

# ACSYNT :

## un corpus oral du français contemporain

Elisabeth DELAIS-ROUSSARIE

et Didier BOURIGAULT, Injoo CHOI-JONIN, Cécile FABRE, Lucia MOLINU, Magali ROUQUIER, Jean-Michel TARRIER

❖	<u>Présentation générale</u>	Page 2
❖	<u>Contenu des CDs</u>	Page 11
❖	<u>Enregistrement et numérisation des données</u>	Page 24
❖	<u>Transcription des données : modalités et conventions</u>	Page 31
❖	<u>Documentation linguistique diverse</u>	Page 41
❖	<u>Identité des locuteurs</u>	Page 43
❖	<u>Quelques liens utiles</u>	Page 103

Ce corpus a pu être constitué grâce à deux financements distincts attribués pour soutenir un projet de recherche visant à étudier les liens entre syntaxe et prosodie dans différents styles de français oral, à savoir :

- **ACI jeune chercheur 2001 / projet 67031** (« *Etude variationniste de l'accentuation et des interfaces Phonologie/ Syntaxe/ Sémantique en français contemporain : Descriptions et Modélisations* »)
- **JEMSTIC Jeune chercheur / CNRS** (Financement accordé à des projets de recherche de jeunes chercheurs dans le cadre du Département STIC du CNRS)

# Présentation générale

## Le corpus ACSYNT

Le corpus ACSYNT est un corpus oral du français contemporain de 124 000 mots. Il est composé de trois types de documents :

- des textes lus ;
- des présentations monologuées préparées ;
- des entretiens guidés.

La répartition entre les différents types de document sur l'ensemble du corpus est la suivante :

	Textes lus	Présentation préparée	Entretien guidé	TOTAL
Nombre de mots	17 858 mots	19 575 mots	86 584 mots	124 017 mots
% sur l'ensemble du corpus	14,4 %	15,8 %	69,8 %	
Nombre de locuteurs enregistrés	23	17	17	23

Chaque document ou sous-document est représenté par une série de trois fichiers distincts ayant le même nom, mais une extension distincte :

- un fichier audio au format et à l'extension WAV ( 22 kHz, 16 bits, Mono) ;
- un fichier à l'extension « TextGrid », lisible sous [PRAAT](#), un logiciel téléchargeable permettant de visualiser simultanément le signal de parole (ou fichier son) et la transcription orthographique correspondante ;
- un fichier texte ayant l'extension TXT et qui renferme sous forme textuelle brute le contenu du fichier audio auquel il est associé.

Pour faciliter l'utilisation du corpus, les noms des fichiers ont tous la même forme, à savoir 4 lettres suivies d'un nombre. En outre, le choix des lettres et des nombre est fait de façon à respecter une certaine syntaxe dont voici les principales caractéristiques :

- les trois premières lettres sont choisies en fonction du nom et du prénom des locuteurs. Dans l'ensemble du corpus, chaque trigramme sera d'ailleurs traité comme le nom des locuteurs, afin de préserver leur anonymat ;

- la quatrième lettre est choisie en fonction du type de document contenu dans le fichier, quelle que soit son extension (wav, txt ou TextGrid). La lettre E est utilisée pour les entretiens guidés, la lettre P pour les présentations monologuées préparées et la lettre T pour les textes lus.
- le nombre qui suit les quatre lettres sert à numéroter le fichier, dans la mesure où, pour chaque type de document (lecture de textes, présentation monologuée préparée et entretien guidé), les données ont été sous segmentées en petits fichiers. Ainsi, pour la lecture de texte, un seul texte est contenu par fichier, bien que les locuteurs aient tous lus trois textes. De même, les entretiens guidés et les présentations ont été segmentés en fichiers n'excédant pas 2 à 3 minutes de parole.

### **Exemples :**

Soit le locuteur ASD, les fichiers associés correspondant aux trois textes lus T1, T2 et T3 ont pour nom :

- ASDT1.wav, ASDT1.txt et ASDT1.TextGrid ;
- ASDT2.wav, ASDT2.txt et ASDT2.TextGrid ;
- ASDT3.wav, ASDT3.txt et ASDT3.TextGrid.

Considérons maintenant le cas de la présentation préparée enregistrée pour le locuteur ROM. Comme cette présentation dure une dizaine de minutes, elle a été sous segmentée en 5 sous fichiers. Les différentes séries de fichiers ont pour nom :

- ROMP1.wav, ROMP1.txt et ROMP1.TextGrid ;
- ROMP2.wav, ROMP2.txt et ROMP2.TextGrid ;
- ROMP3.wav, ROMP3.txt et ROMP3.TextGrid.
- ROMP4.wav, ROMP4.txt et ROMP4.TextGrid ;
- ROMP5.wav, ROMP5.txt et ROMP5.TextGrid.

Pour finir, considérons le cas de l'entretien guidé du locuteur GOL. Sachant que cet entretien a été segmenté en 10 sous fichiers. Les séries de fichiers ont pour nom :

- GOLE1.wav, GOLE1.txt et GOLE1.TextGrid ;
- GOLE2.wav, GOLE2.txt et GOLE2.TextGrid ;
- GOLE3.wav, GOLE3.txt et GOLE3.TextGrid.
- GOLE4.wav, GOLE4.txt et GOLE4.TextGrid ;
- GOLE5.wav, GOLE5.txt et GOLE5.TextGrid
- GOLE6.wav, GOLE6.txt et GOLE6.TextGrid ;
- GOLE7.wav, GOLE7.txt et GOLE7.TextGrid ;
- GOLE8.wav, GOLE8.txt et GOLE8.TextGrid.
- GOLE9.wav, GOLE9.txt et GOLE9.TextGrid ;
- GOLE10.wav, GOLE10.txt et GOLE10.TextGrid

## Les textes lus

Sur l'ensemble du corpus, nous comptons 14,4 % de documents constitués de textes lus par 23 locuteurs. Cela représente donc 17 858 mots et 68 séries de fichiers (une série étant composée des trois types de fichiers différents : un fichier audio, un fichier texte et un fichier TextGrid). Les textes proposés aux locuteurs sont au nombre de trois :

- T1 : un extrait de *l'Equipe* : « **La tempête de glace** »
- T2 : un extrait de *Ouest France* : « **La campagne accélérée** »
- T3 : un extrait de « *Voyage au bout de la nuit* » de Louis-Ferdinand Céline

### **T1 : Extrait de *l'Equipe***

**Extrait de *l'Equipe*.** Vendredi 15 février

« La tempête de glace, voilà comment la presse américaine a baptisé depuis deux jours « l'affaire » de la compétition des couples en patinage artistique. Comme s'il s'agissait de la répétition de l'opération « tempête du désert » accompagnant la guerre du Golfe en 1991. Rien que cela ! C'est que depuis lundi soir et la décision de cinq juges (et non la seule Française) contre quatre d'offrir le titre aux Russes face aux Canadiens, ce qui nous a paru erroné mais pas honteux, il n'est quasiment plus question que de cela dans les médias nord-américains qui découvrent tout à coup que le patinage artistique peut être le théâtre de petits arrangements entre amis. Quelle surprise ! Et surtout quel formidable bal des faux-culs en attendant celui de la compétition de danse qui ouvre ce soir. Nos spécialistes du patinage vous racontent ci-dessous tout ce qu'il faut penser de cette histoire mais la vive ampleur qu'elle a prise vaut également quelques commentaires.

Feindre aujourd'hui de découvrir que le patinage est un sport sujet à caution relève soit de la totale naïveté soit de l'extrême hypocrisie (ou des deux, mon général). Quant à nous il y a malheureusement longtemps que nous savons qu'une vraie réforme de ce sport s'impose d'autant que beaucoup de nos champions en ont souffert dans le passé. Il y a deux ans, le président de la Fédération Française avait proposé que les juges ne dépendent plus des fédérations nationales comme aujourd'hui mais de l'ISU, la Fédération internationale, première étape vers la seule solution possible, la constitution d'un corps de juges pros et indépendants. Et devinez qui a fait capoter cette réforme ? Les Anglo-saxons justement.... »

## **T2 : Extrait de *Ouest France***

**Extrait de *Ouest France*.** Mardi 12 février

« La campagne accélérée :

C'est parti, pour sa quatrième campagne présidentielle ! A vrai dire, ça l'était déjà depuis longtemps... Mais, puisqu'il faut bien formaliser un tel engagement avec des mots, Jacques Chirac les a prononcés : « Oui, je suis candidat ! ». Et comme y ajouter un symbole ne gâte rien, il a annoncé sa candidature, une fois encore, depuis la province : après le Nord (Lille) en 1995, ce fut, hier, dans le Sud, en Avignon, la ville de la « tombeuse » d'Elisabeth Guigou.

Ce faisant, Jacques Chirac semble avoir précipité une annonce prévue pour le mois prochain. Ainsi rassure-t-il ses partisans, inquiets d'un attentisme que beaucoup jugeaient stratégiquement néfaste. Ainsi enlève-t-il un argument de poids à ceux qui lui reprochaient d'éviter le débat. Mais, en même temps, il donne le sentiment d'avoir cédé à la pression ambiante. A moins que son solide instinct politique ne lui ait dicté ce choix du moment. Le poison des affaires, la baisse dans les sondages, les attaques de plus en plus acérées de ses adversaires, la montée en puissance de Chevènement, autant d'éléments qui ont créé un climat défavorable au lancement de sa candidature. Pour tenter de regagner une confiance pour le moins érodée et des électeurs gaullistes en déshérence, une campagne trop courte pouvait se révéler insuffisante. »

## **T3 : Extrait de *Voyage au bout de la nuit*. Louis-Ferdinand Céline**

**Extrait de *Voyage au bout de la nuit*. Louis-Ferdinand Céline**

« Entre deux grosses roches il s'était établi une sorte de cagna, à l'abri, me fit-il remarquer, des tornades de l'Est, les plus mauvaises, les plus rageuses. Je voulus bien admettre que c'était un avantage, mais quant à la case elle-même, c'était sûrement à la dernière catégorie miteuse qu'elle appartenait, demeure presque théorique, effilochée de partout. Je m'attendais bien à quelque chose de ce genre-là en fait d'habitation, mais tout de même la réalité dépassait mes prévisions.

Je dus lui sembler tout à fait navré au copain, car il m'interpella assez brusquement pour me faire sortir de mes réflexions. « Allez donc, vous serez moins mal encore ici qu'à la guerre ! Ici, après tout, on peut se débrouiller ! On bouffe mal, c'est exact, et

pour boire, c'est une vraie boue, mais on peut dormir tant qu'on veut... Pas de canons ici mon ami ! Pas de balles non plus ! En somme c'est une affaire ! » Il parlait un peu dans le même ton que l'Agent général mais des yeux pâles comme ceux d'Alcide, il avait.

(...)

C'est vrai, repris-je, y a pas pire que la guerre !

C'était assez pour le moment comme confidences, je n'avais pas envie d'en dire davantage. Mais ce fut lui qui continua sur le même sujet.

« Surtout maintenant qu'on les fait si longues les guerres...qu'il ajouta. Enfin vous verrez mon ami qu'ici c'est pas très drôle, voilà tout ! Y a rien à faire... C'est comme des espèces de vacances... Seulement voilà des vacances ici ! n'est-ce pas !... Enfin, ça dépend peut-être des natures, j'peux rien dire... » »

## Les présentations monologuées préparées

Les présentations monologuées préparées représentent 15,8 % des données du corpus et comptent 19 575 mots. Elles ont été obtenues auprès de 17 locuteurs (15 femmes et deux hommes) qui avaient pour consigne de préparer une intervention. Selon les locuteurs, les situations dans lesquelles se sont effectués les enregistrements, ainsi que les thèmes développés diffèrent : pour une locutrice (COA), la présentation monologuée préparée n'est autre qu'une intervention dans des Journées d'Etudes scientifiques, alors que pour les autres, elle prend la forme d'un petit exposé de cinq minutes environ dans lequel le locuteur présente sa profession, son cursus universitaire, son domaine de recherche, etc.

Un récapitulatif de type de données disponibles pour chaque locuteur est proposé dans le tableau ci-dessous :

Locuteur	Situation d'enregistrement	Thème développé	Durée approximative de l'enregistrement
Loc1_ASD	Salle de cours, devant l'enquêteur.	Présentation de ce que sont les sciences du langage.	3 min.
Loc2_ASH	Salle de cours, devant l'enquêteur.	Présentation du cursus personnel de la locutrice (Hypokhâgne, Université, etc.).	7 min.

<b>Loc3_BAV</b>	Salle de cours, devant l'enquêteur.	Présentation des qualités et des défauts de la locutrice en vue d'un recrutement.	7 min.
<b>Loc4_BOA</b>	Salle de cours, devant l'enquêteur.	Présentation des perspectives professionnelles de la locutrice (elle souhaite devenir institutrice).	4 min
<b>Loc7_COA</b>	Journées d'Etudes linguistiques, devant un public.	Communication sur la compétence linguistique en terminologie.	32 min.
<b>Loc11_GOL</b>	Salle de cours, devant l'enquêteur.	Présentation de métier d'enseignant d'Espagnol en collège et lycée.	3 min. 30
<b>Loc12_GRC</b>	Salle de cours, devant l'enquêteur.	Présentation du cursus personnel de la locutrice (Hypokhâgne, Université, etc.).	3 min. 30
<b>Loc13_JAS</b>	Salle de cours, devant l'enquêteur.	Présentation des enseignements de la seconde année de DEUG en Sciences du Langage à l'Université de Toulouse-le Mirail.	5 min.
<b>Loc14_JUM</b>	Salle de cours, devant l'enquêteur.	Présentation du cursus personnel de la locutrice (Médecine, Sciences du Langage, etc.).	3 min.
<b>Loc15_LAD</b>	Salle de cours, devant l'enquêteur.	Présentation de ce que sont les sciences du langage.	6 min.
<b>Loc16_LAU</b>	Salle de cours, devant l'enquêteur.	Présentation du métier d'enseignant-chercheur.	8 min. 30
<b>Loc17_MAG</b>	Salle de cours, devant l'enquêteur.	Présentation du cursus de Lettres Modernes (DEUG et Licence) à l'Université de Toulouse-le Mirail.	11 min.
<b>Loc18_NES</b>	Salle de cours, devant l'enquêteur.	Présentation rapide de la Licence TAL proposée en Sciences du Langage à l'Université de Toulouse-le Mirail.	2 min. 30

<b>Loc19_NOM</b>	Salle de cours, devant l'enquêteur.	Présentation d'un travail de Maîtrise de Lettres Modernes sur deux œuvres de Barbey d'Aurévilly.	5 min.
<b>Loc20_PEA</b>	Salle de cours, devant l'enquêteur.	Présentation du cursus personnel de la locutrice.	3 min. 30
<b>Loc22_RIO</b>	Salle de cours, devant l'enquêteur.	Présentation de son activité professionnelle dans l'Animation socio-culturelle.	4 min.
<b>Loc23_ROM</b>	Salle de cours, devant l'enquêteur.	Présentation de son activité professionnelle, en particulier de son activité de recherche en morphologie.	8 min. 30

A titre d'information, voici ci-dessous le document qui a été donné aux différents locuteurs afin qu'ils préparent la présentation monologuée.

### **Document fourni aux locuteurs pour qu'ils préparent l'enregistrement :**

Veillez préparer un court exposé dans lequel vous présenterez clairement vos activités professionnelles.

- Vous décrirez ces activités en expliquant en quoi elles consistent ;
- Vous dépeindrez les différentes tâches que vous effectuez (descriptif de vos recherches et de la démarche, descriptif du cursus ou des enseignements en donnant le nom, l'objectif, ce qui est attendu, etc.) ;
- Vous tâcherez d'être le plus précis et le plus complet possible.

## **Les entretiens guidés**

Les entretiens guidés représentent environ 70 % des données du corpus (86 584 mots et environ 7 heures d'enregistrement). Ces entretiens sont le résultat d'une discussion entre un enquêteur et un locuteur, l'enquêteur menant la discussion en posant des questions au locuteur. Ils ont été enregistrés par trois enquêteurs différents auprès de 17 locuteurs. Les informations sur l'identité et les caractéristiques des locuteurs sont fournies sous la rubrique « [Identité des locuteurs](#) ». Comme la durée des entretiens varie entre 20 et 30 minutes environ, ils ont été segmentés en plusieurs sous fichiers. Pour chaque locuteur, il existe donc plusieurs séries de fichiers « entretien » (entre 10 et 15 fichiers audio, txt et TextGrid).

Les consignes fournies aux enquêteurs et dans lesquelles figurent à titre indicatif les questions qui peuvent être posées au locuteur sont fournies ci-après.



## **Consignes fournies aux enquêteurs afin d'enregistrer les entretiens :**

A titre indicatif, voici le type de questions/ entretiens qui peut être développé :

- Etes-vous né dans la région toulousaine ? Si non, est-ce que vous êtes né loin d'ici ? Où est-ce exactement ?
- Quand êtes vous arrivé à Toulouse ? Est-ce différent de votre région d'origine ? Pourquoi ? Dans votre ville/ village natal, que font les habitants ? Y a t il une école ? un lycée ? un cinéma ?
- Aimez-vous vivre à Toulouse ? Pourquoi ?
- D'où vos parents étaient-ils originaires ?
- que fait/faisait votre père ? Parlait-il une langue régionale ? laquelle ? d'autres langues ? Quelles études a-t-il faites ?
- Votre mère, travaille/travaillait-elle ? Est-elle aussi de la même région ? Parlait-elle une langue régionale ? Quelles études a-t-elle faites ?
- Décrivez la maison de vos parents. Parlez-moi de votre famille.
- Où êtes-vous allé à l'école ? Est-ce que l'école se trouvait loin de chez vous ? Quels jeux faisiez-vous pendant la récréation ? et après l'école ? Avez-vous garder un bon souvenir d'un instituteur (professeur) ? un très mauvais souvenir d'un autre ?
- Avez-vous appris une langue étrangère pendant votre scolarité ? Pouvez-vous vous débrouiller maintenant dans cette langue ?
- Après le collège/lycée, qu'avez-vous fait ?
- Quels métiers avez-vous exercés ou voulez-vous exercer ? Quel est votre métier aujourd'hui ? Pouvez-vous me donner quelques détails ? Est-ce que vous ressentez/ avez peur de la crise économique au sein de votre entreprise ? Quelles sont vos relations avec vos collègues ? Est-ce que vos conditions de travail ont évolué au cours de ces dernières années ? Est-ce que votre entreprise embauche des jeunes ? Comment sont-ils accueillis ?
- Comment vivez-vous ? Etes-vous marié(e) ? Avez-vous avez des enfants ? Quel âge ? Où vont-ils à l'école ? Quels sont leurs matières préférées ?

- Que faites-vous pendant comme activités de loisir? Regardez-vous la TV ? Quel genre d'émissions ? Sur quelle chaîne regardez-vous les nouvelles? Pourquoi? Regardez-vous les nouvelles régionales?
- Pratiquez-vous un sport ? Regardez-vous le sport à la TV? Lequel ?
- Que faites-vous pendant vos vacances? Avez-vous voyagé? Où?
- Faites-vous de la musique ? Ecoutez-vous de la musique ? Quel genre ?

# Contenu des CD :

## CD n° 1 : ACSYNT\_CD1

### Répertoire [DOC] :

Ce répertoire contient la documentation sur le corpus ACSYNT. Cette documentation se présente sous deux formats distinct : HTML ou PDF.

- ❖ ACSYNT\_DOC.html
- ❖ ACSYNT\_Documentation.pdf
- ❖ Répertoire [ACSYNT\_DOC\_FILES] dont le contenu est donné dans le tableau ci-dessous.

Nom des fichiers ou sous-répertoires	Descriptif rapide
<b>Presentation.html</b>	Page WEB présentant le corpus
<b>Contenu.html</b>	Page WEB indiquant l'architecture des CD-ROM
<b>Enregistrement.html</b>	Page WEB fournissant des informations sur les modalités d'enregistrement du corpus ACSYNT
<b>Transcription.html</b>	Page WEB fournissant des informations sur les conventions d'annotation du corpus ACSYNT
<b>Documentationling.html</b>	Page WEB présentant certaines informations relatives au corpus
<b>Locuteurs.html</b>	Page WEB donnant des informations sur les 23 locuteurs du corpus ACSYNT.
<b>Liens.html</b>	Page WEB indiquant quelques liens
<b>Image001.gif</b>	Fichiers « image » associés à la page WEB <b>ACSYNT_DOC.html</b>
<b>Image002.gif</b>	
<b>Image003.gif</b>	
<b>Image004.gif</b>	
<b>Dossier Presentation_files</b>	Ce dossier renferme les fichiers divers associés à <b>Presentation.html</b>
<b>Dossier contenu_files</b>	Ce dossier renferme les fichiers divers associés à <b>Contenu.html</b>
<b>Dossier Enregistrement_files</b>	Ce dossier renferme les fichiers divers associés à <b>Enregistrement.html</b>
<b>Dossier Transcription_files</b>	Ce dossier renferme les fichiers divers associés à <b>Transcription.html</b> , notamment les extraits sonores.
<b>Dossier Documentationling_files</b>	Ce dossier renferme les fichiers divers associés à <b>Documentationling.html</b>
<b>Dossier locuteurs_files</b>	Ce dossier renferme les fichiers divers associés à <b>locuteurs.html</b> , ainsi que les 23 pages WEB consacrées à chaque locuteur.
<b>Dossier Liens_files</b>	Ce dossier renferme les fichiers divers associés à <b>Liens.html</b>

## Répertoire [Textes] :

Ce répertoire contient 68 séries de fichiers élaborés à partir des enregistrements des trois textes lus par 23 locuteurs (soit 204 fichiers)

- ASDT1.wav, ASDT1.txt, ASDT1.TextGrid
- ASDT2.wav, ASDT2.txt, ASDT2.TextGrid
- ASDT3.wav, ASDT3.txt, ASDT3.TextGrid
  
- ASHT1.wav, ASHT1.txt, ASHT1.TextGrid
- ASHT2.wav, ASHT2.txt, ASHT2.TextGrid
- ASHT3.wav, ASHT3.txt, ASHT3.TextGrid
  
- BAVT1.wav, BAVT1.txt, BAVT1.TextGrid
- BAVT2.wav, BAVT2.txt, BAVT2.TextGrid
- BAVT3.wav, BAVT3.txt, BAVT3.TextGrid
  
- BOAT2.wav, BOAT2.txt, BOAT2.TextGrid
- BOAT3.wav, BOAT3.txt, BOAT3.TextGrid
  
- CAFT1.wav, CAFT1.txt, CAFT1.TextGrid
- CAFT2.wav, CAFT2.txt, CAFT2.TextGrid
- CAFT3.wav, CAFT3.txt, CAFT3.TextGrid
  
- CAMT1.wav, CAMT1.txt, CAMT1.TextGrid
- CAMT2.wav, CAMT2.txt, CAMT2.TextGrid
- CAMT3.wav, CAMT3.txt, CAMT3.TextGrid
  
- COAT1.wav, COAT1.txt, COAT1.TextGrid
- COAT2.wav, COAT2.txt, COAT2.TextGrid
- COAT3.wav, COAT3.txt, COAT3.TextGrid
  
- COCT1.wav, COCT1.txt, COCT1.TextGrid
- COCT2.wav, COCT2.txt, COCT2.TextGrid
- COCT3.wav, COCT3.txt, COCT3.TextGrid
  
- DAAT1.wav, DAAT1.txt, DAAT1.TextGrid
- DAAT2.wav, DAAT2.txt, DAAT2.TextGrid
- DAAT3.wav, DAAT3.txt, DAAT3.TextGrid
  
- FEDT1.wav, FEDT1.txt, FEDT1.TextGrid
- FEDT2.wav, FEDT2.txt, FEDT2.TextGrid
- FEDT3.wav, FEDT3.txt, FEDT3.TextGrid

- GOLT1.wav, GOLT1.txt, GOLT1.TextGrid
- GOLT2.wav, GOLT2.txt, GOLT2.TextGrid
- GOLT3.wav, GOLT3.txt, GOLT3.TextGrid
- GRCT1.wav, GRCT1.txt, GRCT1.TextGrid
- GRCT2.wav, GRCT2.txt, GRCT2.TextGrid
- GRCT3.wav, GRCT3.txt, GRCT3.TextGrid
- JAST1.wav, JAST1.txt, JAST1.TextGrid
- JAST2.wav, JAST2.txt, JAST2.TextGrid
- JAST3.wav, JAST3.txt, JAST3.TextGrid
- JUMT1.wav, JUMT1.txt, JUMT1.TextGrid
- JUMT2.wav, JUMT2.txt, JUMT2.TextGrid
- JUMT3.wav, JUMT3.txt, JUMT3.TextGrid
- LADT1.wav, LADT1.txt, LADT1.TextGrid
- LADT2.wav, LADT2.txt, LADT2.TextGrid
- LADT3.wav, LADT3.txt, LADT3.TextGrid
- LAUT1.wav, LAUT1.txt, LAUT1.TextGrid
- LAUT2.wav, LAUT2.txt, LAUT2.TextGrid
- LAUT3.wav, LAUT3.txt, LAUT3.TextGrid
- MAGT1.wav, MAGT1.txt, MAGT1.TextGrid
- MAGT2.wav, MAGT2.txt, MAGT2.TextGrid
- MAGT3.wav, MAGT3.txt, MAGT3.TextGrid
- NEST1.wav, NEST1.txt, NEST1.TextGrid
- NEST2.wav, NEST2.txt, NEST2.TextGrid
- NEST3.wav, NEST3.txt, NEST3.TextGrid
- NOMT1.wav, NOMT1.txt, NOMT1.TextGrid
- NOMT2.wav, NOMT2.txt, NOMT2.TextGrid
- NOMT3.wav, NOMT3.txt, NOMT3.TextGrid
- PEAT1.wav, PEAT1.txt, PEAT1.TextGrid
- PEAT2.wav, PEAT2.txt, PEAT2.TextGrid
- PEAT3.wav, PEAT3.txt, PEAT3.TextGrid
- RICT1.wav, RICT1.txt, RICT1.TextGrid
- RICT2.wav, RICT2.txt, RICT2.TextGrid
- RICT3.wav, RICT3.txt, RICT3.TextGrid
- RIOT1.wav, RIOT1.txt, RIOT1.TextGrid
- RIOT2.wav, RIOT2.txt, RIOT2.TextGrid
- RIOT3.wav, RIOT3.txt, RIOT3.TextGrid

- ROMT1.wav, ROMT1.txt, ROMT1.TextGrid
- ROMT2.wav, ROMT2.txt, ROMT2.TextGrid
- ROMT3.wav, ROMT3.txt, ROMT3.TextGrid

## Répertoire [Presentation\_Part1] :

Ce répertoire contient 13 sous répertoires (un par locuteur) dans lesquels sont enregistrées les séries de fichiers construites à partir des présentations monologuées préparées. Le détail des fichiers est donné dans le tableau ci-après :

Sous répertoire	Noms et extension des fichiers
<b>Loc1_ASD</b>	<u>Fichiers audio</u> : ASDP1.wav, ASDP2.wav <u>Fichiers TXT</u> : ASDP1.txt, ASDP2.txt <u>Fichiers TextGrid</u> : ASDP1.TextGrid, ASDP2.TextGrid
<b>Loc2_ASH</b>	<u>Fichiers audio</u> : ASHP1.wav, ASHP2.wav, ASHP3.wav, ASHP4.wav, ASHP5.wav <u>Fichiers TXT</u> : ASHP1.txt, ASHP2.txt, ASHP3.txt, ASHP4.txt, ASHP5.txt <u>Fichiers TextGrid</u> : ASHP1.TextGrid, ASHP2.TextGrid, ASHP3.TextGrid, ASHP4.TextGrid, ASHP5.TextGrid
<b>Loc3_BAV</b>	<u>Fichiers audio</u> : BAVP1.wav, BAVP2.wav, BAVP3.wav, BAVP4.wav <u>Fichiers TXT</u> : BAVP1.txt, BAVP2.txt, BAVP3.txt, BAVP4.txt <u>Fichiers TextGrid</u> : BAVP1.TextGrid, BAVP2.TextGrid, BAVP3.TextGrid, BAVP4.TextGrid
<b>Loc4_BOA</b>	<u>Fichiers audio</u> : BOAP1.wav, BOAP2.wav, BOAP3.wav, BOAP4.wav <u>Fichiers TXT</u> : BOAP1.txt, BOAP2.txt, BOAP3.txt, BOAP4.txt <u>Fichiers TextGrid</u> : BOAP1.TextGrid, BOAP2.TextGrid, BOAP3.TextGrid, BOAP4.TextGrid
<b>Loc7_COA</b>	<u>Fichiers audio</u> : COAP1.wav, COAP2.wav, COAP3.wav, COAP4.wav, COAP5.wav, COAP6.wav, COAP7.wav, COAP8.wav, COAP9.wav, COAP10.wav, COAP11.wav, COAP12.wav, COAP13.wav, COAP14.wav, COAP15.wav, COAP16.wav, COAP17.wav, COAP18.wav <u>Fichiers TXT</u> : COAP1.txt, COAP2.txt, COAP3.txt, COAP4.txt, COAP5.txt, COAP6.txt, COAP7.txt, COAP8.txt, COAP9.txt, COAP10.txt, COAP11.txt, COAP12.txt, COAP13.txt, COAP14.txt, COAP15.txt, COAP16.txt, COAP17.txt, COAP18.txt <u>Fichiers TextGrid</u> : COAP1.TextGrid, COAP2.TextGrid, COAP3.TextGrid, COAP4.TextGrid, COAP5.TextGrid, COAP6.TextGrid, COAP7.TextGrid, COAP8.TextGrid, COAP9.TextGrid, COAP10.TextGrid, COAP11.TextGrid, COAP12.TextGrid, COAP13.TextGrid, COAP14.TextGrid, COAP15.TextGrid, COAP16.TextGrid, COAP17.TextGrid, COAP18.TextGrid

<b>Loc11_GOL</b>	<u>Fichiers audio</u> : GOLP1.wav, GOLP2.wav, GOLP3.wav, GOLP4.wav, GOLP5.wav <u>Fichiers TXT</u> : GOLP1.txt, GOLP2.txt, GOLP3.txt, GOLP4.txt, GOLP5.txt <u>Fichiers TextGrid</u> : GOLP1.TextGrid, GOLP2.TextGrid, GOLP3.TextGrid, GOLP4.TextGrid, GOLP5.TextGrid
<b>Loc12_GRC</b>	<u>Fichiers audio</u> : GRCP1.wav, GRCP2.wav <u>Fichiers TXT</u> : GRCP1.txt, GRCP2.txt <u>Fichiers TextGrid</u> : GRCP1.TextGrid, GRCP2.TextGrid
<b>Loc13_JAS</b>	<u>Fichiers audio</u> : JASP1.wav, JASP2.wav, JASP3.wav <u>Fichiers TXT</u> : JASP1.txt, JASP2.txt, JASP3.txt <u>Fichiers TextGrid</u> : JASP1.TextGrid, JASP2.TextGrid, JASP3.TextGrid
<b>Loc14_JUM</b>	<u>Fichiers audio</u> : JUMP1.wav, JUMP2.wav <u>Fichiers TXT</u> : JUMP1.txt, JUMP2.txt <u>Fichiers TextGrid</u> : JUMP1.TextGrid, JUMP2.TextGrid
<b>Loc15_LAD</b>	<u>Fichiers audio</u> : LADP1.wav, LADP2.wav, LADP3.wav <u>Fichiers TXT</u> : LADP1.txt, LADP2.txt, LADP3.txt <u>Fichiers TextGrid</u> : LADP1.TextGrid, LADP2.TextGrid, LADP3.TextGrid
<b>Loc16_LAU</b>	<u>Fichiers audio</u> : LAUP1.wav, LAUP2.wav, LAUP3.wav, LAUP4.wav, LAUP5.wav <u>Fichiers TXT</u> : LAUP1.txt, LAUP2.txt, LAUP3.txt, LAUP4.txt, LAUP5.txt <u>Fichiers TextGrid</u> : LAUP1.TextGrid, LAUP2.TextGrid, LAUP3.TextGrid, LAUP4.TextGrid, LAUP5.TextGrid
<b>Loc17_MAG</b>	<u>Fichiers audio</u> : MAGP1.wav, MAGP2.wav, MAGP3.wav, MAGP4.wav, MAGP5.wav, MAGP6.wav <u>Fichiers TXT</u> : MAGP1.txt, MAGP2.txt, MAGP3.txt, MAGP4.txt, MAGP5.txt, MAGP6.txt <u>Fichiers TextGrid</u> : MAGP1.TextGrid, MAGP2.TextGrid, MAGP3.TextGrid, MAGP4.TextGrid, MAGP5.TextGrid, MAGP6.TextGrid
<b>Loc18_NES</b>	<u>Fichiers audio</u> : NESP1.wav, NESP2.wav, NESP3.wav <u>Fichiers TXT</u> : NESP1.txt, NESP2.txt, NESP3.txt <u>Fichiers TextGrid</u> : NESP1.TextGrid, NESP2.TextGrid, NESP3.TextGrid

## CD n° 2 : ACSYNT\_CD2

### Répertoire [DOC] :

Ce répertoire contient la documentation sur le corpus ACSYNT. Cette documentation se présente sous deux formats distinct : HTML ou PDF.

- ❖ ACSYNT\_DOC.html
- ❖ ACSYNT\_Documentation.pdf

- ❖ Répertoire [ACSYNT\_DOC\_FILES] dont le contenu est donné dans le tableau ci-dessous.

Nom des fichiers ou sous-répertoires	Descriptif rapide
<b>Presentation.html</b>	Page WEB présentant le corpus
<b>Contenu.html</b>	Page WEB indiquant l'architecture des CD-ROM
<b>Enregistrement.html</b>	Page WEB fournissant des informations sur les modalités d'enregistrement du corpus ACSYNT
<b>Transcription.html</b>	Page WEB fournissant des informations sur les conventions d'annotation du corpus ACSYNT
<b>Documentationling.html</b>	Page WEB présentant certaines informations relatives au corpus
<b>Locuteurs.html</b>	Page WEB donnant des informations sur les 23 locuteurs du corpus ACSYNT.
<b>Liens.html</b>	Page WEB indiquant quelques liens
<b>Image001.gif</b>	Fichiers « image » associés à la page WEB <b>ACSYNT_DOC.html</b>
<b>Image002.gif</b>	
<b>Image003.gif</b>	
<b>Image004.gif</b>	
<b>Dossier Presentation_files</b>	Ce dossier renferme les fichiers divers associés à <b>Presentation.html</b>
<b>Dossier contenu_files</b>	Ce dossier renferme les fichiers divers associés à <b>Contenu.html</b>
<b>Dossier Enregistrement_files</b>	Ce dossier renferme les fichiers divers associés à <b>Enregistrement.html</b>
<b>Dossier Transcription_files</b>	Ce dossier renferme les fichiers divers associés à <b>Transcription.html</b> , notamment les extraits sonores.
<b>Dossier Documentationling_files</b>	Ce dossier renferme les fichiers divers associés à <b>Documentationling.html</b>
<b>Dossier locuteurs_files</b>	Ce dossier renferme les fichiers divers associés à <b>locuteurs.html</b> , ainsi que les 23 pages WEB consacrées à chaque locuteur.
<b>Dossier Liens_files</b>	Ce dossier renferme les fichiers divers associés à <b>Liens.html</b>

## Répertoire [Presentation\_Part2] :

Ce répertoire contient 4 sous répertoires (un par locuteur) dans lesquels sont enregistrées les séries de fichiers construites à partir des présentations monologuées préparées. Le détail des fichiers est donné dans le tableau ci-après :

Sous répertoire	Noms et extension des fichiers
<b>Loc19_NOM</b>	<u>Fichiers audio</u> : NOMP1.wav, NOMP2.wav, NOMP3.wav <u>Fichiers TXT</u> : NOMP1.txt, NOMP2.txt, NOMP3.txt <u>Fichiers TextGrid</u> : NOMP1.TextGrid, NOMP2.TextGrid, NOMP3.TextGrid



<b>Loc20_PEA</b>	<u>Fichiers audio</u> : PEAP1.wav, PEAP2.wav, PEAP3.wav, PEAP4.wav <u>Fichiers TXT</u> : PEAP1.txt, PEAP2.txt, PEAP3.txt, PEAP4.txt <u>Fichiers TextGrid</u> : PEAP1.TextGrid, PEAP2.TextGrid, PEAP3.TextGrid, PEAP4.TextGrid
<b>Loc22_RIO</b>	<u>Fichiers audio</u> : RIOP1.wav, RIOP2.wav, RIOP3.wav, RIOP4.wav <u>Fichiers TXT</u> : RIOP1.txt, RIOP2.txt, RIOP3.txt, RIOP4.txt <u>Fichiers TextGrid</u> : RIOP1.TextGrid, RIOP2.TextGrid, RIOP3.TextGrid, RIOP4.TextGrid
<b>Loc23_ROM</b>	<u>Fichiers audio</u> : ROMP1.wav, ROMP2.wav, ROMP3.wav, ROMP4.wav, ROMP5.wav <u>Fichiers TXT</u> : ROMP1.txt, ROMP2.txt, ROMP3.txt, ROMP4.txt, ROMP5.txt <u>Fichiers TextGrid</u> : ROMP1.TextGrid, ROMP2.TextGrid, ROMP3.TextGrid, ROMP4.TextGrid, ROMP5.TextGrid

## Répertoire [Entretien\_Part1] :

Ce répertoire contient 8 sous répertoires (un par locuteur) dans lesquels sont enregistrées les séries de fichiers construites à partir des enregistrements des entretiens guidés. Le détail des fichiers est donné dans le tableau ci-après :

Sous répertoire	Noms et extension des fichiers
<b>Loc1_ASD</b>	<u>Fichiers audio</u> : ASDE1.wav, ASDE2.wav, ASDE3.wav, ASDE4.wav, ASDE5.wav, ASDE6.wav, ASDE7.wav, ASDE8.wav, ASDE9.wav, ASDE10.wav, ASDE11.wav, ASDE12.wav <u>Fichiers TXT</u> : ASDE1.txt, ASDE2.txt, ASDE3.txt, ASDE4.txt, ASDE5.txt, ASDE6.txt, ASDE7.txt, ASDE8.txt, ASDE9.txt, ASDE10.txt, ASDE11.txt, ASDE12.txt <u>Fichiers TextGrid</u> : ASDE1.TextGrid, ASDE2.TextGrid, ASDE3.TextGrid, ASDE4.TextGrid, ASDE5.TextGrid, ASDE6.TextGrid, ASDE7.TextGrid, ASDE8.TextGrid, ASDE9.TextGrid, ASDE10.TextGrid, ASDE11.TextGrid, ASDE12.TextGrid

<b>Loc2_ASH</b>	<p><u>Fichiers audio</u> : ASHE1.wav, ASHE2.wav, ASHE3.wav, ASHE4.wav, ASHE5.wav, ASHE6.wav, ASHE7.wav, ASHE8.wav, ASHE9.wav, ASHE10.wav, ASHE11.wav, ASHE12.wav</p> <p><u>Fichiers TXT</u> : ASHE1.txt, ASHE2.txt, ASHE3.txt, ASHE4.txt, ASHE5.txt, ASHE6.txt, ASHE7.txt, ASHE8.txt, ASHE9.txt, ASHE10.txt, ASHE11.txt, ASHE12.txt</p> <p><u>Fichiers TextGrid</u> : ASHE1.TextGrid, ASHE2.TextGrid, ASHE3.TextGrid, ASHE4.TextGrid, ASHE5.TextGrid, ASHE6.TextGrid, ASHE7.TextGrid, ASHE8.TextGrid, ASHE9.TextGrid, ASHE10.TextGrid, ASHE11.TextGrid, ASHE12.TextGrid</p>
<b>Loc3_BAV</b>	<p><u>Fichiers audio</u> : BAVE1.wav, BAVE2.wav, BAVE3.wav, BAVE4.wav, BAVE5.wav, BAVE6.wav, BAVE7.wav, BAVE8.wav, BAVE9.wav, BAVE10.wav, BAVE11.wav</p> <p><u>Fichiers TXT</u> : BAVE1.txt, BAVE2.txt, BAVE3.txt, BAVE4.txt, BAVE5.txt, BAVE6.txt, BAVE7.txt, BAVE8.txt, BAVE9.txt, BAVE10.txt, BAVE11.txt</p> <p><u>Fichiers TextGrid</u> : BAVE1.TextGrid, BAVE2.TextGrid, BAVE3.TextGrid, BAVE4.TextGrid, BAVE5.TextGrid, BAVE6.TextGrid, BAVE7.TextGrid, BAVE8.TextGrid, BAVE9.TextGrid, BAVE10.TextGrid, BAVE11.TextGrid</p>
<b>Loc4_BOA</b>	<p><u>Fichiers audio</u> : BOAE1.wav, BOAE2.wav, BOAE3.wav, BOAE4.wav, BOAE5.wav, BOAE6.wav, BOAE7.wav, BOAE8.wav, BOAE9.wav, BOAE10.wav, BOAE11.wav</p> <p><u>Fichiers TXT</u> : BOAE1.txt, BOAE2.txt, BOAE3.txt, BOAE4.txt, BOAE5.txt, BOAE6.txt, BOAE7.txt, BOAE8.txt, BOAE9.txt, BOAE10.txt, BOAE11.txt</p> <p><u>Fichiers TextGrid</u> : BOAE1.TextGrid, BOAE2.TextGrid, BOAE3.TextGrid, BOAE4.TextGrid, BOAE5.TextGrid, BOAE6.TextGrid, BOAE7.TextGrid, BOAE8.TextGrid, BOAE9.TextGrid, BOAE10.TextGrid, BOAE11.TextGrid</p>
<b>Loc7_COA</b>	<p><u>Fichiers audio</u> : COAE1.wav, COAE2.wav, COAE3.wav, COAE4.wav, COAE5.wav, COAE6.wav, COAE7.wav, COAE8.wav, COAE9.wav, COAE10.wav, COAE11.wav, COAE12.wav, COAE13.wav, COAE14.wav, COAE15.wav, COAE16.wav, COAE17.wav</p> <p><u>Fichiers TXT</u> : COAE1.txt, COAE2.txt, COAE3.txt, COAE4.txt, COAE5.txt, COAE6.txt, COAE7.txt, COAE8.txt, COAE9.txt, COAE10.txt, COAE11.txt, COAE12.txt, COAE13.txt, COAE14.txt, COAE15.txt, COAE16.txt, COAE17.txt</p> <p><u>Fichiers TextGrid</u> : COAE1.TextGrid, COAE2.TextGrid, COAE3.TextGrid, COAE4.TextGrid, COAE5.TextGrid, COAE6.TextGrid, COAE7.TextGrid, COAE8.TextGrid, COAE9.TextGrid, COAE10.TextGrid, COAE11.TextGrid, COAE12.TextGrid, COAE13.TextGrid, COAE14.TextGrid, COAE15.TextGrid, COAE16.TextGrid, COAE17.TextGrid</p>

<b>Loc11_GOL</b>	<p><u>Fichiers audio</u> : GOLE1.wav, GOLE2.wav, GOLE3.wav, GOLE4.wav, GOLE5.wav, GOLE6.wav, GOLE7.wav, GOLE8.wav, GOLE9.wav, GOLE10.wav</p> <p><u>Fichiers TXT</u> : GOLE1.txt, GOLE2.txt, GOLE3.txt, GOLE4.txt, GOLE5.txt, GOLE6.txt, GOLE7.txt, GOLE8.txt, GOLE9.txt, GOLE10.txt</p> <p><u>Fichiers TextGrid</u> : GOLE1.TextGrid, GOLE2.TextGrid, GOLE3.TextGrid, GOLE4.TextGrid, GOLE5.TextGrid, GOLE6.TextGrid, GOLE7.TextGrid, GOLE8.TextGrid, GOLE9.TextGrid, GOLE10.TextGrid</p>
<b>Loc12_GRC</b>	<p><u>Fichiers audio</u> : GRCE1.wav, GRCE2.wav, GRCE3.wav, GRCE4.wav, GRCE5.wav, GRCE6.wav, GRCE7.wav, GRCE8.wav, GRCE9.wav, GRCE10.wav, GRCE11.wav, GRCE12.wav, GRCE13.wav</p> <p><u>Fichiers TXT</u> : GRCE1.txt, GRCE2.txt, GRCE3.txt, GRCE4.txt, GRCE5.txt, GRCE6.txt, GRCE7.txt, GRCE8.txt, GRCE9.txt, GRCE10.txt, GRCE11.txt, GRCE12.txt, GRCE13.txt</p> <p><u>Fichiers TextGrid</u> : GRCE1.TextGrid, GRCE2.TextGrid, GRCE3.TextGrid, GRCE4.TextGrid, GRCE5.TextGrid, GRCE6.TextGrid, GRCE7.TextGrid, GRCE8.TextGrid, GRCE9.TextGrid, GRCE10.TextGrid, GRCE11.TextGrid, GRCE12.TextGrid, GRCE13.TextGrid</p>
<b>Loc13_JAS</b>	<p><u>Fichiers audio</u> : JASE1.wav, JASE2.wav, JASE3.wav, JASE4.wav, JASE5.wav, JASE6.wav, JASE7.wav, JASE8.wav, JASE9.wav, JASE10.wav, JASE11.wav, JASE12.wav, JASE13.wav</p> <p><u>Fichiers TXT</u> : JASE1.txt, JASE2.txt, JASE3.txt, JASE4.txt, JASE5.txt, JASE6.txt, JASE7.txt, JASE8.txt, JASE9.txt, JASE10.txt, JASE11.txt, JASE12.txt, JASE13.txt</p> <p><u>Fichiers TextGrid</u> : JASE1.TextGrid, JASE2.TextGrid, JASE3.TextGrid, JASE4.TextGrid, JASE5.TextGrid, JASE6.TextGrid, JASE7.TextGrid, JASE8.TextGrid, JASE9.TextGrid, JASE10.TextGrid, JASE11.TextGrid, JASE12.TextGrid, JASE13.TextGrid</p>

## CD n° 3 : ACSYNT\_CD3

### Répertoire [DOC] :

Ce répertoire contient la documentation sur le corpus ACSYNT. Cette documentation se présente sous deux formats distinct : HTML ou PDF.

- ❖ ACSYNT\_DOC.html
- ❖ ACSYNT\_Documentation.pdf
- ❖ Répertoire [ACSYNT\_DOC\_FILES] dont le contenu est donné dans le tableau ci-dessous.

Nom des fichiers ou sous-répertoires	Descriptif rapide
<b>Presentation.html</b>	Page WEB présentant le corpus
<b>Contenu.html</b>	Page WEB indiquant l'architecture des CD-ROM
<b>Enregistrement.html</b>	Page WEB fournissant des informations sur les modalités d'enregistrement du corpus ACSYNT
<b>Transcription.html</b>	Page WEB fournissant des informations sur les conventions d'annotation du corpus ACSYNT
<b>Documentationling.html</b>	Page WEB présentant certaines informations relatives au corpus
<b>Locuteurs.html</b>	Page WEB donnant des informations sur les 23 locuteurs du corpus ACSYNT.
<b>Liens.html</b>	Page WEB indiquant quelques liens
<b>Image001.gif</b>	Fichiers « image » associés à la page WEB <b>ACSYNT_DOC.html</b>
<b>Image002.gif</b>	
<b>Image003.gif</b>	
<b>Image004.gif</b>	
<b>Dossier Presentation_files</b>	Ce dossier renferme les fichiers divers associés à <b>Presentation.html</b>
<b>Dossier contenu_files</b>	Ce dossier renferme les fichiers divers associés à <b>Contenu.html</b>
<b>Dossier Enregistrement_files</b>	Ce dossier renferme les fichiers divers associés à <b>Enregistrement.html</b>
<b>Dossier Transcription_files</b>	Ce dossier renferme les fichiers divers associés à <b>Transcription.html</b> , notamment les extraits sonores.
<b>Dossier Documentationling_files</b>	Ce dossier renferme les fichiers divers associés à <b>Documentationling.html</b>
<b>Dossier locuteurs_files</b>	Ce dossier renferme les fichiers divers associés à <b>locuteurs.html</b> , ainsi que les 23 pages WEB consacrées à chaque locuteur.
<b>Dossier Liens_files</b>	Ce dossier renferme les fichiers divers associés à <b>Liens.html</b>

## Répertoire [Entretien\_Part2] :

Ce répertoire contient 9 sous répertoires (un par locuteur) dans lesquels sont enregistrées les séries de fichiers construites à partir des enregistrements des entretiens guidés. Le détail des fichiers est donné dans le tableau ci-après :

Sous répertoire	Noms et extension des fichiers
<b>Loc14_JUM</b>	<p><u>Fichiers audio</u> : JUME1.wav, JUME2.wav, JUME3.wav, JUME4.wav, JUME5.wav, JUME6.wav, JUME7.wav, JUME8.wav, JUME9.wav, JUME10.wav, JUME11.wav, JUME12.wav, JUME13.wav, JUME14.wav</p> <p><u>Fichiers TXT</u> : JUME1.txt, JUME2.txt, JUME3.txt, JUME4.txt, JUME5.txt, JUME6.txt, JUME7.txt, JUME8.txt, JUME9.txt, JUME10.txt, JUME11.txt, JUME12.txt, JUME13.wav, JUME14.wav</p> <p><u>Fichiers TextGrid</u> : JUME1.TextGrid, JUME2.TextGrid, JUME3.TextGrid, JUME4.TextGrid, JUME5.TextGrid, JUME6.TextGrid, JUME7.TextGrid, JUME8.TextGrid, JUME9.TextGrid, JUME10.TextGrid, JUME11.TextGrid, JUME12.TextGrid, JUME13.TextGrid, JUME14.TextGrid</p>
<b>Loc15_LAD</b>	<p><u>Fichiers audio</u> : LADE1.wav, LADE2.wav, LADE3.wav, LADE4.wav, LADE5.wav, LADE6.wav, LADE7.wav, LADE8.wav, LADE9.wav, LADE10.wav, LADE11.wav, LADE12.wav, LADE13.wav</p> <p><u>Fichiers TXT</u> : LADE1.txt, LADE2.txt, LADE3.txt, LADE4.txt, LADE5.txt, LADE6.txt, LADE7.txt, LADE8.txt, LADE9.txt, LADE10.txt, LADE11.txt, LADE12.txt, LADE13.txt</p> <p><u>Fichiers TextGrid</u> : LADE1.TextGrid, LADE2.TextGrid, LADE3.TextGrid, LADE4.TextGrid, LADE5.TextGrid, LADE6.TextGrid, LADE7.TextGrid, LADE8.TextGrid, LADE9.TextGrid, LADE10.TextGrid, LADE11.TextGrid, LADE12.TextGrid, LADE13.TextGrid</p>
<b>Loc16_LAU</b>	<p><u>Fichiers audio</u> : LAUE1.wav, LAUE2.wav, LAUE3.wav, LAUE4.wav, LAUE5.wav, LAUE6.wav, LAUE7.wav, LAUE8.wav, LAUE9.wav, LAUE10.wav, LAUE11.wav, LAUE12.wav, LAUE13.wav</p> <p><u>Fichiers TXT</u> : LAUE1.txt, LAUE2.txt, LAUE3.txt, LAUE4.txt, LAUE5.txt, LAUE6.txt, LAUE7.txt, LAUE8.txt, LAUE9.txt, LAUE10.txt, LAUE11.txt, LAUE12.txt, LAUE13.txt</p> <p><u>Fichiers TextGrid</u> : LAUE1.TextGrid, LAUE2.TextGrid, LAUE3.TextGrid, LAUE4.TextGrid, LAUE5.TextGrid, LAUE6.TextGrid, LAUE7.TextGrid, LAUE8.TextGrid, LAUE9.TextGrid, LAUE10.TextGrid, LAUE11.TextGrid, LAUE12.TextGrid, LAUE13.TextGrid</p>

<p><b>Loc17_MAG</b></p>	<p><u>Fichiers audio</u> : MAGE1.wav, MAGE2.wav, MAGE3.wav, MAGE4.wav, MAGE5.wav, MAGE6.wav, MAGE7.wav, MAGE8.wav, MAGE9.wav, MAGE10.wav, MAGE11.wav, MAGE12.wav, MAGE13.wav, MAGE14.wav, MAGE15.wav, MAGE16.wav</p> <p><u>Fichiers TXT</u> : MAGE1.txt, MAGE2.txt, MAGE3.txt, MAGE4.txt, MAGE5.txt, MAGE6.txt, MAGE7.txt, MAGE8.txt, MAGE9.txt, MAGE10.txt, MAGE11.txt, MAGE12.txt, MAGE13.txt, MAGE14.txt, MAGE15.txt, MAGE16.txt</p> <p><u>Fichiers TextGrid</u> : MAGE1.TextGrid, MAGE2.TextGrid, MAGE3.TextGrid, MAGE4.TextGrid, MAGE5.TextGrid, MAGE6.TextGrid, MAGE7.TextGrid, MAGE8.TextGrid, MAGE9.TextGrid, MAGE10.TextGrid, MAGE11.TextGrid, MAGE12.TextGrid, MAGE13.TextGrid, MAGE14.TextGrid, MAGE15.TextGrid, MAGE16.TextGrid</p>
<p><b>Loc18_NES</b></p>	<p><u>Fichiers audio</u> : NESE1.wav, NESE2.wav, NESE3.wav, NESE4.wav, NESE5.wav, NESE6.wav, NESE7.wav, NESE8.wav, NESE9.wav, NESE10.wav, NESE11.wav</p> <p><u>Fichiers TXT</u> : NESE1.txt, NESE2.txt, NESE3.txt, NESE4.txt, NESE5.txt, NESE6.txt, NESE7.txt, NESE8.txt, NESE9.txt, NESE10.txt, NESE11.txt</p> <p><u>Fichiers TextGrid</u> : NESE1.TextGrid, NESE2.TextGrid, NESE3.TextGrid, NESE4.TextGrid, NESE5.TextGrid, NESE6.TextGrid, NESE7.TextGrid, NESE8.TextGrid, NESE9.TextGrid, NESE10.TextGrid, NESE11.TextGrid</p>
<p><b>Loc19_NOM</b></p>	<p><u>Fichiers audio</u> : NOME1.wav, NOME2.wav, NOME3.wav, NOME4.wav, NOME5.wav, NOME6.wav, NOME7.wav, NOME8.wav, NOME9.wav, NOME10.wav, NOME11.wav</p> <p><u>Fichiers TXT</u> : NOME1.txt, NOME2.txt, NOME3.txt, NOME4.txt, NOME5.txt, NOME6.txt, NOME7.txt, NOME8.txt, NOME9.txt, NOME10.txt, NOME11.txt</p> <p><u>Fichiers TextGrid</u> : NOME1.TextGrid, NOME2.TextGrid, NOME3.TextGrid, NOME4.TextGrid, NOME5.TextGrid, NOME6.TextGrid, NOME7.TextGrid, NOME8.TextGrid, NOME9.TextGrid, NOME10.TextGrid, NOME11.TextGrid</p>

<b>Loc20_PEA</b>	<p><u>Fichiers audio</u> : PEAE1.wav, PEAE2.wav, PEAE3.wav, PEAE4.wav, PEAE5.wav, PEAE6.wav, PEAE7.wav, PEAE8.wav, PEAE9.wav, PEAE10.wav, PEAE11.wav, PEAE12.wav, PEAE13.wav, PEAE14.wav</p> <p><u>Fichiers TXT</u> : PEAE1.txt, PEAE2.txt, PEAE3.txt, PEAE4.txt, PEAE5.txt, PEAE6.txt, PEAE7.txt, PEAE8.txt, PEAE9.txt, PEAE10.txt, PEAE11.txt, PEAE12.txt, PEAE13.txt, PEAE14.txt</p> <p><u>Fichiers TextGrid</u> : PEAE1.TextGrid, PEAE2.TextGrid, PEAE3.TextGrid, PEAE4.TextGrid, PEAE5.TextGrid, PEAE6.TextGrid, PEAE7.TextGrid, PEAE8.TextGrid, PEAE9.TextGrid, PEAE10.TextGrid, PEAE11.TextGrid, PEAE12.TextGrid, PEAE13.TextGrid, PEAE14.TextGrid</p>
<b>Loc22_RIO</b>	<p><u>Fichiers audio</u> : RIOE1.wav, RIOE2.wav, RIOE3.wav, RIOE4.wav, RIOE5.wav, RIOE6.wav, RIOE7.wav, RIOE8.wav, RIOE9.wav, RIOE10.wav, RIOE11.wav, RIOE12.wav</p> <p><u>Fichiers TXT</u> : RIOE1.txt, RIOE2.txt, RIOE3.txt, RIOE4.txt, RIOE5.txt, RIOE6.txt, RIOE7.txt, RIOE8.txt, RIOE9.txt, RIOE10.txt, RIOE11.txt, RIOE12.txt</p> <p><u>Fichiers TextGrid</u> : RIOE1.TextGrid, RIOE2.TextGrid, RIOE3.TextGrid, RIOE4.TextGrid, RIOE5.TextGrid, RIOE6.TextGrid, RIOE7.TextGrid, RIOE8.TextGrid, RIOE9.TextGrid, RIOE10.TextGrid, RIOE11.TextGrid, RIOE12.TextGrid</p>
<b>Loc23_ROM</b>	<p><u>Fichiers audio</u> : ROME1.wav, ROME2.wav, ROME3.wav, ROME4.wav, ROME5.wav, ROME6.wav, ROME7.wav, ROME8.wav, ROME9.wav, ROME10.wav, ROME11.wav, ROME12.wav, ROME13.wav, ROME14.wav, ROME15.wav, ROME16.wav</p> <p><u>Fichiers TXT</u> : ROME1.txt, ROME2.txt, ROME3.txt, ROME4.txt, ROME5.txt, ROME6.txt, ROME7.txt, ROME8.txt, ROME9.txt, ROME10.txt, ROME11.txt, ROME12.txt, ROME13.txt, ROME14.txt, ROME15.txt, ROME16.txt</p> <p><u>Fichiers TextGrid</u> : ROME1.TextGrid, ROME2.TextGrid, ROME3.TextGrid, ROME4.TextGrid, ROME5.TextGrid, ROME6.TextGrid, ROME7.TextGrid, ROME8.TextGrid, ROME9.TextGrid, ROME10.TextGrid, ROME11.TextGrid, ROME12.TextGrid, ROME13.TextGrid, ROME14.TextGrid, ROME15.TextGrid, ROME16.TextGrid</p>

# Enregistrement et numérisation des données

## Enregistrement des données

### Généralités

La totalité des données du corpus a été enregistrée grâce à la participation de 23 locuteurs de la région Midi-Pyrénées et de 3 enquêteurs. Les enregistrements ont été effectués dans l'enceinte de l'Université de Toulouse – le Mirail, plus précisément dans des salles de cours ou dans des bureaux d'enseignants. Bien que des précautions aient été prises afin de minimiser les bruits ambiants (les ordinateurs ont été éteints, les téléphones débranchés, etc.), des bruits de fonds apparaissent dans les enregistrements, surtout que les salles ne sont pas toutes insonorisées. Dans la majorité des cas, deux personnes étaient présentes dans la salle lors de l'enregistrement : un enquêteur (ou un chercheur participant au projet de constitution du corpus) et un locuteur. Il existe cependant une exception : l'enregistrement de la présentation monologuée préparée pour le locuteur n°7 COA. Dans ce dernier cas, en effet, l'enregistrement a été effectué dans une salle où se déroulaient les Journées d'Etudes Scientifiques.

Les enregistrements se sont généralement déroulés en trois temps distincts :

- ❖ dans un premier temps, l'enquêteur recherchait des locuteurs auxquels il remettait les documents nécessaires à la préparation des enregistrements, à savoir les textes à lire (T1, T2 et T3) et les consignes pour la préparation des présentations monologuées préparées (cf. « [Présentation du corpus](#) ») ;
- ❖ lors de la séance d'enregistrement, il était demandé à chaque locuteur de remplir une fiche signalétique et une autorisation. Cette tâche avait un double objectif :
  - obtenir des informations sur l'identité du locuteur (origines sociales et géographiques, âge, etc.) ;
  - se protéger juridiquement et permettre une utilisation des données enregistrées.Un exemplaire de chaque document est proposé dans la section suivante.
- ❖ dans un troisième temps, les enregistrements étaient effectués. Chaque locuteur avait à remplir trois tâches distinctes, à savoir :
  - lire les trois textes T1, T2 et T3 (cf. « [Présentation du corpus](#) ») ;
  - effectuer un entretien guidé, c'est à dire répondre à diverses questions qui lui étaient posées par l'enquêteur (cf. « [Présentation du corpus](#) ») ;
  - faire une présentation monologuée qui a été préalablement préparée par le locuteur grâce aux consignes qui lui ont été remises (cf. « [Présentation du corpus](#) »).



Afin de mettre en confiance les locuteurs, les tâches ont été généralement présentées dans un ordre de difficulté croissante : l'entretien guidé en premier, suivi de la tâche de lecture, et de la présentation monologuée préparée.

### Formulaires remplis par chaque locuteur

Chaque locuteur a rempli une fiche signalétique dans laquelle il fournit plusieurs informations sur ces origines sociales et géographiques, sa famille, etc. Ces documents nous ont servi pour construire les fiches d'identité des locuteurs (cf. [« Identité des locuteurs »](#)). A titre indicatif, voici un exemple de fiche signalétique.

Fiche signalétique
Prénom, Nom :
Date de naissance :
Lieu de naissance :
Domiciles successifs (en nombre d'années)
Domicile actuel :
Professions successives :
Profession actuelle :
Situation familiale
Marié(e)
enfants, âge, scolarisation
Père de l'informateur, âge approximatif
lieu d'origine
profession
études, langues parlées (étrangères ou régionales)
Mère de l'informateur, âge approximatif
lieu d'origine
profession
études, langues parlées (étrangères ou régionales)

Études du locuteur (préciser jusqu'à quel âge et quel type d'études)

Langues parlées par le locuteur :

Activités culturelles, loisirs, voyages

Renseignements complémentaires (partie remplie par l'enquêteur) :

Date(s) de l'enregistrement, durée(s) approximative(s).

Lieu des enregistrement(s)

En outre, afin de permettre la transcription des enregistrements et leur diffusion, chaque locuteur a rempli une autorisation dont voici une reproduction.

## **Autorisation**

Les études sur le français parlé spontané sont faites dans les universités en France et à l'étranger. Elles sont précieuses pour la connaissance du langage et pour le développement des industries de la langue. Il est fondamental de faire des progrès dans ce domaine, d'enregistrer des situations spontanées, de les transcrire, de les analyser et de diffuser certains enregistrements ou transcription, tout en préservant votre anonymat.

Afin de pouvoir utiliser et diffuser les données collectées dans le cadre du projet ACI JEUNE CHERCHEUR 2002 (projet n°67031 / « Etude variationniste de l'accentuation et des interfaces Phonologie/ Syntaxe / Sémantique en français contemporain : Descriptions et Modélisations »), nous vous demandons une autorisation officielle.

Je  
soussigné(e) \_\_\_\_\_ autorise :

1. L'enregistrement par un membre du projet d'ACI « Etude variationniste de l'accentuation et des interfaces Phonologie/ Syntaxe / Sémantique en français contemporain : Descriptions et Modélisations » des données suivantes :

- a. La lecture des trois textes
- b. L'entretien guidé entre l'enquêteur et moi-même
- c. La présentation de mes activités professionnelles
- d. Un extrait de cours de Travaux dirigés
- e. Un extrait de cours magistral
- f. Une conversation entre un autre collègue et moi-même.

2. La transcription des enregistrements ci-dessus mentionnés par un membre du projet d'ACI « Etude variationniste de l'accentuation et des interfaces Phonologie/ Syntaxe / Sémantique en français contemporain : Descriptions et Modélisations »

3. L'édition à des fins scientifiques ou pédagogiques des transcriptions ou des données sonores enregistrées , en préservant mon anonymat.

4. La diffusion et la représentation des transcriptions et des données sonores enregistrées, mais en préservant mon anonymat.

Date :

Signature :

NOM :

Adresse :

## Modalités d'enregistrement des différents types de documents

Les modalités d'enregistrement ont été différentes selon qu'il s'agissait de l'entretien guidé, de la lecture des textes ou de la présentation monologuée préparée. Pour les entretiens, les enquêteurs avaient accès à un document sur lequel figuraient les questions pouvant être posées à chaque locuteur (une présentation en est proposée dans la rubrique « [Présentation du corpus](#) »). Les questions portent principalement sur la vie du locuteur, son lieu d'habitation, son enfance, sa famille, ses loisirs. Elles ont été choisies dans la mesure où elles étaient susceptibles de conduire chaque locuteur à parler librement en développant ses réponses. D'une façon générale, les entretiens se déroulaient de façon à ce que le locuteur se sente à l'aise et réponde longuement aux questions, l'enquêteur n'intervenant que très peu. L'enregistrement des entretiens a généralement été fait en premier dans la mesure où il permettait de « mettre à l'aise » le locuteur. Ont ensuite été enregistrées les tâches de lecture de textes et de présentation préparée. Pour la lecture de textes, chaque locuteur a eu les textes à l'avance et a pu en prendre connaissance. En outre, lui seul indiquait s'il se sentait prêt à commencer l'enregistrement de tel ou tel texte (T1, T2, T3). Il avait reçu pour consigne, qu'en cas d'erreur ou de « bafouillage », il devait simplement reprendre l'énoncé ou le mot erroné comme il l'entendait. Pour la présentation monologuée préparée, les locuteurs avaient reçu un document dans lequel était indiqué ce qui était attendu et sur quel thème

ils pouvaient construire leur présentation (cf. rubrique « [Présentation du corpus](#) »). Au moment de l'enregistrement, ils indiquaient à l'enquêteur quand ils souhaitaient commencer leur exposé.

## Support et matériel

Les enregistrements ont été faits :

- ❖ soit sur support numérique (DAT) en utilisant deux configurations « matériel » distinctes (présentées ci-après) et des cassettes DAT « **Maxell** », DM 120 minutes ;
- ❖ soit sur support analogique en utilisant la configuration « matériel » présentée ci-après et des K7 audio « **TDK** », modèle SA-X 90 (type II)

### Configurations « matériel » numériques :

Pour les enregistrements sur support numérique, une présentation des deux configurations utilisées est faite ci-après.

#### ❖ Configuration DAT 1 :

Pour certains enregistrements, le matériel utilisé consiste en :

- une platine DAT portable **TEAC DAP20** ;
- un microphone avec prise et câble XLR **AKG C1000** en mode hyper cardioïde.

Dans ce cas, le microphone était dirigé vers le locuteur avec une perche, tandis que l'enquêteur se trouvait à l'opposé, séparé par une table. Dans cette configuration, les propos des enquêteurs ne sont donc pas toujours clairement audibles.

#### ❖ Configuration DAT 2 :

Pour certains enregistrements, notamment lors des tâches de lecture de textes, le matériel utilisé a consisté en :

- un lecteur-enregistreur DAT portable **SONY DTC-D8** ;
- un micro-cravate omnidirectionnel à électret **VIVANCO EM 216** avec une prise Jack 3,5.

Avec cette configuration, le micro-cravate a été fixé sur le locuteur, mais comme l'enquêteur se trouvait à une distance limitée de ce dernier (moins de 1 mètre), les productions des deux protagonistes sont correctement enregistrées et audibles.

### Configuration « matériel » analogique

Pour les enregistrements analogiques, le matériel utilisé est présenté ci-dessous.

#### ❖ Configuration ANA :

Le matériel utilisé pour les enregistrements sur support analogique consiste en :

- un magnétophone K7 de reportage monophonique **MARANTZ PMD 222** ;
- un microphone avec prise et câble XLR **AKG C1000S** en mode hyper cardioïde.

Pour effectuer les enregistrements avec cette configuration « matériel », le microphone a toujours été fixé sur un pied de micro situé sur une table à distance équivalente de l'enquêteur et du locuteur.

Comme les enregistrements n'ont pas tous été faits avec le même matériel et sur les mêmes types de support, le tableau ci-dessous indique la configuration retenue pour chaque type d'enregistrement distinct.

	<b>ASD</b>	<b>ASH</b>	<b>BAV</b>	<b>BOA</b>	<b>CAF</b>	<b>CAM</b>	<b>COA</b>	<b>COC</b>	<b>DAA</b>	<b>FED</b>	<b>GOL</b>
<b>Textes</b>	DAT 2	ANA	DAT 1	DAT 2	DAT 2	ANA	DAT 2	DAT 1	ANA	DAT 2	DAT 2
<b>Entretien</b>	DAT 2	ANA	DAT 1	DAT 2	N / A	N / A	ANA	N / A	N / A	N / A	DAT 2
<b>Présentation</b>	DAT 2	ANA	DAT 1	DAT 2			ANA				DAT 2

	<b>GRC</b>	<b>JAS</b>	<b>LAD</b>	<b>LAU</b>	<b>MAG</b>	<b>NES</b>	<b>NO M</b>	<b>PEA</b>	<b>RIC</b>	<b>RIO</b>	<b>ROM</b>
<b>Textes</b>	ANA	DAT 2	DAT 2	DAT 1	ANA	DAT 2	DAT 2	ANA	DAT 2	ANA	ANA
<b>Entretien</b>	ANA	DAT 2	DAT 2	DAT 1	ANA	DAT 2	DAT 2	ANA	N / A	ANA	ANA
<b>Présentation</b>	ANA	DAT 2	DAT 2	DAT 1	ANA	DAT 2	DAT 2	ANA		ANA	ANA

(PS : N/A signifie que les données concernées ne sont pas disponibles)

## Numérisation et préparation des données :

Les données ont été traitées différemment selon que les enregistrements ont été effectués sur support numérique (cf. configuration DAT1 et DAT2) ou sur support analogique (configuration ANA).

- ❖ les données enregistrées sur DAT ont directement été transférées sur un ordinateur PC équipé d'une carte son **SoundBlaster Platinum Audigy EX**. Cela s'est fait à partir du lecteur-enregistreur portable **SONY DTC-D8** à l'aide d'un câble de transfert de données optique. Comme les données ainsi transférées étaient au format stéréophonique, avec une fréquence d'échantillonnage de 44,1 kHz, elles ont été converties pour avoir les caractéristiques suivantes :

- fréquence d'échantillonnage : 22 kHz ;
- 16 bits ;

- mono ;
  - format : wav.
- ❖ les données enregistrées en analogique ont été numérisées sur un ordinateur PC équipé d'une carte son **SoundBlaster Platinum Audigy EX**. Pour ce faire, les enregistrements K7 ont été transférés d'un lecteur de K7 **Thomson (TM 9136)** vers l'ordinateur à l'aide d'une câble de transfert audio (prise Jack 3,5 / 6,35). La numérisation s'est faite de façon à ce que les données aient les caractéristiques suivantes :
- fréquence d'échantillonnage : 22 kHz ;
  - 16 bits ;
  - mono ;
  - format : wav.

Une fois que l'ensemble des données a été transféré sur l'ordinateur et converti pour avoir un format de 22 kHz 16 bits mono, les fichiers volumineux ont été sous segmentés en petits fichiers d'une durée n'excédant pas trois minutes. Cette segmentation a surtout été effectuée pour les fichiers contenant les enregistrements des entretiens guidés et des présentations monologuées préparées (cf. « [Présentation du corpus](#) »). La segmentation a été effectuée sous PRAAT manuellement lors de pauses, et si possible de fin de tours de parole.

En outre, afin de préserver l'anonymat des locuteurs, le signal acoustique correspondant à l'élocution de certains noms propres a été filtré (Filtre passe-bas Butterworth d'ordre 10 – fréquence de coupure : 660 Hz) de façon à rendre inaudible le contenu segmental de l'énoncé, tout en conservant la mélodie (F0). Ce filtrage a été effectué à l'aide du logiciel « Cool Edit Pro ». Dans ce cas, la transcription des noms propres dans les fichiers TextGrid et TXT est codée de façon à conserver l'accès à leurs caractéristiques prosodiques (nombre de syllabes et nature des syllabes). Voici quelques exemples de codage.

### Exemples de codage :

- ❖ « Paris » prononcé [paRi] sera codé **CV.CV** ; (codage SAMPA, cf. <http://www.phon.ucl.ac.uk/home/sampa/french.htm>)
- ❖ « Espelette » prononcé [Esp@lEt] sera codé **VC.CV.CVC** ;
- ❖ « Marseille » prononcé [maRsEj] sera codé **CVC.CVC**.

# Transcription des données : Modalités et Conventions

Les contenus des fichiers audio ont été transcrits orthographiquement, et cela quelle que soit la tâche concernée : lecture de texte, présentation monologuée ou entretien guidé. La transcription orthographique est proposée sous deux formats distincts (TextGrid et texte brut) et permet de savoir exactement ce qui est contenu dans les fichiers audio, notamment pour les lectures de textes en cas d'hésitations, de reprises, etc. Dans ce document, nous allons présenter comment ont été effectuées ces transcriptions, et quelles conventions ont été retenues.

## Modalités retenues pour la transcription des données :

Deux points importants doivent être clarifiés :

- ❖ Qui a effectué les transcriptions orthographiques ?
- ❖ Comment ces transcriptions ont été faites, avec quels outils ?

### Qui a transcrit les données ?

En ce qui concerne le premier point, il est important de noter que toutes les transcriptions ont été effectuées par deux personnes : une première personne avait pour tâche de transcrire orthographiquement le contenu des enregistrements en respectant des conventions particulières, tandis qu'une seconde avait pour but de vérifier la transcription, c'est à dire :

- ❖ de vérifier la correction orthographique et grammaticale de la transcription ;
- ❖ de contrôler l'adéquation entre ce qui est enregistré et la transcription proposée ;
- ❖ de veiller à ce que les conventions soient respectées.

Les transcriptions ont été faites par des étudiants de 2<sup>ème</sup> cycle et des chercheurs qui avaient reçu une formation préalable dans le but :

- ❖ de se familiariser avec l'outil utilisé pour transcrire, à savoir le logiciel [PRAAT](#) développé par Paul Boersma. Ce logiciel permet de transcrire et d'annoter des fichiers audio, les annotations étant alignées sur le signal. Il a été retenu pour plusieurs raisons, parmi lesquelles nous pouvons citer le fait que :
  - PRAAT est un logiciel d'analyse de la parole multi-plateforme, facilement téléchargeable et bien documenté ;
  - la visualisation des différentes 'tires' d'annotation (étiquetage morpho-syntaxique, annotation prosodique, transcription orthographique) est possible, moyennant le développement de quelques utilitaires facilitant les conversions des fichiers ;
  - l'alignement sur le signal et son écoute sont toujours possibles.

- ❖ de connaître les conventions retenues pour les transcriptions, et cela quel que soit le type de document (lecture de texte, présentation monologuée préparée ou entretien guidé).

A l'issue de la formation, un document reprenant les points essentiels a été remis à chaque transcripateur.

Les transcriptions effectuées par ces transcripateurs ont ensuite été intégralement revues par un expert (chercheur post-doctorant ayant fait de nombreux travaux de transcription orthographique de corpus oraux dans l'ensemble de sa formation universitaire et doctorale). Cet expert a revu l'ensemble des transcriptions qu'il a corrigées si nécessaire, et a fourni après avoir revu chaque locuteur une synthèse dans laquelle il indiquait ce qui avait été corrigé, quels noms propres avaient été utilisés et quelles difficultés il avait rencontrées.

### Comment les données ont été transcrites ? quels choix ont été faits ?

Pour effectuer la transcription orthographique des données nous avons décidé de nous conformer autant que possible aux recommandations exprimées dans les documents de travail du groupe d'experts européens EAGLES (cf. EAGLES (1996a) *Recommendations on corpus encoding*. EAG-TCWG-CES/R-F. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale ; EAGLES (1996b) *Recommendations for the morphosyntactic annotation of corpora*. EAG-TCWG-MAC/R. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale ; EAGLES (1996c) *Preliminary recommendations for the syntactic annotation of corpora*. EAG-TCWG-SASG1/P-B. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale ; EAGLES (1996d) *Preliminary recommendations on corpus typology*. EAG-TCWG-CTYP/P. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale ; EAGLES (1996e) *Preliminary recommendations on spoken texts*. EAG-TCWG-SPT/P. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale ; EAGLES (1996f) *Preliminary recommendations on text typology*. EAG-TCWG-TTYP/P. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale). Nous justifions ce choix par le fait que nous étions soucieux de rendre ces données échangeables et exploitables par tous.

Comme une transcription orthographique a pour fonction d'encoder symboliquement toutes les informations comprises dans le signal sonore, elle doit fournir différents types d'informations qui ont été clairement listées dans le cadre des travaux du TEI et du CES (EAGLES). Dans les transcriptions proposées, nous avons respecté cette liste. En plus du contenu des énoncés produits par les différents interlocuteurs, plusieurs informations ont été encodées. Les voici, classées en plusieurs catégories :

- ❖ les éléments du contenu, à savoir :
  - les événements vocaux semi-lexicaux, c'est à dire les pauses remplies (filled pauses), les hésitations. Pour ces événements, l'orthographe retenue doit être unifiée sur l'ensemble du corpus ;
  - les événements vocaux non lexicaux, c'est à dire les éternuements, les rires, la toux, les soupirs, etc. ;
  - les événements sonores non communicatifs et non vocaux, c'est à dire les bruits produits par les auditeurs, les claquements de porte, le téléphone, les sonneries, etc.



- ❖ les éléments relatifs aux locuteurs en présence :
  - l'identité des locuteurs doit être indiquée dans une en-tête. La description des locuteurs doit comporter leur nom (et la façon de le coder), mais également leur âge, leur origine géographique, etc. ;
  - les tours de parole ou changement de locuteur ;
  - les chevauchements.
- ❖ les éléments relatifs aux productions du locuteur, à savoir :
  - les omissions lors de la lecture des textes ;
  - les réparations ou reprises, les répétitions, etc. ;
  - les fragments de mots, si ces derniers ne sont pas intégralement prononcés ;
  - les fragments inintelligibles. Dans ce cas, une distinction entre 'devinés', 'incertains' et inintelligibles' peut être faite.

Par ailleurs, le TEI et le CES (ou les recommandations du groupe d'experts EAGLES) ont formulé des recommandations sur la façon d'encoder orthographiquement les énoncés produits et les éléments que nous venons de présenter. Les voici :

- les mots produits (y compris les onomatopées) sont transcrits en utilisant l'orthographe standard ;
- seules les formes contractées admises dans le dictionnaire sont utilisées ;
- les frontières de phrases ou d'énoncés sont indiquées par des ponctuations fortes (point, point d'interrogation et point d'exclamation). Ces signes de ponctuation sont d'ailleurs les seuls utilisés ;
- aucune ponctuation (virgule et point virgule) n'est utilisée à l'intérieur des énoncés ;
- les citations de textes lus ou le discours direct est indiqué par des guillemets simples ;
- les sigles et acronymes sont transcrits tels qu'ils ont été produits (lus, épelés ou sous forme intégrale 'parti socialiste') ;
- les chiffres sont transcrits orthographiquement en tenant compte des variantes (*nonante* ou *quatre-vingt dix*).

Nous allons suivre ces recommandations et nous utiliserons les conventions typographiques présentées dans le paragraphe suivant à l'aide d'exemples précis.

Par ailleurs, le logiciel PRAAT a été utilisé pour effectuer les transcriptions orthographiques.

## Conventions de transcription

Comme nous venons de le mentionner, les transcriptions orthographiques du corpus ACSYNT ont été faites en suivant scrupuleusement les recommandations du groupe EAGLES. Ceci étant, des choix ont été faits, notamment :

- ❖ dans l'alignement sous PRAAT entre transcription orthographique et signal sonore ;

- ❖ dans la façon de noter les tours de parole et les chevauchements ;
- ❖ dans l'usage de l'orthographe et la typographie.

### L'alignement de la transcription sur le signal :

Les transcriptions ont été effectuées sous PRAAT, dans la mesure où ce logiciel :

- en permettant des écoutes réitérées de portions de signal, facilite la tâche des transcripteur ;
- n'impose aucun schéma préalable pour les annotations, mais permet de visualiser différentes annotations en parallèle (orthographe, étiquetage grammatical, analyse prosodique, etc.).

La transcription orthographique associée à chaque fichier son est saisie par chaque transcripteur sous PRAAT dans une tire créée par le transcripteur et nommée « **Orthographe** ». En plus de cette tire « **Orthographe** », tout transcripteur crée une tire nommée « **commentaires** », dans laquelle il indique les difficultés rencontrées, la prononciation réalisée s'il la juge intéressante, etc.

Les intervalles ou segmentations à effectuer varient en fonction des types de document à transcrire. Voici en gros ce qui peut être dit :

- ❖ **pour le texte** : une segmentation et un alignement au niveau de chaque phrase du texte sont effectués. La transcription correspondant à l'extrait de texte lu dans l'intervalle est reprise dans la version électronique du texte à lire, mais sont insérées, si nécessaire, les informations suivantes : répétitions, reprises ou mots incomplets, hésitations, etc. ;
- ❖ **pour le monologue** : une segmentation en énoncés (ou en unités perçues comme des phrases) est faite. La transcription associée à chaque intervalle doit respecter les recommandations présentées précédemment (ponctuation forte à la fin de l'énoncé, mais aucun autre signe de ponctuation, notation des bruits, des hésitations, des faux départs, respect de l'orthographe standard, etc.) ;
- ❖ **pour l'entretien** : une segmentation en tours de parole est faite. Ceci étant, si un tour de parole est très long, il peut être segmenté en plusieurs intervalles plus petits contenant un ou deux énoncés. Au début de chaque changement de locuteur, il faut noter le locuteur qui prend la parole avant de transcrire. L'enquêteur est indiqué **E** et le locuteur **L**.

**Exemple :**

<b>E</b> : Est-ce que vous habitez à Toulouse depuis longtemps ?	<b>L</b> : Non, je suis arrivé il y a deux ou trois ans pour commencer mes études.	<b>E</b> : Vous étiez où avant ?
--	--	----------------------------------

## Changement de locuteurs, tours de parole et chevauchements :

Tous les tours de parole ou changement de locuteurs sont indiqués. Sous PRAAT, un nouvel intervalle est créé à chaque changement de locuteur. Celui qui prend la parole est alors indiqué au début de l'intervalle de la façon suivante :

E : désigne l'enquêteur. Typographiquement, il s'agit d'un « E » immédiatement suivi de deux points ;

L : désigne le locuteur cible. Il s'agit d'un « L » suivi de deux points.

### Exemple :

<b>E:</b> [claquement de langue] Et euh vous êtes arrivée quand à Toulouse quand en fait ?	<b>L:</b> euh bè pour la première année de DEUG donc c'était il y a deux ans et demi.	<b>E:</b> Et vous retournez souvent à Albi ou bien comment ça se passe en fait ?
--	---	--



En ce qui concerne les chevauchements, ils sont toujours indiqués entre chevrons, avec l'initiale de la personne qui interrompt suivie de deux points (**L** : pour le locuteur, **E** : pour l'enquêteur). Mais, selon les cas envisagés, les consignes et conventions de notation à respecter sont données ci-après :

Cas 1 : E fait une intervention et L marque seulement son approbation

E: Hum hum. Donc euh bè votre père vous pouvez euh dire d'abord vos grand-parents si c'était un petit village si c'était Gaillac <L: Hum.> ville qu'est-ce qu'il fait lui enfin décrire un peu tout ça. <L : hum hum>



Cas 2 : E conserve la parole, malgré l'intervention de L à deux reprises

E: Humhum et euh <L: à mon sens.> donc vous étiez tout le temps à l'école en fait <L: Ouais.> c'êt- ça s'organisait comment une journée en en gros ?



Bien que L interrompe E à deux reprises, ce dernier garde la parole jusqu'à la fin du tour.

Cas 3 : L conserve la parole, même si E intervient assez longuement.

L : Euh <E: Bon Almodovar> bon j'ai bien aimé euh hein ?



Cas 4 : L interrompt E et prend ensuite la parole.

E: Il y avait pas la possibilité de faire des choses à deux trois quoi c'était vraiment réglementé. <L: Non c'était très>

L: rare c'est vraiment que si on devait faire vraiment un truc à deux pour un prof <E: Hum hum.> dans ce cas là on avait le droit mais euh sinon non après en troisième on avait droit à une salle en fait où on pouvait bosser à plusieurs mais ça on a eu du mal un peu à l'avoir.



L produit « non c'était très » bien que E n'ait pas fini son intervention. En ce sens, L interrompt E pour ensuite prendre la parole seule. Une autre intervalle est alors créé sous PRAAT.

### L'usage de l'orthographe et la typographie :

Conformément aux recommandations du groupe d'experts EAGLES, les transcriptions ont été faites en respectant l'orthographe standard. Seules les formes contractées données dans les dictionnaires de langue (Petit Robert et Larousse) ont été acceptées. Aucun trucage n'a été admis pour donner un caractère plus « oral » à la transcription.

#### **Exemples (les codages API sont au format [SAMPA](#))**

- [Ssepa] sera transcrit « **je sais pas** », et non \**j'sais pas*. Il faut cependant noter que le *ne* de négation n'est pas ajouté s'il n'a pas été produit par le locuteur.
- [ivwapasakOmsa] sera transcrit « **Il voit pas ça comme ça** » et non « *i voit pas ça comme ça* ».
- [japadkwa] sera transcrit « **Il y a pas de quoi** » et non « *y a pas d'quoi* ».

En conclusion, aucune invention orthographique permettant de simuler l'oral n'est utilisée. Ceci étant, seules les unités produites, mais de façon contractées, sont transcrites.

En cas de **faux départs** et de **reprises**, sont transcrits orthographiquement tous les items produits par le locuteurs.

**Exemple :**

L: Oui oui. *Ben euh dans les puisque j'y suis allée euh d'abord à l'a- au enfin à l'appartement où habitaient mes parents d'abord et ensuite ap- à la- à la maison. C'était pas la même école mais dans les deux cas euh j'y allais à pied oui.*



**Les mots tronqués** sont indiqués par un tiret directement à la suite du mot, sans espace.

**Exemple :**

E: Et après au **co-** euh collège c'était aussi un collège privé ? <L: Ouais>

**Les mots repris** sont retranscrits autant de fois qu'ils sont dits.

**Exemple :**

L: Euh pff. Globalement j'ai plutôt des mauvais souvenirs [rires]. J'é- J'étais pas une fana d'école je. C'était **des des** écoles de soeurs où il y avait une discipline importante et euh. J'étais pas du tout une fana d'école.



**Les sigles** sont transcrits de façon à indiquer s'ils sont lus ou épelés. S'ils sont épelés, chaque lettre est suivie d'un point, en revanche, s'ils sont lus, les lettres se suivent.

**Exemples (les codages API sont au format [SAMPA](#))**

- Dans «*La S.N.C.F. est à nouveau en grève.*», le sigle SNCF est transcrit **S.N.C.F** puisqu'il est prononcé [EsEnseEf] ;
- Dans l'énoncé «*Le CNET embauche du personnel.*», le sigle est prononcé [knEt], donc transcrit **CNET** ;

- L'énoncé prononcé [itRavajopaRtisosjalist] est transcrit « ***Il travaille au parti socialiste.*** » , tandis que celui prononcé [itRavajopeEs] est transcrit « ***il travaille au P.S.*** ».

**Les syllabes inaudibles** sont indiquées entre parenthèses, avec une croix par syllabe.

**Exemple :**

E: Des agriculteurs ? < L: Enfin paysan euh oui.> hum ils faisaient  
quoi comme type de (XX) ? < L: Ben un petit peu tout. >



**Les hésitations** et les **onomatopées** sont notées en respectant l'orthographe standard.

**Les discours rapportés** sont indiqués par des guillemets simples.

**Exemple :**

L: Il m'a dit 'tu es complètement idiot d'avoir accepté ce boulot' et  
je crois bien qu'au fond il avait raison.

**Les bruits** sont indiqués entre crochets.

**Exemple :**

E: En plus il aurait fallu faire un DEUG B où on physique math  
enfin <L: Voilà.  
[ grincement de chaise] Hum.>



Quant aux **chiffres**, ils sont transcrits en toute lettre.

Pour la transcriptions des textes, toutes les erreurs, les reprises et les hésitations sont indiquées dans la transcription.

# Exemples de transcription

## Extrait de la transcription de la lecture de texte 1 par le locuteur GRC

Feindre aujourd'hui de découvrir que le part- que le patinage est un sport sujet à caution relève soit de la totale naïveté soit de l'extrême hypocrisie (ou des deux, mon général).

Quant quant à nous il y a malh- malheureus- quant à nous il y a malheureusement longtemps que nous savons qu'une vraie réforme de ce sport s'impose d'autant que beaucoup de nos champions en ont souffert dans le passé.

Il y a deux ans, le président de la Fédération Française avait proposé que les juges ne dépendent plus des fédérations nationales comme aujourd'hui mais de l'I.S.U, la Fédération internationale, première étape vers la seule solution possible, la constitution d'un corps de juges pros et indépendants.

Et devinez qui a fait capoter cette réforme ?

Les Anglo-saxons justement....



## Extrait de la transcription de la présentation préparée faite par le locuteur LAU

Bon et puis euh dans cette partie euh enseignement aussi euh il y a une partie euh d'aide aux étudiants.

Bon on a souvent des entretiens avec eux euh parfois on les on les reçoit même pour leur donner euh un peu d'aide de soutien etc.

[inspiration] Donc cette partie là aussi fait partie de notre de notre métier d'enseignant de recevoir les étudiants [elle tourne la page] entre les cours.

Alors pour la deuxième euh partie donc la partie recherche qui euh là aussi on peut considérer qu'elle peut constituer disons un tiers du temps en temps euh normal quand on a une charge administrative quand même assez importante

[inspiration] Alors en fonction du domaine de recherche eh bien on va euh pendant ses é- pendant ses études disons et le sujet de thèse on continue euh sa généralement sa recherche dans dans le dans le domaine où on a fait sa thèse

[inspiration] on a un laboratoire d'accueil c'est-à-dire que on travaille bien sûr en en collaboration avec d'autres euh d'autres enseignants qui sont sur le même domaine.

[longue inspiration] Donc euh voilà ça c'est pour une partie



**Extrait de la transcription de la présentation préparée faite par le locuteur GOL**

L: Enfin assez vite quoi j'avais quand même sept huit ans mais bon c'est vrai c'est le souvenir euh que je garde donc c'était quand même une assez grande maison euh au départ c'était à Fonsorbes puis après euh ça a été Saint-Orens on s'est rapproché puis bon là c'est le nord de de Toulouse et bon donc c'est vrai que c'est toujours des des maisons assez confortables quoi.

L: Et puis bon sinon mes parents bon ben c'était le euh un peu le euh le type euh le cliché portugais si on peut dire le bon travailleur très attaché au travail euh c'est euh bon et puis voilà.

E: Et vous avez des frères et soeurs ou ? <L: oui>

L: J'ai un frère donc c'est vrai qu'ils étaient très attachés aux études ils voulaient ils nous ont poussés euh à faire des études pour pas qu'on euh bon pour qu'on s'élève socialement bon c'était important pour eux et donc c'est vrai que mon frère a fait des études aussi bon.





# Documentation linguistique diverse

Le chercheur chargé de revoir l'ensemble des transcriptions afin :

- ❖ de vérifier leur correction tant orthographique que typographique (suivi des consignes et conventions, etc.) ;
- ❖ d'uniformiser l'orthographe retenue pour transcrire certains mots de l'oral

a également fait des listes permettant de noter :

- ❖ quels noms propres sont utilisés dans l'ensemble du corpus ;
- ❖ comment sont codées plusieurs informations de l'oral.

Nous proposons dans cette section de fournir l'ensemble de ces listes.

## Lexique des noms propres :

Dans cette liste nous faisons figurer les noms propres qui ne figurent pas dans le *Larousse des noms propres* et dont l'orthographe n'a pas pu être vérifiée. Les noms sont classés par locuteur, le fichier où ils figurent étant indiqué entre parenthèses.

- Loc1\_ASD :
  - Reading (nom d'un musicien ou groupe de rock). (ASDE 10)
- Loc4\_BOA :
  - Ségenne (Lieu). (BOAE 4)
- Loc11\_GOL :
  - Mot portugais difficile à transcrire orthographiquement. A été retenue l'orthographe suivante « *Bérebaiêche* », mais cela fait penser à la région portugaise Beira Baixa. (GOLE 1)
- Loc13\_JAS :
  - Digne d'Aval (nom de commune). (JASE 6)
- Loc14\_JUM :
  - Carmaux (nom de commune). (JUME 4)
  - Pierre Ernet (Nom d'un pâtissier). (JUME 12)
  - Berkley (Université de San Francisco, USA). (JUME 14)
- Loc17\_MAG :
  - Roullens (nom de village). (MAGE 1, 4, 5, 7 et 8)
  - Vallon-Pont-d'Arc (nom de commune). (MAGE 3 et 4)
  - Saint Léon (nom de commune). (MAGE 4)
  - Villarzel (nom de commune). (MAGE 4)
  - Cazilhac (nom de commune). (MAGE 4)

- Loc22\_RIO :
  - Laboratoire « *cacahuètes* ». (RIOE 1)

## Liste et orthographe retenue pour les mots de l'oral et autres formes de codages :

### ❖ Orthographe retenue pour les mots de l'oral :

- bon
- ben
- bé
- bè,
- hum hum, mh
- euh,
- oh,
- ah,
- eh,
- ouh,
- pouh,
- ouf,
- pouf
- pff
- eh bien
- ouais
- tè

### ❖ Orthographe et codage retenus pour des indications diverses :

- raclement de gorge,
- toux,
- rires,
- sourire,
- claquement de langue,
- inspiration,
- bruit de bouche,
- reniflement,
- soupir,
- étouffement,
- bruit de fond (parfois précisions sur le type : moteur d'avion, chant d'oiseau, etc.).

# Identité des locuteurs

Dans cette section, plusieurs informations concernant les locuteurs et les informateurs sont exposées, cela afin d'avoir une idée précise sur les caractéristiques sociales et géographiques des locuteurs enregistrés dans les documents du corpus. En outre, des informations plus développées sur la vie et la famille des locuteurs sont accessibles à partir des liens.

## Locuteurs enregistrés pour les lectures de textes

	NOM	SEXE	AGE	PROFESSION	ORIGINE GEOGRAPHIQUE	ENQUETEUR
1	<a href="#">Loc1 ASD</a>	F	22 ans	Etudiante en Sciences du Langage	Tarn (Gaillac et Albi)	RED
2	<a href="#">Loc2 ASH</a>	F	21 ans	Etudiante en Lettres Modernes	Tarn (Castres)	RED
3	<a href="#">Loc3 BAV</a>	F	24 ans	Etudiante en Sciences du Langage	Tarn (Castres et environs)	TJM
4	<a href="#">Loc4 BOA</a>	F	21 ans	Etudiante en Lettres Modernes	Haute-Garonne (Toulouse)	RED et CIJ
5	<a href="#">Loc5 CAF</a>	F	20 ans	Etudiante en Lettres Modernes	Ariège (près de Pamiers)	RED et CIJ
6	<a href="#">Loc6 CAM</a>	F	21 ans	Etudiante en Lettres Modernes	Tarn (Albi)	RMO
7	<a href="#">Loc7 COA</a>	F	42 ans	Chargée de recherches	Aveyron et Haute-Garonne (Rodez et Toulouse)	RED
8	<a href="#">Loc8 COC</a>	F	21 ans	Etudiante en Sciences du Langage	Aveyron (St Affrique, Laguiole, Espalion) et un peu Toulouse	TJM
9	<a href="#">Loc9 DAA</a>	F	37 ans	Maître de Conférences	Toulouse	RMO
10	<a href="#">Loc10 FED</a>	F	23 ans	Etudiante en Lettres Modernes	Haute-Garonne (Montastruc)	RED et CIJ
11	<a href="#">Loc11 GOL</a>	F	38 ans	Enseignante de Portugais dans le secondaire	Haute-Garonne (Toulouse)	RED et CIJ
12	<a href="#">Loc12 GRC</a>	F	21 ans	Etudiante en Lettres Modernes	Lot et Garonne (Villeneuve sur Lot)	RED
13	<a href="#">Loc13 JAS</a>	F	20 ans	Etudiante en Sciences du Langage	Aude (Limoux et Carcassonne)	RED
14	<a href="#">Loc14 JUM</a>	F	21 ans	Etudiante en Sciences du Langage	Haute-Garonne (Colomiers et Toulouse)	TJM
15	<a href="#">Loc15 LAD</a>	L	27 ans	Etudiante en Sciences du Langage	Haute-Garonne (Toulouse)	RED et TJM

16	<a href="#">Loc16 LAU</a>	F	46 ans	Maître de Conférences	Tarn (Carmaux et environs) Haute-Garonne (Toulouse)	TJM
17	<a href="#">Loc17 MAG</a>	F	24 ans	Etudiante en Lettres Modernes	Aude (Roullens)	RED
18	<a href="#">Loc18 NES</a>	F	21 ans	Etudiante en Sciences du Langage	Tarn (Mazamet)	RED
19	<a href="#">Loc19 NOM</a>	F	22 ans	Etudiante en Lettres Modernes	Haute-Garonne (Toulouse)	RED et CIJ
20	<a href="#">Loc20 PEA</a>	F	22 ans	Etudiante en Lettres Modernes	Lot et Haute-Garonne (Cahors et Toulouse)	RED et CIJ
21	<a href="#">Loc21 RIC</a>	F	21 ans	Etudiante en Lettres Modernes	Est de la France (10 ans) et Toulouse	RED et CIJ
22	<a href="#">Loc22 RIO</a>	F	21 ans	Etudiante en Lettres Modernes	Tarn et Haute-Garonne (Mazamet et Toulouse)	RMO
23	<a href="#">Loc23 ROM</a>	L	59 ans	Professeur des Universités	Aude et Haute-Garonne (Carcassonne et Toulouse)	RED

## Locuteurs enregistrés pour les présentations monologuées préparées

	NOM	SEXE	AGE	PROFESSION	ORIGINE GEOGRAPHIQUE	ENQUETEUR
1	<a href="#">Loc1 ASD</a>	F	22 ans	Etudiante en Sciences du Langage	Tarn (Gaillac et Albi)	RED
2	<a href="#">Loc2 ASH</a>	F	21 ans	Etudiante en Lettres Modernes	Tarn (Castres)	RED
3	<a href="#">Loc3 BAV</a>	F	24 ans	Etudiante en Sciences du Langage	Tarn (Castres et environs)	TJM
4	<a href="#">Loc4 BOA</a>	F	21 ans	Etudiante en Lettres Modernes	Haute Garonne (Toulouse)	RED et CIJ
5	<a href="#">Loc7 COA</a>	F	42 ans	Chargée de recherches	Aveyron et Haute-Garonne (Rodez et Toulouse)	RED
6	<a href="#">Loc11 GOL</a>	F	38 ans	Enseignante de Portugais dans le secondaire	Haute Garonne (Toulouse)	RED et CIJ
7	<a href="#">Loc12 GRC</a>	F	21 ans	Etudiante en Lettres Modernes	Lot et Garonne (Villeneuve sur Lot)	RED
8	<a href="#">Loc13 JAS</a>	F	20 ans	Etudiante en Sciences du Langage	Aude (Limoux et Carcassonne)	RED

9	<a href="#">Loc14_JUM</a>	F	21 ans	Etudiante en Sciences du Langage	Haute-Garonne (Colomiers et Toulouse)	TJM
10	<a href="#">Loc15_LAD</a>	L	27 ans	Etudiante en Sciences du Langage	Haute Garonne (Toulouse)	RED et TJM
11	<a href="#">Loc16_LAU</a>	F	46 ans	Maître de Conférences	Tarn (Carmaux et environs) Haute-Garonne (Toulouse)	TJM
12	<a href="#">Loc17_MAG</a>	F	24 ans	Etudiante en Lettres Modernes	Aude (Roullens)	RED
13	<a href="#">Loc18_NES</a>	F	21 ans	Etudiante en Sciences du Langage	Tarn (Mazamet)	RED
14	<a href="#">Loc19_NOM</a>	F	22 ans	Etudiante en Lettres Modernes	Haute-Garonne (Toulouse)	RED et CIJ
15	<a href="#">Loc20_PEA</a>	F	22 ans	Etudiante en Lettres Modernes	Lot et Haute-Garonne (Cahors et Toulouse)	RED et CIJ
16	<a href="#">Loc22_RIO</a>	F	21 ans	Etudiante en Lettres Modernes	Tarn et Haute Garonne (Mazamet et Toulouse)	RMO
17	<a href="#">Loc23_ROM</a>	L	59 ans	Professeur des Universités	Aude et Haute Garonne (Carcassonne et Toulouse)	RED

## Locuteurs enregistrés pour les entretiens guidés

	NOM	SEXE	AGE	PROFESSION	ORIGINE GEOGRAPHIQUE	ENQUETEUR
1	<a href="#">Loc1_ASD</a>	F	22 ans	Etudiante en Sciences du Langage	Tarn (Gaillac et Albi)	RED
2	<a href="#">Loc2_ASH</a>	F	21 ans	Etudiante en Lettres Modernes	Tarn (Castres)	RED
3	<a href="#">Loc3_BAV</a>	F	24 ans	Etudiante en Sciences du Langage	Tarn (Castres et environs)	TJM
4	<a href="#">Loc4_BOA</a>	F	21 ans	Etudiante en Lettres Modernes	Haute Garonne (Toulouse)	RED et CIJ
5	<a href="#">Loc7_COA</a>	F	42 ans	Chargée de recherches	Aveyron et Haute-Garonne (Rodez et Toulouse)	RED
6	<a href="#">Loc11_GOL</a>	F	38 ans	Enseignante de portugais dans le secondaire	Haute Garonne (Toulouse)	RED et CIJ

7	<a href="#">Loc12_GRC</a>	F	21 ans	Etudiante en Lettres Modernes	Lot et Garonne (Villeneuve sur Lot)	RED
8	<a href="#">Loc13_JAS</a>	F	20 ans	Etudiante en Sciences du Langage	Aude (Limoux et Carcassonne)	RED
9	<a href="#">Loc14_JUM</a>	F	21 ans	Etudiante en Sciences du Langage	Haute-Garonne (Colomiers et Toulouse)	TJM
10	<a href="#">Loc15_LAD</a>	L	27 ans	Etudiante en Sciences du Langage	Haute Garonne (Toulouse)	RED et TJM
11	<a href="#">Loc16_LAU</a>	F	46 ans	Maître de Conférences	Tarn (Carmaux et environs) Haute-Garonne (Toulouse)	TJM
12	<a href="#">Loc17_MAG</a>	F	24 ans	Etudiante en Lettres Modernes	Aude (Roullens)	RED
13	<a href="#">Loc18_NES</a>	F	21 ans	Etudiante en Sciences du Langage	Tarn (Mazamet)	RED
14	<a href="#">Loc19_NOM</a>	F	22 ans	Etudiante en Lettres Modernes	Haute-Garonne (Toulouse)	RED et CIJ
15	<a href="#">Loc20_PEA</a>	F	22 ans	Etudiante en Lettres Modernes	Lot et Haute-Garonne (Cahors et Toulouse)	RED et CIJ
16	<a href="#">Loc22_RIO</a>	F	21 ans	Etudiante en Lettres Modernes	Tarn et Haute Garonne (Mazamet et Toulouse)	RMO
17	<a href="#">Loc23_ROM</a>	L	59 ans	Professeur des Universités	Aude et Haute Garonne (Carcassonne et Toulouse)	RED

# Locuteur n° 1 : ASD

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripteur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 22 ans

**Sexe** : F

**Profession** : Etudiante en Licence de Sciences du Langage (TAL)

**Niveau d'étude** : Bac + 3

**Langues maîtrisées** : Français (L1) ; Anglais et Espagnol (appris à l'école)

**Lieu de naissance et domiciles successifs** : La locutrice a surtout habité dans le Tarn, puisqu'elle est née et est restée 5 ans à Albi, puis a vécu 14 ans à Gaillac, d'où est originaire son père. Depuis le début de ses études, elle habite Toulouse.

## Les parents de la locutrice :

### Son père :

**Age** : Information non disponible. Il est décédé.

**Profession** : Médecin

**Etudes suivies** : Bac + 5 (Doctorat)

**Langues maîtrisées** : Information non disponible

**Lieu d'origine** : Tarn (Gaillac)

### Sa mère :

**Age** : 46 ans

**Profession** : Comptable

**Etudes suivies** : Bac + 2 (IUT Technique de Commercialisation)

**Langues maîtrisées** : Information non disponible

**Lieu d'origine** : Dordogne (Sarlat la Caneda)

## Profil géographique de la locutrice :

La locutrice est originaire du Tarn, principalement du Nord du département (Gaillac et Albi).

## Informations fournies par le transcripateur expert :

### Entretien

#### ❖ Transcription :

- **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**
  - ajout de morphèmes
  - transcription d'une séquence ou d'un mot pour un autre, par exemple « ça combien » a été remplacé par « c'est à combien » (ASDE1), « qu'est qui fait lui par qu'est-ce qu'il fait lui » (ASDE2), « mauvais » par « bon » (ASDE5)
  - changement de genre : « un scierie » a été remplacé par « une scierie » (ASDE4), « dures » par « durs » (ASDE4), « strictes » par « stricts » (ASDE4).
  - transcription de syllabes qui étaient notées inaudibles : l'atmosphère (ASDE2)
- **Doubles interprétations introduites dans la "tire commentaire" :**
  - hésitation sur la présence du « ne » de négation : « on (n') a plus vraiment d'endroit » (ASDE2)
  - hésitation entre « parce qu'on me euh on m'expliquait » et « parce que comme euh on m'expliquait »

#### ❖ Qualité audio:

La qualité audio est bonne. On entend des bruits d'avion mais ils n'empêchent pas l'audition de l'entretien.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

### Exposé

#### ❖ Transcription:

- **Corrections introduites dans la "tire orthographe"**  
Ajout de morphèmes.

#### ❖ Qualité audio:

La qualité audio est bonne. On entend des bruits d'avion et une sirène d'ambulance mais cela ne gêne pas l'audition de l'exposé.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.



## Textes

### ❖ Transcriptions:



#### **Remarque:**

Dans le troisième texte (lecture d'un extrait du Voyage au bout de la nuit), le transcrip-teur utilise les mêmes trucages orthographiques que Céline: j' peux rien..., y a pas, y a rien.



### Qualité audio:

La qualité audio est bonne.



### Caractéristiques du locuteur:

Rien de particulier à signaler.

# Locuteur n° 2 : ASH

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc. ;
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ les informations fournies par le transcripateur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age :** 21 ans

**Sexe :** F

**Profession :** Etudiante en 2<sup>nd</sup> Cycle de Lettres Modernes

**Niveau d'étude :** Bac + 3

**Langues maîtrisées :** Français (Langue première) ; anglais et allemand (appris à l'école).

**Lieu de naissance et domiciles successifs :** La locutrice a surtout vécu dans le Tarn, à Castres, où elle est née et est restée une quinzaine d'années. Elle a ensuite été à Albi et à Toulouse pour poursuivre ses études.

## Les parents de la locutrice :

### Son père :

**Age :** 49 ans

**Profession :** Directeur d'une banque à Castres

**Etudes suivies :** Bac + 4 (Ecole supérieure de Commerce)

**Langues maîtrisées :** Anglais et Allemand, mais également Occitan (avec les parents)

**Lieu d'origine :** Castres (Tarn)

### Sa mère :

**Age :** 47 ans

**Profession :** Infirmière

**Etudes suivies :** Bac + 3 (Ecole d'infirmière)

**Langues maîtrisées :** Information non disponible

**Lieu d'origine :** Gaillac (Tarn)

## Profil géographique de la locutrice :

La locutrice est originaire du Tarn, tout comme ses deux parents d'ailleurs.

## Informations fournies par le transcripteur expert :

### Entretien

#### ❖ Transcription :

- **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**
  - dans la très grande majorité des cas, ajout d'un morphème que le transcripteur n'avait pas noté: déterminant, préposition, connecteur, pronom, etc.
  - le transcripteur a reconstitué un ou plusieurs morphèmes alors que le locuteur ne faisait qu'une amorce, par exemple il a transcrit estompé au lieu de estom- (ASHE 5)
  - Transcription d'un mot pour un autre, exemple: heure pour œuvre (ASHE 9)
- **Doubles interprétations introduites dans la "tire commentaire":**
  - elles concernent principalement la présence ou l'absence du "ne" de négation. Dans certains cas, le transcripteur les a transcrits alors qu'il paraît très délicat de décider, par exemple : on (n') avait jamais fait, on (n') en avait jamais fait, on (n') a même pas le temps, etc. (ASHE 9)
  - Hésitations sur une forme verbale: par exemple j'étais ou j'ai été (ASHE 2)

#### ❖ Qualité audio:

La qualité audio est bonne. Il y a très peu de bruits de fond. On peut noter cependant quelques syllabes inaudibles.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

### Exposé

#### ❖ Transcription:

- **Corrections introduites dans la "tire orthographe"**

Les principales difficultés de transcriptions sont liées aux répétitions de prépositions. D'autre part certains morphèmes n'ont pas été transcrits (un connecteur, une onomatopée).
- **Doubles interprétations dans la "tire commentaire":**
  - Hésitation sur la présence d'un morphème: "ne" de négation (ASHP2), déterminant (ASHP1)
  - Hésitation entre deux prépositions: surtout au lycée ou en lycée (ASHP2)

- Hésitation sur une forme verbale: c'était ou ça a été (ASHP5)

❖ **Qualité audio:**

La qualité audio est bonne. Il n'y a pas de bruits de fond. Seulement une ou deux syllabes inaudibles (ASHP4)

❖ **Caractéristiques du locuteur:**

Rien de particulier à signaler.

## Textes

❖ **Transcriptions:**

➤ **Remarque:**

Dans ASHT3 (lecture d'un extrait du Voyage au bout de la nuit), le transcripateur utilise les mêmes trucages orthographiques que Céline: j' peux rien..., y a pas, y a rien.

➤ **Doubles interprétations dans la "tire commentaire":**

Hésitation sur une forme verbale: il s'était établi ou il s'est établi (ASHT3). Ici, seul le recours au texte permet de trancher.

❖ **Qualité audio:**

La qualité audio est bonne. Il n'y a pas de bruits de fond.

❖ **Caractéristiques du locuteur:**

Rien de particulier à signaler.

# Locuteur n° 3 : BAV

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc. ;
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ les informations fournies par le transcripateur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 24 ans

**Sexe** : F

**Profession** : Etudiante en DEUG de Sciences du Langage et surveillante d'externat

**Niveau d'étude** : Bac + 2

**Langues maîtrisées** : Français (Langue première) ; allemand, anglais et italien (appris à l'école).

**Lieu de naissance et domiciles successifs** : Bien que la locutrice habite actuellement entre Pau et Toulouse, elle a surtout vécu dans le Tarn (Longuegineste, Castres, etc. pendant 17 ans).

## Les parents de la locutrice :

### Son père :

**Age** : 49 ans

**Profession** : VRP puis animateur en centres de vacances

**Etudes suivies** : Jusqu'à la seconde.

**Langues maîtrisées** : Français, mais également Occitan

**Lieu d'origine** : Longuegineste (Tarn)

### Sa mère :

**Age** : 48 ans

**Profession** : Coordinatrice du tournoi international de Rugby junior.

**Etudes suivies** : Bac

**Langues maîtrisées** : Français, allemand et anglais appris à l'école.

**Lieu d'origine** : Vitry sur Seine (94)

## Profil géographique de la locutrice :

La locutrice est originaire du Tarn, où elle a été scolarisée et a passé 17 ans de sa vie. Son père est également originaire du Tarn, mais sa mère est de la région parisienne.

## Informations fournies par le transcripteur expert :

### Entretien

#### ❖ Transcription :

- **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**
  - ajout de morphèmes (pronoms, connecteurs, déterminants, prépositions, verbes...) et d'amorces
  - transcription d'une séquence ou d'un mot pour un autre, par exemple « qu'est-ce qui fait votre père » a été remplacé par « qu'est-ce qu'il fait votre père » (BAVE3), « que fait-il » par « ou que faisait-il » (BAVE3), « oralement » par « en allemand » (BAVE8), « également » par « véritablement » (BAVE9), « à tout ce qui est manuel » par « aux travaux manuels » (BAVE11), etc.
  - transcription de syllabes qui étaient notées inaudibles : « g-géographiquement » (BAVE1), « ils étaient originaires d'où » (BAVE2), « donc elle n'est pas elle n'est » (BAVE3), « c'est jeune ça » (BAVE4), « en voiture » (BAVE6), « c'est pas pareil » (BAVE8), etc.
- **Doubles interprétations introduites dans la "tire commentaire" :**
  - hésitation sur la présence du « ne » de négation : « on (n')a pas pu le faire » « on (n') a pas pu déménager le piano » (BAVE10)

#### ❖ Qualité audio:

La qualité audio est assez bonne dans l'ensemble mais les interventions de l'enquêteur sont souvent peu audibles. Un bruit d'avion gêne ponctuellement l'audition de l'entretien.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

## Exposé

#### ❖ Transcription:

- **Corrections introduites dans la "tire orthographe"**  
Ajout de morphèmes

#### ❖ Qualité audio:

La qualité audio est bonne.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

## Textes

### ❖ Transcriptions:

#### ➤ Remarque:

Dans la lecture de l'extrait de *Voyage au bout de la nuit* (T3), le transcripteur utilise les mêmes trucages orthographiques que Céline: j' peux rien..., y a pas, y a rien.

### ❖ Qualité audio:

La qualité audio est bonne.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

# Locuteur n° 4 : BOA

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripateur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 21 ans

**Sexe** : F

**Profession** : Etudiante en Licence de Lettres Modernes

**Niveau d'étude** : Bac + 3

**Langues maîtrisées** : Français (L1) ; Anglais et Allemand (appris à l'école)

**Lieu de naissance et domiciles successifs** : Bien que la locutrice soit née à Paris, nous pouvons dire qu'elle est de Toulouse puisqu'elle y est arrivée alors qu'elle était âgée de deux ans et qu'elle y est toujours restée.

## Les parents de la locutrice :

### Son père :

**Age** : 45 ans

**Profession** : Représentant commercial

**Etudes suivies** : Bac

**Langues maîtrisées** : Français (Langue première) et anglais (notions)

**Lieu d'origine** : Sarthe, le Mans.

### Sa mère :

**Age** : 45 ans

**Profession** : Cadre de Gestion

**Etudes suivies** : Bac + 2 (BTS secrétariat)

**Langues maîtrisées** : Français (langue première), anglais et allemand

**Lieu d'origine** : Yonne (Auxerre)

## Profil géographique de la locutrice :

La locutrice a vécu et a été scolarisée à Toulouse, mais il est important de noter que ses deux parents n'étaient pas originaires de la région Midi-Pyrénées.



# Informations fournies par le transcripteur expert :

## Entretien

### ❖ Transcription :

- **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**  
Ajouts de morphèmes: déterminants, connecteurs, pronoms
- **Doubles interprétations introduites dans la "tire commentaire":**
  - hésitation entre deux morphèmes homophones: et donc /euh, eux / vous vous entendez bien avec eux (BOAE 4)
  - hésitation sur la présence d'une préposition (BOAE9) ou d'un déterminant (BOAE10)

### ❖ Qualité audio:

La qualité audio est bonne. On entend un avion qui décolle en bruit de fond mais cela ne rend pas l'entretien inaudible.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

## Exposé

### ❖ Transcription:

- **Corrections introduites dans la "tire orthographe"**  
L'orthographe de plusieurs communes a été corrigé.

### ❖ Qualité audio:

La qualité audio est bonne. Il y a peu de bruit de fond et aucun passage inaudible.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

## Textes

### ❖ Transcriptions:

#### ➤ **Remarque:**

Dans BOAT 3 (extrait de *Voyage au bout de la nuit* (T3), le transcripteur utilise les mêmes trucages orthographiques que Céline: *j' peux rien..., y a pas, y a rien.*

### ❖ Qualité audio:

La qualité audio est bonne. Il n'y a peu de bruits de fond.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

# Locuteur n° 5 : CAF

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.

## Caractéristiques de la locutrice :

**Age** : 20 ans

**Sexe** : F

**Profession** : Etudiante en Licence de Lettres Modernes

**Niveau d'étude** : Bac + 3

**Langues maîtrisées** : Français (L1) ; Anglais et espagnol (appris à l'école)

**Lieu de naissance et domiciles successifs** : Bien que la locutrice habite à Toulouse depuis trois ans puisqu'elle y fait ses études, elle est originaire de l'Ariège, d'Arvigna, un village près de Pamiers, où elle a toujours habité.

## Les parents de la locutrice :

### Son père :

**Age** : 57 ans

**Profession** : sans emploi

**Etudes suivies** : Bac

**Langues maîtrisées** : Français (Langue première) et occitan (Langue seconde, parlée avec ses grands-parents)

**Lieu d'origine** : Aude.

### Sa mère :

**Age** : 52 ans

**Profession** : Professeur des Ecoles (institutrice)

**Etudes suivies** : Bac

**Langues maîtrisées** : Français (Langue première) et occitan (Langue seconde, parlée avec ses grands-parents)

**Lieu d'origine** : Ariège (Arvigna)

## Profil géographique de la locutrice :

La locutrice a vécu et a été scolarisée à Arvigna en Ariège, d'où était également originaire sa mère.

# Locuteur n° 6 : CAM

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.

## Caractéristiques de la locutrice :

**Age** : 21 ans

**Sexe** : F

**Profession** : Etudiante en Licence de Lettres Modernes

**Niveau d'étude** : Bac + 3

**Langues maîtrisées** : Français (L1) ; Anglais et allemand (appris à l'école)

**Lieu de naissance et domiciles successifs** : Bien que la locutrice soit née dans le Finistère à Landernau, elle peut être considérée comme venant du Tarn, plus précisément d'Albi, puisqu'elle y est arrivée âgée de 18 mois et y est restée jusqu'à maintenant.

## Les parents de la locutrice :

### Son père :

**Age** : 50 ans

**Profession** : Agent de direction à la Mutualité agricole

**Etudes suivies** : Bac + 2 (Etudes de Sciences Sociales)

**Langues maîtrisées** : Français (Langue première) ; occitan d'Aveyron (notions) ; Allemand et Espagnol

**Lieu d'origine** : Aveyron.

### Sa mère :

**Age** : 48 ans

**Profession** : Infirmière libérale

**Etudes suivies** : Bac + 3 (école d'Infirmière)

**Langues maîtrisées** : Français (Langue première) ; Espagnol et anglais

**Lieu d'origine** : Aveyron

## Profil géographique de la locutrice :

La locutrice a vécu et a été scolarisée dans le Tarn, à Albi ; ses parents, eux, sont originaires d'Aveyron.

# Locuteur n° 7 : COA

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripateur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 42 ans

**Sexe** : F

**Profession** : Chargée de Recherches au CNRS

**Niveau d'étude** : Bac + 5 (Doctorat)

**Langues maîtrisées** : Français (L1) ; Anglais et espagnol (appris à l'école)

**Lieu de naissance et domiciles successifs** : La locutrice est née et a vécu à Rodez (Aveyron de sa naissance à ses 19 ans, puis elle s'est installée à Toulouse où elle vit toujours.

## Les parents de la locutrice :

### Son père :

**Age** : 75 ans

**Profession** : Retraité, ancien ingénieur agricole

**Etudes suivies** : Bac + 3 (Etudes supérieures)

**Langues maîtrisées** : Français (Langue première) ; occitan d'Aveyron

**Lieu d'origine** : Aveyron, St Affrique

### Sa mère :

**Age** : 78 ans

**Profession** : Infirmière à la retraite

**Etudes suivies** : Bac + 3 (école d'Infirmière)

**Langues maîtrisées** : Français ; Occitan d'Aveyron

**Lieu d'origine** : Aveyron, Millau

## Profil géographique de la locutrice :

La locutrice a vécu et a été scolarisée à Rodez dans l'Aveyron, d'où sont originaires ses parents. Elle s'est ensuite installée sur Toulouse.

## Informations fournies par le transcripteur expert :

### Entretien

#### ❖ Transcription :

Quelques difficultés sont liées aux chevauchements de paroles

#### ➤ **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**

- le plus souvent, ajout d'un ou plusieurs morphèmes qui n'avaient pas été transcrits : connecteurs, prépositions, déterminants, amorces, etc.
- transcription d'un mot pour un autre, exemple *Provence* au lieu de *province* (COA16)
- à plusieurs reprises, le transcripteur avait noté « un » au lieu de « hein »
- plusieurs trucages orthographiques ont été enlevés

#### ➤ **Doubles interprétations introduites dans la "tire commentaire" :**

Hésitation sur la présence du « ne » de négation : *on (n')était pas resté* (COA14)

#### ❖ Qualité audio :

La qualité audio est bonne.

#### ❖ Caractéristiques du locuteur :

Rien de particulier à signaler.

### Exposé

#### ❖ Transcription :

Les principales difficultés sont liées à la transcription de répétitions et d'amorces.

#### ➤ **Corrections introduites dans la "tire orthographe"**

Ajout d'amorces et de morphèmes : adverbe, préposition, déterminant, connecteur, verbe

#### ➤ **Doubles interprétations introduites dans la "tire commentaire" :**

Hésitation sur la présence du « ne » de négation : *on (n')a pas* (COAP2, COAP17 et COAP18), *on (n')a plus* (COAP5)

#### ❖ Qualité audio :

La qualité audio est bonne.

- ❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler.

## Textes

- ❖ **Transcriptions:**
  - **Remarque:**  
Dans COAT 3 (extrait de *Voyage au bout de la nuit* (T3), le transcripteur utilise les mêmes trucages orthographiques que Céline: *j' peux rien...*, *y a pas*, *y a rien*.
- ❖ **Qualité audio:**  
La qualité audio est bonne.  
.
- ❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler.

# Locuteur n° 8 : COC

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.

## Caractéristiques de la locutrice :

**Age :** 21 ans

**Sexe :** F

**Profession :** Etudiante en DEUG de Sciences du Langage

**Niveau d'étude :** Bac + 2

**Langues maîtrisées :** Français et Espagnol (appris à l'école)

**Lieu de naissance et domiciles successifs :** Née à Bergerac (24) et a ensuite vécu en Aveyron (successivement St Affrique, Laguiole et Espalion). Est à Toulouse depuis trois ans pour ses études.

## Les parents de la locutrice :

### Son père :

**Age :** 43 ans

**Profession :** Gendarme

**Etudes suivies :** BEP d'électricien

**Langues maîtrisées :** Français

**Lieu d'origine :** Millau (Aveyron)

### Sa mère :

**Age :** 43 ans

**Profession :** Aide médico-psychologique

**Etudes suivies :** Baccalauréat

**Langues maîtrisées :** Français

**Lieu d'origine :** Forbach (Est de la France)

## Profil géographique de la locutrice :

Bien que la locutrice soit née en Dordogne, elle a surtout vécu en Aveyron, d'où son père est originaire. Elle vit à Toulouse depuis le début de ses études universitaires (3 ans). Sa mère est originaire de l'Est de la France.

# Locuteur n° 9 : DAA

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.

## Caractéristiques de la locutrice :

**Age** : 37 ans

**Sexe** : F

**Profession** : Maître de Conférences à l'Université de Toulouse – le Mirail

**Niveau d'étude** : Bac + 5 (Doctorat)

**Langues maîtrisées** : Français (L1) ; Anglais et espagnol (appris à l'école) ; allemand (notions)

**Lieu de naissance et domiciles successifs** : La locutrice est née et a vécu à Toulouse pendant toute sa formation. Elle a ensuite habité pendant 4 ans à Nantes avant de revenir sur Toulouse.

## Les parents de la locutrice :

### Son père :

**Age** : décédé à 62 ans

**Profession** : Ingénieur CNRS

**Etudes suivies** : Bac + 5 (Doctorat)

**Langues maîtrisées** : Français (Langue première) ; occitan compris mais peu parlé ; anglais et espagnol (appris à l'école)

**Lieu d'origine** : Aude, Carcassonne

### Sa mère :

**Age** : 63 ans

**Profession** : professeur d'Espagnol

**Etudes suivies** : Bac + 3 (Licence d'Espagnol)

**Langues maîtrisées** : Français ; Occitan (compris mais peu parlé) ; espagnol courant

**Lieu d'origine** : Carcassonne

## Profil géographique de la locutrice :

La locutrice a vécu et a été scolarisée à Toulouse ; ses deux parents en revanche sont originaires de Carcassonne.



# Locuteur n° 10 : FED

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.

## Caractéristiques de la locutrice :

**Age** : 23 ans

**Sexe** : F

**Profession** : Etudiante en Licence de Lettres Modernes

**Niveau d'étude** : Bac + 3

**Langues maîtrisées** : Français (L1) ; Anglais (appris à l'école)

**Lieu de naissance et domiciles successifs** : La locutrice est née à Chateauroux, mais a très vite déménagé dans la Région Midi-Pyrénées (elle avait 18 mois). Elle habite depuis son plus jeune âge à Montastruc, en Haute-Garonne, près de Toulouse.

## Les parents de la locutrice :

### Son père :

**Age** : 52 ans

**Profession** : Cadre SNCF

**Etudes suivies** : Bac

**Langues maîtrisées** : Français (Langue première) ; occitan ; espagnol (appris à l'école)

**Lieu d'origine** : Ariège

### Sa mère :

**Age** : 49 ans

**Profession** : Secrétaire

**Etudes suivies** : Bac

**Langues maîtrisées** : Français ; Occitan ; anglais (appris à l'école)

**Lieu d'origine** : Ariège

## Profil géographique de la locutrice :

La locutrice a vécu et a été scolarisée à Montastruc, près de Toulouse ; ses deux parents sont originaires de l'Ariège.

# Locuteur n° 11 : GOL

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripteur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 38 ans

**Sexe** : F

**Profession** : Enseignante de Portugais en collège et lycée et étudiante en Licence de Lettres Modernes

**Niveau d'étude** : Bac + 4

**Langues maîtrisées** : Portugais (L1), Français (L2)

**Lieu de naissance et domiciles successifs** : La locutrice est née au Portugal, mais a habité Toulouse tout de suite après sa naissance et y est restée jusqu'à ses 26 ans. Elle y a donc été scolarisée. Ensuite, elle a habité Lille pendant 6 ans ; puis est revenue sur Toulouse.

## Les parents de la locutrice :

### Son père :

**Age** : 62 ans

**Profession** : Artisan

**Etudes suivies** : Etudes primaires

**Langues maîtrisées** : Portugais

**Lieu d'origine** : Portugal

### Sa mère :

**Age** : 59ans

**Profession** : mère au foyer (sans profession)

**Etudes suivies** : Etudes primaires

**Langues maîtrisées** : Portugais, français

**Lieu d'origine** : Portugal

## Profil géographique de la locutrice :

La locutrice a appris le français à Toulouse où elle est arrivée dès son plus jeune âge et où elle a passé toute sa scolarité. Il faut cependant noter que sa langue première parlée en famille est le Portugais.

## Informations fournies par le transcripteur expert :

### Entretien

#### ❖ Transcription :

- **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**
  - dans la très grande majorité des cas, ajout d'un morphème que le transcripteur n'avait pas noté: onomatopée, pronom, déterminant, connecteur, etc.
  - Transcription d'un mot pour un autre, exemple: *émigré* pour *immigré* (GOLE 1)
- **Doubles interprétations introduites dans la "tire commentaire":**
  - Hésitations sur une forme verbale: par exemple *j'ai* ou *j'ai eu* (GOLE 8), *que je sois* ou *que je suis* (GOLE 10)
  - Hésitation entre la réalisation d'une forme verbale ou d'une amorce : *c'est c'était* ou *c'ét- c'était* ((GOLE 5)
  - Hésitations sur la présence d'un déterminant (GOLE 4, 5 et 7)
  - Hésitation sur la présence du "ne" de négation, par exemple *on n'avait pas* (GOLE 5 et 6)

Il reste des difficultés liées aux chevauchements de paroles.

Difficulté également à orthographier une région portugaise (GOLE 1). Seule région dont le nom ressemble : Beira Baixa.

#### ❖ Qualité audio:

La qualité audio est bonne. Il y a très peu de bruits de fond.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

### Exposé

#### ❖ Transcription:

- **Corrections faites :**
  - Ajout de morphèmes manquants : adjectif « petit » (GOLP2), connecteur (GOLP3)
  - multi-transcription pour le "ne" de négation dans *on (n') arrive pas* (GOLP3)

- ❖ **Qualité audio:**  
La qualité audio est bonne. Il y a peu de bruit de fond et aucun passage inaudible.
- ❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler.

## Textes

- ❖ **Transcriptions:**
  - **Remarque:**  
Dans GOLT 3 (extrait de *Voyage au bout de la nuit* (T3), le transcripteur utilise les mêmes trucages orthographiques que Céline: *j' peux rien...*, *y a pas*, *y a rien*.
- ❖ **Qualité audio:**  
La qualité audio est bonne. Il n'y a pas de bruits de fond.
- ❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler.

# Locuteur n° 12 : GRC

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripateur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 21 ans

**Sexe** : F

**Profession** : Etudiante en Licence de Lettres Modernes

**Niveau d'étude** : Bac + 3

**Langues maîtrisées** : Français (L1) ; Allemand et Anglais (appris à l'école)

**Lieu de naissance et domiciles successifs** : La locutrice a surtout habité dans le Lot et Garonne, puisqu'elle est née et est restée 18 ans à Villeneuve sur Lot, puis elle est venue sur Toulouse pour faire ses études.

## Les parents de la locutrice :

### Son père :

**Age** : 48 ans.

**Profession** : Métreur

**Etudes suivies** : Bac

**Langues maîtrisées** : Français (L1), Anglais et Espagnol (appris à l'école)

**Lieu d'origine** : Lot et Garonne

### Sa mère :

**Age** : 46 ans

**Profession** : Secrétaire

**Etudes suivies** : Bac

**Langues maîtrisées** : Français et Espagnol

**Lieu d'origine** : Lot et Garonne

## Profil géographique de la locutrice :

La locutrice est originaire du Lot et Garonne, comme ses parents d'ailleurs.

# Informations fournies par le transcripteur expert :

## Entretien

### ❖ Transcription :

- **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**
  - ajout de morphèmes (pronoms, connecteurs, déterminants, prépositions, verbes...) et d'amorces
  - transcription d'une séquence ou d'un mot pour un autre, par exemple « enfin » a été remplacé par « c'est » (GRCE1), « se » par « ça » (GRCE2), « et comment c'était » par « elle est comment cette maison » (GRCE3), « des qualités » par « qu'elle était bien » (GRCE5), « avez » par « aviez » (GRCE7), « excepté pour des raisons financières » par « et que c'est plus des raisons financières » (GRCE11), « jobs d'été » par « jobs d'étudiant » (GRCE13), « c'est plein de petits villages » par « c'est un petit village » (GRCE13), etc.
  - transcription de syllabes qui étaient notées inaudibles : « comme act- » (GRCE1), « une maison » (GRCE3), « dont vous gardez un très mauvais souvenir » (GRCE5), « j'allais dire » (GRCE6), « vous vous connaissiez » (GRCE11), « oh c'est quand même » (GRCE12), « vous savez » (GRCE13)
  - corrections dans la notation de sigle : CAP a été remplacé par C.A.P (GRCE2), BEP par B.E.P (GRCE2), CE1 par C.E.1 (GRCE5), TNT par T.N.T (GRCE12), WWF par V.V.F (GRCE13)
- **Doubles interprétations introduites dans la "tire commentaire" :**
  - hésitation sur la présence du « ne » de négation : « on (n') était pas toujours ensemble » (GRCE6), « on (n') avait pas du tout été sensibilisé » (GRCE9)
  - hésitation entre deux séquences verbales : « ils ont encore la ferme » et « ils sont encore à la ferme » (GRCE4)

### ❖ Qualité audio:

La qualité audio est bonne.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

## Exposé

### ❖ Transcription:

- **Corrections introduites dans la "tire orthographe"**  
ajouts de morphèmes (déterminant, pronom)

### ❖ Qualité audio:

La qualité audio est bonne.

- ❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler.

## Textes

- ❖ **Transcriptions:**
  - **Remarque:**  
Dans le troisième texte (extrait de *Voyage au bout de la nuit* (T3), le transcripateur utilise les mêmes trucages orthographiques que Céline: *j' peux rien...*, *y a pas*, *y a rien*.
- ❖ **Qualité audio:**  
La qualité audio est bonne.  
.
- ❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler.

# Locuteur n° 13 : JAS

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripateur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 20 ans

**Sexe** : F

**Profession** : Etudiante en DEUG 2 de Sciences du Langage

**Niveau d'étude** : Bac + 2

**Langues maîtrisées** : Français (L1) ; Anglais et Espagnol (appris à l'école) ; Italien (notions)

**Lieu de naissance et domiciles successifs** : La locutrice a surtout habité dans le Languedoc-Roussillon (Limoux et Carcassonne), puisqu'elle est arrivée dans cette région-là alors qu'elle n'avait que 18 mois et qu'elle y est restée jusqu'à ses 19 ans.

## Les parents de la locutrice :

### Son père :

**Age** : 47 ans.

**Profession** : Animation socio-culturelle.

**Etudes suivies** : Bac

**Langues maîtrisées** : Français (L1), Anglais et Espagnol (appris à l'école)

**Lieu d'origine** : Région Midi-Pyrénées

### Sa mère :

**Age** : 45 ans

**Profession** : Neuropsychologue

**Etudes suivies** : Bac + 4

**Langues maîtrisées** : Français (L1), Anglais, Espagnol et Italien

**Lieu d'origine** : Hautes Pyrénées



## Profil géographique de la locutrice :

Bien qu'elle n'y soit pas née, nous pouvons considérer que la locutrice est originaire du Languedoc-Roussillon, plus précisément de Carcassonne. Ses parents en revanche sont originaires de la région Midi-Pyrénées.

## Informations fournies par le transcripteur expert :

### Entretien

#### ❖ Transcription :

Les principales difficultés sont liées aux répétitions et aux hésitations des locuteurs.

#### ➤ **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**

- Ajout d'un morphème que le transcripteur n'avait pas noté: pronom, déterminant, connecteur, préposition etc.
- Transcription d'une séquence pour une autre, exemple: *ici par indiscret* pour *si c'est pas indiscret* (JASE 2)
- Le transcripteur a parfois eu recours à des trucages orthographiques qui ont été enlevés

#### ➤ **Doubles interprétations introduites dans la "tire commentaire":**

- Indécision sur la présence ou absence du "ne" de négation: on (n') était pas on (n') était pas vraiment beaucoup (JASE )

#### ❖ Qualité audio:

La qualité audio est bonne. . Il y a très peu de bruits de fond.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

### Exposé

#### ❖ Transcription:

Les principales difficultés sont liées à la transcription des hésitations et des répétitions.

#### ❖ Qualité audio:

La qualité audio est bonne. Il y a peu de bruit de fond et aucun passage inaudible.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

### Textes

#### ❖ Transcriptions:

Quelques hésitations et quelques reprises pendant la lecture.

➤ **Remarque:**  
Dans JAST 3 (extrait de *Voyage au bout de la nuit* (T3), le transcrip-teur utilise les mêmes trucages orthographiques que Céline: *j' peux rien...*, *y a pas*, *y a rien*.

- ❖ **Qualité audio:**  
La qualité audio est bonne. Il n'y a peu de bruits de fond.  
.
- ❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler.

# Locuteur n° 14 : JUM

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripteur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 21 ans

**Sexe** : F

**Profession** : Etudiante en DEUG de Sciences du Langage

**Niveau d'étude** : Bac +2

**Langues maîtrisées** : Français (L1) ; Anglais, allemand et italien (appris à l'école)

**Lieu de naissance et domiciles successifs** : La locutrice est née à Toulouse, elle a ensuite vécu à Colomiers pendant 20 ans. Elle est à Toulouse depuis un an.

## Les parents de la locutrice :

### Son père :

**Age** : 48 ans

**Profession** : Médecin

**Etudes suivies** : Bac + 5 (Doctorat de Médecine)

**Langues maîtrisées** : Français (Langue première) ; anglais parlé

**Lieu d'origine** : Tarbes

### Sa mère :

**Age** : 46 ans

**Profession** : Représentante de Presse

**Etudes suivies** : Bac + 3

**Langues maîtrisées** : Français ; anglais courant

**Lieu d'origine** : Toulon

## Profil géographique de la locutrice :

La locutrice a vécu et a été scolarisée à Colomiers. Son père est originaire de Tarbes et sa mère de Toulon.

# Informations fournies par le transcripteur expert :

## Entretien

### ❖ Transcription :

Les principales difficultés sont liées aux hésitations, aux reprises, aux répétitions.

#### ➤ **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**

- ajout de morphèmes (pronoms, connecteurs, déterminants, prépositions) et d'amorces
- transcription d'une séquence ou d'un mot pour un autre, par exemple « je vais » a été remplacé par « je voulais » (JUME1), « mis » par « lu » (JUME1), « ils exercent » par « il exerce » (JUME3), « comme elle doit se trouver » par « ben l'école se trouvait » (JUME 5), « j'ai de la chance » par « j'ai eu de la chance » (JUME13), etc.
- transcription de syllabes qui étaient notées inaudibles : « la lecture des textes » (JUME1), « oh pourquoi pas » (JUME1), « d'où ils sont » (JUME3), « comme comme musique » (JUME11), etc.
- corrections de noms propres : « Coulomiers » a été remplacé par « Colomiers », « Yasmina Aréza » par « Yasmina Réza » (GRCE10)

### ❖ Qualité audio:

Dans l'ensemble la qualité audio est plutôt bonne mais les interventions de l'enquêteur sont souvent peu audibles voire inaudibles. Un bruit d'avion gêne l'audition de l'entretien pendant quelques secondes (JUME8).

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

## Exposé

### ❖ Transcription:

#### ➤ **Corrections introduites dans la "tire orthographe"**

Ajout de morphèmes

#### ➤ **Doubles interprétations introduites dans la "tire commentaire":**

Hésitation sur la présence du « ne » de négation : on (n')a pas (COAP2, COAP17 et COAP18), on (n')a plus (COAP5)

### ❖ Qualité audio:

La qualité audio est bonne.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

## Textes

### ❖ Transcriptions:

#### ➤ Remarque:

Dans JUMT 3 (extrait de *Voyage au bout de la nuit* (T3), le transcripteur utilise les mêmes trucages orthographiques que Céline: *j' peux rien...*, *y a pas*, *y a rien*.

### ❖ Qualité audio:

La qualité audio est bonne.

.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

# Locuteur n° 15 : LAD

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripteur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques du locuteur :

**Age** : 27 ans

**Sexe** : M

**Profession** : Etudiante en Licence de Sciences du Langage

**Niveau d'étude** : Bac + 3

**Langues maîtrisées** : Français (L1) ; Anglais (appris à l'école)

**Lieu de naissance et domiciles successifs** : Le locuteur est né à Toulouse et y a toujours habité.

## Les parents du locuteur :

### Son père :

**Age** : 55 ans.

**Profession** : Information non communiquée

**Etudes suivies** : Information non communiquée

**Langues maîtrisées** : Information non communiquée

**Lieu d'origine** : Gironde (Bordeaux)

### Sa mère :

**Age** : 50 ans

**Profession** : Secrétaire

**Etudes suivies** : Bac

**Langues maîtrisées** : Français (L1)

**Lieu d'origine** : Toulouse

## Profil géographique du locuteur :

Le locuteur est né et a toujours vécu à Toulouse.

# Informations fournies par le transcripteur expert :

## Entretien

### ❖ Transcription :

Les principales difficultés sont liées aux hésitations, aux reprises, aux répétitions.

#### ➤ **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**

- dans la très grande majorité des cas, ajout d'un morphème que le transcripteur n'avait pas noté: connecteurs, déterminants, prépositions, onomatopées.
- le transcripteur a reconstitué un ou plusieurs morphèmes alors que le locuteur ne faisait qu'une amorce, par exemple il a transcrit *fait* au lieu de *f* (LADE1)...
- Transcription d'un mot pour un autre, exemple: *toc* pour *top*(LAD1), *maternel* pour *paternel* (LADE1), *alors* pour *bon* (LADE1), *venus* pour *revenus* (LADE1)...

#### ➤ **Doubles interprétations introduites dans la "tire commentaire":**

- hésitation sur la présence d'un morphème , par exemple : *je regrette pas le choix que j'ai fait* ou *je le regrette pas le choix que j'ai fait* (LADE8) *c'est une chambre de vingt mètres carré considérée comme studio* ou *comme un studio* (LADE9)
- Indécision sur la présence du « ne » de négation : *on (n') est pas forcément conscient* (LADE12)

### ❖ Qualité audio:

La qualité audio est bonne.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

## Exposé

### ❖ Transcription:

Beaucoup de difficultés sont liées à la transcription des hésitations, des amorces et des répétitions que fait le locuteur

### ❖ Qualité audio:

La qualité audio est bonne.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

## Textes

### ❖ Transcriptions:

#### ➤ Remarque:

Dans LADT 3 (extrait de *Voyage au bout de la nuit* (T3), le transcripteur utilise les mêmes trucages orthographiques que Céline: *j' peux rien...*, *y a pas*, *y a rien*.

### ❖ Qualité audio:

La qualité audio est bonne.

.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.



# Locuteur n° 16 : LAU

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripateur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 46 ans

**Sexe** : F

**Profession** : Enseignant-Chercheur, Maître de Conférences

**Niveau d'étude** : Bac + 5 (Doctorat)

**Langues maîtrisées** : Français (L1) ; Anglais et espagnol (appris à l'école)

**Lieu de naissance et domiciles successifs** : La locutrice est née et a passé une grande partie de son enfance dans le Tarn, près de Carmaux. Elle habite Toulouse depuis plus de vingt ans.

## Les parents de la locutrice :

### Son père :

**Age** : 67 ans

**Profession** : Retraité, ancien gendarme

**Etudes suivies** : Certificat d'études

**Langues maîtrisées** : Français (Langue première) ; occitan

**Lieu d'origine** : Carmaux (Tarn)

### Sa mère :

**Age** : 67 ans

**Profession** : Sans profession

**Etudes suivies** : Certificat d'études

**Langues maîtrisées** : Français ; Occitan

**Lieu d'origine** : Carmaux (Tarn)

## Profil géographique de la locutrice :

La locutrice a vécu et a été scolarisée dans le Tarn, près de Carmaux, d'où sont originaires ses deux parents. Elle habite Toulouse depuis une vingtaine d'année.

## Informations fournies par le transcripteur expert :

### Entretien

#### ❖ Transcription :

- **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**
  - ajout d'amorces, de morphèmes et de syntagmes dans des passages assez peu audibles.
  - des trucages orthographiques ont été enlevés, par exemple *t'achètes* a été remplacé par *tu achètes* (LAUE11)
  - transcription d'un mot pour un autre, par exemple *goût* pour *coût* (LAUE7) etc.
- **Doubles interprétations introduites dans la "tire commentaire":**
  - hésitation sur le « ne » de négation : *on (n')en fait plus* (LAUE13)
  - hésitation entre *dont* et *donc* : *en occitan dont tu me dis / en occitan donc tu me dis*

#### ❖ Qualité audio:

Dans l'ensemble la qualité audio est bonne mais les interventions de l'enquêteur sont souvent peu audibles voire inaudibles.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler, excepté la prononciation de année : [ãne].

## Exposé

#### ❖ Transcription:

- **Corrections introduites dans la "tire orthographe"**
  - ajout de morphèmes : connecteur (LAUP5)
  - transcription d'un mot pour un autre : *donc* remplacé par *dans* (LAUP4)
- **Doubles interprétations introduites dans la "tire commentaire":**
  - Hésitation sur la présence du « ne » de négation : *on (n') a pas* (LAUP3)
  - Hésitation sur la présence d'une interjection : *alors cet enseignement (eh) bien il s'agit...* (LAUP2)
  - Hésitation entre une amorce et un déterminant : *l'a- l'année / la l'année* (LAUP3)

- ❖ **Qualité audio:**  
La qualité audio est bonne.
- ❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler.

## Textes

- ❖ **Transcriptions:**
  - **Remarque:**  
Dans le troisième texte (extrait de *Voyage au bout de la nuit* (T3), le transcripteur utilise les mêmes trucages orthographiques que Céline: *j' peux rien...*, *y a pas*, *y a rien*.
- ❖ **Qualité audio:**  
La qualité audio est bonne.
- ❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler.

# Locuteur n° 17 : MAG

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripteur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 24 ans

**Sexe** : F

**Profession** : Etudiante en Maîtrise de Lettres Modernes

**Niveau d'étude** : Bac + 4

**Langues maîtrisées** : Français (L1) ; Anglais (appris à l'école)

**Lieu de naissance et domiciles successifs** : La locutrice est née à Carcassonne et a habité à Roullens dans l'Aude pendant 18 ans. Elle est ensuite venue sur Toulouse pour faire ses études.

## Les parents de la locutrice :

### Son père :

**Age** : 54 ans.

**Profession** : Sans

**Etudes suivies** : Brevet (Etudes Primaires)

**Langues maîtrisées** : Français (L1)

**Lieu d'origine** : Ardèche (Vallon Pont d'Arc)

### Sa mère :

**Age** : 49 ans

**Profession** : Infirmière en hôpital

**Etudes suivies** : Bac + 3

**Langues maîtrisées** : Français (L1)

**Lieu d'origine** : Aveyron (Millau)

## Profil géographique de la locutrice :

La locutrice est originaire du Languedoc-Roussillon, plus précisément de Carcassonne et sa région. Ses parents en revanche ne sont pas originaires de la région de Carcassonne.

## Informations fournies par le transcripteur expert :

### Entretien

#### ❖ Transcription :

- **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**
  - dans la très grande majorité des cas, ajout d'un morphème que le transcripteur n'avait pas noté: déterminant, préposition, connecteur, pronom, etc.
  - le transcripteur a reconstitué un ou plusieurs morphèmes alors que le locuteur ne faisait qu'une amorce, par exemple il a transcrit *parents* au lieu de *par-* (MAGE16)
  - Transcription d'un mot pour un autre, exemple: *informelle* pour *formel* (MAGE15), *des* pour *les* (MAGE9), *nous* pour *non* (MAGE11), etc.
  - Correction de sigles : U.R.S.S a été remplacé par E.R.S.S (MAGE14), D.E.U.G par DEUG (MAGE1)
- **Doubles interprétations introduites dans la "tire commentaire":**
  - elles concernent principalement la présence ou l'absence du "ne" de négation, exemples : *on (n')en faisait pas trop* (MAGE9), *on (n')était pas les premiers* (MAGE11), *on (n')a il y a plus de d'appel* (MAGE12), *on (n')a pas forcément* (MAGE13)
  - Hésitation sur la présence d'un connecteur : *des lotos (ou) des choses comme ça* (MAGE1)
  - Hésitations sur une forme verbale: *ça a été* ou *ça était* (MAGE14)

#### ❖ Qualité audio:

La qualité audio est bonne.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler, excepté la prononciation de année : [âne].

## Exposé

#### ❖ Transcription:

- **Corrections introduites dans la "tire orthographe"**
  - transcription d'un mot pour un autre : *plaisait* a été remplacé par *tentait* (MAGP6), *nous* par *ne* (MAGP2)
  - Sigle : D.E.U.G a été remplacé par DEUG

- **Doubles interprétations introduites dans la "tire commentaire":**
  - Hésitation sur la présence du « ne » de négation : *on (n')a pas vraiment d'ouvrages* (MAGP3)
  - Hésitation sur une forme verbale : *c'est / c'était* (MAGP5)

❖ **Qualité audio:**  
La qualité audio est bonne.

❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler, excepté la prononciation de année : [âne].

## Textes

❖ **Transcriptions:**  
➤ **Remarque:**  
Dans MAGT3 (extrait de *Voyage au bout de la nuit* (T3), le transcrip-  
tueur utilise les mêmes trucages orthographiques que Céline: *j' peux rien..., y a pas,*  
*y a rien.*

❖ **Qualité audio:**  
La qualité audio est bonne.

❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler.

# Locuteur n° 18 : NES

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripateur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 21 ans

**Sexe** : F

**Profession** : Etudiante en Licence de Sciences du langage (mention TAL)

**Niveau d'étude** : Bac + 3

**Langues maîtrisées** : Français (L1) ; Anglais et Allemand (appris à l'école)

**Lieu de naissance et domiciles successifs** : La locutrice est née à Mazamet dans le Tarn et y a habité jusqu'à ce qu'elle vienne sur Toulouse pour faire ses études supérieures.

## Les parents de la locutrice :

### Son père :

**Age** : 46 ans.

**Profession** : Employé de banque

**Etudes suivies** : Bac

**Langues maîtrisées** : Français (L1) ; Anglais et Allemand (appris à l'école) ; Espagnol (notions) ; Occitan (parlé)

**Lieu d'origine** : Tarn (Mazamet)

### Sa mère :

**Age** : 46 ans

**Profession** : Institutrice

**Etudes suivies** : Bac

**Langues maîtrisées** : Français (L1) ; Occitan (L2), Anglais et espagnol (appris à l'école)

**Lieu d'origine** : Tarn (Mazamet)

## Profil géographique de la locutrice :

La locutrice est originaire de Mazamet dans le Tarn, tout comme ses parents.

## Informations fournies par le transcripateur expert :

### Entretien

#### ❖ Transcription :

- **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**
  - ajout de morphèmes (connecteurs, adjectifs) et d'amorces.
  - le transcripateur a reconstitué un ou plusieurs morphèmes alors que le locuteur ne faisait qu'une amorce, par exemple *depuis* au lieu de *dep-* (NESE2), *grands-parents* au lieu de *grands-p-* (NESE3)
  - NESE6 : *Déjà j'avais vu qu'il y avait une euh une partie TAL donc il y a l'informatique* a été remplacé par *Déjà j'avais vu qu'il y avait une euh une partie TAL donc euh liée à l'informatique*
- **Doubles interprétations introduites dans la "tire commentaire" :**
  - hésitation sur la présence d'une amorce : *et après elle est (re-) elle a redemandé* (NESE3)

#### ❖ Qualité audio:

La qualité audio est bonne.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

### Exposé

#### ❖ Transcription:

- **Corrections introduites dans la "tire orthographe"**
  - ajout d'un morphème : parce que **que** ce soit en ...

#### ❖ Qualité audio:

La qualité audio est bonne.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

### Textes

#### ❖ Transcriptions:

##### ➤ **Remarque:**

Dans NEST3 (extrait de *Voyage au bout de la nuit* (T3), le transcripateur utilise les mêmes trucages orthographiques que Céline: *j' peux rien..., y a pas, y a rien.*



❖ **Qualité audio:**  
La qualité audio est bonne.

❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler.

# Locuteur n° 19 : NOM

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripateur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 22 ans

**Sexe** : F

**Profession** : Etudiante en Maîtrise de Lettres Modernes

**Niveau d'étude** : Bac + 4

**Langues maîtrisées** : Français (L1) ; Anglais (courant) ; Allemand (appris à l'école)

**Lieu de naissance et domiciles successifs** : La locutrice est née à Agen mais n'y a pas vécu. Elle est arrivée à Toulouse âgée de 5 ans et y est restée jusqu'à maintenant.

## Les parents de la locutrice :

### Son père :

**Age** : 47 ans.

**Profession** : Cadre

**Etudes suivies** : Bac + 2 (BTS)

**Langues maîtrisées** : Français (L1) ; Anglais (appris à l'école)

**Lieu d'origine** : Lot et Garonne

### Sa mère :

**Age** : 53 ans

**Profession** : sans profession

**Etudes suivies** : Information non communiquée

**Langues maîtrisées** : Français

**Lieu d'origine** : Lot et Garonne

## Profil géographique de la locutrice :

Bien que la locutrice soit originaire du Lot et Garonne (Agen), nous pouvons considérer qu'elle vient de Toulouse puisque c'est là qu'elle a entièrement été scolarisée. Ses parents, eux, sont originaires du Lot et Garonne.

## Informations fournies par le transcripteur expert :

### Entretien

#### ❖ Transcription :

- **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**
  - ajout de morphèmes et d'amorces.
  - Certains syntagmes ont été remplacés par d'autres, par exemple *pas du tout* par *pas plus que ça* (NOM2), etc.
  - Certains passages notés par des syllabes inaudibles (XXX) ont été transcrits, par exemple : *vous pouvez en* (NOME6), *j'allais dire* (NOME11)
  - Les trucages orthographiques du type *t'es* pour *tu es* (NOME7) ont été enlevés
- **Doubles interprétations introduites dans la "tire commentaire" :**
  - hésitation entre deux formes verbales : *je rencontre* / *je rencontrais* (NOME5)

#### ❖ Qualité audio:

La qualité audio est bonne. A plusieurs reprises on entend des bruits d'avion mais cela ne gêne pas l'audition de l'entretien.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

## Exposé

#### ❖ Transcription:

- **Corrections introduites dans la "tire orthographe"**
  - ajout d'un connecteur (NOMP3)
  - *un* remplacé par *donc* (NOMP3)
- **Doubles interprétations introduites dans la "tire commentaire" :**

Hésitation sur la présence d'un morphème : *deux œuvres une de Barbey d'Aurévilly qui (est) donc me concerne* (NOMP2), *comme je vous (le) disais ce matin* (NOMP3)

#### ❖ Qualité audio:

La qualité audio est bonne.

- ❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler.

## Textes

- ❖ **Transcriptions:**
  - **Remarque:**  
Dans le troisième texte (extrait de *Voyage au bout de la nuit* (T3), le transcrip-teur utilise les mêmes trucages orthographiques que Céline: *j' peux rien...*, *y a pas*, *y a rien*.
- ❖ **Qualité audio:**  
La qualité audio est bonne.
- ❖ **Caractéristiques du locuteur:**  
Rien de particulier à signaler.

# Locuteur n° 20 : PEA

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripateur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 22 ans

**Sexe** : F

**Profession** : Etudiante en Maîtrise de Lettres Modernes

**Niveau d'étude** : Bac + 4

**Langues maîtrisées** : Français (L1) ; Anglais et espagnol (notions)

**Lieu de naissance et domiciles successifs** : La locutrice est née à Toulouse et y a toujours vécu.

## Les parents de la locutrice :

### Son père :

**Age** : 52 ans.

**Profession** : Employé des Postes et télécommunications (PTT)

**Etudes suivies** : Information non communiquée.

**Langues maîtrisées** : Information non communiquée

**Lieu d'origine** : Lot (Cahors)

### Sa mère :

**Age** : 49 ans

**Profession** : assistante maternelle

**Etudes suivies** : Certificat d'études

**Langues maîtrisées** : Français (L1), notions d'Italien

**Lieu d'origine** : Tarn et Garonne (Montauban)

## Profil géographique de la locutrice :

La locutrice est originaire de Toulouse. Quant à ses parents, ils viennent de la région Midi-Pyrénées.

# Informations fournies par le transcripteur expert :

## Entretien

### ❖ Transcription :

- **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**
  - ajout d'un morphème (préposition, déterminant, adverbe, verbe), d'onomatopées et d'amorces.
  - Ajout de syntagmes qui n'avaient pas été transcrits : *on parle* (PEAP4), *moi j'étais petite et euh* (PEAE3), *faire un stage* (PEAE4), *qui fait de très bons cours* (PEAE9), *que j'aime, plusieurs matières* (PEAE10)
  - Transcription d'un mot par un autre : *d'abord* remplacé par *d'ailleurs* (PEAE3), *partie* remplacé par *période* (PEAE5), etc.
  - des trucages orthographiques ont été enlevés
- **Doubles interprétations introduites dans la "tire commentaire" :**
  - Hésitation sur la présence du « ne » de négation : on (n')entend enfin on comprend pas (PEAE3), on (n')enseigne pas (PEAE8)
  - Hésitation sur la présence d'une préposition : j'avais des not- (de) super notes (PEAE6)
  - Hésitation entre deux formes verbales : j'ai été déçue / j'étais déçue (PEAE9)

### ❖ Qualité audio:

La qualité audio est bonne.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler, excepté la prononciation de année : [âne].

## Exposé

### ❖ Transcription:

- **Corrections introduites dans la "tire orthographe"**

Ajout de morphèmes : une préposition, un verbe
- **Doubles interprétations introduites dans la "tire commentaire" :**

Hésitation sur la transcription du « ne » de négation : si on (n')aime pas lire (PEAP3)

### ❖ Qualité audio:

La qualité audio est bonne.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

## Textes

### ❖ Transcriptions:

#### ➤ Remarque:

Dans PEAT3 (extrait de *Voyage au bout de la nuit* (T3), le transcripteur utilise les mêmes trucages orthographiques que Céline: *j' peux rien...*, *y a pas*, *y a rien*.

### ❖ Qualité audio:

La qualité audio est bonne.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

# Locuteur n° 21 : RIC

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.

## Caractéristiques de la locutrice :

**Age** : 21 ans

**Sexe** : F

**Profession** : Etudiante en Licence de Lettres Modernes

**Niveau d'étude** : Bac + 3

**Langues maîtrisées** : Français (L1) ; Espagnol (courant) ; Anglais et Créole martiniquais (notions)

**Lieu de naissance et domiciles successifs** : La locutrice est née dans l'Est de la France et y est restée jusqu'à ses 10 ans (Nancy et Liverdun). Elle a ensuite vécu à Toulouse pendant 6 ans, puis à la Martinique pendant 3 ans. Elle est revenue sur Toulouse pour ses études supérieures.

## Les parents de la locutrice :

### Son père :

**Age** : 51 ans.

**Profession** : Militaire retraité

**Etudes suivies** : BEP mécanique.

**Langues maîtrisées** : Français (L1), « patois » vendéen (L2)

**Lieu d'origine** : Vendée

### Sa mère :

**Age** : 51 ans

**Profession** : Secrétaire médicale à la retraite

**Etudes suivies** : Bac

**Langues maîtrisées** : Français (L1)

**Lieu d'origine** : Lorraine (Nancy)

## Profil géographique de la locutrice :

Il est difficile d'attribuer un profil géographique à cette locutrice dans la mesure où elle a vécu pendant longtemps dans deux régions dont les parlers sont très différents, à savoir la Lorraine et la région Toulousaine.



# Locuteur n° 22 : RIO

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripteur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques de la locutrice :

**Age** : 21 ans

**Sexe** : F

**Profession** : Etudiante en Licence de Lettres Modernes

**Niveau d'étude** : Bac + 3

**Langues maîtrisées** : Français (L1) ; Espagnol et Anglais (appris à l'école)

**Lieu de naissance et domiciles successifs** : La locutrice est née dans le Tarn à Mazamet. Elle y est ensuite restée pendant quatre ans. Puis, après un séjour de trois ans à Paris, elle s'est installée à Toulouse où elle demeure maintenant depuis plus de 12 ans.

## Les parents de la locutrice :

### Son père :

**Age** : 54 ans.

**Profession** : artiste peintre

**Etudes suivies** : Information non communiquée.

**Langues maîtrisées** : Français (L1), Espagnol (courant)

**Lieu d'origine** : Paris

### Sa mère :

**Age** : 53 ans

**Profession** : Infographiste

**Etudes suivies** : Bac + 4

**Langues maîtrisées** : Français (L1) ; Allemand

**Lieu d'origine** : Paris

## Profil géographique de la locutrice :

Il est difficile d'attribuer un profil géographique à cette locutrice, mais il faut noter qu'elle a habité longtemps dans la région Midi-Pyrénées (4 ans dans le Tarn et 12 ans à Toulouse). Ses parents sont originaires de Paris.

## Informations fournies par le transcripteur expert :

### Entretien

#### ❖ Transcription :

Les principales difficultés de transcriptions sont liées à des chevauchements de paroles.

#### ➤ **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**

- transcription d'un mot pour un autre ou d'une séquence pour une autre, exemples : *désossé* (de Montaigne) au lieu de *les Essais* (de Montaigne) RIOE1, *oui voilà* au lieu de *bon bè voilà* RIOE1, *décidément* au lieu de *et sinon* RIOE3, *après* au lieu de *à peu près* RIOE5, *inscrite* au lieu de *enseignante* RIOE8, *c'est pareil* au lieu de *parfait* RIOE9 etc.
- ajout d'un morphème qui n'était pas transcrit : préposition, connecteur, déterminant.

#### ➤ **Doubles interprétations introduites dans la "tire commentaire" :**

- hésitation entre deux formes verbales : *j'habitais* et *j'ai habité* RIOE3
- hésitation sur la présence d'une préposition : *(de) la cuisine* RIOE10
- Indécision sur la transcription du « ne » de négation : *on (n') est pas obligé* RIOE9, *on (n') en a plus du tout entendu parler* RIOE10

#### ❖ Qualité audio :

La qualité audio est bonne. On entend épisodiquement de la musique et des voix en bruit de fond mais cela ne rend pas l'entretien inaudible.

#### ❖ Caractéristiques du locuteur :

Rien de particulier à signaler.

### Exposé

#### ❖ Transcription :

Les principales difficultés sont liées à la transcription des hésitations, des reprises.

#### ➤ **Corrections introduites dans la "tire orthographe" :**

ajout de prépositions qui n'avaient pas été transcrites

#### ➤ **Doubles interprétations introduites dans la "tire commentaire" :**

- hésitation sur la présence d'une préposition : *et non (de) mon activité étudiante* RIOP1
- hésitation entre plusieurs morphèmes : *euh c'est* ou *enfin* RIOP2

- hésitation entre deux formes verbales : *ça m'a facilité* ou *ça me facilitait*  
RIOP3

❖ **Qualité audio:**

La qualité audio est bonne.

❖ **Caractéristiques du locuteur:**

Rien de particulier à signaler, excepté la prononciation de année : [âne].

## Textes

❖ **Transcriptions:**

➤ **Remarque:**

Dans le texte T3 (extrait de *Voyage au bout de la nuit* (T3), le transcripteur utilise les mêmes trucages orthographiques que Céline: *j' peux rien...*, *y a pas*, *y a rien*.

❖ **Qualité audio:**

La qualité audio est bonne.

❖ **Caractéristiques du locuteur:**

Rien de particulier à signaler.

# Locuteur n° 23 : ROM

Plusieurs types d'informations concernant le locuteur seront fournis dans cette page, notamment :

- ❖ son identité personnelle : âge, niveau d'étude, profession, etc.
- ❖ son profil géographique qui se déduira des villes et des régions où il a vécu, ainsi que des origines géographiques de ses parents ;
- ❖ l'identité de ses parents, leurs origines géographiques, leur profession, etc.
- ❖ les informations fournies par le transcripateur expert après relecture des transcriptions relatives aux productions du locuteur.

## Caractéristiques du locuteur :

**Age** : 59 ans

**Sexe** : M

**Profession** : Professeur des Universités

**Niveau d'étude** : Bac + 5 (Doctorat et HDR)

**Langues maîtrisées** : Français (L1) ; Espagnol et Anglais (courant) ; Italien et Occitan (un peu) ; Allemand (notions)

**Lieu de naissance et domiciles successifs** : Le locuteur est né et a vécu à Carcassonne (Aude) jusqu'à ses dix-huit ans. Ensuite, il a vécu à Paris pendant 6 ans et à Nantes pendant 9 ans. Depuis maintenant plus de vingt ans il habite dans la région toulousaine.

## Les parents du locuteur :

### Son père :

**Age** : Décédé.

**Profession** : Employé de Banque

**Etudes suivies** : Bac + 2.

**Langues maîtrisées** : Français (L1), Allemand et Occitan (un peu)

**Lieu d'origine** : Aude

### Sa mère :

**Age** : 81 ans

**Profession** : Sans profession

**Etudes suivies** : Bac

**Langues maîtrisées** : Français (L1) ; Occitan (non pratiqué mais pourrait être courant) ; anglais

**Lieu d'origine** : Aude

## Profil géographique du locuteur :

Bien que le locuteur ait beaucoup « bougé », nous pouvons considérer qu'il est originaire de l'Aude puisque c'est là qu'il a été scolarisé.

## Informations fournies par le transcripteur expert :

### Entretien

#### ❖ Transcription :

- **Corrections ajoutées dans la "tire orthographe" lors de la relecture :**
  - ajout de nombreuses amorces et de morphèmes (déterminant, connecteur, « ne » de négation, préposition, pronom...)
  - certains termes ont été remplacés par d'autres : *est-ce que vous habitez* par *est-ce que vous êtes né* (ROME1), *se souvenir* par *me souvenir* (ROME4), *à l'avance* par *en avance* (ROME5), *c'était organisé* par *ça s'organisait* (ROME6), etc.
  - correction dans la notation de sigle : CM deux, CE un ont été remplacés par C.M.2, C.E.1 (ROME5), CES a été remplacé par C.E.S. (ROME7), DESS par D.E.S.S
  - certaines séquences notées par des syllabes inaudibles (XXX) ont été transcrites : de ville (ROME4), un petit bébé encore (ROME13)
- **Doubles interprétations introduites dans la "tire commentaire" :**
  - Hésitation entre deux formes verbales : j'ai été scolarisé / j'étais scolarisé (ROME5)
  - Hésitation sur la présence du « ne » de négation : on (n')avait pas cours (ROME6), on (n') a pas eu de poste (ROME10)
  - Hésitation entre deux mots : ville / vie (ROME1)

#### ❖ Qualité audio:

La qualité audio est bonne.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

### Exposé

#### ❖ Transcription:

- **Corrections introduites dans la "tire orthographe"**  
*quand l'enseignant en lycée* a été remplacé par *quand j'enseignais en lycée* (ROMP1)

#### ❖ Qualité audio:

La qualité audio est bonne.

#### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

## Textes

### ❖ Transcriptions:

Correction dans la notation d'un sigle : ISU a été remplacé par I.S.U

### ➤ Remarque:

Dans le texte T3 (lecture d'un extrait du *Voyage au bout de la nuit*), le transcripteur utilise les mêmes trucages orthographiques que Céline: *j'peux rien..., y a pas, y a rien.*

### ❖ Qualité audio:

La qualité audio est bonne.

### ❖ Caractéristiques du locuteur:

Rien de particulier à signaler.

# Quelques liens utiles :

## Informations relatives à la constitution des corpus :

- ❖ Text Encoding Initiative : <http://www.tei-c.org/>
- ❖ Corpus Encoding Standard et EAGLES : <http://www.cs.vassar.edu/CES/>
- ❖ EAGLES : <http://coral.lili.uni-bielefeld.de/EAGLES/>
- ❖ Page concernant les outils élaborée par le LDC (Linguistic data Consortium) : <http://www ldc.upenn.edu/annotation/>

## Sites à partir desquels quelques outils d'aide à la transcription de corpus oraux sont rapatriables :

- ❖ <http://cslu.cse.ogi.edu/toolkit/>
- ❖ <http://www.nist.gov/speech/atlas/>
- ❖ <http://chilides.psy.cmu.edu/>
- ❖ <http://www.fon.hum.uva.nl/praat/>
- ❖ <http://www ldc.upenn.edu/mirror/Transcriber/>

## Quelques corpus ou bases de données du français

- ❖ <http://valibel.fltr.ucl.ac.be/pfc.htm>
- ❖ <http://www.up.univ-mrs.fr/delic/>
- ❖ <http://bach.arts.kuleuven.ac.be/elicop/>