



HAL
open science

L'apport des données participatives pour l'étude linguistique des français du monde : le cas de l'opposition /a~/

Mathilde Hutin, Marc Allasonnière-Tang

► **To cite this version:**

Mathilde Hutin, Marc Allasonnière-Tang. L'apport des données participatives pour l'étude linguistique des français du monde : le cas de l'opposition /a~/ . *Journal of French Language Studies*, 2023, pp.1-24. 10.1017/s0959269523000200 . hal-04299068

HAL Id: hal-04299068

<https://hal.science/hal-04299068>

Submitted on 22 Nov 2023

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.

RESEARCH NOTE

L'apport des données participatives pour l'étude linguistique des français du monde : le cas de l'opposition /a~ɑ/

Mathilde Hutin¹  and Marc Allasonnière-Tang²

¹F.R.S.-FNRS, Université Catholique de Louvain, Louvain-la-Neuve, Belgique and ²CNRS, MNHN, EA (UMR 7206), Paris, France

Emails: mathilde.hutin@uclouvain.be, marc.allasonniere-tang@mnhn.fr

(Received 10 January 2023; revised 19 June 2023; accepted 21 June 2023)

ABSTRACT

French is a language spoken by hundreds of millions of speakers in Europe, Africa, and America. Such widespread use favours variation, yet large homogeneous corpora allowing to account for this variation worldwide are scarce and would in any case necessitate non-negligible financial and human resources, as did for instance the project *Phonologie du Français Contemporain*. In this study, we present a possible alternative – crowdsourcing. We introduce *Lingua Libre*, Wikimedia's open linguistic library, and use it to describe the variation of a phonemic opposition between two low vowels, /a/ and /ɑ/, in several varieties of French. The recordings of 38 speakers from 26 survey points are processed automatically and compared to values from past research. Results show that the platform has the potential to provide results mostly congruent with those of professional field recordings. The study concludes on the advantages and limitations of the platform and proposes suggestions for its improvement.

Keywords: Crowdsourcing; French; dialectal variation; phonology; phonetics; corpus linguistics

RÉSUMÉ

Le français est une langue parlée par plusieurs centaines de millions de locuteurs en Europe, en Afrique et en Amérique. Une telle dispersion favorise la variation, mais de grands corpus unifiés permettant de rendre compte de cette variation à l'échelle mondiale restent rares, et dans tous les cas nécessitent des efforts financiers et humains non négligeables, à l'instar du projet de *Phonologie du Français Contemporain*. Dans cet article, nous présentons une alternative possible : les données participatives. Pour ce faire, nous présentons *Lingua Libre*, la médiathèque linguistique participative de Wikimedia France, et l'utilisons pour décrire la variation sur une opposition phonémique entre deux voyelles ouvertes, /a/ et /ɑ/, dans de nombreuses variétés de français. Les données de 38 locuteurs provenant de 26 points d'enquête sont traitées automatiquement et comparées aux mesures présentées dans la littérature passée. Les résultats montrent que la plateforme a le potentiel de donner des résultats conformes à ceux des études de terrain professionnelles. L'article conclut sur les avantages et les limites de la plateforme, tout en proposant des pistes d'amélioration.

1. Introduction

Le français est une langue romane occidentale parlée par environ 300 millions de locuteurs courants à travers le monde¹ et officielle dans une trentaine de pays. La particularité de la répartition géographique de cette langue est son caractère discontinu : elle est traditionnellement parlée dans le bloc européen (Belgique, Luxembourg, France, Suisse et Monaco), mais aussi en Amérique du Nord (dans plusieurs régions du Canada et des États-Unis), et en Afrique (du Nord comme subsaharienne) (Smith 2016). Or les grands territoires, à fortiori lorsqu'ils favorisent le contact avec de nombreuses autres langues, favorisent la variation linguistique à tous les niveaux : lexical, bien entendu, mais aussi syntaxique et, pour ce qui nous intéresse, phonétique et phonologique.

Étudier une telle variation représente un défi de taille pour les dialectologues, qui restent tributaires des données. Ces dernières viennent essentiellement d'enquêtes de terrain, dont beaucoup, pour le français, ont été récoltées dans le cadre du programme de recherche *Phonologie du Français Contemporain* (PFC, Durand et al. 2002). Pourtant, avec l'avènement des nouvelles technologies, on est en droit de se demander si une nouvelle approche n'est pas possible. En effet, outre les enregistrements de voix « classiques » qui ont permis la création de corpus comme PFC (49 enquêtes de terrain en 2019), BREF (120h de journaux lus à haute voix, Lamel et al. 1991) ou NCCFr (31h de parole informelle, Torreira et al. 2010), les types de récolte se sont diversifiés. La possibilité d'enregistrer des conversations téléphoniques a donné lieu à des corpus comme MASK (35h de parole spontanée, Lamel et al. 1995), les émissions de télévision et de radio à des corpus comme ESTER (80h de parole journalistique, Galliano et al. 2005) ou ETAPE (42,5h de débat, Gravier et al. 2012), et la possibilité d'enregistrer les manifestations publiques à des corpus comme AssNat (126h de débat à l'Assemblée nationale en France et au Québec, ANF 2011, ANQ 2011), entre autres.

Plus récemment encore, l'accès de plus en plus généralisé à Internet semble répondre à la contrainte d'étendue géographique dans l'étude du français. En particulier, les outils participatifs, dits aussi collaboratifs, permettent de centraliser les compétences, les connaissances, et le bon-vouloir du grand public, donc à moindre coût, tant en argent qu'en énergie, pour les chercheurs. L'usage d'Internet pour mener des enquêtes à grande échelle a fait ses preuves dans de nombreux domaines, en premier lieu la lexicographie (et dans une moindre mesure la grammaire et la phonologie), avec par exemple les enquêtes en ligne du *Français de nos régions*² (Avanzi et al. 2016), qui se fondent essentiellement sur des tâches de jugement. Avec les micros intégrés aujourd'hui à presque tous les ordinateurs et l'accès à Internet depuis les téléphones portables, Internet peut même être utilisé pour récolter des données vocales, ce qui ouvre la porte à un nouveau type d'étude en phonologie et même, en phonétique (pour une revue des avantages, inconvénients et défis de l'usage du smartphone et d'Internet en linguistique, voir Hilton & Leeman 2021).

¹Ces chiffres proviennent du recensement opéré par l'Organisation Internationale de la Francophonie datant de 2018 mais ne distingue pas les locuteurs éduqués en français et les locuteurs proprement natifs.

²<https://francaisdenosregions.com>

Notre étude vise donc à explorer des données participatives, en l'occurrence en explorant la réalisation de l'opposition /a~ɑ/ dans une tâche de lecture de mots isolés. Les données participatives permettent-elles des études linguistiques fiables ? Nous comparerons autant que faire se peut les analyses et la finesse des résultats avec les recherches existantes portant sur des données provenant d'enquêtes de terrain faites par des spécialistes.

L'article se présente comme suit. Dans un premier temps, nous revenons sur l'opposition /a~ɑ/ dans les diverses variétés de français et dressons un tableau détaillé pour étayer nos propres analyses. Dans la section 3, nous présentons notre corpus, *Lingua Libre*, et notre méthodologie. La section 4 est consacrée à l'analyse fondée sur les données extraites de *Lingua Libre*. Enfin, la cinquième et dernière section est consacrée à la conclusion et à la discussion des résultats.

2. L'Opposition /a~ɑ/ dans les Différentes Variétés de Français à travers le Monde

2.1. La voyelle ouverte à double timbre : problème de définition

Un des points de variation phonétique du français concerne la réalisation de ses voyelles, qui traduit, parfois, une distinction d'ordre phonologique entre ses catégories vocaliques. En effet, selon les variétés, on dénombre entre 14 voyelles phonologiques (en français parisien par exemple) et 23 (en français laurentien) (Lyche 2010, Tchkhovrebova 2019). Dans ce qui suit, nous nous intéressons à la distinction entre /a/, voyelle traditionnellement notée comme ouverte antérieure voire centrale, et /ɑ/, voyelle traditionnellement notée comme ouverte postérieure, donnant lieu, dans les variétés où elle est opérationnelle, à des paires minimales de type « patte~pâte ».

L'opposition /a~ɑ/, qui aurait émergé de l'évolution d'une ancienne opposition entre un /a/ bref et un /a:/ long (Martinet & Walter 1973), est difficile à cerner. Martinet & Walter (1973) constatent notamment un manque d'uniformité dans la distribution de cette opposition parmi leurs 17 locuteurs et locutrices du français parisien, et l'opposition provoque la confusion chez de nombreux auteurs³. On appelle parfois l'archiphonème /A/ une « voyelle à double timbre », au même titre que /E/, qui regroupe /e/ et /ɛ/, /O/, qui regroupe /o/ et /ɔ/ et /œ/, qui regroupe /œ/ et /ø/⁴ (cf. Hansen & Juillard 2011). Or la définition même comme « voyelle à double timbre » est trompeuse : elle se fonde sur la loi de position, qui suggère que l'opposition est neutralisée dans certains types de syllabe. Ainsi, supposément, la distinction /a~ɑ/ serait, dans les variétés où elle est opérationnelle, distinctive en syllabe fermée, comme en (1), mais pas en syllabe ouverte, comme en (2).

(1) « patte » /pat/ vs « pâte » /pat/

(2) « bas » [ba] ou [ba]

³« Some phoneticians claim that there are two distinct as in French, but evidence from speaker to speaker and sometimes within the speech of a single speaker is too contradictory to give empirical support to this claim ». (Casagrande 1984, p. 20)

⁴On pourrait ajouter également l'archiphonème /Ē/ qui regrouperait /ê/ et /ë/ (Léon 1992).

En réalité, comme nous le montrons ci-dessous en Section 2.2, la structure syllabique joue certes dans certaines variétés un rôle dans la distribution de /a/ et /ɑ/, mais l'opposition peut aussi se décrire, dans d'autres variétés, comme une distinction phonémique classique, valable dans toutes les positions, comme en (3).

(3) « bas » /ba/ vs « bât » /bɑ/

Les membres de cette opposition auraient fusionné en une seule et même catégorie en français parisien, comme /ɛ̃~œ̃/ (par exemple dans la paire minimale « brin~brun ») (Léon 1992). La distinction entre /a/ et /ɑ/ était encore assez présente en français métropolitain jusque dans les années 1940 : Martinet (1945, p. 76), dans son étude sur la prononciation des officiers français dans un camp de prisonniers en Allemagne pendant la Seconde Guerre Mondiale, démontre que la distinction était encore faite à 90% au moins par les locuteurs, bien que la tendance soit à la baisse (99% chez les plus âgés, 96% chez les locuteurs d'âge moyen, et 92% chez les plus jeunes). Jusque dans les années 1950 au moins, la distinction est aussi un marqueur social : Reichstein (1960) compare la prononciation de quelques variables phonologiques du français, telles que les oppositions /a~ɑ/, /e~ɛ/ et /ɛ̃~œ̃/, et montre que, en général, la prononciation des enfants des classes privilégiées se conforme significativement plus à la norme de prestige, qui maintient l'opposition, que celle des enfants des autres couches sociales. Les études de Deyhime (1967a, b) ou Martinet (1969) confirment qu'à cette époque, la distinction, qui n'a lieu que dans les syllabes finales fermées, continue de se perdre dans le français parisien. Armstrong (2001), Hansen & Juillard (2011) et Østby (2016) confirment qu'elle se perd davantage encore jusque dans les années 2000 et 2010, y compris dans la haute bourgeoisie parisienne, pourtant conservatrice.

2.2. La variété des réalisations à travers le monde

La situation dans les français du monde est extrêmement complexe. Comme le montrent les paragraphes qui suivent, certaines variétés n'ont pas ou plus l'opposition phonémique (notamment en France Métropolitaine, en Afrique et possiblement en Louisiane), d'autres ont conservé une opposition essentiellement positionnelle (dans le Nord-Est de la France), d'autres encore ont conservé l'opposition uniquement en syllabe fermée (dans les campagnes françaises, en Belgique, dans le canton de Neuchâtel en Suisse, et au Québec et au Nouveau Brunswick au Canada) et d'autres enfin ont maintenu l'opposition dans toutes les positions (dans le Nord-Ouest de la France, dans le canton de Vaud en Suisse et en Alberta au Canada). Cependant, l'opposition, lorsqu'elle a lieu, ne se manifeste pas toujours de la même manière : elle est implémentée, selon les variétés, tantôt par une distinction sur l'axe antéro-postérieur, à divers degrés selon les langues, parfois avec de la diphthongaison, tantôt par la longueur, tantôt par les deux. Enfin, des facteurs sociolinguistiques, notamment l'âge, le milieu social ou le style de parole, influencent parfois les réalisations. Les paragraphes qui suivent tentent de dresser un état des lieux aussi détaillé que possible et de montrer par la même occasion la complexité de la situation et l'importance d'avoir accès à des données propres et nombreuses.

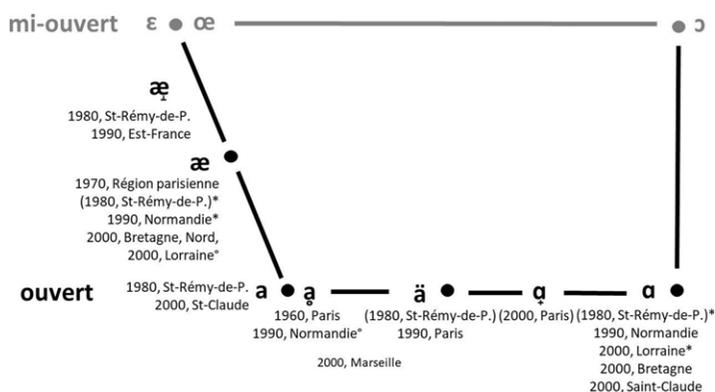


Figure 1. Réalisations possibles de /a/ et /a/ en français de France dans la deuxième moitié du XXe siècle : dans les années 1960 à Paris (Martinet 1969), 1970 en région parisienne (François 1974, Lennig 1978, Mettas 1979), 1980 à Saint-Rémy-de-Provence dans les Bouches-du-Rhône (Hilt 1986), 1990 à Paris (Fougeron & Smith 1999), en Normandie (Hauchecorne & Ball 1997) et dans l'est de la France (Armstrong 1993), 2000 dans le Nord et l'Alsace (Woehrling 2009), en Lorraine et en Bretagne (Boughton 2005), à Paris (Hansen 2012), à Saint-Claude dans le Haut-Jura (Arnaud 2006) et à Marseille (Coquillon & Turcsan 2012). Les voyelles orales mi-ouvertes du français sont indiquées en gris à titre de repère. Les encarts spécifient pour chaque réalisation la variété où elle a été observée : les variantes entre parenthèses sont des variantes rares, les variantes avec * ont été observées essentiellement en syllabe fermée et les variantes avec ° ont été observées essentiellement après /w/.

Pour faciliter la lecture des premiers paragraphes ci-après, consacrés au français de France, on peut se reporter au graphique ci-dessus, qui montre à quel point la variété des réalisations phonétiques complique l'observation de l'opposition phonémique (Figure 1).

Selon Walter (1988), jusqu'au milieu du XXe siècle, l'opposition est maintenue en français métropolitain entre un /a/ réalisé à l'avant, et un /a/ réalisé postérieur, et probablement arrondi. Aujourd'hui, sur l'ensemble de la France, d'après Picoche & Marchello-Nizia (1994 : 208), l'opposition /a~ɑ/ serait parfois maintenue, mais essentiellement dans des monosyllabes. Pour la majeure partie des locuteurs du français standard de France, cependant, à l'instar des locuteurs parisiens, /a/ et /ɑ/ ont fusionné en une seule catégorie /a/ et la distinction phonémique ne fait surface que par pression normative, où cependant [ɑ] n'est jamais réalisé clairement postérieur mais seulement un peu plus postérieur ou un peu plus long que [a] (Hansen 2012). Bien qu'un petit nombre de locuteurs conservateurs, surtout âgés, maintienne l'opposition phonémique, la réalisation [ɑ] n'est plus guère utilisée que comme une variante allophonique de /a/ pour véhiculer ou tourner en ridicule une impression d'élégance ou de prétention. Martinet (1969 : 202–203) suggère que le /a/ parisien, dans les années 1960, est réalisé arrondi, [ɶ]. Cependant, selon une étude sur différents sociolectes (Coveney 2001, p. 76), la catégorie unique /a/ est réalisée antérieure, c'est-à-dire [æ]. Dans les années 1970, on aurait en effet prononcé un /a/ hérité du /æ/, c'est-à-dire plus avancé et plus haut, chez les descendants de l'aristocratie parisienne (Mettas 1979) et chez les Parisiens de classe moyenne (Lennig 1978) ou encore à Argenteuil (François 1974). Plus récemment, cette réalisation [æ] est attestée chez les locuteurs du Nord et d'Alsace (Woehrling 2009 :

82), mais la description du français d'une jeune locutrice parisienne décrit la réalisation de /a/ comme centrale, [ä] (Fougeron & Smith 1999). Enfin, une étude sur les données du corpus PFC analyse les valeurs formantiques F1 et F2 et la durée dans l'opposition *patte~pâte* chez 83 locuteurs et confirme que, en général, le contraste entre /a/ et /a/ se perd en France métropolitaine (Berns 2015). Les données participatives d'Avanzi (2020), recueil de jugements métalinguistiques, confirment que l'opposition *patte~pâte* ne serait réalisée que dans 27,9% des cas en France métropolitaine et que l'opposition se perd chez les plus jeunes.

Pour ce qui est d'autres variétés citadines du Nord de la France, Avanzi (2020) montre que l'opposition /a~a/ en syllabe fermée est surtout opérationnelle dans le quart Nord-Est de la France. Boughton (2005) compare les productions de locuteurs de Nancy, en Lorraine dans le Nord-Est, et celles des locuteurs de Rennes, en Bretagne, c'est-à-dire dans le Nord-Ouest. En Lorraine, la distinction /a~a/ est maintenue sur des critères positionnels, mais l'opposition ne semble plus phonémique. La voyelle /a/ est prononcée plus avancée, [æ], en syllabe ouverte (ex. « papa » prononcé [pæpæ]), et parfois légèrement réhaussée [æ̣] (notamment après /w/ comme dans « quoi », « moi »...), et [a] peut s'entendre en syllabe fermée, en particulier devant obstruante (Armstrong 1993 : 59), et surtout en syllabe finale de mot (ex. « camarade » prononcé [kamarad]). En Haute-Bretagne, en revanche, l'opposition /a~a/ est généralement maintenue quelle que soit la structure syllabique : ainsi, en syllabe ouverte, « moi » se prononce [mwæ] (avec /a/ réalisé avancé et rétracté également) mais « mois » se prononce [mwa], et en syllabe fermée, « mal » se prononce [mal] et « mâle » se prononce [mäl]. Un peu plus au Nord, en Normandie, l'opposition /a~a/ semble encore vivace au Havre dans les années 1990 (Hauchecorne & Ball 1997) et à La Bonneville (commune rurale) et Darnétal (commune urbaine) dans les années 2000 (Hall 2008). Par opposition à une voyelle /a/ réalisée plus ouverte, la voyelle /a/ est régulièrement réalisée [æ] en syllabe finale fermée, surtout lorsqu'elle est suivie de la rhotique, mais aussi en syllabe ouverte, surtout après /w/, où elle est réalisée arrondie [ɶ] (Hauchecorne & Ball 1997), ce qui serait généralisé à toute la Normandie (Walter 1982). Hall (2008) trouve aussi une légère différence d'aperture, avec /a/ réalisé plus ouvert que /a/.

Plus au sud, dans le cas spécifique du français parlé à Saint-Rémy-de-Provence, une petite commune des Bouches-du-Rhône, Hilt (1986) trouve que /a/ a tendance à être réalisé relativement antérieur, [a, æ], dans toutes les positions, avec, en syllabe ouverte, quelques rares réalisations plus centralisées [ä] ou au contraire antérieures réhaussées [æ̣, æ̣], et en syllabe fermée quelques rares réalisations [æ] également, et exceptionnellement [a] (uniquement sur le mot « pâte »). Les autres variétés rurales de France, en revanche, auraient généralement conservé l'opposition lexicale /a~a/ en syllabe fermée (ex. « patte~pâte », « tache~tâche » etc.) mais [a] apparaît aussi en syllabe ouverte, y compris dans des mots fréquents comme la particule de négation « pas », prononcé [pa] (Fagyal et al. 2006). Plus au sud encore, mais dans une métropole comme Marseille, la distinction /a~a/ n'est plus opérationnelle, et le phonème unique /a/ peut avoir des réalisations variables entre une voyelle antérieure et une voyelle plus centrale (Coquillon & Turcsan 2012). Berns (2015) confirme que les locuteurs de Marseille et Rodez et les Aveyronnais vivant à Paris montrent une différence minime de F1, F2 et durée en syllabe fermée (bien que les locuteurs de Rodez montrent une distinction de F2 légèrement supérieure aux deux

autres variétés). Avanzi (2020) montre que les locuteurs et locutrices de la moitié sud de la France ne rapportent pas prononcer différemment *patte* et *pâte*, *malle* et *mâle*, ni *tache* et *tâche*.

Enfin, plus à l'est, dans la commune de Saint-Claude dans le Haut-Jura, à la frontière avec la Suisse, l'opposition /a~ɑ/ est maintenue chez les 20-35 ans et chez les plus de 55 ans, et le /a/ est réalisé plus antérieur chez les plus jeunes (Arnaud 2006).

En français de Belgique, la distinction de timbre est globalement perdue, mais l'opposition est maintenue par la longueur, surtout au centre et à l'est de la Wallonie, moins à l'ouest (Pohl 1985, Walter 1988, Francard 2008, Berns 2019). Ainsi, « tache » se prononce [taʃ] mais « tâche » se prononce [ta:ʃ] (Hambye et al. 2003, Woehrling 2009), surtout chez les locuteurs plus âgés du centre et de l'est du pays (Hambye & Simon 2012). La longueur est cependant parfois combinée à une légère variation de timbre (Hambye & Simon 2012), surtout à l'ouest (Berns 2019). En tous cas, les locuteurs et locutrices interrogés par Avanzi (2020) rapportent distinguer /a/ et /ɑ/ dans 75 à 90% des cas sur l'ensemble du territoire.

En Suisse, selon Fagyal et al. (2006), la situation est similaire à celle des variétés rurales de France métropolitaine, à savoir que /a/ s'oppose phonémiquement à /ɑ/ en syllabe fermée mais qu'il y a de la variation en syllabe ouverte.

En réalité, le français de Suisse parlé dans les années 1970 présente de la variation entre les cantons dans les syllabes ouvertes en fin de mot : la distinction « rat~ras » par exemple est à peu près stable à cette époque dans les cantons de Vaud et de Genève, où 80% des locuteurs environ maintiennent l'opposition, et aussi à Fribourg, mais de façon moins solide puisque seulement 65% des locuteurs maintiennent l'opposition (Métral 1977). Dans le français de Suisse parlé dans les années 2000, cependant, l'opposition « rat~ras » n'est plus maintenue qu'à 50% à Vaud, 45% à Fribourg, 10% à Genève et Neuchâtel, et elle aurait disparu du Jura et du Valais (Schouwey 2008). Plus récemment enfin, il a été montré que les locuteurs de Nyon (Vaud), quel que soit leur âge, font toujours une différence de qualité vocalique (mais pas de longueur), mais pas ceux de Neuchâtel (Racine & Andreassen 2012).

Pour ce qui est des syllabes fermées finales de mot, dans les années 1970, la distinction « patte~pâte » par exemple est maintenue à plus de 80% dans le Jura et à Neuchâtel, et à près de 70% dans le Valais, mais s'y manifeste davantage par une opposition de longueur ; 50% des locuteurs du Vaudois perçoivent une distinction à la fois en timbre et en longueur, et enfin les locuteurs de Fribourg et Genève sont divisés entre [pa:t] et [pa:t] pour « pâte » (Métral 1977). Trente ans plus tard, ce contraste est maintenu à 100% à Neuchâtel et dans le Jura, à 95% dans le canton de Vaud, et à 60% dans le Valais (Schouwey 2008). Plus récemment encore, les valeurs des deux premiers formants sont significativement plus hautes pour /a/ que pour /ɑ/ dans le français parlé à Nyon, mais pas dans celui parlé à Neuchâtel, où on trouve plutôt une distinction de longueur (Racine & Andreassen 2012), ou dans celui parlé à Genève, qui néanmoins ne montre pas une grande distinction de longueur (Berns 2019).

Aujourd'hui en Suisse, donc, l'opposition /a~ɑ/ est variable selon les régions mais pas selon l'âge (Racine & Andreassen 2012), ce qui laisse à penser que la

situation est assez stabilisée, autant dans les cantons où l'opposition est maintenue que dans ceux où elle ne l'est pas.

Au Canada, d'une manière générale, l'opposition /a~ɑ/ est opérationnelle dans la plupart des variétés : au Québec mais aussi en Acadie, en Ontario/Manitoba dans ~80% à ~100% des cas en syllabes fermées au moins (Avanzi 2020). Cependant, /ɑ/ est généralement réalisé moins ouvert (Walker 1984: 76) et semble être la seule voyelle postérieure à diphtonguer. Pour Dumas (1986), l'opposition phonologique repose ainsi exclusivement sur la longueur, dont le timbre postérieur ne serait qu'une propriété dérivée.

En français laurentien, parlé dans la région de Québec, la distinction /a~ɑ/ est toujours robuste, et soumise en partie à la loi de position, mais /ɑ/ y est souvent prononcé arrondi et légèrement allongé [ɔ:], possiblement diphtongué [aw] en syllabe fermée ou réhaussé [ɔ] en syllabe ouverte (Côté 2012).

En français acadien, parlé à l'est du pays, au Nouveau Brunswick et en Nouvelle-Écosse, /ɑ/ et /a/ s'opposent toujours régulièrement, avec une opposition de timbre (variable, mais /ɑ/ est généralement plus postérieur) et de longueur (/ɑ/ est plus long) (Cichocki 2012). Ainsi, /ɑ/ est réalisé [æ, a], par exemple dans « tabernacle », [tabɛrnæk]. Dans les syllabes ouvertes en position finale de mot, /ɑ/ a souvent un allophone postérieur [ɑ], par exemple dans « éclater », [eklatɛ] vs « éclat », [ekla]. Le phonème /ɑ/, pour sa part, est réalisé [ɑ, ɔ] et peut être diphtongué en syllabe fermée finale, par exemple dans « Jacques », [ʒɑ:k, ʒawk]. Cependant, les réalisations de ces voyelles varient selon l'influence de l'orthographe, du style de parole, de l'environnement et de la classe du mot. Ainsi, en lecture (puisque c'est la condition de production qui nous intéresse), on prononce [wa] en syllabe ouverte (par exemple dans « moi ») et dans les syllabes finales fermées par /R/ (comme « avoir »), alors qu'on trouve [wa] en position non-finale (comme dans « boisson »).

À l'Ouest du Canada aussi, en Alberta, où le français est nettement minoritaire, l'opposition /a~ɑ/ est distinctive, et une tendance généralisée à diphtonguer les voyelles longues s'applique également à la voyelle ouverte postérieure (Walker 2012). Ainsi, le /ɑ/ de « lâche » par exemple se prononce [aw].

Pour en conclure avec l'Amérique du Nord, beaucoup plus au sud, en français cajun, parlé notamment à Ville Platte en Louisiane aux États-Unis, /ɑ/ est réalisé extrêmement postérieur, comme un /ɔ/, et plus encore après /w/ (Papen & Rottet 1997). Cependant, la grande variation dans les réalisations des deux voyelles ouvertes invite Klingler & Lyche (2012) à postuler une allophonie.

Au Maghreb, dans les années 1960, une vingtaine de locuteurs et locutrices d'Algérie, de Tunisie et du Maroc rapportent faire la distinction entre « patte » et « pâte » dans 55% des cas (Deyhime 1967a). Aujourd'hui cependant, l'opposition est généralement perdue, avec moins de 25% des 340 répondants de l'étude d'Avanzi (2020) qui maintiennent l'opposition en syllabe fermée, et ce dans chacun des trois pays.

En Afrique subsaharienne, le français est souvent influencé par les autres langues des locuteurs. Ainsi, la distinction /a~ɑ/ est marginale, voire inexistante, en République Centrale d'Afrique (Bordal 2012, Tchkhovrebova 2019) et au Mali (Lyche & Skattum 2012, Tchkhovrebova 2019), deux pays où le français est acquis

tardivement, par l'éducation. C'est néanmoins aussi le cas en Côte d'Ivoire, où le français est une langue officielle largement maîtrisée par des natifs, puisqu'à Abidjan (99% de locuteurs selon l'OIF), le taux d'opposition /a~ɑ/ est également proche de 0% dans les données de PFC analysées par Tchkhovrebova (2019). Dans le français parlé au Sénégal, la distinction /a~ɑ/ a quasiment disparu : 9 des 12 participants de l'étude d'Akissi Boutin et al. (2012) prononcent « mal » et « mâle » ou « patte » et « pâte » de la même façon, en l'occurrence comme une voyelle centrale [ä], ni antérieure ni postérieure. La distinction n'est conservée marginalement que par des locuteurs âgés, et est parfois remplacée par une opposition de longueur /a~ɑ:/ (et même chez le plus âgé des participants, /a~ɑ:/).

3. Données et Méthodologie

Pour cette étude préliminaire, nous avons sélectionné un ensemble de vocables présentant, dans les variétés où l'opposition /a~ɑ/ est distinctive, des paires minimales (Tableau 1).

Le processus d'extraction des données se présente comme suit. D'abord, les enregistrements des mots ciblés en Tableau 1 sont détectés puis téléchargés de la base de données *Lingua Libre*,⁵ médiathèque linguistique participative portée par Wikimedia France et lancée en 2015. En tant qu'outil participatif, n'importe quel internaute peut s'inscrire, créer un profil de locuteur en renseignant quelques métadonnées de base, et s'enregistrer en train de lire des listes de mots. La fiabilité des informations transmises par les contributeurs ne peut donc pas être garantie. Il est également possible que les locuteurs, conscients de pérenniser des prononciations représentatives de leur dialecte, soient sujets à un biais de désirabilité sociale et s'alignent inconsciemment avec une norme plus standard, ou au contraire forcent le trait sur ce qu'ils perçoivent comme typique de leur dialecte. Cependant, ce biais est peut-être moins important dans *Lingua Libre* que dans d'autres plateformes participatives comme *Common Voice* de Mozilla, pour laquelle des internautes ayant pour tâche de valider les enregistrements d'autres internautes reçoivent la consigne explicite de rejeter les enregistrements présentant par exemple des réductions segmentales, pourtant courantes en français.

Durant la lecture de la liste de mots, le dispositif repère les pauses, ce qui permet à l'enregistrement de s'arrêter à la fin du mot et de se relancer au mot suivant, générant ainsi avec une grande facilité des séries de courts fichiers sons, correspondant chacun à un mot. Ces fichiers sont censés être tous nommés selon le modèle suivant : langue–pseudonyme du locuteur–mot. Ainsi, l'enregistrement fra.-Guilhelma-avril.wav par exemple commence par le code ISO 639-3 de la langue, ici « fra. » pour « français », suivi du pseudonyme du locuteur, ici « Guilhelma », et enfin le mot prononcé, ici « avril ».

En février 2022, *Lingua Libre* contenait près de 680 000 enregistrements en 147 langues par 737 locuteurs. Parmi eux, le français est de loin la langue la mieux représentée, avec 239 457 enregistrements produits par 255 locuteurs différents, c'est-à-dire plus du tiers des données, loin devant la deuxième langue de *Lingua Libre*, le polonais, qui compte un peu moins de 80 000 enregistrements. L'avantage

⁵https://lingualibre.org/wiki/LinguaLibre:Main_Page

Tableau 1. Items-tests répartis selon la voyelle sous-jacente /a/ ou /ɑ/

/a/	/ɑ/
En syllabe fermée finale	
Anne	âne, ânes
bal	Bâle
mal	mâle, mâles
pack, packs	Pâques
patte, pattes	pâte, pâtes
tache, taches	tâche, tâches
En syllabe ouverte finale	
bas	bât
ma	mât, mâts
rat, rats	ras
En syllabe ouverte non-finale	
battons	bâton, bâtons
bailler, bayer (+ conj.)	bâiller (+ conj.)

dans les variétés où l'opposition est distinctive d'après la littérature détaillée en Section 2.2, triés par structure syllabique et position dans le mot.

de cette base de données néanmoins est qu'elle continue de s'accroître avec le temps puisque la plateforme est toujours ouverte. De plus, les données sont sous licence libre et accessibles à tous gratuitement. Cependant, à notre connaissance, en-dehors du présent projet de recherche (Hutin & Allasonnière-Tang 2022a, b, c, d), *Lingua Libre* n'a été utilisé qu'une seule fois pour une publication académique, en l'occurrence pour faire une estimation de la transparence de l'orthographe de 17 langues avec un réseau de neurones (Marjou 2021).

Les enregistrements téléchargés sont ensuite segmentés et alignés à l'aide de WebMAUS (Kisler et al. 2017). WebMAUS⁶ est la version en ligne en libre accès du logiciel MAUS⁷ (Schiel 1999), qui permet d'aligner le contenu d'un enregistrement avec sa transcription orthographique. Pour ce faire, MAUS crée un graphe d'hypothèses de prononciation fondé sur la transcription orthographique de l'enregistrement à l'aide d'un convertisseur de graphème en phonème. Durant ce processus, la transcription orthographique est convertie en alphabet phonétique SAMPA (Speech Assessment Methods Phonetic Alphabet, Wells 1997). Le signal sonore est ensuite aligné avec le graphe d'hypothèses, et l'alignement avec la probabilité la plus élevée est sélectionné. Pour donner un aperçu de la précision de l'alignement sur MAUS, ce dernier correspond à 95% avec les alignements réalisés manuellement (Kipp et al. 1997). Dans notre cas, la vérification d'un sous-ensemble des données (10 /a/ et 10 /ɑ/) montre que les bornes de début de phones varient de

⁶<https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic>

⁷<https://www.bas.uni-muenchen.de/forschung/Verbmobil/VM14.7eng.html>

Tableau 2. Un exemple des données extraites et compilées de *Lingua Libre*. Les rangées représentent les occurrences des voyelles /a/ et /ɑ/

Voyelle	F1	F2	Durée	Mot	Locuteur	Genre	Commune	Pays
/a/	929	1749	0.159	mâle	Guilhelma	femme	Cesseras	France
/ɑ/	1106	2205	0.149	Anne	Pamputt	homme	Brétigny-sur-Orge	France

0,009688 seconde entre l'alignement automatique de WebMAUS et l'alignement manuel, et les bornes de fin de 0,01110 seconde, soit une moyenne d'environ 0,01 seconde.

Les voyelles /a/ et /ɑ/ sont extraites des enregistrements grâce à l'alignement effectué. Ces enregistrements extraits sont ensuite analysés en termes de formants et de durée. Pour chaque enregistrement de chaque voyelle, les valeurs F1 et F2 sont extraites au milieu du son sur une fenêtre de 50 millisecondes afin d'atténuer l'influence du bruit induit par le contexte dans les enregistrements⁸. Au cours de ce processus d'extraction et d'analyse de données, les packages R suivants sont utilisés : emuR (Winkelmann et al. 2021), PraatR (Albin 2014) et la suite tidyverse (Wickham 2017). Le code et les données utilisées pour l'analyse sont disponibles dans la documentation complémentaire de l'article⁹.

Un échantillon des données est présenté dans le Tableau 2. Pour chaque occurrence de /a/ et de /ɑ/, les informations suivantes sont extraites : la transcription de la voyelle, ses formants et sa durée ; le mot d'origine de la voyelle ; les informations du locuteur. Ces informations contiennent l'identifiant du locuteur, son genre et sa région ainsi que son pays¹⁰.

Pour ce qui est de l'origine géographique détaillée des locuteurs, elle est présentée dans le Tableau 3 ci-dessous. Nous les avons regroupées en région correspondant plus ou moins aux régions pour lesquelles nous avons des données dans la littérature. Les points de données pour lesquels les locuteurs n'ont pas communiqué de détails personnels ne sont pas inclus dans le tableau (14 locuteurs pour 10 /a/ et 12 /ɑ/).

Nos données comprennent 67 occurrences de /a/ et 96 occurrences de /ɑ/, produites par 38 locuteurs sur 26 communes, qu'on peut regrouper en 7 régions au sens large. On remarque, notamment, que la Belgique et le Nord-Ouest de la France ne sont pas représentés, ce qui est dommage puisque ces deux régions, d'après la littérature, non seulement maintiennent l'opposition mais ne la manifestent pas de la même façon. On note aussi que les données d'Afrique sont presque inexistantes, avec seulement 3 occurrences en provenance du Tchad produites par un seul locuteur. Ces lacunes dans les points de données seront discutées en Section 5.

⁸Nous avons également effectué l'extraction en considérant la moyenne des formants sur la durée entière de l'enregistrement. Les résultats étaient similaires.

⁹Les données et le code utilisés peuvent être téléchargés au référentiel suivant : <https://osf.io/z8whx/>

¹⁰Ces métadonnées ne sont pas toujours communiquées par les locuteurs : nous avons choisi de ne pas tenir compte des données pour lesquelles la localisation n'était pas précisée.

Tableau 3. Nombre d'items répartis par voyelle sous-jacente /a/ ou /a/ et par région d'origine des locuteurs et locutrices

Commune	Région fine	Région large	Nombre de locuteurs	/a/	/a/
Chicoutimi	Québec	Canada	1	0	4
Sayabec	Québec	Canada	1	1	0
Shawinigan	Québec	Canada	1	6	1
Total			3	7	5
Brétigny-sur-Orge	Région parisienne	Centre	1	1	1
Pantin	Région parisienne	Centre	1	1	1
Paris	Région parisienne	Centre	9	11	11
Saint-Barthélémy-d'Anjou	Pays de la Loire ¹¹	Centre	1	0	4
Total			12	13	17
Cornimont	Vosges	Nord-Est	1	5	26
Muntzenheim	Lorraine	Nord-Est	1	1	2
Nancy	Lorraine	Nord-Est	1	1	0
Strasbourg	Alsace	Nord-Est	1	1	1
Vosges	Vosges	Nord-Est	2	5	8
Total			6	13	37
Montpouillan	Nouvelle Aquitaine	Sud-Ouest	1	0	1
Nouvelle-Aquitaine	Nouvelle Aquitaine	Sud-Ouest	1	0	1
Toulouse	Occitanie	Sud-Ouest	2	8	4
Cesseras	Occitanie	Sud-Ouest	1	2	5
Montpellier	Occitanie	Sud-Ouest	1	0	1
Occitania	Occitanie	Sud-Ouest	1	1	0
Total			7	11	12
Grenoble	Rhône-Alpes	Sud-Est	1	0	1
Lyon	Rhône-Alpes	Sud-Est	3	16	13
Saint-Etienne	Rhône-Alpes	Sud-Est	1	0	2
Total			5	16	16
Canton du Valais	Valais	Suisse	1	4	3

(Continued)

¹¹Pour des raisons méthodologiques, nous avons fait le choix de regrouper la région parisienne et les Pays de la Loire sous le label "Centre". Nous sommes conscients que ce choix est discutable. Néanmoins, rien dans la littérature décrite en 2.2 ne nous permet de considérer ce regroupement comme abusif : les données d'Avanzi (2020) semblent même indiquer qu'il est justifié.

Tableau 3. (Continued.)

Commune	Région fine	Région large	Nombre de locuteurs	/a/	/ɑ/
Gaillard	Haute-Savoie	Suisse ¹²	1	1	1
Geneva	Genève	Suisse	1	0	1
Lausanne	Vaud	Suisse	1	0	3
Total			4	5	8
N'Djamena	Tchad	Afrique	1	2	1
Total			1	2	1
Grand total			38	67	96

4. Résultats

Les résultats montrent que, tous locuteurs confondus, les voyelles du groupe /a/ et celles du groupe /ɑ/ sont effectivement réalisées différemment, comme on peut le voir en Figure 2. La distinction se manifeste sur F1, où /a/ a tendance à être réalisé plus bas de façon statistiquement significative ($t = 3.02$, $df = 170.38$, $p = 0.003$), sur F2, où /a/ a tendance à être réalisé plus haut, mais de façon non statistiquement significative ($t = -0.48$, $df = 166.81$, $p = 0.633$) et en longueur, puisque /a/ a tendance à être réalisé plus court, ce qui est surprenant, mais ce résultat n'est pas non plus significatif ($t = 0.93$, $df = 170.92$, $p = 0.355$) et pourrait aussi être dû à la segmentation automatique. Ces résultats sont également obtenus en utilisant des modèles linéaires généralisés à effets mixtes (présentés dans le matériel complémentaire).

Le fait que F1 soit le facteur phonétique le plus significatif suppose que l'opposition /a~ɑ/ en français est davantage une opposition de hauteur, ou d'aperture, qu'une question d'antéro-postériorité: la réalisation d'un /a/ est plus ouverte que celle d'un /ɑ/ plutôt que plus postérieure. Mais qu'en est-il des différentes variétés de français?

Dans ce qui suit, nous observons l'opposition /a~ɑ/ par région, afin de montrer si on l'observe effectivement dans les régions où la littérature témoigne qu'elle est maintenue, et si on ne l'observe effectivement pas dans les régions où la littérature témoigne qu'elle a disparue. Pour des raisons d'effectifs, pour les présents calculs, nous avons regroupé entre elles les localités des Vosges, de Lorraine et d'Alsace sous le label "Nord-Est", les localités de Rhône-Alpes sous le label "Sud-Est", et celles de Nouvelle-Aquitaine et d'Occitanie sous le label "Sud-Ouest", mais aussi la région parisienne et les pays de la Loire sous le label "Centre" ($n=30$), la Haute-Savoie et les différentes localités de Suisse sous le label "Suisse" ($n=13$), et enfin les différentes localités du Canada sous le label "Canada" ($n=12$). Le label "Tchad" fait référence à l'unique point d'enquête dont nous disposons pour l'heure en Afrique francophone. Etant donnée la rareté des données, nous présentons ici des analyses quantitatives

¹²Nous sommes conscients que, malgré sa localisation limitrophe, Gaillard n'est pas en Suisse mais en France. Nous le regroupons ici avec les localités suisses en raison de sa proximité géographique avec elles et le fait que, dans le cas de l'opposition /a~ɑ/, « la frontière politique qui isole la Wallonie et la Romandie de l'Hexagone n'en est pas une sur le plan linguistique » (Avanzi 2020 : 56).

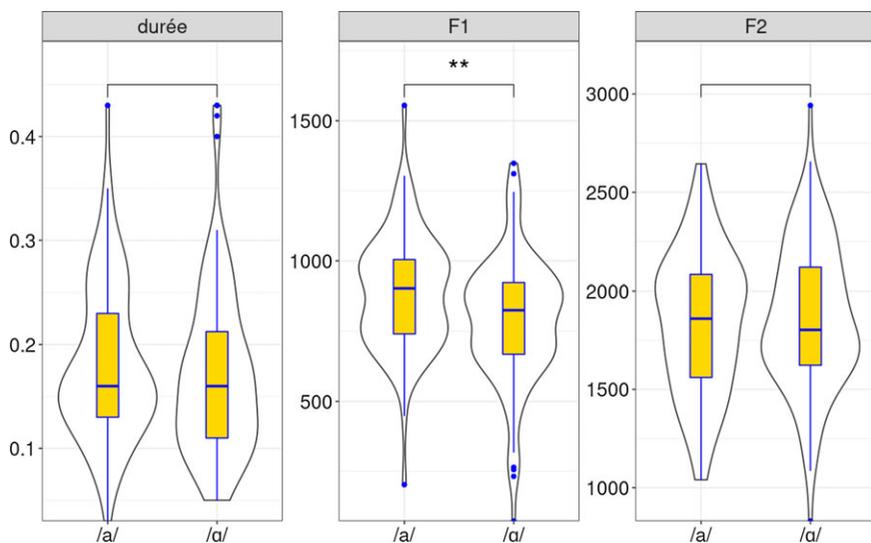


Figure 2. Différence de réalisations entre /a/ et /a/ en fonction de la durée du segment (en secondes) et de F1 et F2 (en Hertz) toutes variétés confondues. Les valeurs statistiques indiquent le résultat d'un t-test. Les comparaisons sans symboles relatifs à la valeur p indiquent des différences non significatives.

générales, c'est-à-dire toutes positions confondues, que nous affinons ensuite en opposant syllabe fermée et syllabe ouverte (indépendamment de la position dans le mot), et nous complétons les résultats avec des observations d'ordre qualitatif.

Dans la Figure 3, qui détaille les résultats par région et par paramètre, on peut voir que, tous types de syllabe confondus (les boîtes rouges), il n'y a pas de différences de réalisation statistiquement significatives entre les voyelles du groupe /a/ et celles du groupe /a/ pour F2 (dernière ligne) ni, contrairement aux résultats de la Figure 2¹³, pour F1 (ligne du milieu), sauf dans le Nord-Est de la France. La durée en revanche est significativement différente entre /a/ et /a/ toutes syllabes confondues seulement au Canada et dans le Sud-Ouest de la France.

Lorsqu'on observe les données par région, le point le plus surprenant concerne l'opposition de longueur en syllabe ouverte que l'on observe dans le Nord-Est, le Sud-Est et le Sud-Ouest de la France et en Suisse. Ces résultats ont ceci de surprenant que (i) dans la littérature, lorsque l'opposition /a~a/ est maintenue dans ces régions, elle l'est plutôt en syllabe fermée (Fagyal et al. 2006), et (ii) qu'on attendrait que /a/ soit réalisé plus court que /a/, c'est-à-dire le schéma inverse à celui qu'on observe. Ceci étant, l'opposition /a~a/ est maintenue parfois dans les monosyllabes (Picoche & Marchello-Nizia 1994 : 208) et il est possible qu'on ait ici un effet de fréquence : les items choisis pour représenter /a/ dans cette position sont tous nettement plus rares que leurs homologues du groupe /a/. Il est aussi possible que l'aligneur segmente le signal de façon incorrecte sur /a/, qui est une voyelle

¹³Cette différence peut s'expliquer par un effet de la faible quantité de données par région. En séparant les données par région, il reste peu de points disponibles pour chaque comparaison (les chiffres détaillés par région sont en Tableau 3), ce qui rend les comparaisons facilement non-significatives.

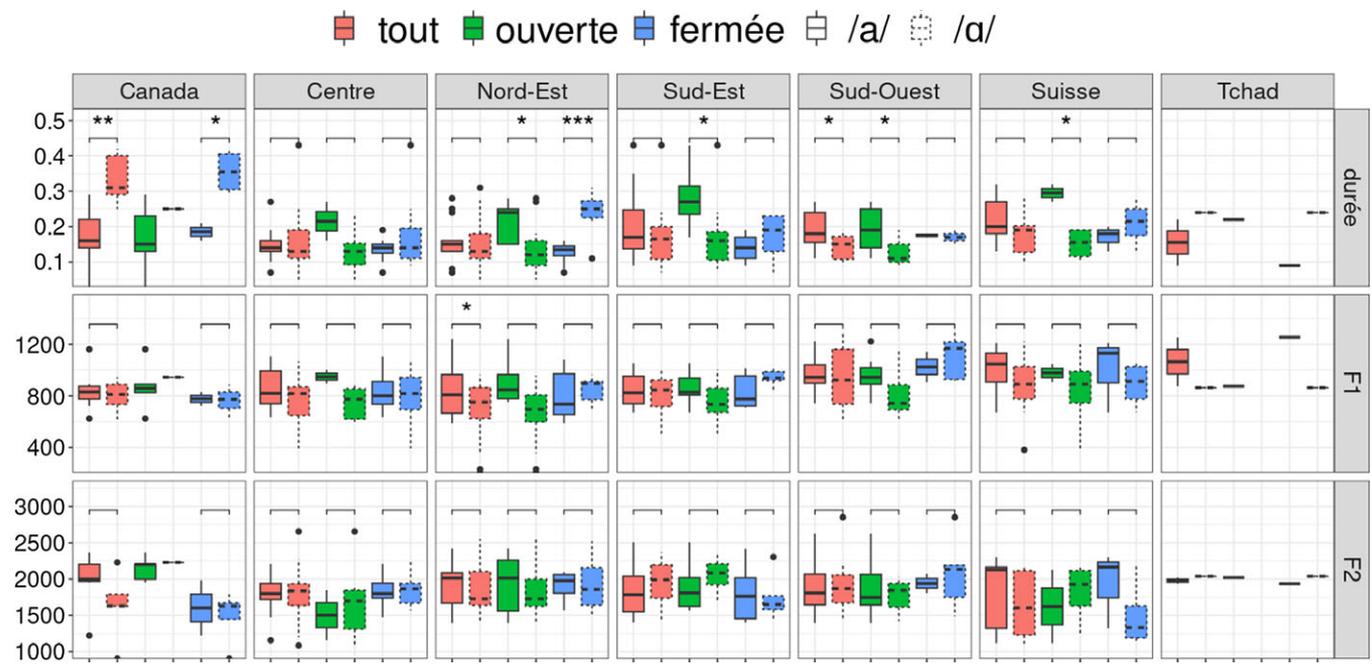


Figure 3. Visualisation par région et syllabe ouverte/fermée de la différence de réalisation entre /a/ et /ɑ/ en fonction de la durée du segment (en seconde) et de F1 et F2 (en Hertz). Les valeurs statistiques indiquent le résultat d'un t-test avec correction de Bonferroni. Les comparaisons sans symboles relatifs à la valeur *p* indiquent des différences non significatives. Les comparaisons sans accolades indiquent des comparaisons non réalisées suite au manque de données.

Tableau 4. Valeurs moyennes de F1 et F2 (en Hertz) et de durée (en secondes) pour les réalisations de /a/ et de /ɑ/ par régions d'origine des locuteurs

Région	F1 (Hertz)		F2 (Hertz)		durée (s)	
	/a/	/ɑ/	/a/	/ɑ/	/a/	/ɑ/
Canada	833,05	697,02	1866,58	1579,92	0,17	0,33
Centre	849,65	751,41	1809,69	1817,58	0,15	0,16
Nord-Est	848,94	706,79	1984,95	1886,98	0,16	0,15
Sud-Est	821,78	826,82	1799,74	2000,39	0,20	0,17
Sud-Ouest	986,27	946,06	1827,56	1897,71	0,19	0,14
Suisse	1015,79	877,92	1765,07	1629,84	0,22	0,18
Tchad	1025,72	872,03	1975,34	2042,27	0,15	0,24

moins fréquente en français que /a/. Cependant, outre ces quelques problèmes techniques, les résultats par région sont généralement conformes à la littérature.

Ainsi, dans le Centre de la France, l'absence de différence de réalisation sur F2 et sur la durée est conforme à la littérature passée (Hansen 2012, Berns 2015, 2019). Le seul résultat surprenant est la différence significative de F1 entre les syllabes ouvertes avec /a/ et celles avec /ɑ/ (opposition du type «ma ~ mâ»), qui à notre connaissance n'a été décrite dans les études passées pour aucune variété, sauf le français de Normandie¹⁴. Notre explication est que la distinction entre /a/ et /ɑ/ s'est perdue encore davantage depuis les enregistrements de PFC utilisés dans les études passées, et que la façon qu'ont les locuteurs du Centre de la France, en 2020, de marquer la différence, surtout en lecture, entre les deux voyelles, consiste à exagérer l'aperture de la bouche, signe plus ostentatoire visuellement de distinction dans la production, plutôt qu'à reculer la langue.

De plus, comme on peut le voir dans le Tableau 4 ci-dessus, dans la variété du Centre de la France, la voyelle unique /a/ est réalisée plutôt avancée, comme un [æ], conformément à ce qui est proposé dans la littérature des années 1970 (Mettas 1979, Lennig 1978, François 1974) mais moins aux données plus récentes sur une seule locutrice de Fougeron & Smith (1999).

Plus au sud, notamment dans le Sud-Est ($t = 2.48$, $df = 14.29$, $p = 0.026$) et le Sud-Ouest ($t = 2.66$, $df = 13.28$, $p = 0.019$) de la France, on trouve la différence de durée en syllabe ouverte déjà évoquée, mais peu de différences dans les formants des deux voyelles, ni en syllabe ouverte ni en syllabe fermée, conformément à la littérature (Coquillon & Turcsan 2012), sauf pour F1 en syllabe fermée dans le Sud-Est, montrant là encore une tendance à accentuer l'aperture pour prononcer [ɑ]. En

¹⁴La différence de hauteur a été identifiée par Hall (2008) sur le français de Normandie, où, contrairement à nos résultats toutes syllabes confondues, /a/ est généralement plus haut (c'est-à-dire avec un F1 plus bas) que /ɑ/. Cette divergence entre nos résultats est sans doute liée à l'équilibrage des données dans l'étude de Hall (2008) et dans la nôtre puisque le phénomène semble sensible au type de syllabe. Ainsi, les résultats de Hall sont effectivement contraires à nos résultats *en syllabe ouverte*, mais conformes à nos résultats *en syllabe fermée*. Malheureusement, Hall ne propose pas d'analyse détaillée en syllabes ouvertes vs syllabes fermées.

revanche, nos observations nous invitent à postuler une réalisation plus postérieure que la littérature.

Dans le Nord-Est de la France, l'opposition est maintenue, surtout sur des critères positionnels, conformément aux observations de Boughton (2005). Ainsi, on observe notamment une différence de durée en syllabe fermée comme en syllabe ouverte, et on peut voir, en Figure 3, que la distinction de F1 s'inverse selon le type de syllabe : /a/ a une valeur plus élevée que /a/ en syllabe ouverte, mais plus basse en syllabe fermée. L'opposition en revanche n'est pas significative pour F2, bien qu'on voie un F2 plus élevé pour /a/ que pour /a/, c'est-à-dire que [a] semble réalisé plus postérieur que [a].

En Suisse, on ne voit d'opposition statistiquement significative que pour la durée en syllabe ouverte, même si on constate aussi, à l'œil nu, une différence de F1 et F2 en syllabe fermée. Le manque de résultats statistiquement significatifs pour cette région est sans doute dû au fait que la Suisse présente plusieurs comportements phonologiques différents selon la localité précise (Métral 1977, Schouwey 2008, Racine & Andreassen 2012). Or nos locuteurs suisses viennent tous d'un canton différent. Rappelons aussi que des résultats similaires sont visibles chez leurs voisins du Sud-Est de la France, sinon que la différence de F1 en syllabe fermée y est statistiquement significative, et même un peu plus loin, dans le Sud-Ouest, du moins pour les syllabes ouvertes.

Au Canada, l'opposition est effectivement visible dans nos données : la voyelle /a/ semble réalisée très ouverte et très avancée, comme un [æ], et la voyelle /a/ moins ouverte, plus postérieure et plus longue comme un [a:], mais l'opposition n'est statistiquement significative que sur la durée ($t = -3.59$, $df = 9.43$, $p = 0.005$). Bien qu'on attende effectivement, dans le français laurentien, une telle opposition de longueur (Côté 2012), elle est généralement couplée à une opposition de timbre qui n'est pas tranchée dans nos données, ce qui est peut-être dû au fait qu'elle relève souvent de la diphthongaison ou de l'arrondissement, c'est-à-dire d'une différence sur le troisième formant.

Enfin, les données sont trop peu nombreuses pour le français parlé en Afrique, avec seulement trois occurrences en provenance du Tchad. A titre purement indicatif, on peut seulement observer que les 2 occurrences de /a/ ont une réalisation extrêmement ouverte et avancée, tandis que la seule occurrence de /a/ a une réalisation encore plus antérieure.

Pour résumer, dans nos données, /a/ et /a/ s'opposent de façon significative essentiellement en termes de durée en syllabe fermée (comme dans « patte~pâte ») au Canada et dans le Nord-Est de la France, et en syllabe ouverte (comme dans « ma~mât ») dans le sud européen (France sauf Centre et Suisse). Cette opposition de longueur est parfois couplée à une opposition d'aperture en France, en syllabe ouverte (Centre et Nord-Est) ou fermée (Sud-Est). Ces résultats ne confirment donc pas les observations sur l'opposition sur l'axe antéro-postérieur. Cela peut-être dû à un biais dans notre échantillon, soit parce que les items ne sont pas assez équilibrés, soit parce que les locuteurs ont lissé leur prononciation, mais cela pourrait aussi être indicatif du fait que l'usage a évolué entre les années 2000 (données PFC) et 2020 (données *Lingua Libre*). Des études plus approfondies devraient permettre ultérieurement de savoir si l'opposition d'antéro-postériorité décrite par le passé a effectivement été remplacée par une opposition d'aperture comme nos données le laissent présumer.

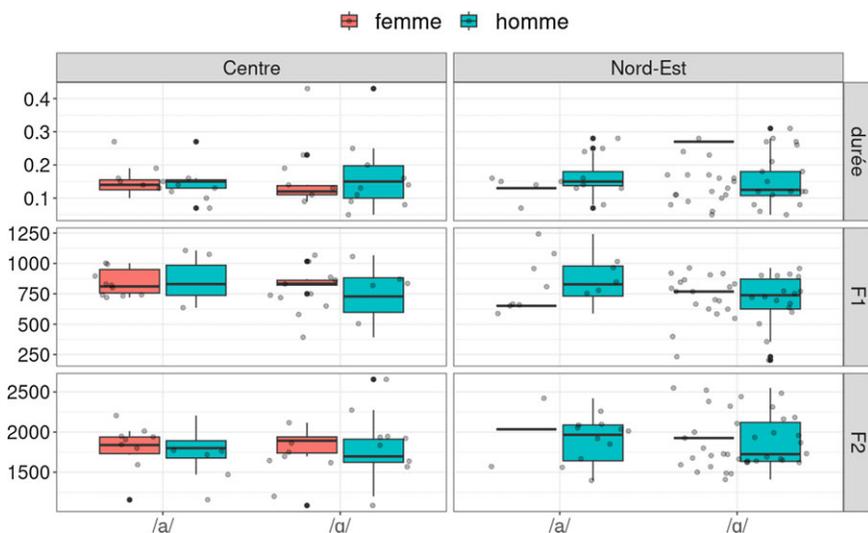


Figure 4. Visualisation des valeurs de durée, de premier formant (F1) et de deuxième formant (F2) au Centre (colonne gauche) et au Nord-Est (colonne droite) entre les femmes (boîtes roses) et les hommes (boîtes bleues).

Enfin, *Lingua Libre* donne la possibilité aux participants et participantes de renseigner leur genre. Cette métadonnée est intéressante car, dans les phénomènes qui apparaissent ou, dans notre cas, disparaissent de la langue, on attend de la variation entre les hommes et les femmes (Labov 2001). Certains ne le font pas, cependant, ce qui empêche d'avoir assez de données pour opérer des tests de signifiante, mais nous pouvons tout de même observer des tendances. En l'occurrence, nous proposons d'observer l'effet du genre dans le Centre (5 femmes et 5 hommes), où nous n'avons pas trouvé d'opposition phonémique entre /a/ et /ɑ/, et dans le Nord-Est de la France (1 femme et 5 hommes), où nous avons trouvé une opposition au moins en syllabe fermée¹⁵.

Comme on peut le voir dans la Figure 4 ci-dessus, dans le Centre de la France, les hommes réalisent des /a/ similaires à ceux des femmes à tous points de vue, sauf peut-être pour F2, tandis qu'ils semblent réaliser /ɑ/ légèrement plus long et avec des valeurs F1 et F2 légèrement inférieures. Les valeurs des femmes et des hommes paraissent cependant très proches dans cette variété. Dans le Nord-Est de la France, en revanche, la locutrice réalise /a/ plus court, avec un F1 plus bas et un F2 plus haut que les hommes, et surtout elle réalise /ɑ/ beaucoup plus long, avec un F1 similaire mais un F2 beaucoup plus haut.

Ces premières intuitions sur ces quelques données devront être étayées avec des données de terrain plus complètes, par exemple lorsque *Lingua Libre* aura été augmenté. Néanmoins, il semble que dans le Nord-Est de la France, l'opposition est

¹⁵Nous concédons que comparer les données du Centre, où il n'y a pas d'opposition phonémique entre /a/ et /ɑ/, avec les données du Canada, où il y en a une, aurait été plus pertinent, mais les locuteurs du Canada n'étaient pas assez nombreux (n = 3) et tous se sont identifiés comme hommes.

maintenue surtout chez la locutrice, ce qui pourrait indiquer que l'opposition /a~ɑ/ y est en voie de disparition et n'est maintenue que par hypercorrection, souvent plus visible chez les femmes. Ce résultat devra être confirmé avec des données plus complètes et plus équilibrées dans des recherches ultérieures.

Ce que montre cette micro-étude sociophonétique, c'est que *Lingua Libre* est une plateforme potentiellement puissante pour des études en socio-phonétique mais que cette potentialité ne sera effective que si les champs du questionnaire sont remplis par une majorité de participants et participantes. Nous espérons que cette tentative aura le mérite de mettre en lumière le besoin d'informer les contributeurs et contributrices sur la nécessité de renseigner ces champs.

5. Conclusion et Discussion

Dans cette étude exploratoire, nous tentons de montrer le potentiel des données participatives pour l'étude phonétique et phonologique des langues, ici du français, et de sa variation à travers le monde. Nous explorons le cas particulier de l'opposition entre la voyelle antérieure ou centrale /a/ et la voyelle postérieure et parfois allongée /ɑ/, encore phonémique dans plusieurs variétés de français. Sur un total de 163 occurrences produites par 38 locuteurs en provenance de 7 régions du monde, nous montrons que *Lingua Libre*, la médiathèque linguistique libre de Wikimédia France, a le potentiel de donner des résultats conformes aux descriptions offertes par des études de terrain professionnelles. Avec ce même corpus, nous pouvons donc envisager d'étudier l'opposition des autres voyelles à double timbre /E/, /O/ et /ɛ/ (voire /Ē/) selon les régions du monde, mais aussi d'éclaircir, notamment, le rôle de l'orthographe, en l'occurrence de la présence de l'accent circonflexe, sur la prononciation de <a> vs <â>.

L'étude montre aussi un certain nombre de limites de cet outil. Du point de vue de la gestion de la plateforme, il faut souligner des problèmes techniques, par exemple des noms de fichier non conformes au gabarit, qui empêchent le traitement automatique et obligent à éliminer des données. De plus, la plateforme ne permet de travailler que sur des données de lecture de mots isolés, qui sont problématiques à de nombreux égards. Ce problème pourra être en partie surmonté à l'avenir puisque les équipes de *Lingua Libre* commencent à communiquer plus largement sur les possibilités techniques offertes par la plateforme, notamment le fait que la durée de la pause détectable par le système pour segmenter les fichiers sons peut être augmentée, ce qui permet de lire des phrases et même des textes plus longs sans que les pauses prosodiques ne causent de découpage intempestif. Enfin, la plateforme ne permet pas d'obtenir des informations démographiques extrêmement précises. Ainsi, les linguistes qui s'intéressent aux données de *Lingua Libre* n'ont pas accès à plus de détails sur la provenance des locuteurs, leur immersion, l'âge et la durée de l'exposition à la langue, etc. Nous espérons que ce travail permettra aux créateurs et créatrices de la plateforme de prendre la mesure des besoins des scientifiques en matière de métadonnées et les invitera à proposer des questionnaires plus précis, tout en maintenant bien sûr l'impossibilité de désanonymiser les données (Sweeney 2000).

L'outil présente aussi des limites inhérentes à toute opération participative. Tout d'abord, qui s'intéresse à ces données est forcé d'accepter un risque calculé en raison

du fait que les métadonnées, notamment concernant le genre ou le lieu d'origine, puissent être renseignées de façon incomplète ou incorrecte. Dans la présente étude, nous n'avons pas tenu compte du genre, qui était renseigné de façon trop partielle pour donner des résultats significatifs. En revanche, nous envisageons de reformater les données afin de pouvoir inclure ce facteur dans nos futures analyses. Ensuite, le ou la linguiste est tributaire du bon-vouloir des personnes, et à ce titre ne peut travailler que sur une quantité de données imposée. Or, malgré de nombreuses données en français dans *Lingua Libre*, c'est encore trop peu pour regarder en détail un phénomène aussi fin que l'opposition /a~ɑ/, problème que l'on trouve aussi dans les corpus mais qui ne se pose pas (ou moins) dans les enquêtes de terrain. Par exemple, dans notre analyse, il n'est pas toujours possible d'obtenir les mots voulus pour chaque locuteur. Si les données étaient suffisantes, il serait davantage possible de réunir des données similaires à celles d'une expérience avec des stimuli contrôlés et prononcés par chaque locuteur. Enfin, sur une note similaire, l'inégalité des points d'enquête nous a empêché d'observer le comportement linguistique dans plusieurs régions qu'il aurait été intéressant de comparer à la littérature, notamment le Nord-Ouest de la France, la Belgique et l'Afrique. Cependant, ce problème est aussi, dans une moindre mesure, celui auquel se confrontent les linguistes qui travaillent sur des données de terrain, où les français d'Afrique ou des Etats-Unis, par exemple, sont clairement sous-représentés.

Enfin, notre étude souffre aussi des problèmes liés à l'automatisation de la méthode. En effet, certains résultats sont surprenants, voire clairement contraires aux attentes, comme /a/ réalisé plus long que /ɑ/ dans plusieurs régions de France et en Suisse. Il est possible que l'aligneur WebMAUS ne soit pas assez précis pour traiter des données aussi fines. Dans de futures études, nous envisageons de produire une segmentation manuelle comme référence et de la comparer à plusieurs logiciels d'alignement (Montreal Forced Aligner de McAuliffe et al. (2017), EasyAlign de Goldman (2011), SPPAS de Bigi (2012), etc.) pour identifier le meilleur.

Cependant, nous pensons que cette étude montre le potentiel que représente cette plateforme. En effet, elle peut être améliorée pour mieux gérer les noms de fichiers ou s'étendre pour traiter d'autres types de données que la parole lue, comme l'a projeté un temps la plateforme *Donnez votre français à la science* (Glikman et al. 2018). Quant au problème de données lacunaires, en termes de quantité comme de variété, il se résoudra certainement de lui-même, puisque cette base de données est vouée, par sa nature même, à continuer de grandir. Pour l'heure, les résultats de nos analyses sont relativement conformes à ce que nous attendions des observations passées, et ne pourront que s'affiner à l'avenir.

Acknowledgements. Cette recherche a été en partie financée par l'Institut DATAIA et la MSH Paris-Saclay dans le cadre du projet OTELO - *OnTologies pour l'Enrichissement de l'analyse Linguistique de l'Oral* (porté par Ioana Vasilescu et Fabian Suchanek), ainsi que par le projet ANR EVOGRAM: *The role of linguistic and non-linguistic factors in the evolution of nominal classification systems* (ANR-20-CE27-0021, porté par Marc Allasonnière-Tang), et par le F.R.S.-FNRS dans le cadre de l'appel Bourses et Mandats octroyé au projet PPaDisM - *Phonetic Patterns in Discourse Markers* (porté par Mathilde Hutin).

Les auteurs souhaitent également remercier la communauté Wikimedia de son intérêt pour ce projet, et en particulier Lucas Lévêque d'avoir mis à notre disposition sa connaissance de l'outil *Lingua Libre*.

Competing Interests. Les auteurs n'ont aucun conflit d'intérêt à déclarer.

References

- Akissi Boutin, Béatrice, Gess, Randall & Guèye, Gabriel Marie.** 2012. French in Senegal after three centuries: A phonological study of Wolof speakers' French. *Phonological Variation in French: Illustrations from Three Continents*, edited by Randall Gess, Chantal Lyche & Trudel Meisenburg, John Benjamins Publishing Company. 45–71.
- Albin, Aaron.** 2014. PraatR: An architecture for controlling the phonetics software “Praat” with the R programming language. *Journal of the Acoustical Society of America*, 135(4), 2198. <https://doi.org/10.1121/1.4877175>
- ANF.** 2011. Assemblée nationale de France. <http://www.assemblee-nationale.fr/index.asp>. Website of Assemblée nationale de France.
- ANQ.** 2011. Assemblée nationale du Québec. <http://www.assnat.qc.ca/fr/index.html>. Website of Assemblée nationale du Québec.
- Armstrong, Nigel.** 1993. *A study of phonological variation in French secondary school pupils*. Unpublished PhD thesis, University of Newcastle upon Tyne.
- Armstrong, Nigel.** 2001. *Social and stylistic variation in spoken French: a comparative approach*. Vol. 8. John Benjamins Publishing.
- Arnaud, Vincent.** 2006. *La dimension variationniste du français en usage à Saint-Claude (Haut-Jura): une étude acoustique des voyelles orales des « gens d'en haut »*, Thèse de doctorat, Université Laval (Québec) et Université de Franche-Comté (France).
- Avanzi, Mathieu.** 2020. Géographie des tendances centripètes et centrifuges du français en francophonie : le cas des oppositions phonologiques /a~/~/a/ et /ɛ~/~/œ/, in Teresa Cabré & Mònica Güell (dir.), *Norme et diversité linguistique: la gestion normative dans des contextes pluricentriques*. *Francophonie et catalanophonie*. Barcelone/Paris: Institut d'Estudis Catalans, pp. 33–74.
- Avanzi, Mathieu, Barbet, Cécile, Glikman, Julie, Peuvergne, Julie.** 2016. Présentation d'une enquête pour l'étude des régionalismes du français. *Actes du Congrès mondial de linguistique française* [en ligne, non paginé]. Tours.
- Berns, Janine.** 2015. Merging low vowels in metropolitan French. *Journal of French Language Studies* 25(3): 317–338. <https://doi.org/10.1017/S0959269515000174>
- Berns, Janine.** 2019. Low vowel variation in three French-speaking countries. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 64(1): 1–31. <https://doi.org/10.1017/cnj.2018.23>
- Bigi, Brigitte.** 2012. SPPAS: A tool for the phonetic segmentations of Speech. *The eighth international conference on Language Resources and Evaluation (LREC)*. Istanbul, Turkey, 1748–1755.
- Bordal, Guri.** 2012. A phonological study of French spoken by multilingual speakers from Bangui, the capital of the Central African Republic. *Phonological Variation in French: Illustrations from Three Continents*, edited by Randall Gess, Chantal Lyche & Trudel Meisenburg, John Benjamins Publishing Company. 23–44.
- Boughton, Zoë.** 2005. Accent leveling and accent localisation in northern French: Comparing Nancy and Rennes. *Journal of French Language Studies*, 15, 235–256. <https://doi.org/10.1017/S0959269505002140>
- Casagrande, Jean.** 1984. *The Sound System of French*. Washington D.C.: Georgetown University Press.
- Cichoki, Wladyslaw.** 2012. An overview of the phonetics and phonology of Acadian French spoken in northeastern New Brunswick (Canada). *Phonological Variation in French: Illustrations from Three Continents*, edited by Randall Gess, Chantal Lyche & Trudel Meisenburg, John Benjamins Publishing Company. 211–233.
- Coquillon, Annelise & Turcsan, Gabor.** 2012. An overview of the phonological and phonetic properties of Southern French Data from two Marseille surveys. *Phonological Variation in French: Illustrations from Three Continents*, edited by Randall Gess, Chantal Lyche & Trudel Meisenburg, John Benjamins Publishing Company. 105–127.
- Côté, Marie-Hélène.** 2012. Laurentian French (Quebec): Extra vowels, missing schwas and surprising liaison consonants. *Phonological Variation in French: Illustrations from Three Continents*, edited by Randall Gess, Chantal Lyche & Trudel Meisenburg, John Benjamins Publishing Company. 235–274.
- Coveney, Aidan.** 2001. *The sounds of contemporary French*. *Articulation and diversity*. Exeter: Elm Bank Publications.
- Deyhime, Guiti.** 1967a. Enquête sur la phonologie du français contemporain. *La Linguistique*, 3 (1): 97–108.
- Deyhime, Guiti.** 1967b. Enquête sur la phonologie du français contemporain. *La Linguistique*, 3 (2): 57–84.

- Dumas, Denis.** 1986. Le statut des « deux a » en français québécois. *Revue québécoise de linguistique*, 15(2), 167–196. <https://doi.org/10.7202/602566ar>
- Durand, Jacques, Laks, Bernard & Lyche, Chantal** (2002). La phonologie du français contemporain: usages, variétés et structure. In C. Pusch & W. Raible (eds), *Romanistische Korpuslinguistik – Korpora und gesprochene Sprache/Romance Corpus Linguistics – Corpora and Spoken Language*. Tübingen: Gunter Narr Verlag, 93–106. PDF (Durand/Laks/Lyche 2002)
- Fagyal, Zsuzsanna, Kibbee, Douglas, Jenkins, Frederic.** 2006. *French: A linguistic introduction*. Cambridge University Press.
- Fougeron, Cécile & Smith, Caroline L.** 1999. French. *Handbook of the International Phonetic Association*, 78–81. Cambridge: Cambridge University Press.
- Francard, Michel.** 2008. Variétés de français en Belgique. In *La langue française dans sa diversité*, ed. Direction des relations publiques du ministère de la culture, des communications et de la condition féminine, 39–57. Québec: Bibliothèque et archives nationales du Québec.
- François, Denise.** 1974. Français parlé : analyse des unités phonétiques et significatives d'un corpus recueilli dans la région parisienne. Paris, S.E.L.A.F.
- Galliano, Sylvain, Geoffrois, Edouard, Mostefa, Djamel, Choukri, Khalid, Bonastre, Jean-François & Gravier, Guillaume.** 2005. ESTER Phase II Evaluation Campaign for the Rich Transcription of French Broadcast News. *Interspeech* 2005, 2453–2456.
- Glikman, Julie, Benzitoun, Christophe, Goldman, Jean-Philippe, Scherrer, Yves, Avanzi, Mathieu & Boula de Mareüil, Philippe.** 2018. *Donnez votre Français à la Science !* Internet et la documentation de la diversité linguistique : présentation de la plateforme et premiers résultats. *Actes du 6^e Congrès Mondial de Linguistique Française – CMLF*. SHS Web Conference Vol. 46, 2018. <https://doi.org/10.1051/shsconf/20184602003>
- Goldman, Jean-Philippe.** 2011. Easyalign: An automatic phonetic alignment tool under Praat. *Proceedings of Interspeech* 2011. 3233–3236.
- Gravier, Guillaume, Adda, Gilles, Paulson, Niklas, Carré, Matthieu, Giraudel, Aude & Galibert, Olivier.** 2012. The ETAPE corpus for the evaluation of speech-based TV content processing in the French language. *Proceedings of LREC Eighth International Conference on Language Resources and Evaluation*.
- Hall, Damien J.** 2008. *A Sociolinguistic Study of the Regional French of Normandy*. Thèse de doctorat, Université de Pennsylvanie.
- Hambye, Philippe, Francard, Michel & Simon, Anne-Catherine.** 2003. Phonologie du français en Belgique. Bilan et perspectives. *La Tribune internationale des langues vivantes*, 33, 56–63.
- Hambye, Philippe & Simon, Anne-Catherine.** 2012. The variation of pronunciation in Belgian French From segmental phonology to prosody. *Phonological Variation in French: Illustrations from Three Continents*, edited by Randall Gess, Chantal Lyche & Trudel Meisenburg, John Benjamins Publishing Company. 129–149.
- Hansen, Anita Berit.** 2012. A study of young Parisian speech: Some trends in pronunciation. *Phonological Variation in French: Illustrations from Three Continents*, edited by Randall Gess, Chantal Lyche & Trudel Meisenburg, John Benjamins Publishing Company. 151–172.
- Hansen, Anita Berit & Juillard, Caroline.** 2011. La phonologie parisienne à trente ans d'intervalle – Les voyelles à double timbre. *Journal of French Language Studies*, 21(3), 313–359. <https://doi.org/10.1017/S0959269510000347>
- Hauchecorne, Fabrice & Ball, Rodney.** 1997. L'accent du Havre : un exemple de mythe linguistique. *Langage & société* 82, 5–25.
- Hilt, Anne.** 1986. Une analyse auditive sur les voyelles à double timbre du français parlé à Saint-Rémy-de-Provence. *Revue Romane* 21, 2.
- Hilton, Nanna Haug & Leemann, Adrian.** 2021. Editorial: using smartphones to collect linguistic data. *Linguistics Vanguard*, 7(s1), 2021. <https://doi.org/10.1515/lingvan-2020-0132>
- Hutin, Mathilde & Allasonnière-Tang, Marc.** 2022a. Languages Worldwide and the World Wide Web: Crowdsourcing on the Internet to Explore Linguistic Theories. In *Diversity of Methods and Materials in Digital Human Sciences: Proceedings of the Digital Research Data and Human Sciences DRDHum Conference 2022* (pp. 136–147). Jyväskylä, Finland: University of Jyväskylä.
- Hutin, Mathilde, & Allasonnière-Tang, Marc.** 2022b. Investigating phonological theories with crowd-sourced data: The inventory size hypothesis in the light of Lingua Libre. *Proceedings of the 19th*

- SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology (pp. 23–28). <https://doi.org/10.18653/v1/2022.sigmorphon-1.3>
- Hutin, Mathilde & Allasonnière-Tang, Marc.** 2022c. Crowd-sourcing for less-resourced languages: Lingua Libre for Polish. In Proceedings of SIGUL 2022 (pp. 41–47). Marseille, France: European Language Resources Association (ELRA).
- Hutin, Mathilde & Allasonnière-Tang, Marc.** 2022d. Operation LiLi: Using crowd-sourced data and automatic alignment to investigate the phonetics and phonology of less-resourced languages. *Languages*, 7(3). <https://doi.org/10.3390/languages7030234>
- Kisler, Thomas., Reichel, Uwe., & Schiel, Florian.** 2017. Multilingual processing of speech via web services. *Computer Speech & Language*, 45, 326–347. <https://doi.org/10.1016/j.csl.2017.01.005>
- Kipp, Andreas., Wesenick, Maria-Barbara., & Schiel, Florian.** 1997. Pronunciation modeling applied to automatic segmentation of spontaneous speech. *Proceedings of Eurospeech*, 1023–1026.
- Klinger, Thomas A. & Lyche, Chantal.** 2012. “Cajun” French in a non-Acadian community A phonological study of the French of Ville Platte, Louisiana. *Phonological Variation in French: Illustrations from Three Continents*, edited by Randall Gess, Chantal Lyche & Trudel Meisenburg, John Benjamins Publishing Company, 275–313.
- Labov, William.** 2001. *Principles of linguistic change, volume 2: Social factors*, Oxford: Blackwell.
- Lamel, Lori, Gauvain, Jean-Luc, & Eskénazi, Maxine.** 1991. BREF, a large vocabulary spoken corpus for French. *Proceedings of the European Conference on Speech Technology*, EuroSpeech, pp. 505–508, Genoa.
- Lamel, Lori, Rosset, Sophie, Bennacef, Samir, Bonneau-Maynard, Héléne, Devillers, Laurence & Gauvain, Jean-Luc.** 1995. Development of spoken language corpora for travel information. *Proceedings of the European Conference on Speech Technology*, EUROSPEECH, Madrid.
- Lennig, Matthew.** 1978. *Acoustic Measurement of Linguistic Change: The Modern Paris Vowel System*, Ph.D. dissertation, University of Pennsylvania.
- Léon, Pierre.** 1992. *Phonétisme et prononciation du français*. Armand Colin, Paris.
- Lyche, Chantal.** 2010. Le français de référence: éléments de synthèse. *Les variétés du français parlé dans l'espace francophone: ressources pour l'enseignement*. S. Detey, J. Durand, B. Laks and C. Lyche. Paris/Gap, Ophrys: 143–165.
- Lyche, Chantal & Skattum, Ingse.** 2012. The phonological characteristics of French in Bamako, Mali: A sociolinguistic approach. *Phonological Variation in French: Illustrations from Three Continents*, edited by Randall Gess, Chantal Lyche & Trudel Meisenburg, John Benjamins Publishing Company, 73–101.
- Marjou, Xavier.** 2021. OTEANN: Estimating the Transparency of Orthographies with an Artificial Neural Network. *Proceedings of the Third Workshop on Computational Typology and Multilingual NLP* (pp 1–9), Online. Association for Computational Linguistics.
- Martinet, André.** 1945. *La Prononciation du français contemporain*, Genève, Droz, 1945.
- Martinet, André.** 1969. «C'est jeuli, le Mareuc!». *Le français sans fard*. Paris, Presses Universitaires de France.
- Martinet, André & Walter, Henriette.** 1973. *Dictionnaire de la prononciation française dans son usage réel*. Paris: France-Expansion.
- McAuliffe, Michael, Socolof, Michaela, Mihuc, Sarah, Wagner, Michael, & Sonderegger, Morgan.** 2017. Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Proceedings of Interspeech 2017*. 498–502.
- Métral, Jean-Pierre.** 1977. Le vocalisme du français en Suisse romande: considérations phonologiques. *Cahiers Ferdinand de Saussure* 31, 145–176.
- Mettas, Odette.** 1979. *La prononciation parisienne: Aspects phonique d'un sociolecte parisien (du Faubourg Saint-Germain à la Muette)*, Paris: Conseil International de la Langue Française et Laboratoire des Langues et Civilisations à Tradition Orale du CNRS.
- Østby, Kathrine Asla.** 2016. *Les voyelles orales à double timbre dans le parler de la haute bourgeoisie parisienne: analyse acoustique et diachronique*. Thèse de doctorat. Universités d'Oslo & Paris-Nanterre.
- Papen, Robert A. & Kevin J. Rottet.** 1997. Cajun French in two Louisiana parishes. *French and Creole in Louisiana*, ed. Albert Valdman, 71–108. New York: Plenum.
- Picoche, Jacqueline & Christiane Marchello-Nizia** (1994). *Histoire de la langue française*. Paris, Nathan.
- Pohl, Jacques.** 1985. Le français de Belgique est-il belge? *Présence francophone* 27, 9–19.

- Racine, Isabelle & Andreassen, Helene N.** 2012. A phonological study of a Swiss French variety Data from the canton of Neuchâtel. *Phonological Variation in French: Illustrations from Three Continents*, edited by Randall Gess, Chantal Lyche & Trudel Meisenburg, John Benjamins Publishing Company. 173–207.
- Reichstein, R.** 1960. Étude des variations sociales et géographiques des faits linguistiques. *Word* 16, 55–99.
- Schiel, Florian.** 1999. Automatic Phonetic Transcription of Non-Prompted Speech. In Ohala, J.J. et al. (eds), *Proceedings of the XIVth International Congress of Phonetic Sciences: ICPhS 99*; San Francisco, 1–7 August 1999, 607–610.
- Schouwey, Véronique.** 2008. *Les variantes cantonales dans la prononciation des voyelles du français en Suisse romande*. Mémoire de Master en Logopédie, Université de Genève.
- Smith, John Charles.** 2016. French and Northern Gallo-Romance. *The Oxford Guide to the Romance Languages*, edited by Adam Ledgeway and Martin Maiden, Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199677108.003.0018>
- Sweeney, L.** 2000. Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, *Data Privacy Working Paper* 3. Pittsburgh 2000.
- Tchkhovrebova, Ekaterina.** 2019. *La variation dans l'inventaire vocalique du français contemporain: approche dialectométrique*. Mémoire de Master, Université de Lausanne.
- Torreira Francisco, Adda-Decker Martine & Ernestus Mirjam.** 2010. The Nijmegen Corpus of Casual French. *Speech Communication*, Elsevier: North-Holland, 52(3), 201–212. <https://doi.org/10.1016/j.specom.2009.10.004>
- Walker, Douglas.** 1984. *The pronunciation of Canadian French*. Ottawa: University of Ottawa Press.
- Walker, Douglas.** 2012. Albertan French phonology French in an anglophone context. *Phonological Variation in French: Illustrations from Three Continents*, edited by Randall Gess, Chantal Lyche & Trudel Meisenburg, John Benjamins Publishing Company. 341–368.
- Walter, Henriette.** 1982. *Enquête phonologique et variétés régionales du français*. Paris, Presses Universitaires de France.
- Walter, Henriette.** 1988. *Le français dans tous les sens*. Paris: Editions Robert Laffont.
- Wells, John C.** 1997. SAMPA computer readable phonetic alphabet. In Gibbon, D., Moore, R. & Winski, R. (eds.), *Handbook of Standards and Resources for Spoken Language Systems*. Mouton de Gruyter, Part IV.
- Wickham, Hadley.** 2017. Tidyverse: Easily install and load the Tidyverse. *R Package Version*, 1.2.1. <https://CRAN.R-project.org/package=tidyverse>
- Winkelmann, Raphael, Jaensch, Klaus, Cassidy, Steve, & Harrington, Jonathan.** 2021. EmuR: Main Package of the EMU Speech Database Management System. *R Package Version*, 2.3.0.
- Woehrling, Cécile.** 2009. *Accents régionaux en français: perception, analyse et modélisation à partir de grands corpus*. Informatique [cs]. Université Paris Sud – Paris XI, 2009. Français. <tel-00617248>

Cite this article: Hutin M and Allasonnière-Tang M (2023). L'apport des données participatives pour l'étude linguistique des français du monde: le cas de l'opposition /a~ɑ/. *Journal of French Language Studies*, 1–24. <https://doi.org/10.1017/S0959269523000200>