

# *Outils informatiques d'aide à la transcription*

---

## TRAVAUX PRATIQUES

Enseignant **Jean-Yves ANTOINE**  
(Jean-Yves.Antoine AT univ-tours.fr)

# Transcription orthographique de parole: Transcriber

## Présentation

Transcriber est un outils d'aide à la transcription de corpus oraux développé par la DGA (Claude Barras, Direction Générale de l'Armement) et le LDC américain. Ce gratuiciel (*freeware*) permet d'éditer la plupart des formats de signaux de parole et offre une interface interactive très bien conçue pour écouter et transcrire en parallèle ces corpus oraux. Par rapport à d'autres outils existants, Transcriber permet de transcrire plus spécifiquement des longs dialogues, au cours desquels la parole de chaque locuteur peut se chevaucher avec celle d'autres interlocuteurs ou simplement d'autres sons. En sortie, les transcriptions sont sauvegardées en format XML, suivant une DTD qui est propre au système. Cet outil peut être récupéré à l'adresse URL suivante :

<http://trans.sourceforge.net/en/presentation.php>.

## 1 Prise en main du logiciel : observer une transcription

En guise de première découverte, nous allons dans un premier temps nous intéresser à une transcription déjà réalisée. Lancez tout d'abord le logiciel Transcriber: l'interface générale du système apparaît à l'écran. En sélectionnant l'option File/Open Trans du menu, ouvrez la transcription frint980428.trns que vous aurez préalablement recopié sur votre compte, en même temps que le fichier audio frint980428.wav. Ces fichiers se trouvent dans le répertoire Demos et correspondent à l'enregistrement d'un journal radio.

The screenshot shows the Transcriber 1.4.2 interface. The main text area contains the following transcription:

```
filler
Simon Tivolle + Patricia Martin
1: ouais .
2: sûr ?
Simon Tivolle
ah bon ?[rire] non . blague , blague de Patricia .
[i] France-Inter , [rire-] il est 7 heures [-rire] .
nontrans
[no speaker]
filler
Patricia Martin
le journal , Simon Tivolle :
```

Below the text is a playback control bar with buttons for play, stop, and other functions. Below that is an audio waveform. At the bottom, there is a segmented timeline with labels for different parts of the audio, such as 'filler', 'music', 'nontrans', and 'les titres'. The timeline is marked with time values from 0 to 20.

Les informations concernant la transcription apparaissent alors dans trois zones distinctes:

- 1 En haut de fenêtre, la zone de transcription où l'on observe la succession des tours de parole entre les deux interlocuteurs (Patricia Martin et Simon Tivolle). Cette dans cette zone que l'on

rédigera la transcription. On remarquera à cette occasion qu'il est possible de représenter les tours de parole avec chevauchement (parole simultanée), de même que la superposition de sons non vocaux (ici, l'indicatif du journal radio).

- 2 La zone en milieu de fenêtre permet la visualisation du signal de parole. Cet affichage se limite à l'enveloppe de l'énergie du signal et ne propose pas de spectrogramme : cet affichage limité est suffisant pour les besoins de la transcription automatique puisqu'on a recours à lui que pour segmenter les tours de parole (recherche des zones de silence en particulier). On remarque que Transcriber a trouvé automatiquement le fichier audio qui correspond à la transcription (fichier `frint980428.wav`): comme nous le verrons plus loin, cette information est présente dans l'entête du fichier XML de transcription (`frint980428.tris`). Précisons enfin que dans cet exemple, les deux interlocuteurs ont été enregistrés sur une seule voix sonore (son mono). Il est cependant possible de gérer avec Transcriber des corpus audio enregistrés sur deux voies séparées (vous pouvez par exemple ouvrir le fichier de démonstration `know.tris`: vous verrez que deux signaux de parole simultanés s'affichent dans cette zone).
- 3 Enfin, la zone en bas de fenêtre donne une représentation condensée de la structuration (segmentation en tours de parole) de la transcription: c'est dans cette zone qu'on segmente tout d'abord le signal en tours de parole avant de transcrire chaque segment. Chaque ligne de la zone de segmentation représente une information particulière: nature des sons non vocaux (ici, il s'agit de musique), segmentation en zones transcrites (`filler`) et non transcrites (`nontrans`), succession des interlocuteurs (ou chevauchement un `no speaker` le cas échéant) et enfin début de la transcription orthographique de chaque segment. Une fois encore, toutes les informations décrites ici sont codées dans le fichier XML de transcription.

### 1.1 Ecoute du signal de parole, synchronisation et sélection temporelle

**Ecoute globale** — Histoire de découvrir, sur un exemple encore assez simple, les difficultés de la transcription orthographique, nous allons tout d'abord écouter le signal correspondant à cette transcription. Pour cela, il vous suffit de cliquer le bouton d'écoute (triangle vert) de la zone d'affichage audio (merci d'utiliser votre casque audio pendant les écoutes).

On voit directement l'intérêt de l'interface Transcriber: la transcription défile automatiquement au cours de l'écoute, de même qu'un curseur vertical parcourt le signal de parole et que les segments concernés sont alternativement mis en surbrillance dans la fenêtre du bas. Cette fonctionnalité permet de vérifier à la volée la qualité d'une transcription. Pour des zones plus difficiles, il faut recourir à une écoute limitée à un segment.

La synchronisation temporelle que nous venons d'observer est essentielle à la transcription manuelle des corpus de parole. Une fois encore, elle est permise par des informations de synchronisation présentes dans le fichier XML de transcription.

**Ecoute et sélection d'un ou plusieurs segment** — Grâce à la synchronisation transcription / signal, il est possible de limiter l'écoute à un seul segment: pour cela, cliquer sur le tour de parole de votre choix dans la zone de segmentation, et cliquez sur le bouton d'écoute.

En faisant glisser le pointeur de souris sur plusieurs zones, vous pouvez sélectionner plusieurs segments en une seule fois.



Cliquez sur le bouton d'avance ou de recul rapide de la zone de signal: à quels déplacements correspondent-ils?

**Sélection et déplacement temporel** — Lorsque aucun segment n'a été défini, la sélection d'une position temporelle ou d'une zone d'écoute ne peut se faire que directement dans la zone de signal :

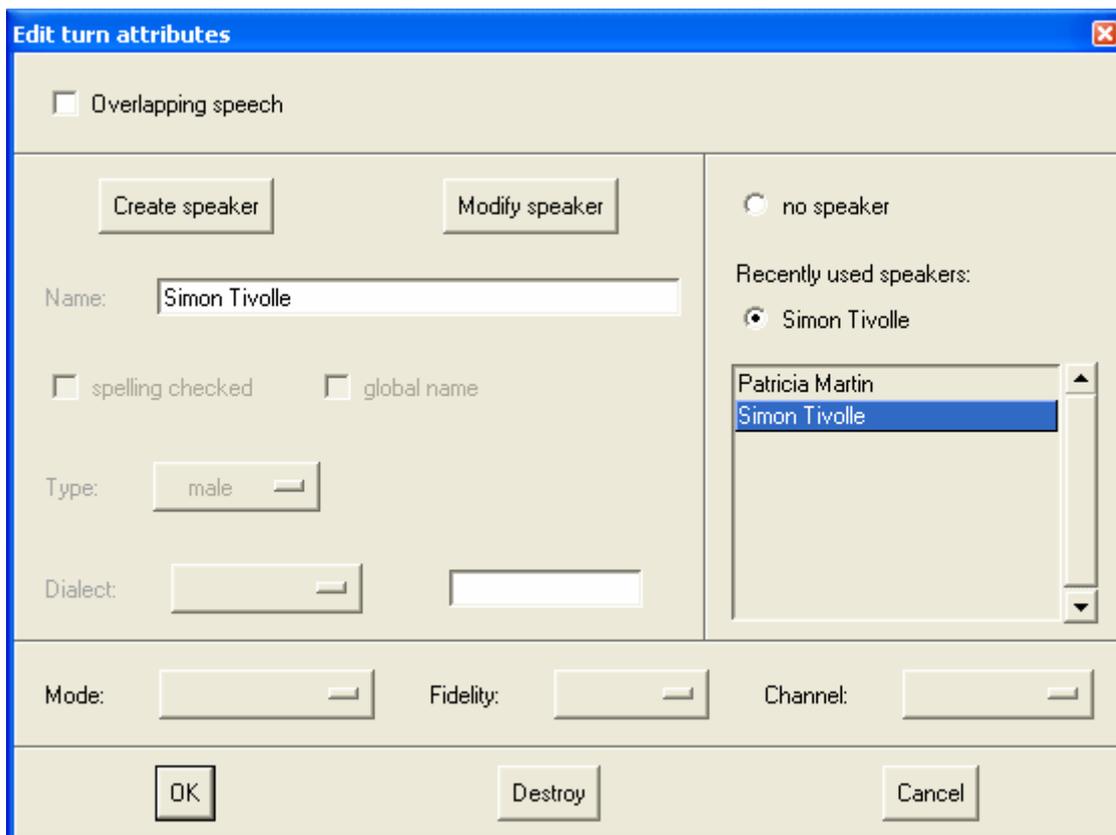
- 1 pour sélectionner une position, il suffit de cliquer avec le bouton gauche de la souris à l'endroit recherché: le curseur temporel se positionne directement,
- 2 pour sélectionner une zone temporelle, on clique avec le bouton gauche sur une position de début puis, sans relâcher, on déplace le pointeur souris jusqu'à la position finale. La zone en cours de sélection apparaît en gris: elle est définie par relâchement du bouton. Vous pouvez modifier la position de la fin de la zone ainsi définie par la combinaison **shift+clik gauche** et déplacement du curseur jusqu'à la position souhaitée.

Essayez ainsi d'écouter une zone quelconque du signal puis d'étendre cette zone de sélection.

## 1.2 Modification et sauvegarde d'une transcription

Maintenant que nous savons manipuler un signal de parole sous Transcriber, nous allons commencer à intervenir sur la transcription elle-même. Dans un premier temps, nous nous contenterons de modifier la transcription existante.

**Transcription orthographique** — Les signes de ponctuation de l'écrit (point, virgule, point d'interrogation...) sont rarement pertinents à l'oral. Il serait donc avisé de modifier toutes ces marques dans la transcription<sup>1</sup>. Pour cela, il vous suffit de travailler dans la fenêtre de transcription comme avec un éditeur de texte classique...



**Ajout de locuteurs** — Le transcripateur a fait une erreur sur le patronyme du présentateur du journal radiophonique: "Tivolle" ne prend en fait qu'un seul "l". Pour modifier cette information, sélectionnez un des boutons introduisant ce locuteur dans un tour de parole. La boîte de dialogue figurée sur la droite s'affiche à l'écran.

On remarque tout d'abord à droite une zone qui permet de définir le locuteur du tour de parole sélectionné (ou l'absence de locuteur sur ce segment). Nous utiliserons cette fonctionnalité plus tard quand nous créerons nous même une nouvelle transcription.

Pour l'heure, nous utiliserons l'option `Modify Speaker` pour corriger l'orthographe du nom considéré, en remarquant qu'il existe également une option de création de nouveau locuteur.

Validez votre modification ("OK"). Qu'observez-vous sur la transcription ? Là encore, c'est la sturcturation XML du document qui permet cette modification à la volée...

**Frontières de segments** — Enfin, il est également possible de modifier la limite temporelle d'un segment déjà défini. Pour cela il faut se placer sur la frontière du segment considéré et la déplacer par la combinaison **Ctrl + clic gauche**.

---

<sup>1</sup> En réalité, la présence de ponctuation est justifiée ici. En effet, les journaux télévisés ou radiodiffusés correspondent le plus souvent à de la parole lue, le journaliste se contentant de relire le texte qu'il a écrit auparavant. Ce registre de langue orale (au sens de Biber) est en fait très peu d'autres registres de l'écrit, d'où le recours parfaitement légitime à des signes de ponctuation.

Agrandissez légèrement, par exemple, un des segments correspondant à de la musique sans transcription.

**Enregistrement d'une transcription** — Après toute modification, vous pouvez enregistrer votre transcription à l'aide des commandes du menu `File / Save` ou `File / Save As`.

Enregistrez sur votre propre compte la transcription ainsi modifiée.

### 1.3 Transcription et XML

Avant de créer notre propre transcription, nous allons tout d'abord étudier le fichier XML de transcription créé par Transcriber, afin de découvrir les principales caractéristiques de ce format.

Fermez Transcriber et rendez-vous dans le répertoire où vous avez sauvegardé votre fichier de transcription. Le navigateur Internet Explorer sait visualiser des fichiers HTML (langage prédéfini par une norme) comme des fichiers XML (DTD propre à chaque document).

- 1 Ouvrez donc le fichier de transcription (.trs) à l'aide d'Internet Explorer. Qu'observez-vous? D'où provient le problème rencontré ?
- 2 Afin de visualiser correctement ce fichier XML, vous pouvez placer sur votre compte la DTD Transcriber que vous trouverez sur la page WWW consacrée à cet enseignement: fichier `trans-13.dtd`.
- 3 Une fois votre fichier correctement affiché, tentez de retrouver parmi les balises XML les informations concernant les différents locuteurs, les différents tours de parole ainsi que les informations de synchronisation qui leur sont associées. La notion de section, que nous rencontrons également ici, ne sera pas étudiée dans le cadre de ce cours.

## 2 Réalisation *ab nihilo* d'une transcription

Maintenant que nous avons compris les principes généraux qui gouvernent la manipulation de Transcriber, nous allons réaliser intégralement une première transcription. Celle-ci concernera un extrait du corpus `ECOLE_MASSY`, distribué par le laboratoire LI dans le cadre du projet `PAROLE_PUBLIQUE` ([http://www.info.univ-tours.fr/~antoine/parole\\_publicue](http://www.info.univ-tours.fr/~antoine/parole_publicue)). Ce corpus regroupe des enregistrements de dialogue entre des enfants d'école primaire et leur institutrice. Les dialogues rencontrés ici sont relativement aisés à transcrire, puisqu'on n'observe aucune situation de parole spontanée.

### 2.1 Création d'une nouvelle transcription

Pour commencer la transcription, il faut tout d'abord disposer du fichier audio correspondant au dialogue. Récupérez ce fichier (fichier `audio_cl.wav`) sur le forum ou sur ma page WWW, et recopiez-le sur votre compte.

Lancez ensuite Transcriber, puis créez une nouvelle transcription en choisissant l'option `File/New Trans` du menu. Sélectionnez le fichier audio dans la fenêtre de dialogue `Open Audio File` qui s'ouvre à l'écran : le signal de parole est alors affiché dans la fenêtre de travail. Vous pouvez déjà écouter une fois ce dialogue : en dépit d'une qualité sonore perfectible (bruit de fond), le dialogue entre l'enfant et son institutrice est parfaitement audible.

La transcription d'un signal de parole se fait en quatre étapes successives :

1. Renseignement des informations concernant le dialogue transcrit (caractéristiques des locuteurs, de la situation de dialogue...). Ces informations seront présentes dans l'entête du fichier XML de transcription.
2. Segmentation du dialogue en tours de parole
3. Transcription segment par segment
4. Vérification de la transcription.

Nous allons suivre chacune de ces étapes.

### 2.2 Renseignements sur le dialogue

**Caractéristiques générales** — La définition des caractéristiques du dialogue se fait en sélectionnant le menu `File / Edit Episode Attributes`. Au minimum, il est nécessaire de préciser le nom

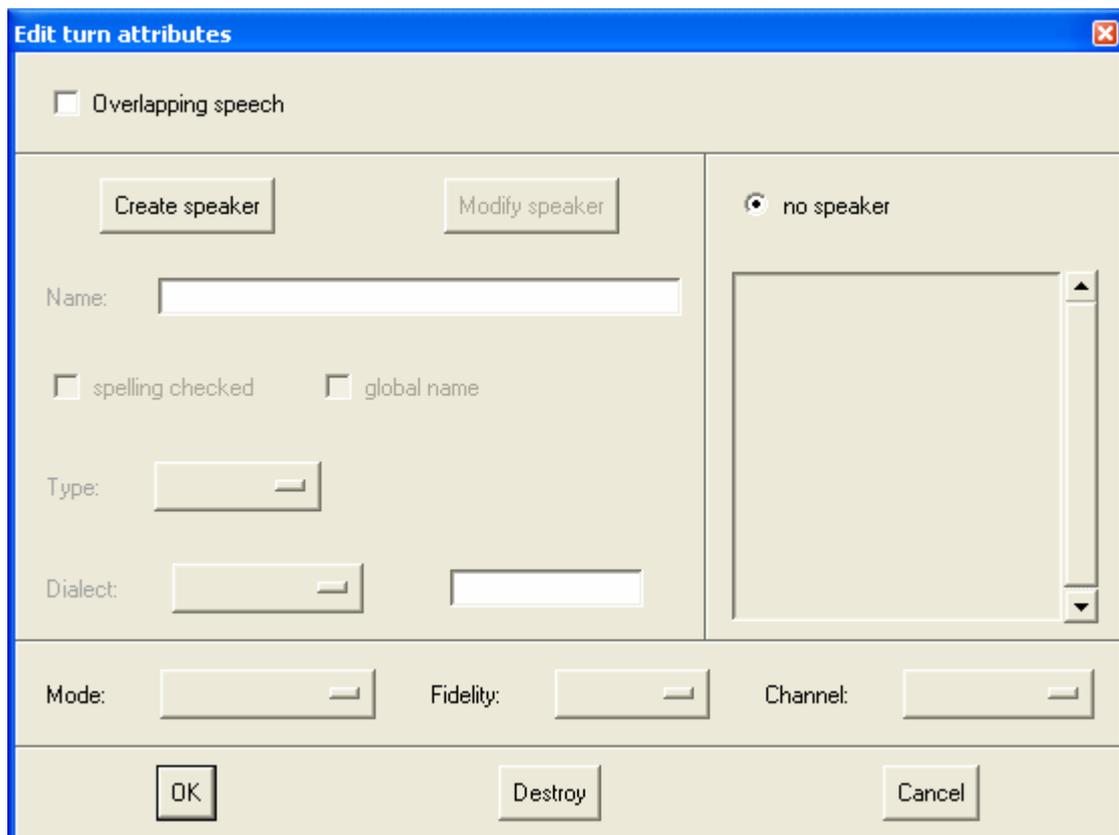
de la personne qui a réalisé la transcription, pour pouvoir la consulter le cas échéant (reprise ultérieure d'une transcription par exemple). Il est également utile de définir la situation de dialogue :

- 1 la langue principale utilisée dans le dialogue (utile par exemple pour les corpus multilingues)
- 2 un descriptif du dialogue (champ `program`) : ici, par exemple, on précisera corpus Ecole Massy.

Les autres champs (versions, date de transcription, nom du fichier de transcription) sont générés automatiquement.

Renseignez donc les informations concernant votre identité, la langue et le programme.

**Définition des locuteurs** — Dans le cas d'une transcription de dialogue, il faut également définir l'identité des différents interlocuteurs. Pour cela, sélectionnez les menu `Segmentation/Edit Turn Attributes` : vous pouvez créer de nouveaux locuteurs en sélectionnant le bouton `Create Speaker`.



La localisation de ce menu peut surprendre. Elle est au contraire très logique : dans les situations les plus complexes, il est difficile de prévoir le nombre de locuteurs qui vont intervenir. C'est donc à chaque nouveau tour de parole qu'on choisira soit de définir un nouveau locuteur, soit d'associer un locuteur déjà défini. Lorsque nous connaissons, comme ici, le nombre et l'identité des locuteurs, autant réaliser cette opération une fois pour toute...

Définissez donc l'identité des deux locuteurs. Il n'est pas inutile de rappeler ici une règle déontologique de base de constitution de corpus oraux : à l'exception des personnes publiques, qui jouissent d'un statut particulier, il est impératif de ne pas révéler l'identité des personnes qui ont été enregistrées. On choisira donc comme identités `institutrice` et `enfant`, et nom pas le patronyme réel des personnes.

## 2.3 Segmentation

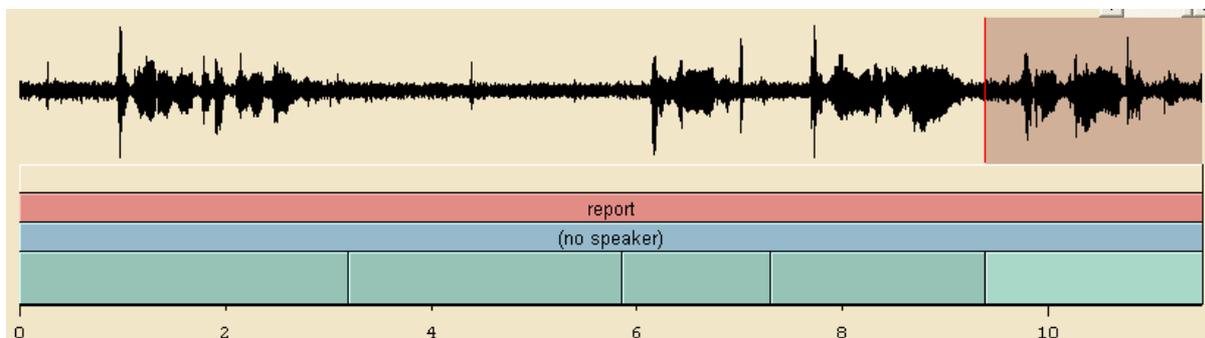
Une fois ces informations définies, on peut passer à la seconde étape de segmentation de l'énoncé.

**Structuration d'une transcription** — Transcriber distingue quatre niveaux récursifs de structuration de transcription (cette structure se retrouve dans les fichiers XML de transcription créés).

1. **Episode** ou **enregistrement** — il correspond au dialogue dans sa totalité : à un fichier de transcription correspond donc un *épisode* (*enregistrement* dans la version française)
2. **Section** — L'enregistrement est découpé en *sections*, délimitant les parties où il y a dialogue de celles qui correspondent à des parties non transcrites (musique, sonals, etc...).
3. **Tours** de parole — À l'intérieur d'une section on va distinguer les *tours de parole* (*speech turns*) qui correspondent à des changements de locuteur (soit que le locuteur change, soit qu'on passe à un dialogue avec chevauchements).
4. **Segments** — Les tours de paroles sont eux-mêmes découpés en segments plus courts au moyen de points de synchronisation. La transcription proprement dite n'intervient qu'à ce niveau de structuration.

L'utilisation de segments n'est utile que si on veut caractériser des événements particuliers au sein des tours de parole. De même, l'utilisation de segments n'a pas de sens si les dialogues ne comportent que des échanges langagiers sans sons additionnels.

**Segmentation initiale** — Pour segmenter le signal de parole, il suffit de se placer à chaque position temporelle correspondant à un changement acoustique (changement de locuteur, silence, musique,...) et d'aller à la ligne (touche `Enter`, `Entrée` ou `Return` de votre clavier) : on constate que le signal est bien découpé en différentes zones successives.



Découpez donc votre signal de parole en segments acoustiquement cohérents.

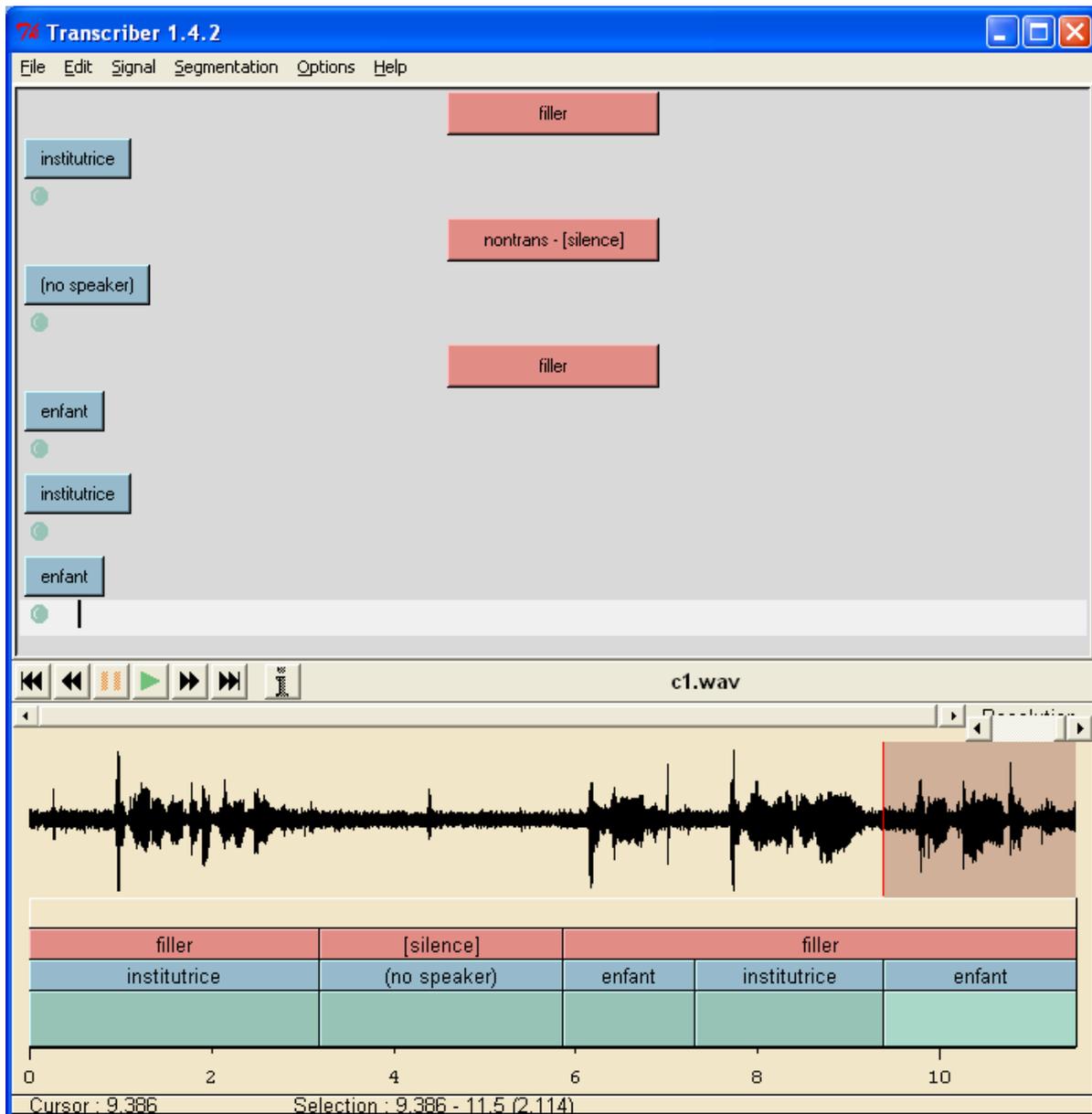
**Segmentation en tours de parole et sections** — Une fois cette segmentation de base réalisée, on peut associer des débuts de tours de parole ou de section à chacune de ses frontières. Pour cela, il suffit de choisir les options suivantes :

- 1 sélection du menu `Segmentation/Create Turn` (ou directement `Ctrl-t`) pour définir un nouveau tour de parole (choisissez l'identité du locuteur et validez).
- 2 sélection du menu `Segmentation/Create Section` (ou directement `Ctrl-r`) pour définir un nouvelle section (choisissez le type section : `filler` pour de la parole, `nontrans` pour du silence, `report` pour du bruit additionnel). Dans la zone `topic`, vous pouvez préciser la nature de cette section (musique, etc...).

Un segment initial peut bien entendu servir de frontière à la fois à un tour de parole et à une section : dans ce cas, il faut répéter les deux actions correspondantes. Notez enfin que si vous vous trompez sur une étiquette de tour de parole ou de section, vous pouvez modifier ces propriétés en cliquant sur le bouton correspondant dans la zone de transcription.

Suivant un principe de transcription général, définissez les sections comme les alternances marquées entre zone de parole et autres (silence dans notre cas) : ne pas s'attacher à rendre compte de pauses silencieuses de courte durée. Définissez de même les tours de parole entre les deux locuteurs ou les zones de silence. Vous remarquerez que si vous définissez les tours de parole suivant le déroulement du signal de parole, Transcriber vous propose alternativement une alternance de locuteur entre tours de parole successifs.

A la fin de cette opération, la structure de la transcription est complète et ne demande qu'à être complétée.



## 2.4 Transcription

Le signal étant segmenté, il ne vous reste plus qu'à transcrire chaque tour de parole, en respectant les conventions orthographiques étudiées en cours.

**Ajout de points de synchronisation temporelle (*breakpoints*)** — Si nécessaire, la transcription peut encore être affinée. Certains tours de parole comportent en effet des marques de respiration ou des raclements de gorge marqués dont il serait bon de rendre compte.

Pour ce faire, nous allons définir des points de synchronisation qui permettront de découper chaque tour de parole en parties successives de parole et de bruits acoustiques. Par exemple, placez-vous au début du dialogue : vous pouvez entendre une pause marquée d'une respiration avant que l'institutrice ne prononce l'énoncé *qu'as-tu choisi comme activité*.

Définissez alors un point de synchronisation à la fin de cette pause : il vous suffit pour cela d'entrer à nouveau un retour à la ligne (Entrée ou Return de votre clavier) ou de sélectionner le menu Segmentation / Insert breakpoint.

Vous constatez qu'une nouvelle ligne de transcription a été créée dans le tour de parole : indiquez-y la présence d'une respiration par le symbole [r] défini dans les conventions de transcription de Transcriber (consulter le guide du transcripteur sur le site WWW du logiciel)

**Remarque** — on ne cherchera pas ici à distinguer une inspiration d'une expiration ! En règle générale, une respiration aussi peu marquée n'est même pas marquée sauf besoins spécifiques...

De même, utilisez une marque de synchronisation pour indiquer le raclement de gorge de l'enfant lors du dernier tour de parole. Ce raclement s'apparente déjà à une onomatopée assez distincte : transcrivez là par *hum hum*, par exemple.

The screenshot shows the Transcriber 1.4.2 interface. The top part displays a transcription with speaker labels and time markers. Below the transcription is a control bar with playback buttons and a progress indicator for 'c1.wav'. The bottom part features a waveform and a segmentation table.

filler		[silence]		filler	
institutrice	(no speaker)	enfant	institutrice	enfant	
[r]	qu'as tu choisi comme activité	un film	quel film veux tu voir	hum hum	"les Razmoket"

Cursor : 10.19      Selection : 10.19 - 11.5 (1.31)

## 2.5 Vérification

Réécoutez attention votre transcription (traduction orthographique, limite des segments). Si vous jugez que votre transcription est complète, vous pouvez l'enregistrer.

## 3 Transcription de dialogues avec chevauchements

Vous avez maintenant toutes les connaissances suffisantes transcrire aussi proprement que possible un fichier de dialogue oral qui pourra comporter des cas de chevauchement. Ces situations ne changent rien à la méthodologie de transcription. Simplement, lors de la définition d'un tour de parole, il vous faut définir le locuteur comme correspondant à un chevauchement (*overlapping speech*)

puis choisir les locuteurs concernés (ils peuvent être plus de deux !). Transcrire affiche alors autant de lignes de transcription que nécessaire dans le tour de parole.

Pour réaliser ce type de transcription, téléchargez sur la page du programme PAROLE PUBLIQUE un fichier audio du corpus OTG ou du corpus Accueil\_UBS.

The screenshot shows the 'Edit turn attributes' dialog box with the following settings:

- Overlapping speech
- Choose first speaker
- Choose second speaker
- Buttons: Create speaker, Modify speaker
- Name: enfant
- spelling checked,  global name
- Type: [dropdown]
- Dialect: native, fr
- Mode: [dropdown], Fidelity: [dropdown], Channel: [dropdown]
- Buttons: OK, Destroy, Cancel
- no speaker:
- Recently used speakers:  enfant,  institutrice
- Speaker list: enfant (selected), institutrice

The screenshot shows the 'Edit turn attributes' dialog box with the following settings:

- Overlapping speech
- no speaker
- Buttons: Create speaker, Modify speaker
- Name: institutrice
- spelling checked,  global name
- Type: [dropdown]
- Dialect: native, fr
- Mode: [dropdown], Fidelity: [dropdown], Channel: [dropdown]
- Buttons: OK, Destroy, Cancel
- Recently used speakers:  enfant,  institutrice
- Speaker list: enfant, institutrice (selected)