#### Journées PFC-FLoRAL, 17-19 novembre 2016

# PFC-Québec: État des lieux et perspectives

Marie-Hélène Côté<sup>1,2</sup> • Hugo Saint-Amant Lamy<sup>1,2</sup> • Peter Milne<sup>3</sup>

<sup>1</sup>Université de Lausanne, <sup>2</sup>Université Laval, <sup>3</sup>McGill University



Social Sciences and Humanities Research Council of Canada Conseil de recherches en sciences humaines du Canada



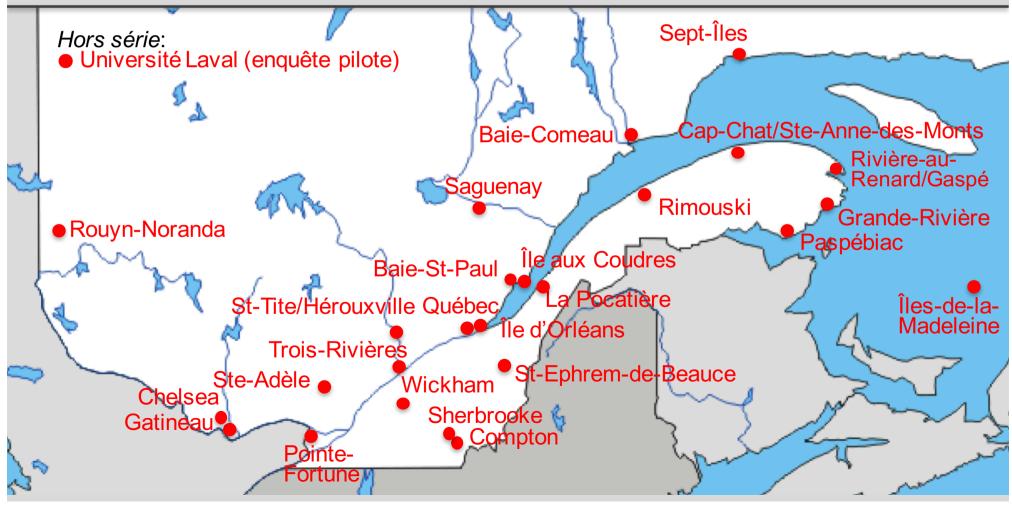
### **PLAN**

Situation des enquêtes (Marie-Hélène)

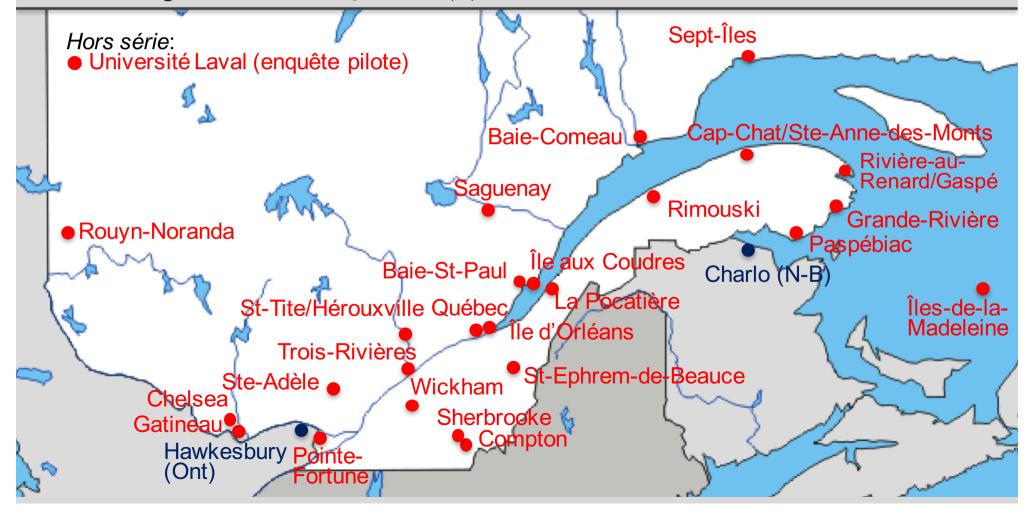
Traitement des données (Hugo)

Complétées au Québec (26)



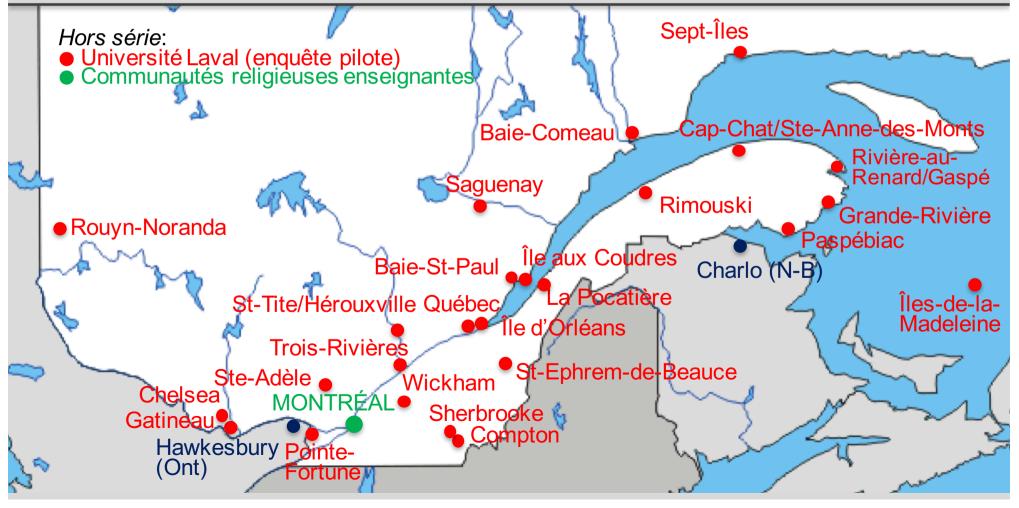


- Complétées au Québec (26)
- Complétées hors-Québec (2)



- Complétées au Québec (26)
- Complétées hors-Québec (2)
- En cours (2)





- Complétées au Québec (26)
- Complétées hors-Québec (2)

- En cours (2)
- Prévues (6)

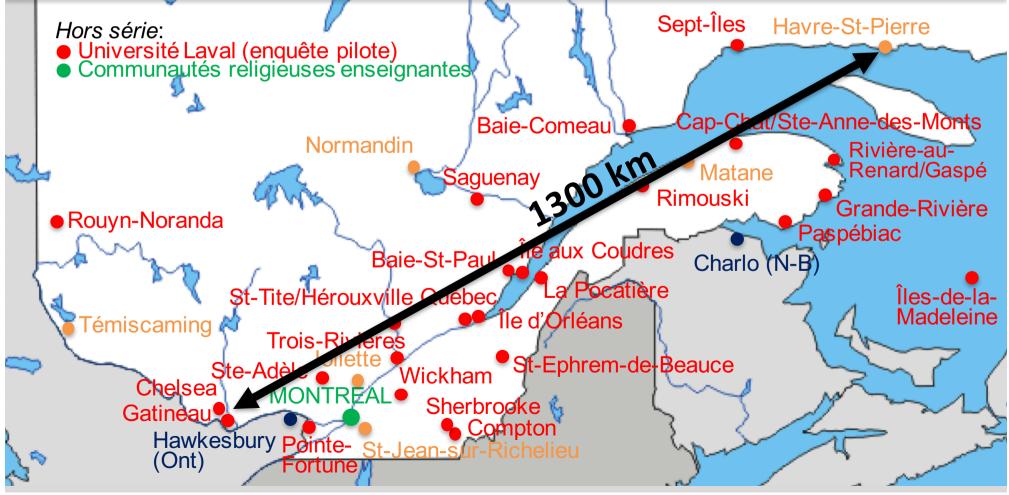




- Complétées au Québec (26)
- Complétées hors-Québec (2)

- En cours (2)
- Prévues (6)





# Merci aux enquêteurs.trices!

- Hugo Saint-Amant Lamy (coc, cqe, cqf, cqg, cqi, cqk, cqz + formation des autres)
- Myriam Paquet-Gauthier (cql, cqn, cqp, cqq, cqr, cqx, cqy + 2 à venir)
- Julie-Anne Veilleux (cqt, cqu, cqw, cry)
- Marie-Pier Picard (cqc16, cry + à venir)
- Nicolas Royer-Artuso (cnb, cqo, cqv)
- Marie-Claude Séguin (cqc, cqd, cqe)
- Karine Groulx (coc, crx)
- Julie Bilodeau (cqs)
- Caroline Sigouin (cqm)
- Élizabeth Morin-Lessard (cqj)
- Jessica Martineau (cqh)

- Marie-Claude Tremblay (cqd)
- Anne-Sophie Bergeron (cqu)
- Rémi Belleau (cqd)
- Luc Baronian et ses étudiants (cqb)
- Jacques Durand (cqa)

Merci à la Faculté des arts de l'Université d'Ottawa pour son soutien financier. Merci aux nombreux transcripteurs!



Social Sciences and Humanities Research Council of Canada

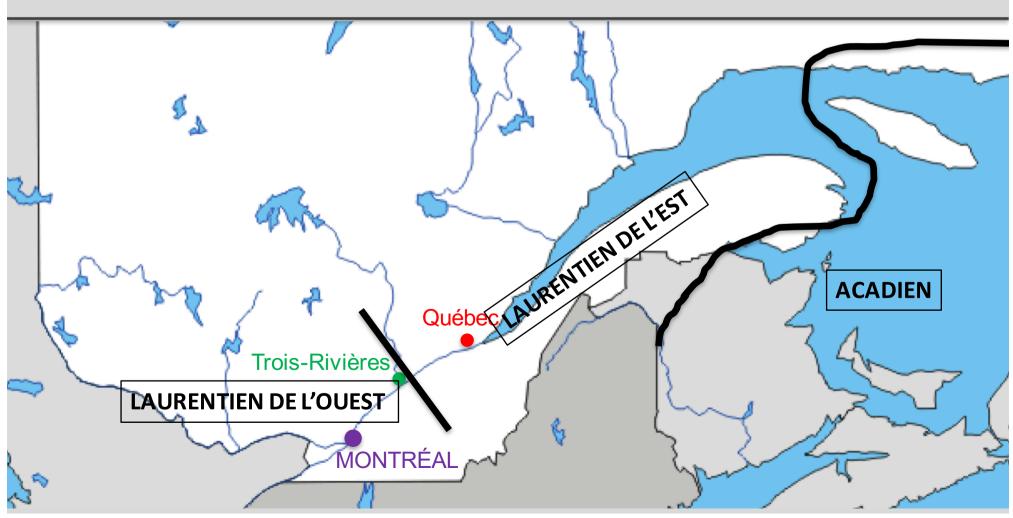
Conseil de recherches en sciences humaines du Canada



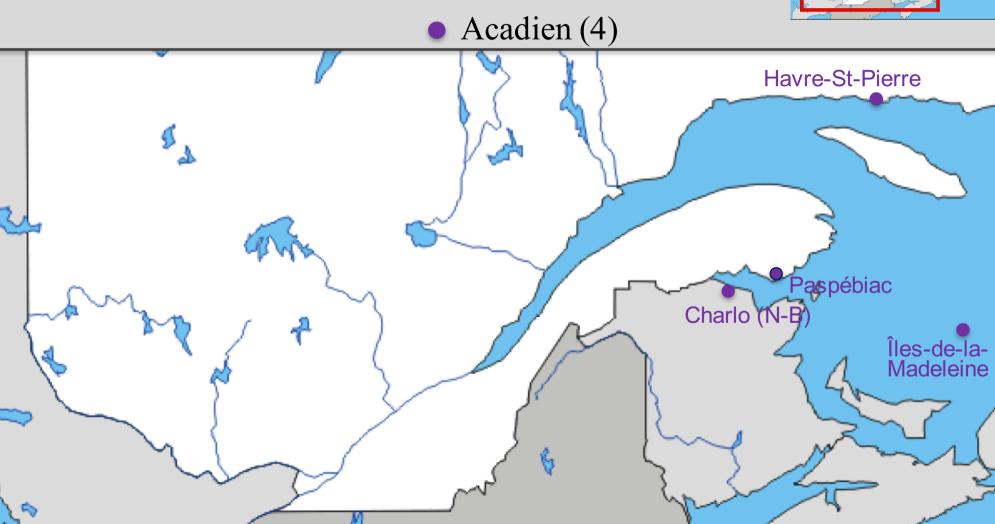




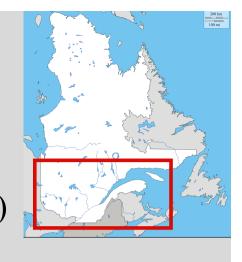


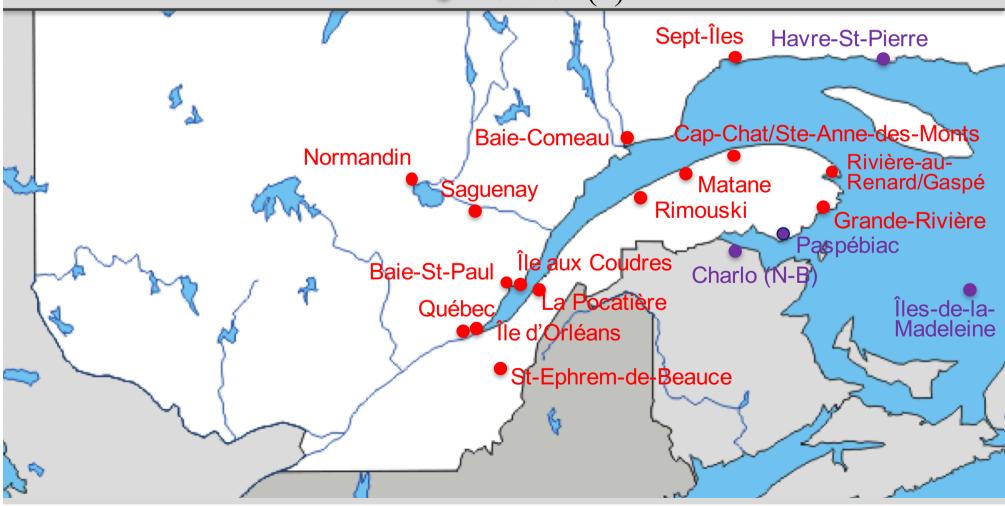


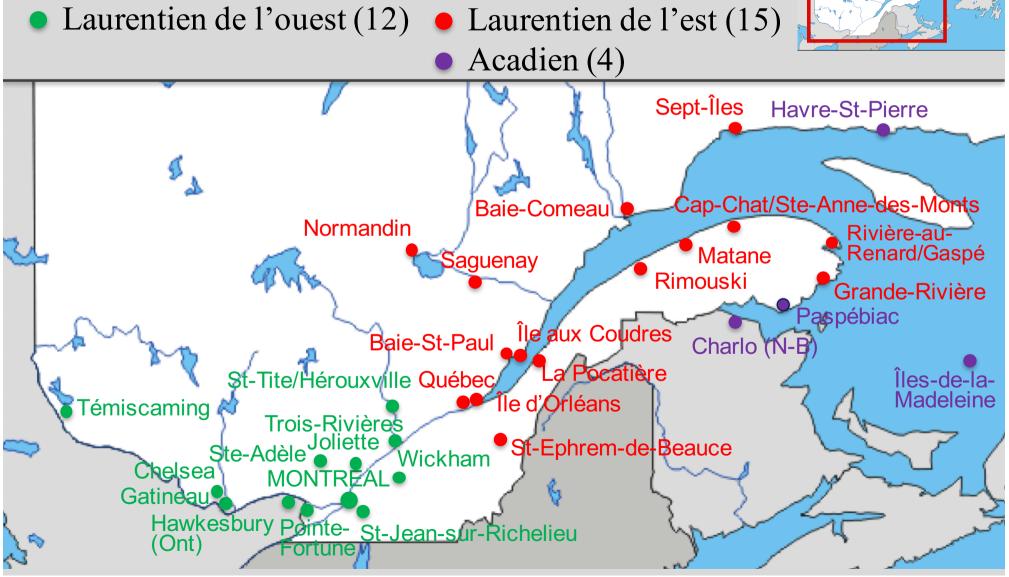




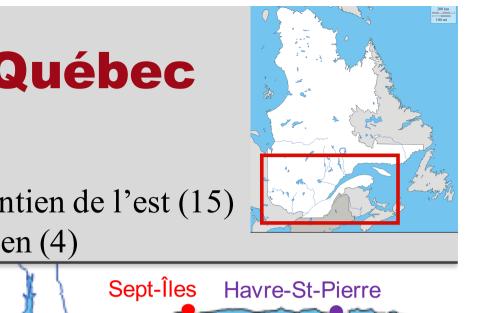
- Laurentien de l'est (15)
- Acadien (4)





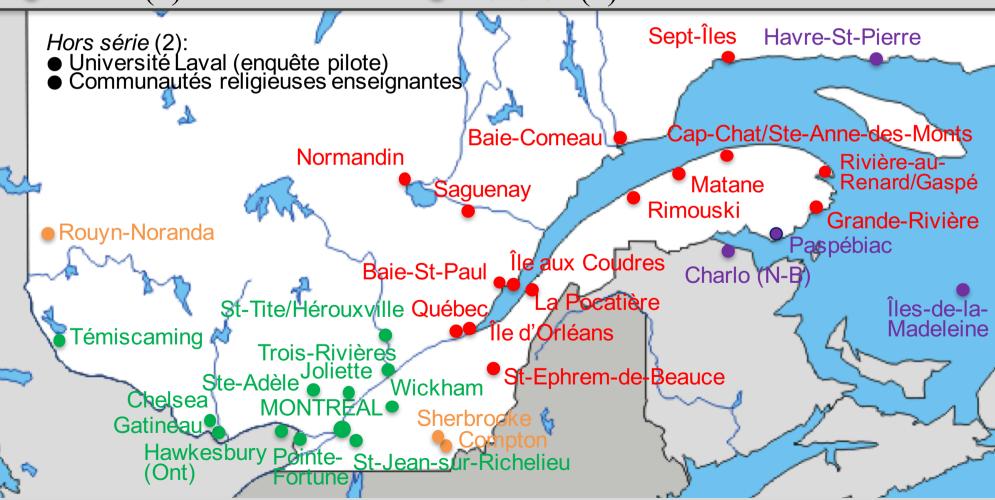


- Laurentien de l'ouest (12)
   Laurentien de l'est (15)
- Mixte (3)Acadien (4)





- Laurentien de l'ouest (12)
   Laurentien de l'est (15)
- Mixte (3)Acadien (4)



	Laurentien ouest	Laurentien est	Mixte	Acadien	Hors- série	TOTAL
Points actuels	9	13	3			30
Points prévus	12	15	3	4	2	36
Locuteurs actuels	121	177	42	39	15	394
Locuteurs prévus	157	201	42	51	21	472
Locuteurs transcrits	84	105	13	13	9	223

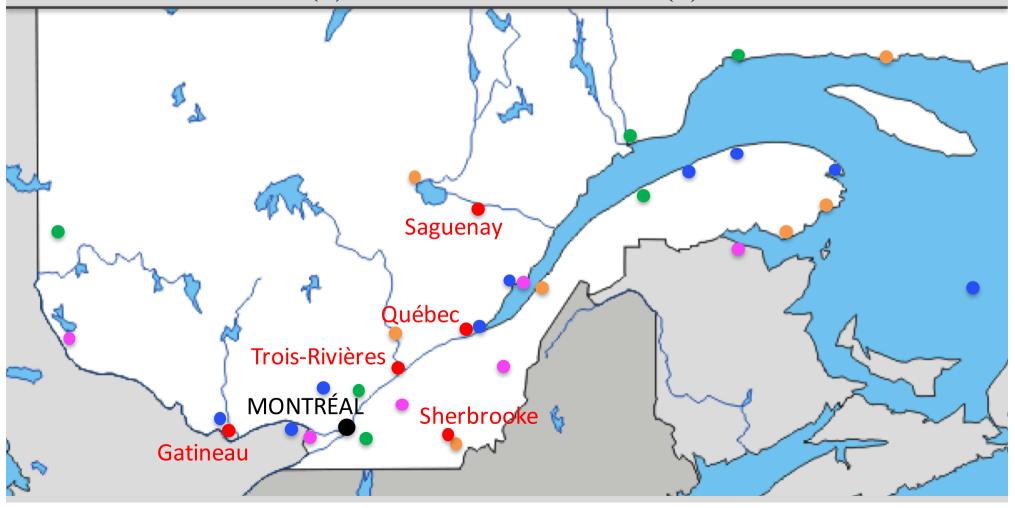
Hommes	Femmes	TOTAL ACTUEL
183	211	394

1910	1920	1930	1940	1950	1960	1970	1980	1990	TOTAL ACTUEL
1	10	43	49	70	40	48	66	67	394

- Montréal (~4M)
- Autres villes +120 000 (5)
- 20 000-100 000 (6)

- **5** 000-18 000 (9)
- **9** 3 000-5 000 (7)
- **•** 500-2 500 (6)





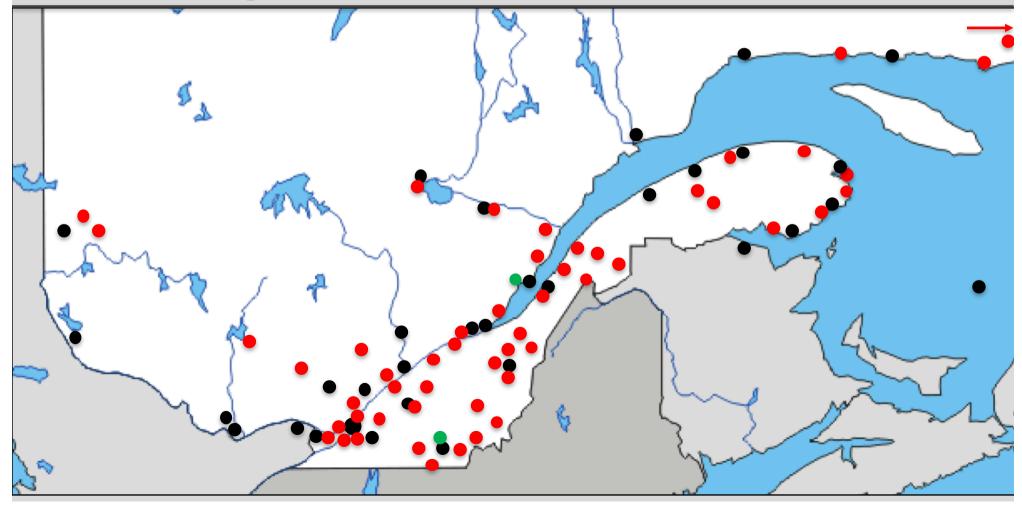
### **ALEC**

Dulong, Gaston & Gaston Bergeron. 1980. Le parler populaire du Québec et de ses régions voisines: atlas linguistique de l'Est du Canada, vol. 1-10. Québec: Gouvernement du Québec, Ministère des communications / Office de la langue française.

- Enquête lexicologique menée entre 1969 et 1973: « 2500 concepts (ou questions) portant sur la vie traditionnelle et matérielle » (vol. 1: 2).
- Locuteurs âgés (moyenne d'âge: 73) dans 169 points d'enquête en milieu rural.
- Transcriptions phonétiques d'un intérêt certain mais limité (Larochelle 1989; Morin 1996; Friesner 2010)
- Enregistrements:
  - « A la fin de l'enquête, on réalise avec un témoin principal l'enregistrement d'une entrevue d'une vingtaine de minutes de façon à conserver un échantillon sonore de la prononciation de cet endroit » (vol. 1: 25)
  - Ces enregistrements n'avaient jamais été exploités (on ignorait leur existence ou ils étaient crus perdus ou inexistants)
  - Seuls des échantillons de 59 points d'enquête ont pu être retrouvés (aux Archives de folklore et d'ethnologie de l'Université Laval) et associés à un point d'enquête particulier; quelques enregistrements ont été retrouvés, mais n'ont pu être localisés.
  - Enregistrements en cours de transcription, selon le protocole PFC (conversations guidées).
  - Ils nous permettent de reculer d'une génération dans la description du français québécois, appréhendé dans sa diversité géographique.

- Points d'enquête PFC (34)
- Enregistrements de l'ALEC (57)
- Points d'enquête PFC = ALEC (2)

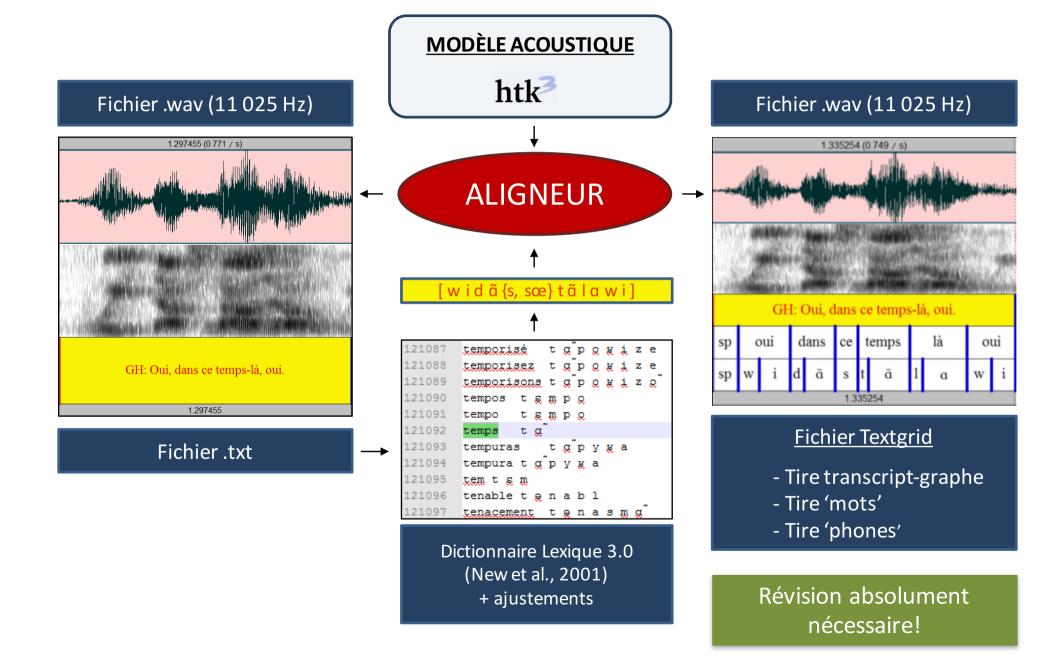




### Traitement des données

- Transcription phonétique
  - Aligneur automatique
  - Transcription des listes (liste PFC, liste laurentienne)
  - Transcription des conversations
- Codage assisté
  - Schwa
  - Liaison
- Système multi-tires

### Aligneur automatique (Milne, 2012)



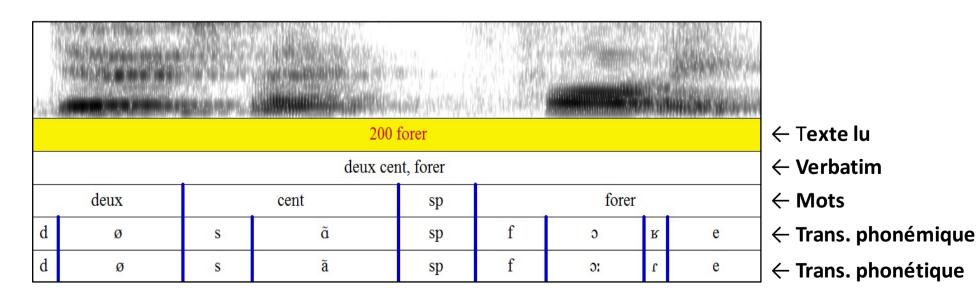
# Transcription des listes

#### Tire phonémique

- Transcription du dictionnaire de l'aligneur
- Comparaisons & recherches

### Tire phonétique

- Transcription des réalisations réelles par assistants
- Description de la variation



# Transcription des listes

### Application dialectométrique

- Gabmap (Nerbonne et al., 2011)
- Mesure distance linguistique & comparaison avec données géographiques
- Identification de zones dialectales

#### **Chaînes segmentales**

		ba	ileii	ne	connaissent					
Québec	b	а	ı	3	n	k	Э	n	3	S
St-Tite	b	а	I	3	n	k	Э	n	3	S
Ste-Adèle	b	а	I	3	n	k	Э	n	ε	S

#### Valeurs numériques

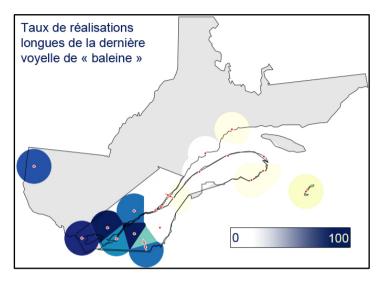
(% [3])	baleine	connaissent			
Québec	0 (z=-2.27)	100 (z=0.22)			
St-Tite	67 (z=-0.37)	75 (z=-0.22)			
Ste-Adèle	93 (z=0.37)	0 (-1.55)			

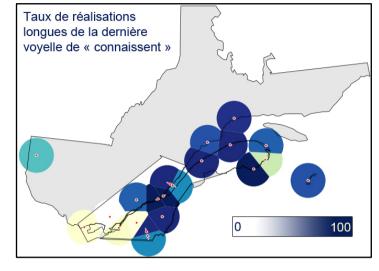
ш	
C	
Ž	
4	U
⋖	Ĭ
$\vdash$	_
Ċ	

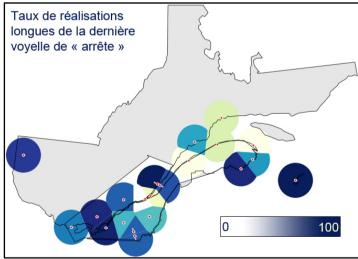
	Québec	St-Tite	Ste-Adèle
Québec	0	0.5	1
St-Tite	0.5	0	0.5
Ste-Adèle	1	0.5	0

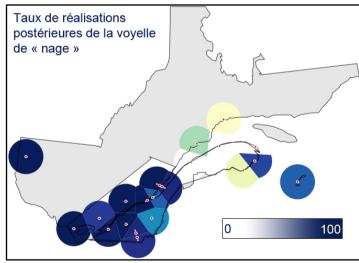
	Québec	St-Tite	Ste-Adèle
Québec	0	1.17	2.20
St-Tite	1.17	0	1.04
Ste-Adèle	2.20	1.04	0

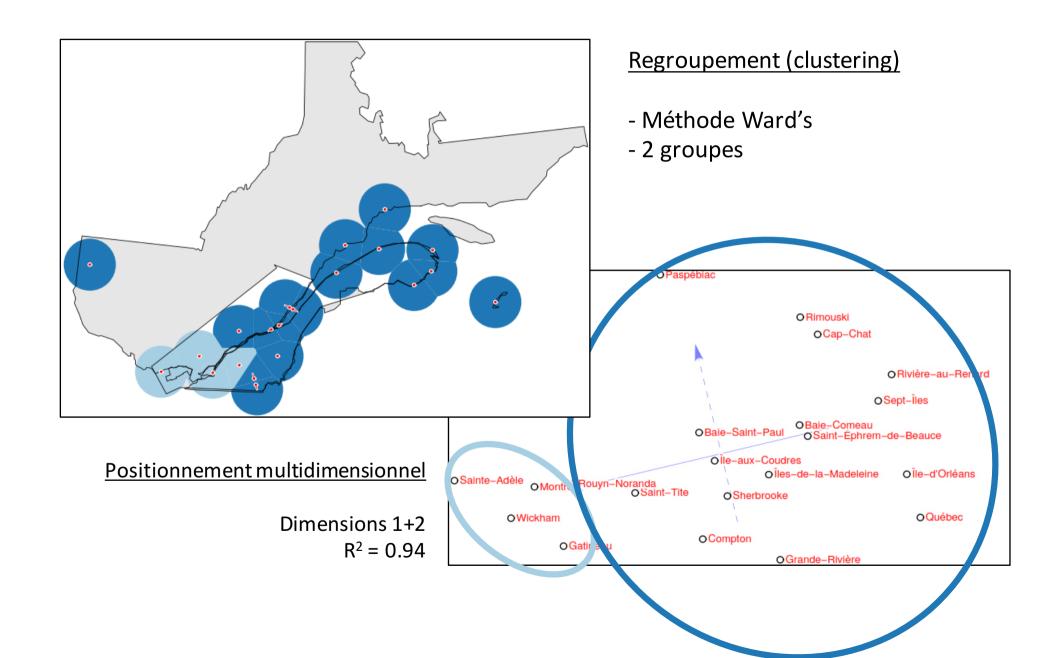
#### Ex.:

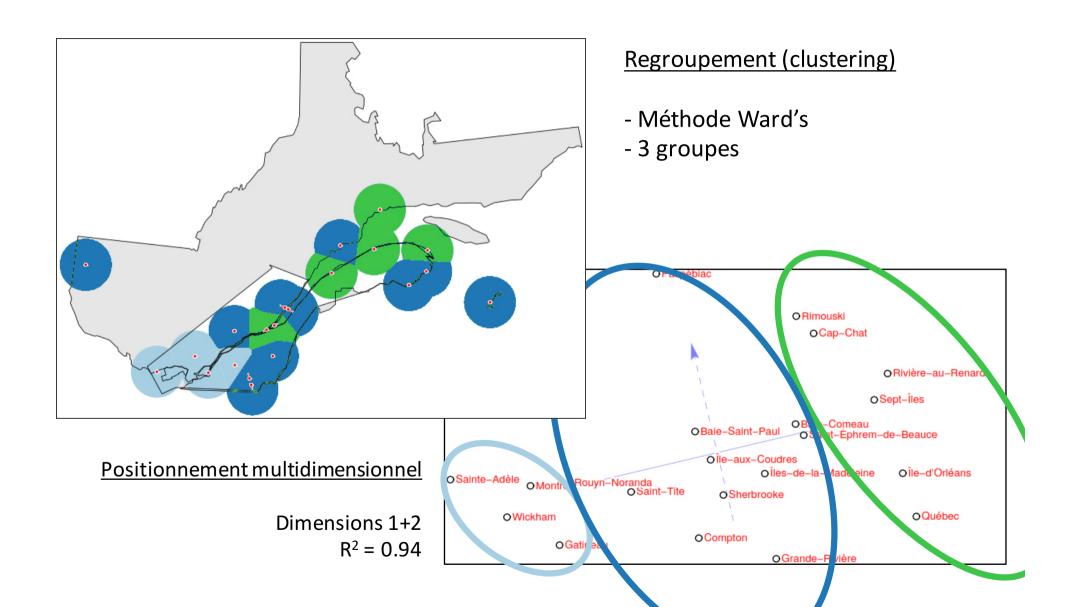


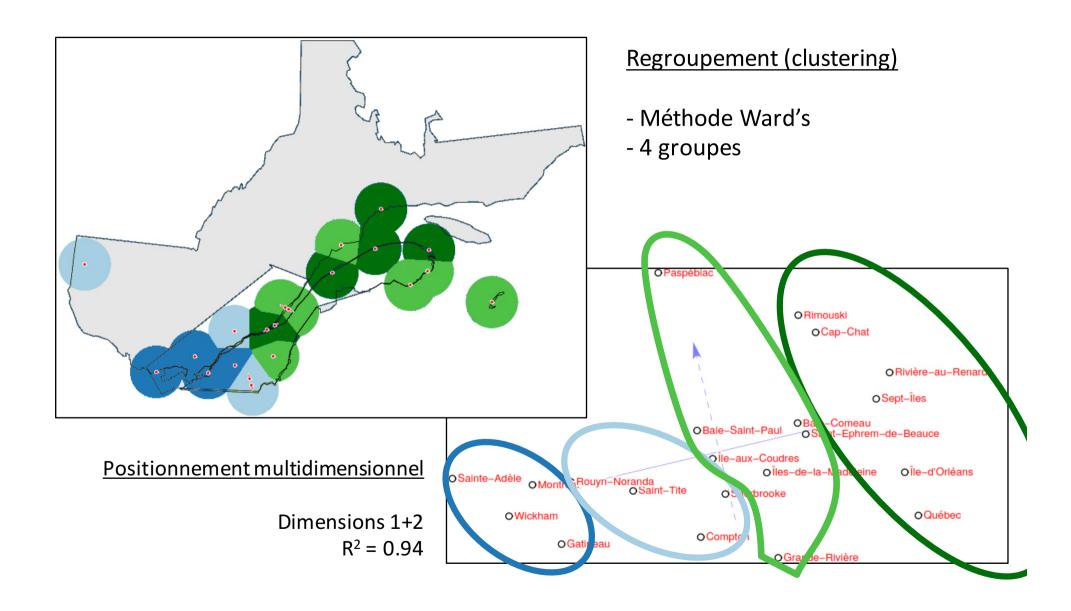


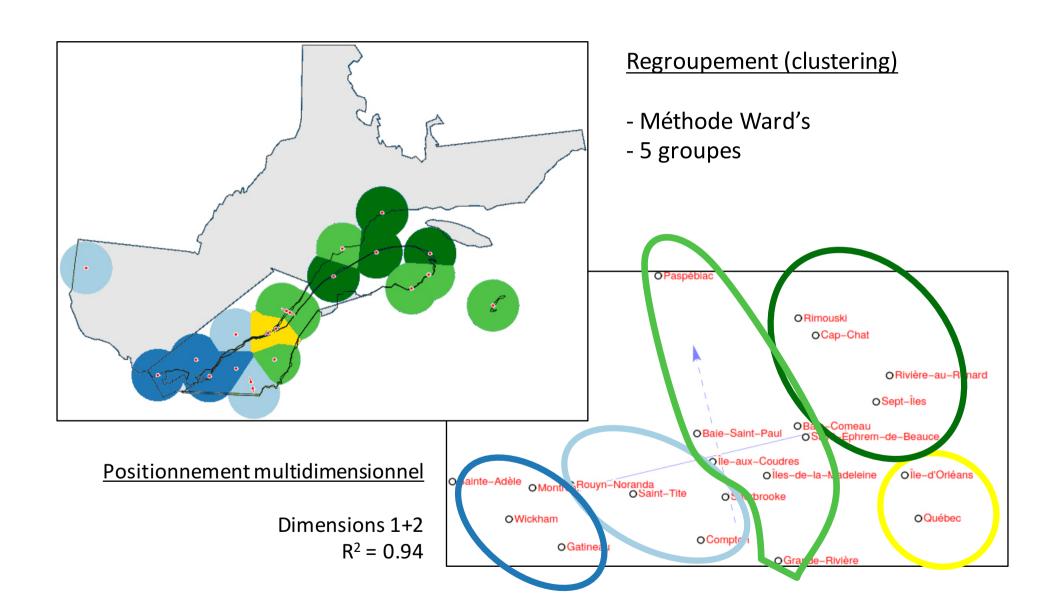


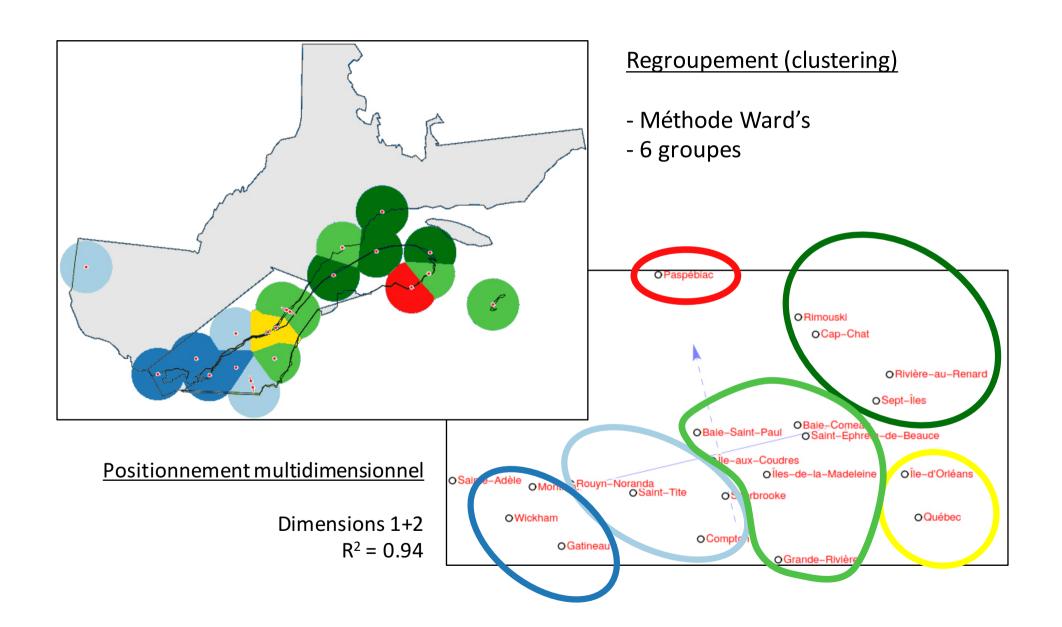












# Transcription des conversations

- Transcription de 5 minutes par locuteur
  - Seulement interventions du locuteur
  - Seulement séquences alignables (sans bruits, chevauchement, etc.)
  - Transcription phonémique (output de l'aligneur)
  - Transcription phonétique (manuelle)

																	(90)			
			(	GH:	Je p	eux b	ara	agoi	iine	r e	n ital	ien	un	peu.	<e< td=""><td>Q1</td><td>:En</td><td>ita</td><td>alie</td><td>n, oui?&gt;</td></e<>	Q1	:En	ita	alie	n, oui?>
			·			·	Je	peu	x b	ara	goui	ner	en	italie	n u	n p	eu.			
				Je	peu	ıx bara	igo	uin	er e	n it	alien	l								(chevauchement)
sp	je	p	eux			bara	igo	uin	er			en			ita	alie	1			
sp	3	p	ø	b	a	a	g	w	i	n	e	ã	n	i	t	a	1	j	ĩ	
sp	ſ	p	ø	b	a	a	g	w	i	n	e	ã	n	i	t	a	1	j	ẽ	

- ← Trans. orthographique
- ← Seulement locuteur
- ← Séquences alignables
- $\leftarrow$  Mots
- ← Trans. phonémique
- ← Trans. phonétique

## Transcription des conversation

### Application

- Traitement automatique des données alignées
- Portrait acoustique du français laurentien (timbre des voyelles, durée, etc.)

i		42 000	Y	14 000	u	14 000
е		61 000	ø	21 000	0	10 000
ε/3	3	54 000 8 000	œ	34 000	Э	22 000
а		58 000			a	11 000
ẽ	1	0 000	œ	3 000	õ	16 000
					ã	25 000

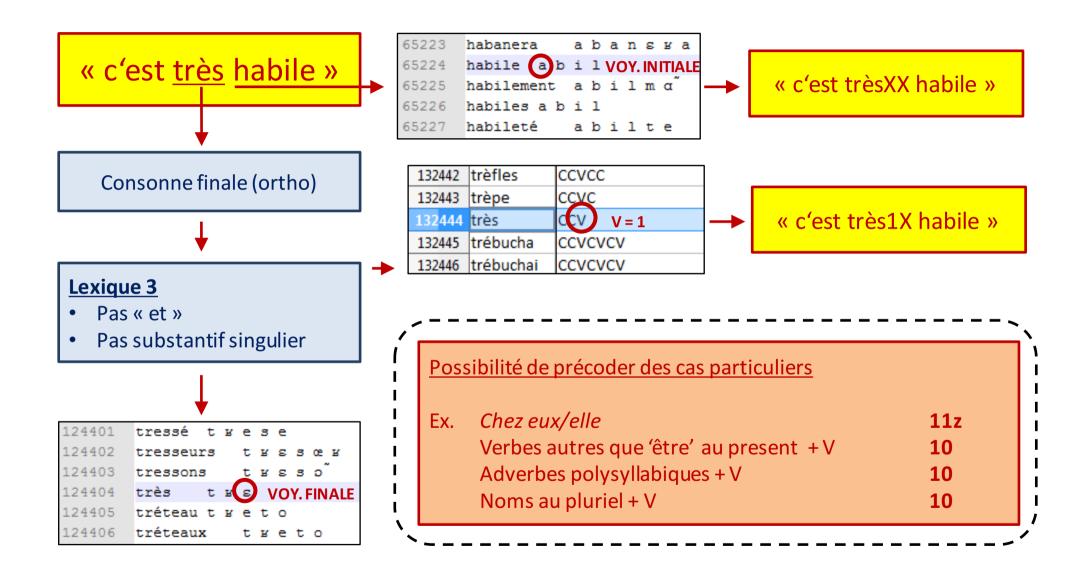
	~1965 phones / locuteur (5min) X 394 locuteur									
<u>Total: 774 000 phones</u>										
р	28 000	t	35 000	k	30 000					
b	10 000	d	30 000	g	2 000					
f	11 000	S	48 000	ſ	4 000					
V	17 000	Z	3 000	3	22 000					
m	31 000	n	12 000	ŋ	160					
		I	47 000	R	44 000					

Nb. occurrences attendu

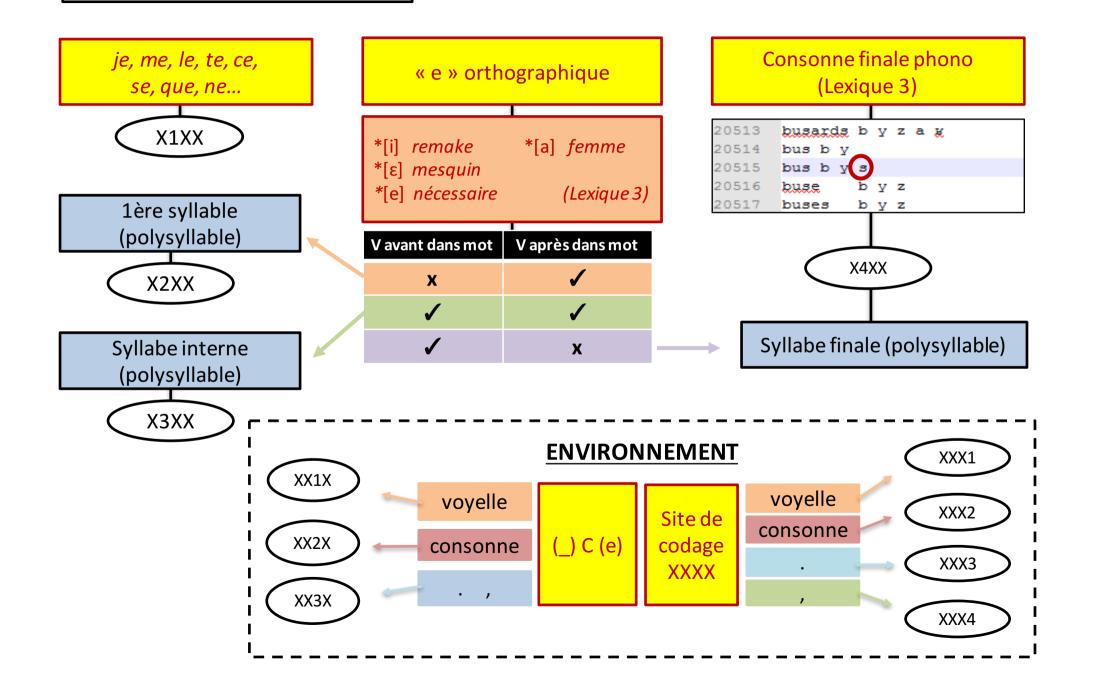
### Assistant codeur

- Identification automatique des sites de codage
  - Dans le textgrid, à partir des transcriptions révisées
  - Basé sur même dictionnaire phonétique que l'alignement automatique
- Facilite et accélère le travail du codeur
- Objectif n'est pas de remplacer le codage manuel
  - Jugement sur présence/absence de liaison ou de schwa, métathèses, liaisons atypiques, etc.

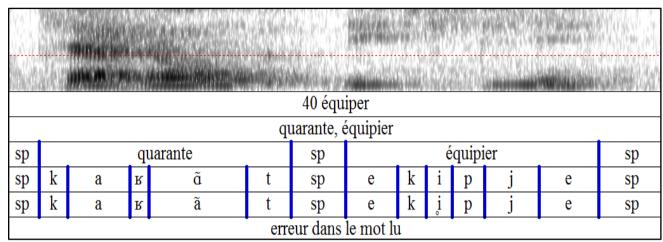
### **CODAGE LIAISON**



### CODAGE SCHWA



#### **Textgrid complet - Listes**



← Texte lu

← Verbatim

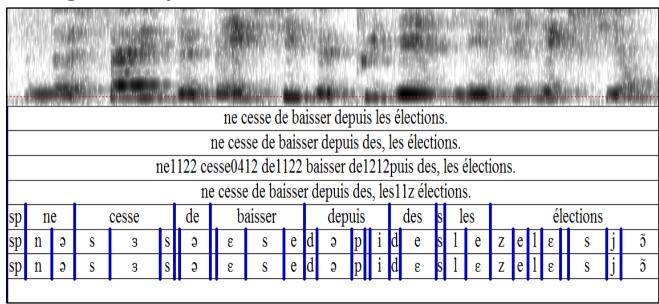
← Mots

← Transcription phonémique

← Transcription phonétique

← Commentaires

#### **Textgrid complet - Texte**



- ← Texte lu
- ← Verbatim
- ← Codage schwa
- ← Codage liaison
- ← Mots
- ← Transcription phonémique
- ← Transcription phonétique
- ← Commentaires

#### **Textgrid complet**

- 1. Transcription cible
- 2. Transcription verbatim
- 3. Transcription anonymisée
- 4. Codage schwa
- 5. Codage liaison
- 6. Transcription locuteur
- 7. Transcription alignable
- 8. Alignement (mots)
- 9. Alignement (phonèmes)
- 11. Transcription phonétique
- 12. Commentaire
- 13. Tire anonymisation (hum)

# Textgrid aligné / anonymisé

- 1. Transcription cible
- 2. Transcription verbatim
- 3. Transcription anonymisée
- 4. Codage schwa
- 5. Codage liaison
- 6. Transcription locuteur
- 7. Transcription alignable
- 8. Alignement (mots)
- 9. Alignement (phonèmes)
- 10. Transcription phonétique

#### **Textgrid PFC**

- 1. Transcription cible
- 2. Transcription verbatim
- 3. Transcription anonymisée
- 4. Codage schwa
- 5. Codage liaison

Listes

Listes + texte

**Texte** 

Listes + conv.

Conversations

Texte + conv.

Listes + texte + conversations

### Références

- Dulong, G., & Bergeron, G. (1980). Le parler populaire du Québec et de ses régions voisines, Atlas linguistique de l'Est du Canada, collection Études et dossiers Québec, 19-80. Québec, Ministère des Communications, Éditeur officiel du Québec, et Office de la langue française.
- Friesner, M. (2010). Une prononciation «tsipéquement» québécoise?: La diffusion de deux aspects stéréotypés du français canadien. *The Canadian Journal of Linguistics/La revue canadienne de linguistique*, 55(1), 27-53.
- Larochelle, P. (1989). Analyse géolinguistique de quelques données phonétiques de l'Atlas linguistique de l'est du Canada. Mémoire de maîtrise, Université de Sherbrooke.
- Milne, P. M. (2001). The effects of syllable position on allophonic variation in Québec French/R. *University of Pennsylvania Working Papers in Linguistics* 18.2:9.
- Morin, Y. C. (1996). The origin and development of the pronunciation of French in Québec. In H. Fé Nielsen & L. Schosler (réds). *The origins and development of emigrant languages*. 243-275.
- Nerbonne J., Rinke Colen J., Gooskens C., Kleiweg P., & Leinonen T. (2011). Gabmap A
  Web Application for Dialectology. *Dialectologia*, Special Issue II.
- New B., Pallier C., Ferrand L. & Matos R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE, L'Année Psychologique, 101, 447-462. http://www.lexique.org
- Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Journal of the Acoustical Society of America*, 123(5), 3878.

### Annexe

- Méthode de regroupement de Ward
  - Méthode de la variance minimale
  - À chaque étape, regroupement des ensembles qui résulte dans la plus petite augmentation de la variance dans le nouvel ensemble.
  - Tend à créer des regroupements de tailles comparables
  - Legendre & Legendre (1998)