



**HAL**  
open science

## Peut-on distinguer perceptivement huit accents régionaux en français parlé en Europe? Une réponse à base de crowdsourcing

Mathieu Avanzi, Philippe Boula de Mareüil

### ► To cite this version:

Mathieu Avanzi, Philippe Boula de Mareüil. Peut-on distinguer perceptivement huit accents régionaux en français parlé en Europe? Une réponse à base de crowdsourcing. XXXIIe Journées d'Études sur la Parole (JEP 2018), 2018, Avignon, France. 10.21437/jep.2018-43 . hal-03321515

**HAL Id: hal-03321515**

**<https://hal.science/hal-03321515>**

Submitted on 17 Aug 2021

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

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



# Peut-on distinguer perceptivement huit accents régionaux en français parlé en Europe ? Une réponse à base de *crowdsourcing*

Mathieu Avanzi<sup>1</sup> & Philippe Boula de Mareuil<sup>2</sup>

(1) FNRS & VALIBEL, Univ. catholique de Louvain, Louvain-la-Neuve, Belgique

(2) LIMSI, CNRS & Univ. Paris-Saclay, Orsay, France

mathieu.avanzi@uclouvain.be ; philippe.boula.de.mareuil@limsi.fr

## RESUME

---

Nous présentons ici les résultats d'une enquête portant sur huit accents régionaux en français parlé en Europe : dans le nord, le sud-est et le sud-ouest de la France, en Alsace, Bretagne, Corse, Suisse et Belgique. Un test perceptif à grande échelle a été mené, à base de *crowdsourcing*, utilisant l'Internet pour recueillir les réponses de plus d'un millier de participants afin d'examiner si des auditeurs des régions ou pays susmentionnés sont capables de distinguer les accents correspondants. Les résultats suggèrent que ce sont plutôt trois grandes catégories d'accents (Nord-Ouest, Est et Sud) qui peuvent être discriminées.

## ABSTRACT

---

**Can eight regional accents be distinguished perceptually in European French? A crowdsourcing-based answer**

This paper presents the results of a survey addressing eight regional accents in European French: in the North, the South-East and the South-West of France, in Alsace, Brittany, Corsica, Switzerland and Belgium. A large-scale perception test was conducted, based on crowdsourcing, using the Internet to collect the responses of a thousand of participants, to examine whether listeners from the regions or countries mentioned are capable of distinguishing the corresponding accents. Results suggest that three broad categories of accents (North-West, South and East) can be distinguished.

**MOTS-CLES :** accents régionaux, géolinguistique, dialectologie perceptive, *crowdsourcing*

**KEYWORDS:** regional accents, geolinguistics, perceptual dialectology, crowdsourcing

---

## 1 Introduction

L'espace figure parmi les tout premiers facteurs pris en compte, tant par la dialectologie traditionnelle que par la sociolinguistique variationniste, pour expliquer l'hétérogénéité des pratiques langagières (Gadet, 2007 ; Mufwene & Vigouroux, 2012). Il nous donne des étiquettes pour dénommer aussi bien les langues (ou dialectes) que les accents, par exemple en français : ainsi va-t-on parler d'alsacien, de corse et d'accents corse ou alsacien en français, même si coexistent en Alsace des parlers alémaniques et franciques, en Corse des parlers d'origine toscane ou génoise. L'accent alsacien est-il distinct d'accents belges ou suisses ; l'accent corse est-il distinct d'accents continentaux du Midi de la France ? Les aires des dialectes et des accents ne coïncident pas nécessairement ; si les dialectes sont assez bien documentés et cartographiés depuis la fin du XIX<sup>e</sup> siècle (Gilliéron & Edmont, 1902–1910), les accents à l'intérieur de la langue

nationale (définis comme des ensembles de traits de prononciation liés à l'origine linguistique, géographique ou sociale des locuteurs d'une langue plus ou moins standardisée) n'ont suscité l'intérêt des linguistes que plus récemment.

Les accents régionaux, définis avant tout par la perception et les représentations qu'on en a, ont notamment été investis par la dialectologie perceptive (Preston, 1989 ; Innàccaro & Dell'Aquila, 2001, *inter alia*), dont un des objectifs est d'étudier dans quelle mesure différentes variétés d'une même langue peuvent être identifiées et catégorisées. Cette approche a été développée pour le français, avec des tests de perception incluant 5 points d'enquête en France (Woehrling & Boula de Mareüil, 2006), 3 en Belgique (Boula de Mareüil & Bardiaux, 2011), 4 en Suisse (Racine *et al.* 2013), 4 en Afrique de l'Ouest (Boula de Mareüil & Boutin, 2011), 8 au Québec (Remysen, 2016), mais chaque expérience était assez limitée, n'impliquant que quelques dizaines d'auditeurs. Une méthodologie à base de *crowdsourcing* (production participative à grande échelle) a depuis permis de passer à quelques centaines voire quelques milliers d'auditeurs, auxquels étaient présentés des accents français d'Amérique du Nord, d'Afrique et d'Europe (Boula de Mareüil *et al.*, 2017), mais chaque variété était représentée par un seul locuteur.

D'autres expériences ont été conduites, centrées sur la France, la Suisse et la Belgique, touchant également une large audience, à travers des sites grand public ou les réseaux sociaux (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 (de 5 régions différentes dans chacun de ces pays) ont été présentés à des centaines d'auditeurs de ces mêmes pays. Dans une première série d'expériences, les auditeurs devaient identifier le pays d'origine des locuteurs : ils ont obtenu 60 % d'identification correcte en moyenne, avec des effets significatifs de l'origine des auditeurs, de l'âge, du statut socio-économique et de l'origine des locuteurs. 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 trois pays (avec un choix forcé entre 5 possibilités). Les résultats, quoiqu'au-dessus du niveau du hasard, se sont révélés moins bons, avec 31 % d'identification correcte en moyenne. Des analyses complémentaires ont été conduites pour évaluer le rôle de l'origine des auditeurs, de l'âge et de l'origine des locuteurs, ainsi que de leur interaction. Elles ont montré des patrons de réponses asymétriques entre les trois pays étudiés : la France (ou, en France, Paris), semble agir comme un pôle d'attraction et un catalyseur d'unification. Les locuteurs les plus jeunes, en particulier, sont plus souvent associés à sa façon de parler, quand leur accent n'est pas clairement identifiable. La Suisse, cependant, semble mieux résister que la Belgique à ce processus d'homogénéisation.

Dans l'étude que nous venons de résumer, toutefois, ne figurait ni l'accent alsacien, qui peut être confondu avec les accents belges et suisses (Woehrling, 2009), ni les accents méridionaux du sud-est et du sud-ouest de la France, ou encore de Corse. Dans cet article, nous nous proposons de réfléchir à une question en apparence simple : peut-on, pour le français parlé en Europe, distinguer les huit accents suivants : Nord, Bretagne, Alsace, Suisse, Belgique, Sud-Ouest, Sud-Est et Corse ? Comme tout dépend du *on*, il est nécessaire d'avoir un nombre suffisant d'auditeurs de chacune de ces régions (ou chacun de ces pays), d'où l'importance du *crowdsourcing*. Pour répondre à cette question, nous présenterons le corpus ainsi que la méthode employés (section 2) et les résultats obtenus, globalement et en fonction de l'origine des auditeurs (section 3), avant de conclure (section 4).

## 2 Corpus et méthode

Outre la Suisse et la Belgique, six régions ont été retenues dans cette étude : le Nord, la Bretagne à l'Ouest, l'Alsace à l'Est, le Sud-Ouest, le Sud-Est et la Corse. Des traits de prononciation, a priori, peuvent en effet distinguer le français parlé dans ces régions (Avanzi, 2017) : nous nous contenterons de citer le mot *moins*, dont la consonne finale tend à se faire entendre dans le Sud-Ouest. Pour la France, les enregistrements utilisés dans cette étude 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). Souvent âgés, ces locuteurs étaient comme les Belges et les Suisses bien ancrés dans la région où ils ont été enregistrés, y ayant passé la plus grande partie de leur vie. Pour chacune des huit régions étudiées, nous avons sélectionné quatre locuteurs, à partir desquels nous avons extrait des segments de parole d'une durée comprise entre 10 et 15 secondes, en veillant 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 (*euh* d'hésitation, répétitions, etc.). L'expérience d'identification a été mise au point avec le logiciel Qualtrics, lequel permet de faire des sondages en ligne, à partir de n'importe quel navigateur, sur ordinateur, téléphone portable multifonction ou tablette. Les participants, contactés via les réseaux universitaires et les réseaux sociaux, devaient 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. Pour ce faire, ils devaient appuyer sur le bouton du lecteur multimédia intégré, et cliquer, après écoute du stimulus, n'importe où à l'intérieur de la région colorée de la carte reproduite dans la Figure 1. 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.

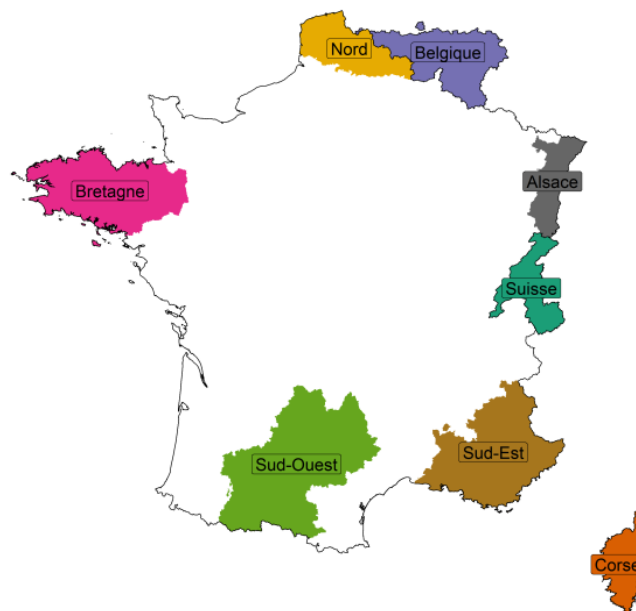


FIGURE 1: Carte présentant les régions et leurs labels proposée aux auditeurs lors de l'expérience.

Pour éviter que l'enquête ne soit trop longue, dans le but de recueillir les réponses d'un maximum de participants, seuls 16 extraits sonores parmi les 32 que comptait l'expérience étaient présentés à chaque sujet. Les stimuli étaient sélectionnés semi-aléatoirement, un algorithme permettant que

chacun des fichiers sons soit joué le même nombre de fois. À la fin de l'expérience, il était demandé aux participants d'indiquer leur âge, leur sexe, le pays et le code postal de la localité dans laquelle ils avaient passé la plus grande partie de leur jeunesse. Un champ libre leur permettait d'en dire plus quant à leur connaissance des accents régionaux. Au total, 1575 francophones ayant déclaré avoir passé la plus grande partie de leur jeunesse en France, Suisse ou Belgique ont pris part à cette expérience. 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, nous avons codé l'origine géographique des participants dans notre base de données. À cette fin, nous avons considéré la Suisse et la Belgique comme des régions à part entière. 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 : pour l'Alsace, la Bretagne et la Corse, nous avons considéré les participants originaires des régions éponymes ; pour le Sud-Ouest, nous avons considéré les participants originaires de Midi-Pyrénées, pour le Nord ceux de Nord-Pas-de-Calais et pour le Sud-Est ceux de Provence-Alpes-Côte-D'azur. Les sujets se répartissent comme rapporté dans la Table 1.

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

TABLE 1 : Participants à l'enquête, avec, leur région d'origine, leur nombre, le pourcentage par rapport au total, leur sexe et leur âge.

Il est intéressant de constater que l'échantillon est relativement bien équilibré en ce qui concerne le sexe des participants, et que malgré le format de l'expérience (accessible seulement en ligne et diffusée majoritairement sur 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.

### 3 Résultats

#### 3.1 Statistiques

Les analyses statistiques ont été conduites avec le logiciel R version 3.3.2 (R Development Core Team, 2016). Pour vérifier si les variétés ont été identifiées au-dessus du niveau du hasard, nous avons réalisés 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* — R package *lme4* (Bates *et al.*, 2013) —, dans lesquels nous avons inclus les locuteurs et les auditeurs comme effets 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 (MDS) de la librairie *MASS* (Venables & Ripley, 2002). Tous les graphiques ont été réalisés avec le package *ggplot2*.

### 3.2 Résultats globaux

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 obtenir la matrice de confusion de la Table 2.

		origine prédite							
		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	<b>21,8</b>	21,0	8,1
	Nord	10,8	5,8	<b>37,2</b>	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>	<b>42,0</b>	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 2 : Matrice de confusion, tous participants et toutes régions confondus (%).

En moyenne, avec 35,9 % d'identification correcte, les huit régions sont reconnues bien mieux qu'au hasard (12,5 %, tous les tests *t* donnant des valeurs de  $p < 0,001$ ). Pour vérifier si les différences entre les scores obtenus 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 stimuli et les auditeurs étant les variables aléatoires. Les résultats de ce test 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 : 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 cette interaction significative, 8 modèles distincts ont 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 la variable « réponse des auditeurs » était toujours significative (tous les modèles mettent en évidence un effet significatif de réponse, avec des valeurs de  $p < 0,0001$ ). Pour rendre plus fluide la lecture, nous ne commenterons ici que les différences impliquant les pourcentages de réponses les plus élevés pour chaque région (en grisé dans la Table 2, les valeurs correctes sont en gras).

Pour l'Alsace, la Belgique, la Bretagne, le Sud-Ouest et la Suisse, les valeurs les plus hautes (celles en grisé dans la table 2) 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 (21,0 %). 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 est significativement différent de tous les autres scores à l'intérieur de cette même région, sauf pour la Corse (les Corses étant confondus avec des locuteurs du Sud-Est et du Sud-Ouest), pour le Nord (où c'est la réponse Bretagne qui arrive en tête des suffrages) et pour 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 (non représenté ici par manque de place) permet de visualiser. Les trois variétés du Sud-Ouest, du Sud-Est et de Corse forment un premier groupe : elles ont souvent été confondues entre elles, la différence entre la Corse et les deux régions du sud du continent étant 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) et les variétés du nord-ouest de la France (Nord et Bretagne).

### 3.3 Effet de l'origine des auditeurs

Dans un second temps, nous avons examiné 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. Pour plus de clarté, nous avons isolé les réponses obtenues pour les locuteurs de chacune des huit régions et avons comparé les scores des auditeurs habitant ces régions à ceux qui n'y habitaient pas : 8 modèles linéaires généralisés avec la réponse VRAI/FAUX comme variable dépendante, l'interaction avec l'origine des auditeurs (locaux ou non-locaux) et leur réponse (8 possibilités), les stimuli et les auditeurs étant entrés comme variables aléatoires. Les scores d'identification correcte des locaux et des autres, pour chacune des huit régions, sont données dans la Table 3. Toutes les différences sont significatives ( $p < 0,001$ , hormis pour la Corse, où aucun test statistique n'a pu être mené, compte tenu du trop faible nombre de participants — les valeurs sont donc données à titre indicatif. Nous avons ensuite réalisé un MDS sur la base des réponses des locaux uniquement, pour chacune des huit régions, ce qui nous a permis de visualiser aisément les confusions commises par les auditeurs.

On voit dans la Table 3 que les locaux sont meilleurs que les autres pour reconnaître l'accent de leur région. Si l'on compare à présent les 8 MDS présentés dans la Figure 2, 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 participants du Nord, mais également des participants du Sud.

	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 3 : Taux d'identification correcte de chaque région, selon l'origine des auditeurs (%).

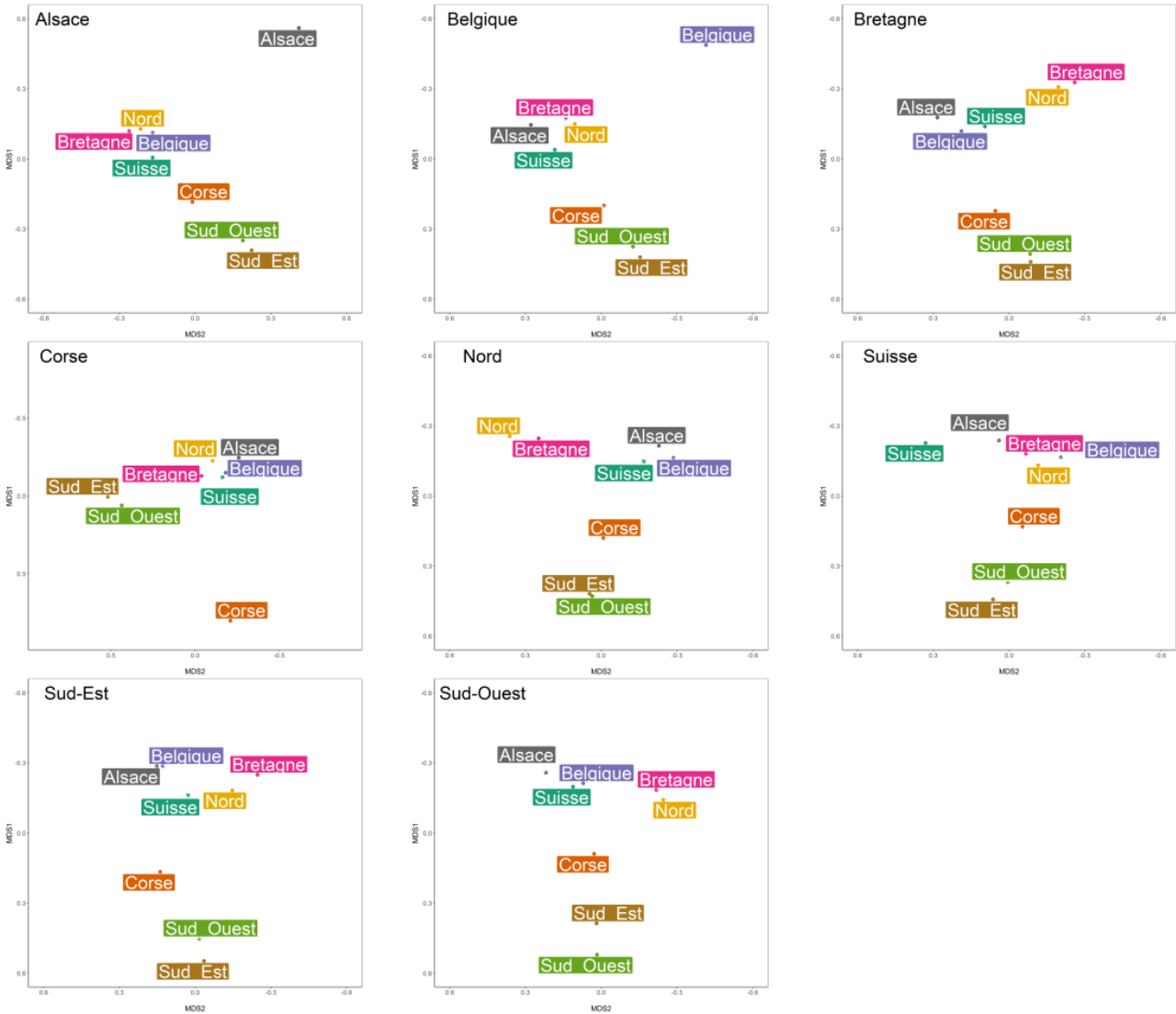


FIGURE 2 : Echelonnements multidimensionnels pour les huit régions étudiées, d'après les réponses obtenues par les locaux (leur origine étant à chaque fois indiquée en haut à gauche).

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

Les résultats présentés dans la section 3.2 ont permis de mettre le doigt sur des confusions constantes entre le Sud-Est et le Sud-Ouest ; ceux de la section 3.3 ont révélé que ces confusions sont également faites par les participants de ces régions. Dans cette dernière partie, nous avons cherché à évaluer et comparer les capacités des auditeurs du sud de la France 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 4 : 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 participants (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 stimuli et les auditeurs étant entrés comme variables aléatoires). Il a montré une absence d'effets simples : 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 ( $p<0,0001$ ) : comparer 64,9 % avec 43,5 %. A contrario, 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 faire la part 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 %.

## 4 Conclusion

Pour reprendre la question posée dans le titre, « peut-on perceptivement identifier des accents d'Alsace, de Belgique, de Bretagne, de Corse, du Sud-Est, du Sud-Ouest et de Suisse », la réponse se doit d'être nuancée. Globalement, ce sont plutôt trois grands groupes qui se distinguent : Nord-Est (incluant la Suisse et la Belgique), Nord-Ouest et Sud (incluant la Corse). Si ce résultat confirme des expériences antérieures (Woehrling, 2009, *inter alia*) et les généralise, le nombre de participants que le *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 %. En dépit du grand nombre d'auditeurs qui ont pris part à ce test d'identification, ce travail met également en lumière le fait que le *crowdsourcing* n'est pas la panacée : le nombre de Corses reste insuffisant, suggérant que cette approche ne doit pas se substituer à des enquêtes de terrain. Au contraire, les deux démarches sont complémentaires et peuvent se féconder l'une l'autre. Dans tous les cas, une analyse des stimuli est à mener, de même qu'une analyse des résultats par locuteur, afin de cerner si des prototypes se dégagent pour les différents accents.

## Remerciements

Mathieu Avanzi a reçu le financement du Fonds National de la Recherche Scientifique (subside n° 24901170). Ce travail s'inscrit également dans les activités du projet « Donnez votre français à la science » (DFS) du programme « Langues et Numérique » 2016 de la Délégation Générale à la Langue Française et aux Langues de France (DGLFLF). Nous remercions les locuteurs et les nombreux auditeurs qui ont rendu possible ce travail.

## Références

- AVANZI, M. (2017). *Atlas du français de nos régions*. Paris : Armand Colin.
- AVANZI, M. & BOULA DE MAREÛIL, P. (2017). Identification of regional French accents in (northern) France, Belgium and Switzerland. *Journal of Linguistic Geography*, 5/1, 17–40.
- BATES, D. M., MÄCHLER, M., BOLKER, B. & WALKER, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. *R package*.
- BOULA DE MAREÛIL, P. & AKISSI BOUTIN, B. (2011). Évaluation et identification perceptives d'accents ouest-africains en français. *Journal of French Language Studies*, 21, 361–379.
- BOULA DE MAREÛIL, P. & BARDIAUX, A. (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 47–50.
- BOULA DE MAREÛIL, P., SCHERRER, Y. & GOLDMAN, J.-P. (2017). Combien d'accents en français ? Focus sur la France, la Belgique et la Suisse. *Bulletin suisse de linguistique appliquée*, 104, 91–103.
- GADET, F. (2007). *La variation sociale en français*. Paris/Gap : Ophrys.
- GILLIÉRON, J. & EDMONT, E. (1902–1910). *Atlas linguistique de la France (ALF)*. Paris : Champion.
- INNÀCCARO, G. & DELL'AQUILA, V. (2001). Mapping languages from inside: notes on perceptual dialectology. *Social and Cultural Geography*, 2/3, 265–280.
- MUFWENE, S. & VIGOUROUX, C. (2012). Individuals, populations, and timespace: Perspectives on the ecology of language *Cahiers de linguistique*, 38/2, 111–138.
- PRESTON, D. R. (1989). *Perceptual dialectology*. Dordrecht : Foris.
- RACINE, I., SCHWAB, S. & DETEY, S. (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*. Mons : CIPA, 41–59.
- R DEVELOPMENT CORE TEAM. (2016). R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
- REMYSEN, W. (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)*. Milan : LED, 31–57.
- VENABLES, B. & RIPLEY, B. D. (2002). *Modern Applied Statistics with S*. New York : Springer.
- WOERHLING, C. (2009). *Accents régionaux en français. Perception, analyse et modélisation à partir de grands corpus*. Thèse de doctorat de l'Université Paris Sud, Orsay.
- WOERHLING, C. & BOULA DE MAREÛIL, P. (2006). Identification d'accents régionaux en français : perception et analyse. *Revue Parole*, 37, 25–65.