

TEI et analyses linguistique

January 2010



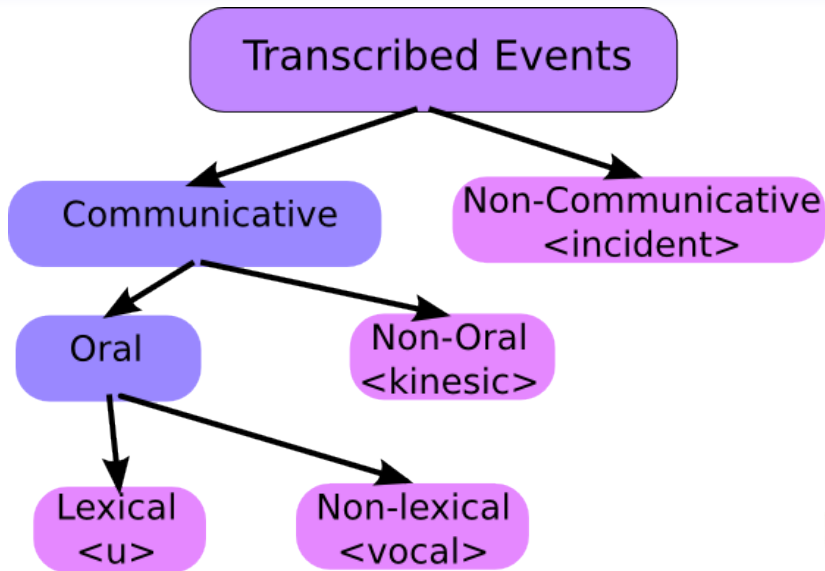
L'oral

Dans un texte oral on peut reconnaître :

- des énoncés (utterances)
- des pauses
- des phénomènes vocalisés mais pas lexicalisés, par exemple la toux, des quasi-mots comme "hein", "uh" etc.
- des phénomènes *kinésiques* (non vocalisés, non lexicaux, mais servant à communiquer) notamment les gestes
- des événements entièrement non linguistiques, mais ayant un effet sur le discours, par exemple un camion qui passe etc.
- des événements écrits, que l'on peut donc transcrire, par exemple des titres ou des diapositives affichés pendant une communication
- des changements dans la qualité de la voix, par exemple de volume

Un sous-ensemble des phénomènes de discours qui sont d'utilité
lexicale

En résumé...



Propositions du module spoken

Des éléments supplémentaire dans l'en-tête <broadcast>,
<equipment>, <recording>, <recordingStmt>,
<scriptStmt>

Des éléments pour la transcription de l'oral <incident>, <kinesic>,
<pause>, <shift>, <u>, <vocal>, <writing>

Des attributs supplémentaires @dur@start@end@sync etc.

Par exemple...

```
<u who="#Jan">mmm délicieux</u>
<incident>
  <desc>téléphone sonne</desc>
</incident>
<u who="#Kim">j'y vais</u>
<u who="#Tom">ya longtemps <vocal>
  <desc>tousse</desc>
</vocal> jne
fume plus</u>
<u who="#Bob">
  <vocal>
    <desc>sniffs</desc>
  </vocal>il se croit dur
</u>
<vocal who="#Ann">
  <desc>grognement</desc>
</vocal>
<u who="#Tom">oueh
<kinesic>
  <desc>fait un geste avec le doigt</desc>
</kinesic>
</u>
<u who="#Bob">donc j'aurais dû <vocal who="#Ann">
  <desc>faisant tss-tss</desc>
  > faire quoi</u>
```

Le concept d'"énoncé"

- une séquence de discours d'un seul locuteur
- peut être regroupé dans des sections `<div>`
- peut être fragmenté dans des segments `<seg>` ou `<s>`
- L'attribut `@who` sert à indiquer le locuteur

Exemple utilisant d'autres éléments TEI

```
<u who="#mar">you never <pause/> take this cat for  
show and tell  
<pause/> meow meow</u>  
<u who="#ros">yeah well I dont want to</u>  
<incident>  
  <desc>toy cat has bell in tail which continues  
    to make a tinkling sound</desc>  
</incident>  
<u who="#ros">because it is so old</u>  
<u who="#mar">how <choice>  
  <orig>bout</orig>  
  <reg>about</reg>  
</choice>  
<emph>your</emph> cat <pause/>yours is <emph>new</emph>  
<kinesic>  
  <desc>shows Father the cat</desc>  
</kinesic>  
</u>  
<u trans="pause" who="#fat">thats <pause/> darling</u>  
<u who="#mar">no <emph>mine</emph> isnt old  
mine is just um a little dirty</u>
```

Changements de voix

- Problème de chevauchement classique !
 - on peut se servir de la balise `<shift>` comme `<milestone>` pour indiquer les frontières...
 - ... ou des `<seg>` typés
- applicable également au "code shifting"

```
<u who="#LB">
  <shift feature="loud" new="f"/>Elizabeth
</u>
<u who="#EB">Yes</u>
<u who="#LB">
  <shift feature="loud"/>Come and try this <pause/>
  <shift feature="loud" new="ff"/>come on
<shift feature="code" new="fr-mru"/> 'tin va!

</u>
<listPerson type="speakers">
  <person xml:id="LB"/>
  <person xml:id="EB"/>
</listPerson>
```


Liste non exhaustive de caractéristiques en prose

(basée sur Boase, Survey of English Usage, 1990)

tempo	rapide, lent, de plus en plus rapide, de plus en plus lent, etc.
volume	fort, faible, de plus en plus fort, de plus en plus faible
hauteur	aigu, grave,...
tension	lié, tendu, staccato, legato...
rythme	régulier, irrégulier...
qualité de la voix	murmures, voix enrouée, voix de fausset, gloussements, sanglots, bâillements, soupirs...

Les chercheurs ont besoin de définir leur propre terminologie

<shift/> : exemple

```
<u who="#a">écoutez <shift new="reading"/>Matignon se déclare  
confiant que les problèmes financiers actuels seront  
entièrement maîtrisés fin juin<shift/> mon cul</u>
```

Ou bien :

```
<u who="#a">écoutez  
<incident>  
  <desc>lit à haute voix du journal</desc>  
</incident>mon cul</u>
```

<writing> exemple

```
<u who="#a">regardez ceci</u>  
<writing who="#a" type="newspaper" gradual="false">  
Matignon se déclare  
<soCalled>confiant de maîtriser</soCalled> les problèmes financiers actuels  
</writing>  
<u who="#a">mon cul!</u>
```

Questions relatives à la temporalité

- pour les pauses : élément `<pause>`
- pour la durée : attribut `@dur`
- synchronisation : attribut `@synch`
- chevauchement : attribut `@trans`

<pause> : exemple

<u>Okay <pause dur="PT2M"/>U-m<pause dur="PT75S"/>la scène ouvre
<pause dur="PT50S"/> avec <pause dur="PT20S"/> um <pause dur="PT145S"/>
on voit un arbre okay?</u>

Chevauchement

Mutt: vous avez entendu l - -
Jeff: les résultats?
Mutt: quel désastre !
Jeff: (en même temps) quel miracle!

```
<u who="#mutt">vous avez entendu l</u>  
<u trans="latching" who="#jeff">les résultats</u>  
<u who="#mutt">quel désastre</u>  
<u who="#jeff" trans="overlap">quel miracle </u>
```

Synchronisation

```
<u who="#mutt">vous avez entendu <anchor synch="#t1"/>l</u>
<u who="#jeff" synch="#t1">les résultats</u>
<u who="#mutt" synch="#t2">quel désastre</u>
<u who="#jeff" synch="#t2">quel miracle</u>
<!-- Elsewhere in Document -->
<timeline origin="#t1">
  <when xml:id="t1" since="00001728281"/>
  <when xml:id="t2" since="00001728302"/>
</timeline>
```

<timeline> : exemple

```
<timeline unit="s" origin="#TS-P1">
  <when xml:id="TS-P1" absolute="12:20:01"/>
  <when xml:id="TS-P2" interval="4:05" since="#TS-P1"/>
  <when xml:id="TS-P6"/>
  <when xml:id="TS-P3" interval="1:05" since="#TS-P6"/>
</timeline>
```

```
<u xml:id="TS-U1" start="#TS-P2" end="#TS-P3">This is my
<anchor synch="#TS-P6" xml:id="TS-P6A"/> turn</u>
```

Le début de l'énoncé TS-U1 s'aligne avec le point temporel TS-P2, il est donc 4.5 unités après TS-P1, i.e. à 12:24:06. Sa fin est synchronisée avec le point temporel TS-P3. La transition entre les mots 'my' et 'turn' arrive à un moment synchronisé avec le point temporel TS-P6.

Description des participants

```
<particDesc>
  <listPerson>
    <person xml:id="P-1234" sex="2" age="mid">
      <p>informateur, sexe féminin, bonne éducation, née à Shropshire UK, 12
Jan 1950,
        commerçante parle français couramment. Statut
        socio-économique (SSE) :
        commerçante.</p>
    </person>
    <person xml:id="P-4332" sex="1">
      <persName>
        <surname>Hancock</surname>
        <forename>Antony</forename>
        <forename>Aloysius</forename>
        <forename>St John</forename>
      </persName>
      <residence notAfter="1959">
        <address>
          <street>Railway Cuttings</street>
          <settlement>East Cheam</settlement>
        </address>
      </residence>
      <occupation>comedian</occupation>
    </person>
  </listPerson>
</particDesc>
```

Description des enregistrements

```
<recordingStmt>
  <recording type="audio" dur="P10M">
    <equipment>
      <p>podcast</p>
    </equipment>
    <broadcast>
      <bibl>
        <title>Questions sur la souffrance et la santé au travail :
pénibilité, stress,
          dépression, harcèlement, maladies et accidents...</title>
        <author>France Inter</author>
        <respStmt>
          <resp>Présentateur</resp>
          <name>Alain Bédouet</name>
        </respStmt>
        <respStmt>
          <resp>Personne interrogée</resp>
          <name> Marie Pezé, Docteur en psychologie, psychanalyste, expert
judiciaire ; dirige
            la consultation « souffrance et travail » à l'Hôpital de
Nanterre (92), auteure de
              <title>ils ne mourraient pas tous mais tous étaient frappés</title>,
Editions
                Pearson.</name>
            </resp>
        </resp>
      </bibl>
    </broadcast>
  </recording>
</recordingStmt>
<series>
```

Ou peut-être plus simplement...

```
<recordingStmt>
  <recording type="audio" dur="P15M" xml:id="rec-3001">
    <date>14 Feb 2001</date>
  </recording>
  <recording type="audio" dur="P15M" xml:id="rec-3002">
    <date>17 Feb 2001</date>
  </recording>
  <recording type="audio" dur="P15M" xml:id="rec-3003">
    <date>22 Feb 2001</date>
  </recording>
</recordingStmt>
```

...et pour le contexte

```
<setting xml:id="KDFSE002" n="063505" who="#PS0M6">  
  <name type="place">Lancashire: Morecambe </name>  
  <locale> at home </locale>  
  <activity> watching television </activity>  
</setting>
```