

Projet « Les Vocaux »: Outils et formats

Julie Glikman

Université de Strasbourg (UR LiLPa)

&

Nicolas Mazziotta

Université de Liège (UR Traverses)

TraSoGal – Liège, 24 juin 2022

Les
VOCALAUX

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 vous 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

Objectifs du projet « Les Vocaux »

- ▶ Constituer un corpus rassemblant des messages vocaux

Introduction

Objectifs du projet « Les Vocaux »

- ▶ Constituer un corpus rassemblant des messages vocaux
- ▶ Annoter les messages :
 - ▶ phonétique (phones, syllabation, pauses, alignement au signal)
 - ▶ lemmatisation
 - ▶ PDD et morphologie
 - ▶ syntaxe dépendancielle

Introduction

Objectifs du projet « Les Vocaux »

- ▶ Constituer un corpus rassemblant des messages vocaux
- ▶ Annoter les messages :
 - ▶ phonétique (phones, syllabation, pauses, alignement au signal)
 - ▶ lemmatisation
 - ▶ PDD et morphologie
 - ▶ syntaxe dépendancielle
- ▶ Distribuer le corpus dans des formats utilisables

Introduction

Objectifs du projet « Les Vocaux »

- ▶ Constituer un corpus rassemblant des messages vocaux
- ▶ Annoter les messages :
 - ▶ phonétique (phones, syllabation, pauses, alignement au signal)
 - ▶ lemmatisation
 - ▶ PDD et morphologie
 - ▶ syntaxe dépendancielle
- ▶ Distribuer le corpus dans des formats utilisables

Membres du projet

- ▶ Julie Glikman (porteuse), Université de Strasbourg
- ▶ Christophe Benzitoun, Université de Lorraine
- ▶ Camille Fauth, Université de Strasbourg
- ▶ Nicolas Mazziotta, Université de Liège

Objectif de l'exposé et plan

Objectif

- ▶ Présentation des méthodes de recueil et de traitement
- ▶ En cours de test sur un corpus « martyr »

Objectif de l'exposé et plan

Objectif

- ▶ Présentation des méthodes de recueil et de traitement
- ▶ En cours de test sur un corpus « martyr »

Plan

Introduction

Recueil des données

Chaîne de traitement

Formats et Outils

Conclusion

Recueil des données

Introduction

Recueil des données

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
- ▶ genre : 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)

Chaîne de traitement

Introduction

Recueil des données

Chaîne de traitement

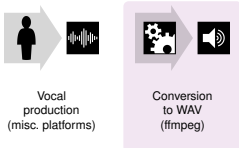
Formats et Outils

Conclusion

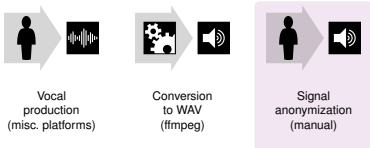
Chaîne de traitement



Chaîne de traitement



Chaîne de traitement



Chaîne de traitement

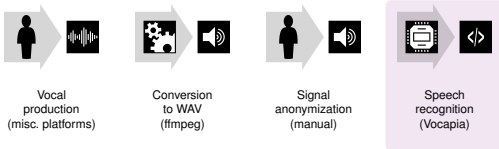
The screenshot displays the Audacity audio editing software interface. At the top, the menu bar includes options like 'Fichier', 'Édition', 'Sélectionner', 'Affichage', 'Transport', 'Pistes', 'Générer', 'Effets', 'Analyse', 'Outils', and 'Aide'. Below the menu is a toolbar with various editing tools. The main workspace shows a waveform for a track named 'Loc_190_03'. A blue selection box is placed over a segment of the waveform. A 'Tonalité' (Tone) dialog box is open in the foreground, with the following settings:

- Forme d'onde (w): Sinusoïde
- Fréquence (Hz): 300
- Amplitude (0-1): 0,8
- Durée: 00 h 00 m 00,348 s

Buttons at the bottom of the dialog include 'Gestion (m)', 'Pré-écoute', 'Valider', 'Annuler', and a help icon. The bottom status bar shows 'Taux du projet (Hz): 48000', 'Se coller à (magnétique): Éteint', and 'Début et fin de la sélection: 00 h 00 m 00,000 s - 00 h 00 m 00,000 s'.

Anonymisation manuelle (Audacity)

Chaîne de traitement



Chaîne de traitement

The screenshot shows the YobiYoba web application interface. A modal window titled "Lancer une transcription" is open, allowing the user to start a transcription. The modal has three tabs: "Transcrire" (selected), "Aligner une transcription", and "Détection de langues".

Transcrire

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

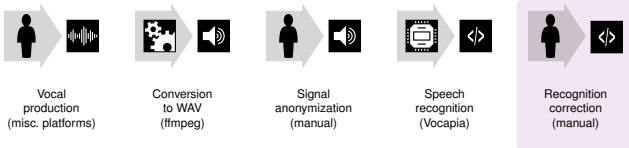
Annuler Lancer la transcription

The background interface shows "Mes fichiers" (My files) with a sidebar for file management and a main area displaying a list of audio files:

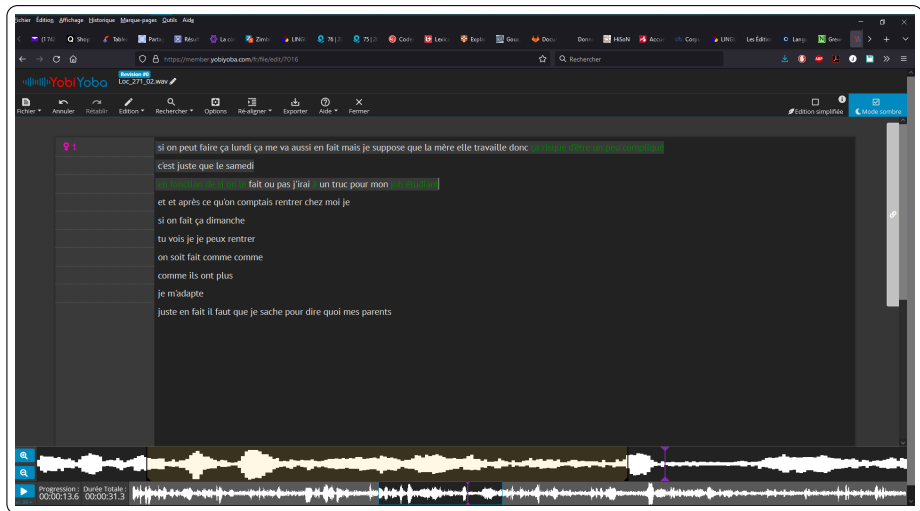
| Fichier | Date | Statut | Action |
|----------------|-----------------------|--------|-----------|
| Loc_188_04.wav | 23 mai 2022, 17:39:28 | hier | Consulter |
| Loc_188_03.wav | 23 mai 2022, 17:39:27 | hier | Consulter |
| Loc_76_02.wav | 23 mai 2022, 17:39:15 | hier | Consulter |

Transcription automatique (Vocapia sur Yobiyoba)

Chaîne de traitement



Chaîne de traitement



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

Progression : Durée Totale : 00:00:13.6 00:00:31.3

Correction manuelle (Vocapia)

Chaîne de traitement



Vocal production
(misc. platforms)



Conversion to WAV
(ffmpeg)



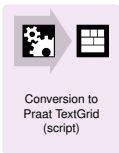
Signal anonymization
(manual)



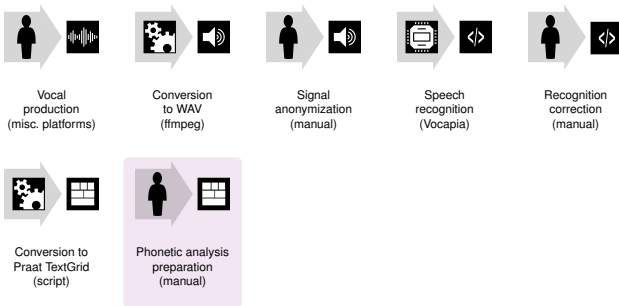
Speech recognition
(Vocapia)



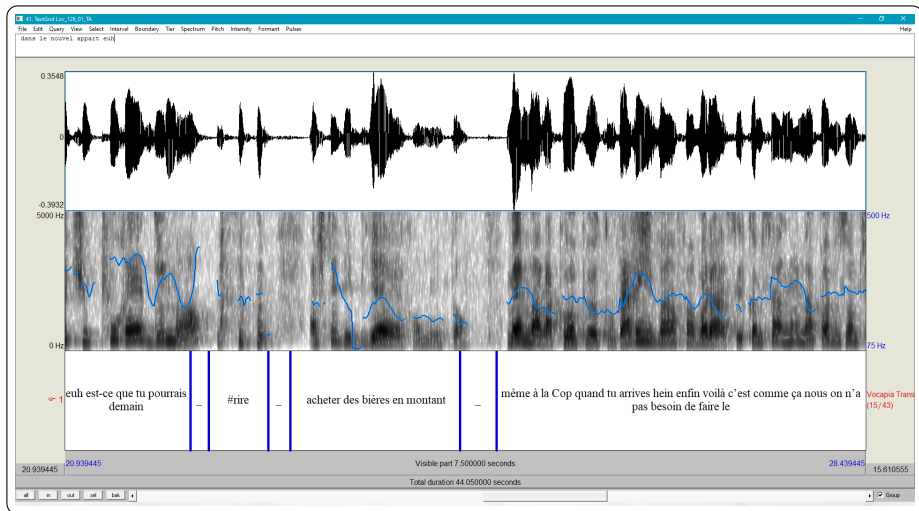
Recognition correction
(manual)



Chaîne de traitement



Chaîne de traitement



Préparation manuelle à l'analyse phonétique

Chaîne de traitement



Vocal
production
(misc. platforms)



Conversion
to WAV
(ffmpeg)



Signal
anonymization
(manual)



Speech
recognition
(Vocapia)



Recognition
correction
(manual)



Conversion to
Praat TextGrid
(script)

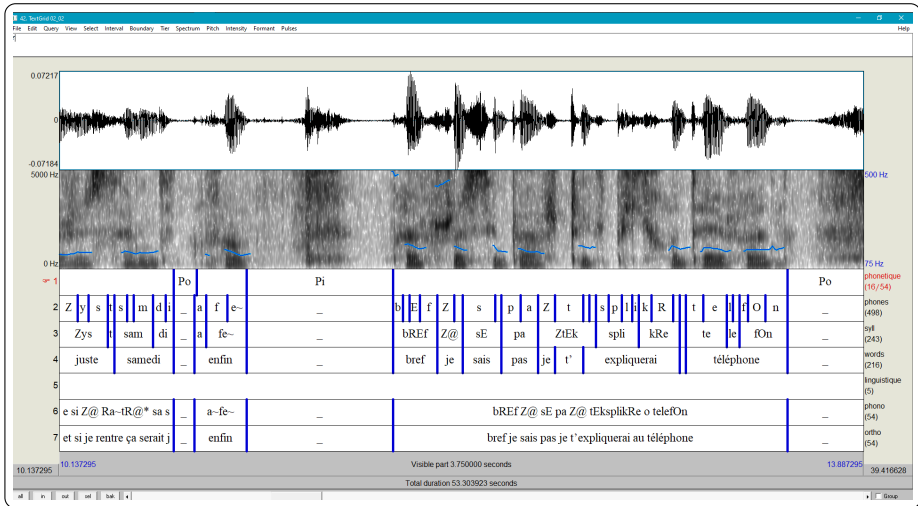


Phonetic analysis
preparation
(manual)



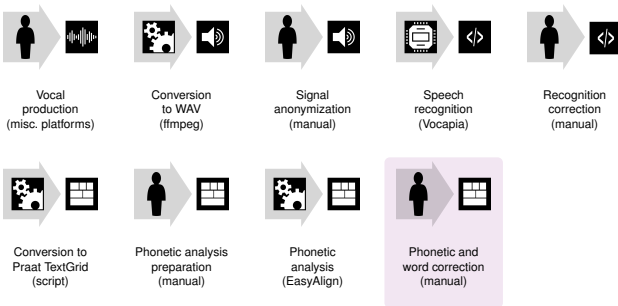
Phonetic
analysis
(EasyAlign)

Chaîne de traitement

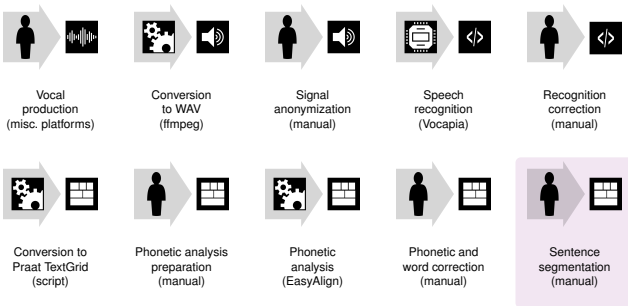


Analyse phonétique automatique (EasyAlign)

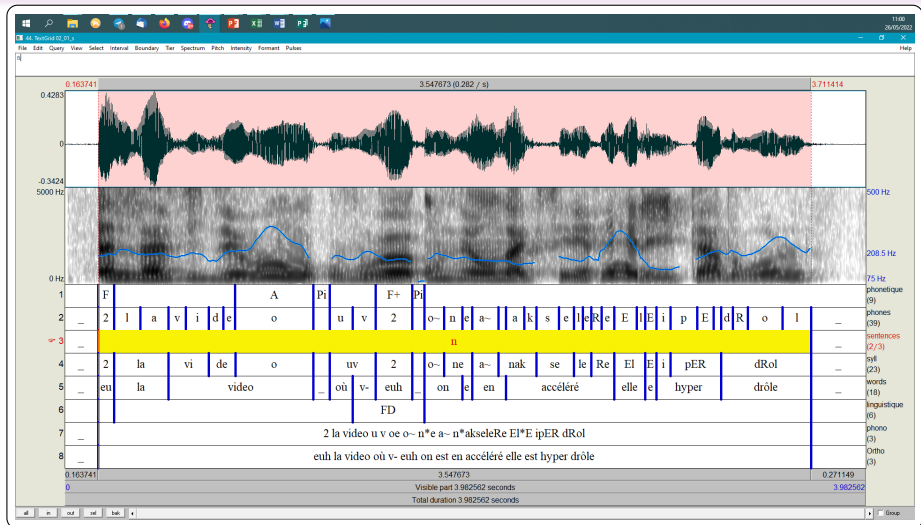
Chaîne de traitement



Chaîne de traitement

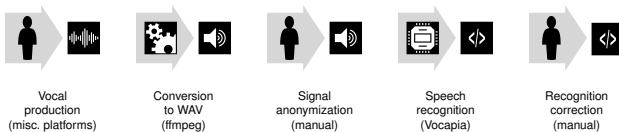


Chaîne de traitement

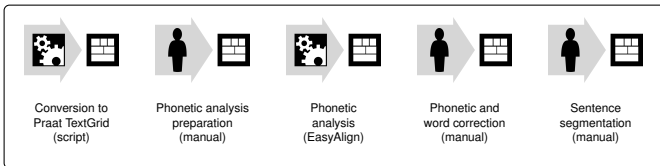


Découpage manuel en phrases

Chaîne de traitement



Praat



Chaîne de traitement



Vocal production
(misc. platforms)



Conversion to WAV
(ffmpeg)



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)

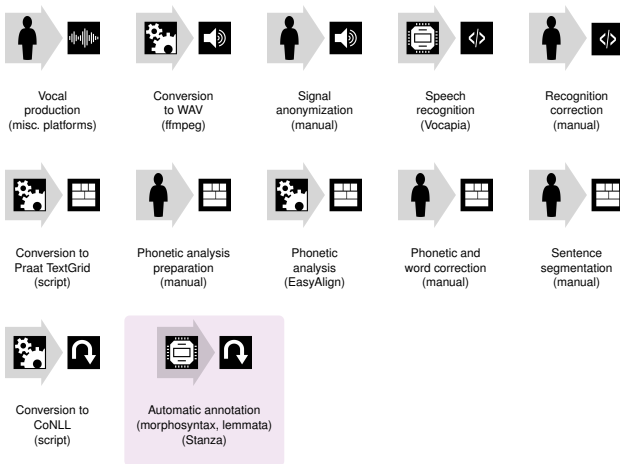


Sentence segmentation
(manual)

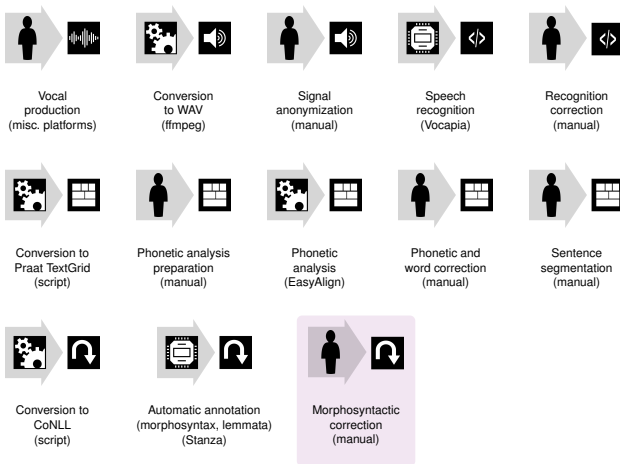


Conversion to CoNLL
(script)

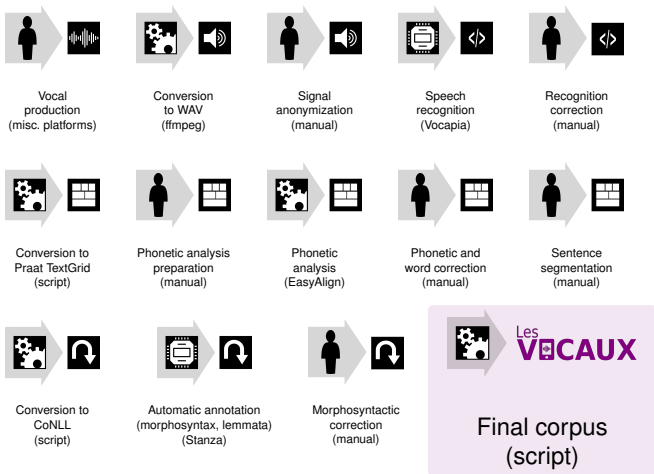
Chaîne de traitement



Chaîne de traitement



Chaîne de traitement



Formats et Outils

Introduction

Recueil des données

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 VoCAUX 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 Feuille Données Outils Fenêtre Aide

AE2 1.146

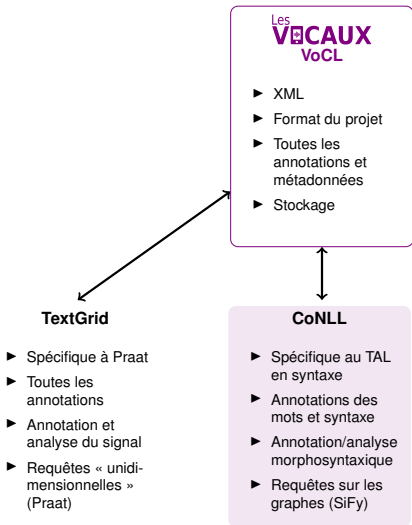
Ce document est ouvert en mode lecture seule.

| | 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 | 1,146 | 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 | | | | | | | |

- Requêtes « unidimensionnelles » (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

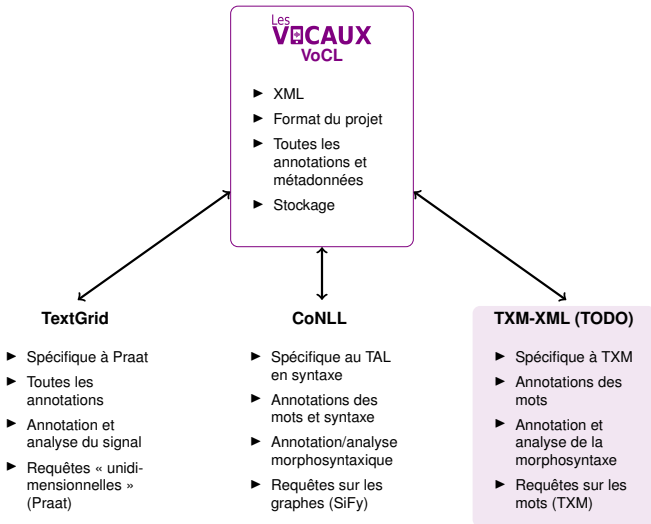
Output format: On screen

Results : 24 hits

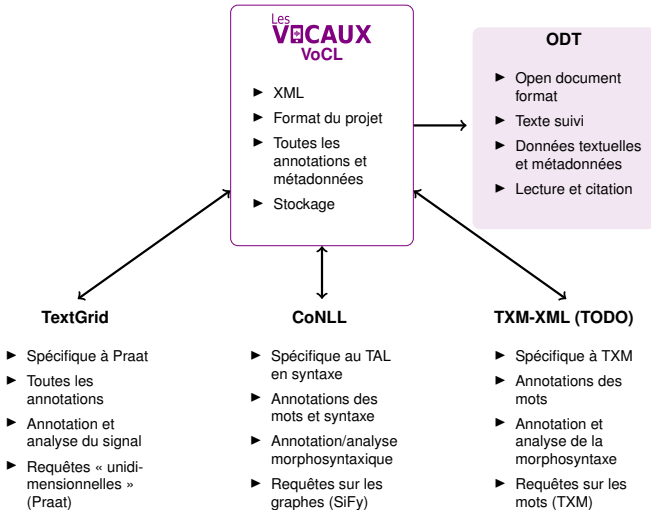
| N | FOC.lemma | FOC.upos | | | |
|---|-----------|----------|-------------------|-------------|--|
| 1 | il | PRON | et après | je | t' appelle demain |
| 2 | il | PRON | mais ça | tu | me rediras |
| 3 | il | PRON | mais vu que #name | elle | perd souvent les trucs du coup je vous l' enverrai |

Requête dans Syfi

Formats et Outils



Formats et Outils



Formats et Outils

Fichier Édition Affichage Insertion Format Styles Tableau Formulaire Outils Fenêtre Aide

Normal Liberation Se 12

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

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
- ▶ Chaîne de traitement (beaucoup d'étapes, mixage manuel/automatique)

Conclusion

Nous avons présenté :

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

Conclusion

Nous avons présenté :

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

Nous vous demandons :

Conclusion

Nous avons présenté :

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

Nous vous demandons :

- ▶ Comment exploiteriez-vous pareil corpus ?

Conclusion

Nous avons présenté :

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

Nous vous demandons :

- ▶ Comment exploiteriez-vous pareil corpus ?
- ▶ Quelles constructions vous intéressent ?

Conclusion

Nous avons présenté :

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

Nous vous demandons :

- ▶ Comment exploiteriez-vous pareil corpus ?
- ▶ Quelles constructions vous intéressent ?
- ▶ Quel est votre avis sur ce choix technique ?
 - ▶ Anonymisation « beep » : [LesVocaux, 190-02](#)
 - ▶ Anonymisation « euheuh » : [LesVocaux, 190-02](#)

Conclusion

Nous avons présenté :

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

Nous vous demandons :

- ▶ Comment exploiteriez-vous pareil corpus ?
- ▶ Quelles constructions vous intéressent ?
- ▶ Quel est votre avis sur ce choix technique ?
 - ▶ Anonymisation « beep » : [LesVocaux, 190-02](#)
 - ▶ Anonymisation « euheuh » : [LesVocaux, 190-02](#)

<glikman@unistra.fr> & <nicolas.mazziotta@uliege.be>