

# Les "Vocaux" constituent-ils un genre ?

Julie Glikman et Nicolas Mazziotta

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

Toulouse, le 6 avril 2023



# 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 (Snapchat, Messenger, Whatsapp, etc., mais également messagerie standard)
  - 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 (?)

# Qu'est-ce qu'un « vocal » ?

- **Type de communication peu étudié + besoin**

- ▶ Multiplication des corpus oraux  
(<https://www.ortolang.fr/>)
- ▶ SMS bien étudiés (*sms4science*, *What's up Switzerland*)
- ▶ Peu d'études sur les vocaux (exceptions rares, comme Estade et le Maire 2018)
- ▶ Parole spontanée naturelle toujours **difficile d'accès** (récolte sous forme d'entretiens)

# 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** :
    - ▶ Distribution relative des pauses et des piles : Ph × MS
    - ▶ Typologie des allongements : Ph × Le et Ph × MS
    - ▶ Comportement syntaxique des apostrophes, marqueurs de discours, etc. : Le × MS

# Objectifs du projet « Les Vocaux »

- ▶ Y compris **interférences à « trois voies »**
  - ▶ Réalisation des marqueurs de discours : Le × Ph × MS
  - ▶ « Usure phonique » [SHi], [Sepa], liaisons etc. : Le × Ph × MS

Distribuer le corpus dans des formats utilisables :  
**Praat, CoNLL, Texte, XML pour TXM**

# 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

**Question : Les vocaux constituent-ils un genre ?**

- Caractériser le type de production
- À partir du corpus « Les Vocaux » (en cours de constitution)
  - Présentation de la méthodologie et premier bilan du recueil
  - Premières observations sur le corpus



# Plan de l'exposé

- Introduction
  - Qu'est-ce qu'un vocal?
  - Objectifs
  - Plan de l'exposé
- Méthodologie
  - Le corpus Les Vocaux
- Caractériser le type de production
  - Oral / écrit : code et concept (Koch & Oesterreicher 2001)
  - Caractéristiques des Vocaux
- Conclusion

# Méthodologie

# Deux étapes de recueil des données

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

# Martyr

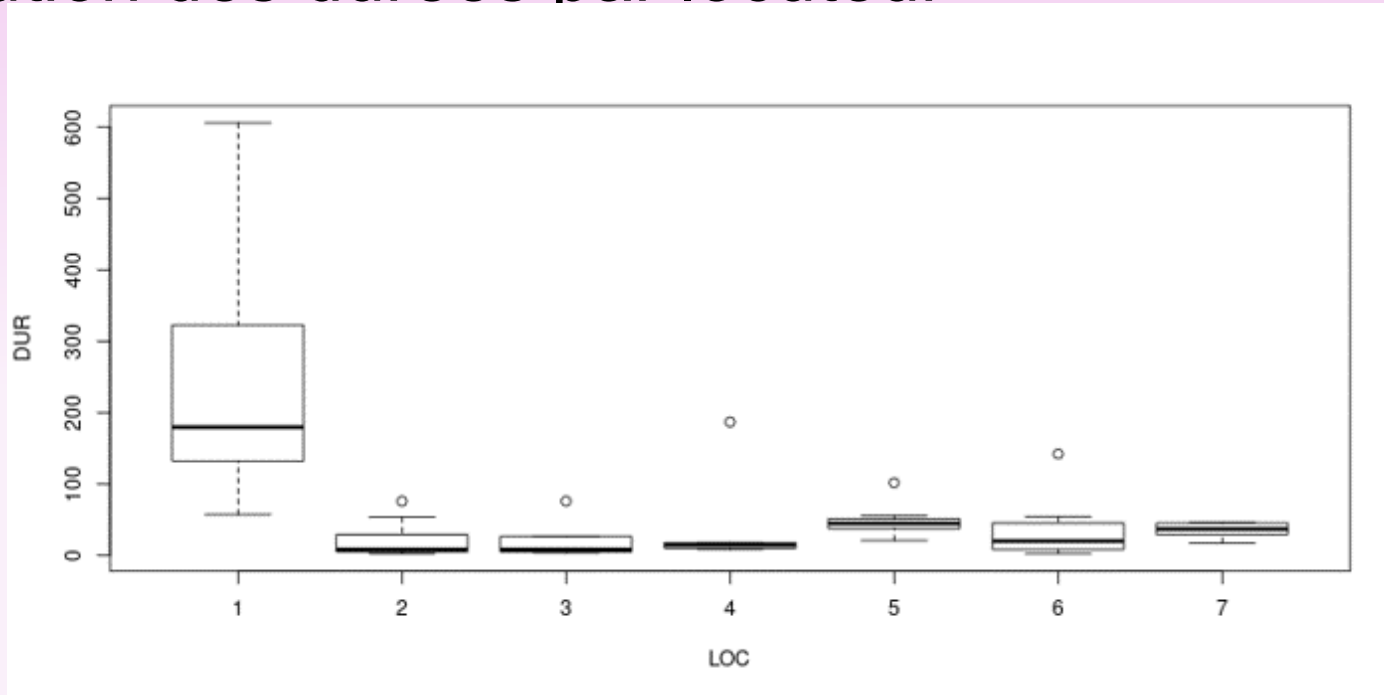
- ▶ 7 participantes volontaires
- ▶ Entretiens individuels sur les pratiques
- ▶ Partage des vocaux (sept.-nov. 2021)
- ▶ Recueil échantillon d'entretiens (matériaux contrastifs)

## • **Données recueillies**

- ▶ Nombre de locutrices : 7
- ▶ Origine : Strasbourg et Paris
- ▶ Nombre de vocaux : 66
- ▶ Durée totale : 1h00m44.73s

# Martyr

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



# Campagne 2022

- ▶ **Crowdsourcing** (cp. *Français de nos régions* et *Donnez votre français à la science*)
  - ▶ **Avantages** : Recueil rapide d'un grand nombre de données réparties dans différents lieux
  - ▶ **Inconvénients** : Nombre 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)
  - ▶ **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 obligatoire au questionnaire

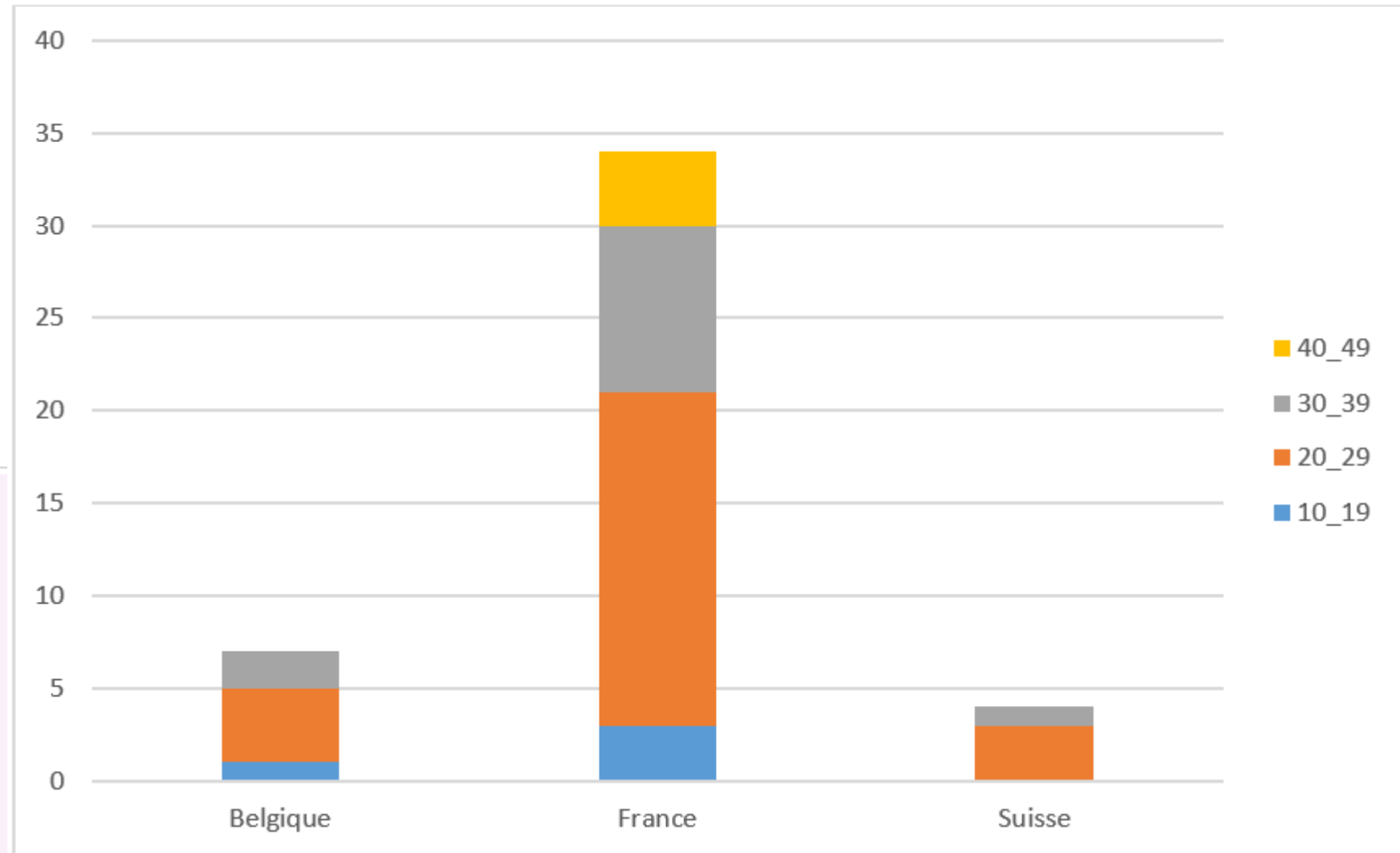
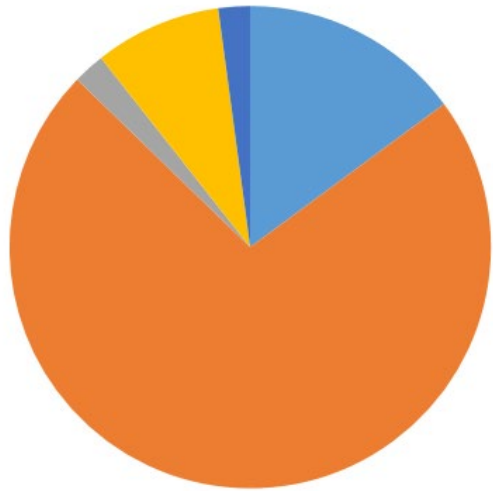
# Campagne 2022

- ▶ **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)
- ▶ **Bilan de la campagne :**
- ▶ Campagne de publicité ayant surtout attiré des **étudiants et collègues** (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

Total



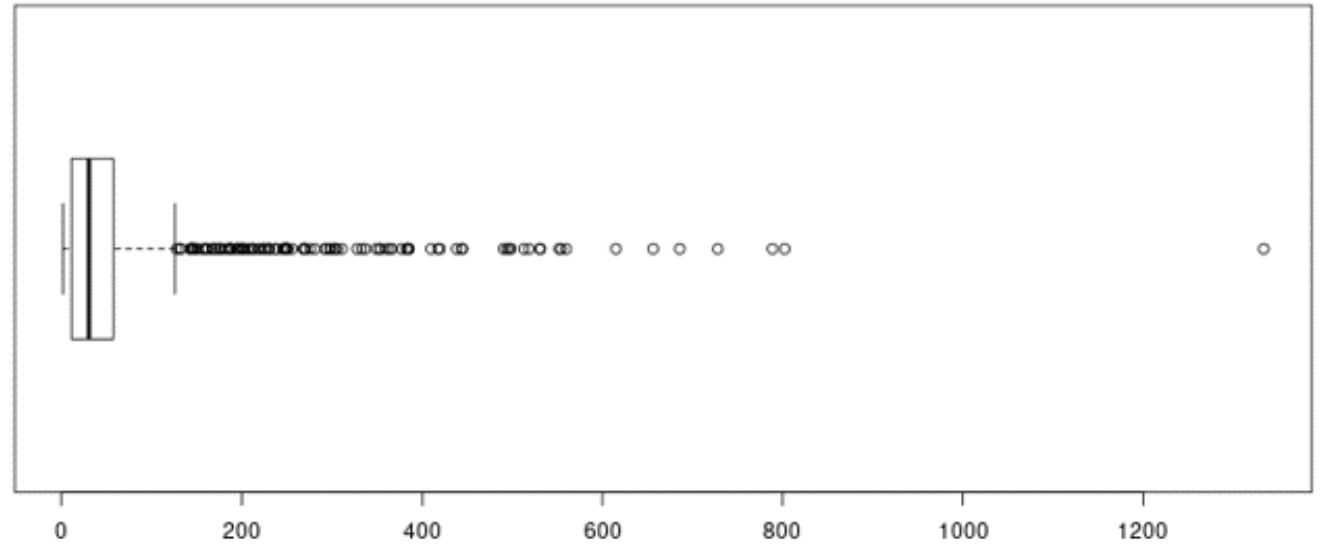
# Campagne 2022

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

# **Caractériser le type de production**

Rappels théoriques

# Code et concept

- Clarifie les confusions, comme par exemple « oral » et « familier », en différenciant « code phonique » et « oralité conceptionnelle »
- **Code:**
  - Matérialité, médium de transmission
  - **Binaire**
- **Concept:**
  - Conception, représentation lié à la situation de communication
  - **Continuum**

# Code : binaire

- Phonie / graphie (ou code phonique / code graphique)
  - Forme matérielle, réalisation concrète
  - Conditions de production
  - Conditions de réception
  - Caractéristiques propres à chaque code
    - Formelles
    - Morphologiques
    - Structures

# Code

## Phonie

- Unité: phonème
- Caractéristique: linéarité temporelle du signal sonore
- Sens: prosodie, accent, débit (vitesse), hauteur, pauses, chuchotements...
- Conditions: pas de retour en arrière (sauf dans enregistrement, mais *inscription*), plutôt proximité
- Prononciation: [cuvã] vs [cuv]
- Morphologie: [li] vs [leli]
- Syntaxe: [typaR]
- Brouillon: plutôt en temps réel

## Graphie

- Unité: graphème
- Caractéristique: spatial
- Sens : mise en page (spatialisation qui peut construire du sens), signes spéciaux (ponctuation, italique, polices ex. effet *ancien* ou **riolo** de certaines polices...),
- Conditions: retour en arrière, plutôt distance
- Orthographe: /couvent/
- Morphologie: /lits/
- Syntaxe: /Tu pars?/
- Brouillon: plutôt non accessible

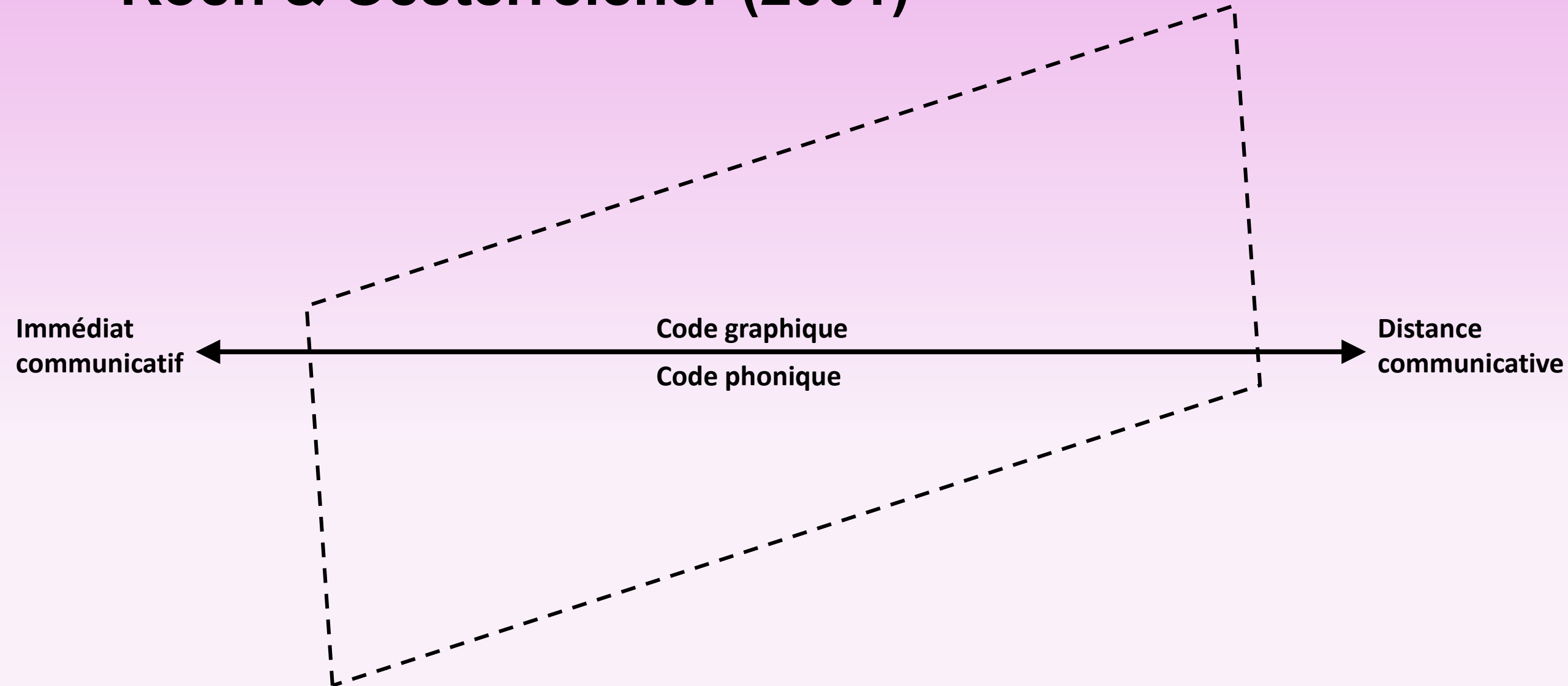
# **Concept: continuum** (Koch & Oesterreicher 2001)

- **Caractériser le comportement communicatif des interlocuteurs par rapport aux déterminants situationnels et contextuels**
  - Situation, contexte de production ou de réception, rapport entre les interlocuteurs, thématique...
  - Se rapproche davantage du type de texte et du registre, mais proposent ici une grille d'analyse
  - Continuum: peut être mixte, plus ou moins proche de l'un ou l'autre pôle
- **Immédiat communicatif versus distance communicative**



- Distinction:
  - Au niveau du moyen de communication:
    - Code graphique / code phonique
  - Au niveau du mode de communication (« conception »)
    - Langage écrit ↔ langage oral
    - Distance ↔ proximité (« immédiat » chez Koch & Oesterreicher)
      - Distance: différée, hors situation, délai entre émission et réception, coupure avec la situation d'émission
      - Proximité: immédiate et en situation, référents communs...

# Koch & Oesterreicher (2001)





# 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
- ...

# Différentes situations

- Écrit élaboré:
  - temps d'élaboration du message, le scripteur peut revenir sur ce qu'il écrit: caractère construit, élaboré
- Oral spontané:
  - le locuteur élabore et émet son message presque au même moment: corrections effectuées à l'oral et qui laissent des traces (reprises, hésitations, ruptures de constructions): effet brouillon
- Influence aussi sur la linguistique:
  - Proximité:
    - Déictiques, gestes et informations non verbales
  - Distance:
    - Anaphoriques, temps et lieu...

# Caractériser les situations

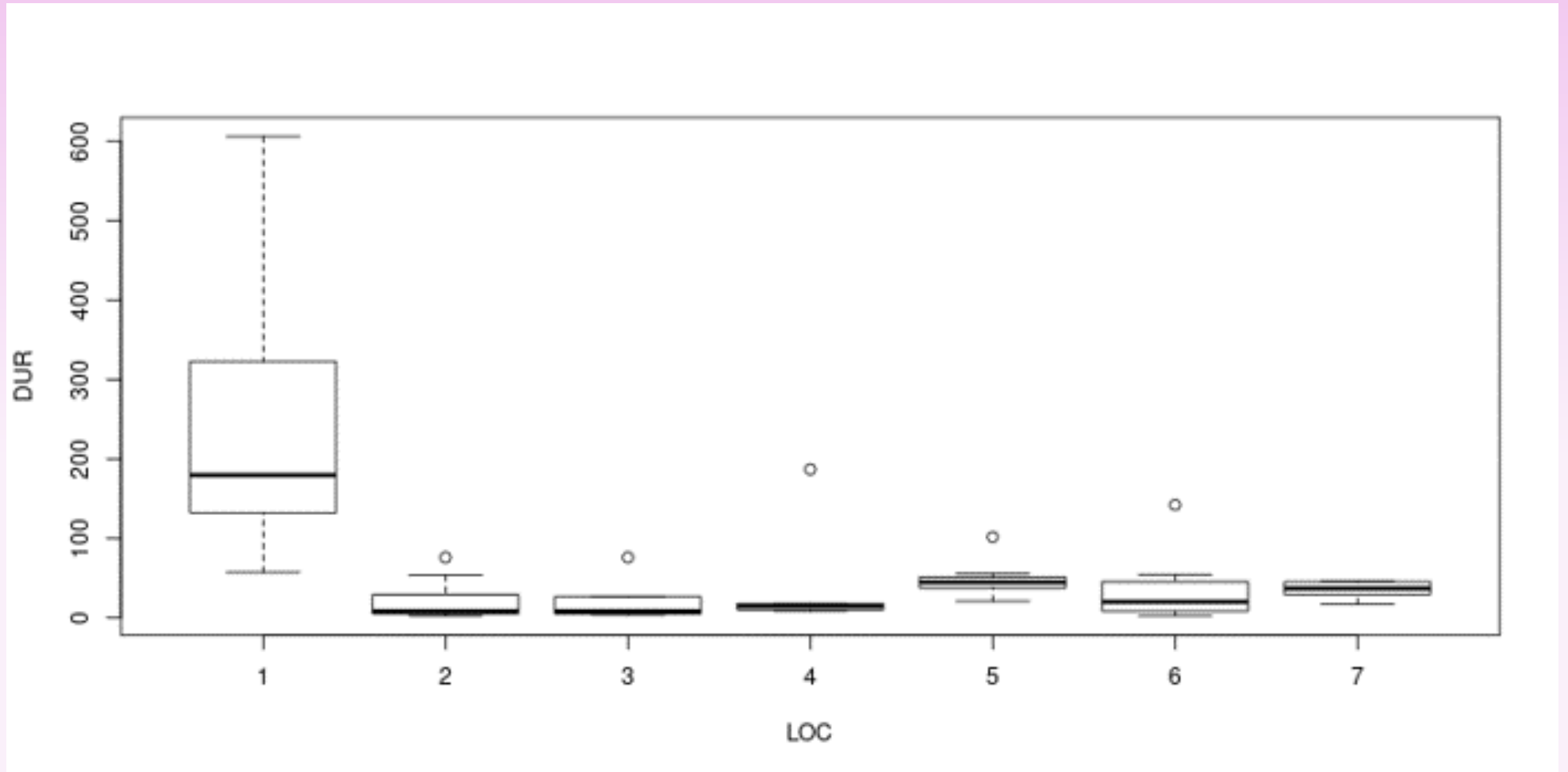
- Situations de communication prototypiques:
  - Écrit littéraire élaboré : distance
  - Conversation orale spontanée : proximité
- Mais de nombreux cas intermédiaires
  - Texte lu
  - Lettre à un ami
  - Sms, chat, forum...
- Penser aussi les nouveaux modes de communication:
  - Sms vocaux, tiktok...

# **Caractériser le type de production**

Premières observations du corpus Les Vocaux

# Caractéristiques formelles : Durée

Martyr:



# Caractéristiques formelles : Durée

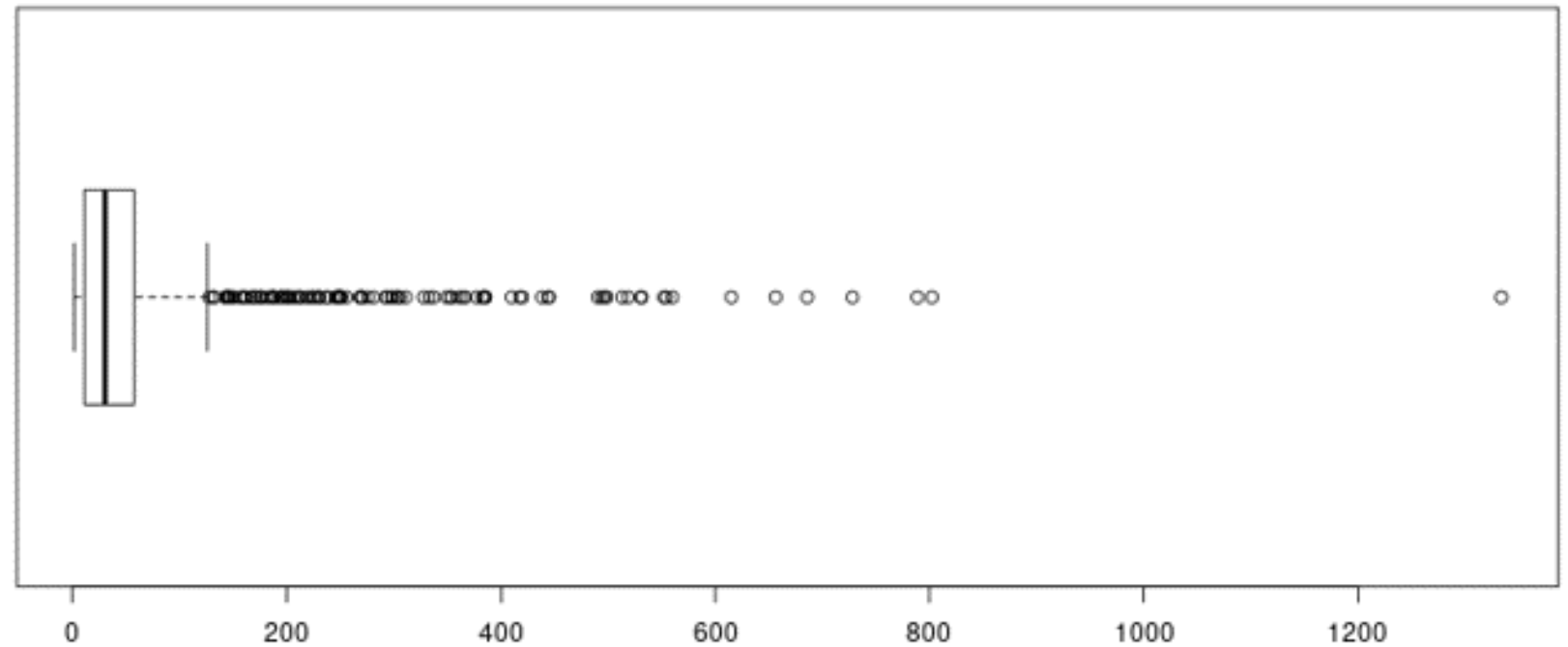
Id	Âge	Lieu enfance	Vocaux retenus	Durée totale	Durée moyenne
1	41	Paris	2	370,96	185,46 (24,44)
2	20	Haut-Rhin	14	198,07	14,15 (16,98)
3	23	Haut-Rhin	5	50,9	10,18 (9,18)
4	22	Belgique flamande / Alsace	5	60,32	12,06 (3,67)
5	21	Moselle	9	343,07	38,12 (6,13)
6	26	Haut-Rhin / Normandie	11	247,12	22,46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



# Caractéristiques formelles : Durée

Campagne 2022:



➤ Très variable

# 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 ... (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  
... (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 ...  
(06\_01)



... ça marche je suis en train de partir du parking de l'uni justement ... alors je ... je t'appelle  
quand je suis parké puis comme ça tu me dis où tu es ... allez ... à toute ... (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)

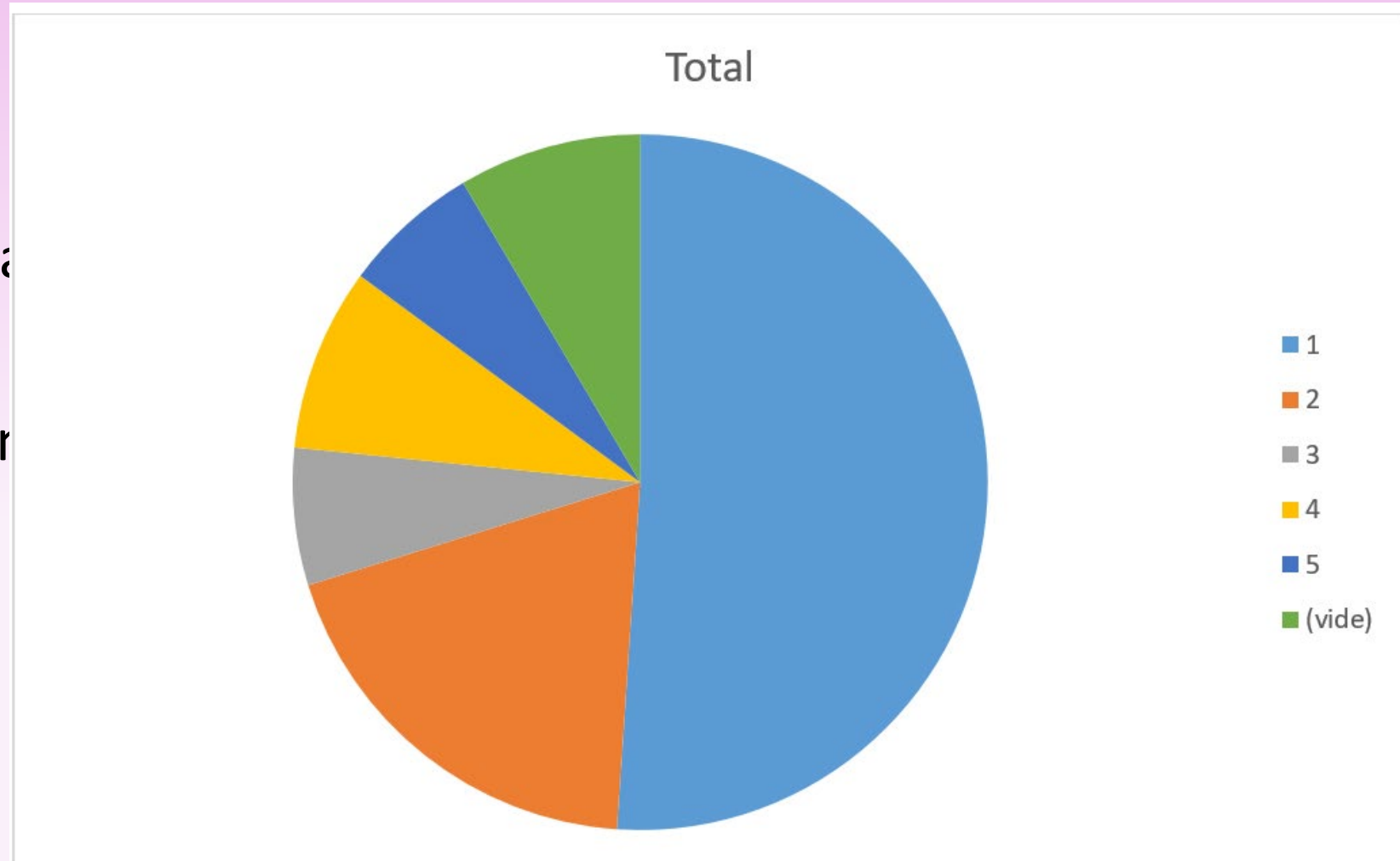
## ➤ Piétinements, amorce

... 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  
(01\_03)



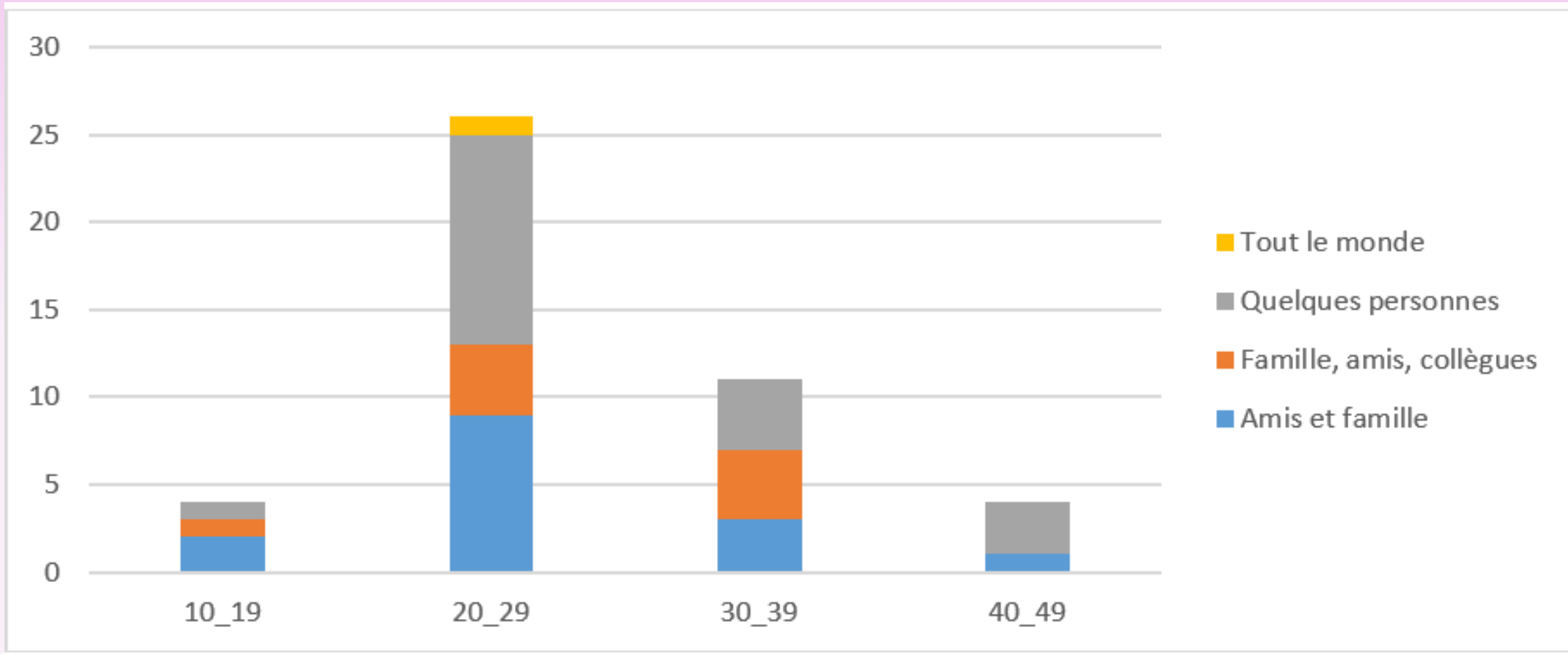
# Accès à l'oral spontané non surveillé?

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



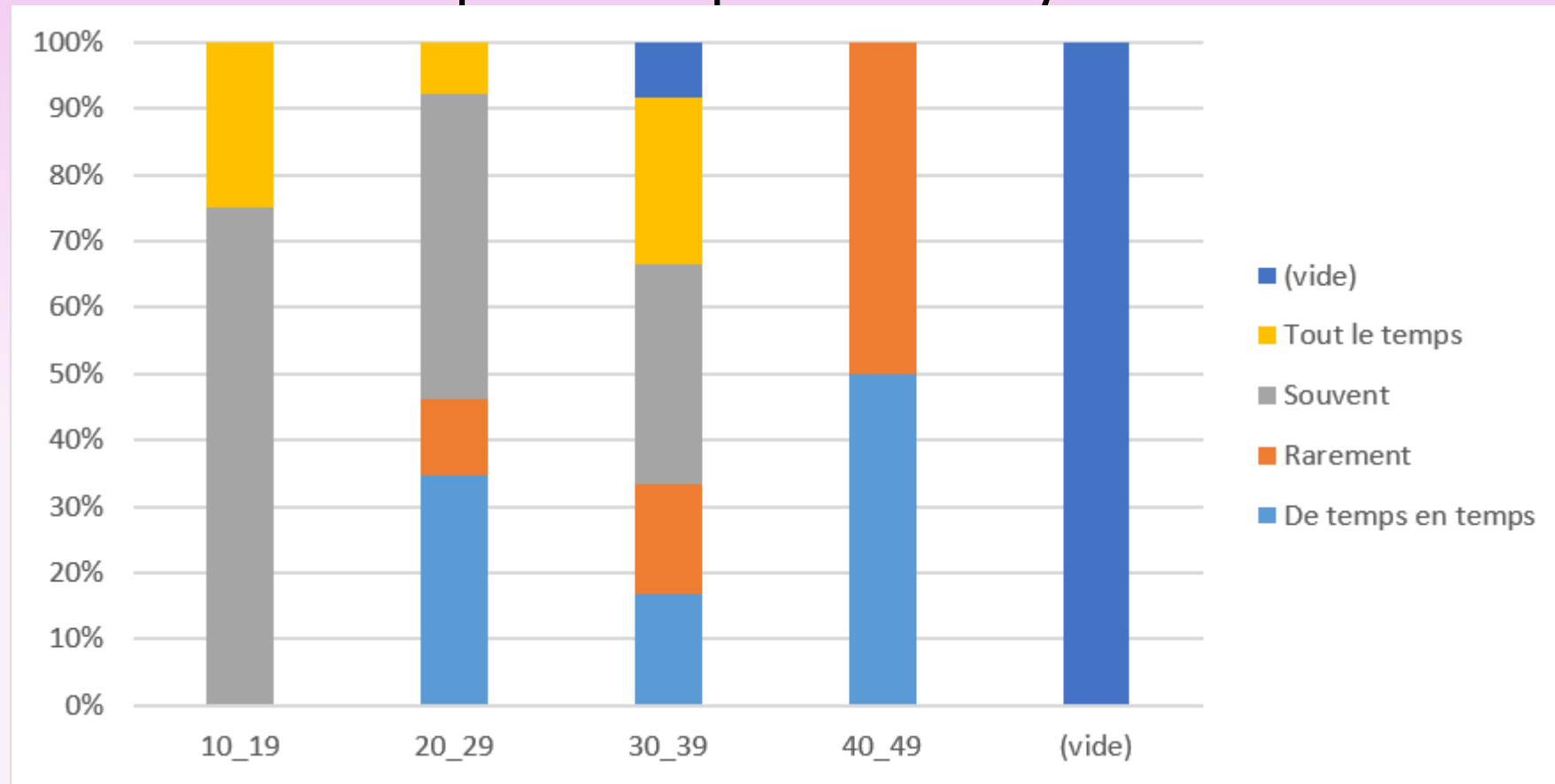
# Accès à l'oral spontané non surveillé?

- 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 ... (33\_01)



# Constructions typiques

## • Formes en apostrophe

- ▶ ... meuf ... #rire ... je viens de relire ... que vendredi soir j'avais euh j'avais appelé plein de fois (LesVocaux 06-04)
- ▶ je suis trop dég qu'ils aient enlevé How I meet your mother de ... de Netflix meuf (LesVocaux, 02-06) 
- ▶ juste ... meuf ... histoire hyper drôle genre j'étais soule au lieu de faire un virement de cinq euros j'ai fait un virement de cinq cent euros (LesVocaux, 02-12)
- ▶ finalement les gars je vais pas aller euh chez #name parce que j'ai rendez-vous chez le médecin l'après-midi (105\_04)

# Constructions typiques

## • Formes interrogatives diverses

- ▶ alors je dois faire comment ? (LesVocaux, 02-07)
- ▶ ouais comment on peut s'organiser du coup ? (LesVocaux, 07-02)
- ▶ tu fais la gueule ou quoi ? (LesVocaux, 06-08)
- ▶ est-ce que nous on peut prendre un train jusqu'à euh ... jusqu'à #city jusqu'à #city ? (LesVocaux, 02-05)
- ▶ qu'est-ce qui t'a déclenché cette pensée ? (LesVocaux, 06-03)

# 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 (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 (126\_03)

# Constructions typiques

- **Discours sur le discours**

- Oui coucou. Bon ben je te fais un petit message. (106\_07)
- ... je suis super contente pour toi ... je ... t'envoie ... ce message ... (01\_04)
- ... coucou chéri je te fais un petit vocal avant de reprendre la route je me suis arrêtée deux minutes . (33\_01)

# 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 (180)
- Alors pour ce soir j'viens chez vous vers 19h20-19h30 puis on ira chercher <PRE\_3>!

Voilà! Biiisoux (477)

- Meuf ta francais demain ? (75)
- Non mais je parlais de nous deux la tu vois ^^ (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 (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 (61\_22)





**Conclusion**

# Les vocaux sont-ils un *genre*?

- **Définition de *genre* ?**

- Tout fait de langue est en même temps fait dialectal, fait de niveau et fait de style (du moins, en tant que fait appartenant à la langue commune et au niveau et au style « neutres ») et devrait être « indexé » comme tel dans le lexique et dans une grammaire de la variété de la langue historique. (Coseriu 1998 : 14)
- tout énoncé comporte en même temps une dimension diatopique, diastratique et diaphasique, s'inscrit dans le continuum entre immédiat et distance, représente une modalité médiale et appartient à un genre textuel donné. Autrement dit : tout énoncé fait appel à chacune des dimensions diasystématiques qui sont ainsi à tout moment coprésentes. (Glessgen et Schøsler 2018 : 17)
- Les sujets s'expriment d'une manière ou d'une autre parce que des paramètres extérieurs influent sur les situations de communication – paramètres classiquement divisés en diatopiques, diastratiques et diaphasiques, éventuellement diamésiques, et aussi parce qu'ils répondent à des stratégies pragmatiques particulières. (Siouffi 2020 : 68)

# Les vocaux sont-ils un genre?

- Toute production s'inscrit dans ces différents paramètres
- L'inscription dans un *genre* peut induire des contraintes linguistiques (cf. effets de genre dans l'étude de la variation)
- Les vocaux :
  - Genre au sens mode de production: contraintes liées aux modes de production, diffusion, réception
  - Mais pas au sens contenu, visée, registre, style (langue de spécialité...), etc.
  - Plutôt mode discursif?
- Distinguer les caractéristiques formelles (fixes) et les caractéristiques fonctionnelles (variables)

# Caractéristiques fixes

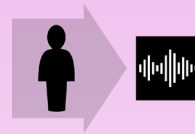
- **Code phonique**
- Monologue
  - Pas de chevauchement ni d'interruption possibles
  - Coopération communicative faible dans le mode de production
    - mais potentiellement forte dans le contenu = questions / réponses, thématisation...
- Séparation spatiale et (potentiellement) temporelle
- Détachement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

# Caractéristiques variables

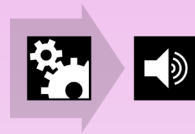
- Variables diatopique, diaphasique, diastratique
- Contenu
- Visée
- Style (au sens large)

# Pistes d'exploitation

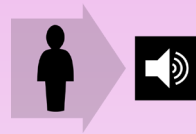
- ▶ Variation diatopique (origine des sujets parlants)
- ▶ Variation diaphasique
- ▶ Microdiachronie (structures émergentes, modes)



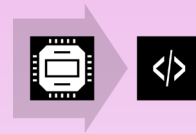
Vocal production  
(misc. platforms)



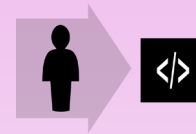
Conversion to WAV  
(audacity)



Signal anonymization  
(manual)



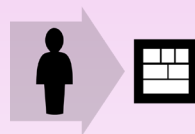
Speech recognition  
(Vocapia)



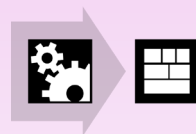
Recognition correction  
(manual)



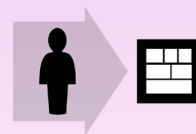
Conversion to Praat TextGrid  
(script)



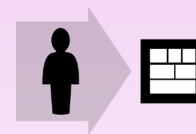
Phonetic analysis preparation  
(manual)



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)