

Le projet *Les Vocaux*: Premières analyses

Julie Glikman et Nicolas Mazziotta

U. Strasbourg, LiLPa & ATILF ; U. Liège, Traverses

Montpellier, le 15 mai 2023

Les
V**CAUX**

Qu'est-ce qu'un « vocal » ?

- Concept de « SMS vocal », « message vocal », « note vocale » ou « vocal » :
 - Large variété de plateformes de communication les prenant en charge
 - Populaires et relayés dans la presse
 - Éventuellement pour critiquer, mais signe de visibilité
 - Tendance à se répandre dans les pratiques de tous
- Nouveau mode de communication :
 - Parallèle à l'envoi de SMS
 - Différent des appels
 - Différent des messages sur le répondeur ? (hypothèse à vérifier)

Qu'est-ce qu'un « vocal » ?

- Exemple:



euh la vidéo . . . où v- euh . . . on est en accéléré elle est hyper drôle . . . (LesVocaux, 02-01)

- En tant que actes d'énonciation:
 - Messages enregistrés
 - La récolte peut commencer par un simple transfert (cf. recueil des sms)
 - Contexte de production non surveillée (?) => réécoute des messages?
 - Accès à l'oral spontané (?)
- SMS bien étudiés (sms4science, What's up Switzerland)
- Peu d'études sur les vocaux (exceptions rares, comme Estade et le Maire 2018)

Objectifs du projet « Les Vocaux »

- Constituer un corpus rassemblant des messages vocaux
 - **Annotations** nécessaires :
 - Ph **phonétique** (phones, syllabation, pauses, alignement au signal)
 - Le **lemmatisation**
 - MS **morphosyntaxe** (PDD, morphologie, syntaxe dépendancielle)
 - Permettront d'évaluer les **interférences**
 - Distribuer le corpus dans des **formats** utilisables

Objectifs du projet « Les Vocaux »

Membres du projet

- Julie Glikman (porteuse), Université de Strasbourg
- Christophe Benzitoun, Université de Lorraine
- Camille Fauth, Université de Strasbourg
- Nicolas Mazziotta, Université de Liège
 - + Collaborateurs ponctuels
 - + Stagiaires étudiants

Objectif de l'exposé

- Présentation de la méthodologie et premier bilan du recueil
- Caractérisations générales des données
- Premières analyses sur le corpus
 - *Écoute*
 - *Avoir été / être allé*
 - *(croire / penser)*

Méthodologie et bilan du recueil

Deux étapes de recueil des données

1. Recueil expérimental préalable : « Martyr »
2. Recueil à grande échelle : « Campagne 2022 »

Martyr

- 7 participantes volontaires
 - Entretiens individuels sur les pratiques
 - Partage des vocaux (sept.-nov. 2021)
 - Recueil échantillon d'entretiens (matériaux contrastifs)
- **Données recueillies**
 - Nombre de locutrices : 7
 - Origine : Strasbourg et Paris
 - Nombre de vocaux : 66
 - Durée totale : 1h00m44.73s

Campagne 2022

- **Crowdsourcing** (cp. *Français de nos régions* et *Donnez votre français à la science*)
- **Avantages** : Recueil rapide d'une grande quantité de données réparties dans différents lieux
- **Inconvénients** : Quantité de données dépend de la réussite de la campagne ; pas de contrôle sur l'échantillon
- Campagne de recueil limitée dans le temps (avril – septembre 2022)
- **Problématique du recueil** : Nécessité de respecter la RGPD (voix = donnée sensible)

Campagne 2022

- ▶ **Question de l'anonymat** : voix vs données
 - ▶ **Besoin** : recueil du consentement explicite et éclairé des participants (simple transfert des messages ne suffit pas) comprenant le **nom et numéro de téléphone** du participant (pour faire le lien)
 - ▶ **Engagement** : non transfert des données privées (numéro de téléphone...); anonymisation des noms et indications reconnaissables dans le vocal ; mais pas de modification de la voix
 - ▶ **Solution adoptée** :
 - ▶ Questionnaire en ligne (consentement + métadonnées + habitudes d'utilisation)
- + transfert des vocaux à un numéro de téléphone

Métadonnées

- ▶ speaker : numéro d'identification du locuteur
 - ▶ gender : genre
 - ▶ age : âge
 - ▶ current location : lieu de vie actuel
 - ▶ current country : pays où se trouve le lieu de vie
 - ▶ home location : ville natale
 - ▶ home country : pays natal
 - ▶ message usage : fréquence d'utilisation des vocaux
 - ▶ filename : nom du fichier (référence du texte)
- Réponses non obligatoires au questionnaire

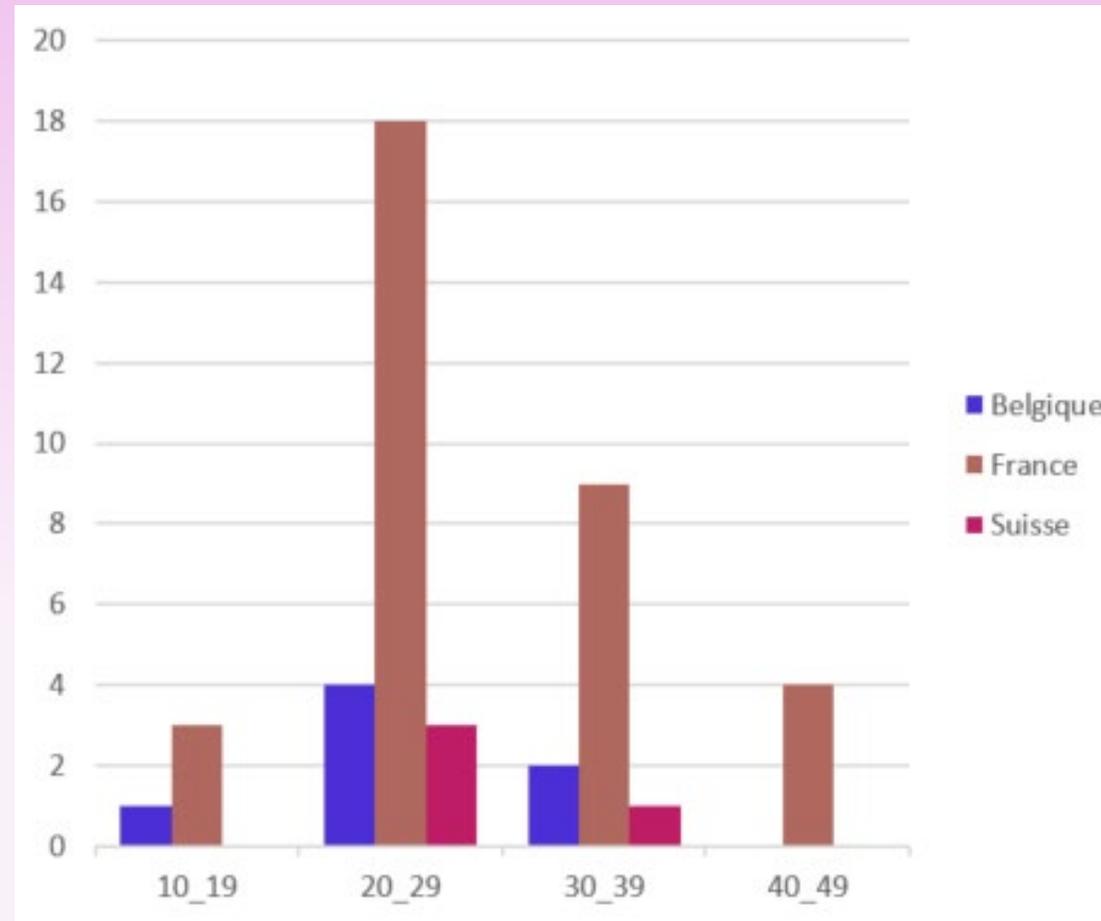
Campagne 2022

- ▶ **Bilan de la méthode de recueil :**
- ▶ Perte de participants en cours de route :
 - ▶ **445** accès au formulaire en ligne
 - ▶ **133** consentements
 - ▶ **117** questionnaires remplis jusqu'au métadonnées
 - ▶ **45** locuteurs ayant envoyé des vocaux
- ▶ Risque de filtrage des sms transmis
- ▶ Absence d'indication de l'interlocuteur (mais récupérable dans certains cas)

Campagne 2022

- ▶ **Bilan de la campagne :**
- ▶ Campagne de publicité ayant surtout attiré des **étudiants** (Facebook, Instagram, annonces aux cours)
 - ▶ **29** locuteurs entre 17 et 30 ans ; **11** locuteurs entre 31 et 39 ; **1** locuteur de 11 ans ; **4** entre 42 et 44 ans
 - ▶ **28** locuteurs de France ; **8** Belgique ; **4** Suisse ; **1** Canada
 - ▶ **35** femmes ; **7** hommes ; **1** non binaire
- ▶ Disproportion entre locuteurs : de 1 à plus de 100 vocaux transmis

Campagne 2022



Campagne 2022

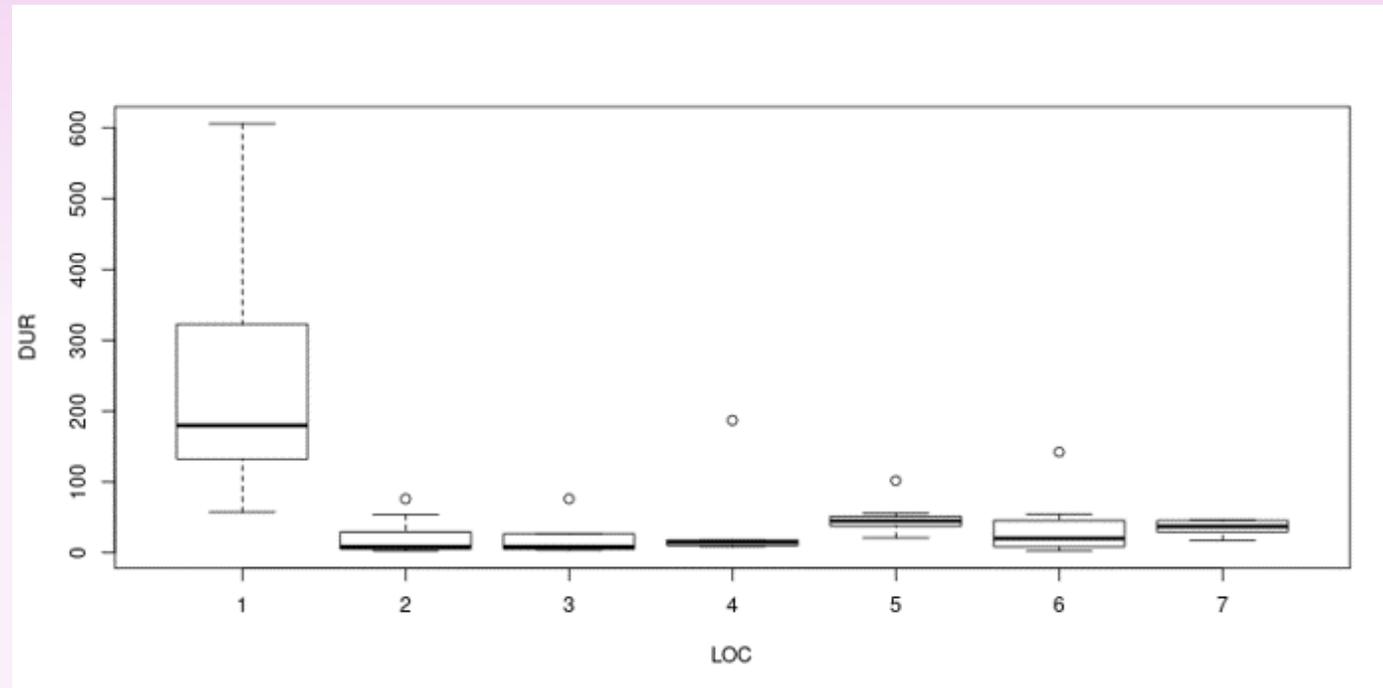
- ▶ Nombre de locuteurs : 45
- ▶ Nombre de vocaux : 1160
- ▶ Durée totale : 19h06m49.2s
- ▶ Une distribution variable des durées (non évaluées par locuteur)

Caractéristiques globales des données

Premières observations du corpus Les Vocaux

Martyr

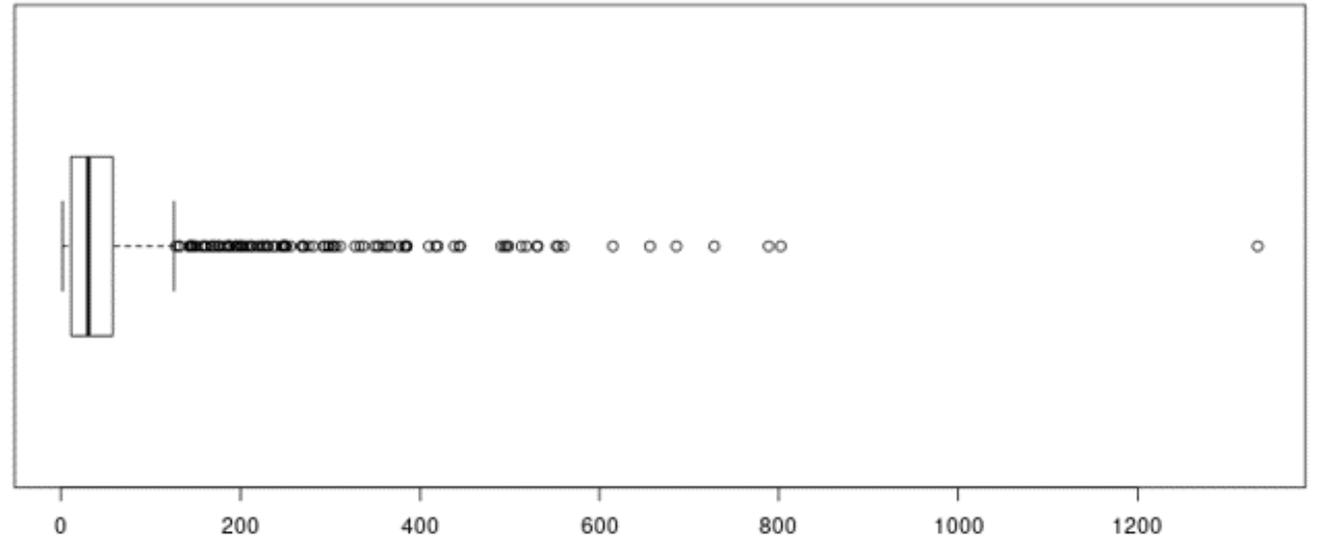
- Une distribution variable des durées (cas atypique)
 - Distribution des durées par locuteur



Campagne 2022

- Distribution des durées

➤ 580 vocaux de durée typique, 5 :10 :30.16



Min.	1st Qu.	Median	Mean	3rd Qu.	Max
1.313	10.992	30.241	59.318	57.561	1333.678

Caractéristiques phonétiques

- Pauses, hésitations, allongements, taux de disfluence...

➤ Usages
différents et
stratégies
individuelles

	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 [ø] = entre deux pauses silencieuses ; F+ = [ø] produit en fin de mot ; F- = [ø] produit avant un mot ; A = allongement final d'une voyelle ; Dis/DT = ratio disfluences/durée totale)

Caractéristiques phonétiques

- **Pauses, hésitations, allongements, taux de disfluence...**

... 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 ... (Les Vocaux, 03-03)



... papa ... j' ai trouvé un portefeuille ... et tu sais au mec et je dois lui envoyer par la poste ... mais il habite en Suisse je crois ... alors je dois faire comment ... je dois mettre un timbre spécial ... (Les Vocaux, 02-09)



- **Prosodie, Accent (à faire)**

... 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 ... (Les Vocaux, 06-01)



... ça marche je suis en train de partir du parking de l'uni justement ... alors je ... je t'appelle quand je suis parquée puis comme ça tu me dis où tu es ... allez ... à toute ... (Les Vocaux, 93-03)



Caractéristiques typiques de l'oral spontané

« lorsque nous produisons des discours non préparés, nous les composons au fur et à mesure de leur production, en laissant des traces de cette production » (Blanche-Benveniste *et al.* 1990 : 17)

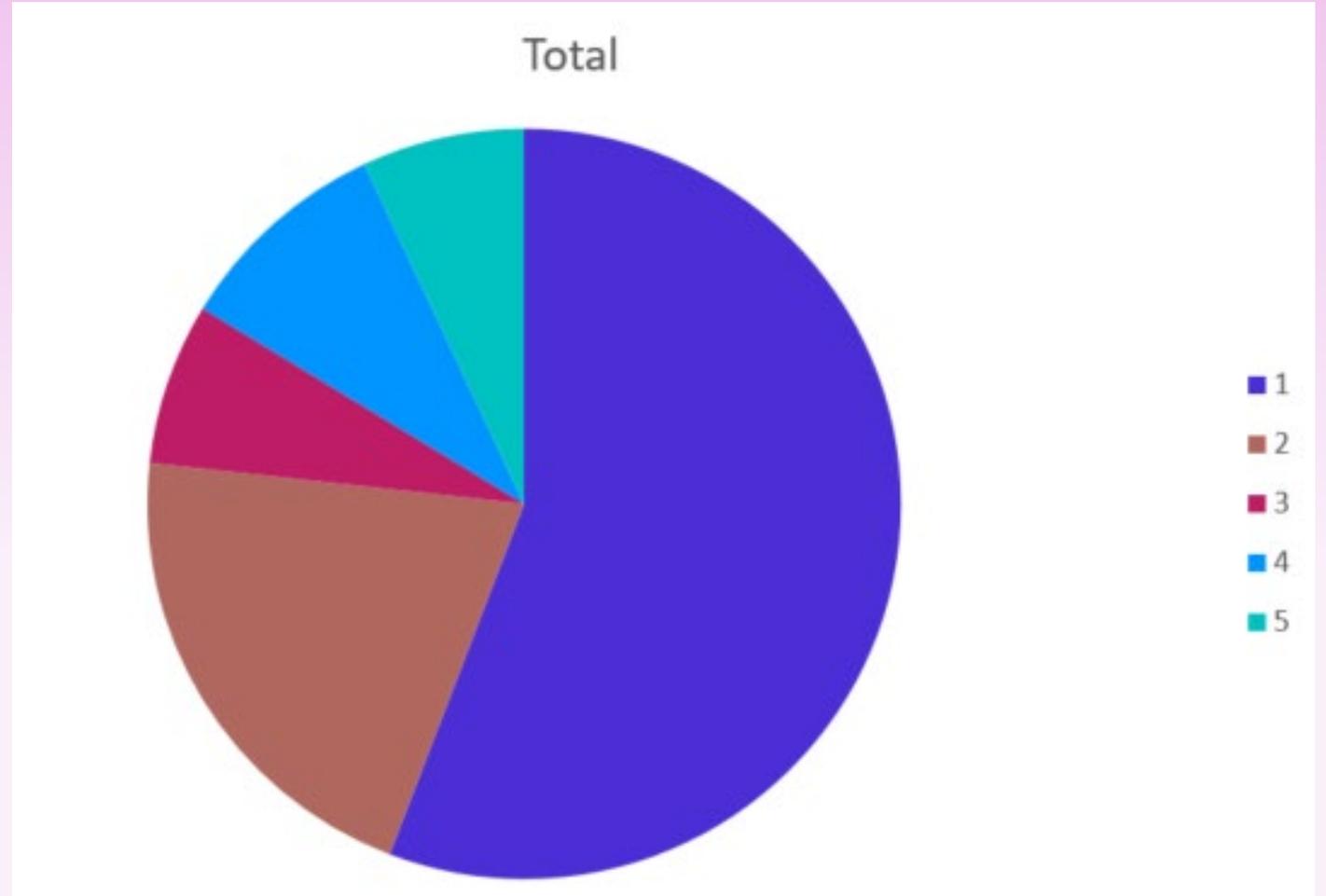
➤ Piétinements, amorce

... oui euh alors sur #name ... ben c' est c' est c' est je suis un peu embêtée pour toi que #name te fasse euh ... te f- te fasse vous fasse des difficultés euh (Les Vocaux, 01-03)



Accès à l'oral spontané non surveillé?

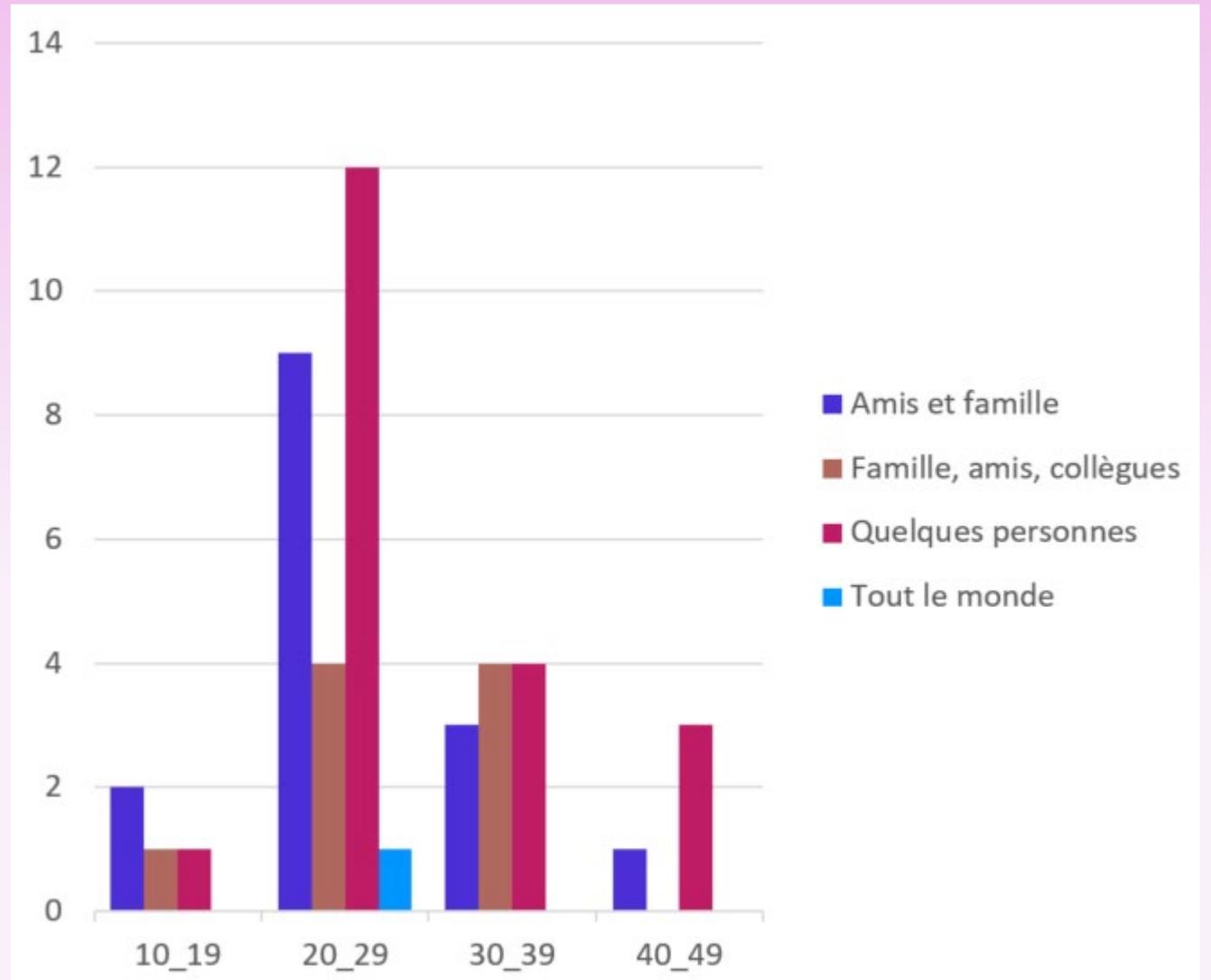
- Questionnaire :
à quelle fréquence
réécoutez-vous un vocal
avant de l'envoyer?
(1 jamais, 5 très souvent)



Accès à l'oral spontané non surveillé?

- Questionnaire :

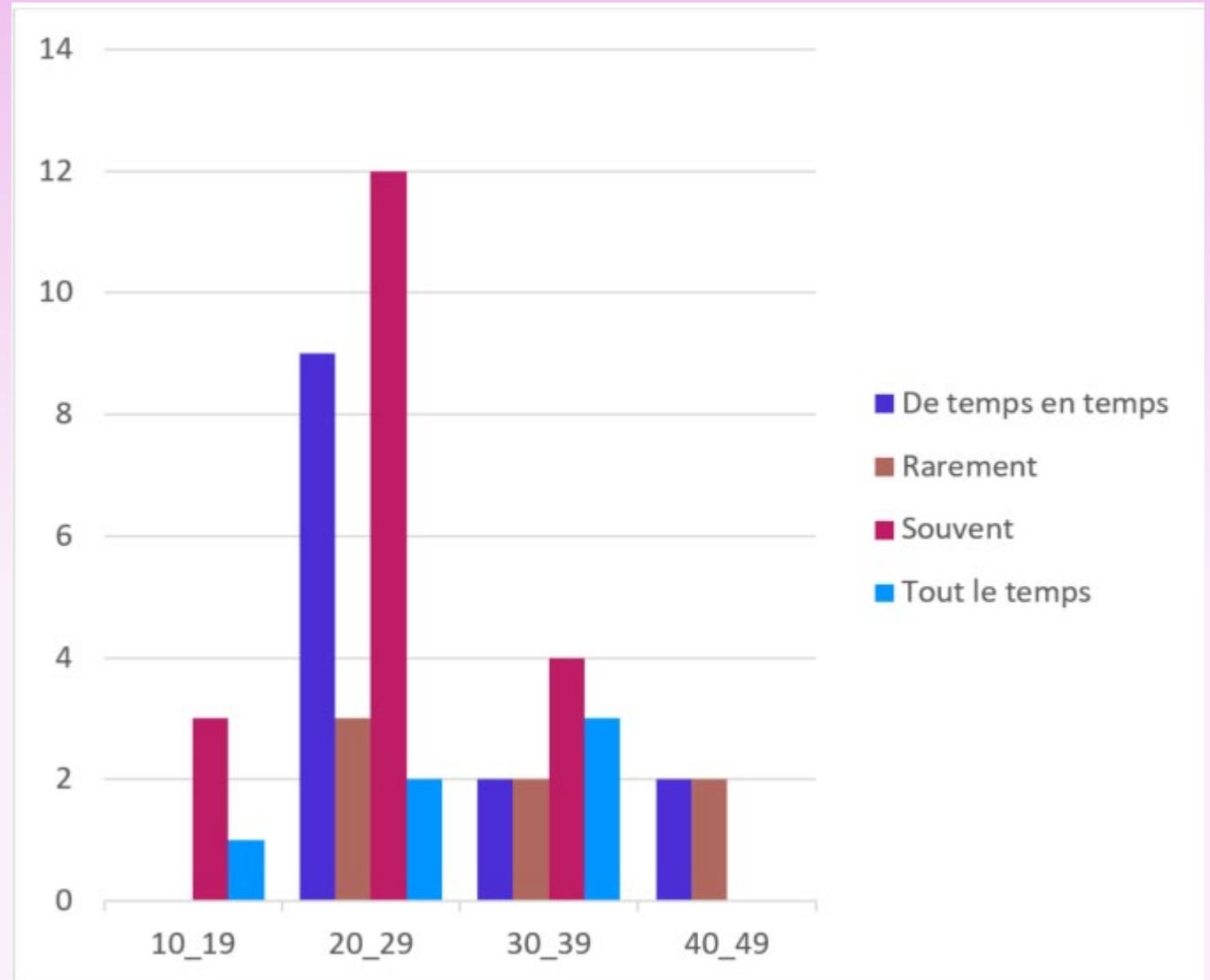
À qui envoyez-vous
des vocaux?



Accès à l'oral spontané non surveillé?

- Questionnaire :

À quelle fréquence
envoyez-vous des vocaux?



Types discursifs variés

- **Discours/Narration** :

vingt heures cinquante-sept je sors à peine du . . . du magasin . . . parce que du coup euh la caisse ne marchait pas donc j'ai dû tout retaper à la main . . . et en fait euh il y avait un souci parce que [. . .] (LesVocaux, 05-06)

- **Interactions** :

bon ben tant pis ça y est tu fais la gueule ou quoi (LesVocaux, 08-02)

- **Discours cité** :

il sort à #name euh . . . ouais euh vous pensez à bien garder les distances de sécurité et euh alors elle dit bah oui vous voyez bien qu'ils sont espacés . . . oui oui non mais je vois je vois hein (LesVocaux, 05-08)

- **MAIS** :

➤ Jamais de chevauchement ni d'interruption (≠ situations dialogiques)

Types discursifs variés : registres

- ah ouais non mais je viens d'écouter mais euh mais euh le le roi des fils de pute c'est pas possible . . . enfin je vais pas commencer à insulter sur le groupe avec tes parents mais . . . mais c'est un gros chien le mec putain à quel mais . . . mais genre je comprends pas (LesVocaux, 188-05)
- . . . alcool cocktail c'est euh . . . c'est euh . . . un substance . . . tout . . . du coup . . . l'alcool . . . est un ingrédient . . . du cocktail . . . ça fonctionne . . . donc . . . le mot ingrédient . . . fonctionne . . . pour euh . . . la euh méronymie substance tout . . . par contre est ce que l'alcool . . . est un membre du cocktail . . . non . . . donc . . . le mot membre . . . ne fonctionne pas pour la rémo- la méronymie . . . substance tout (LesVocaux, 188-06)
- ... coucou chéri je te fais un petit vocal avant de reprendre la route je me suis arrêtée deux minutes ... juste une petite seconde pour te ... h- pour te serrer dans mes bras et te faire plein de bisous ... voilà tu me manques beaucoup j'aimerais bien partager mon café avec toi là tout de suite maintenant ... et puis euh voilà je suis contente que ... que ce soit pas le covid ça va passer plus vite j'espère ... tu me manques tu me manques tu me manques ... je t'embrasse très fort très fort ... à très bientôt j'espère ... bisous ... (Les Vocaux 33-01)



Constructions typiques

• Formes en apostrophe

- je suis trop dég qu'ils aient enlevé How I meet your mother de ... de Netflix meuf j'aurais trop bien reregardé (Les Vocaux, 02-06) 
- finalement les gars je vais pas aller euh chez #name parce que j'ai rendez-vous chez le médecin l'après-midi (Les Vocaux, 105-04)

• Discours sur le discours

- Oui coucou. Bon ben je te fais un petit message. (Les Vocaux, 106-07)
- ... je suis super contente pour toi ... je ... t'envoie ... ce message ... (Les Vocaux, 01-04)
- ... coucou chéri je te fais un petit vocal avant de reprendre la route je me suis arrêtée deux minutes . (Les Vocaux, 33-01)

Constructions typiques

- **Marqueurs de discours propositionnels et/ou marques de structuration de l'énoncé**
 - j'avais j'avais besoin de réconfort tu vois (LesVocaux, 03-03)
 - en fait je t'explique nous on avait garé la caisse là ... on avait fait un tour ... du petit centre-ville du port (LesVocaux, 02-07)
 - ... merci merci euh ben écoute journée de cours bon ça va (Les Vocaux, 05-07)
 - et puis voilà quoi ... donc euh ben bon courage pour rentrer si tu es encore en train de rentrer ... puis à tout bientôt (Les Vocaux, 126-03)

Typiques de l'oral ou de la proximité?

- **Cf. corpus de SMS (Pankhurst et al. 2013)**

- Hello. Ça va bien et toi? Écoute ça aurait été avec plaisir mais demain avec <PRE_3> on va faire du canyoning. Bisous (SMS 180)
- Alors pour ce soir j'viens chez vous vers 19h20-19h30 puis on ira chercher <PRE_3>'

Voilà! Biiisoux (SMS 477)

- Meuf ta francais demain ? (SMS 75)
- Non mais je parlais de nous deux la tu vois ^^ (SMS 508)

Concept: caractériser le continuum (Koch & Oesterreicher 2001)

Oralité : Immédiat communicatif

- **Communication privée**
- **Interlocuteur intime**
- Émotionnalité forte
- Ancrage actionnel et situationnel
- Ancrage référentiel dans la situation
- Coprésence spatio-temporelle
- Coopération communicative intense
- Dialogue
- **Communication spontanée**
- **Liberté thématique**
- ...

Scripturalité: Distance communicative

- Communication publique
- Interlocuteur inconnu
- Émotionnalité faible
- **Détachement actionnel et situationnel**
- **Détachement référentiel de la situation**
- **Séparation spatio-temporelle**
- **Coopération communicative minimale**
- **Monologue**
- Communication préparée
- Fixation thématique
- ...

Autres paramètres

- **Adaptation à l'interlocuteur**

- ... mec ouais ça va ça va juste euh ... #rire ... tu vois je t' avais dit que j' allais aller en France voir une de mes potes que j' ai rencontrée en Erasmus ... ben du coup j' y suis pas allée parce donc en gros hum ... mardi elle m' envoie un message en me disant ouais coucou euh pour te dire que voilà j' ai été en contact avec une personne euh testée positive na na na ... je sais pas si du coup tu as toujours envie de venir ... et puis moi j' étais là en mode ... bah bon voilà une personne ... pour le moment en Belgique quand tu es cas contact mais que tu es vacciné tu dois même pas faire de quarantaine tu fais juste attention aux symptômes et c' est la même chose en France (Les Vocaux, 61-10)



- ... euh coucou ... #name ... alors c' est pour te préparer psychologiquement au fait que demain ... euh si tu viens avec moi à la réunion ... euh en fait moi je dois aller euh un peu avant midi ... au domaine militaire déposer une tente avec #name ... et après ça ... euh je passe au collège parce qu' en fait on fait réunion avec les athénas ... et les mecs ... et donc on va faire un petit débrief euh ... de la réunion être sûr que ... euh tout est ok machin avant le CU donc euh (Les Vocaux, 61-22)



Premières analyses

Les Vocaux

- Corpus provisoire:
- env 60 000 mots, 5h de données
- Proximité communicative - code oral

locuteur	Nb vocaux	durée
Loc_01	6	00:28:13
Loc_02	14	00:03:33
Loc_03	5	00:00:50
Loc_04	5	00:01:05
Loc_05	9	00:06:13
Loc_06	11	00:04:14
Loc_07	9	00:05:19
Loc_33	2	00:01:12
Loc_44	22	00:05:03
Loc_61	75	00:37:13
Loc_70	50	00:16:18
Loc_76	4	00:02:32
Loc_86	18	00:14:39
Loc_93	6	00:07:56
Loc_108	1	00:00:39
Loc_126	4	00:02:46
Loc_188	7	00:08:23
Loc_190	4	00:09:54
Loc_271	1	00:00:31
Loc_105	32	00:16:18
Loc_106	26	01:52:25
Loc_131	19	00:12:30
Loc_193	9	00:04:29
Loc_212	4	00:08:44
Total	343	5:10:59
Dont 2022	284	4:21:32

88milSMS

- Corpus 88milSMS (Panckhurst et al. 2013)
 - plus de 88 000 SMS authentiques en français à Montpellier, en 2011
 - env 2 000 000 mots
 - Proximité communicative – code écrit

Premières analyses

Écoute

(Dostie 1998, 2004, Rodriguez Somolinos 2003, Diwersy et Grutschus 2014, Mazziotta et Glikman en préparation)

Les Vocaux

- Salutation / début
 - Oui coucou ben **écoute** ouais j'ai bien écouté c'est très fort (Les Vocaux, 106-16)
 - coucou #name alors **écoute** je suis navrée pour toi que tu aies reçu tes résultats que tu as pas été prise (Les Vocaux, 131-15)
- Marqueur de pile
 - euh ... mon ... père **écoute** mon père est ... ho- ... a ... subit une première opération (Les Vocaux, 01-02)
- En DD
 - elle m'a dit ben **écoute** je te rappelle (Les Vocaux, 212-03)
- Changement de sujet
 - je lui ai acheté un petit quelque chose donc j'espère que ça va lui plaire [...] c'est vraiment un petit truc un petit truc un petit jouet en bois, enfin voilà que que je trouvais trop chou enfin bref bref bref non je te dis pas plus **écoute** moi je me suis décidée là à la der' de partir en Sardaigne ce week-end (Les Vocaux, 131-17)
- Dont conclusion
 - Tu n'es pas trop tracassée parce que c'est pas trop ton habitude à cette heure-là d'être encore debout enfin je sais pas donc donc voilà ben **écoute** je t'embrasse hein et puis je te dis à bientôt (Les Vocaux, 106-20)
- Poursuite de sujet
 - On sait pas ce qu'il a quoi tu vois je, même pas te dire où est-ce qu'il a mal ou quoi et enfin bref et puis et puis après bah **écoute** ça allait mieux on a eu très très peur (Les Vocaux, 106-23)

88milSMS

- Début / après salutation
 - Oui ça va bien et toi ? Ben **écoute** pas grand chose, là jv au boulot jusqu'à 22h.. sinon pu beaucoup de temps pour sortir et toi ? Tu t es remis avec ton ex alors ? Ou tj juste plan q ? (sms 87219)
 - Ben **écoute** pour ce matin jy peux rien (sms 86384)
- Poursuite de sujet? Après confirmation?
 - Ah ok.. Certes ! Bon ben **écoute** au moins il t'aidera aussi, possiblement lol (sms 84171)
- Changement de sujet
 - Je sais pas l'appareil c'est genre 30/40 euros je suppose (ce qui est deja assez cher vu le prix d'origine), et les films a partir de 25 euros je pense. Pour 10 ou 15 films. C'est donc un budget ! Pour le neuf ben **écoutes** des examens bientôt, un gars qui me plait bien, un voyage a berlin a fevrier, paris et disney en mars, et ce soir vernissage en mode arts plastiques voila voila (sms 84035)

Premières analyses

Avoir été / être allé

(Steuckardt et Dal Bo 2018, Glikman et Patard 2022)

Être pour aller

- Lieu

- ... j' étais avec des potes ... avec #name et tout ... on est allés au à la Péniche Mécanique parce que j' étais jamais allée (Les Vocaux, 06-06)
- ... en fait euh ben j'ai toujours euh voulu aller euh en Ecosse ... et euh du coup ben je jamais été en Ecosse #rire ... parce que bah avant j'avais euh pas trop de salaire et puis covid et tout ça mais je pense que d'ici cet été euh ça sera un peu euh ... plus ou moins euh ... fini on va dire enfin ... pas fini mais euh mm voilà on se comprend ... et puis euh du coup je sais pas si toi tu as déjà été en écosse ou pas (Les Vocaux, 93-01)



- Infinitif

- ... mais je sais pas je connaissais pas c'est une hier je suis allée souper avec une amie (Les Vocaux, 93-04)
- et donc on a été chercher du fumier (Les Vocaux, 106-15)

- Impersonnel (type « ça va »)

- et du coup toi euh comment ça a été ta session ... ma chère (Les Vocaux, 61-12)

Être pour aller

	aller	être	Total général
ça impersonnel		6 (100%)	6
infinitif	12 (60%)	8 (40%)	20
lieu	13 (61,90%)	8 (38,10%)	21
Total général	25	22	47

Premières analyses

Je crois, je pense, je trouve, j'espère

(Blanche-Benveniste 1989, Blanche-Benveniste et Willems 2007, Schneider 2007, Glikman 2009, Ferreira 2022)

Les Vocaux

Étiquettes de lignes	croire	espérer	penser	trouver	Total
final			3		3
incise	3		3	3	9
incise antéposée		1			1
je_queP	22	45	92	31	190
parataxe	4	2	6		12
post	9	4	11		24
Total général	38	52	115	34	239



	croire	espérer	penser	trouver
Recteur	26	47	98	31
Incident	12	5	17	3

Les Vocaux

- QueP et parataxe:
 - et lui il est à 250 ça fait 253 francs pour pour les trois nuits donc voilà **je pense que** ça peut être un bon deal (Les Vocaux, 131-11)
 - ... en fait **je crois qu'**il faudrait que tu fasses quand même pas mal de billets (Les Vocaux, 61-58)
 - parce que du coup dans une pièce il y a quatre lits simples et du coup **je pense on**  dormira là (Les Vocaux, 07-08)
 - il y a aussi une classe **je crois c'** est des cp ou ce1 ... ou les deux (Les Vocaux, 188-02)
 - j' ai trop peur euh d' arriver qu' il y ait une file monstre mais en fait euh pas du tout **je crois** il y avait trois personnes devant moi (Les Vocaux, 61-17)
 - voilà **j'espère** c'est pas trop chiant ... (Les Vocaux, 61-58)
- portée prospective sur l'ensemble de l'énoncé

Les Vocaux

- Possibilité de variation: personne, temps, négation, enchâssement, visée
 - bref euh je savais pas comment l'envoyer euh le renvoyer en lui disant que tu m'avais dit ... que ... euh ouais **je trouvais que** ça faisait un peu des réseaux tant pis tant ... pis ... (Les Vocaux, 70-13)
 - et en fait ... euh je me rends compte que j'ai mal anticipé mes ... cours ... *parce ... que ... je pensais que* j'avais deux ... fois une classe ... et en fait ... euh je m'aperçois maintenant que c'est l'autre ... que j'ai deux fois (Les Vocaux, 70-12)
 - par contre j'ai ... trop le seum *parce que ... je crois qu'*il y a eu un incendie à ... à la station jeanne d'arc j'allais prendre le métro (Les Vocaux, 70-20)
 - j' ai cru que tu étais réveillée et que du coup je pouvais t' appeler ... **je pensais pas que** tu étais déjà ... partie ... dans le bus et tout (Les Vocaux, 86-18)

Les Vocaux

- mais bref c'est vrai que ouais je me sens **je pense pas que** j'ai le covid long mais c'est vrai que je suis encore un peu euh un peu fatiguée (Les Vocaux, 93-06)
- ben ouais justement l'autre euh c'est à l'autre de prouver et **je trouve pas** moi parce qu'il avait fait fau- que Logan il avait fait foirer une épreuve justement parce que (Les Vocaux, 105-25)
- ou sinon **j'ai pensé** aussi ce qu'on pourrait faire ... c'est euh mm tu sais ces pyjamas (Les Vocaux, 93-06)
- ah oui du coup avec #name **on avait euh pensé** ... euh ce qu' on pourrait faire mardi pro du coup c'est que ... euh ... on pourrait euh aller manger une glace en ville (Les Vocaux, 07-06)
- combien **tu penses que** je dois donner pour voilà ce genre de de prestation (Les Vocaux, 106-20)
- bref ... tout ça pour te demander **si tu crois que** ce serait possible que tu fasses un brownie (Les Vocaux, 61-37)



Les Vocaux

- Incises: portée rétrospective, éventuellement restreinte
 - pendant 15 jours j'ai au moins une épreuve par jour **je pense** donc je suis à fond dans les révisions (Les Vocaux, 106-29)
 - soirée tranquillo tous les deux à. Mère et fils va se mater un film **je pense** et et donc voilà (Les Vocaux, 106-12)
- Interne
 - ... je te dirai quand euh quand le truc il arrive chez moi je vais recevoir un mail euh **je pense** demain ou après-demain quoi (Les Vocaux, 93-05)
- Dans une pile
 - Mais pour répondre à ton vocal de ce matin enfin de non pas de oui dans de hier **je crois** oui de hier (Les Vocaux, 106-04)
- Finale
 - je sais pas quelle bonne action j' ai faite mais il faut que je continue sur ce droit chemin **je pense** ... (Les Vocaux, 61-11)
- Initiale?
 - **J'espère** à bientôt (Les Vocaux, 106-26)

Les Vocaux

- Possibilité de variation: négation?
 - et du coup il y a mon CV et ma lettre de motiv' mais bon ça faut juste ... jeter un coup d'oeil mais a priori euh il y a pas de faute ... enfin **je crois pas** ... et euh sinon ... le vrai truc que faut regarder ... c'est euh ... l'orthotypographie en gros (Les Vocaux, 02-08)
 - si tu avais ... enfin tu vois ... si tu avais été prise là en fait ... bien précisément ... en fait tu y aurais passé ton mois d'aout ... et euh ... et tu aurais tu aurais moins travaillé à tes trucs à toi ... ou tu aurais pas travaillé à des trucs à toi enfin **je pense pas** ... et euh ... alors que là ... bon ... c'est un peu le rush (Les Vocaux, 01-01)
- Visée interrogative type disjoint ou emploi absolu?
 - enfin moi je trouve que c'est bien mais quand même non **vous trouvez pas** oh je sais pas (Les Vocaux, 105-17)



88milSMS

- Différentes constructions possibles

- Que_P

- Ben oui **je pense qu** elle ne veut pas rentrer parce qu elle va se sentir seule ! ? ! : '((sms 58473)

- Parataxe:

- Donc **je pense** je vais pas en faire beaucoup (sms 20020)
 - Franchement **je pensais pas** t'allais repondre a 2h : p (sms 34834)

- Incises

- Ca sera fini en mai **je pense**, fin d'année ! (sms 27749)

- Disjointe / réponse à une question

- ouais **je pense** ... (sms 54612)
 - Nan **je pense pas**. (sms 34578)
 - Ah **je pensais pas**. Donc normal que tu sois debout ^ ^ (sms 21239)

Discussion

- Lien entre parenthétiques et autres marqueurs propositionnels?
 - Caractéristiques communes de la position incise
 - Absence de possibilité de modification
 - Différents points d'insertion possible
 - Mais
 - Interprétation sémantique (portée) et valeur pragmatique différentes
 - MD davantage intégrés à l'énoncé (enchâssement)?
 - du coup je me dis **tu vois** comme il y aura différentes monnaies ben ce qu'on peut faire c'est genre (Les Vocaux, 61-62)
 - parce que j'ai pris rendez-vous parce que **tu sais** dans certains magasins avec le truc du Covid faut faire la queue (Les Vocaux, 106-07)

Discussion

- Constructions à l'initiale mais à fonctionnement différent
 - Écoute
 - écoute euh là je suis à gifi euh du côté de compans caffarelli (Les Vocaux, 70-22)
 - Je t'explique
 - pas de souci j' ai pas cours le vendredi moi ... mais du coup je t' explique ... je serai sûrement dans une des petite bibliothèques (Les Vocaux, 61-06)
- Quel paradigme considérer?
 - Selon position ?
 - Selon verbes ?
- Moins de cas de parataxe qu'attendu dans les vocaux : lié à l'échantillon?
- Plus nombreux cas de parataxe dans les SMS: lié proximité plutôt que code?
- Constructions qui conservent des propriétés verbales à l'initiale
 - Différent d'autres verbes en position initial à fonctionnement spécifique (*écoute, je t'explique*)
- Position incise qui perd ses possibilités de modification
 - Mais conserve un régime sous-entendu, à la différence des MD propositionnels

Discussion

- Combinaisons

- Donc voilà mais je vais aller à la recherche de ce couteau et **tu sais écoute** donc je suis dans le rangement c'est ça voilà. (Les Vocaux, 106-18) 
- enfin **tu vois** les les mettre dans ... d- dans une euh ... dans un état de conscience ... qui est en fait euh ... propice à ... à une certaine forme d'introspection ... euh qui est euh ... à laquelle ils arrivent pas ... quand ils sont euh ... quand ils sont dans leur état normal ou peut-être même quand il y a pas ... quand il y a pas de rituel **enfin tu vois je pense que** le ... la dimension euh ... la dimension rituelle de l' hypnose elle ... elle crée un espace ... enfin **c' est ce que je m' imagine** en tout cas (Les Vocaux, 01-04) 

Conclusion

Caractéristiques fixes

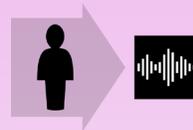
- **Code phonique**
- Monologue
 - Pas de chevauchement ni d'interruption possibles
 - Coopération communicative faible dans le mode de production
 - mais potentiellement forte dans le contenu = questions / réponses, thématisation...
- Séparation spatiale et (potentiellement) temporelle
- Détachement actionnel et situationnel
- Communication de type spontané

Caractéristiques variables

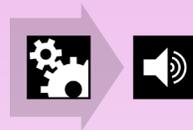
- Variables diatopique, diaphasique, diastratique
- Contenu
- Visée
- Style (au sens large)

Pistes d'exploitation

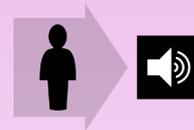
- ▶ Spécificités liées au code vs liées à la distance / proximité communicative
- ▶ Variation diatopique (origine des sujets parlants)
- ▶ Variation diaphasique et interpersonnelle
- ▶ Microdiachronie (structures émergentes, modes)
- ▶ Perspectives : une campagne 2024 incluant SMS et vocaux ?



Vocal production
(misc. platforms)



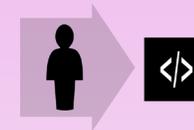
Conversion to WAV
(audacity)



Signal anonymization
(manual)



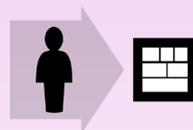
Speech recognition
(Vocapia)



Recognition correction
(manual)



Conversion to Praat TextGrid
(script)



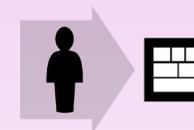
Phonetic analysis preparation
(manual)



Phonetic analysis
(EasyAlign)



Phonetic and word correction
(manual)



Sentence segmentation
(manual)



Conversion to CoNLL
(script)



Automatic annotation
(morpho, syntax, lemmata)
(Trankit)



Morpho, syntax, lemmata
correction
(manual)

Les **V**OCAUX

Final corpus
(script)