

Conventions de l'annotation des reformulations dans les transcriptions de l'oral

L'annotation se fait sous format XML.

Éléments du document XML

- **Marqueur** : <MRP> (voir le document avec la typologie des marqueurs)
- **Catégories syntaxiques des segments paraphrasés** : <N>, <A> etc.

catégorie seule : *N*, *A*, *V*, *Prep*, *Pr* (pronom), *PRES* (présentateur : *il y a*, *c'est*, *voici*, etc), *P* (phrase/proposition), *Adv*, *Ger*,

catégorie avec des modificateurs : *NP*, *VP*, *AP*, *PP*

*Différence entre N/NP : *N* pour les noms sans déterminants et NP pour tout le reste, pareil pour les autres catégories syntaxiques

* La catégorie syntaxique annotée est suivie d'un numéro :

1 - pour indiquer qu'il s'agit d'un segment source qui va être paraphrasé, il se trouve avant le marqueur

2 - pour indiquer qu'il s'agit d'un segment paraphrasé qui se trouve après le marqueur

Attributs du document XML

Les attributs se mettent sur le deuxième segment, c'est-à-dire que c'est la catégorie syntaxique du deuxième segment qui possède les attributs.

- **rel_lex="hypero / syno / hypo / ant / instance / mero (mot1/mot2) "**

L'attribut rel_lex montre les relations que peuvent entretenir les unités lexicales d'un segment 1 (S1) avec segment 2 (S2). Les valeurs possibles sont :

- **hypero** : hyperonymie lorsqu'un mot ou une séquence de mots d'un S1 est un hyperonyme d'un mot ou d'une séquence de mots d'un S2. IL s'agit des relations hiérarchiques du type : sorte_de

```
<NP1>des pièces de théâtre qui passent</NP1> <MRD>des euh</MRD>
<NP2 modif_lex="rempl(pièces de théâtre/spectacles) suppr(qui passent)" rel_lex="hypero(pièces de théâtre/spectacles)" rel_pragm="gener">des spectacles</NP2> on va plutôt euh euh
```

- **hypo** : hyponymie, l'inverse du cas précédent, lorsqu'un mot ou une séquence de mots d'un S1 est un hyponyme d'un mot ou d'une séquence de mots d'un S2.
- **syno** : synonymie, un mot ou une séquence de mots d'un S1/S2 est un synonyme d'un mot ou d'une séquence de mots d'un S2/S1

*Parfois, le trait de synonymie entre les deux unités lexicales n'est pas évident mais on le met quand même car c'est comme cela qu'il est perçu par le locuteur :

```
<API>basée euh sur le capitalisme</API> enfin la société française
<MDR>disons</MDR> euh euh <AP2 rel_lex="syno(capitalisme/ valeurs erronées)"
```

modif_lex="remp(capitalisme/valeurs erronées)" rel_pragm="prec">basée sur les valeurs euh euh erronées</AP2>

- **anto** : antonymie lorsqu'un mot ou une séquence de mots d'un S1/S2 est un antonyme d'un mot ou d'une séquence de mots d'un S2/S1

<P1>on est bien détendu</P1> <P2
 modif_lex="remp(détendu/énervé) ajout(de la de sa journée)
 ajout(ne pas)" rel_pragm="oppos"
 rel_lex="anto(détendu/énervé)">on n'est pas énervé de la de sa
 journée</P2>

- **mero** : meronymie, lorsqu'un mot ou une séquence de mots d'un S1/S2 est un meronyme d'un mot ou d'une séquence de mots d'un S2/S1. Généralement il s'agit d'une relation partie_de mais nous y avons ajouté une relation plus large fondée sur les associations que les mots peuvent avoir entre eux
- **instance** : lorsqu'un mot ou une séquence de mots d'un S1/S2 est une instance d'un mot ou d'une séquence de mots d'un S2/S1. Il s'agit souvent des entités nommées (noms propres ou groupes nominaux) qui permettent de référencer un élément

- **rel_pragm="def/denom/explic/justif/exempl/prec/res/para"**

L'attribut rel_pragm montre les relations que peuvent entretenir les unités lexicales d'un segment 1 (S1) avec le segment 2 (S2) au niveau pragmatique. Il s'agit d'une fonction pragmatique qui répond à la question pour quelle raison le locuteur fait le recours à une reformulation. Les valeurs possibles sont :

- **def** : définition, lorsqu'un terme (simple ou composé) dans un S1 est défini dans une S2. Il s'agit souvent des termes spécialisés du domaine.
- **explic** : explication, lorsque le locuteur essaie d'expliquer qqch à son interlocuteur. Pour vérifier cette fonction, on peut remplacer le marqueur par *parce que*
- **justif** : justification. Cette fonction ressemble à la précédente mais dans ce cas, le locuteur présuppose que ce qu'il a dit dans S1 peut être mal perçu par son interlocuteur, il doit donc le justifier. On peut remplacer le marqueur, dans certains cas, par *c'est pourquoi*
- **prec** : précision. Une fonction assez large dont les cas sont nombreux. Il s'agit d'une envie du locuteur d'ajouter une information dans le but de l'éclaircir. lorsqu'on met qqch de plus petit=précis (=>méronymie)
- =>**prec_qualif** : précision sur la qualification, le locuteur qualifie ce qui a été dit
 puis après euh on a bon on est confronté évidemment <PP1>aux
 jeunes</PP1> <MRD>euh</MRD> <PP2
 modif_lex="ajout(délinquants)" rel_lex="hypo(jeunes/ jeunes
 délinquants)" rel_pragm="prec_qualif">aux jeunes
 délinquants</PP2>
 =>**prec_quant** (précision quantitative)
 l'hiver non bah <PRES1>y a beaucoup</PRES1> <PRES2
 modif_lex="remp(beaucoup/de plus en plus)"
 rel_lex="syno(beaucoup/de plus en plus)"
 rel_pragm="prec_quant">y a de plus en plus</PRES2> de fraudes

- **res** : résultat. Le locuteur résume le contenu du S1 ou donne en abrégé une conséquence de ce qui a été dit dans le S1. Le marqueur peut être remplacé dans ce cas pour *en somme*, *en bref*.
- **exempl** : exemplification, lorsque le locuteur donne un exemple dans un S2 d'une entité mentionné dans le S1. Il peut s'agit d'une entité nommé éventuellement.
- **dénom** : dénomination. C'est le cas qui ressemble au précédent. La différence est que contrairement à l'exemple où l'existence des autres entités du même type est présumée, ici la fonction consiste dans l'attribution d'un nom à une entité unique (agglomération=>Orléans)
- **para** : paraphrase lorsqu'on ne constate aucune différence sémantique entre les deux segments.
- **oppos** : opposition (lorsqu'il s'agit de l'opposition sémantique et pas juste l'ajout / la suppression de négation ne...pas)

```
<PRES1>c'est      pas      un      concours</PRES1>      <PRES2
modif_lex="suppr(pas)      rempl(un      concours/un      examen)"
rel_lex="hypo" rel_pragm="">c'est un examen</PRES2>
<P1>on      est      bien      détendu</P1>      <P2
modif_lex="rempl(détendu/énervé) ajout(de la de sa journée)
ajout(ne      pas)"      rel_pragm="oppos"
rel_lex="anto(détendu/énervé)">on n'est pas énervé de la de sa
journée</P2>
<P1>qui      a      douze      ans      bientôt      le      dix      avril</P1>
<MRCONS>donc</MRCONS>      <P2      modif_lex="rempl(qui/il)
ajout(pas) suppr(bientôt) suppr(le dix avril) ajout(encore)"
rel_pragm="oppos">il a pas encore douze ans</P2> quand je dis
qu'il a douze ans il est pas content oui
```

*Les relations pragmatiques (rel_pragm) doivent être toujours annotées.

- **modif_morph="flex/deriv"**

L'attribut modif_morph montre les modifications ou les liens que les unités lexicales d'un segment 1 (S1) et d'un segment 2 (S2) entretiennent au niveau morphologique. Les valeurs possibles sont :

- **flex** : flexion, lorsqu'un mot d'un S1 réapparaît dans un S2 sous une autre forme fléchie.
- **deriv** : dérivation, lorsque les S1 et S2 contiennent des mots de la même famille morphologique

- **modif_synt="passif/actif"**

L'attribut modif_synt montre les modifications qui se font entre les S1 et S2 au niveau syntaxique. Il s'agit (pour l'instant) que d'une type de modification passif/actif ou actif/passif. Une seule valeur possible est :

- passif/actif

Remarques supplémentaires

* On peut rencontrer plusieurs reformulations dans le même énoncé. Dans ce cas,

- chaque segment reçoit le numéro par ordre apparition.

*esl01-059_B-u8.txt|3|disons|enfin je enfin je enfin euh <ADV1>personnellement</ADV1>
<MDR>disons</MDR> le <P2 rel_lex="syno(personnellement/où je suis)" rel_pragm="prec"> dans
la branche euh où où je suis</P2> pour l' instant ce qui m' intéresse surtout c' est qu' il y a enfin de
vous voyez enfin le le règlement euh le les modifications le le les circulaires qui passent euh
<NP3>les points euh le- litigieux</NP3> <MDR>disons</MDR> enfin <PRES4
rel_lex="hypero(les points le litigieux/travail de recherche)" rel_pragm="justif"> il y il y a pas mal
de euh de travail euh de recherche</PRES4> <MDR>disons</MDR> euh dans ce ce dans ce
domaine-ci quoi c' est euh c' est surtout ça qui est intéressant parce que sinon enfin euh
<VP5>contrôlait le les dossiers</VP5> euh seulement <MDR>disons</MDR> c' est d' un <PRES6
rel_lex="hypo(contrôlait les dossiers/travail de routine)" rel_pragm="prec">c' est un travail de de
routine</PRES6> pratiquement*

- si les reformulations se chevauchent, on redouble l'énoncé

1308.657|1319.654|1'hiver non bah <PRES1>y a beaucoup</PRES1> <PRES2
modif_lex="rempl(beaucoup/de plus en plus)"
rel_lex="syno(beaucoup/de plus en plus)" rel_pragm="prec_quant">y a
de plus en plus</PRES2> de fraudes les gens fraudent de plus en plus ce
qui ce qui peut être compréhensible hein vu euh vu le prix des euh le
tarif des euh des billets de train c'est vrai qu'on a de plus en plus
de fraudeurs en costume cravate
1308.657|1319.654|1'hiver non bah y a beaucoup y a <NP1>de plus en plus
de fraudes</NP1> <P2 modif_morph="deriv(fraudes/fraudent)"
rel_pragm="para">les gens fraudent de plus en plus</P2> ce qui ce qui
peut être compréhensible hein vu euh vu le prix des euh le tarif des
euh des billets de train c'est vrai qu'on a de plus en plus de
fraudeurs en costume cravate
ou
1507.04|1512.546|bah <NP1>les chiens d'Orléans</NP1> <NP2
rel_lex="syno(les chiens d'Orléans/ des gens froids)"
rel_pragm="def">des gens froids<NP2> pas accueillants euh
1507.04|1512.546|bah les chiens d'Orléans des gens <Adj1>froids<Adj1>
<Adj2 modif_lex="rempl(froids/pas accueillants)"
rel_lex="syno(froids/pas accueillants)" rel_pragm="para">pas
accueillants</Adj2> euh

* on n'annote pas les reformulations dans les tours de paroles différents :

1692.025|1703.547|hm hm bah j'ai une question justement euh euh à
propos des jeunes euh on y revient et donc euh on va dire que hm dans
la moyenne globalement est-ce que vous trouvez qu'on parle bien ?
1703.547|1711.427|euh dans l'ensemble euh ou- les trente euh les trente
et plus oui