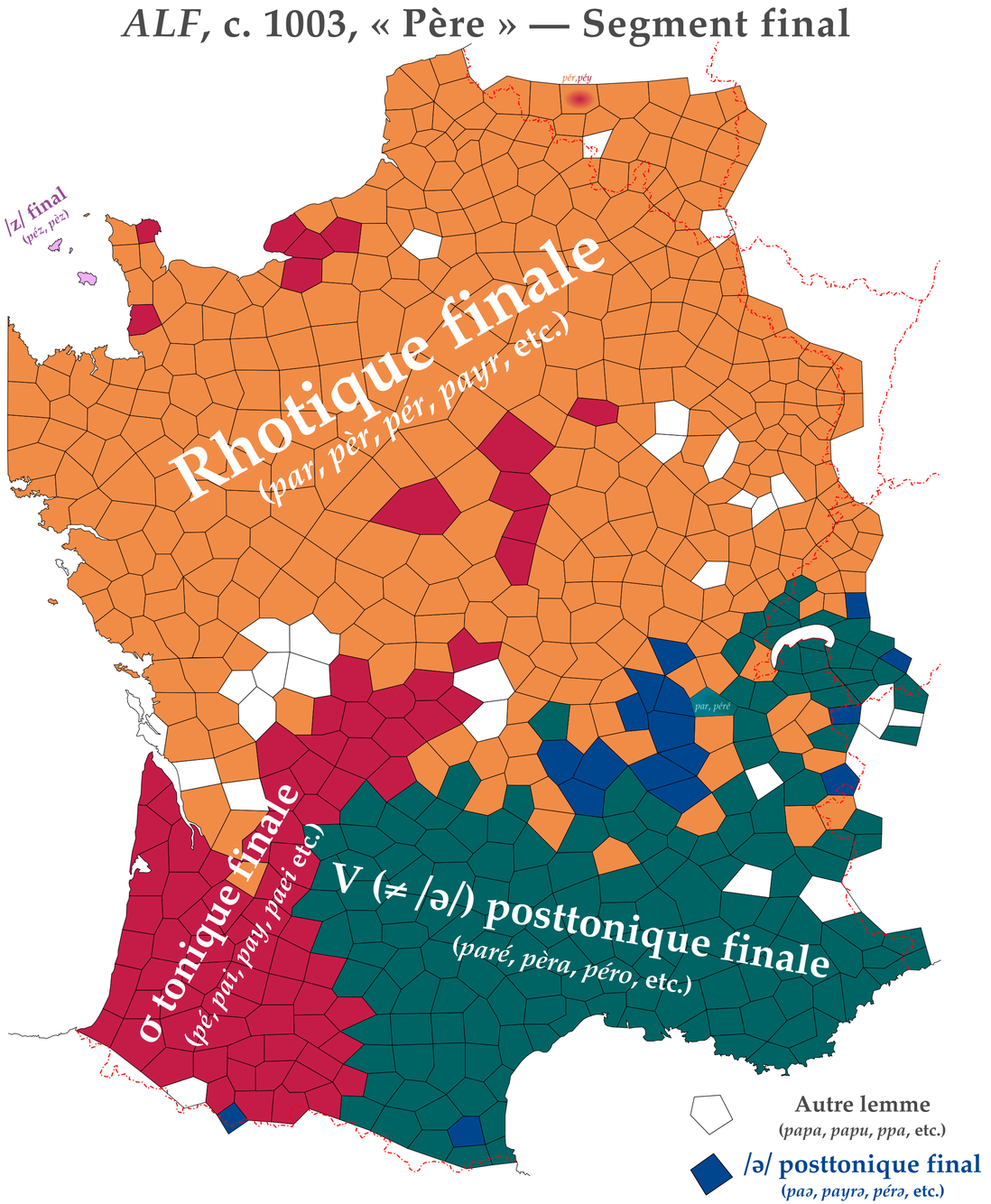


(source : Premat & Boula de Mareüil 2019)

Variation dialectale

Des cartes dans R ?

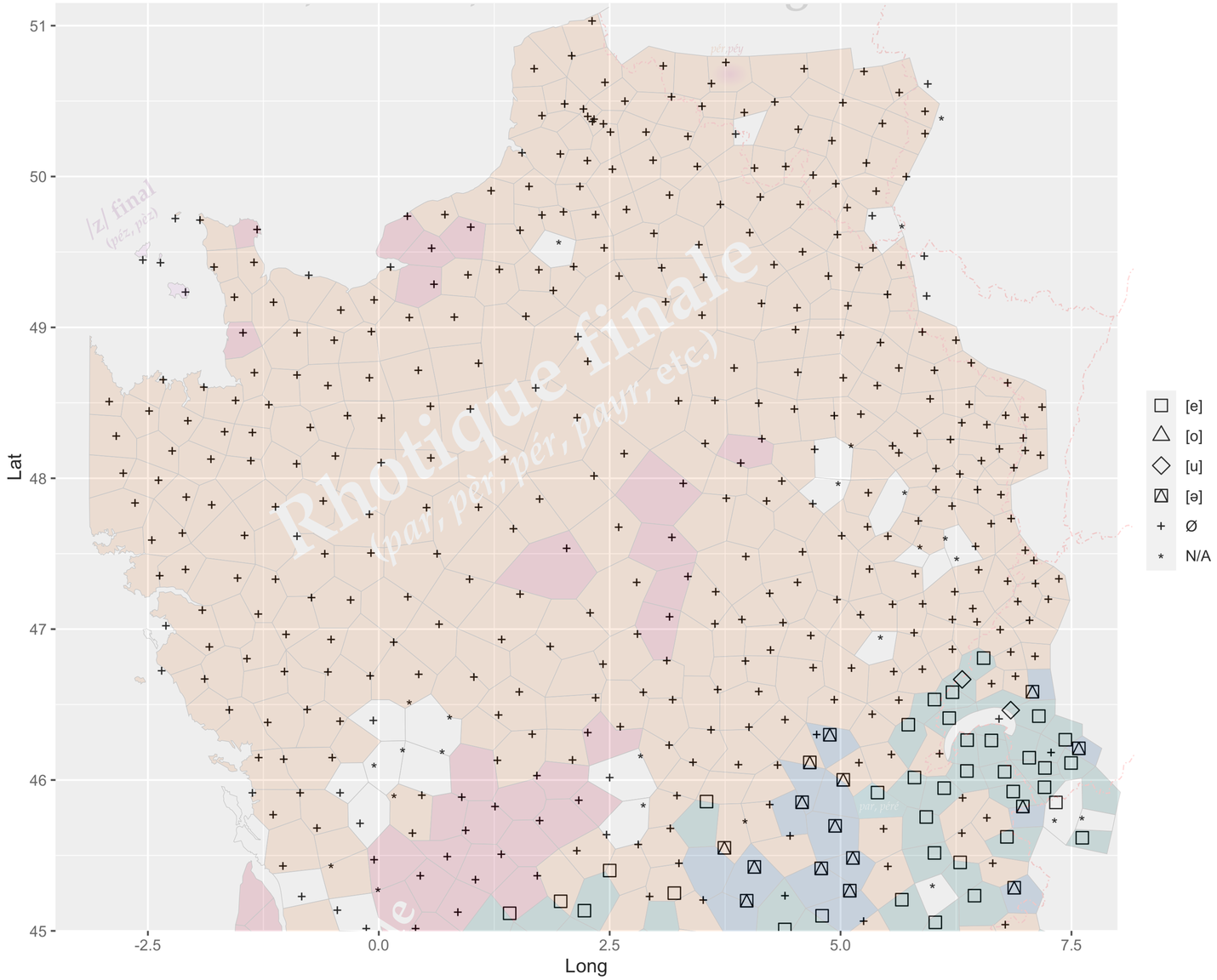
- Une carte *est* un graphique
- Dans le monde réel, chaque point/région est défini·e par :
  - des coordonnées (GPS, etc.)
    - ici, un polygone tracé autour du village dont vient la forme
  - une esthétique
    - ici, par type de réalisation de la V finale atone
  - + un label



Variation dialectale

Des cartes dans R ?

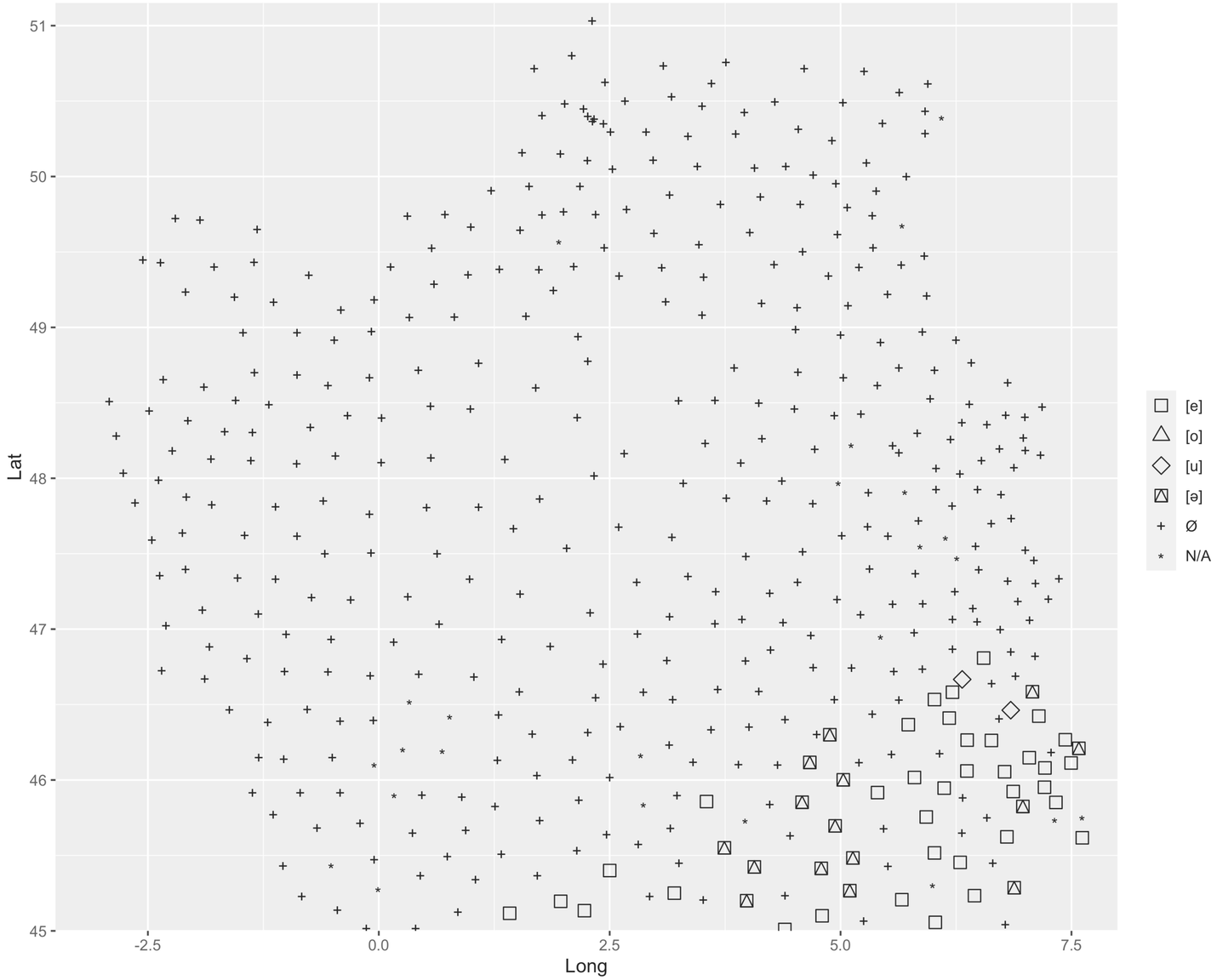
- On peut donc superposer un nuage de point à une carte
- Si :
  - *x* = longitude
  - *y* = latitude
  - forme = typologie *ad hoc*



Variation dialectale

Des cartes dans R ?

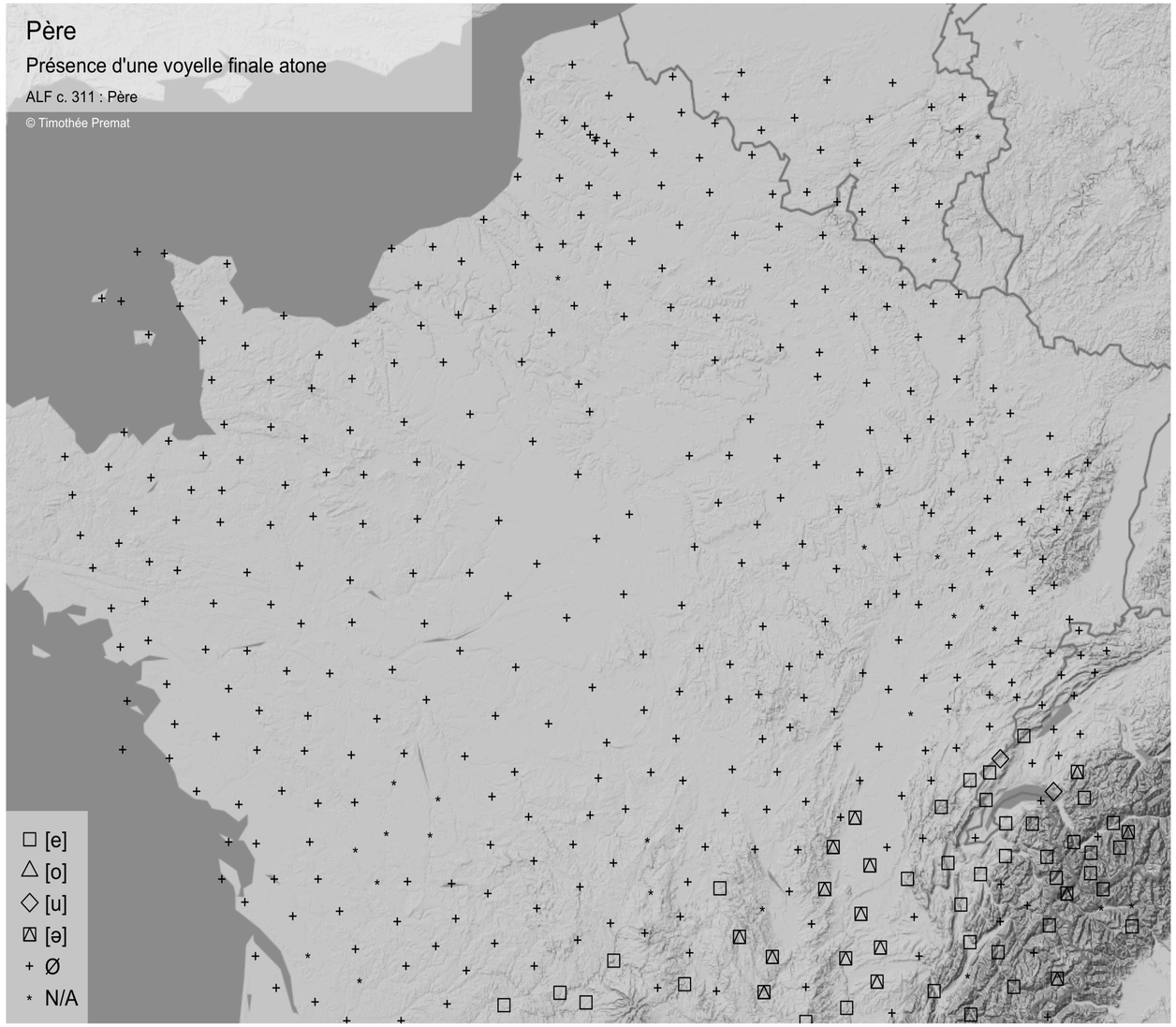
- On peut donc superposer un nuage de point à une carte
- Si :
  - *x* = longitude
  - *y* = latitude
  - forme = typologie *ad hoc*



Variation dialectale

Des cartes dans R ?

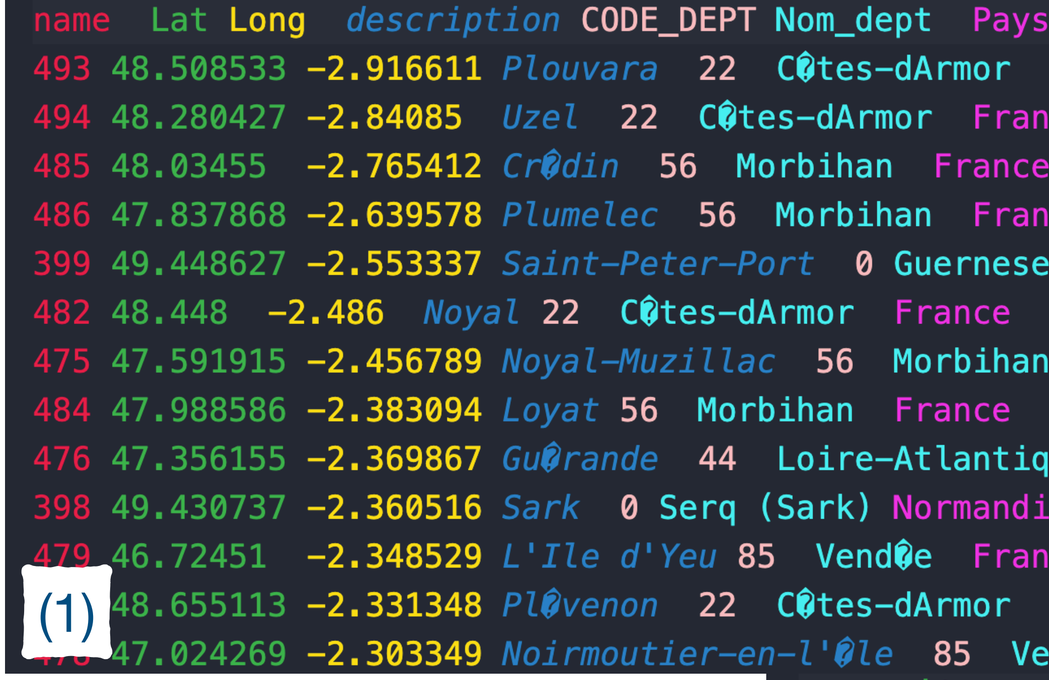
- Il suffit ensuite d'ajouter un fond de carte
- aligné automatique par des *packages* R sur les coordonnées Long, Lat.
- Ici, fond de carte Google Maps personnalisé. Énormément de fonds de cartes différents possibles, paramétrables, etc.



Variation dialectale

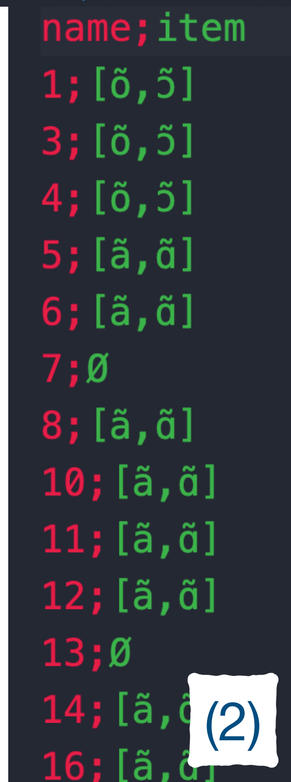
Workflow (procédure de travail)

- Saisie des données dans un tableur
  - d’une part : chaque point : long., lat. (1)
  - chaque point : catégorie de la typologie *ad hoc* (2)
- Écriture du script R
  - définition du fond de carte et des esthétiques
- Exécution du script R
- Pour faire une nouvelle carte, il suffit de préparer les données dans un tableur et de régler les esthétiques, et… c’est tout !



	name	Lat	Long	description	CODE_DEPT	Nom_dept	Pays
493		48.508533	-2.916611	Plouvara	22	C?tes-dArmor	
494		48.280427	-2.84085	Uzel	22	C?tes-dArmor	Fran
485		48.03455	-2.765412	Cr?din	56	Morbihan	France
486		47.837868	-2.639578	Plumelec	56	Morbihan	Fran
399		49.448627	-2.553337	Saint-Peter-Port	0	Guernese	
482		48.448	-2.486	Noyal	22	C?tes-dArmor	France
475		47.591915	-2.456789	Noyal-Muzillac	56	Morbihan	
484		47.988586	-2.383094	Loyat	56	Morbihan	France
476		47.356155	-2.369867	Gu?rande	44	Loire-Atlantiq	
398		49.430737	-2.360516	Sark	0	Serq (Sark)	Normandi
479		46.72451	-2.348529	L'Ile d'Yeu	85	Vend?e	Fran
		48.655113	-2.331348	Pl?venon	22	C?tes-dArmor	
		47.024269	-2.303349	Noirmoutier-en-l'?le	85	Ve	

(1)



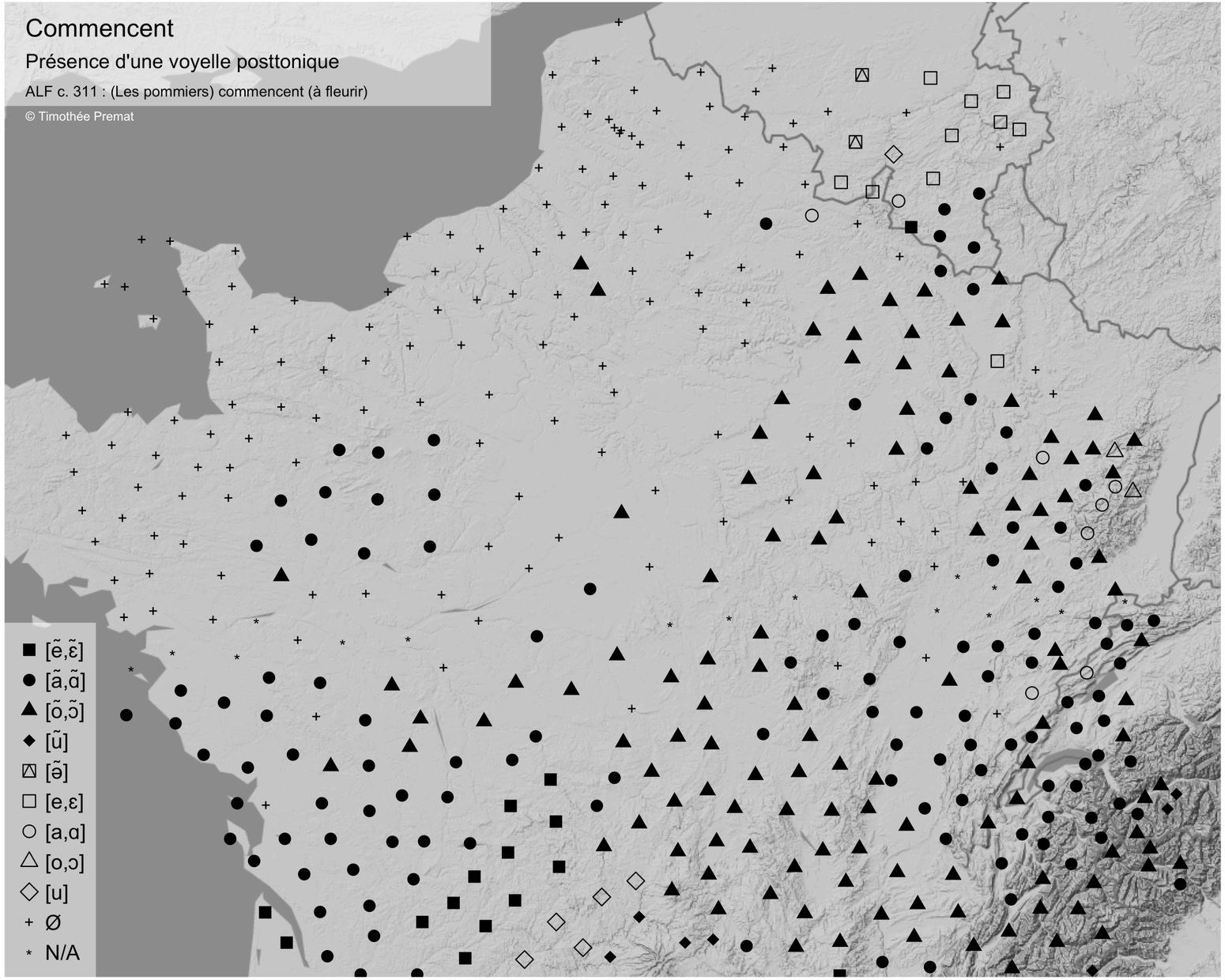
```
name;item
1;[õ,ɔ̃]
3;[õ,ɔ̃]
4;[õ,ɔ̃]
5;[ã,ɑ̃]
6;[ã,ɑ̃]
7;Ø
8;[ã,ɑ̃]
10;[ã,ɑ̃]
11;[ã,ɑ̃]
12;[ã,ɑ̃]
13;Ø
14;[ã,ɑ̃]
16;[ã,ɑ̃]
```

(2)

Variation dialectale

Autres exemples

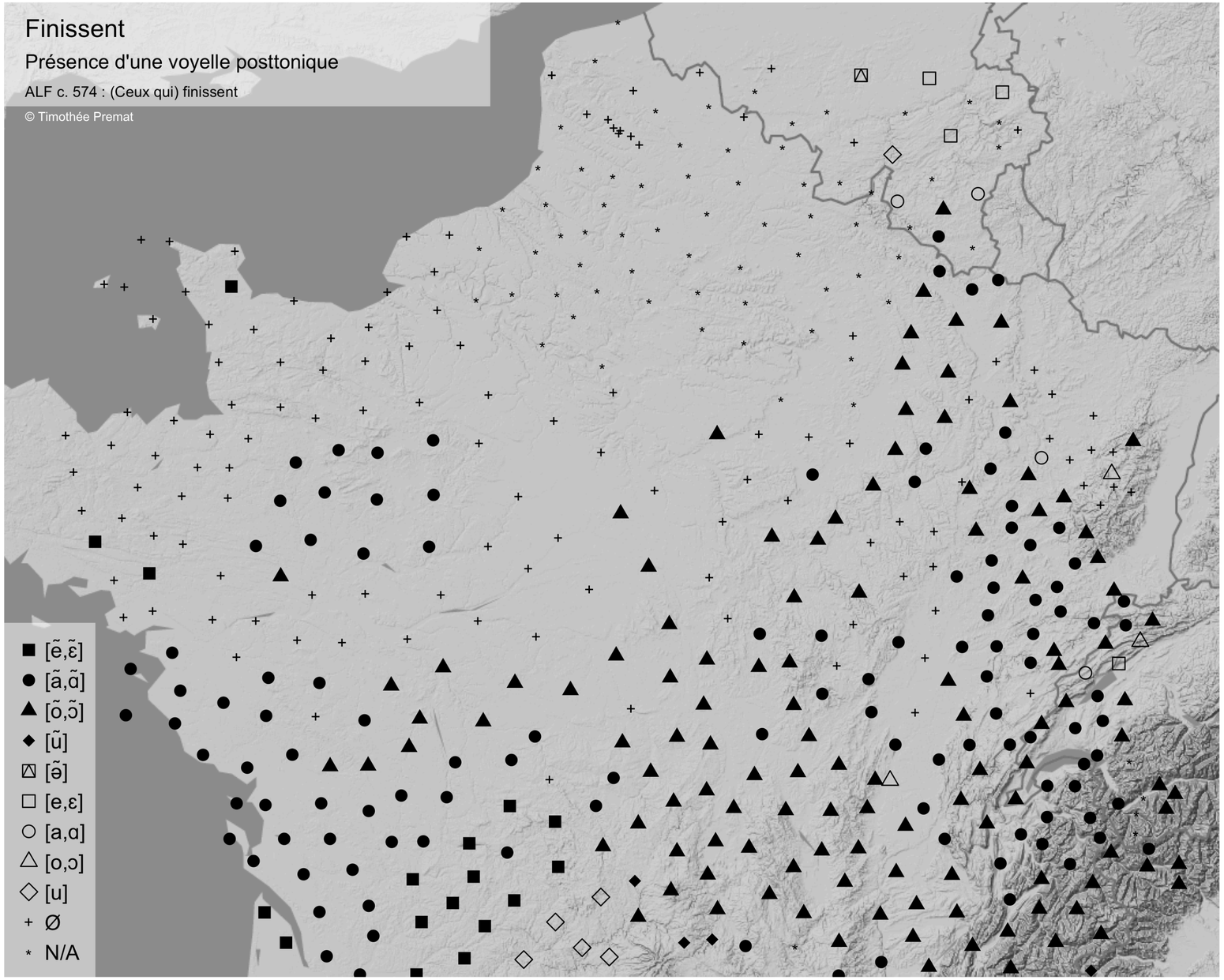
- Présence/absence et type de voyelle pour *-ent* dans *ils commencent*
- Données de l’ALF



Variation dialectale

Autres exemples

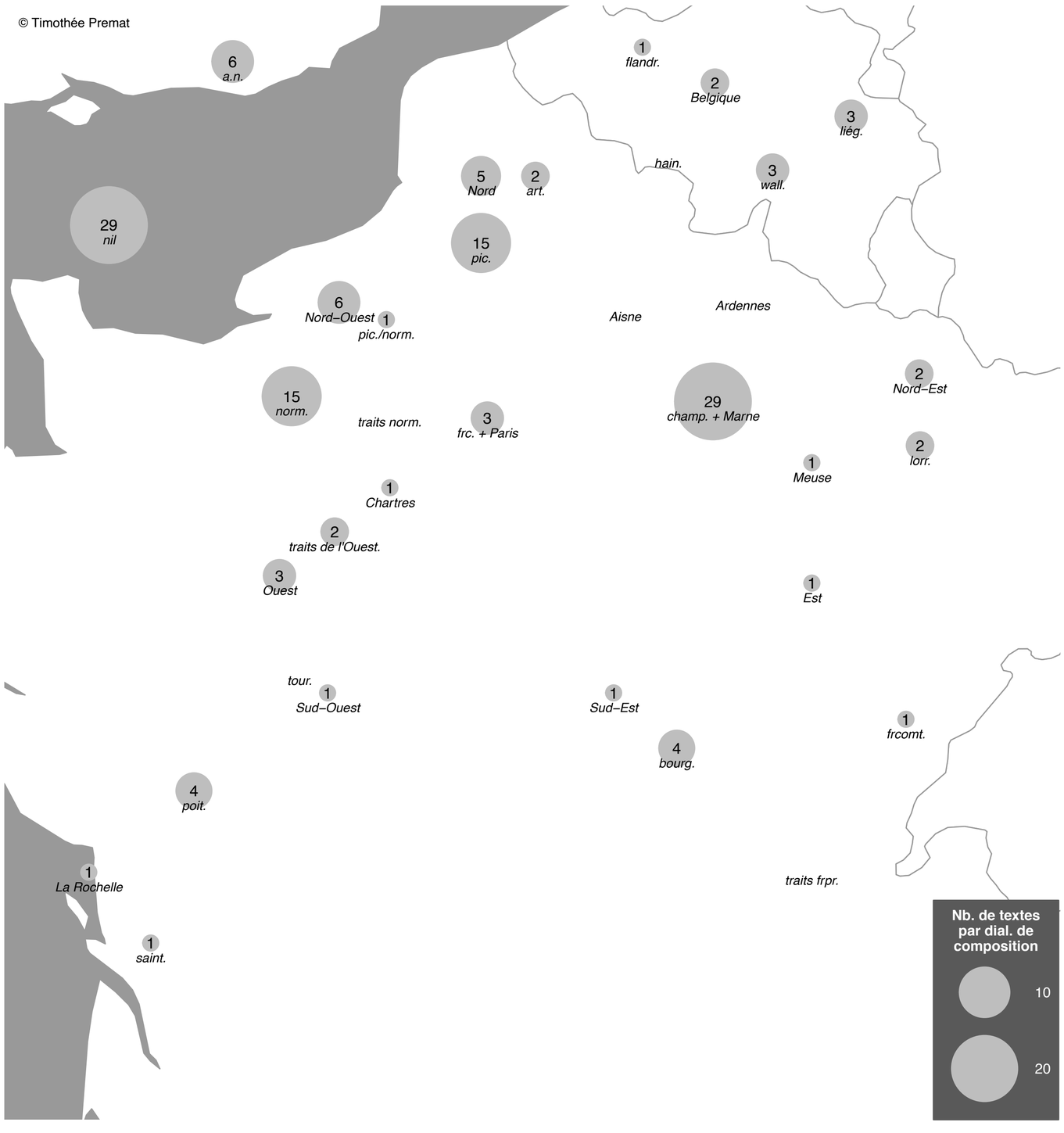
- Présence/absence et type de voyelle pour *-ent* dans *ils finissent*
- Données de l’ALF



Composition d’un corpus

NCA

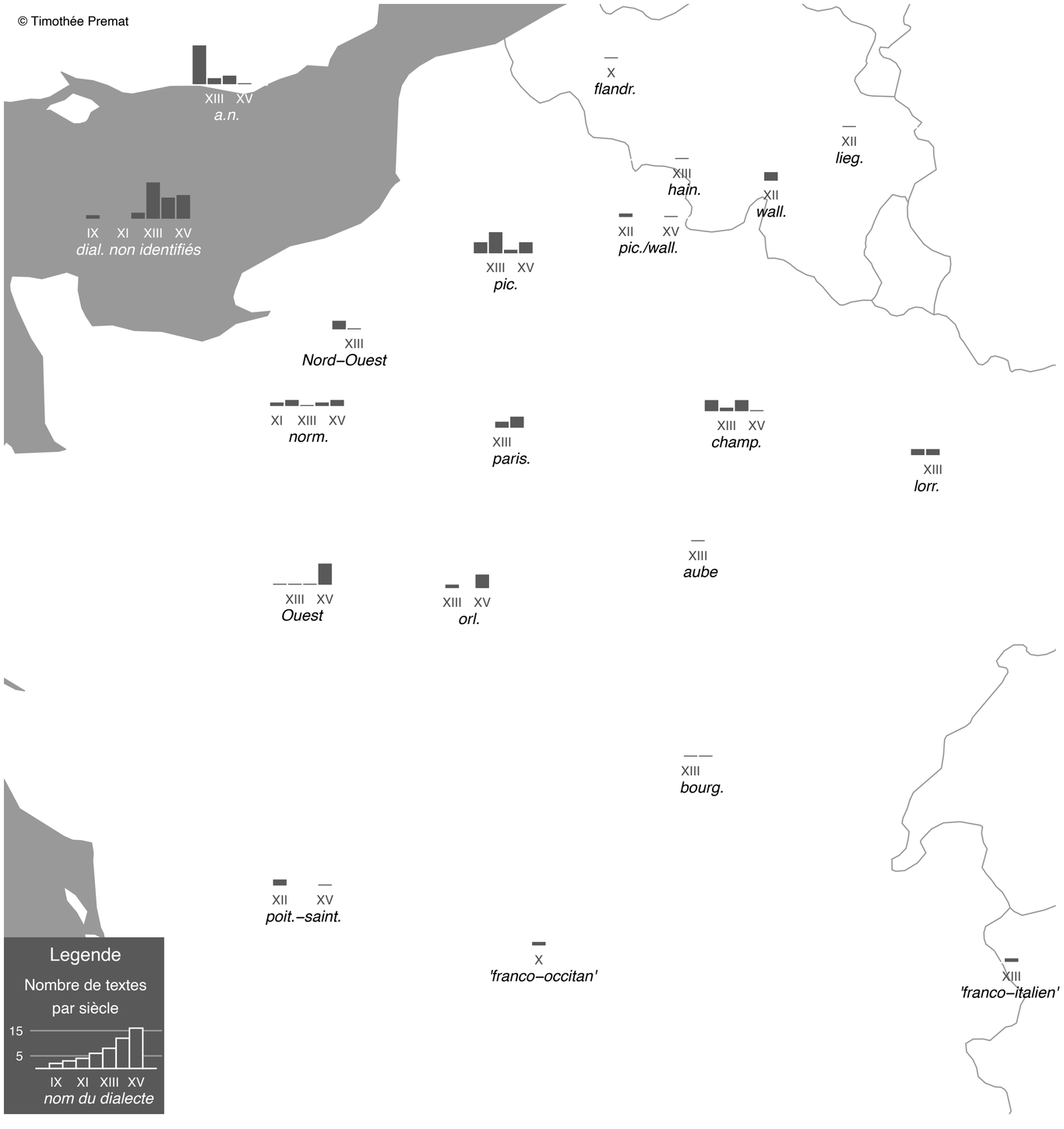
- *Nouveau Corpus d’Amsterdam*
  - Stein, Kunstmann & Gleßgen (2006)
- Objectif : représenter la distribution des textes du corpus
  - par dialecte
- Solution : placer un point pour chaque dialecte, dont la taille représente le nombre de textes



Composition d’un corpus

BFM

- *Base de Français Médiéval*
  - Guillot-Barbance, Heiden & Lavrentiev (2017)
- Objectif : représenter la distribution des textes du corpus
  - par dialecte **et**
  - par siècle
- Solution : placer un histogramme (graphique en barres) représentant la date pour chaque dialecte



Applications

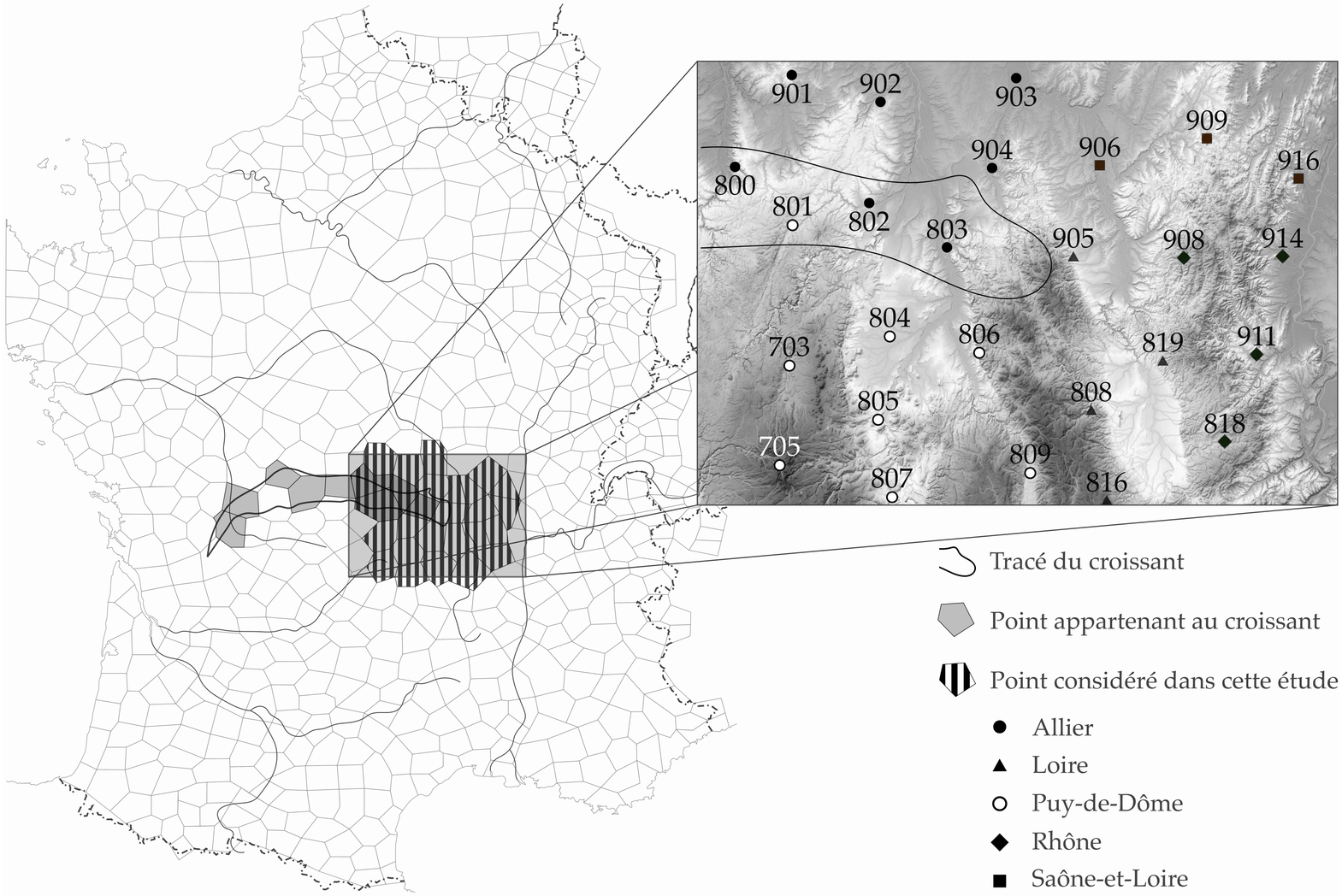
Applications non cartographiques de R

Variation dialectale en phonologie

Variation dialectale

Représentation non géographique

- Russo & Premat :
  - article sur la variation des voyelles finales là où se rencontrent l'oïl, l'occitan et le francoprovençal.
  - Données ← ALF
    - 4 mots par points par type de voyelle
  - Simple à cartographier si les points sont cohérents
  - Plus difficiles si 1 points a différends traitements de la V, ou s'il a des traitements intermédiaires

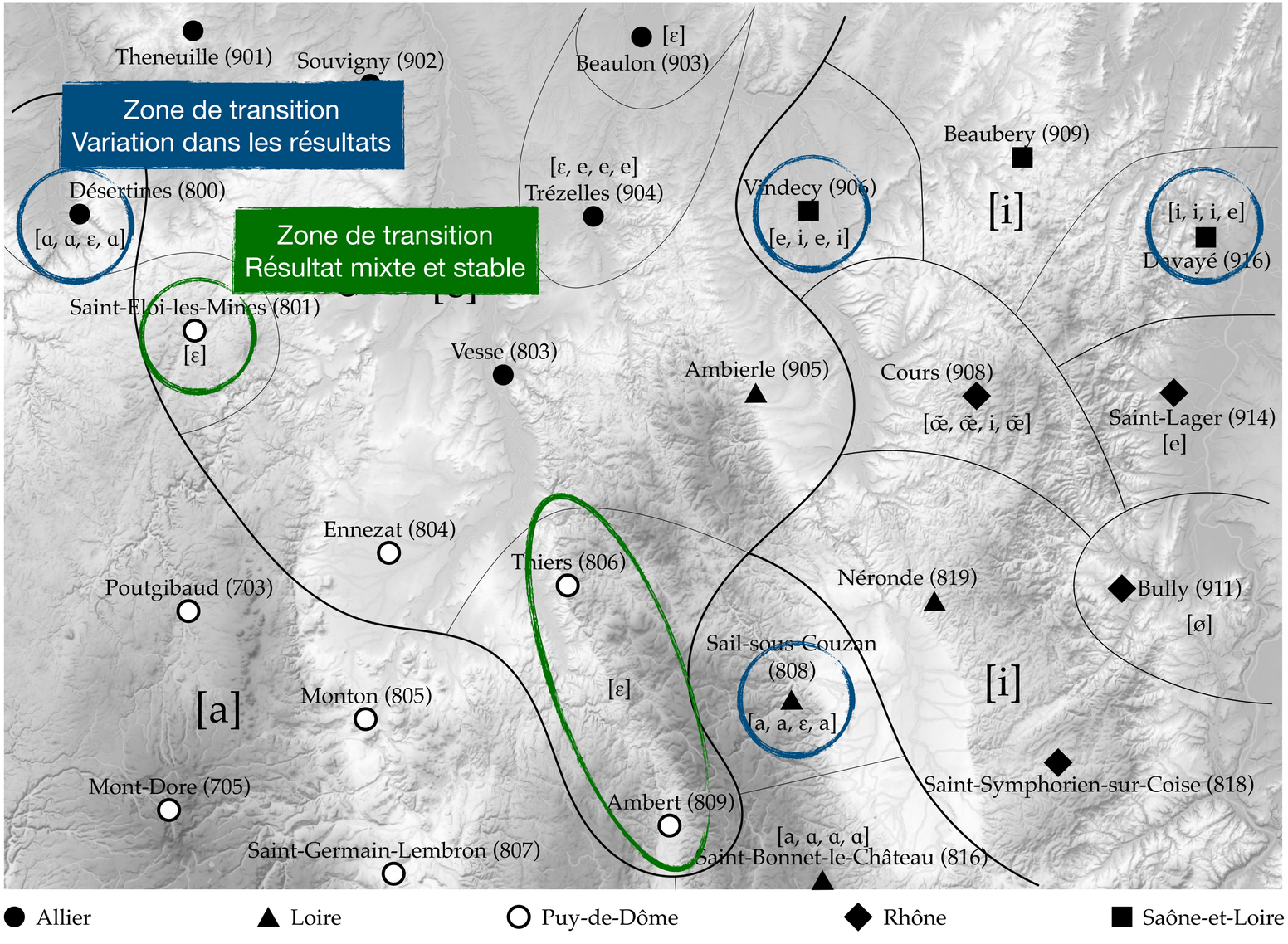


(source : Russo & Premat)

Variation dialectale

Représentation non géographique

- Exemple : Traitement de -ARE suivi d'un élément palatalisant
- Mots :
  - *nettoyer* *NITIDIĀRE
  - *pêcher* *PISCĀRE
  - *pisser* *PISSIĀRE
  - *purger* PURGĀRE
- La carte montre l'*espace géographique*, pas la *proximité dialectale*
- Objectif : montrer à quel diasystème se rattache chaque point

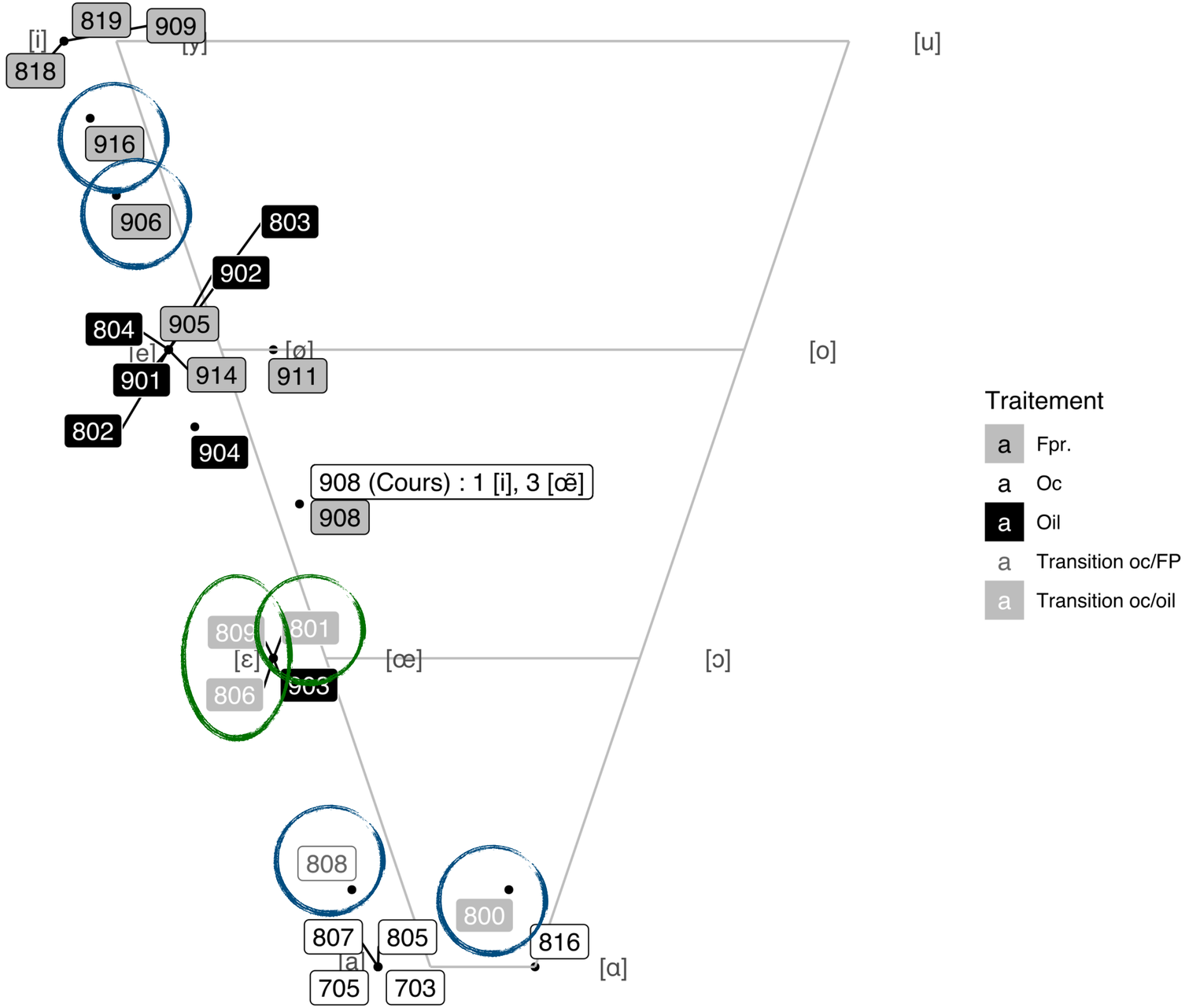


(source : Russo & Premat)

Variation dialectale

Représentation non géographique

- Comment *projeter dans l'espace* des voyelles ?
  - Sur le trapèze vocalique
  - Représente les caractéristiques acoustiques des voyelles (F1, F2)
- Les **résultats mixtes stables** sont naturellement *entre* les diasystèmes
- Comment représenter **les points qui ont une variation** ? En faisant la moyenne de leurs résultats.
- Peut permettre de mesurer à quel point un point est proche de chaque diasystème…



(source : Russo & Premat)

Applications

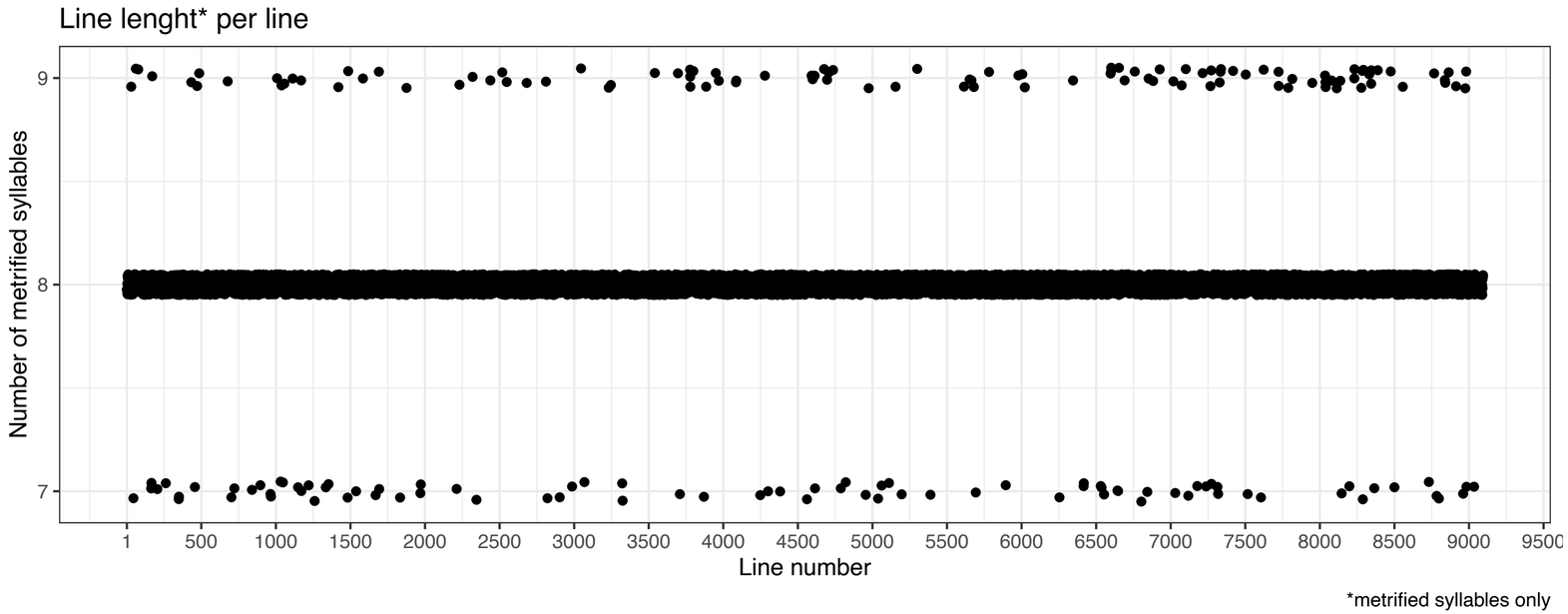
Applications non cartographiques de R

Autres types de variations

Métrique

Analyse des vers d’un texte

- PAM : Programme d’Analyse Métrique (Poggio & Premat 2019)
- Passe en revue un texte et signale quels sont les vers qui sont ‘faux’ (qui n’ont pas le bon nombre de syllabes)
- Objectif : avoir une vue d’ensemble de où sont situés les vers ‘faux’ dans le texte

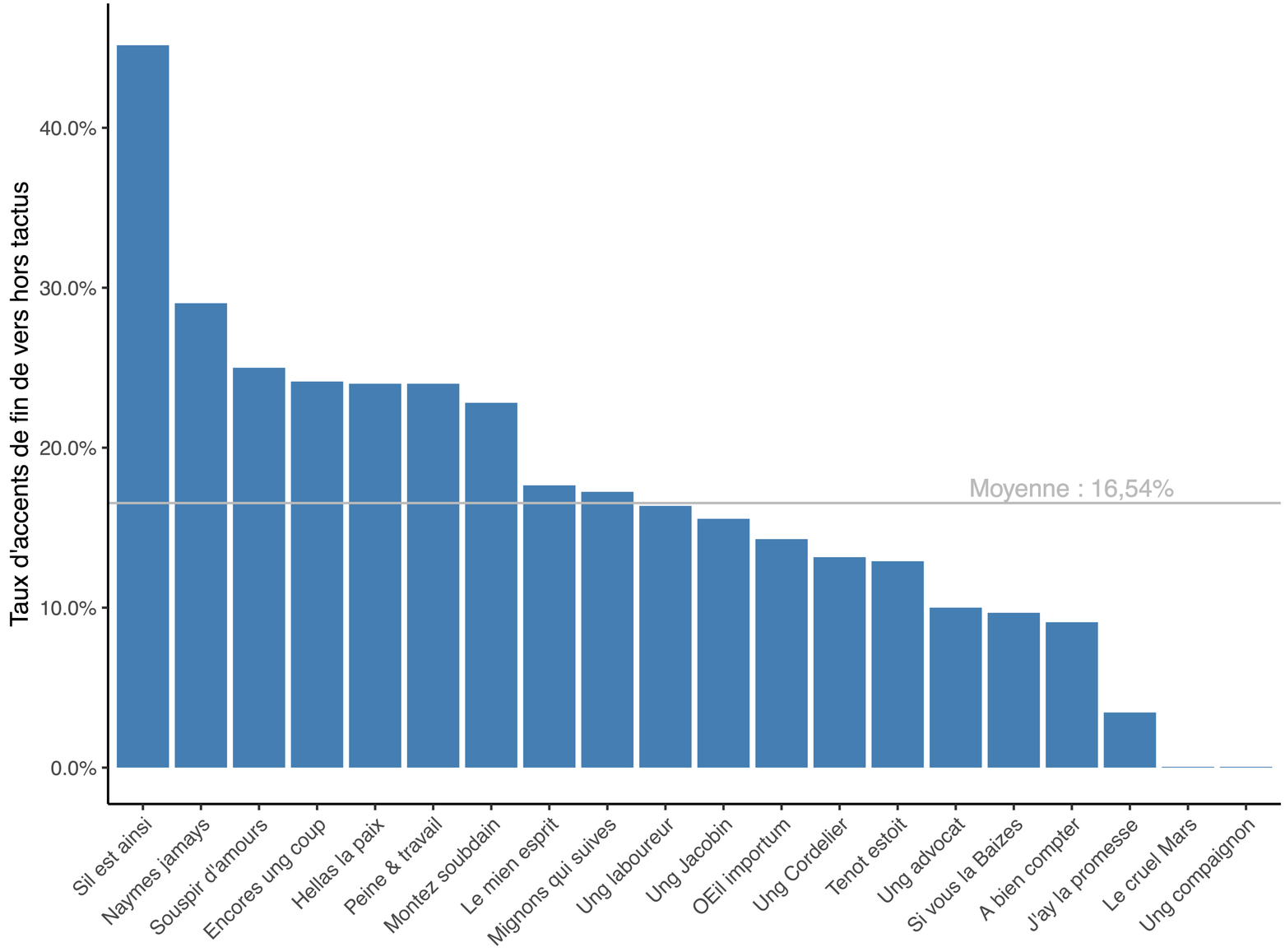


(Texte : *Cristal et Clarie*, éd. Lydie Louison)

Musication

Représenter la variation entres les voix

- Henry Fresneau : compositeur du XVIe s.
- Polyphonie à 4 voix
- Objectif : mesurer le taux d'alignement entre accents de fin de vers et temps forts musicaux
- En distinguant voix par voix
- En en fonction de différents facteurs

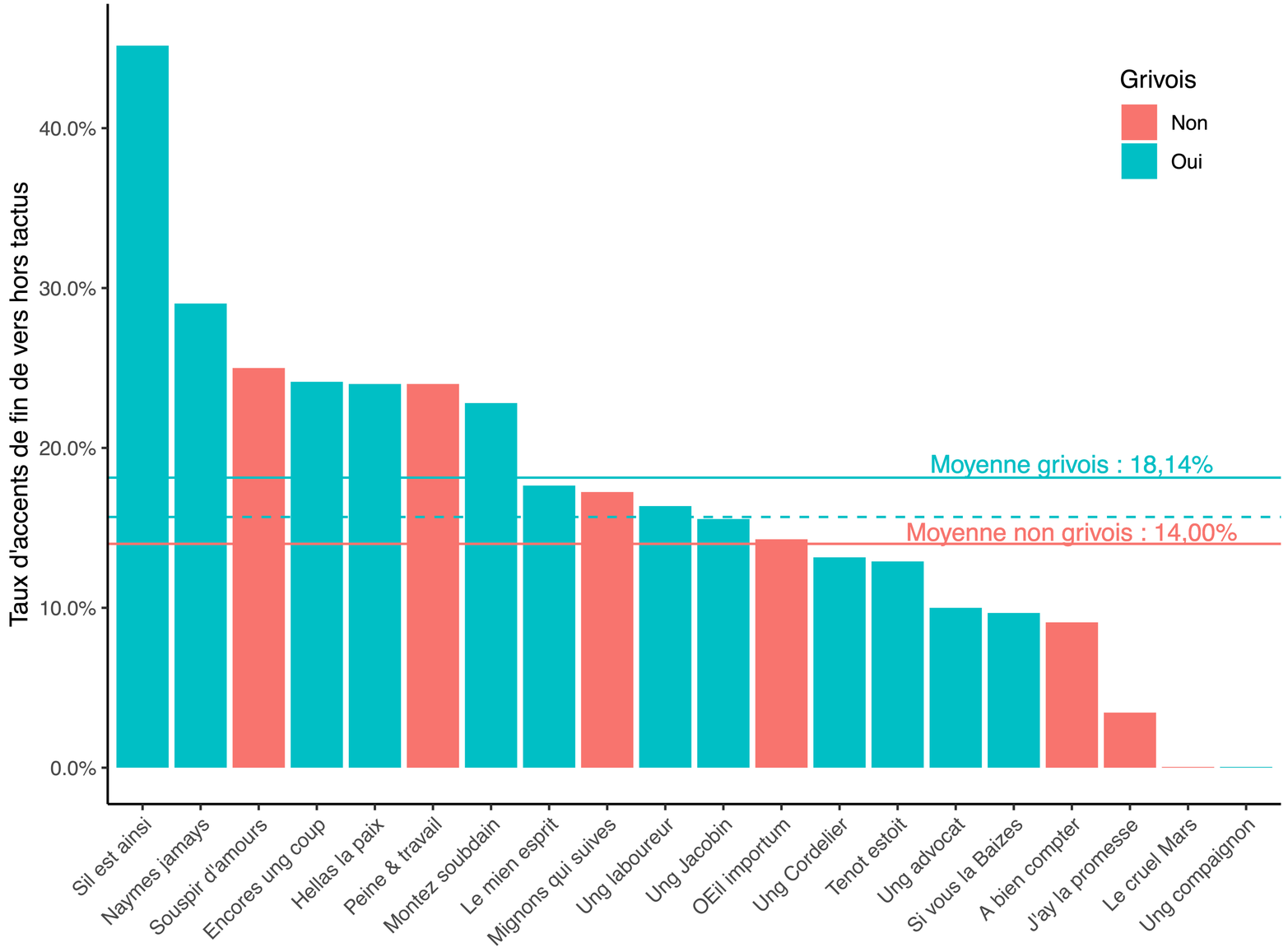


(Source : Chouvion, Premat & Verner)

Musication

Représenter la variation entres les voix

- Facteur :
  - Caractère grivois ou non du texte
  - Ligne pointillée : chansons grivoises sauf *Sil est ainsi*
  - Conclusion : le registre ne change pas significativement le taux d'alignements

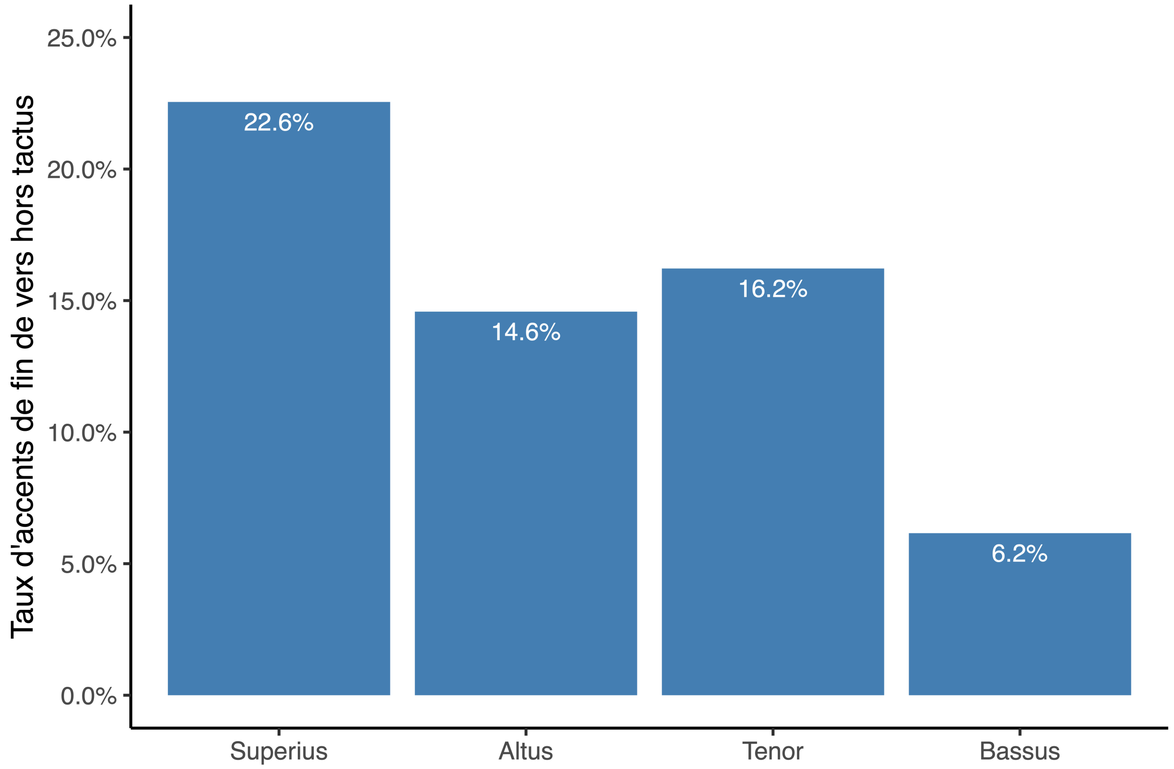


(Source : Chouvion, Premat & Verner)

Musication

Représenter la variation entres les voix

- Facteur :
  - Voix
- Conclusion :
  - La voix de basse a bien moins d’accents de vers non alignés avec des temps forts
  - Mais est-ce juste une moyenne, ou est-ce vrai dans chaque chanson ?



(Source : Chouvion, Premat & Verner)

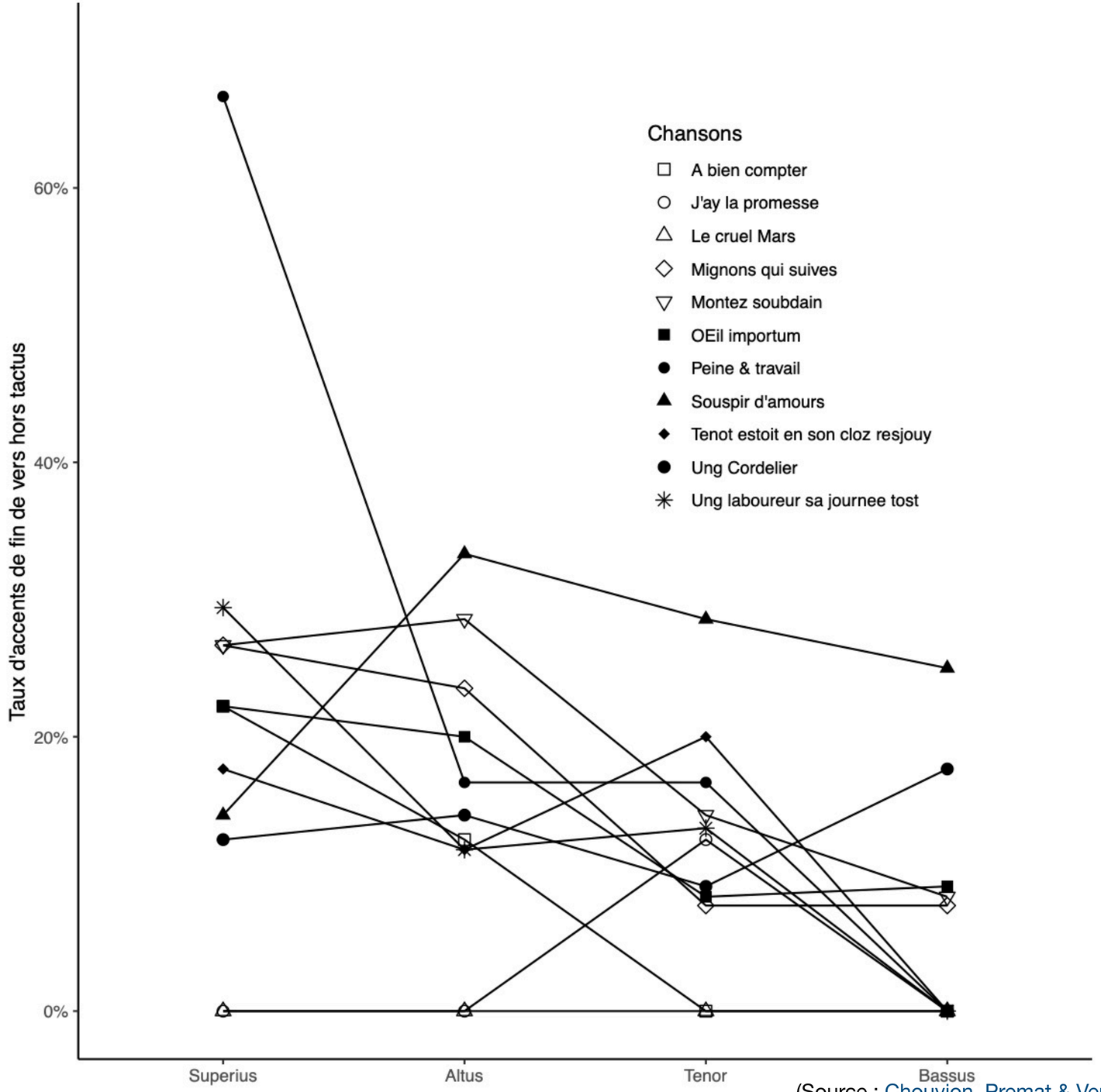
Musication

Représenter la variation entres les voix

Sur 11 chansons

- Oui :
  - 6 chansons
    % bassus < % tenor
  - 4 chansons
    % bassus < % autres voix
- Non :
  - 1 chansons
    % bassus > % tenor
  - 4 chansons
    % bassus ≈ % tenor

Donc **non**, la basse n'a pas nécessairement moins d'accents de vers non alignés avec des temps forts que les autres voies !



(Source : Chouvion, Premat & Verner)

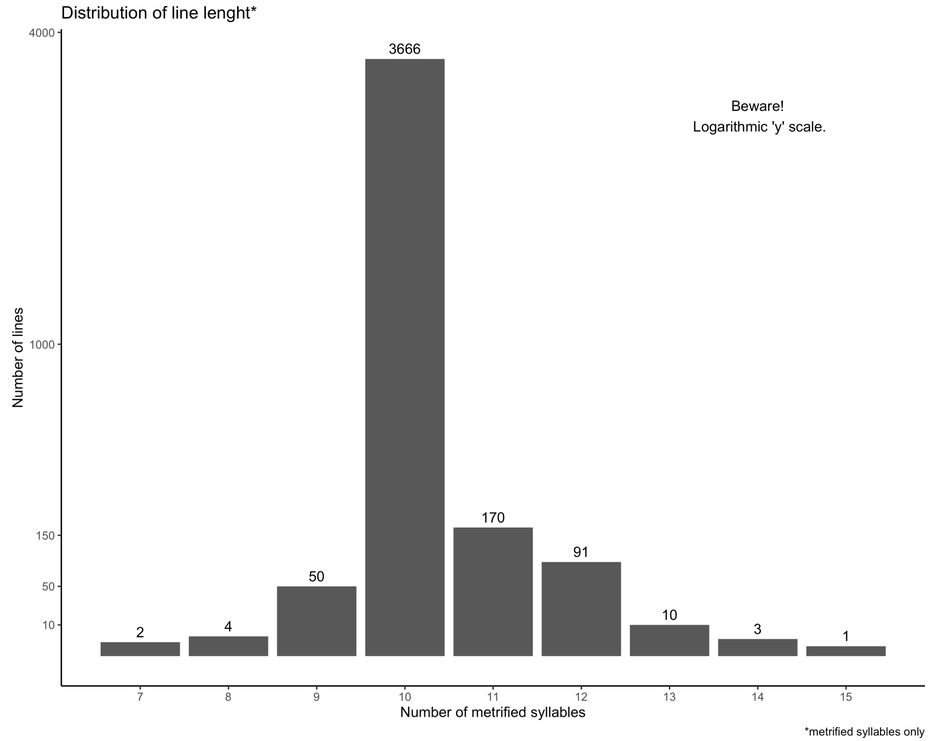
1 *seul* script

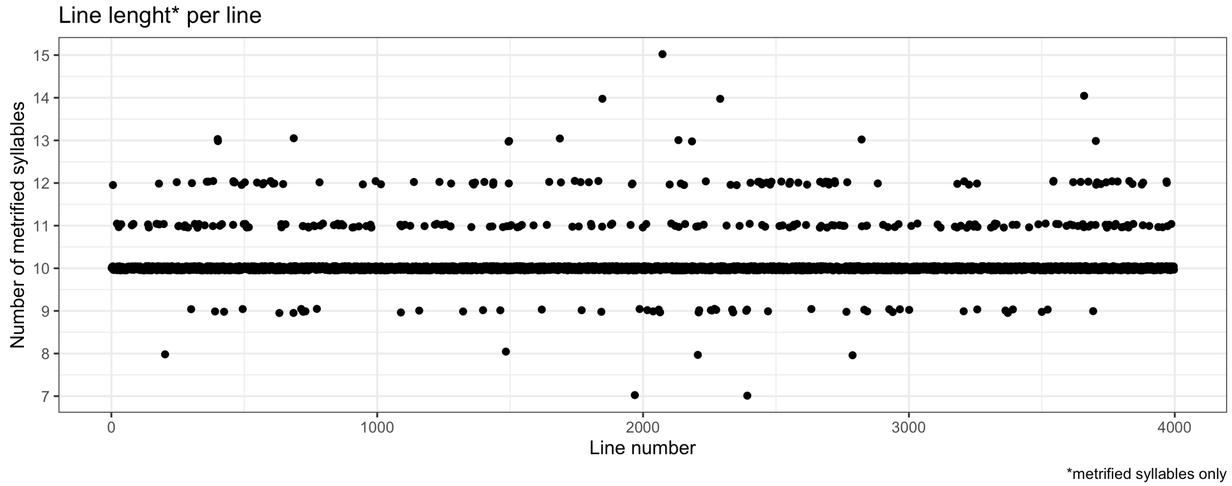
Économie du *workflow*

PAM

Un script pour toutes les données

- Le PAM produit un fichier par texte
- Corpus :
  - des dizaines de textes
  - plusieurs graphes à produire pour chaque texte
- Un seul script suffit !





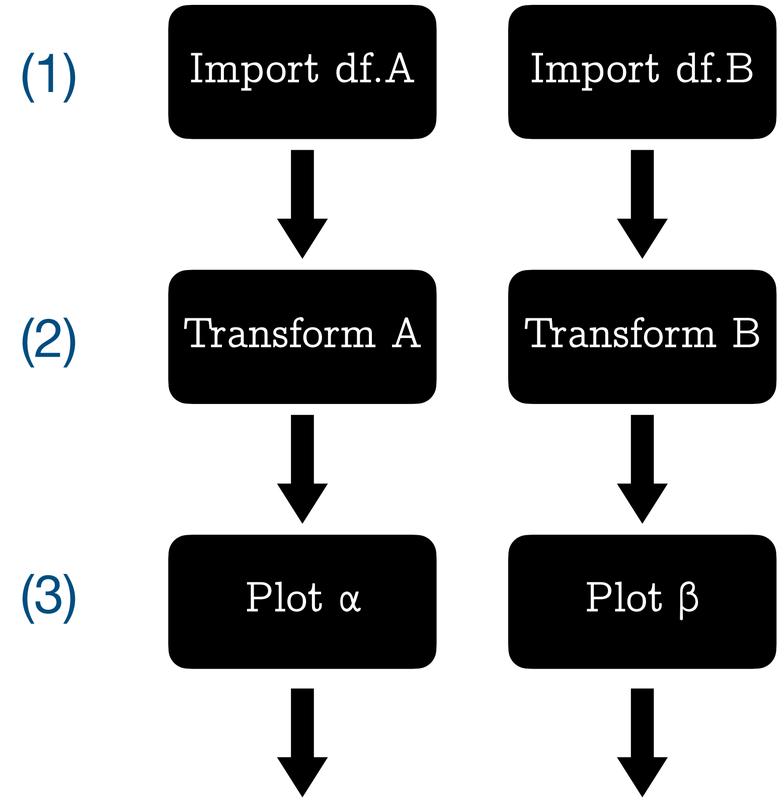
(Texte : *Chanson de Roland*)

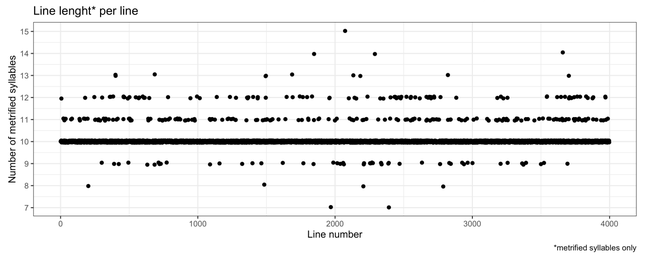
PAM

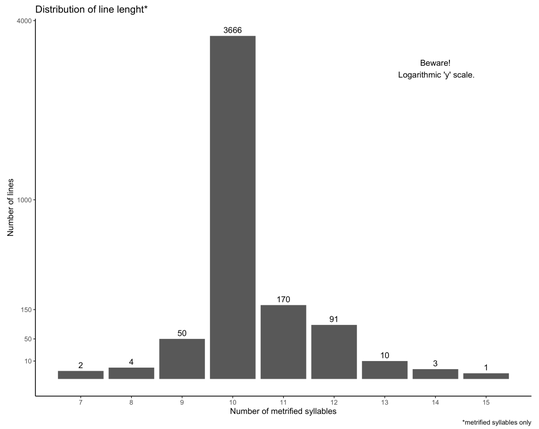
Un script pour toutes les données

Plusieurs graphiques pour un texte

- Dans *un seul script* :
  - importer toutes les données nécessaires (1)
  - appliquer toutes les transformations nécessaires (2)
  - produire toutes les représentations voulues (3)





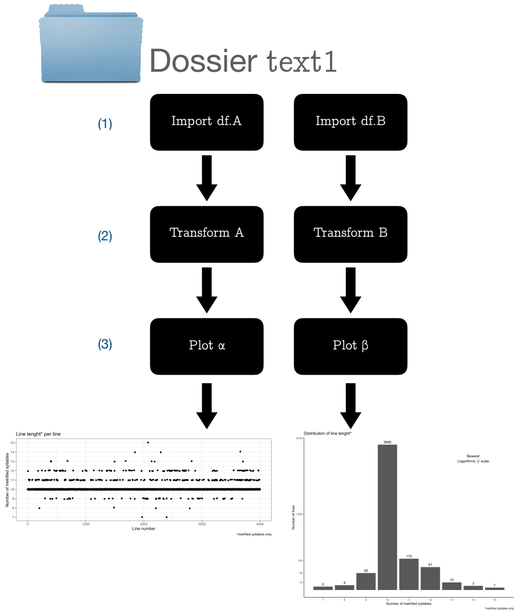


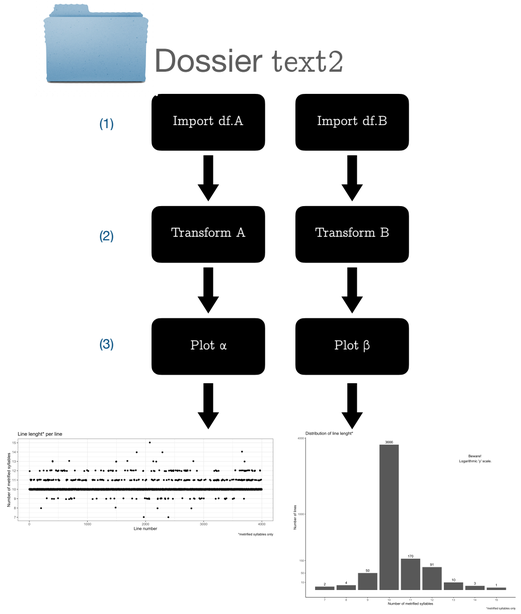
PAM

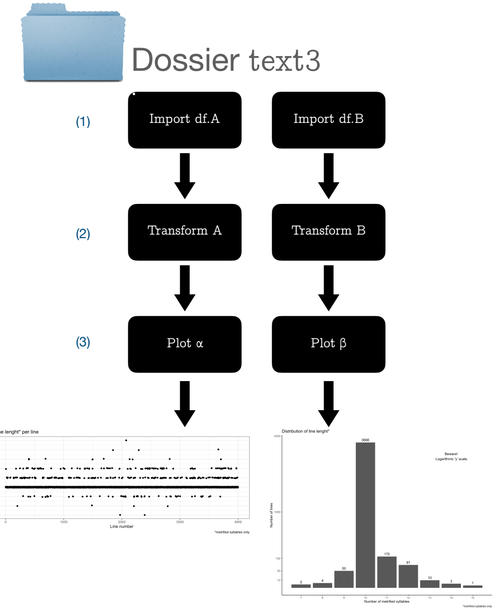
Un script pour toutes les données

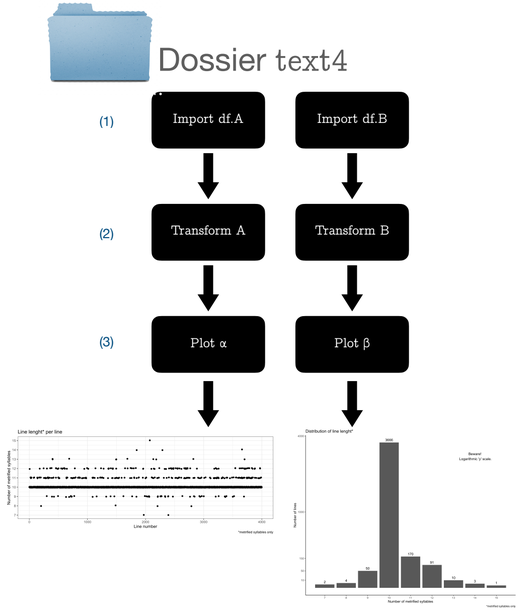
Plusieurs graphiques pour plusieurs textes

- Créer un répertoire par texte
- Placer le script dans chaque répertoire
- Placer les fichiers de données de chaque texte dans son répertoire
  - /!\ Ils doivent avoir toujours le même nom (p.ex. pas d'ID du texte dans le nom)
- Et lancer le script à chaque fois









PAM

Un script pour toutes les données

Automatisation du processus :

- pas d’erreurs de manipulations
- pas de centaines de manipulations répétitives
- très rapide !

Conclusion

Représenter des données linguistiques avec R

- Représenter la variation de données permet
  - de ne pas se laisser tromper par leur apparence superficielle
  - de ne pas être perdu dans leur détail fin
- R est un bon outil pour faire cela
  - Cartes, graphiques
  - Beaucoup de ressources en ligne, une forte communauté
  - Il permet une automatisation du processus ⇾ énorme gain de temps
- Face à des données difficiles à représenter, il faut être inventif, et R permet cela la plupart du temps

Merci de votre attention !

Chanson de Roland, édité par Gérard Moignet, Paris, Bordas, 1972. Publié en ligne par la Base de français médiéval, http://catalog.bfm-corpus.org/roland. Dernière révision le 2014-02-24.

ANDERSON, Stephen R. (1982). « The Analysis of French Shwa: Or, how to Get Something for Nothing », in : *Language* 3(58), pp. 534-573.

CHOUVION, Sophie, Timothée PREMAT & Axelle VERNER (éval. en cours). "La musication dans les chansons d'Henry Fresneau, Grammaire d'association, paraphonologie et forme polyphonique au XVIe siècle", in : *Textus & Musica*.

GUILLIÉRON, Jules & Edmond EDMONT (1902-1910). *Atlas Linguistique de la France*, Paris : Champion.

GUILLOT-BARBANCE, Céline, Serge HEIDEN & Alexei LAVRENTIEV (2017). "Base de français médiéval : une base de référence de sources médiévales ouverte et libre au service de la communauté scientifique". In : *Diachroniques* 7, pp. 168-184.

POGGIO, Enzo & Timothée PREMAT (2019). "Le PAM, un Programme d'Analyse Métrique pour le français médiéval », in : *Actes des Rencontres lyonnaises des jeunes chercheurs en linguistique historique*, sous la dir. de Timothée Premat & Ariane Pinche, Lyon : Diachronies contemporaines, pp. 59-70.

PREMAT, Timothée & Philippe BOULA DE MAREÜIL (2019), "The "rolled" /R/ in French and some regional languages of France", présentation à *R-atics 6*, Paris, 07/11/2019.

RUSSO, Michela & Timothée PREMAT (accepté ; à paraître). "Voyelles finales et traits-ϕ à la rencontre des diasystèmes d'oïl, d'oc et du francoprovençal", in : *Verbum*, "Modélisation diasystémique et typologie", numéro sous la dir. de Jean Léo Léonard.

STEIN, Achim, Pierre KUNSTMANN & Martin-D. GLEßGEN (2006). "Nouveau Corpus d'Amsterdam. Corpus informatique de textes littéraires d'ancien français (ca 1150-1350)". Établi par Anthonij Dees (Amsterdam : 1987), remanié par Achim Stein, Pierre Kunstmann et Martin-D. Gleßgen, Stuttgart : Institut für Linguistik/Romanistik.

Plus de scripts et d'exemples sur mon site :
https://sites.google.com/view/timothee-premat/programmes

Diapo : ⟨halshs-03116096⟩