

# Une nouvelle voie d'accès au français populaire : Les Vocaux

Julie Glikman<sup>\*,\*\*\*</sup>, Camille Fauth<sup>\*</sup>,  
Nicolas Mazziotta<sup>\*\*</sup>, Christophe Benzitoun<sup>\*\*\*</sup>

<sup>\*</sup>Université de Strasbourg, LiLPa (France)

<sup>\*\*</sup>Université de Liège, Traverses (Belgique)

<sup>\*\*\*</sup>Université de Lorraine, ATILF (France)

13<sup>e</sup> congrès des francoromanistes – Vienne, 21-24 sept. 2022

Les  
**V****CAUX**

# Introduction

**Qu'est-ce qu'un « vocal » ?**

# Introduction

## Qu'est-ce qu'un « vocal » ?

- ▶ Concept de « SMS vocal », « message vocal », « note vocale » ou « vocal » :

# Introduction

## 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)

# Introduction

## 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

# Introduction

## 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
  - ▶ Se substituent à l'envoi de SMS et aux appels

# Introduction

## 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
  - ▶ Se substituent à l'envoi de SMS et aux appels
- ▶ Exemple : [LesVocaux, 02-01](#)

# Introduction

## 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
  - ▶ Se substituent à l'envoi de SMS et aux appels
- ▶ Exemple : [LesVocaux, 02-01](#)
  - ▶ euh la vidéo ... où v- euh ... on est en accéléré elle est hyper drôle ... (LesVocaux, 02-01)



# Introduction

## Caractéristiques communicatives

# Introduction

## Caractéristiques communicatives

- ▶ langue **spontanée** dans un contexte de production non dirigée, aux échanges de type privés, non surveillés,

# Introduction

## Caractéristiques communicatives

- ▶ langue **spontanée** dans un contexte de production non dirigée, aux échanges de type privés, non surveillés,
- ▶ En exploitant Koch et Oesterreicher (2001) :

# Introduction

## Caractéristiques communicatives

- ▶ langue **spontanée** dans un contexte de production non dirigée, aux échanges de type privés, non surveillés,
- ▶ En exploitant Koch et Oesterreicher (2001) :
  - ▶ Code oral

# Introduction

## Caractéristiques communicatives

- ▶ langue **spontanée** dans un contexte de production non dirigée, aux échanges de type privés, non surveillés,
- ▶ En exploitant Koch et Oesterreicher (2001) :
  - ▶ Code oral
  - ▶ **Proximité** communicative ( $\neq$  publications Tiktok/instagram) : communication privée, interlocuteur connu (pas nécessairement intime), globalement spontanée, émotionnalité potentiellement élevée

# Introduction

## Caractéristiques communicatives

- ▶ langue **spontanée** dans un contexte de production non dirigée, aux échanges de type privés, non surveillés,
- ▶ En exploitant Koch et Oesterreicher (2001) :
  - ▶ Code oral
  - ▶ **Proximité** communicative ( $\neq$  publications Tiktok/instagram) : communication privée, interlocuteur connu (pas nécessairement intime), globalement spontanée, émotionnalité potentiellement élevée
  - ▶ **Distance** communicative ( $\neq$  communication téléphonique) : monologues, coopération communicative faible (pas de chevauchement, ni de coproduction), généralement pas de coprésence spacio-temporelle

# Introduction

## Caractéristiques communicatives

- ▶ langue **spontanée** dans un contexte de production non dirigée, aux échanges de type privés, non surveillés,
- ▶ En exploitant Koch et Oesterreicher (2001) :
  - ▶ Code oral
  - ▶ **Proximité** communicative ( $\neq$  publications Tiktok/instagram) : communication privée, interlocuteur connu (pas nécessairement intime), globalement spontanée, émotionnalité potentiellement élevée
  - ▶ **Distance** communicative ( $\neq$  communication téléphonique) : monologues, coopération communicative faible (pas de chevauchement, ni de coproduction), généralement pas de coprésence spacio-temporelle

→ Catégorie **spécifique**

# Introduction

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

- ▶ Multiplication des corpus oraux  
(<https://www.ortolang.fr/>)



# Introduction

## Type de communication peu étudié + besoin

- ▶ Multiplication des corpus oraux  
(<https://www.ortolang.fr/>)
- ▶ SMS bien étudiés (*sms4science*, *What's up Switzerland*)

# Introduction

## 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)

# Introduction

## 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)

# Introduction

## Objectif de l'exposé

- ▶ Présentation des méthodes de recueil, et de construction du corpus
- ▶ Premières observations (tests sur un corpus « martyr »)

# Plan

Introduction

Recueil des données

Premières observations

Chaîne de traitement

Formats et Outils

Conclusion

# Recueil des données

Introduction

**Recueil des données**

Premières observations

Chaîne de traitement

Formats et Outils

Conclusion

# Recueil des données

## Nature technique des énonciations

- ▶ En tant qu'actes d'énonciation, ce sont des messages **enregistrés**
- ▶ La récolte peut commencer par un simple transfert

# Recueil des données

## Nature technique des énonciations

- ▶ En tant qu'actes d'énonciation, ce sont des messages **enregistrés**
- ▶ La récolte peut commencer par un simple transfert

## Cadre de récolte



# Recueil des données

## Nature technique des énonciations

- ▶ En tant qu'actes d'énonciation, ce sont des messages **enregistrés**
- ▶ La récolte peut commencer par un simple transfert

## Cadre de récolte

- ▶ **Crowdsourcing** (cp. *Français de nos régions* et *Donnez votre français à la science* (Glikman et al.)

# Recueil des données

## Nature technique des énonciations

- ▶ En tant qu'actes d'énonciation, ce sont des messages **enregistrés**
- ▶ La récolte peut commencer par un simple transfert

## Cadre de récolte

- ▶ **Crowdsourcing** (cp. *Français de nos régions* et *Donnez votre français à la science* (Glikman et al.)
- ▶ Nécessité de respecter la RGPD (voix = donnée sensible)

# Recueil des données

## Nature technique des énonciations

- ▶ En tant qu'actes d'énonciation, ce sont des messages **enregistrés**
- ▶ La récolte peut commencer par un simple transfert

## Cadre de récolte

- ▶ **Crowdsourcing** (cp. *Français de nos régions* et *Donnez votre français à la science* (Glikman et al.)
- ▶ Nécessité de respecter la RGPD (voix = donnée sensible)
- ▶ Campagne de recueil limitée dans le temps (avril – septembre 2022)

# Recueil des données

## Nature technique des énonciations

- ▶ En tant qu'actes d'énonciation, ce sont des messages **enregistrés**
- ▶ La récolte peut commencer par un simple transfert

## Cadre de récolte

- ▶ **Crowdsourcing** (cp. *Français de nos régions* et *Donnez votre français à la science* (Glikman et al.)
- ▶ Nécessité de respecter la RGPD (voix = donnée sensible)
- ▶ Campagne de recueil limitée dans le temps (avril – septembre 2022)
- ▶ Campagne de publicité ayant surtout attiré des **étudiants et collègues** (Facebook, Instagram, annonces aux cours)

# Recueil des données

## 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)

# Premières observations

Introduction

Recueil des données

**Premières observations**

Chaîne de traitement

Formats et Outils

Conclusion

# Premières observations

## **1<sup>re</sup> approche « intuitive » : types discursifs**

Des types discursifs variés :

# Premières observations

## 1<sup>re</sup> approche « intuitive » : types discursifs

Des 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)



# Premières observations

## 1<sup>re</sup> approche « intuitive » : types discursifs

Des 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)

# Premières observations

## 1<sup>re</sup> approche « intuitive » : types discursifs

Des 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)

# Premières observations

## **1<sup>re</sup> approche « intuitive » : phénomènes saillants**

Nous pouvons observer des phénomènes :

# Premières observations

## **1<sup>re</sup> approche « intuitive » : phénomènes saillants**

Nous pouvons observer des phénomènes :

- ▶ Particulièrement fréquents dans le corpus

# Premières observations

## **1<sup>re</sup> approche « intuitive » : phénomènes saillants**

Nous pouvons observer des phénomènes :

- ▶ Particulièrement fréquents dans le corpus
- ▶ Typiques de l'oral

# Premières observations

## **1<sup>re</sup> approche « intuitive » : phénomènes saillants**

Nous pouvons observer des phénomènes :

- ▶ Particulièrement fréquents dans le corpus
- ▶ Typiques de l'oral
- ▶ Peu normés (non conformes aux descriptions normées)

# Premières observations

## **1<sup>re</sup> approche « intuitive » : phénomènes saillants**

Nous pouvons observer des phénomènes :

- ▶ Particulièrement fréquents dans le corpus
- ▶ Typiques de l'oral
- ▶ Peu normés (non conformes aux descriptions normées)
- ▶ Peu étudiés (ou à explorer davantage)

# Premières observations

## 1<sup>re</sup> approche « intuitive » : phénomènes saillants

Nous pouvons observer des phénomènes :

- ▶ Particulièrement fréquents dans le corpus
- ▶ Typiques de l'oral
- ▶ Peu normés (non conformes aux descriptions normées)
- ▶ Peu étudiés (ou à explorer davantage)

⇒ Autant de **constructions**, **emplois de mots** et de **phénomènes prosodiques** que nous allons évoquer



# Premières observations

## Structures interrogatives

Structures **fréquentes** dans le corpus (dont interrogatives *in situ*, en *est-ce que*,... ) – Coveney 2020

# Premières observations

## Structures interrogatives

Structures **fréquentes** dans le corpus (dont interrogatives *in situ*, en *est-ce que*,...) – Coveney 2020

- ▶ 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)

# Premières observations

## **Pseudo-clivées**

Également **typiques** de l'oral – Roubaud 2000

# Premières observations

## Pseudo-clivées

Également **typiques** de l'oral – Roubaud 2000

- ▶ euh ce qu'on pourrait faire mardi pro du coup ... c'est que ... euh ... on pourrait euh aller manger une glace en ville (LesVocaux, 07-06)
- ▶ comme si euh comme si le seul ... le seul truc ... qui était ennuyeux dans cette affaire ... euh c'est qu'il ait énervé sa mère (LesVocaux, 01-03)

## Premières observations

### ***meuf* en apostrophe**

À intégrer à la problématique des **termes d'adresse** (y compris question de la diachronie courte) – Bruno 2013, Lagorgette 2003, Lagorgette 2006

# Premières observations

## ***meuf* en apostrophe**

À intégrer à la problématique des **termes d'adresse** (y compris question de la diachronie courte) – Bruno 2013, Lagorgette 2003, Lagorgette 2006

- ▶ ... 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 euro j'ai fait un virement de cinq cent euro (LesVocaux, 02-12)

# Premières observations

## ***meuf* en apostrophe**

À intégrer à la problématique des **termes d'adresse** (y compris question de la diachronie courte) – Bruno 2013, Lagorgette 2003, Lagorgette 2006

- ▶ ... 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 euro j'ai fait un virement de cinq cent euro (LesVocaux, 02-12)

*meuf* pas évalué dans la thèse de Wang (2016) sur le paradigme des appellatifs comme *Mademoiselle*

# Premières observations

## **(In)subordination, (in)dépendance**

*parce que P* non causales microsyntaxiquement indépendantes –  
Debaisieux 1994

- ▶ mais bref je voulais savoir si ... hum ... tu trouvais ça ... chouette parce qu'en fait j'ai de la peinture à doigts ... qu'on devait utiliser pour l'anniversaire à #name (LesVocaux, 05-04)
- ▶ mais en tout cas mais vraiment montez parce que c'est vraiment trop beau (LesVocaux, 02-07)



# Premières observations

## (In)subordination, (in)dépendance

*parce que* P non causales microsyntaxiquement indépendantes –  
Debaisieux 1994

- ▶ mais bref je voulais savoir si ... hum ... tu trouvais ça ... chouette  
parce qu'en fait j'ai de la peinture à doigts ... qu'on devait utiliser  
pour l'anniversaire à #name (LesVocaux, 05-04)
- ▶ mais en tout cas mais vraiment montez parce que c'est vraiment  
trop beau (LesVocaux, 02-07)

Autres formes dont il faut **évaluer le paradigme** : *même que, sauf que, alors que* P non temporelle, etc.

## Premières observations

### **Marqueurs discursifs propositionnels**

Question des « recteurs faibles » et de la parataxe – Andersen 2007, Bolly 2010, Galatanu et al. 2014

# Premières observations

## Marqueurs discursifs propositionnels

Question des « recteurs faibles » et de la parataxe – Andersen 2007, Bolly 2010, Galatanu et al. 2014

- ▶ j'avais j'avais besoin de réconfort tu vois (LesVocaux, 03-03)
- ▶ c'était chelou tu vois (LesVocaux, 06-06)
- ▶ euh je t'explique ... euh ... pour ... donc quand on va à Lyon ... pour revenir genre euh ... les trains ils sont ... enfin ils sont chers genre ça fait quand même genre ... euh cinquante balles (LesVocaux, 02-05)
- ▶ 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)

# Premières observations

## Marqueurs discursifs propositionnels

Question des « recteurs faibles » et de la parataxe – Andersen 2007, Bolly 2010, Galatanu et al. 2014

- ▶ j'avais j'avais besoin de réconfort tu vois (LesVocaux, 03-03)
- ▶ c'était chelou tu vois (LesVocaux, 06-06)
- ▶ euh je t'explique ... euh ... pour ... donc quand on va à Lyon ... pour revenir genre euh ... les trains ils sont ... enfin ils sont chers genre ça fait quand même genre ... euh cinquante balles (LesVocaux, 02-05)
- ▶ 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)

Pas de travaux sur la forme *je t'explique* (?)

# Premières observations

## Syntaxe : reformulations en « piles »

Blanche-Benveniste C. et al. 1979, Kahane et Pietrandrea 2012

- ▶ et je trouve que c'est ... hyper bizarre en fait ... euh le mécanisme ... de la désobéissance ou de la ... ouais du ouais du de la désobéissance au sens large disons de pas faire ce qu'on te demande ... chez les enfants et surtout chez les enfants euh ... euh ... un peu tu vois qui sont plus pas des tous petits ... qui vraiment ont beaucoup de mal à résister à ... à ... à l'attrait d'une chose et qui comprennent pas forcément le sens des interdits ... ou ... ou des ordres en fait ... qu'on leur qu'on leur donne (LesVocaux, 01-03)

# Premières observations

## Syntaxe : reformulations en « piles »

Blanche-Benveniste C. et al. 1979, Kahane et Pietrandrea 2012

- ▶ et je trouve que c'est ... hyper bizarre en fait ... euh le mécanisme ... de la désobéissance ou de la ... ouais du ouais du de la désobéissance au sens large disons de pas faire ce qu'on te demande ... chez les enfants et surtout chez les enfants euh ... euh ... un peu tu vois qui sont plus pas des tous petits ... qui vraiment ont beaucoup de mal à résister à ... à ... à l'attrait d'une chose et qui comprennent pas forcément le sens des interdits ... ou ... ou des ordres en fait ... qu'on leur qu'on leur donne (LesVocaux, 01-03)

Nombreuses **piles imbriquées**, nombreux marqueurs de piles, nombreuses marques d'hésitation (pauses, *euh*)

# Premières observations

et je trouve que c' est hyper bizarre en fait euh le mécanisme de la désobéissance  
(ouais) ou de la  
(ouais) du  
du  
de la désobéissance au sens large  
disons de pas faire ce qu'on te demande chez les enfants  
et surtout chez les enfants euh euh un peu tu vois qui sont plus pas des tous petits  
qui vraiment ont beaucoup de mal à résister à  
à l' attrait d' une chose  
et qui comprennent pas forcément le sens des interdits  
ou  
ou des ordres en fait qu' on leur  
qu' on leur donne

## Analyse en grille

## Premières observations

### **Phonétique : question des pauses et des hésitations**

- ▶ Nombreuses pauses (notées <...>) et hésitations



# Premières observations

## Phonétique : question des pauses et des hésitations

- ▶ Nombreuses pauses (notées <...>) et hésitations

⇒ Cela justifie qu'on s'intéresse :

- ▶ Aux **pauses** : interne/externe au syntagme
- ▶ Aux **allongements** : notamment leur positionnement par rapport aux piles notamment
- ▶ Aux **fillers** de type *eah*

# Premières observations

## Phonétique : question des pauses et des hésitations

- ▶ Nombreuses pauses (notées <...>) et hésitations

⇒ Cela justifie qu'on s'intéresse :

- ▶ Aux **pauses** : interne/externe au syntagme
- ▶ Aux **allongements** : notamment leur positionnement par rapport aux piles notamment
- ▶ Aux **fillers** de type *eah*

pour rendre compte des **stratégies de construction** de la chaîne phonique de l'énoncé

# Premières observations

	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	

Données phonétiques (Glikman et Fauth 2022)

# Chaîne de traitement

Introduction

Recueil des données

Premières observations

**Chaîne de traitement**

Formats et Outils

Conclusion

# Chaîne de traitement

## Un corpus pour étudier ces phénomènes

► **Annotations** nécessaires :

Ph **phonétique** (phones, syllabation, pauses, alignement au signal)

Le **lemmatisation**

MS **morphosyntaxe** (PDD, morphologie, syntaxe dépendancielle)

# Chaîne de traitement

## Un corpus pour étudier ces phénomènes

- ▶ **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

# Chaîne de traitement

## Un corpus pour étudier ces phénomènes

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

# Chaîne de traitement

## Un corpus pour étudier ces phénomènes

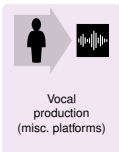
- ▶ **Annotations** nécessaires :
  - Ph **phonétique** (phones, syllabation, pauses, alignement au signal)
  - Le **lemmatisation**
  - MS **morphosyntaxe** (PDD, morphologie, syntaxe dépendancielle)
- ▶ Permettront d'évaluer les **interférences** :
  - ▶ Distribution relative des pauses et des piles : Ph × MS
  - ▶ Typologie des allongements : Ph × Le et Ph × MS
  - ▶ Comportement syntaxique des apostrophes, marqueurs de discours, etc. : Le × MS
- ▶ 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 :

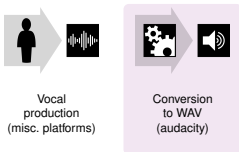
**Praat, CoNLL, Texte, XML pour TXM,**



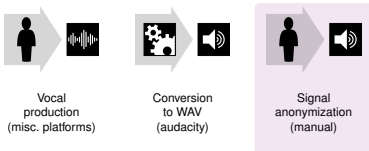
# Chaîne de traitement



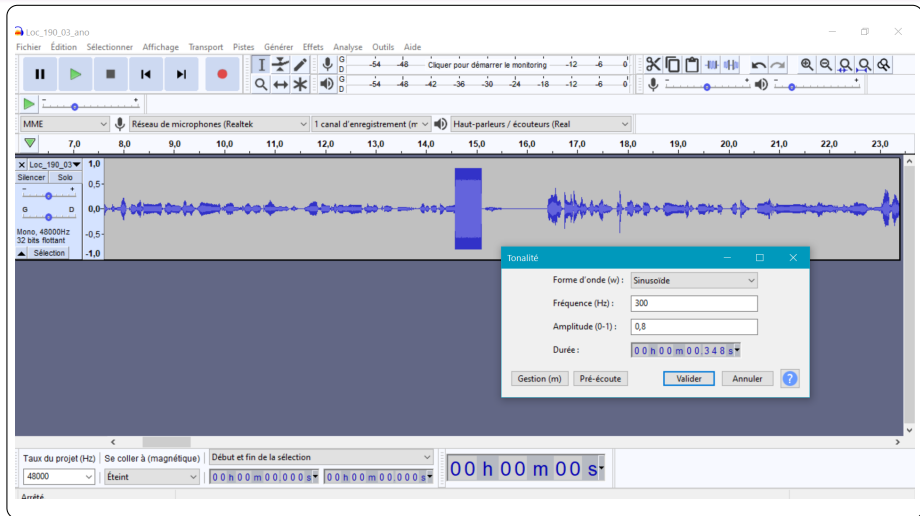
# Chaîne de traitement



# Chaîne de traitement

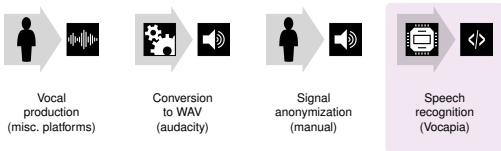


# Chaîne de traitement



Anonymisation manuelle (Audacity)

# Chaîne de traitement



# Chaîne de traitement

The screenshot shows the YobiYoba web interface. The main page is titled "Mes fichiers" and includes a navigation menu with "MES FICHIERS", "MON COMPTE", "API", and "COMPTEUR DE TEMPS : 09H28:36". A modal window titled "Lancer une transcription" is open, showing the following content:

**Lancer une transcription**

Document audio ou vidéo  
Loc\_271\_02.wav

1. Sélectionnez la langue du document (Choisissez "Automatique" seulement si vous ne connaissez pas la langue du document)

Automatique (identification de la langue par le système)

Les dernières langues utilisées : Français, Anglais, Français

Cochez cette case s'il s'agit de véritables enregistrements téléphoniques.

⚠ Attention, les transcriptions issues de traitements d'enregistrements téléphoniques ne contiennent pas de ponctuation.

2. Options d'amélioration de la transcription

Ajouter du vocabulaire (Non disponible)

Contextualisation (Non disponible)

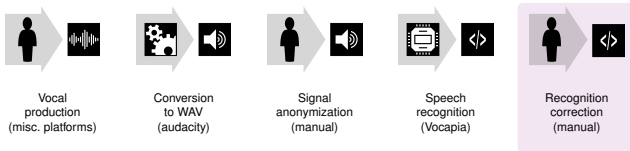
Buttons: Annuler, Lancer la transcription

The background shows a list of files:

Fichier	Date	Statut	Action
Loc_188_04.wav	23 mai 2022, 17:39:28	hier	Consulter
Loc_188_03.wav	23 mai 2022, 17:39:27	hier	Consulter
Loc_76_02.wav	23 mai 2022, 17:39:15	hier	Consulter

## Transcription automatique (Vocapia sur Yobiyoba)

# Chaîne de traitement



# Chaîne de traitement

The screenshot displays the Vocapia software interface. At the top, there is a menu bar with options like 'Fichier', 'Édition', 'Affichage', 'Historique', 'Marque-pages', 'Outils', and 'Aide'. Below the menu is a toolbar with icons for 'Fichier', 'Annuler', 'Rétablir', 'Edition', 'Rechercher', 'Options', 'Ré-aligner', 'Exporter', 'Aide', and 'Fermer'. The main area is divided into two panes. The left pane shows a list of items, with '1' selected. The right pane displays a transcript of a speech recording with several lines of text. The text is: 'si on peut faire ça lundi ça me va aussi en fait mais je suppose que la mère elle travaille donc c'est juste que le samedi fait ou pas j'irai un truc pour mon et et après ce qu'on comptais rentrer chez moi je si on fait ça dimanche tu vois je je peux rentrer on soit fait comme comme comme ils ont plus je m'adapte juste en fait il faut que je sache pour dire quoi mes parents'. The text is highlighted in green, and there are green bars above it indicating the alignment of the text with the audio. Below the text is a waveform of the audio recording. At the bottom, there is a progress bar showing 'Progression: Durée Totale: 00:00:13.6' and '00:00:31.3'. The interface is in French.

Correction manuelle (Vocapia)



# Chaîne de traitement



Vocal  
production  
(misc. platforms)



Conversion  
to WAV  
(audacity)



Signal  
anonymization  
(manual)



Speech  
recognition  
(Vocapia)



Recognition  
correction  
(manual)



Conversion to  
Praat TextGrid  
(script)

# Chaîne de traitement



Vocal production  
(misc. platforms)



Conversion to WAV  
(audacity)



Signal anonymization  
(manual)



Speech recognition  
(Vocapia)



Recognition correction  
(manual)

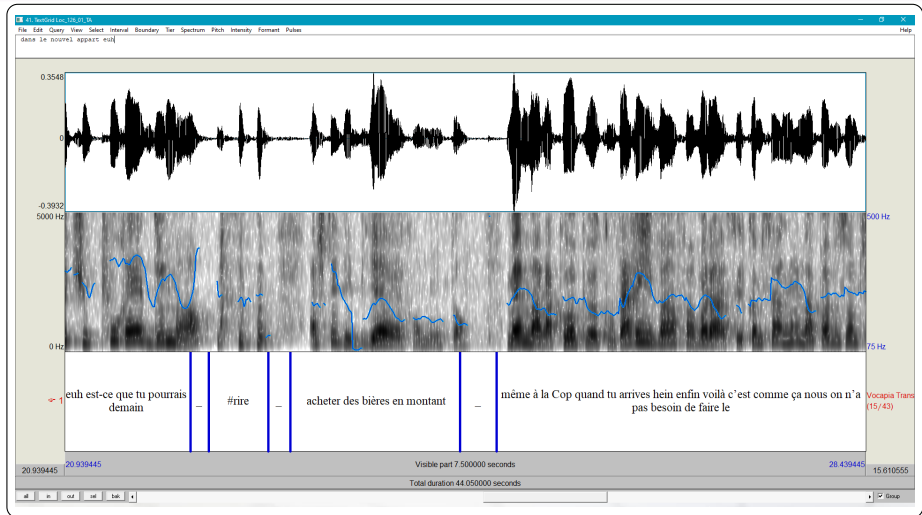


Conversion to Praat TextGrid  
(script)



Phonetic analysis preparation  
(manual)

# Chaîne de traitement



Préparation manuelle à l'analyse phonétique

# Chaîne de traitement



Vocal production  
(misc. platforms)



Conversion to WAV  
(audacity)



Signal anonymization  
(manual)



Speech recognition  
(Vocapia)



Recognition correction  
(manual)



Conversion to Praat TextGrid  
(script)

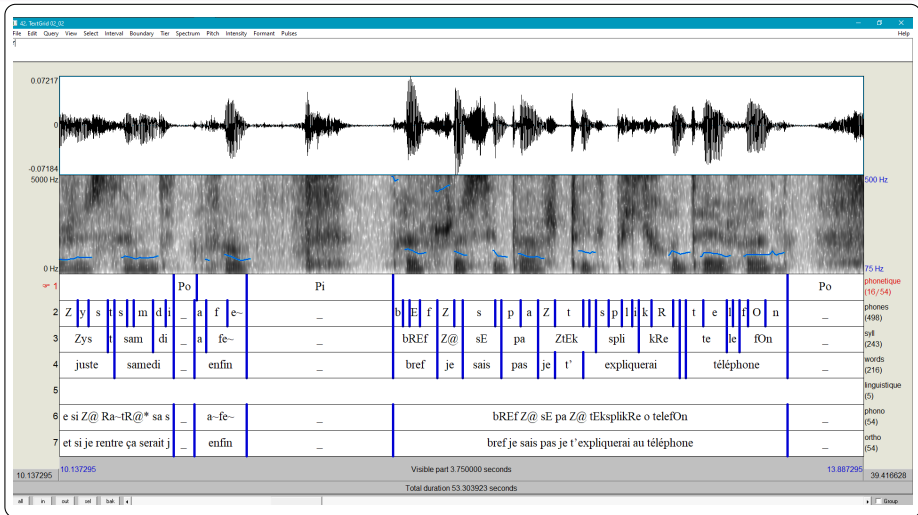


Phonetic analysis preparation  
(manual)



Phonetic analysis  
(EasyAlign)

# Chaîne de traitement



Analyse phonétique automatique (EasyAlign)

# Chaîne de traitement



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)

# Chaîne de traitement



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)

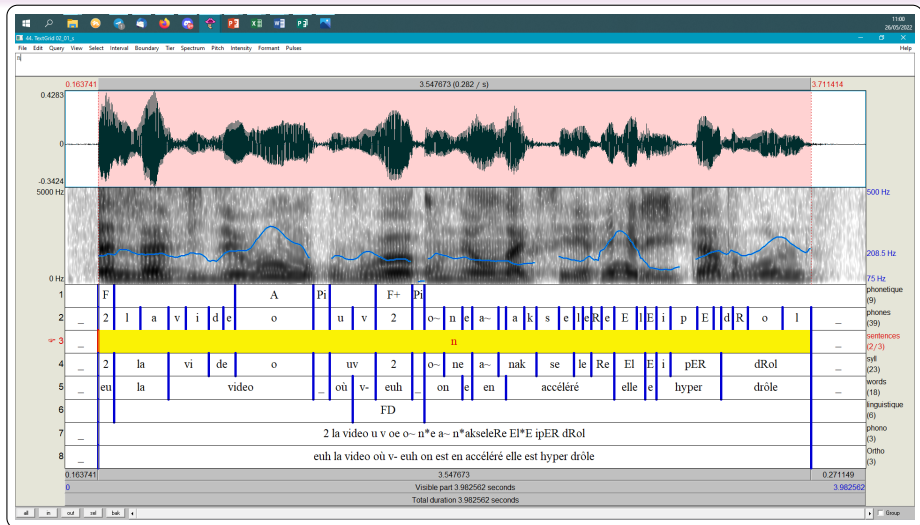


Phonetic and word correction  
(manual)



Sentence segmentation  
(manual)

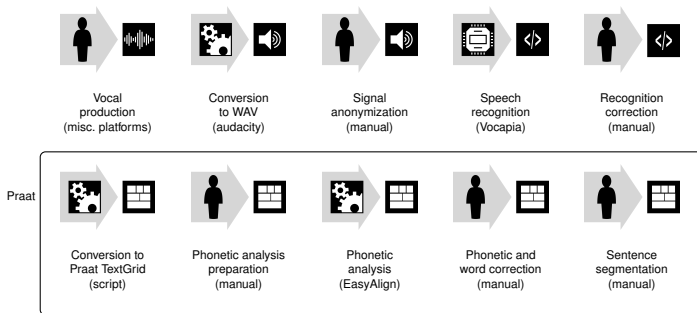
# Chaîne de traitement



Découpage manuel en phrases



# Chaîne de traitement



# Chaîne de traitement



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)

# Chaîne de traitement



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)



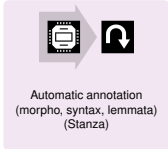
Phonetic and word correction  
(manual)



Sentence segmentation  
(manual)



Conversion to CoNLL  
(script)



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

# Chaîne de traitement



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)



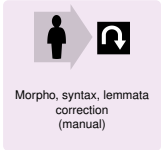
Sentence segmentation  
(manual)



Conversion to CoNLL  
(script)



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



Morpho, syntax, lemmata correction  
(manual)

# Chaîne de traitement



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)  
(Stanza)



Morpho, syntax, lemmata  
correction  
(manual)



Final corpus  
(script)

# Formats et Outils

Introduction

Recueil des données

Premières observations

Chaîne de traitement

**Formats et Outils**

Conclusion

# Formats et Outils

Les  
**VECAUX**  
VoCL

- ▶ XML
- ▶ Format du projet
- ▶ Toutes les annotations et métadonnées
- ▶ Stockage

# Formats et Outils

## Les VECAUX VoCL

- ▶ XML
- ▶ Format du projet
- ▶ Toutes les annotations et métadonnées
- ▶ Stockage

### TextGrid

- ▶ Spécifique à Praat
- ▶ Toutes les annotations
- ▶ Annotation et analyse du signal
- ▶ Requêtes « unidimensionnelles » (Praat)



# Formats et Outils

Fichier Édition Affichage Insertion Format Styles Feuilles Données Outils Fenêtre Aide

AE2 1.146

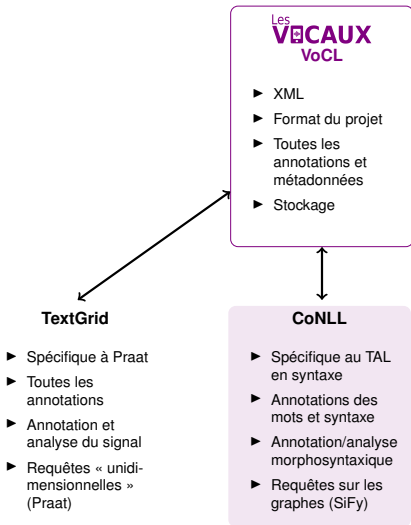
Ce document est ouvert en mode lecture seule.

	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
1	0,287		05_04_A	F-	0,445		05_01_A	A	0,471		05_01_A		2	0,27	05_01_A	FD	0,433		05_02_A	M	0,356
2	0,33						05_01_A	A	0,433		05_01_A		2	0,867	05_02_A	FD	1,144		05_02_A	M	0,152
3	0,26						05_01_A	A	1,239		05_01_A		2	0,187	05_04_A	FD	0,353		05_04_A	M	0,117
4	0,27						05_03_A	A	0,25		05_01_A	a		0,08	05_05_A	FD	0,205		05_04_A	M	0,37
5	0,154						05_03_A	A	0,84		05_01_A	a		0,05	05_08_A	FD	0,367		05_07_A	M	0,123
6	0,405						05_03_A	A	0,611		05_01_A	a		0,05	05_09_A	FD	0,12		05_07_A	M	0,219
7	0,867						05_03_A	A	0,23		05_01_A	a		0,03					05_07_A	M	0,449
8	0,187						05_03_A	A	0,28		05_01_A	a		0,079					05_08_A	M	0,322
9	0,408						05_04_A	A	0,408		05_01_A	a		0,139					05_08_A	M	0,115
10	0,474						05_04_A	A	0,275		05_01_A	a-		0,1							
11	0,148						05_04_A	A	0,3		05_01_A	a-		0,101							
12	0,675						05_04_A	A	0,361		05_01_A	a-		0,13							
13	1,24						05_04_A	A	0,377		05_01_A	a-		0,08							
14	0,427						05_05_A	A	0,25		05_01_A	a-		0,03							
15	0,223						05_05_A	A	0,604		05_01_A	be		0,145							
16	0,657						05_05_A	A	0,287		05_01_A	bje-		0,128							
17	0,381						05_05_A	A	0,18		05_01_A	d@		0,06							
18	0,342						05_05_A	A	0,463		05_01_A	d@		0,06							
19	0,342						05_06_A	A	0,621		05_01_A	d@		0,15							
20	0,16						05_06_A	A	0,695		05_01_A	d@		0,06							
21	0,154						05_07_A	A	0,416		05_01_A	d@		0,08							
22	0,454								0,4567143		05_01_A	de		0,142							
23	0,5										05_01_A	di		0,26							
24	0,394										05_01_A	di		0,18							
25	0,303										05_01_A	do		0,102							
26	0,32										05_01_A	do-		0,14							
27	0,388										05_01_A	do-		0,133							
28	0,125										05_01_A	do-k		0,194							
29	0,205										05_01_A	do-k		0,181							
30	0,348										05_01_A	do-k		0,17							

(Praat)

## Décomptes Praat

# Formats et Outils



localhost/~mzi/syfi/

Menu Open paste basket Get current state

SPARQL Query GREW Query

```

1 % In order to build the concordance and to sort results,
2 % a node named "FOC" must be defined. Nodes numbered C1, C2
3 % C3, etc. are highlighted in the treeview
4
5 pattern {
6   V -[nsubj]-> FOC;
7 }
            
```

Powered by Bruno Guillaume's Grew

**Sort keys**

Ref. node	Sort key	Actions
FOC	Lemma	↑ ↓ 🗑
FOC	UPOS	↑ ↓ 🗑

+

**Options**

Output format: On screen

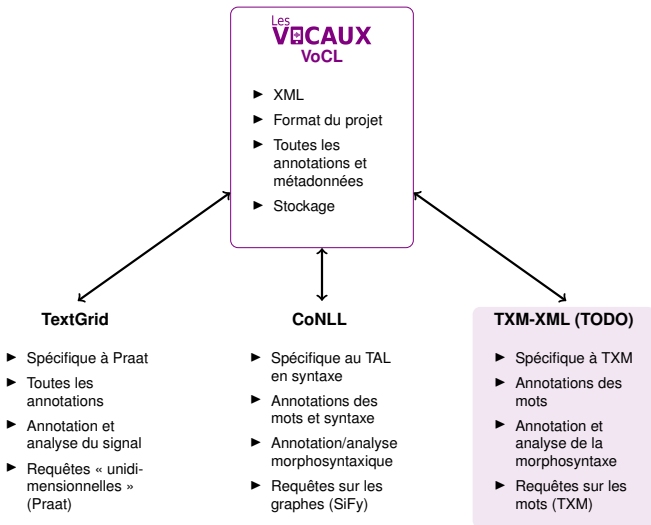
**Results : 24 hits**

N	FOC.lemma	FOC.upos			
1	il	PRON	et après	je	t' appelle demain
2	il	PRON	mais ça	tu	me rediras
3	il	PRON	mais vu que #name	elle	perd souvent les trucs du coup je vous l' enverrai

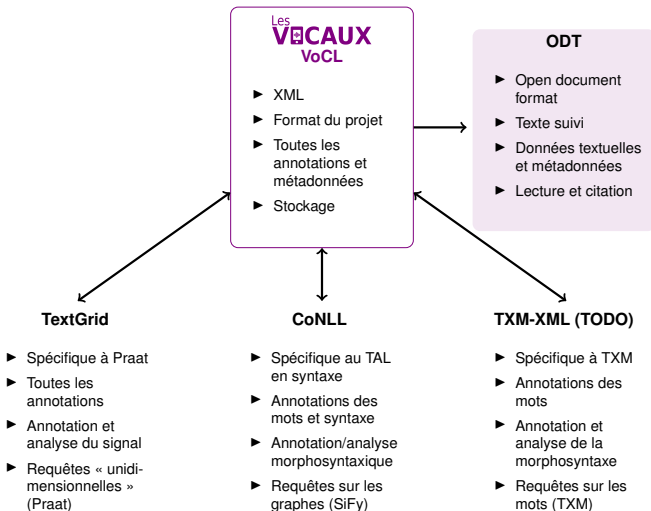
Diagram illustrating the Syfi tree structure for the sentence "et après je t' appelle demain". The root node is "appelle" (VERB). The FOC node is "je" (PRON). The tree structure shows the following relations:

- cc (et) connected to the root.
- advmod (après) connected to the root.
- obl:arg (t') connected to the root.
- advmod (demain) connected to the root.
- FOC (je) connected to the root.

# Formats et Outils



# Formats et Outils



# Formats et Outils

The screenshot shows a text editor window with a menu bar (Fichier, Édition, Affichage, Insertion, Format, Styles, Tableau, Formulaire, Outils, Fenêtre, Aide) and a toolbar. The text content is as follows:

**02\_01**

- speaker: 02
- genre: F
- age: 20
- current\_location: Strasbourg
- current\_country: France
- home\_location: Haut-Rhin
- home\_country: France
- message\_usage: frequent
- filename: 02\_01

\_ euh la video \_ où v- euh \_ on est en accéléré elle est hyper drôle \_

**02\_02**

- speaker: 02
- genre: F
- age: 20
- current\_location: Strasbourg

Lecture « simple »

# Conclusion

Introduction

Recueil des données

Premières observations

Chaîne de traitement

Formats et Outils

**Conclusion**

# Conclusion

**Nous avons présenté :**



# Conclusion

## Nous avons présenté :

- ▶ Intérêt/type des données

# Conclusion

## **Nous avons présenté :**

- ▶ Intérêt/type des données
- ▶ Phénomènes saillants

# Conclusion

## **Nous avons présenté :**

- ▶ Intérêt/type des données
- ▶ Phénomènes saillants
- ▶ Chaîne de traitement (beaucoup d'étapes, mixage manuel/automatique)

# Conclusion

## Nous avons présenté :

- ▶ Intérêt/type des données
- ▶ Phénomènes saillants
- ▶ Chaîne de traitement (beaucoup d'étapes, mixage manuel/automatique)
- ▶ Format variable des sorties (permet l'interopérabilité)

# Conclusion

## **Nous avons présenté :**

- ▶ Intérêt/type des données
- ▶ Phénomènes saillants
- ▶ Chaîne de traitement (beaucoup d'étapes, mixage manuel/automatique)
- ▶ Format variable des sorties (permet l'interopérabilité)

## **Pistes d'exploitation**

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

# Conclusion

## Nous avons présenté :

- ▶ Intérêt/type des données
- ▶ Phénomènes saillants
- ▶ Chaîne de traitement (beaucoup d'étapes, mixage manuel/automatique)
- ▶ Format variable des sorties (permet l'interopérabilité)

## Pistes d'exploitation

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

# Conclusion

## Nous avons présenté :

- ▶ Intérêt/type des données
- ▶ Phénomènes saillants
- ▶ Chaîne de traitement (beaucoup d'étapes, mixage manuel/automatique)
- ▶ Format variable des sorties (permet l'interopérabilité)

## Pistes d'exploitation

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

```
<glikman@unistra.fr>  
<cfauth@unistra.fr>  
<nicolas.mazziotta@uliege.be>  
<christophe.benzitoun@univ-lorraine.fr>
```