Les propositions de la TEI pour l'oral

Novembre 2011

Approchons-nous, petit à petit de l'*oral*

L'oral peut correspondre (du point de vue de la transcription, à des choses diverses...

- des transcription de discours
- des transcription d'interview (dans une revue par exemple)
- des dialogues dans un roman
- du théatre
- du *vrai* oral

Je vous propose qu'on s'approche progressivement de ce 'vrai' oral.

Un discours politique

Discours du Président de la République à l'occasion de la cérémonie de remise de décorations du 11 novembre 2012 |
Mesdames, Messieurs,
C'est un moment exceptionnel que nous allons vivre aujourd'hui à l'occasion du 11 novembre 2012.
Exceptionnel, non pas parce que c'est la première fois que je préside cette cérémonie mais parce qu'il y a devant moi cinq éminentes personnalités. .
Ces personnalités, c'est Mme Marie-José CHOMBART DE LAUWE, ensuite M. Pierre DAIX, M. Daniel CORDIER, Mme Andrée GROS et M. Jean-François GUTHMANN.
Remise des insignes de Grand'Croix de la Légion d'Honneur à Yvonne (dite Marie-José) CHOMBART DE LAUWE.

Encodage de ce type de discours

- Pas de problème particulier ;
  - `<title>`
  - `<p>`
  - `<div>`
- A l'exception peut-être de la description d'une action (remise des insignes...) en plein milieu du discours.
- Et à l'exception aussi des méta-données ; on a probablement envie de mettre un `<settingDesc>` dans le header du fichier

On s'approche de l'oral : la description de la situation

Attention : module *corpus* (language corpora)

```
<setting>
  <date>11 novembre 2012</date>
  <locale>palais de l'élysée</locale>
  <activity>Remise de décorations</activity>
</setting>
```

On décrit ici la situation dans laquelle s'effectue le discours ; on est dans `<profileDesc>`/`<settingDesc>`.

On continue de s'approcher de l'oral : une interview publiée

Le Figaro Magazine - Le moment est-il venu pour vous d'annoncer votre candidature?
Nicolas Sarkozy - J'ai dit que le rendez-vous approchait: il approche. Sous la Ve République, de tous les candidats potentiels à l'élection présidentielle, il en est un qui a plus de devoirs et moins de droits, c'est le Président..
-Savez-vous quand vous annoncerez votre décision?
-Si la question est de savoir si j'ai réfléchi, sur le fond et sur la forme, ma réponse est oui. Comme pour toutes les décisions importantes, je prends le temps d'une réflexion longue et approfondie.

Encodage de cette interview

- On a (probablement) envie de repérer qui dit quoi.
- On n'a toujours pas de l'*oral*, mais on souhaite pouvoir attribuer des segments à des locuteurs.
  - Il nous faudra donc une description des participants
  - Un repérage parmi ces participants

Décrire des participants

A la surprise générale : `<particDesc>` (on est toujours dans *corpus*
Et, toujours à la surprise générale, les participants sont des `<person>`, de préférence dans un `<listPerson>`
Exemple :

```
<particDesc>
  <person  xml:id="NS">
    <persName>
      <forename>Nicolas</forename>
      <surname>Sarkozy</surname>
    </persName>
  </person>
  <org  xml:id="FM">
    <orgName>Le figaro magazine</orgName>
    <desc>Ou plus probablement un journaliste de la
rédaction</desc>
  </org>
</particDesc>
```

On peut maintenant repérer nos participants

- Attribut : 'who' qui pointera à l'intérieur du `<particDesc>`
- Comme on n'a toujours pas du *vrai* oral, ça pourrait être `<sp>` + `<speaker>` (module 'core').

Exemple :

```
<sp  who="#NS">
  <speaker>Nicolas Sarkozy</speaker>
  <p>- J'ai dit que le rendez-vous approchait:  il
approche.  Sous la Ve République, de tous les candidats
potentiels à l'élection présidentielle, il en est un qui
a plus de devoirs et moins de droits, c'est le
Président.<gap/>.</p>
</sp>
```

Attribut ‘who’

Classe d’attributs ‘att.ascribed’
Les membres de cette classe d’attribut :

- Des éléments pour l’oral
- L’élement `<change>` (cf révisions)
- `<sp>` (théatre, discours rapporté direct
- `<move>` (théatre)

Ce qui nous manque pour avoir du *vrai* oral :

Jusqu'à présent, on a vu des exemples avec:

- des locuteurs
- une situation

mais on n'a pas encore vraiment de l'oral. On aurait pu regarder du théatre (on aurait eu des *didascali*) mais on n'aurait toujours pas eu de *vrai* oral.
Le *vrai* oral correspond aux cas où on *transcrit* de l'oral !
D'où le nom du module 'spoken' : *transcription of speech*.

Transcriptions d'oral

Quand on transcrit de l'oral, on souhaite que le contenu textuel soit la transcription de l'oral !
Ainsi, des descriptions d'évènements annexes (cf: le discours de Hollande et la remise des médailles) seront décrits (si on le souhaite), mais seront contenus dans des descriptions (`<desc>`).
Si on avait fait une transcription (à partir d'un enregistrement) du discours de Hollande, on aurait alors transcrit cette partie par :

```
<kinesic  who="#HO">
  <desc>Remise des insignes de Grand'Croix de la Légion
d'Honneur à Yvonne (dite Marie-José) CHOMBART DE
LAUWE.</desc>
</kinesic>
```

Quelques particularité de l'oral

- Contrairement aux exemples précédents, la notion de paragraphe n'a pas grand sens
- En revanche, la notion de `<div>` peut parfaitement faire sens
- techniquement, il faut quelque chose qui puisse aller dans une `<div>` et qui :
  - puisse être attribué à un locuteur
  - soit connexe dans le temps (on veut pouvoir lire la transcription)
- Le quelque chose en question s'appelle `<u>`

Un exemple minimal


```
<u  who="#LB">Bertrand tu es en retard</u>
<u  who="#BG">comme d'habitude</u>
<u  who="#LB">Oui comme d'habitude</u>
```

- le plus souvent, les transcriptions sont sans ponctuation (c'est un choix)
- il y aura besoin d'annotation supplémentaires (propres à la transcription de l'oral) qui seraient l'équivalent ; par exemple :
  - u de type 'interrogatif'
  - `<pause>`
  - etc.

Qu'est-ce qu'un <u>

Il n'y a pas forcément de consensus, parmi les (théories/transcripteurs/chercheurs) quand à ce que seraient des *tours de parole*, *interventions*, etc. La TEI est agnostique de ce point de vue.
un <u> est donc

- attribuable à un locuteur
- connexe (un début et une fin)

Une (mauvaise) transcription :

```
<u  who="#LB">Bertrand tu es en retard
<anchor  xml:id="a1"/>
  <anchor  xml:id="a2"/>Oui comme d'habitude</u>
<u  who="#BG"  start="#a1"  end="#a2">comme d'habitude</u>
```

Introduction, remarques préliminaires et évidences

Particularités de l'oral :

- organisé dans le temps
- des phénomènes particuliers à l'oral
  - (reprises, hésitations, etc.)
  - des annotations propres à l'oral (pauses, évènements vocaux (rires), gestes éventuels, etc.)

Oral et TEI

- Première constation : la TEI n'est actuellement que très peu utilisée par les communautés qui travaillent sur l'oral !
  Une raison essentielle :
  - les outils

  Travailler sur l'oral impose d'utiliser un outil pour la transcription. Aucun des outils courament utilisé ne donne de TEI directement en sortie.
- Deuxième constation : le module TEI consacré à l'oral date !

Pourquoi s'intéresser à la TEI et l'oral ?

(on y reviendra !)

- Interopérabilité !!!!
  - travail sur le texte de la transcription (avec l'éternelle question de qu'est-ce que le texte !)
  - en particulier, travail sur une transcription annotée !
- De plus en plus, on a envie d'un format pivot dans lequel on accumule des annotations.

le temps, et le reste...

Thomas Schmidt (journal TEI) propose de différentier :

- la macrostructure (i.e. l'alignement temporel)
- la micro-structure (le reste)

Je vais suivre plus ou moins ce plan, mais avant d'aligner temporellement, il faut savoir ce qu'on aligne !

- des points situés dans le temps (`<anchor>` en TEI)
- des intervalles

la balise <u>

Dans de l'oral, des locuteurs produisent des segments de parole. La caractérisation de ces segments n'est pas consensuelle !

- énoncés ?
- tours de parole ?
- propositions, actes illocutoire, etc.

comme (presque) toujours, la TEI est agnostique ! -> balise <u> à laquelle on n'impose formellement que :

- la connexité dans le temps (un <u> a un debut et une fin)
- l'attribution à un (parfois des) locuteurs
- en général... séparé du reste par du silence (mais ce n'est pas une nécessité absolue)

<u> suite

donc <u> a :

- un début et une fin (supposées)
- un locuteur (ou un groupe de locuteurs qui disent la même chose)
- contient quelque chose de locutoire (pas uniquement un geste par exemple)

morale : <u> peut faire penser à "utterance", mais certainement pas en surchangeant "utterance" d'un contenu théorique qu'elle ne prétend pas avoir
Si on souhaite faire dire plus à <u>, il faudra être explicite via une autre annotation, ou éventuellement dans le <encodingDesc>

Un élément "ponctuel" <anchor>

- <u> Désigne un segment temporel (un intervalle). Pour aborder la question de l'alignement temporel, on a également besoin d'un élement ponctuel.
- On a un tel élement en TEI : <anchor>. Evidemment, le modèle de contenu de <anchor> est *empty*

L'alignement temporel

- La TEI permet de représenter aussi bien des corpus alignés sur un signal sonore que de simples transcriptions.
- Une simple transcription pose déjà des problèmes d'alignement (chevauchements de parole)
- Le modèle permet donc la synchronisation au niveau de la transcription d'une part, et l'alignement sur le signal d'autre part (nécessite un outil de transcription).

Exemple (synchro):

```
<u  who="#L1">Je pense que ça fait
<anchor  xml:id="over"/> beaucoup</u>
<u  who="#L2"  start="#over">pas mal</u>
```

Deux types de transcriptions de parole

Globalement, on a donc deux types possibles de transcription :

- des transcriptions "faites à la main"
- des transcriptions alignées (temporellement) sur le signal. Concrètement, on sait dans ce cas qu'un énoncé donné se trouve à un temps précis dans le signal de parole

Ces deux types de transcription on en commun :

- Des annotations spécifiques à la parole (micro-structure)
- Une part de synchronisation (macro-structure)

D’un point de vue pratico-pratique

- La synchronisation sur le signal de parole se traduit dans le modèle tei via des scripts
- Les conventions de transcription sont souvent une autre histoire...

En Résumé :

- La plupart des outils, synchro temporelle OK et traduisible, conventions de transcription = problème
- *à la main*, conventions de transcription : OK, synchronisation = problème

Alignement temporel sur un signal

L'idée est d'avoir une `<timeline>` contenant des instants (`<when>`) sur laquelle on aligne des instants (start et end sur `<u>` et synch sur `<anchor>`)
Exemple :

```
<timeline>
  <when  xml:id="I1"/>
  <when  xml:id="I2"/>
  <when  xml:id="I3"/>
</timeline>
<u  who="#L1"  start="#I1"  end="#I2"/>
<u  who="#L1">
  <anchor  synch="#I3"/>
</u>
```

Organisation en TEI

Elements de la classe att.timed (ex : `<u>`). Ont un attribut 'start' et un attribut 'end'.
Ce sont les éléments :

- `<u>`
- `<pause>`
- `<kinesic>`
- `<incident>`
- `<writing>`

Remarque : 'synch' est défini dans Linking, segmentation and alignment (module linking) de même que `<seg>`, `<timeline>` et `<when>`.

Les modules pour l'oral

- spoken (broadcast, equipment, incident, kinesic, pause, recording, recordingStmt, scriptStmt, shift, u, vocal, writing)
- linking (anchor, timeline, when + att.global.linking -> synch)
- corpus pour définir le `<setting>` et `<settingDesc>`

Eléments ponctuels vs intervales

ponctuels :

- `<shift>`
- `<anchor>`

intervalles :

- `<u>`
- `<pause>`
- `<kinesic>`
- `<incident>`
- `<vocal>`

Les façons de faire...

Il y a beaucoup de possibilité de transcrire les corpus oraux, et ce à différents niveaux de finesse. cf. prop de Thomas d'aller vers une proposition commune TEI et ISO d'un sous-ensemble (une personalisation) de la TEI qui permette une interopérabilité plus facile entre formats de sortie d'outils.

Les éléments supplémentaires à la transcription

<u> contient la transcription elle-même. Pour autant, d'autres élements peuvent être transcrits :

- Des bruits produits à l'aide de l'apareil phonatoire : rire, toux, etc. Element : <vocal>
- Des bruits autres : <incident>
- Des gestes en général : <kinesic>
- Le fait de montrer un texte imprimé : <writing>

De plus, on peut noter les <pause>
Remarque : on obtient la transcription en ne considérant que les <u>, les autres élements contiennent (en général) un sous-élément <desc>

Faux départs, reprises, etc.

Beaucoup de systèmes de transcription adoptent des conventions relatives aux mots incomplets, aux faux départs, etc.
Dans un cadre TEI, ce type de phénomènes peut tomber sous le coup de deux types d'analyses :

- Déclarer explicitement si quelque chose est un mot ou pas
- Régulariser

Régularisations

L'élément TEI correspondant est `<choice>`. On peut y faire entrer typiquement `<orig>` et `<reg>`. Evidemment, rien n'empêche de typer les `<orig>` et `<reg>`. Rien n'empêche non plus de proposer une transcription phonètique.
Exemple :

```
<u>
  <choice>
    <orig>Y</orig>
    <reg>il</reg>
  </choice> m'a dit <choice>
    <orig>qui</orig>
    <reg>qu'il</reg>
  </choice> voulait venir
</u>
```

RQ : on peut se servir de `<orig>` et `<reg>` sans utiliser `<choice>`

Les faux départs, hésitations, etc.

La question est de savoir exactement ce qu'on veut en faire

En général on peut vouloir :

- les marquer (les annoter comme des faux départs
- les régulariser pour faciliter le passage d'outils

La solution (habituelle en TEI), les supprimer... sans les supprimer ! On les marque donc par une opération d'édition.

```
<u who="#BG">je <del cause="repetition">je</del> me
demande ce qu'il en est</u>
```

Les conventions de transcription

La plupart des outils de transcription gèrent essentiellement l'alignement temporel ; Du coup, des convenstion de trnscription orthographiques se sont développées.
Exemples :

- En France, conventions DELIC (exemple : pause = +, ++ ou +++
- Conventions HIAT
- etc.

Problème : il n'est pas toujours simple (ni même possible) de traduire automatiquement en TEI (ou de passer automatiquement des unes aux autres).

Exercices

Prendre la version texte du débat Royal/Sarkozy (corpus de démo d'Exmaralda) et l'encoder en TEI
Au choix : prendre la version TEI et la mettre en TEI !

Autre exercice

Aller chercher un corpus sur TCOF ; on a les méta données (en html) -> les encoder en TEI, on peut aussi récupérer la version transcriber, le wav et les conventions de transcription