



**HAL**  
open science

## Un nouvel accès à la parole spontanée : les vocaux

Julie Glikman, Camille Fauth

► **To cite this version:**

Julie Glikman, Camille Fauth. Un nouvel accès à la parole spontanée : les vocaux. JEP 2022, 2022, Noirmoutier, France. pp.154-162, 10.21437/JEP.2022-17 . hal-04312630

**HAL Id: hal-04312630**

**<https://hal.science/hal-04312630>**

Submitted on 6 Dec 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.



## Un nouvel accès à la parole spontanée : les vocaux

Julie Glikman<sup>1</sup> Camille Fauth<sup>1</sup>

(1) LiLPa, Université de Strasbourg, LiLPa UR 1339, F-67084 Strasbourg, France  
glikman@unistra.fr, cfauth@unistra.fr

### RÉSUMÉ

---

Le projet Les Vocaux vise la constitution d'un grand corpus de données inédites de français parlé spontané : les messages audio échangés dans des fils de conversation, le plus souvent appelés « des vocaux » par les utilisateurs. Dans cette contribution, nous présentons les premiers résultats d'une étude exploratoire sur les vocaux, nous permettant de tester la faisabilité et l'intérêt de la constitution d'un corpus de ces données inédites de français parlé spontané. Nous avons ainsi recueilli près d'une heure de données de vocaux, auprès de sept participantes. Nous présentons ici la méthodologie mise en place, ainsi que les principales caractéristiques phonétique, lexicale, syntaxique et discursive de ces vocaux. Les analyses montrent à la fois la diversité et la richesse des données, qui peuvent intéresser différents domaines d'étude de la parole.

### ABSTRACT

---

#### **A new access to spontaneous speech: voice mails.**

The "Les Vocaux" (i.e. voice mail) project aims at building a large corpus of original data of spontaneous spoken French: the audio messages exchanged in conversation threads, most often called "voice mail" by users. In this contribution, we present the first results of an exploratory study on voice mail, allowing us to test the feasibility and the interest of building a corpus of these data. For this study, we collected nearly one hour of data from seven participants. We present here the methodology used, as well as the main phonetic, lexical, syntactic and discursive characteristics of these voice mails. The analyses show both the diversity and the variety of the data, which can be of use in different areas of speech studies.

---

**MOTS-CLÉS :** corpus, oral spontané, message vocal, nouveaux modes de communication, communication médiée par smartphone.

**KEYWORDS:** corpus, spontaneous speech, voice mail, new modes of communication, communication mediated by smartphone.

---

## 1 Introduction

Malgré la multiplication des corpus oraux (voir par exemple : <https://www.ortolang.fr/>), l'accès à la parole spontanée naturelle, hors de tout contexte d'entretien linguistique, reste encore difficile et les données de ce types sont encore trop peu nombreuses. L'objectif du projet *Les Vocaux* est de combler ce manque, par la constitution d'un corpus de données de français parlé spontané inédites : les « vocaux », également appelés « notes vocales » ou « messages vocaux ». En effet, nous avons observé dans la vie de tous les jours le développement de l'utilisation d'enregistrements vidéo ou audio, diffusés sous forme de messages, à destinataires uniques ou multiples, via des plateformes de

type Snapchat, Instagram, des logiciels de messagerie type WhatsApp ou Messenger, ou encore directement via le système de messagerie du téléphone mobile. Ces productions sont de véritables « sms vocaux/vidéo », se substituant à l'envoi de sms, sans pour autant constituer un retour à la conversation téléphonique. L'importance de ce phénomène est confirmée par les articles de presse que l'on peut trouver à ce propos (voir par exemple Lavoustet, 2019). Ces productions sont le parallèle, en code oral, du sms écrit, dont l'étude s'est développée grâce au projet sms4science<sup>1</sup> ou grâce au projet What's up, Switzerland<sup>2</sup>. Ces messages constituent ainsi pour le chercheur un accès privilégié à la langue spontanée dans un contexte de production non dirigée, aux échanges de type privés, non surveillés, nécessaires à l'étude de la langue naturelle. Leur exploitation est rendue possible de par leur nature de productions enregistrées sur un support (le message lui-même). Pourtant, c'est une ressource jusqu'ici inexploitée. La seule étude sur les vocaux, à notre connaissance, est celle d'Estade & le Maire (2018), qui n'est pas fondée sur un recueil systématique et repose sur des données restreintes, et cherche surtout à définir le type de communication, qu'elles qualifient « d'intime » selon la grille d'analyse proposée par Koch & Oesterreicher (2001). Nous utiliserons également ce cadre pour caractériser plus finement ces productions. Les auteurs distinguent le médium employé, binaire (code oral ou écrit), et le concept, continu entre la distance ou la proximité communicative. Les vocaux, de code oral, comportent plusieurs caractéristiques relevant de la proximité communicative : communication privée, destinée à un interlocuteur connu (mais pas nécessairement intime, certaines personnes interrogées utilisant les vocaux en contexte professionnel), globalement spontanée (la plupart des plateformes ne permettant pas la réécoute du message avant envoi), dont l'émotionnalité peut être élevée (ils sont cependant rarement employés pour des événements très graves, d'après nos entretiens). Toutefois, d'autres critères tendent vers le pôle de la distance communicative : il s'agit de monologues et la coopération communicative est faible (les messages étant enregistrés par chaque locuteur séparément, il n'y a pas de chevauchement, ni de co-production du message, etc.), il n'y a pas coprésence spatiale, ni a priori temporelle. Ainsi, ils partagent des caractéristiques des sms écrits, tout en relevant du code oral. Ils se distinguent également des conversations téléphoniques, qui ont des attributs les rapprochant davantage du pôle de la proximité communicative (dialogue, coprésence temporelle et coopération communicative intense). Enfin, bien qu'en partageant certaines caractéristiques, les vocaux diffèrent également sensiblement des messages vidéos type « réels » sur TikTok ou Instagram, où la communication est davantage préparée et l'interlocuteur public. Ainsi, les vocaux constituent une catégorie spécifique dans l'échelle communicative, dont nous nous proposons ici d'étudier les caractéristiques formelles.

Dans cette contribution, nous présentons les résultats obtenus dans l'étude préparatoire (voir §4), afin de montrer l'intérêt de cette ressource pour l'étude de la langue. Les questions méthodologiques et techniques inhérentes à l'édition d'un tel corpus seront également soulevées. Nous commencerons par présenter le protocole de recueil et traitement des données mis en place pour cette étude exploratoire (§2), puis nous présenterons les premiers résultats (§3). Nous verrons ainsi que, malgré leur taille restreinte, les données recueillies laissent entrevoir les apports qu'un corpus de vocaux peut avoir dans différents domaines pour l'étude de la parole spontanée.

## 2 Protocole de recueil et traitement des données exploratoires

Avant le lancement du projet à grande échelle, nous avons effectué une étude préliminaire auprès d'un nombre réduit de participants et établi un corpus martyr, afin de tester le protocole, la robustesse de la chaîne de traitement et les éléments d'annotation. Cette partie est consacrée à la méthodologie

---

1 <http://www.sms4science.org/>, voir Panckhurst et al., 2013 pour le corpus 88milSMS recueilli à Montpellier.

2 Stark (2016-2018) [www.whatsup-switzerland.ch](http://www.whatsup-switzerland.ch).

suivie lors de la constitution de ce corpus test. Nous présenterons dans un premier temps les aspects liés au recueil, puis le traitement appliqué aux données.

Après un appel à volontaires, sept locutrices (voir table 1) ont accepté de participer à l'étude. Dans un premier temps, nous avons fait passer un entretien à chaque participante, pour interroger leur pratique d'utilisation des vocaux : le nom utilisé pour les désigner, le canal d'envoi, la fréquence d'utilisation, les destinataires habituels, le contexte d'envoi, le contexte d'écoute, le rapport vocal / sms / appel, le choix d'utilisation ou non d'un vocal, la réécoute/correction des vocaux, le jugement sur la durée moyenne d'un vocal. À la fin de l'entretien, la participante a raconté un événement de sa vie sans interruption de l'enquêtrice, afin de recueillir un échantillon de parole comparable (non exploité ici). Les entretiens ont été enregistrés avec un enregistreur numérique Tascam DR-05X dans une pièce calme, ou à distance via la plateforme Big Blue Button (<https://bbb.unistra.fr/>). Nous avons également posé quelques questions pouvant permettre d'interpréter les données : âge, occupation actuelle, lieu d'habitation actuel et lieu de son enfance, langue maternelle. Les participantes ont été informées de la teneur de l'étude et ont consenti à partager leurs données dans ce cadre en complétant un formulaire ad-hoc. Chaque participante nous a ensuite partagé les vocaux de son choix (de 5 à 14 vocaux selon les cas). Les vocaux, reçus dans différents formats (.ogg, .mp3, .mov, .mp4 dans le cas de captures vidéo), ont été triés, convertis en fichiers son (au format .wav) via le logiciel VLC, et anonymisés manuellement. Nous avons notamment exclu les vocaux qui comportaient un deuxième locuteur n'ayant pas signé le formulaire de consentement. Si la capture vidéo comportait plusieurs envois de suite, les fichiers ont été découpés afin de correspondre à un envoi par fichier. Nous avons fait le choix de ne pas recourir à une procédure d'anonymisation par transformation de la voix (Fang et al., 2019 ; Maouche et al., 202 ; Pavankumar Dubagunta et al., 2022, i.a.), en prenant le parti que hors contexte, la reconnaissance stricte du locuteur est difficile (Hollien, 2002). Nous avons en revanche fait le choix d'anonymiser les noms / prénoms et les éventuels lieux liés à l'activité de la locutrice en générant un bruit via le logiciel Audacity. La table 1 récapitule les principales données du corpus. Les mesures de durée ont été effectuées de la première à la dernière syllabe prononcée.

Id	Âge	Lieu enfance	Vocaux retenus	Durée totale	Durée moyenne
1	41	Paris	2	370,96	185,46 (24,44)
2	20	Haut-Rhin	14	198,07	14,15 (16,98)
3	23	Haut-Rhin	5	50,9	10,18 (9,18)
4	22	Belgique flamande / Alsace	5	60,32	12,06 (3,67)
5	21	Moselle	9	343,07	38,12 (6,13)
6	26	Haut-Rhin / Normandie	11	247,12	22,46 (19)
7	19	Alsace	9	298,97	32,21 (9,99)

TABLE 1 : Présentation du corpus - les durées sont en secondes, les valeurs entre parenthèses correspondent aux écarts-types.

Nous avons ensuite procédé à la transcription orthographique. Pour cette étape, nous avons testé l'outil de reconnaissance de la parole Vocapia (<https://www.vocapia.com/>) et l'éditeur en ligne YobiYoba qui utilise la même technologie (<https://www.yobiyoba.com/fr/>), avec la collaboration de Lori Lamel (Limsi), en parallèle de la transcription manuelle. Les résultats étaient meilleurs en

cochant l'option « conversation téléphonique ». Certains messages étaient entièrement reconnus de manière correcte, et d'autres comportaient plusieurs erreurs (mots mal identifiés, même provenant du vocabulaire standard comme « fond de teint » (transcrit « temps »), mots omis, notamment plusieurs « voilà » conclusifs)<sup>3</sup>. Globalement, les résultats étaient tout de même meilleurs que nous l'aurions pensé, étant donné le type de données et la qualité du signal. Le gain de temps sur une transcription manuelle reste cependant encore à établir. Les transcriptions automatiques alignées, au format de sortie xml, ont ensuite été converties au format Textgrid pour être traitées dans Praat (Boersma, Weenink, 2021) grâce à un script de conversion créé par Nicolas Mazziotta (Université de Liège). Les transcriptions orthographiques ont été ainsi soit corrigées soit directement faites dans Praat. Les conventions adoptées ont été de respecter l'orthographe traditionnelle pour les mots, sans trace des éventuelles réductions phonétiques, notamment pour faciliter des recherches plein texte : *enfin* pour [fɛ̃], *je suis* pour [ʃi], *je vais* pour [zve] et tout en restant au plus proche des formes effectivement perceptibles : omission du *ne* de négation sauf quand audible, *y a* sauf si *il* audible.

Les données ont ensuite été alignées au mot, à la syllabe et au phonème grâce à EasyAlign (Goldman, 2012). Nous nous sommes concentrées dans un premier temps sur l'alignement à la syllabe et au phonème, entièrement corrigé manuellement<sup>4</sup>. Nous avons ensuite indiqué sous forme d'annotations spécifiques dans Praat les pauses, distinguées entre pauses internes au syntagme syntaxique (Pi) et pauses externes, y compris dans le cas de coordination de syntagme (Po), les *fillers* (définis comme apparaissant entre deux pauses) (F), allongements (même phonème) (A) et *fillers* + / - (phonème différent allongé collé à un autre phonème à la fin ou au début du mot) (F+/-), les faux départs (FD), termes d'adresse (AD) et formules de salutation (SA).

### 3 Premiers résultats

#### 3.1 Qualité des vocaux

Comme l'on pouvait s'y attendre, la qualité des fichiers reçus diffère considérablement d'un vocal à l'autre. Enregistrés parfois dans la rue, en marchant, les vocaux peuvent contenir des bruits parasites (sirènes de pompier, bruits environnants), ou de mauvaises captures du signal (micro éloigné ou main sur le micro), pouvant rendre le message inaudible. En dehors d'une participante qui déclare envoyer principalement des vocaux en marchant (voire « en faisant ses courses »), les autres participantes disent au contraire éviter au maximum les endroits bruyants, ou les endroits où il y a du monde autour (tram, etc.). Les bruits de respiration sont également parfois perceptibles. Cependant, dans la majorité des cas, la compréhension n'est pas gênée et le signal est bon, ce qui rendrait possible des analyses prosodiques plus fines grâce au repérage des courbes intonatives. Les participantes ont confirmé que cela arrive régulièrement de recevoir des vocaux partiellement inaudibles. Cela confirme le caractère spontané de ces messages, non corrigés avant envoi, souvent parce que la plateforme utilisée ne le permet pas (seuls certains systèmes permettent de réécouter le message avant de l'envoyer). Après envoi, on peut supprimer le message de la conversation, mais cela se voit, et donc les usagers préfèrent éviter cette solution. Certaines participantes déclarent réécouter le message après envoi

---

3 Ex. 06\_01 (12 sec.) : transcription automatique : « j'ai mon entretien là à dix-sept heures trente/ trop peur je suis trop stresse depuis ce matin laisse tomber/ horrible je trop je préfère morte/ bon », corrigé : *\_j'ai mon entretien là à dix-sept heures trente \_j'ai trop peur je suis trop en stress depuis ce matin laisse tomber \_pfff\_ horrible je stresse trop je stresse à mort ça me \_wahou.*

4 La suite de la chaîne de traitement s'appuiera sur l'alignement au mot orthographique, et consistera en une annotation en partie du discours et une annotation syntaxique, mais cet aspect ne sera pas développé ici.

(mais pas toutes, ni systématiquement), pour éventuellement corriger un oubli par un nouveau message<sup>5</sup>.

### 3.2 Utilisation des vocaux

Les participantes représentent un large éventail de types d'usage des vocaux : rarement ou au contraire très fréquemment utilisés, avec des locuteurs spécifiques ou au contraire n'importe quel locuteur, uniquement en contexte privé (famille, amis) ou également en contexte professionnel, dans des discussions de groupe comme pour des destinataires uniques. Ceci confirme, selon nous, l'intérêt de constituer un large corpus de vocaux, qui pourraient fournir un échantillon varié de parler spontané, dans différents registres. Concernant la durée, les configurations techniques des plateformes utilisées peuvent avoir aussi une influence. En effet, certaines plateformes limitent les vocaux à une minute maximum, tandis que d'autres n'ont pas de limites, les locutrices pouvant aussi choisir de faire un seul message long (c'est le cas d'une locutrice, qui nous a fourni des vocaux allant jusqu'à 10 min.), ou plusieurs messages courts (nous avons ainsi reçu des cas de vocaux qui se suivent, comme un seul envoi en plusieurs messages, nous les avons identifiés autant que possible par une sous-numérotation). Le choix d'envoi des vocaux, selon notre analyse des entretiens menés, semble lié en premier lieu à l'objectif communicationnel (pour « raconter quelque chose de long »), avant les aspects pratiques (quand impossibilité d'écrire), et la volonté de transmettre une émotion, sauf pour les locutrices qui emploient dans tous les cas majoritairement des vocaux, y compris pour des messages brefs (« ok »), et se fait le plus souvent au détriment du sms écrit plutôt que de la conversation téléphonique, classée à part (et, pour certaines, rarement utilisée, ou dans des contextes spécifiques)<sup>6</sup>. Toutefois, si l'on demande si elles enverraient un vocal pour annoncer un événement grave (décès par ex.) ou important, la réponse est globalement négative (certaines ajoutent un message écrit pour signifier l'importance d'écouter un vocal urgent, ou inversement pour dire que le vocal n'est pas urgent). Enfin, pour la question de l'écoute des vocaux reçus, bien que l'écoute en accéléré est maintenant possible selon les plateformes, elle n'est pas systématiquement utilisée<sup>7</sup>. Si le message ne peut pas être instantanément écouté, un message écrit est envoyé pour le préciser (l'important étant d'éviter de laisser un message sans réponse, ce qui porte le nom de « mettre un vu », la plupart des plateformes permettant maintenant de signaler à l'émetteur si le message a été « vu », c'est-à-dire affiché, sur le téléphone du destinataire).

### 3.3 Caractéristiques phonétiques et prosodiques

Les caractéristiques phonétiques et prosodiques reflètent l'hétérogénéité des usages précédemment évoquée (voir table 1). Ainsi, le vocal le plus court de notre corpus compte 7 syllabes pour une durée de 2,54 secondes tandis que le plus long dure 202,75 secondes (714 syllabes). La durée des vocaux est donc très variable d'un locuteur à l'autre mais également parfois d'un message à l'autre comme en témoigne les écarts types (valeurs entre parenthèses dans la colonne DMO de la table 1). Si les ajouts de « euh » sont les phénomènes d'hésitation les plus fréquents dans notre corpus, il est

---

5 Nous avons également observé cette pratique de correction par ajout dans les conversations de sms écrits, ce qui constitue à notre avis une piste intéressante à étudier pour le sentiment linguistique des locuteurs.

6 Les entretiens ont également mis au jour le fait que le sms (écrit ou même vocal) remplace le courriel, et est considéré comme moins intrusif. Cela est d'autant plus vrai que des plateformes comme Instagram permettent les envois de messages privés, reçus sur le téléphone et donc considérés comme des sms, tout en passant uniquement par un profil public, sans nécessairement avoir les coordonnées téléphoniques de la personne.

7 La question de la réception et compréhension en écoute accélérée sera une piste intéressante à explorer par la suite. De même, d'un point de vue communicationnel, il est surprenant de noter que toutes les locutrices ont déclaré qu'il leur arrive, même rarement, de ne pas écouter la totalité des vocaux reçus, notamment en cas de « retard » à rattraper sur l'écoute, ou quand le vocal est long et « n'est pas important ».

intéressant de noter qu'ils peuvent être observés entre deux pauses (25 occurrences), directement collés au début d'un mot (26 occurrences) mais que c'est évidemment en position finale, même lorsque le mot ne contient pas de schwa final qu'ils sont principalement produits (218 occurrences). Leurs durées varient beaucoup. Les allongements d'autres voyelles restent minoritaires dans notre corpus (80 occurrences) et de façon intéressante, certaines locutrices n'en produisent quasiment pas comme la locutrice 1 qui préfère l'ajout de schwa finaux (66 occurrences) même lorsque le mot se termine par une voyelle. Enfin, de nombreuses pauses, syntaxiques et non syntaxiques, ont été annotées faisant du vocal un exemple intéressant de l'oral spontané dans un contexte non contrôlé. Cette caractéristique est également observable dans le ratio disfluences (somme des durées de Pi, F, F+, F- et A) / durée totale qui permet de voir que pour la locutrice 1 qui enregistre ses vocaux dans la rue en vaquant à ses occupations se plaçant ainsi dans une situation de double tâche, les éléments disfluents représentent 24,6% du corpus. Ce corpus, de taille modeste, permet déjà d'identifier des usages différents mais également des stratégies individuelles dans la production ce type de parole.

	Syl	Po	Pi	F	F+	F-	A	Dis/DT
Loc1	1296	82	84	8	66	6	6	24,60
Loc2	791	50	27	7	26	4	7	16,30
Loc3	230	8	7	0	10	0	1	14,01
Loc4	250	10	3	0	6	1	6	15,44
Loc5	1508	56	26	3	42	1	21	12,83
Loc6	1003	79	36	1	25	2	20	12,82
Loc7	1321	79	44	6	43	12	19	17,71
Totaux	6399	364	227	25	218	26	80	

TABLE 2 : Caractéristiques phonétiques du corpus par locuteurs tous vocaux confondus (Syl = Nombre de syllabes ; Po = Pause entre deux syntagmes ; Pi = pause à l'intérieur d'un syntagme ; F [∅] = entre deux pauses silencieuses ; F+ = [∅] produit en fin de mot ; F- = [∅] produit avant un mot ; A = allongement final d'une voyelle ; Dis/DT = ratio disfluences/durée totale)

### 3.4 Caractéristiques lexicale, syntaxique, discursive et énonciative

Les données recueillies, malgré la taille restreinte de l'étude, permettent déjà d'observer un certain nombre de phénomènes, montrant l'intérêt d'un tel corpus dans différents domaines pour des études futures. Comme l'on pouvait s'y attendre, on retrouve les éléments habituels qui caractérisent l'oral spontané (Blanche-Benveniste et al., 1990), tels que les faux départs (piétinements, anticipation, amorce avec ou sans reformulation, cf. ex. 1), travail de dénomination, hésitations, y compris dans les messages courts. On observe également la présence de marques de structuration du discours, comme l'emploi de *voilà* conclusif (Col et al., 2015), de nombreux marqueurs de discours (*enfin, bah, tu vois, genre, du coup, en mode ...* (ex. 1, 5, 7) (cf. i.a. Rossari & Jayez, 2000, Rosier, 2002, Chauveau-Thoumelin, 2016, Rodriguez Somolinos (éd.), 2011, 2016, Glikman et al. (éds) 2019), et des marques de l'interaction et des discours sur la construction du message même : termes d'adresses, formules de salutation d'ouverture ou de clôture, annonces de fin du message (ex. 2), voire discours sur l'interaction même (ex. 3). On trouve aussi bien des messages de type interactionnel (ex. 6) que des épisodes narratifs. Ces épisodes narratifs peuvent en outre fournir des exemples de discours rapporté au style direct (ou non) à l'oral (ex. 5, 7) (Caillat, 2013, Verine, 2005), difficilement élicitable en contexte d'entretien linguistique. Les vocaux peuvent également permettre l'étude de la diversité des formes de l'interrogation à l'oral (ex. 4, 6). Enfin, évidemment, les données nous donnent accès

à un large éventail de lexique, de différents registres, pour des éléments parfois difficiles à documenter, comme *meuf* en terme d'adresse (ex. 7), *go* (au sens de « allez ») (ex. 7), *sortir* comme introducteur de discours direct (ex. 5), *trop*, *en stress* (ex. 8).

1. pour me reconforter de ce partiel désastreux je me suis achetée \_ une tartelette aux noix et \_ un bretzel \_ j'avais j'avais besoin de réconfort tu vois (03\_03, durée totale 9,54 sec.)
2. [...] je t'envoie ce petit message (01\_03)
3. oh la la du coup comme j'ai dit que j'avais la flemme de parler tu me réponds à l'écrit en une phrase [...] (06\_08\_1)
4. [...] et qu'est-ce qui t'a déclenché cette pensée \_ explique-moi \_ t'es pas moche \_ je te le dis (06\_03)
5. [...] et il dit bonjour alors on est là genre bonjour \_ et euh \_ il sort à #name euh \_ ouais euh vous pensez à bien garder les distances de sécurité hein euh alors elle dit ben oui vous voyez bien qu'ils sont espacés \_ oui oui non mais je vois je vois hein [...] et il est parti on était là genre \_ ok #rire gênant (05\_08)
6. tu fais la gueule ou quoi (06\_08\_2)
7. meuf [...] comme la veille il m'avait dit de venir j'ai dû me dire genre vas y go [...] (06\_04)
8. [...] j'ai trop peur je suis trop en stress depuis ce matin laisse tomber [...] (06\_01)

## 4 Perspectives

Nous avons montré ici quelques exploitations possibles, à différents niveaux d'analyse. Cette ressource ouvre la possibilité à d'autres études de par son caractère très « écologique ». Des collaborations pourront être envisagées dans les domaines de la linguistique informatique ou le traitement du signal (reconnaissance automatique et débruitage du signal par exemple). Les analyses menées sur le corpus exploratoire montrent bien l'intérêt de la constitution d'un corpus de vocaux à grande échelle et les apports que ce corpus peut avoir pour la communauté dans différents domaines. Cette étude nous a également permis de tester le protocole, qui a été ajusté pour la création du corpus final. Pour le recueil à grande échelle, nous avons lancé un appel à la foule (*crowdsourcing*), à travers la diffusion d'un appel à participation sur les réseaux sociaux, comme cela a été fait dans d'autres projets (voir Avanzi et al., 2016 ; Glikman et al., 2018 ; Panckhurst et al., 2013 ; Stark, 2016-2018). Le consentement des participants est recueilli par un formulaire en ligne, qui présente toutes les implications de la participation à l'étude, et permet également de recueillir les informations socio-culturelles nécessaires à l'interprétation des données (âge, lieu d'habitation actuel et d'enfance, usage et pratique des vocaux). Les participants sont informés des éléments collectés et consentent à l'utilisation de leurs vocaux à usage de la recherche. Ils ont alors accès au numéro mis en place pour la collecte et partagent les vocaux de leur choix. L'ensemble du protocole a été validé par la déléguée à la protection des données de l'Université de Strasbourg.

Outre la transcription orthographique et l'alignement à la syllabe et au phonème, les données seront annotées en parties du discours et en syntaxe. Le corpus sera distribué librement dans différents formats compatibles avec différents logiciels de traitement des données.

## Remerciements

Les auteures remercient Lori Lamel (Limsi) pour nous avoir fourni la transcription automatique des vocaux via Vocapia et permis de tester l'éditeur YobiYoba et P. Boula de Mareuil (Limsi) qui nous a mis en contact avec L. Lamel. Cette recherche a été menée dans le cadre d'une délégation CNRS



effectuée par J. Glikman au laboratoire ATILF (UMR 7118). Les membres du projet Les Vocaux sont J. Glikman (LiLPa), C. Fauth (LiLPa), C. Benzitoun (ATILF) et N. Mazziotta (Université de Liège).

## Références

AVANZI M., BARBET C., GLIKMAN J., PEUVERGNE J. (2016). Présentation d'une enquête pour l'étude des régionalismes du français. SHS Web of Conferences (Vol. 27, 03001), 5<sup>e</sup> Congrès Mondial de Linguistique Française (CMLF). <https://doi.org/10.1051/shsconf/20162703001>

BLANCHE-BENVENISTE C. *et al.* (1990). *Le français parlé*. Paris : CNRS éditions.

BOERSMA P., WEENINK D. (2021). *Praat: Doing Phonetics by Computer*, Université d'Amsterdam: Phonetic Sciences.

CAILLAT D. (2013). Le discours rapporté direct dans les conversations orales: un système de balisage multimodal. C. Desoutter, C. Mellet. *Le discours rapporté et ses marques*, 178. Berlin: Peter Lang. 63-80.

CANDEA M. (2000). Contribution à l'étude des pauses silencieuses et des phénomènes dits « d'hésitation » en français oral spontané, Thèse de doctorat, Université de la Sorbonne nouvelle, Paris III.

CHAUVEAU-THOUMELIN P. (2016). De l'exemplification à la catégorisation approximative: étude de la construction [[X] SN genre [Y] SN]. In *SHS Web of Conferences* (Vol. 27, p. 12005), CMLF 2016.

COL G., DANINO C., RAULT J. (2015). Éléments de cartographie des emplois de *voilà* en vue d'une analyse instructionnelle. *Revue de Sémantique et Pragmatique* [En ligne], 37 | 2015, mis en ligne le 01 juin 2016, consulté le 12 janvier 2022. URL : <http://journals.openedition.org/rsp/1015> ; DOI : <https://doi.org/10.4000/rsp.1015>.

ESTADE L., le MAIRE J. (2018). Les messages vocaux sur WhatsApp: une forme hybride de communication. *Cahiers du Centre de Linguistique et des Sciences du Langage*, 2018, 55, 125-134.

FANG F., WANG X., YAMAGISHI J., ECHIZEN I., TODISCO M., EVANS N., BONASTRE J. F. (2019). Speaker anonymization using x-vector and neural waveform models. *arXiv preprint arXiv:1905.13561*.

GLIKMAN J., BENZITOUN C., GOLDMAN J.-P., SCHERRER Y., AVANZI M., BOULA de MAREÛIL P. (2018). Donnez votre Français à la Science ! Internet et la documentation de la diversité linguistique : présentation de la plateforme et premiers résultats, SHS Web of Conferences 46, 02003 (2018), *Congrès Mondial de Linguistique Française 2018*, <https://doi.org/10.1051/shsconf/20184602003>

GLIKMAN J., PARUSSA G., WALTEREIT R. (éds). (2019). Les marqueurs du discours en diachronie du français : nouvelles perspectives. *Studia Linguistica Romanica*, n° 2, 2019, <https://studialinguisticaromanica.org/index.php/slr/issue/view/2>.

GOLDMAN J.-P. (2011). *EasyAlign: An automatic phonetic alignment tool under Praat*, In *Proceeding of Interspeech 2011*. Florence, p. 3233-3236.

HOLLIEN H. (2002). *Forensic voice identification*. London : Academic Press.

KOCH P., OESTERREICHER W. (2001). Gesprochene Sprache und geschriebene Sprache / Langage parlé et langage écrit. in Holtus G., Metzeltin M., Schmitt Ch. (ed). *Lexikon der Romanistischen Linguistik*, Bd. I/2, Tübingen: Niemeyer. 584-627.

- LAVOUSTET O., (2019, 29 mai). «Oui, coucou c'est moi», ou l'art de communiquer par message vocal sur Whatsapp. *Le Temps*. <https://www.letemps.ch/societe/oui-coucou-cest-moi-lart-communiquer-message-vocal-whatsapp>
- MAOUCHE M., MOHAN LAL SRIVASTAVA B., VAUQUIER N., BELLET A., TOMMASI M. et al. (2020). A comparative study of speech anonymization metrics. INTERSPEECH 2020, Oct 2020, Shanghai, China. hal-02907918
- PANCKHURST R., DÉTRIE C., LOPEZ C., MOÏSE C., ROCHE M. et VERINE B. (2013). Sud4science, de l'acquisition d'un grand corpus de SMS en français à l'analyse de l'écriture SMS. *Épistémè - revue internationale de sciences sociales appliquées*, 9 : Des usages numériques aux pratiques scripturales électroniques, 107-138.
- PAVANKUMAR DUBAGUNTA S., van SON R. J.J.H., MAGIMAI.-DOSS M. (2022). Adjustable deterministic pseudonymization of speech, *Computer Speech & Language*, Volume 72, 2022, 101284, ISSN 0885-2308, <https://doi.org/10.1016/j.csl.2021.101284>.
- RODRÍGUEZ SOMOLINOS A. (éd.) (2011). Les marqueurs du discours – approches contrastives. *Langages*, 184. <https://doi.org/10.3917/lang.184.0003>
- RODRIGUEZ SOMOLINOS A. (éd.). (2016). Énonciation et marques d'oralité dans l'évolution du français. *Linx* 73. Nanterre : Presses universitaires de Paris Nanterre.
- ROSIER L. (2002). Genre: le nuancier de sa grammaticalisation. *Travaux de linguistique*, (1), 79-88.
- ROSSARI C., JAYEZ J. 2000. Du coup et les connecteurs de conséquence dans une perspective dynamique. *Lingvisticae investigationes* 23.2 (2000). 303-326.
- STARK E. (2016-2018). Projet FSN « What's up, Switzerland? » (Sinergia: CRSIII\_160714). Université de Zurich. [www.whatsup-switzerland.ch](http://www.whatsup-switzerland.ch).
- VERINE B. (2005). Dialogisme interdiscursif et interlocutif du discours rapporté: jeux sur les frontières à l'oral. J. Bres (ed.). *Dialogisme et polyphonie. Approches linguistiques*. Louvain-la-Neuve: De Boeck Supérieur. 187-200.