



# GLOTTOPOL

Revue de sociolinguistique en ligne  
n°31 – janvier 2019

*Accents du français : approches  
critiques*

Numéro dirigé par Maria Candea, Gaëlle  
Planchenault, Cyril Trimaille

## SOMMAIRE

- Maria Candea, Gaëlle Planchenault, Cyril Trimaille : *Avant-propos et présentation du numéro – l’accent qu’on a, l’accent qu’on nous donne, l’accent qu’on est.*
- Alexei Prikhodkine : *Accents régionaux du français : interroger des évidences.*
- Elissa Pustka, Jean-David Bellonie, Marc Chalier et Luise Jansen : « *C’est toujours l’autre qui a un accent* » : *Le prestige méconnu des accents du Sud, des Antilles et du Québec.*
- Mathieu Avanzi, Philippe Boula de Mareüil : *Peut-on identifier perceptivement huit accents régionaux en français ? La réponse des sciences participatives.*
- Kristin Reinke, Luc Ostiguy, Louis Houle, Caroline Émond : *Cachez cet accent qu’on ne saurait entendre : la langue du doublage fait au Québec.*
- Liudmila Smirnova, Alain Dawson : « *La ch’tite famille* » : *derrière le film à accent local, un chantier de politique linguistique ?*
- Médéric Gasquet-Cyrus, Gaëlle Planchenault : *Jouer (de) l’accent marseillais à la télévision, ou l’art de mettre l’accent en boîte.*
- Myriam Dupouy : *Dire (avec) l’accent en formation linguistique obligatoire pour adultes allophones, l’accent comme indicateur d’identité linguistique assignée, subie ou choisie.*

## Traduction

- Rosina Lippi-Green : *Le mythe du non-accent* (1<sup>re</sup> édition 1997), traduit de l’anglais par Gaëlle Planchenault.

## Compte-rendus

- François Gaudin : *Signifier, essai sur la mise en signification (parcours dans l’espace épistémique et dans l’espace communicationnel ordinaire)*, de **Robert Nicolai**, 2017, ENS éditions. Collection Langages, Lyon, ISBN-13978-2-84788-924-6.
- Doyle Calhoun : *De la bouche même des indigènes : Échanges linguistiques en Afrique coloniale*, de **Cécile Van de Avenne**, 2017, Paris, Vendémiaire.
- Catherine Adam : *La Bretagne Linguistique n°21*, **Mannaig Thomas, Nelly Blanchard** (dirs), 2017, CRBC, UBO/UBL, Brest, 320 p., ISBN : 979-10-92331-31-8.
- Fabienne Leconte : *Dessiner les frontières*, sous la direction de **Michelle Auzanneau et Luca Greco**, 2018, ENS éditions, collection Langages, Lyon.
- Anaïs Delabie : *Language, capitalism, colonialism – toward a critical history*, de **Monica Heller et Bonnie S. McElhinny**, 2017, University of Toronto Press, 336 pages, ISBN-13: 978-1442606203.

# PEUT-ON IDENTIFIER PERCEPTIVEMENT HUIT ACCENTS RÉGIONAUX EN FRANÇAIS ? LA RÉPONSE DES SCIENCES PARTICIPATIVES<sup>1</sup>

**Mathieu Avanzi**

**Sorbonne-Université**

**Philippe Boula de Mareüil**

**CNRS & Univ. Paris-Saclay (LIMSI)**

## **Introduction**

« Ceux qui n'ont pas d'accent, je ne puis que les plaindre ! », écrivait Miguel Zamacoïs (1910) dans son poème *L'accent* popularisé par Fernandel. Un accent peut constituer une marque de fabrique, une arme de séduction, une coquetterie pour tel artiste, pour tel politicien soucieux de sa clientèle, tel commerçant attaché à la promotion des produits du terroir : il arrive ainsi que la publicité et le tourisme s'en emparent. D'un autre côté, il est pour certains mal vu ou, pire, vulgaire d'avoir un accent, ce qui peut conduire, par souci de réussite sociale, à atténuer son accent, voire à le masquer, sans même que cela soit toujours conscient (Singy, 1996 ; Boula de Mareüil, 2010 ; Prikhodkine, 2011). À proprement parler, tout le monde a un accent si, par ce vocable, on entend une façon particulière de prononcer une langue (Walter, 1988 : 186 ; Lippi-Green, 2012 : 33). Mais le plus souvent, du moins dans les représentations des néophytes, si quelqu'un a un accent, c'est qu'il n'est « pas d'ici ».

La perception des accents, problématique centrale dans le champ des études en sociophonétique (Preston, 1989 ; Innàccaro & Dell'Aquila, 2001, *inter alia*), a été abordée sous trois angles principaux : celui des représentations et de la cartographie mentale en l'absence d'input linguistique, dans le sillage du courant nommé *folk linguistics* (Preston, 1989, 2005 ; L'Éplattenier-Saugy, 2002) ; celui des attitudes évaluatives et affectives en réponse à des stimulus linguistiques (Rubin, 1992 ; Kuiper, 2005) ; celui de l'aptitude d'auditeurs à discerner divers accents à partir d'un input phonétique, typiquement d'une dizaine ou une quinzaine de secondes, enregistré auprès de locuteurs provenant de différentes localités. Cette dernière approche a plus spécifiquement été appliquée au français, avec des tests de perception incluant différents points d'enquête de la francophonie (Moreau *et al.*,

---

<sup>1</sup> Une version plus brève de ce travail a fait l'objet d'une publication dans les actes des JEP : [https://www.isca-speech.org/archive/JEP\\_2018/pdfs/190907.pdf](https://www.isca-speech.org/archive/JEP_2018/pdfs/190907.pdf).

2008 ; Boula de Mareüil *et al.*, 2017), en France (Hauchecorne & Ball, 1997 ; Armstrong & Boughton, 1997 ; Coquillon, 2005 ; Woehrling & Boula de Mareüil, 2006 ; Pustka, 2007), en Belgique (Boula de Mareüil & Bardiaux, 2011 ; Bardiaux, 2014), en Suisse (Sertling-Miller, 2007 ; Racine *et al.*, 2013 ; Goldman *et al.*, 2014), en Afrique de l’Ouest (Boula de Mareüil & Boutin, 2011 ; Lyche & Skattum, 2012) ou au Canada (Falkert, 2013 ; Remysen, 2014, 2016). Ces études ont notamment permis de mettre en lumière le fait qu’en Europe, les auditeurs natifs étaient capables de distinguer peu ou prou quatre grands groupes d’accents (ou *variétés*, terme plus vague) du français : France septentrionale, France méridionale, Suisse et Belgique.

Ces mêmes études ont montré, qu’à l’intérieur de ces grandes régions, des distinctions plus fines étaient possibles, mais moins systématiques, voire aléatoires. En pratique, il est apparu qu’en France, il était relativement difficile pour des auditeurs naïfs de distinguer entre les accents du Sud-Est et du Sud-Ouest (Woehrling & Boula de Mareüil, 2006), de même qu’entre ceux-ci et celui de Corse (Boula de Mareüil *et al.*, 2017). Le même constat peut être fait pour la partie septentrionale de l’Hexagone : d’après l’étude pionnière d’Armstrong et Boughton (1997), des auditeurs ordinaires sont incapables de distinguer entre locuteurs de Nancy et locuteurs de Rennes, deux villes situées à égale distance de Paris. En Belgique et en Suisse, il a été montré que les locuteurs des localités les plus proches de la frontière française (c’est-à-dire Tournai et Genève) étaient souvent confondus avec des locuteurs originaires de l’Hexagone, mais que les auditeurs n’étaient pas toujours capables de distinguer entre locuteurs originaires de différentes localités plus éloignées de la frontière (Racine *et al.*, 2014 ; Bardiaux, 2014). Enfin, il est ressorti que l’accent alsacien était parfois confondu avec les accents belge et suisse (Woehrling, 2009 ; Boula de Mareüil *et al.*, 2017).

Partant du constat que ces expériences ne permettaient pas d’évaluer systématiquement le rôle de l’âge, du profil socioéconomique et de l’origine des auditeurs dans leur capacité à reconnaître l’origine de locuteurs francophones d’Europe, une étude centrée autour de la France, de la Suisse et de la Belgique a été mise au point, permettant de tenir compte de ces aspects (Avanzi & Boula de Mareüil, 2017). Au total, des échantillons (d’une durée comparable aux tests précédents, soit une douzaine de secondes), provenant de 120 locuteurs<sup>2</sup> (8 originaires de 5 localités différentes pour chacun de ces trois pays, appariés en âge et en sexe<sup>3</sup>) ont été présentés à des centaines d’auditeurs ayant déclaré avoir passé la plus grande partie de leur vie à l’intérieur de ces mêmes pays. Dans une première série d’expériences, les auditeurs étaient invités à identifier le pays d’origine des locuteurs (trois choix : France, Suisse ou Belgique) : ils ont obtenu 60 % d’identification correcte en moyenne (ce qui est bien au-dessus du niveau de hasard, 33 %), mais les scores varient significativement en fonction de l’origine des locuteurs (les Français étant mieux reconnus que les Suisses et les Belges), de leur âge (les locuteurs les plus âgés étant identifiés plus facilement que les plus jeunes), de leur profil socioéconomique (les locuteurs exerçant des métiers manuels étant mieux reconnus que ceux exerçant des professions en lien avec la lecture et l’écriture) et de l’origine des auditeurs (les Français ne distinguent pas facilement les Suisses des Belges, alors que les Belges et les Suisses montrent moins de difficultés à faire la part entre les trois pays)<sup>4</sup>. Dans une deuxième série d’expériences, des auditeurs de France, de Suisse et de Belgique devaient identifier la région d’origine des locuteurs à l’intérieur de chacun de ces

<sup>2</sup> Dans le présent article, les termes employés pour désigner des personnes sont pris au sens générique ; ils ont à la fois valeur de féminin et de masculin.

<sup>3</sup> Il s’agissait de points d’enquête du projet Phonologie du Français Contemporain (Durand *et al.*, 2009). Pour la France, les locuteurs provenaient de Béthune (Nord), Brécey (Manche), Lyon (Rhône), Ogéviller (Meurthe-et-Moselle) et Paris. Pour la Belgique, ils provenaient de Bruxelles, Liège, Gembloux (province de Namur), Marche-en-Famenne (province du Luxembourg) et Tournai (province du Hainaut). Pour la Suisse, ils provenaient de Genève, Fribourg, Neuchâtel, Martigny (canton du Valais) et Nyon (canton de Vaud).

<sup>4</sup> Il s’agit de tendances globales, nous nous permettons de renvoyer le lecteur à l’article cité pour plus de détails sur la façon dont interagissent les variables ici décrites.

trois pays (avec un choix forcé entre 5 possibilités à l'intérieur de chaque pays). Sans surprise, les résultats se sont révélés moins bons, avec 31 % d'identification correcte en moyenne (ce qui reste toutefois au-dessus du niveau du hasard, une chance sur cinq, soit 20 %). Des analyses complémentaires ont été conduites pour évaluer le rôle de l'origine des auditeurs, de celle des locuteurs et de leur âge, ainsi que de leurs interactions. Elles ont montré des patrons de réponses asymétriques entre les trois pays étudiés : la France (ou, en France, Paris), semble agir comme un catalyseur d'unification et un pôle d'attraction — « attirant » les réponses des auditeurs. Les locuteurs les plus jeunes, en particulier, sont plus souvent perçus comme français (voire parisiens), quand leur accent n'est pas clairement identifiable. La Suisse, cependant, semble mieux résister que la Belgique à ce processus d'homogénéisation.

Cette dernière étude, que nous venons de résumer, ne comprenait pas de locuteurs alsaciens (parfois confondus avec des Suisses), ni de locuteurs méridionaux (du sud-est ou du sud-ouest de la France), ni même de Corses. La prise en compte de locuteurs originaires de ces régions est cruciale pour mieux comprendre comment des francophones perçoivent ou confondent les différents accents du français parlé en Europe. Pour cette raison, nous avons mis en place une nouvelle expérience, en vue d'apporter des éléments de réponses à une question en apparence simple. Peut-on, pour le français parlé en Europe, distinguer les huit accents associés aux régions suivantes : Suisse romande, Belgique francophone, Nord de la France, Bretagne, Alsace, Sud-Ouest, Sud-Est et Corse ? Comme tout dépend des sujets qui accomplissent la tâche — on a plus de chance d'identifier correctement un accent dont on est familier (Williams *et al.*, 1999 ; Gooskens, 2005 ; Boomershine, 2006 ; Baker *et al.*, 2009 ; Yan, 2015 ; Avanzi & Boula de Mareüil, 2017 ; Ruch, 2018) —, il est nécessaire d'avoir un nombre suffisant d'auditeurs de chacune de ces régions (ou chacun de ces pays), d'où l'importance de faire appel à des méthodes de type *crowdsourcing*<sup>5</sup>. Pour répondre à cette question, nous présenterons le corpus ainsi que la méthode employés et les résultats obtenus, globalement et en fonction de l'origine des auditeurs. L'analyse sera menée par région et par locuteur, avant de se concentrer sur les régions du sud de la France et de conclure.

## Corpus et méthode

Dans cette étude, nous rapportons les résultats d'une expérience visant à évaluer l'aptitude d'auditeurs à reconnaître l'origine géographique de locuteurs originaires de huit régions francophones d'Europe. Outre la Suisse et la Belgique, six régions de France ont été retenues : le Nord, la Bretagne à l'ouest, l'Alsace à l'est, le Sud-Ouest, le Sud-Est et la Corse<sup>6</sup>. Pour la France, les enregistrements ont été collectés, sur le terrain, auprès de locuteurs qui, en plus du français, parlaient une langue régionale (picard, alsacien, breton ou gallo, occitan, corse). En ce qui concerne la Suisse et la Belgique, les extraits sonores ont été tirés respectivement du corpus OFROM (Avanzi *et al.*, 2016) et de la base VALIBEL (Bolly *et al.*, 2016). Pour chacune des huit régions étudiées, nous avons sélectionné les productions de

<sup>5</sup> On dit aussi « production participative » ou « externalisation ouverte ». Appliquée aux domaines des sciences (on parle alors de « sciences participatives », « sciences citoyennes » ou « sciences collaboratives »), la méthode consiste à faire appel à un large public en vue de réaliser certaines tâches. Si la méthode ne date pas d'aujourd'hui, c'est seulement au début des années 2000 qu'elle a connu un certain essor, avec la démocratisation du web (Eskenazi *et al.*, 2013).

<sup>6</sup> Des traits de prononciation, a priori, peuvent distinguer le français parlé dans ces régions (Carton *et al.*, 1983 ; Avanzi, 2017) : citons le mot *moins*, dont la consonne finale tend à se faire entendre dans le Sud-Ouest mais pas ailleurs. Citons encore des mots comme *grosse* ou *fosse*, où le /O/ tend à être ouvert dans le Nord-Pas-de-Calais et dans tout le sud de la France (Corse exclue), ou *mot* lui-même, où le /O/ peut être ouvert en Belgique et en Suisse (Boula de Mareüil *et al.*, 2017). Ces traits sont présents dans les stimulus, qui peuvent être écoutés à l'adresse <https://perso.limsi.fr/mareuil/glottopol/>.

quatre locuteurs, à partir desquels nous avons extrait des segments de parole d'une durée comprise entre 10 et 15 secondes, en veillant à ce que les échantillons ne contiennent aucun indice lexical orientant l'identification (des mots comme *septante* ou *nonante* entraînant un biais vers la Suisse ou la Belgique, par exemple), ni trop de disfluences (*eah* d'hésitation, répétitions, etc.).

Code	Région	Localité	Âge (ans)	Sexe
AL_1	Alsace	Strasbourg	52	Femme
AL_2	Alsace	Illkirch-Graffenstaden	61	Femme
AL_3	Alsace	Geispolsheim	85	Femme
AL_4	Alsace	Colmar	50	Homme
BE_1	Belgique	Bruxelles	89	Homme
BE_2	Belgique	Liège	84	Homme
BE_3	Belgique	Paliseul	65	Homme
BE_4	Belgique	Dison	72	Homme
BR_1	Bretagne	Pacé	75	Femme
BR_2	Bretagne	Lannion	34	Homme
BR_3	Bretagne	Pacé	78	Homme
BR_4	Bretagne	Beaucé	69	Homme
CO_1	Corse	Corte	43	Femme
CO_2	Corse	Penta-Acquatella	60	Femme
CO_3	Corse	Ajaccio	23	Homme
CO_4	Corse	Ghisunaccia	23	Homme
NO_1	Nord	Denain	56	Femme
NO_2	Nord	Lille	58	Homme
NO_3	Nord	Marly	85	Femme
NO_4	Nord	Valenciennes	94	Homme
SE_1	Sud-Est	Caromb	34	Femme
SE_2	Sud-Est	Nice	33	Homme
SE_3	Sud-Est	Marseille	34	Homme
SE_4	Sud-Est	Maillane	62	Homme
SO_1	Sud-Ouest	Sénaillac	65	Femme
SO_2	Sud-Ouest	Toulouse	52	Homme
SO_3	Sud-Ouest	Caussade	74	Homme
SO_4	Sud-Ouest	Gimat	76	Homme
CH_1	Suisse	Chexbres	94	Femme
CH_2	Suisse	Treyvaux	86	Femme
CH_3	Suisse	Puplinge	71	Femme
CH_4	Suisse	Venthône	79	Homme

**Table 1. Pour chaque région, code, commune d'origine, âge au moment de l'enregistrement et sexe des locuteurs**

La tâche n'est pas aisée, car dans des entretiens à caractère biographique ou dans des conversations entre pairs, les tours de parole sont souvent plus courts, ou alors ils donnent trop d'indice sur l'origine de la personne qui parle (toponymes, références culturelles, etc.). À partir de dizaines d'heure de parole que nous avons à disposition, plus d'une centaine d'échantillons ont été présélectionnés, séparément, par les deux auteurs du présent article. Un accord a ensuite été trouvé, pour limiter ce nombre à 32 échantillons, par élimination progressive. Dans certains cas, il n'y avait guère le choix ; dans d'autres, nous avons retenu en priorité les extraits intéressants d'un point de vue phonétique, représentant des particularités de prononciations aussi « marquées » que possible (voir note 6 *supra*). Nous avons tâché de trouver le meilleur compromis possible entre la diversité des communes dans

chaque région, avec de grandes métropoles (Marseille, Toulouse, Bruxelles) comme de petits villages (notamment en Corse), et la qualité des extraits sonores.

Majoritairement assez âgés (moyenne = 63 ans ; écart type = 20,6 ans), les locuteurs retenus (13 femmes, 19 hommes) étaient bien ancrés dans la région où ils ont été enregistrés, pour y avoir passé la plus grande partie de leur jeunesse. On se plaçait ainsi dans des conditions plutôt favorables pour que leur origine géographique soit bien reconnue, les particularités locales de prononciation ayant tendance à être plus saillantes chez les personnes âgées que chez les plus jeunes (Léon & Léon, 1997 ; Woehrling & Boula de Mareuil, 2006 ; Pustka, 2007). Les locuteurs étaient de milieux socioprofessionnels variés : retraités, fonctionnaires, enseignants, associatifs, (anciens) mineurs, infirmiers, etc. Sur le plan géographique, les locuteurs du nord de la France étaient tous originaires du département du Nord. Ceux du Sud-Est étaient originaires du Vaucluse, des Bouches-du-Rhône et des Alpes-Maritimes. Ceux du Sud-Ouest étaient originaires du Gers, de Haute-Garonne, du Tarn-et-Garonne et du Lot. Les Alsaciens étaient originaires du Haut-Rhin et du Bas-Rhin. Les Bretons étaient originaires des Côtes-d'Armor et d'Ille-et-Vilaine. Les Corses étaient originaires de Haute-Corse et de Corse-du-Sud. Enfin, les Belges étaient originaires de trois provinces différentes (Bruxelles, Liège et Luxembourg), les Suisses de quatre cantons différents (Fribourg, Genève, Vaud et Valais). Le Tableau 1 donne des informations plus précises sur ces locuteurs et leur nom de code, leur localisation par commune étant cartographiée dans la Figure 1.

L'expérience d'identification a été mise au point avec le logiciel Qualtrics, lequel permet de faire des sondages en ligne, accessibles à partir de n'importe quel navigateur ou support (ordinateur, téléphone portable multifonction ou tablette). Les participants, contactés via les réseaux universitaires et les réseaux sociaux, étaient invités à cliquer sur un lien qui les amenait sur la plateforme d'enquête. Ils étaient informés qu'ils allaient entendre plusieurs extraits sonores, et qu'ils devaient identifier l'origine géographique des locuteurs sur la base de l'écoute de ces stimulus. Pour ce faire, ils devaient appuyer sur le bouton du lecteur multimédia intégré au navigateur et cliquer, après écoute du stimulus, n'importe où à l'intérieur de la région colorée de la carte reproduite dans la Figure 2. Les participants pouvaient jouer l'extrait sonore autant de fois qu'ils le souhaitaient, mais ne pouvaient pas revenir en arrière après avoir validé leur réponse.

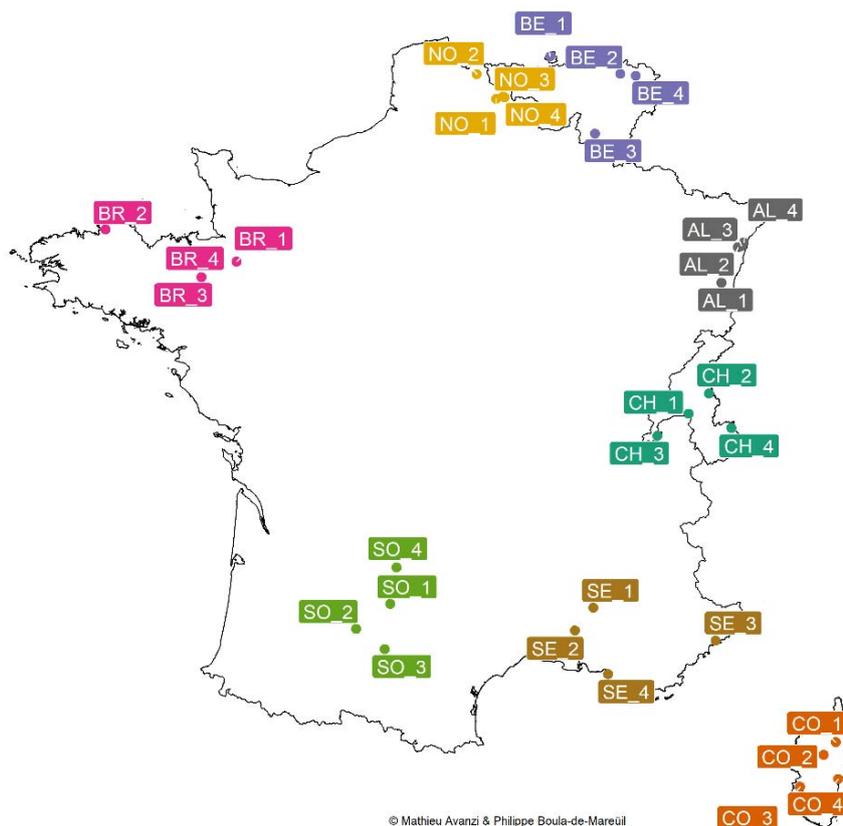


Figure 1. Carte présentant l'origine précise des locuteurs

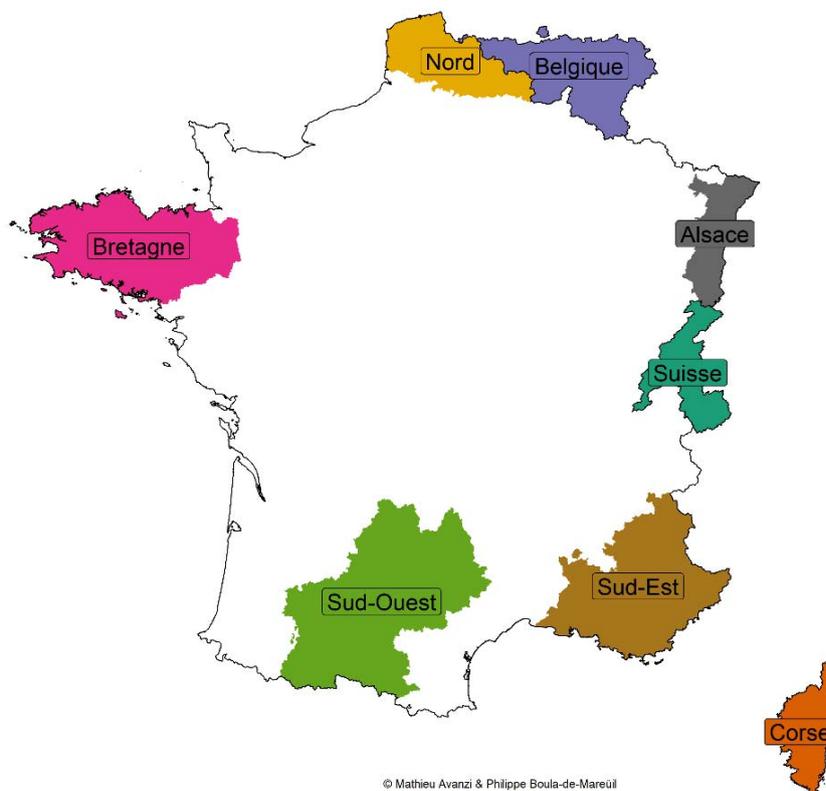


Figure 2. Carte présentant les régions et leurs étiquettes proposées aux auditeurs lors de l'expérience de perception

Un des objectifs de cette étude étant de vérifier l'hypothèse selon laquelle on reconnaît mieux l'accent de sa propre région que celui des autres (voir les références citées *supra*), nous avons codé l'origine géographique des auditeurs dans notre base de données. Nous avons considéré la Suisse et la Belgique comme des régions à part entière, à partir des informations données par les sujets<sup>7</sup>. Pour la France, nous avons eu recours à des tables de correspondance permettant de déterminer la région d'origine des participants à partir des codes postaux qu'ils avaient fournis : pour l'Alsace, la Bretagne et la Corse, nous avons considéré les participants originaires de ces mêmes (anciennes) régions ; pour le Sud-Ouest, nous avons considéré les participants originaires de l'ancienne région Midi-Pyrénées, pour le Nord ceux du département du Nord et pour le Sud-Est ceux de la région Provence-Alpes-Côte-d'Azur. Les auditeurs se répartissent comme rapporté dans la Table 2.

Région	N	%	Sexe		Âge			
			H	F	moyenne	min.	max.	écart type
<b>Alsace</b>	79	5	34	45	29,3	14	69	12,4
<b>Belgique</b>	118	7,5	54	64	28,7	17	80	14,4
<b>Bretagne</b>	145	9,2	69	76	29,8	14	89	12,3
<b>Corse</b>	10	0,6	2	8	35,4	20	70	18,0
<b>Nord</b>	63	4	39	24	30,6	15	71	11,6
<b>Sud-Est</b>	107	6,8	45	62	32,9	16	72	11,8
<b>Sud-Ouest</b>	102	6,5	45	57	31,2	14	67	12,5
<b>Suisse</b>	48	3	30	18	31,5	14	65	13,7
<b>Autre</b>	903	57,3	451	452	32,7	10	87	13
<b>Total</b>	<b>1575</b>	<b>100</b>	<b>769</b>	<b>806</b>	<b>29,3</b>	<b>10</b>	<b>89</b>	<b>12,4</b>

**Table 2. Participants à l'enquête, avec leur région d'origine, nombre, pourcentage par rapport au total, sexe et âge**

L'échantillon, qui s'élève à plus de 1500 auditeurs, est relativement bien équilibré en ce qui concerne le sexe des participants (51,2 % de femmes), et malgré le format de l'expérience (accessible seulement en ligne et diffusée majoritairement à travers les réseaux universitaires et sociaux), notre panel est relativement varié du point de vue de l'âge. Malgré tous nos efforts, nous ne pouvons que constater le nombre décevant de Corses, qui peut s'expliquer par la relativement faible population de l'île de Beauté (la Corse-du-Sud, par exemple, est le deuxième département le moins peuplé de France après la Lozère). En Haute-Corse, de plus, nous n'avons pas souhaité mobiliser les réseaux militants auxquels les locuteurs appartenaient, pour ne pas biaiser les résultats par une reconnaissance de la voix plutôt que de l'accent.

<sup>7</sup> Nous avons choisi de traiter la Belgique et la Suisse comme des régions à part entières (c'est-à-dire de ne pas séparer ces pays en plusieurs entités géographiques) pour deux raisons. Premièrement, il s'agissait de travailler avec des régions de taille similaire (la Suisse est grande comme deux départements français, la Belgique comme trois départements français). Deuxièmement, l'expérience nous a montré qu'il aurait été très difficile (voire impossible), pour des auditeurs français, d'identifier avec précision la province d'un Belge ou le canton d'un Suisse, étant donné que les auditeurs de ces pays y parviennent difficilement eux-mêmes (Avanzi & Boula de Mareuil, 2017).

## Résultats

### Statistiques

Les analyses statistiques ont été conduites avec le logiciel R version 3.3.2 (R Development Core Team, 2018). Pour vérifier si les variétés ont été identifiées au-dessus du niveau du hasard, nous avons effectué différents tests  $t$  en fixant la valeur de  $p$  à 0,99. Pour examiner l'effet de l'origine des locuteurs et/ou des auditeurs sur les scores d'identification obtenus, nous avons utilisé des modèles de régressions logistiques généralisés à effets mixtes, avec une fonction *logit* — librairie *lme4* (Bates *et al.*, 2013) —, dans lesquels nous avons inclus les locuteurs et les auditeurs comme variables aléatoires. Les valeurs de  $p$  ont été obtenues à l'aide des fonctions *drop1* de la librairie *lme4* et *lsmeans* de la librairie éponyme. Enfin, pour visualiser dans un plan cartésien la distance relative entre les variétés perçues, nous avons eu recours à des techniques d'échelonnements multidimensionnels (*Multi Dimensional Scaling*, MDS) de la librairie *MASS* (Venables & Ripley, 2002). Tous les graphiques ont été réalisés dans R avec la librairie *ggplot2*.

### Résultats globaux

#### *Résultats d'identification par région*

Dans un premier temps, nous avons calculé le pourcentage de réponses obtenues pour chaque groupe de locuteurs, sans tenir compte de l'origine géographique des auditeurs. Nous avons ainsi pu générer la Figure 3 à partir de la matrice de confusion présentée dans la Table 3.

En moyenne, avec 35,9 % d'identification correcte, il ressort que les accents des huit régions sont reconnus bien au-delà du hasard (12,5 %, chacun des huit tests  $t$  donnant des valeurs de  $p < 0,001$ ). Il est intéressant de remarquer que le score le plus élevé (45,9 %) a été obtenu par les Bretons : il semble que, leur accent échappant aux stéréotypes associés à d'autres<sup>8</sup>, la réponse Bretagne ait opéré comme une sorte de choix par défaut.

Pour vérifier si les différences entre les scores de reconnaissance sont significatives, nous avons appliqué un modèle dans lequel la réponse (codée comme VRAI/FAUX) était la variable dépendante ; la région des locuteurs (8 possibilités), la réponse des auditeurs (8 possibilités) et leur interaction étaient les prédicteurs, les stimulus et les auditeurs étant entrés comme des variables aléatoires. Les résultats de ce modèle statistique ont révélé qu'il n'y a pas d'effet de région — les Corses, avec 17,9 % d'identification correcte, ne sont pas moins bien reconnus que les Bretons, avec 45,9 % d'identification correcte, pour ne prendre que les deux cas extrêmes — ni d'effet de réponse : en d'autres termes, les participants n'ont pas cliqué plus souvent sur la région Bretagne (16,1 %) que sur la région Corse (7,2 %), pour ne prendre de nouveau que les deux cas extrêmes. Cette absence d'effets simples est vraisemblablement due à la présence d'une interaction significative entre la région des locuteurs et la réponse des auditeurs ( $\chi^2(7) = 3769,1$  ;  $p < 0,0001$ ). Compte tenu de la présence de cette interaction significative, 8 modèles distincts ont ensuite été appliqués (un pour chaque région) afin de mettre au jour d'éventuelles différences entre les scores d'identification correcte à l'intérieur de chacune des huit régions. De ces modèles, il est ressorti que l'effet de la variable « réponse des auditeurs » était toujours significatif (tous les modèles mettent en évidence un effet de réponse avec des valeurs de  $p < 0,0001$ ). Pour rendre plus fluide la lecture, nous ne commenterons ici que les différences impliquant les

---

<sup>8</sup> Il n'existe pas clairement, dans l'imaginaire populaire, de représentations stéréotypées associées à l'accent breton, contrairement à ce qui se passe pour l'accent alsacien, l'accent méridional ou l'accent du Nord-Pas-de-Calais.

pourcentages de réponses les plus élevés pour chaque région (en grisé dans la Table 3, les valeurs correctes étant en gras).

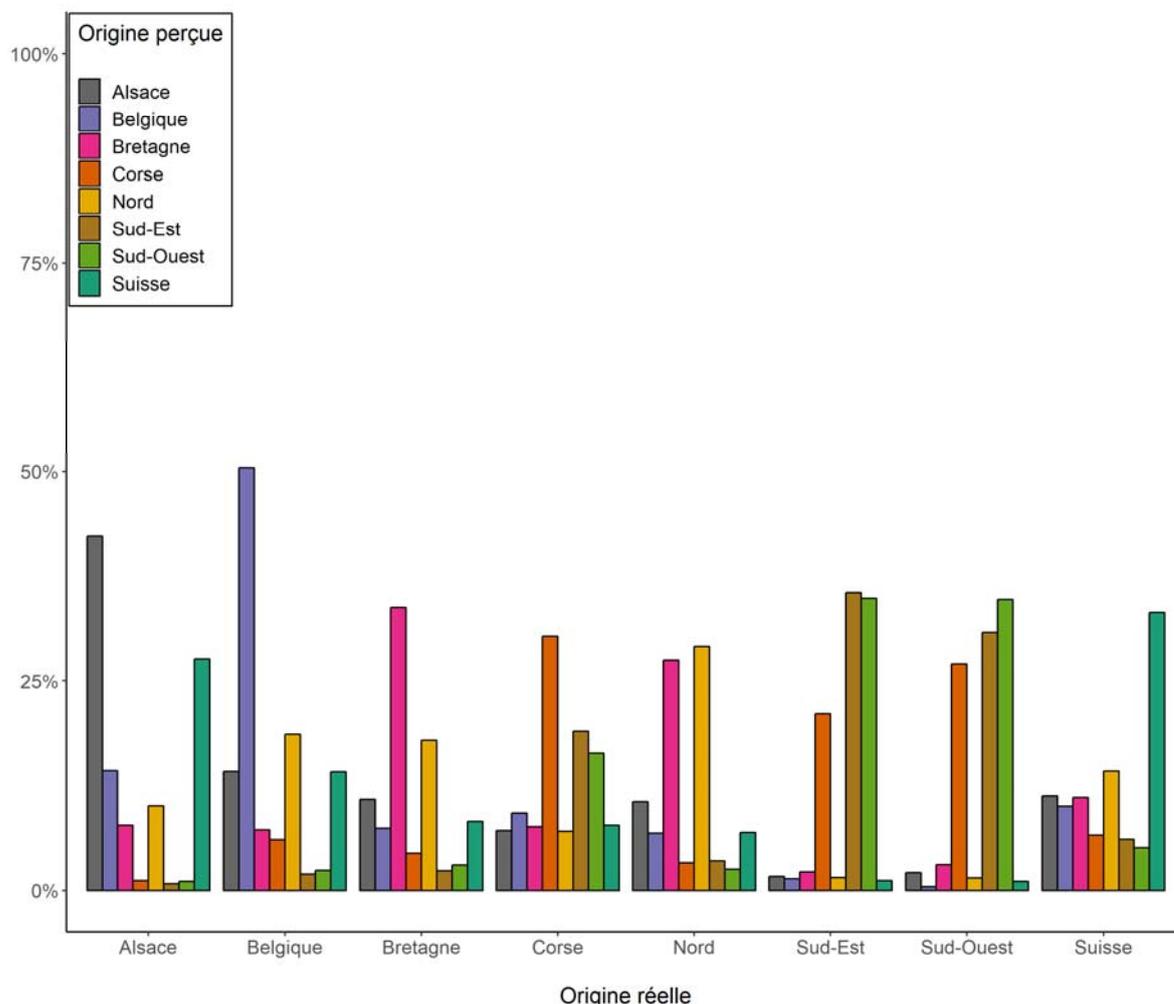


Figure 3. Pourcentage d'identification des locuteurs en fonction de leur région d'origine réelle et de leur région d'origine perçue

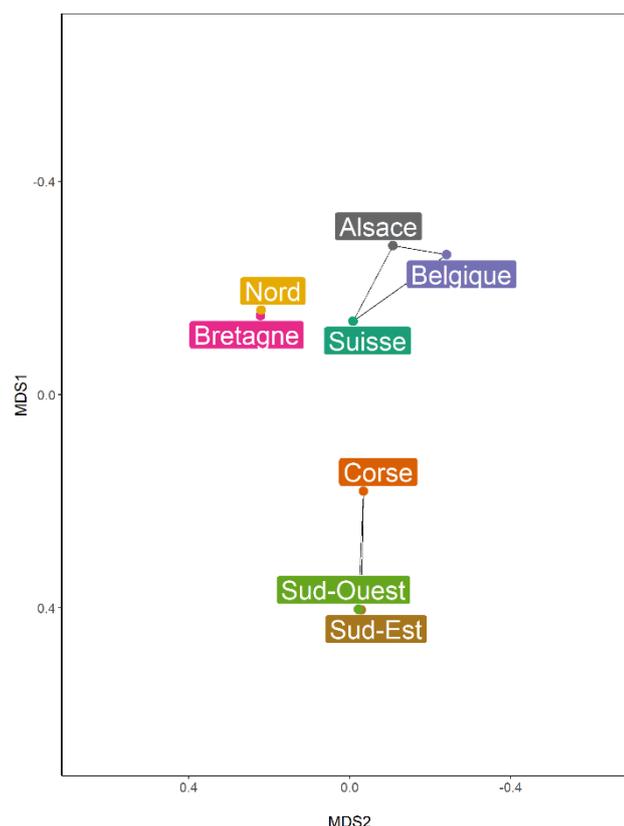
		Origine perçue							
		Alsace	Belgique	Bretagne	Corse	Nord	Sud-Est	Sud-Ouest	Suisse
Origine réelle	Alsace	<b>39,8</b>	11,2	9,6	0,7	9,2	0,9	1,3	27,5
	Belgique	13,2	<b>38,9</b>	8,8	3,3	16,9	2,1	2,8	13,9
	Bretagne	11,2	6,3	<b>45,9</b>	2,7	18,1	2,8	4,0	8,9
	Corse	7,0	7,6	9,9	<b>17,9</b>	6,8	21,8	21,0	8,1
	Nord	10,8	5,8	37,2	2,0	<b>29,2</b>	4,2	3,3	7,4
	Sud-Est	1,5	1,1	2,7	11,7	1,4	<b>38,4</b>	42,0	1,2
	Sud-Ouest	2,0	0,4	3,8	15,2	1,4	33,8	<b>42,3</b>	1,1
	Suisse	11,2	8,3	14,5	3,9	13,9	7,0	6,5	<b>34,9</b>

Table 3. Matrice de confusion entre région d'origine réelle et région d'origine perçue, tous participants confondus (%)

Pour l'Alsace, la Belgique, la Bretagne, le Sud-Ouest et la Suisse, les valeurs les plus hautes sont significativement différentes de toutes celles des lignes parentes, ce qui suggère par exemple que, pour les locuteurs alsaciens, la différence entre les réponses Alsace (39,8 %) et Suisse (27,5 %) n'est pas due au hasard. Pour la Corse, la différence n'est pas significative entre les réponses Corse (17,9 %) et Sud-Est d'une part (21,8 %), Corse et Sud-Ouest d'autre

part (21 %). Pour le Nord, le pourcentage de bonnes réponses (29,2 %) est significativement inférieur ( $p < 0,0001$ ) au pourcentage de réponses Bretagne (37,2 %). Enfin, pour le Sud-Est, il n'y a pas de différence significative entre le pourcentage de bonne réponse (38,4 %) et le pourcentage de réponses Sud-Ouest (42 %). En résumé, il ressort qu'à l'intérieur de chaque région, le score d'identification correcte de ladite région est significativement différent de tous les autres scores à l'intérieur de cette même région. Les exceptions concernent la Corse (les Corses étant confondus avec des locuteurs du Sud-Est et du Sud-Ouest), le Nord (où c'est la réponse Bretagne qui arrive en tête des suffrages) et le Sud-Est (dont les locuteurs sont confondus avec ceux du Sud-Ouest).

Sur le plan perceptif, se dégagent trois groupes de variétés, que l'échelonnement multidimensionnel présenté dans la Figure 4 permet de visualiser directement. Les trois variétés du sud de la France (Sud-Ouest, Sud-Est et Corse) forment un premier groupe, que nous nommerons désormais « groupe Grand Sud » : ces variétés ont souvent été confondues entre elles, la distance entre la Corse et les deux régions du Midi de la France étant toutefois plus importante que celle entre Sud-Est et Sud-Ouest. D'autre part, on trouve les variétés de l'est de la francophonie d'Europe (la Belgique étant perceptivement équidistante de la Suisse et de l'Alsace, formant ce que nous appellerons désormais le « groupe Nord-Est » ; on a enfin les variétés du nord-ouest de la France (Nord et Bretagne), formant ce que nous nommerons désormais le « groupe Nord-Ouest ».



**Figure 4. Échelonnement multidimensionnel permettant de représenter la confusion entre la région d'origine réelle et la région d'origine perçue, tous participants confondus**

#### *Résultats par locuteur*

Les résultats par région, détaillés ci-dessus, masquent une certaine hétérogénéité à l'intérieur de chacune des huit régions étudiées, et certaines confusions non-attendues. Les origines géographiques des locuteurs ayant produit les stimulus n'ont pas été reconnues avec

la même précision (comme on peut le voir dans la Table 4) et n'ont pas toujours été confondues avec les mêmes régions (comme on peut le voir dans la Figure 5).

		Origine perçue							
		Alsace	Belgique	Bretagne	Corse	Nord	Sud-Est	Sud-Ouest	Suisse
Origine réelle	AL_1	<b>44,1</b>	13,7	2,2	0,7	11,3	0,7	1,6	25,8
	AL_2	<b>31,0</b>	10,6	17,3	0,9	8,3	1,2	1,8	28,8
	AL_3	<b>39,7</b>	9,1	11,9	0,4	6,6	0,6	0,7	31,0
	AL_4	<b>44,2</b>	11,2	7,1	0,6	10,3	1,0	1,0	24,4
	BE_1	7,3	<b>68,7</b>	1,4	1,0	7,1	0,3	1,0	13,1
	BE_2	13,0	<b>15,9</b>	24,6	4,7	22,4	4,4	6,6	8,3
	BE_3	14,5	<b>44,4</b>	3,7	2,6	7,3	1,4	1,3	24,8
	BE_4	17,8	<b>25,9</b>	6,4	4,9	<b>30,7</b>	2,4	2,7	9,2
	BR_1	9,5	6,7	<b>49,8</b>	2,1	12,7	3,3	4,9	11,1
	BR_2	8,2	6,1	<b>48,8</b>	4,3	16,9	2,9	4,0	8,8
	BR_3	12,3	6,3	<b>38,0</b>	2,7	27,1	2,2	4,9	6,5
	BR_4	14,6	6,3	<b>47,2</b>	1,7	15,5	3,1	2,3	9,3
	CO_1	6,0	4,6	6,8	<b>15,3</b>	3,1	26,3	<b>27,1</b>	10,8
	CO_2	6,5	2,4	11,3	<b>27,2</b>	4,7	20,1	24,2	3,6
	CO_3	4,9	3,9	15,8	<b>15,3</b>	3,9	<b>28,7</b>	19,8	7,7
	CO_4	10,5	<b>19,0</b>	6,0	<b>13,7</b>	15,1	12,5	12,9	10,2
	NO_1	13,7	9,3	<b>33,5</b>	1,1	<b>28,5</b>	3,0	2,6	8,3
	NO_2	6,4	4,3	<b>58,1</b>	1,5	<b>10,9</b>	4,1	3,4	11,3
	NO_3	10,4	5,8	<b>20,3</b>	1,6	<b>47,2</b>	5,3	4,0	5,5
	NO_4	12,4	3,8	<b>38,2</b>	3,8	<b>29,2</b>	4,3	3,3	5,0
	SE_1	2,4	2,1	4,1	7,3	2,3	<b>42,1</b>	38,3	1,5
	SE_2	0,3	0,3	0,5	6,2	0,1	<b>44,1</b>	<b>48,6</b>	0,0
	SE_3	3,2	1,8	4,6	14,7	2,4	<b>31,5</b>	39,6	2,2
	SE_4	0,5	0,3	1,9	18,6	1,0	<b>35,6</b>	41,1	1,0
	SO_1	1,6	0,3	2,8	30,6	1,7	17,5	<b>44,4</b>	1,2
	SO_2	0,4	0,1	0,5	15,7	0,0	<b>45,1</b>	<b>37,9</b>	0,1
	SO_3	1,0	0,6	1,0	7,4	1,1	40,1	<b>47,8</b>	1,0
	SO_4	5,2	0,5	10,9	6,3	2,7	32,9	<b>39,3</b>	2,1
	CH_1	16,3	8,7	13,8	3,0	17,8	4,0	4,5	<b>32,0</b>
	CH_2	8,1	7,4	3,9	0,5	5,6	0,6	0,8	<b>73,1</b>
	CH_3	19,0	9,0	17,8	3,5	16,4	8,2	6,1	<b>20,1</b>
	CH_4	0,0	8,0	<b>25,0</b>	9,6	17,1	17,2	16,6	<b>6,5</b>

Table 4. Matrice de confusion entre localité d'origine réelle et région d'origine perçue, tous participants confondus (%)<sup>9</sup>

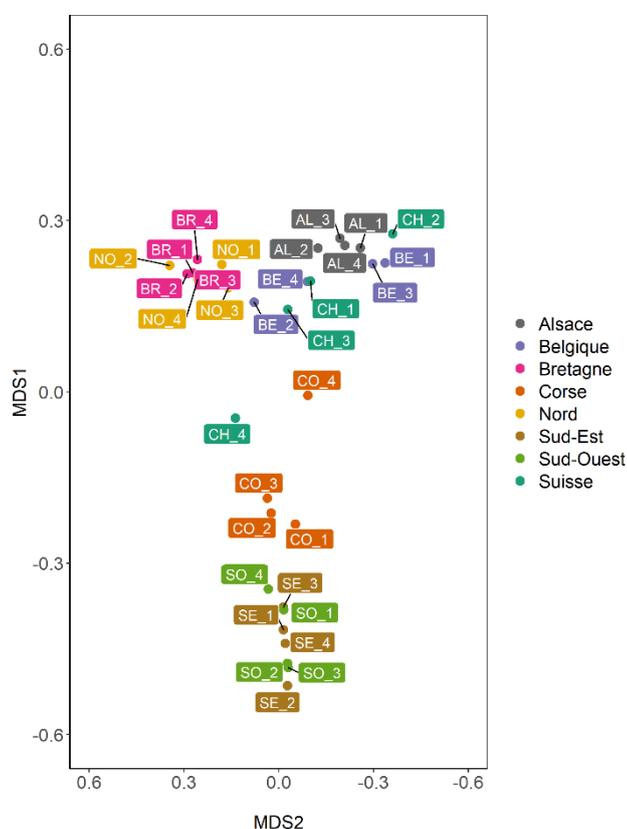
À l'intérieur du groupe Nord-Est, on observe une sorte de continuum, avec à sa tête un noyau formé par trois locuteurs alsaciens (AL\_1, AL\_3 et AL\_4) présentant des scores élevés d'identification correcte (42,6 % en moyenne), le quatrième locuteur (la locutrice AL\_2) n'étant identifié correctement qu'à 31 %. À ce noyau s'agrègent deux locuteurs belges (BE\_1, originaire de Bruxelles et BE\_3, originaire de la province du Luxembourg), mieux reconnus que les autres locuteurs belges (tous deux originaires de la province de Liège). Ces derniers locuteurs (BE\_2 et BE\_4) sont souvent confondus avec des locuteurs du Nord ou de Bretagne : ils se positionnent à la périphérie du groupe Nord-Est, et à proximité du groupe Nord-Ouest. Enfin, de part et d'autre de ce groupe Nord-Est, on trouve le locuteur

<sup>9</sup> Les valeurs en gras indiquent les scores d'identification correcte (les locuteurs Alsaciens ont bien été reconnus comme originaires de l'Alsace) ; les cases en grisé signalent les valeurs les plus hautes obtenues pour chaque ligne.

fribourgeois (CH\_2), locuteur le mieux identifié comme Suisse romand (avec 73,1 % de bonnes réponses). Les locuteurs vaudois (CH\_1) et genevois (CH\_3), souvent confondus avec des Alsaciens, des Bretons ou des locuteurs du Nord, sont perceptivement proches des locuteurs belges les moins bien identifiés. La locutrice valaisanne (CH\_4) est la moins bien identifiée du groupe (6,5 % de bonnes réponses), et en est la plus éloignée : elle se situe dans l'orbite du groupe Grand Sud, ce qui peut s'expliquer par le fait qu'elle produit des voyelles nasales avec un appendice consonantique à la manière des méridionaux.

Dans le groupe Nord-Ouest, on peut voir que, pour la région Nord, la locutrice originaire des environs de Valenciennes (NO\_3) est la seule pour laquelle la réponse majoritaire est la bonne. Les trois autres locuteurs nordistes sont le plus souvent identifiés comme Bretons.

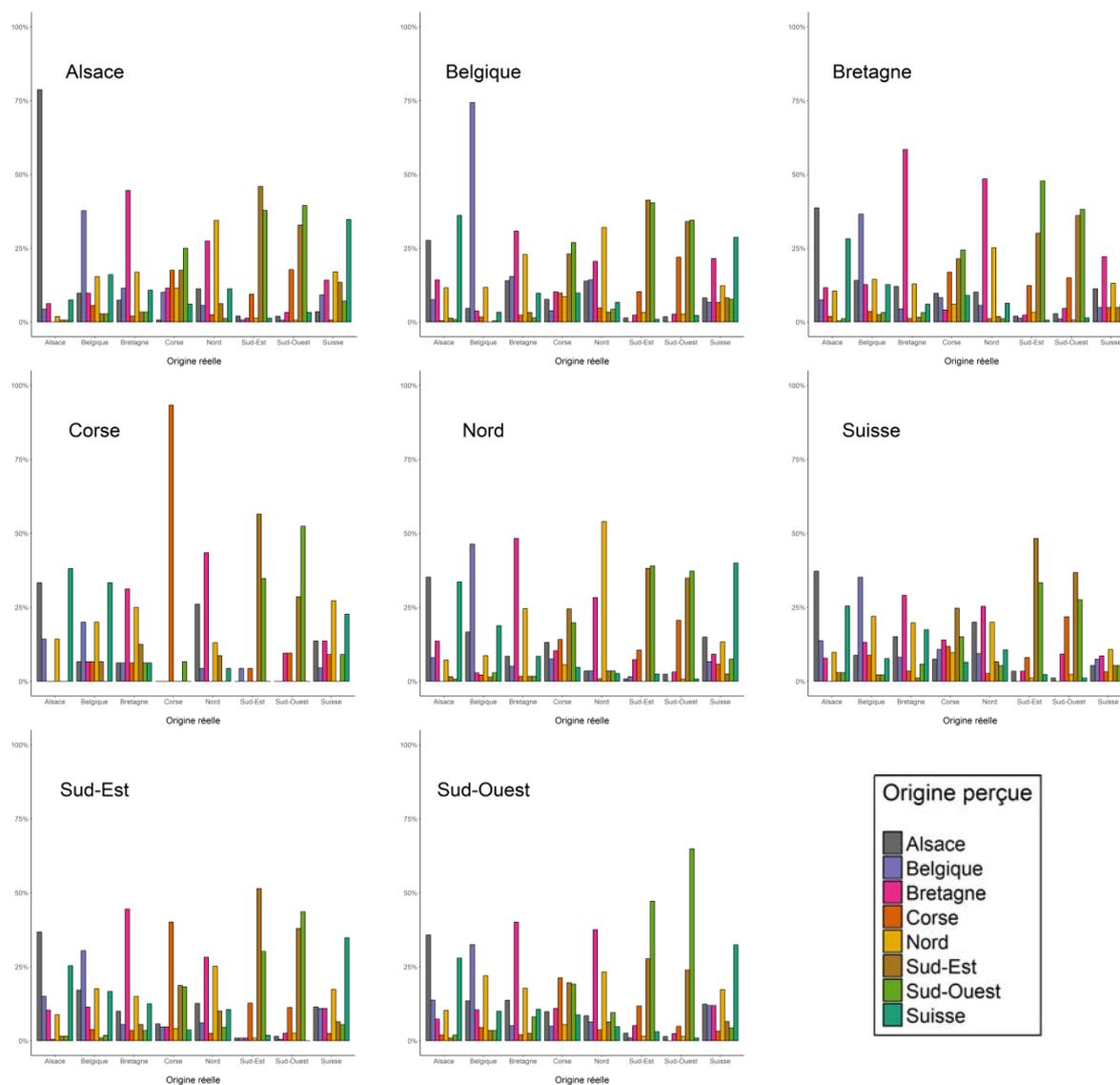
Dans le groupe Grand Sud, on observe de nouveau un continuum entre des locuteurs bien identifiés comme méridionaux (avec des confusions systématiques entre Sud-Est et Sud-Ouest), et des locuteurs moins prototypiques de ce groupe que sont les Corses. Les trois locuteurs originaires de Haute-Corse sont assez bien classés par rapport au quatrième, originaire de Corse-du-Sud (CO\_4), qui se trouve classé dans le groupe Nord-Est (entre CH\_3 et CH\_4).



**Figure 5. Échelonement multidimensionnel permettant de représenter la confusion entre la localité d'origine réelle et la région d'origine perçue, tous participants confondus**

### Effet de l'origine des auditeurs

Dans un second temps, nous avons souhaité examiner l'effet de l'origine des auditeurs sur l'identification d'accents régionaux, notre hypothèse étant qu'un accent devrait être mieux reconnu par des locaux, qui en sont familiers. À cette fin, nous avons isolé les réponses des auditeurs originaires de chacune des régions et représenté graphiquement les résultats — les matrices de confusion qui ont permis de générer ces graphes ne sont pas rapportées ici, par souci d'économie.



**Figure 6. Pourcentage d'identification des locuteurs en fonction de leur région d'origine réelle et de leur région d'origine perçue, d'après les réponses des auditeurs locaux<sup>10</sup>**

Il apparaît assez clairement, à la lecture de la Figure 6, que les tendances par région suivent en général la tendance globale, qu'il est possible de visualiser dans la Figure 3 ci-dessus. On peut néanmoins pointer l'existence de pics de bonnes réponses pour chacune des variétés autochtones en Alsace, Belgique, Corse, et dans le Sud-Ouest, ce qui confirme que les auditeurs de ces régions sont meilleurs quand il s'agit d'identifier leur propre accent que quand il s'agit d'identifier ceux des autres. Ce résultat est assez intuitif : un Alsacien, par exemple, a plus de chance d'identifier correctement un accent alsacien que de discriminer entre des accents du sud-est et du sud-ouest de la France ou que de reconnaître un accent corse dont il serait peut-être moins familier. Pour les autres régions, les résultats sont plus contrastés. Chez les auditeurs de Bretagne, on observe un pic pour les réponses Bretagne (à 58,5 %) qui est à peine plus haut que le bâtonnet correspondant aux réponses Bretagne pour les locuteurs du Nord (48,4 %), ou que le bâtonnet correspondant aux taux de bonne identification des locuteurs du Sud-Est (47,8 %). Chez les auditeurs du Nord, le pic de bonnes réponses pour les locuteurs du Nord est à 55 %, ce qui est à peine plus haut que le taux de

<sup>10</sup> L'origine des auditeurs est à chaque fois indiquée en haut à gauche.

bonnes réponses pour les locuteurs de Bretagne (48,3 %) ou de Belgique (46,3 %). Chez les auditeurs de Suisse, les locuteurs suisses sont reconnus correctement à 53,7 % ce qui est à peine plus que le taux de bonnes réponses pour les locuteurs du Sud-Est (48,2 %). Chez les auditeurs du Sud-Est, enfin, les locuteurs du Sud-Est sont reconnus correctement à 51,4 %, alors que le score de bonne identification pour les locuteurs du Sud-Ouest est de 43,5 %.

	Alsace	Belgique	Bretagne	Corse	Nord	Sud-Est	Sud-Ouest	Suisse
<b>Locaux</b>	78,8	75,3	58,5	93,1	55,0	51,4	64,9	53,7
<b>Non-locaux</b>	37,6	35,9	44,7	17,4	28,1	37,4	40,1	34,2

Table 5. Taux d'identification correcte des locuteurs de chaque région, selon l'origine des auditeurs (%)

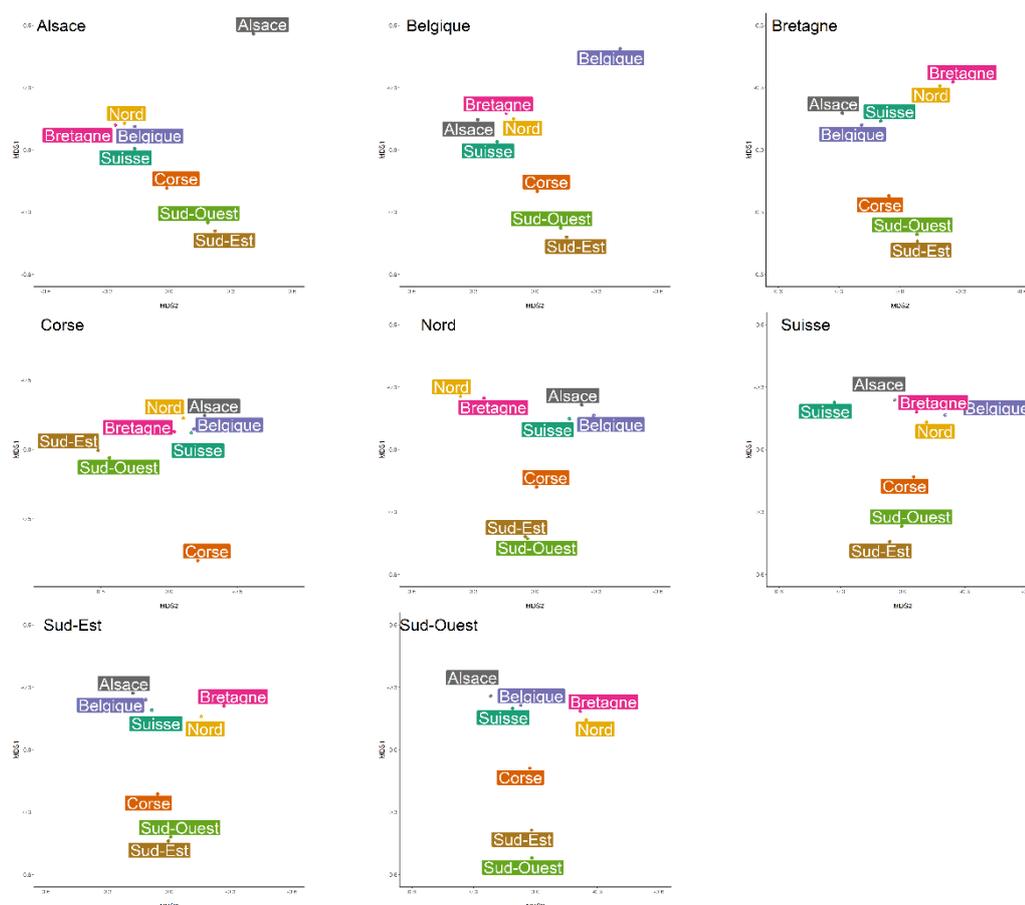


Figure 7. Échelonnements multidimensionnels pour les huit régions étudiées, d'après les réponses obtenues par les locaux (leur origine étant à chaque fois indiquée en haut à gauche)

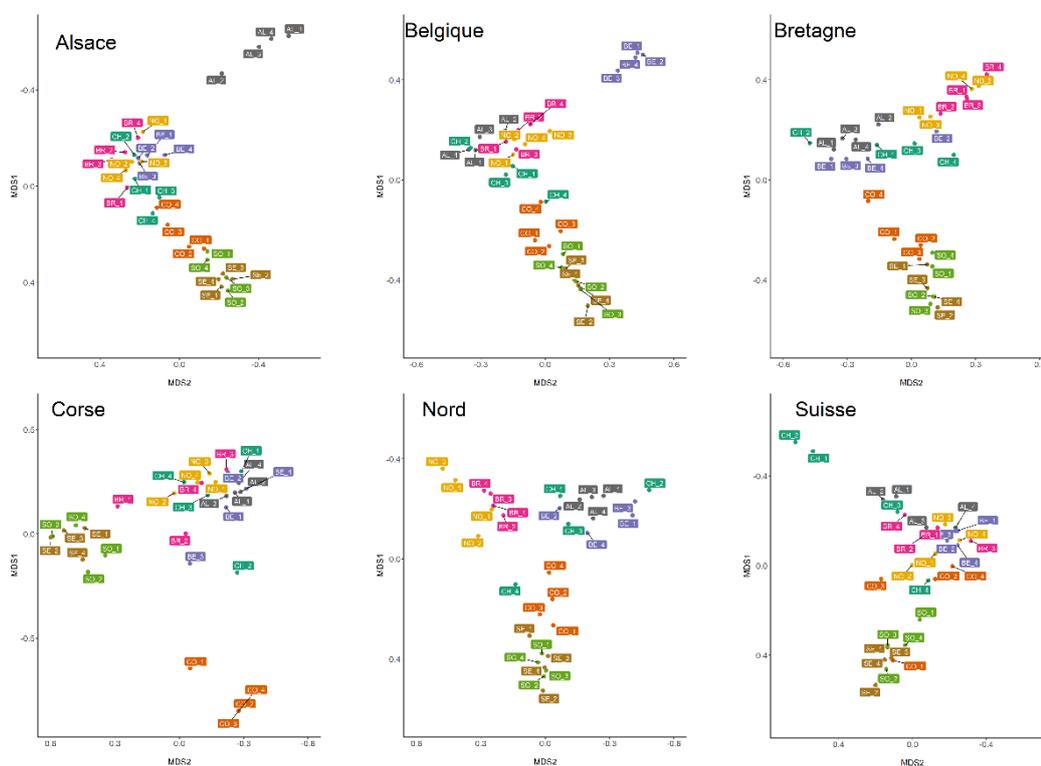
Compte tenu du nombre de régions à comparer, nous n'avons pu conduire de tests statistiques pour vérifier si les différences susmentionnées étaient significatives ou non. Plus simplement, en vue de tester statistiquement la capacité des auditeurs à reconnaître les accents de leur propre région, nous avons isolé les réponses obtenues pour les locuteurs de chacune des huit régions et avons comparé les scores des auditeurs habitant cette région à ceux des autres auditeurs. Les scores d'identification correcte des locaux et des autres, pour chacune des huit régions, sont consignés dans la Table 5. Sur cette base, nous avons réalisé sept<sup>11</sup> modèles linéaires généralisés avec la réponse VRAI/FAUX comme variable dépendante,

<sup>11</sup> Nous n'avons pas pu conduire de test pour la Corse, compte tenu du trop faible nombre de participants — les valeurs sont donc données à titre indicatif.

l'interaction avec l'origine des auditeurs (locaux ou non-locaux) et leur réponse (8 possibilités), les stimulus et les auditeurs étant entrés comme variables aléatoires. Toutes les différences sont significatives ( $p < 0,001$ ). Nous avons ensuite réalisé un MDS sur la base des réponses des auditeurs locaux uniquement, pour chacune des huit régions, ce qui nous a permis de visualiser aisément les confusions commises par les participants.

Si l'on compare à présent les 8 MDS de la Figure 7, on constate que certains groupes d'auditeurs ne confondent jamais l'accent de leur région avec celui des autres. Tel est le cas des Alsaciens, des Belges, des Corses et des Suisses, alors que d'autres groupes d'auditeurs ont plus de peine à distinguer l'accent de leur région de celui de régions proches : tel est le cas des Bretons et des auditeurs du Nord, mais également des auditeurs du sud de la France.

Comme dans la section précédente, nous avons réalisé des échelonnements multidimensionnels en fonction de l'origine des locuteurs, afin d'examiner la variation à l'intérieur des régions, et d'observer si les locuteurs les moins bien identifiés par l'ensemble des auditeurs de l'expérience posaient également problème aux locaux (cf. Figure 8). En Alsace, AL\_2 reste, comme précédemment, la locutrice la plus éloignée des autres. En Belgique, tous les locuteurs sont assez bien reconnus, aucun n'étant spécialement éloigné des autres. On peut faire la même observation dans les régions Nord, Bretagne, Sud-Est et Sud-Ouest : on n'a pas de différence notable entre de grandes métropoles comme Marseille ou Toulouse et de petites villes, et le fait que la locutrice de Provence SE\_1 prononce *moins* sans [s] final n'a pas empêché des confusions avec le Sud-Ouest, où ce [s] se fait généralement entendre (cf. note 5). Aux oreilles des Corses, c'est la locutrice CO\_1 qui est la moins bien reconnue comme corse, alors que dans l'analyse globale, c'était CO\_4 qui était la moins identifiée comme corse. Les Suisses enfin ne sont pas plus performants que l'ensemble des auditeurs de l'enquête, quand il s'agit d'identifier l'origine de CH\_3 et CH\_4.



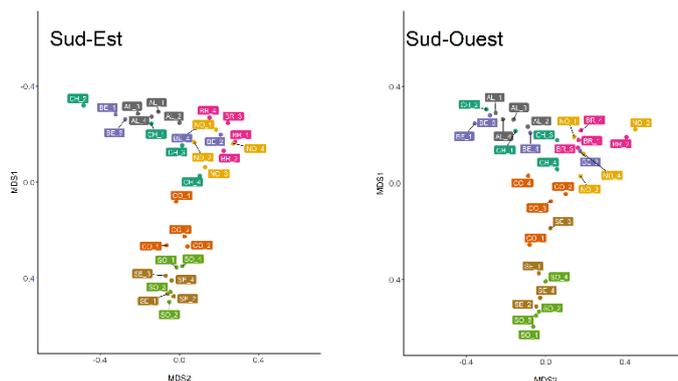


Figure 8. Échelonnements multidimensionnels pour les huit régions étudiées, d'après les réponses des auditeurs locaux (leur origine étant à chaque fois indiquée en haut à gauche)

### Zoom sur les régions du sud de la France

Les résultats présentés ci-dessus ont permis de mettre le doigt sur des confusions constantes entre le Sud-Est et le Sud-Ouest, y compris de la part des auditeurs de ces régions. Dans cette dernière partie, nous avons cherché à évaluer et comparer les capacités des auditeurs du Midi à distinguer les locuteurs du Sud-Est de ceux du Sud-Ouest.

Auditeurs du Sud-Est	Sud-Est	Sud-Ouest	Auditeurs du Sud-Ouest	Sud-Est	Sud-Ouest
Sud-Est	51,4	30,2	Sud-Est	27,7	47,1
Sud-Ouest	37,7	43,5	Sud-Ouest	23,9	64,9

Table 6. Matrices de confusion simplifiées pour les auditeurs du Sud-Est et du Sud-Ouest (%)

Un modèle linéaire généralisé a été appliqué, avec comme variable dépendante la réponse (VRAI/FAUX), l'origine des auditeurs (Sud-Est vs Sud-Ouest), l'origine des locuteurs (Sud-Est vs Sud-Ouest), la réponse sélectionnée (8 possibilités) et toutes les interactions simples, les stimulus et les auditeurs étant entrés comme variables aléatoires. Il a montré une absence d'effets simples : d'après nos résultats, les auditeurs du Sud-Est ne sont pas plus performants que les auditeurs du Sud-Ouest ; les locuteurs du Sud-Est ne sont pas mieux identifiés que ceux du Sud-Ouest, mais il y a une interaction significative entre l'origine et la réponse des auditeurs ( $\chi^2(7) = 71,692$  ;  $p < 0,0001$ ), ainsi qu'une interaction significative entre l'origine des locuteurs et la réponse des auditeurs ( $\chi^2(7) = 40,123$  ;  $p < 0,0001$ ). En résumé, il ressort que les auditeurs du Sud-Ouest sont plus performants que ceux du Sud-Est quand ils identifient des locuteurs du Sud-Ouest ( $p < 0,0001$ ), les premiers obtenant des scores de 64,9 % alors que les seconds atteignent 43,5 % seulement. À l'opposé, les auditeurs du Sud-Est sont meilleurs quand il s'agit d'identifier leur propre variété régionale en comparaison à celle du Sud-Ouest ( $p < 0,01$ ) : comparer 51,4 % avec 27,7 %. Par ailleurs, on constate que les auditeurs du Sud-Est ne sont pas significativement meilleurs lorsqu'il s'agit de distinguer entre les locuteurs du Sud-Ouest et ceux du Sud-Est : comparer 51,4 % avec 43,5 %. *A contrario*, les auditeurs du Sud-Ouest sont clairement meilleurs quand il s'agit de reconnaître leur propre variété à côté de celle du Sud-Est ( $p < 0,001$ ) : comparer 64,9 % avec 27,7 %.

Même si des différences ont pu être notées entre les français de Marseille et de Toulouse (Coquillon, 2005), ce dernier résultat met plutôt en avant des similitudes entre accents du Sud-Est et du Sud-Ouest, rejoignant en cela les conclusions d'autres études (Woehrling, 2009 ; Sichel-Bazin *et al.*, 2012 ; Boula de Mareüil *et al.*, 2017). Ces rapprochements ne sont pas sans rappeler ceux que nous avons mis en évidence au sein du Nord-Est et du Nord-Ouest, sur lesquels nous allons à présent revenir.

## Discussion

Dans cet article, nous avons présenté les résultats d'une enquête en ligne, à laquelle plus de 1 500 sujets originaires des quatre coins de la francophonie d'Europe ont pris part. L'analyse globale des tendances observées nous a permis d'apporter une première réponse, plutôt nuancée, à la question posée dans l'introduction, à savoir « peut-on identifier perceptivement des accents d'Alsace, de Belgique, de Bretagne, de Corse, du Sud-Est, du Sud-Ouest et de Suisse sur la base de la prononciation des locuteurs qui y demeurent ? » — la question corollaire étant « QUI peut identifier QUEL(S) accent(s) ? ».

Nos résultats ont montré que, globalement, ce sont plutôt trois grands groupes qui se distinguent : le Nord-Est (Alsace, Belgique et Suisse), le Nord-Ouest (Bretagne et Nord) et le Grand Sud (Sud-Est, Sud-Ouest et Corse). Si ce résultat confirme des expériences antérieures (Woehrling, 2009, *inter alia*) et les généralise, le nombre de participants que la méthode du *crowdsourcing* nous a permis de recruter apporte un éclairage nouveau, même s'il pouvait être attendu : il est notable que des locaux parviennent, parmi huit possibilités, à reconnaître l'accent de leur région à plus de 50 %. De façon intéressante, l'analyse par locuteur n'a pas permis de faire ressortir des accents plus prototypiques que d'autres (que l'on considère les réponses de l'ensemble des participants ou que l'on ne retienne que celles des locaux), et ce malgré la présence dans les enregistrements de schibboleths, comme le mot *moins*, le timbre de la voyelle /O/, etc. Une analyse acoustique des stimulus reste bien entendu à mener, locuteur par locuteur.

Les enquêtes en ligne, de par leur aspect ludique et leur potentiel médiatique, permettent de réduire la distance qui sépare la recherche universitaire du grand public, auprès duquel elles peuvent contribuer à écorner certains mythes en matière d'accents. Et d'un point de vue scientifique, elles permettent de recueillir une masse de données à des échelles qu'il n'aurait jamais été possible de collecter autrement, en si peu de temps et avec un nombre si réduit de chercheurs. Cela étant dit, ce travail met également en lumière le fait que le *crowdsourcing* n'est pas la panacée, que son usage en dialectologie perceptive demeure questionnable. Premièrement, malgré le format de l'enquête en ligne, le nombre de Corses ayant pris part au sondage demeure insuffisant par rapport aux participants des autres régions, raison pour laquelle il n'a pas toujours été possible de conduire des tests statistiques sur le groupe d'auditeurs corses. Ensuite, il faut souligner que le *crowdsourcing* ne permet pas de mener des enquêtes trop longues, car sinon les participants se lassent assez vite et ne terminent pas le sondage.

En conséquence, nous n'avons pas pu tenir compte, lors de nos analyses, de facteurs importants comme l'âge, le sexe ou le profil socioéconomique des locuteurs. Notre sélection de stimulus ne permettait pas d'avoir un panel représentatif au regard de ces catégories sociolinguistiques (traditionnellement retenues), et l'on est en droit de se demander si les résultats n'auraient pas été différents si ces catégories avaient été contrôlées lors de la sélection des échantillons sonores. Parallèlement, nous n'avons pas pu tenir compte des facteurs comme la mobilité des auditeurs. La méthode du questionnaire en ligne ne permet pas de cerner aisément, avec un petit nombre de questions, le parcours des participants, et partant leur familiarité avec divers accents régionaux. Là aussi, certains facteurs biographiques et sociolinguistiques auraient sans doute permis de mieux expliquer quelques-uns des résultats obtenus. Enfin, si nous n'avons pas été guidés par le souci de toucher un maximum de participants, par une expérience courte, nous aurions pu leur demander, avant le test, d'autoévaluer la capacité à distinguer différents accents. Souvent, il y a un décalage entre ce que les gens pensent pouvoir faire et ce qu'ils font réellement : les sujets imaginent pouvoir distinguer finement des accents régionaux (Woehrling, 2009 ; Avanzi & Boula de Mareüil, 2017), et l'espace perçu ne reflète pas exactement l'espace imaginé (Falkert, 2013). Outre la

rareté de certains indices de prononciation, une cause de cet écart entre les représentations et les performances réelles de sujets à qui on demande d'identifier l'origine d'échantillons de parole réside dans la persistance de certains mythes linguistiques (Hauchecorne & Ball, 1997 ; Boughton, 2006 ; Armstrong & Pooley, 2010). Certains clichés, certaines légendes ont la vie dure, et le discours hérité du passé montre une grande inertie, qui ne peut être ignorée quand on travaille sur les accents et leurs perceptions.

Il n'est pas rare que les aires périphériques se montrent plus conservatrices sur le plan linguistique — a fortiori les zones rurales, liées à l'agriculture, qui préservent mieux les particularismes locaux. Les métropoles, qui brassent les différences traditionnelles, sont davantage source d'innovation — ce qui n'est pas propre à l'Europe francophone (Milroy, 2001) : c'est la métaphore du *melting pot*, où les caractéristiques propres d'une mosaïque initiale se fondent dans un moule unique et s'effacent pour ne plus former qu'un seul et même ensemble. Aujourd'hui, avec l'urbanisation, l'école, la télévision, de nombreuses différences entre accents s'estompent, et on peut toujours s'interroger sur la « représentativité » des locuteurs sélectionnés. En dépit de ses limites, cette étude a malgré tout pu mettre en avant le fait que les accents de grandes villes cosmopolites comme Marseille et Bruxelles ne sont pas nécessairement moins bien identifiés que les accents de locuteurs originaires de petites villes ou de villages. Cette tendance reste à creuser.

Une certaine unification peut être observée parmi les accents français en Europe. En même temps, le développement de la mobilité au sein de la population peut favoriser de nouveaux changements. Si nous nous sommes concentrés ici sur des locuteurs non-mobiles, suivant une conception traditionnelle en dialectologie, cela ne signifie pas que nous nions cette réalité que notre société, depuis des décennies, encourage la mobilité. Les accents changent diachroniquement, comme le langage en général, selon une double dynamique verticale (de filiation) et horizontale (de contact). S'ils tendent à l'uniformisation aujourd'hui, sous la pression notamment de la norme véhiculée par les médias (ce, en particulier dans un pays centralisé comme la France), on peut gager qu'ils sauront se redistribuer, sinon à l'intérieur de territoires contigus, du moins en réseau, comme les accents de banlieue autour des grandes villes (Fagyal, 2010). Si certaines frontières géographiques ont été abolies, d'autres frontières, socioculturelles, se sont maintenues, et priment peut-être actuellement en matière de pratiques langagières. Il est possible que le paysage linguistique de la France, de la Belgique francophone et de la Suisse romande soit organisé davantage autour des différences sociales qu'autour des régions (Armstrong, 2001 ; Armstrong & Jamin, 2002 ; Gasquet-Cyrus & Trimaille, 2017). Une meilleure compréhension des clivages sociaux, des rapports de domination qui existent dans toute société est donc nécessaire pour mieux décrire les changements linguistiques et leur dimension diastratique.

## Remerciements

Mathieu Avanzi a reçu le financement du Fonds National de la Recherche Scientifique (subside n° 24901170). Nous remercions les relecteurs et les éditeurs du numéro pour leurs remarques sur une version préliminaire du travail, ainsi que locuteurs et les nombreux auditeurs qui ont rendu possibles ces enquêtes.

## Bibliographie

ARMSTRONG Nigel, 2001, *Social and stylistic variation in spoken French : a comparative approach*, John Benjamins, Amsterdam/New York.

- ARMSTRONG Nigel & BOUGHTON Zoé, 1997, « Identification and evaluation responses to a French accent: some results and issues of methodology », *Revue PARole*, n°5-6, pp. 7-60.
- ARMSTRONG Nigel & JAMIN Mickaël, 2002, « Le français des banlieues. Uniformity and discontinuity in French of the Hexagon », in K. Salhi (Ed.), *French in and out of France. Language Policies, intercultural antagonisms and dialogue*, Peter Lang, Francfort, pp. 107-136.
- ARMSTRONG Nigel & POOLEY Tim, 2010, *Social and linguistic change in European French*, Palgrave Macmillan, Basingstoke.
- AVANZI Mathieu, 2017, *Atlas du français de nos régions*, Armand Colin, Paris.
- AVANZI MATHIEU, BEGUELIN Marie-José & DIEMOZ Frederica, 2016, « De l'archive de parole au corpus de référence : la base de données orales du français de Suisse romande (OFROM) », *Corpus*, n°16, pp. 309-342.
- AVANZI Mathieu & BOULA DE MAREÛIL Philippe, 2017, « Identification of regional French accents in (northern) France, Belgium and Switzerland », *Journal of Linguistic Geography*, n°5/1, pp. 17-40.
- BAKER Wendy, EDDINGTON David & NAY Lindsey, 2009, « Dialect identification: the effects of region of origin and amount of experience », *American Speech*, n°84, pp. 48-71.
- BARDIAUX Alice, 2014, *La prosodie de quelques variétés de français en Belgique : analyse perceptive et acoustique*, Thèse de doctorat, Université catholique de Louvain, Louvain-la-Neuve.
- BATES Douglas M., MÄCHLER Martin, BOLKER Ben & WALKER Steven, 2013, *lme4: Linear mixed-effects models using Eigen and S4*, R package.
- BOLLY Catherine, CHRISTODOULIDES Georges & SIMON Anne Catherine, 2016, « Disfluences et vieillissement langagier. De la base de données VALIBEL aux corpus outillés en français parlé », *Corpus*, n°16, pp. 193-210.
- BOUGHTON Zoé, 2006, « When perception isn't reality: accent identification and perceptual dialectology in French », *Journal of French Language Studies*, n°16, pp. 277-304.
- BOULA DE MAREÛIL Philippe, 2010, *D'où viennent les accents régionaux ?*, Le Pommier, Paris.
- BOULA DE MAREÛIL Philippe & BOUTIN Béatrice A., 2011, « Évaluation et identification perceptives d'accents ouest-africains en français », *Journal of French Language Studies*, n°21, pp. 361-379.
- BOULA DE MAREÛIL Philippe & BARDIAUX Alice, 2011, « Perception of French, Belgian and Swiss accents by French and Belgian Listeners », *4<sup>th</sup> ISCA Tutorial and Research Workshop on Experimental Linguistics*, Paris, pp. 47-50.
- BOULA DE MAREÛIL Philippe, SCHERRER Yves & GOLDMAN Jean-Philippe, 2017, « Combien d'accents en français ? Focus sur la France, la Belgique et la Suisse », *Bulletin suisse de linguistique appliquée*, n°104, pp. 91-103.
- BOOMERSHINE Amanda, 2006, « Perceiving and processing dialectal variation in Spanish: an exemplar theory approach », in T. Face & C. Klee (Eds.), *Proceedings of the 8<sup>th</sup> Hispanic Linguistics Symposium*, Somerville, pp. 58-72.
- CARTON Fernand, ROSSI Mario, AUTESSERRE Denis, LEON Piere, 1983, *Les accents des Français*, Hachette, Paris.
- COQUILLON Annelise, 2005, *Caractérisation prosodique du parler de la région marseillaise*. Thèse de doctorat, Université de Provence, Aix-en-Provence.
- DURAND Jacques, LAKS Bernard, & LYCHE Chantal, 2009, *Phonologie, variation et accents du français*, Hermès, Paris.

- ESKENAZI Maxine, LEVOW Gina-Anne, MENG Helen, PARENT Gabriel, SUENDERMAN Duendermann, 2013, *Crowdsourcing for Speech Processing : Applications to Data Collection, Transcription and Assessment*, John Wiley & Sons, Ltd, Londres.
- FAGYAL Zsuzsanna, 2010, *Accents de banlieue : aspects prosodiques du français populaire en contact avec les langues de l'immigration*, L'Harmattan, Paris.
- FALKERT Anika, 2013, « La perception de la variation géolinguistique chez les Acadiens de la Côte-Nord (Québec) », in L. Rodriguez & A. Lapierre (dir.), *D'est en ouest : la variation géolinguistique du français au Canada*, Presses universitaires de Saint-Boniface, Winnipeg, pp. 39-53.
- GASQUET-CYRUS Médéric, & TRIMAILLE Cyril, 2017, « Être néo quelque part : la gentrification à Marseille et ses implications sociolinguistiques », *Langage et société*, n°162/4, pp. 81-105.
- GOLDMAN Jean-Philippe, AVANZI Mathieu, & SCHWAB Sandra, 2014, « Perception de l'accent régional dans cinq variétés de français parlées en Suisse romande », *Nouveaux cahiers de linguistique française*, n°31, pp. 285-296.
- GOOSKENS Charlotte, 2005, « How well can Norwegians identify their dialects? », *Nordic Journal of Linguistic*, n°28, pp. 37-60.
- HAUCHECORNE Fabrice & BALL Rodney, 1997, « L'accent du Havre : un exemple de mythe linguistique », *Langage et société*, n°82, pp. 5-25.
- INNACCARO Gabriele & DELL'AQUILA Vittorio, 2001, « Mapping languages from inside: notes on perceptual dialectology », *Social and Cultural Geography*, n°2/3, pp. 265-280.
- KUIPER Lawrence, 2005, « Perception is reality: Parisian and Provençal perceptions of regional varieties of French », *Journal of Sociolinguistics*, n°9, pp. 28-52.
- LÉON Pierre & LÉON Monique, 1997, *La prononciation du français*, Fernand Nathan, Paris.
- L'EPLATTENIER-SAUGY Caroline, 2002, « A perceptual Dialect Study of French in Switzerland », in D. Long & D. R. Preston (Eds.), *Handbook of Perceptual Dialectology*, John Benjamins, Amsterdam/New York, pp. 351-365.
- LIPPI-GREEN Rosina, 2012, *English with an Accent. Language, Ideology and Discrimination in the United States*, Routledge, New York.
- LYCHE Chantal, & SKATTUM Ingse, 2012, « The phonological characteristics of French in Bamako, Mali : A sociolinguistic approach », in R. Gess, C. Lyche, & T. Meisenburg (Eds.), *Phonological Variation in French : Illustrations from Three Continents*, Amsterdam/Philadelphia : John Benjamins, pp. 73-101.
- MILROY James, 2001, « Language ideologies and the consequences of standardization », *Journal of Sociolinguistics*, n°5/4, pp. 530-555.
- MOREAU Marie-Louise, BOUCHARD Pierre, DEMARTIN Stéphanie, GADET Françoise, GUERIN Emmanuelle, HARMEGNIES Bernard, HUET Kathy, LAROUCSI Foued, PRIKHODKINE Alexei, SINGY Pascal, THIAM Ndiassé, TYNE Henri, 2008, *Les accents dans la francophonie : une enquête internationale*, Service de la langue française, Bruxelles.
- PRESTON Denis R., 1989, *Perceptual dialectology*, Foris, Dordrecht.
- PRESTON Denis R., 2005, « What is folk linguistics ? What should you care ? », *Lingua Posnaniensis*, n°47, pp. 143-162.
- PRIKHODKINE Alexei, 2011, *Dynamique normative du français en usage en Suisse romande*, L'Harmattan, Paris.
- PUSTKA Elissa, 2007, *Phonologie et variétés en contact. Aveyronnais et Guadeloupéens à Paris*, Narr, Tübingen.

- R DEVELOPMENT CORE TEAM, 2018, *R : A language and environment for statistical computing*, Vienna (Austria), R Foundation for Statistical Computing.
- RACINE Isabelle, SCHWAB Sandra & DETEY Sylvain, 2013, « Accent(s) suisse(s) ou standard(s) suisse(s) ? Approche perceptive dans quatre régions de Suisse romande », in A. Falkert (dir.), *La perception des accents du français hors de France*, CIPA, Mons, pp. 41-59.
- REMYSEN Wim, 2014, « Les Québécois perçoivent-ils le français montréalais comme une variété topolectale distincte ? Résultats d'une analyse perceptuelle exploratoire », *Revue canadienne de linguistique*, n°59/1, pp. 109-135.
- REMYSEN Wim, 2016, « Langue et espace au Québec : les Québécois perçoivent-ils des accents régionaux ? », in D. Gavinelli & C. Molinari (Eds.), *Lingue, culture, mediazioni, (Espaces réels et imaginaires au Québec et en Acadie : enjeux culturels, linguistiques et géographique)*, LED, Milan, pp. 31-57.
- RUBIN Donald L., 1992, « Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants », *Research in Higher Education*, n°33, pp. 511-531.
- RUCH Hanna, 2018, « The Role of Acoustic Distance and Sociolinguistic Knowledge in Dialect Identification », *Frontiers in Psychology* n°9, p. 818.
- SERTLING-MILLER Jessica, 2007, *Swiss French Prosody: Intonation, Rate, and Speaking Style in the Vaud Canton*, PhD thesis, Illinois University, Urbana-Champaign.
- SICHEL-BAZIN Rafèu, BUTHKE Caroline, MEISENBURG Trudel, 2012, « La prosodie du français parlé à Lacaune : influences du substrat occitan », in A. C. Simon (éd.), *La variation prosodique régionale en français*, De Boeck, Bruxelles, pp. 137-157.
- SINGY Pascal, 1996, *L'image du français en Suisse romande. Une enquête sociolinguistique en Pays de Vaud*, L'Harmattan, Paris.
- VENABLES Bill & RIPLEY Brian D., 2002, *Modern Applied Statistics with S*, Springer, New York.
- YAN Qingyang, 2015, « The perceptual categorization of Enshi mandarin regional varieties », *Journal of Linguistic Geography*, n°3, pp. 1-19.
- WALTER Henriette, 1988, *Le français dans tous les sens*, Robert Laffont, Paris.
- WILLIAMS Angie, GARRETT Peter & COUPLAND Nikolas, 1999, « Dialect recognition », in D. Preston (Ed.), *Handbook of Perceptual Dialectology*, Benjamins, Philadelphia, pp. 345-358.
- WOEHLING Cécile, 2009, *Accents régionaux en français. Perception, analyse et modélisation à partir de grands corpus*, Thèse de doctorat, Université Paris Sud, Orsay.
- WOERHLING Cécile & BOULA DE MAREÜIL Philippe, 2006, *Identification d'accents régionaux en français : perception et analyse*, *Revue PArôle*, n°37, pp. 25-65.
- ZAMACOÏS Miguel, 1910, *La fleur merveilleuse*, pièce en 4 actes, en vers, Comédie française, Paris.

# GLOTTOPOL

Revue de sociolinguistique en ligne

**Comité de rédaction** : Michaël Abecassis, Salih Akin, Sophie Babault, Claude Caitucoli, Véronique Castellotti, Régine Delamotte, Robert Fournier, Stéphanie Galligani, Emmanuelle Huver, Normand Labrie, Foued Laroussi, Benoit Leblanc, Fabienne Leconte, Gudrun Ledegen, Danièle Moore, Clara Mortamet, Alioune Ndao, Isabelle Pierozak, Gisèle Prignitz.

**Rédactrice en chef** : Clara Mortamet.

**Comité scientifique** : Claudine Bavoux, Michel Beniamino, Jacqueline Billiez, Philippe Blanchet, Pierre Bouchard, Ahmed Boukous, Pierre Dumont, Jean-Michel Eloy, Françoise Gadet, Marie-Christine Hazaël-Massieux, Monica Heller, Caroline Juilliard, Jean-Marie Klinkenberg, Jean Le Du, Marinette Matthey, Jacques Maurais, Marie-Louise Moreau, Robert Nicolai, Lambert Félix Prudent, Ambroise Queffélec, Didier de Robillard, Paul Siblot, Claude Truchot, Daniel Véronique.

**Comité de lecture pour ce numéro** : Mickael Abecassis, Michelle Auzanneau, Annette Boudreau, Zoe Boughton, Zsuzsanna Fagyal, Françoise Gadet, Stéphanie Galligani, Marie-Noëlle Guillot, Philippe Hambye, Patricia Lambert, Gregory Miras, Tim Pooley, Wim Remysen.

Laboratoire Dylis – Université de Rouen  
<http://glottopol.univ-rouen.fr>

ISSN : 1769-7425