



# EXEMPLIFICATIONS AVEC ET SANS MARQUEURS : ÉTUDE DE L'ORAL ET DE L'ÉCRIT

Elodie ESTEVES  
Yidi HUANG  
Juliette MASSY  
Aliénor SAUTY DE CHALON



# **Reformulations et exemplifications, définition et mise en perspective :**



## Les reformulations, définition :

**Définition :** procédé de modification dans le discours d'un segment source (S1) par un segment reformulé (S2).

S1 (MARQUEUR | PHASE D'ÉDITION) S2

Exemples :

- 1) L'endométriose **est** une maladie.
- 2) je compte bien aller au bout des choses qu ' il me dise si oui ou non ma  
douleur venait de l ' hernie!!!



## Caractéristiques :

- formes différentes selon la modalité (orale/écrite)
- phase d'édition/marqueur peut être absent<sup>1</sup>,  
=> ce n'est donc pas sur sa présence que repose une reformulation **mais sur le lien sous-jacent qui existe entre le *Segment 1* et le *Segment 2*.**
- ici la définition englobe les définitions, les dénominations, les explications, **les exemplifications**, les paraphrases, etc...

<sup>1</sup>Eshkol-Taravella, I., & Grabar, N. (2018). Reformulations avec et sans marqueurs : étude de trois entretiens de l'oral. *SHS Web of Conferences*, 46, 11003.  
<https://doi.org/10.1051/shsconf/20184611003>



## Exemplifications :

**Définition :** reprise d'un énoncé antérieur par un second segment contenant des exemples.

**Exemple :**

- 1) tu peu rien faire même pas ramasser un simple stylo tomber par terre
- 2) avec euh ce qu'il faut euh une douche euh sur place

**En quoi l'étude statistique comparative  
des exemplifications révèle-t-elle des  
tendances selon les contextes d'émission ?**

—



# **Données et corpus : annotation XML et prétraitement :**



## Nature des données : 2 modalités

**Oral** : 3 transcriptions d'entretiens oraux entre locuteurs témoins et chercheurs à propos de la vie des locuteur à Orléans

- Eslo 33
- Eslo 53
- Eslo 5

**Écrit** : deux fils de messages postés sur des forums de santé :

- 1 avec messages postés par des patients sur un forum consacré aux douleurs de dos
- 1 composé de réponses de médecin à des patients sur le forum *MaSante.net*





## Prétraitement du corpus :

### Correction de l'annotation :

- ajout de balises manquantes et création d'éléments `<reformulation>` pour les document de l'oral,
- ajout de guillemets manquants pour les attributs,
- modification des noms de balises quand un couple de balises ne correspondait pas.

### Mise en forme XML :

- ajout d'un élément racine `<body>`,
- modification du nom de l'attribut quand une balise possédait deux attributs identiques (exemple : `rel_pragm`, `rel_pragm2`),
- l'ajout d'attributs `mod` et `corpus` aux balises `<reformulation>` : distinguer la modalité (*mod*) *oral/écrit* et leur corpus d'origine (*corpus*) *eslo55, eslo33, eslo5, masante, patient*.



# **Modèle d'annotation avec balises XML des données**

# Modèle d'annotation, un aperçu :

## Catégories syntaxiques des segments :

### *Catégories avec modifieurs :*

<PP\*, NP\*, AP\*, VP\*>

### *Catégories seules :*

<A\*, ADV\*, N\*, PRES\*, Pr\*, Ger\*, P\*>

\* = le numéro du segment (généralement 1 ou 2)

## Marqueurs :

- **<MR>** :

→ Quand il n'y en a pas la balise est auto-fermante **<MR/>**.

- **<MRE>** = Marqueurs d'exemplification
- **<MRDENOM>** = Marqueurs de dénomination pour les verbes comme *appeler, nommer*
- **<MRDESIGN>** = Marqueurs de dénomination pour les verbes (pour les expressions de désignations).
- **<MRP>** = paraphrase
- **<MRCONC>** = concluants
- **<MRCOO>** = coordination,
- **<MRCOR>** = (correction).

Les marqueurs de disfluece : DA (amorces), DH (hésitations), DI (interjections), DMD (marqueurs discursif)

# Modèle d'annotation, un aperçu : les relations entre S1 et S2 :



Les relations entre le Segment 1 et le Segment 2 sont notées dans la balise du S2 sous forme d'**attributs**.

- 1) <PRES1>c'est pas tout à fait repos</PRES1> <MR /> <PRES2  
**rel\_lex="anto(repos/beaucoup papiers)" rel\_pragm="explic"**>c'est beaucoup  
papiers</PRES2>  
=> (*reformulation extraite du corpus oral*)

Il y a deux types d'attributs : **les relations et les modifications**

*NB : il peut y avoir un troisième segment*

- 2) <PRES3 rel\_lex="syno(pas pareil/plat)" rel\_pragm="explic">c'est plat</PRES3>.

# Modèle d'annotation, un aperçu : les relations entre S1 et S2 :

## 1. Les relations : deux types :

**Relations lexicales** : hyperonymie / hyponymie / antonymie / synonymie / instance / meronymie.

`<A2 rel_lex="hypero(neurologiques/neuro-psychiatriques)">`


**Relations pragmatiques** : définition / dénomination / explicitation / justification / **exemplification** / précision / résultat / paraphrase / opposition.

`<PP2 rel_pragm="prec">avec la troupe de Boulogne-Billancourt</PP2>`

## 2. Les modifications : deux types :

**Modifications morphologiques** : flexion / dérivation

**Modifications syntaxiques** : passif / actif



# Extraction automatique des reformulations :

```

with open(fichier_traite, 'r') as t:
    text = t.read()
    reformulations = re.findall(r'<reformulation>(.*?)</reformulation>', text, re.DOTALL)

```

```

1 <body>
2 <reformulation mod="oral" corpus="eslo5"><NP1>seize ans</NP1> <NP2 rel_pragm="prec">dix sept ans</NP2></
  reformulation>
3 <reformulation mod="oral" corpus="eslo5"><P1>mon mari a été muté dans la région</P1> on est de Paris <DMD>et
  voilà</DMD> <P2 rel_lex="mero(la région/Orléans)" rel_pragm="prec">il a été muté euh à Orléans</P2></reformulation>
4 <reformulation mod="oral" corpus="eslo5"><VP1>on avait demandé</VP1><MR /> <VP2 rel_lex="mero(demandé/voulait)"
  rel_pragm="explic">on voulait</VP2></reformulation>
5 <reformulation mod="oral" corpus="eslo5"><PP1>sans à priori</PP1> <MR /><PP2 rel_lex="syno(à priori/préjugé)"
  rel_pragm="para">sans sans préjugé</PP2></reformulation>
6 <reformulation mod="oral" corpus="eslo5"><PP1>à Orléans</PP1><MR /> <NP2 rel_lex="mero(Orléans/sud limite Olivet)"
  rel_pragm="prec">sud limite Olivet</NP2></reformulation>
7 <reformulation mod="oral" corpus="eslo5"><P1>on avait choisi Olivet</P1> nos enfants étaient scolarisés là
  <MRCONC>donc</MRCONC> <DH>euh</DH> <DMD>voilà</DMD> <P2 modif_morph="flex(avait/a)" rel_pragm="para">on a choisi de
  vivre à Olivet</P2></reformulation>

```



## Deux angles de travail :

### ANGLE 1:

ORAL

VS

ECRIT

**Module :** Etree, Matplotlib

### ANGLE 2:

ORAL

VS

**ECRIT :** Forum douleurs de dos (patients)

VS

**ECRIT :** Forum MaSante.net (médecins)

**Module :** Regex, Matplotlib



---

# PREMIÈRE APPROCHE:

ORAL

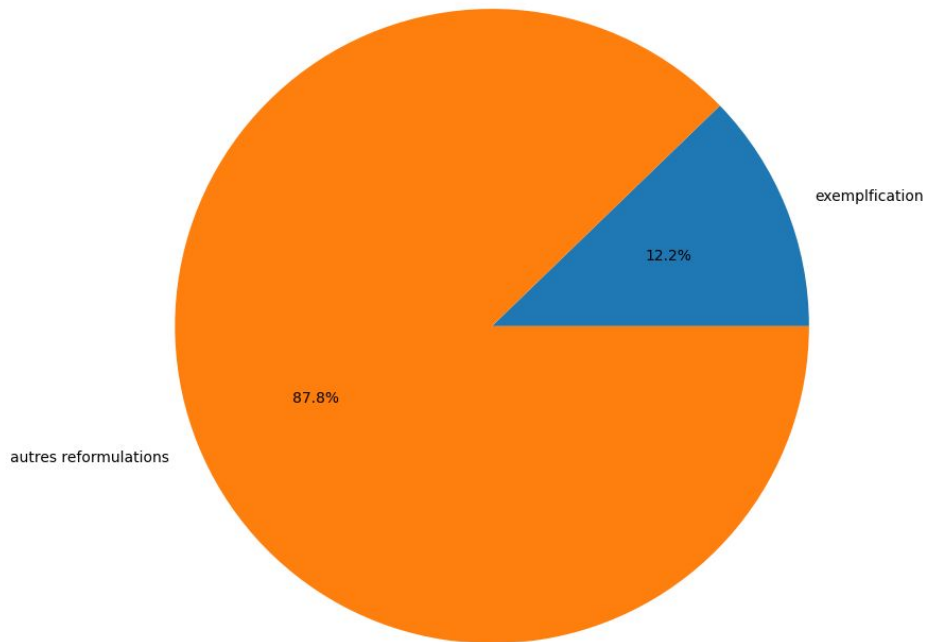
VS

ECRIT

# Les exemplifications, sous ensemble des reformulations :

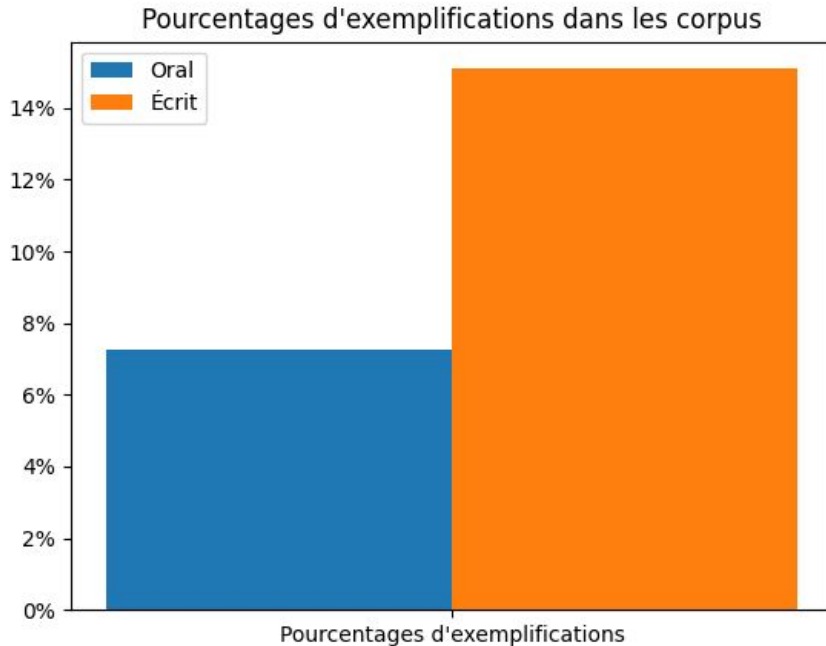


Proportion d'exemplifications dans le corpus total



Nombre total d'exemplifications dans le corpus : 108

# Pourcentage d'exemplifications dans les corpus oraux et écrits :



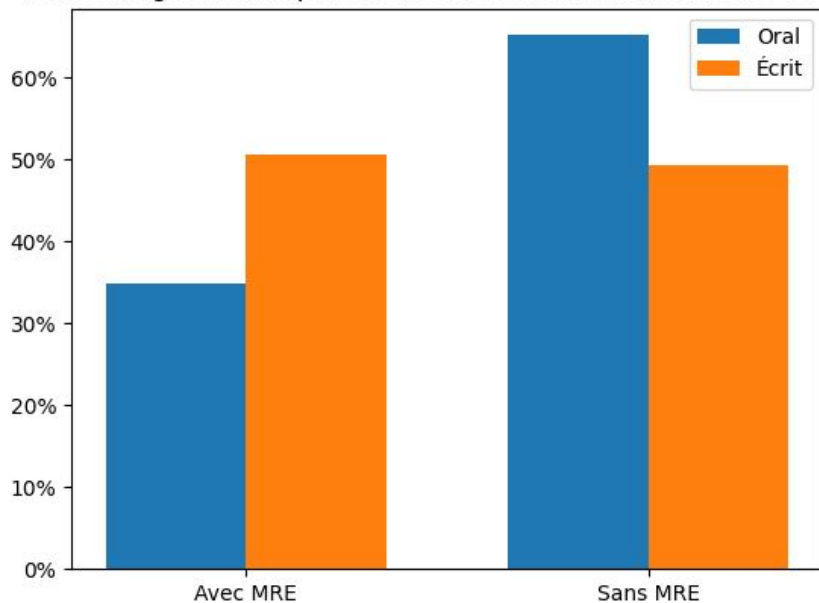
Nombre d'exemplifications à l'écrit : 85

Nombre d'exemplifications à l'oral : 23

Corpus	Proportion relative d'exemplification dans chaque corpus
Ecrit	15,09%
Oral	7,27%

# Les marqueurs d'exemplifications :

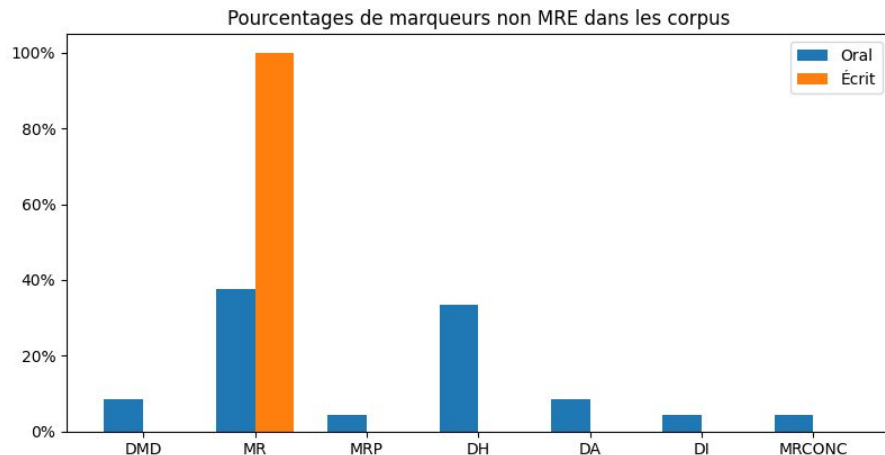
Pourcentages d'exemplifications avec et sans MRE dans les corpus



Exemples de <MRE> :

- écrit : ['type', 'genre', 'par exemple', 'du genre', 'de type', 'dont', 'comme', 'du type', 'en particulier', 'est l'un de ceux-ci', 'notamment', 'et en particulier,']
- oral : ['type', 'par exemple', 'genre', 'style']

# Les autres marqueurs dans les exemplifications :



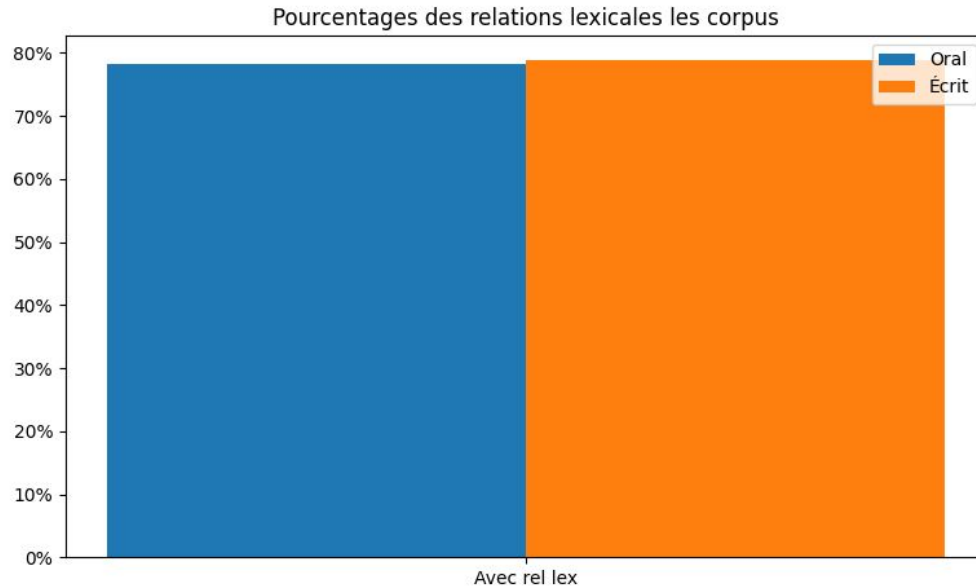
## Corpus écrit :

- <MR> = ponctuation (virgule, point, deux points, etc...) ou verbe être (présentatif)
- **Aucun marqueur**

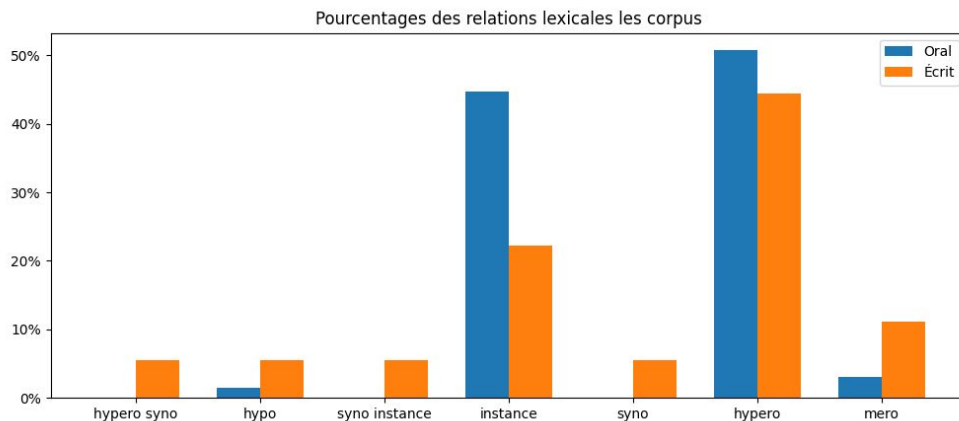
## Corpus oral :

- disfluences, notamment des hésitations (DH), apparaissent parfois avec autres marqueurs
- <MR> de type présentatif
- **Aucun marqueur**

# Les relations lexicales dans les exemplifications :



# Les relations lexicales, nature et répartition :



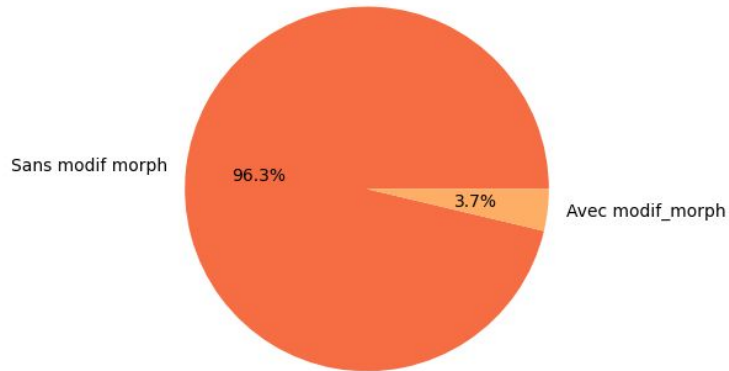
Globalement :

- Majorité d'hyperonymie et d'instances :
  - 1) mes boulots d'été assez physiques maraîchage, fruit et légumes en grande surface
  - 2) les antifongiques **genre :** Econazole, Bifonazole, etc.

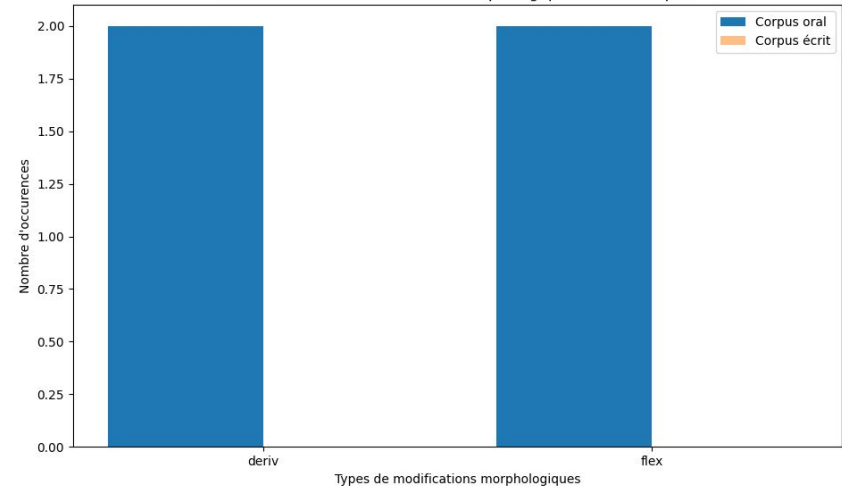
# Les modifications morphologiques, nature et répartition :



Distribution des modifications morphologiques



Distributions des modifications morphologiques dans le corpus



Ex : faut faire tout le nettoyage où euh je vide toute la vitrine et je nettoie tout et je refais



---

## SECONDE APPROCHE :

ORAL

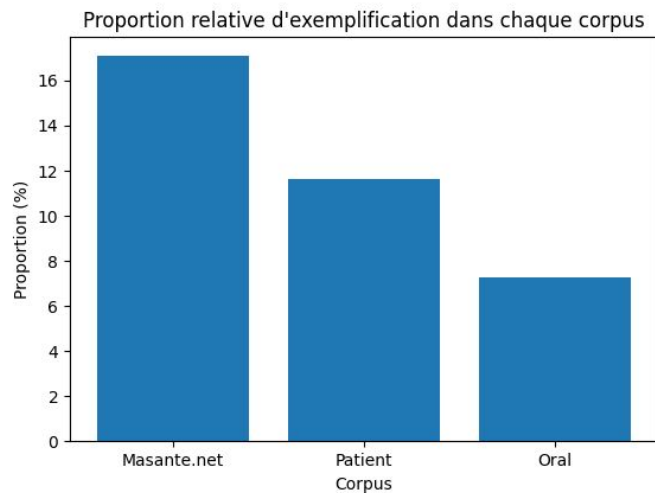
VS

ECRIT : Forum douleurs de dos (patients)

VS

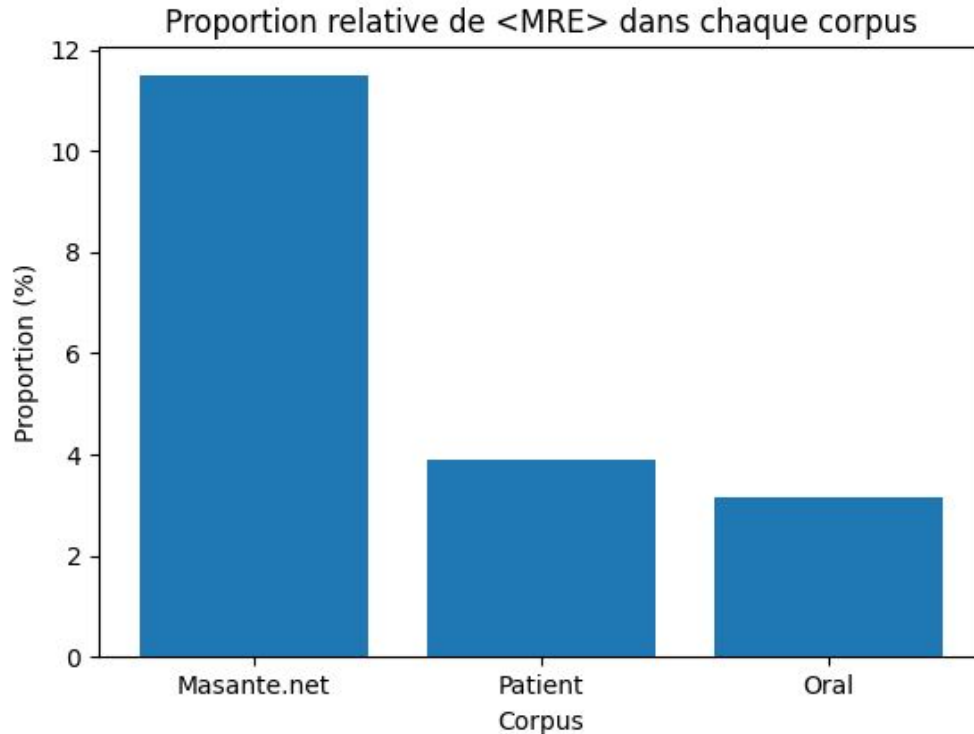
ECRIT : Forum MaSante.net (médecins)

# Les exemplifications, sous ensemble des reformulations :



Corpus	Proportion relative d'exemplification dans chaque corpus
Masante.net	17,08%
Patient	11,65%
Oral	7,27%

# Le corpus patient, plus proche de l'oral ?



Corpus	Proportion relative de <MRE> dans chaque corpus
Masante.net	11,48%
Patient	3,88%
Oral	3,16%

## Nature des marqueurs <MRE> :

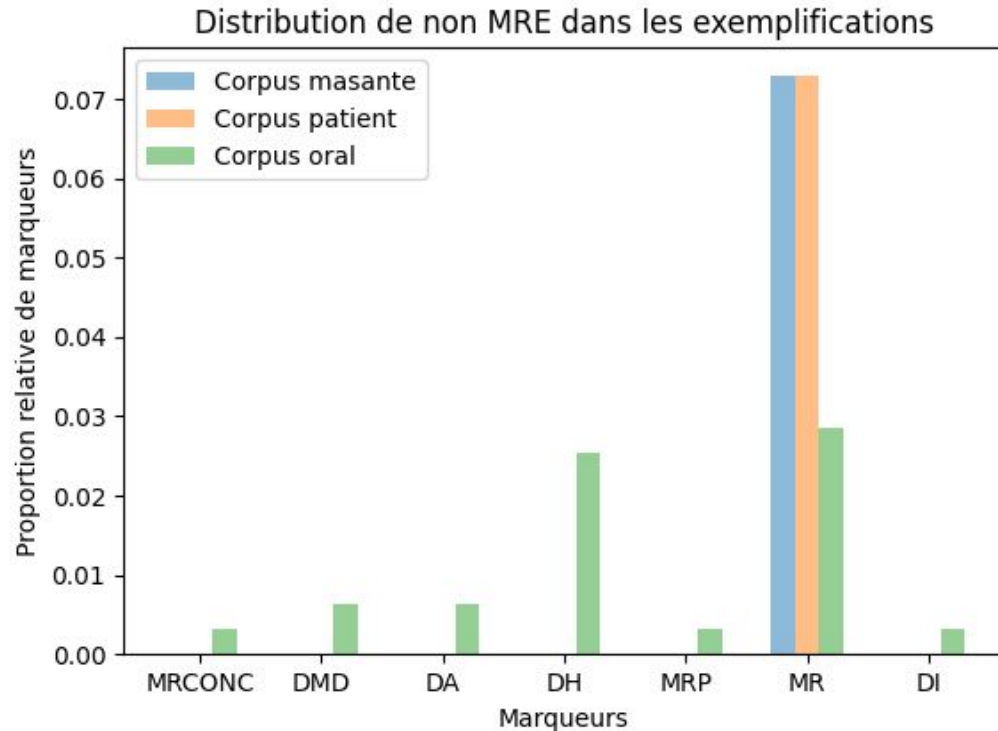


**MRE corpus oraux** : {'par exemple': 5, 'type': 1, 'genre': 2, 'comme': 1, 'style': 2}

**MRE masante** : {'en particulier': 5, 'type': 8, 'de type': 2, 'dont': 1, 'par exemple': 7, 'comme': 13, 'du type': 1, "est l'un de ceux-ci": 1, 'notamment': 1, 'et en particulier': 1, 'genre': 2}

**MRE patient** : {'notamment': 1, 'y compris': 1, 'comme': 2, 'du genre': 1, 'genre': 1, 'du genres': 1, 'par exemple': 1}

## Les autres marqueurs dans les exemplifications :



## Les autres marqueurs dans les exemplifications : exemples :



### Corpus patient :

<P1>j ' ai toutes sortes de traitements</P1><MR type="punct\_cl"></MR><N2 rel\_lex="instance (traitements/cortisone, AINS, paracétamol, dérivé morphinique, morphine, anti épileptique, piqures)" rel\_pragm="exempl">cortisone , AINS , paracétamol , dérivé morphinique , morphine , anti épileptique , piqures et j ' en passe surment</N2>

<NP1>des criteres nationaux</NP1><MR type="punct\_p"> ( </MR><N2 rel\_pragm="exempl">travail cotorep ect ect</N2>

<NP1>un travail très physique</NP1><MR/><P2 rel\_pragm="exempl">ou on doit porter des choses lourdes se baiser souvent</P2>

## Les autres marqueurs dans les exemplifications : exemples :



### Corpus oraux :

<P1>de médias locaux est-ce que vous vous suivez un petit peu</P1>  
<DMD>enfin</DMD> <DH>hm</DH> <P2 rel\_lex="syno(vous suivez/vous lisez)"  
rel\_pragm="exempl">vous lisez par exemple euh la République du Centre</P2>

<PP1>d'aller euh euh enfin de de faire plus que ça</PP1> <MRP>c'est-  
à-dire</MRP> <DH>euh</DH> <PP2 rel\_lex="hypero(faire plus que ça/tutoyer  
quelqu'un)" rel\_pragm="exempl">d'aller jusqu'à tutoyer quelqu'un pour créer on va  
dire une euh une complicité ou</PP2>

<P1>je fais tout ce qu'on font les mamans à la maison</P1>  
<MRCONC>donc</MRCONC> <DH>euh</DH> <NP2 rel\_pragm="exempl">le  
rangement le linge le ménage les courses la cuisine</NP2>

# Les marqueurs d'exemplification dans d'autre type de reformulation :

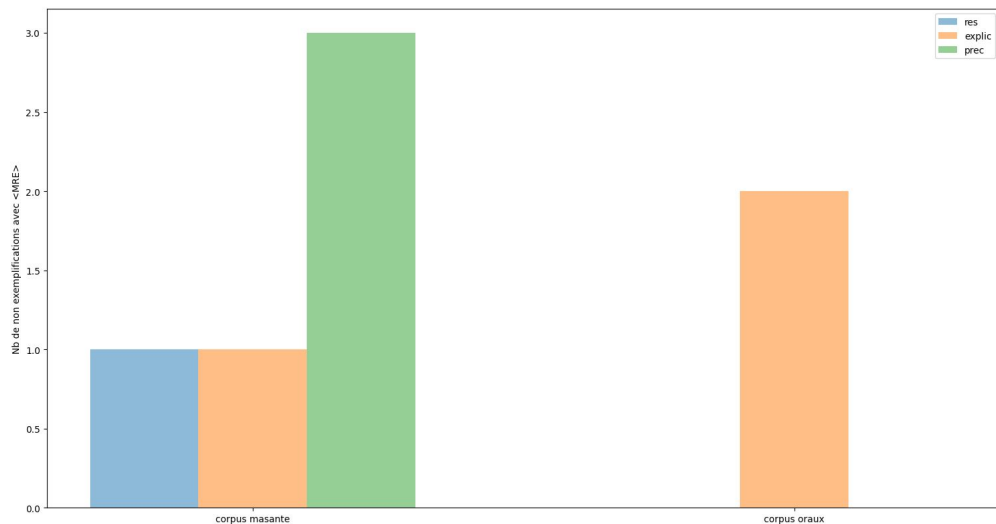


## Corpus Masante.net :

<NP1>un neurologue</NP1>, <MRE>en particulier</MRE> si <MR type="pres">c'est</MR> <NP2 rel\_pragm="prec">un neurologue qui a prescrit votre IRM</NP2>

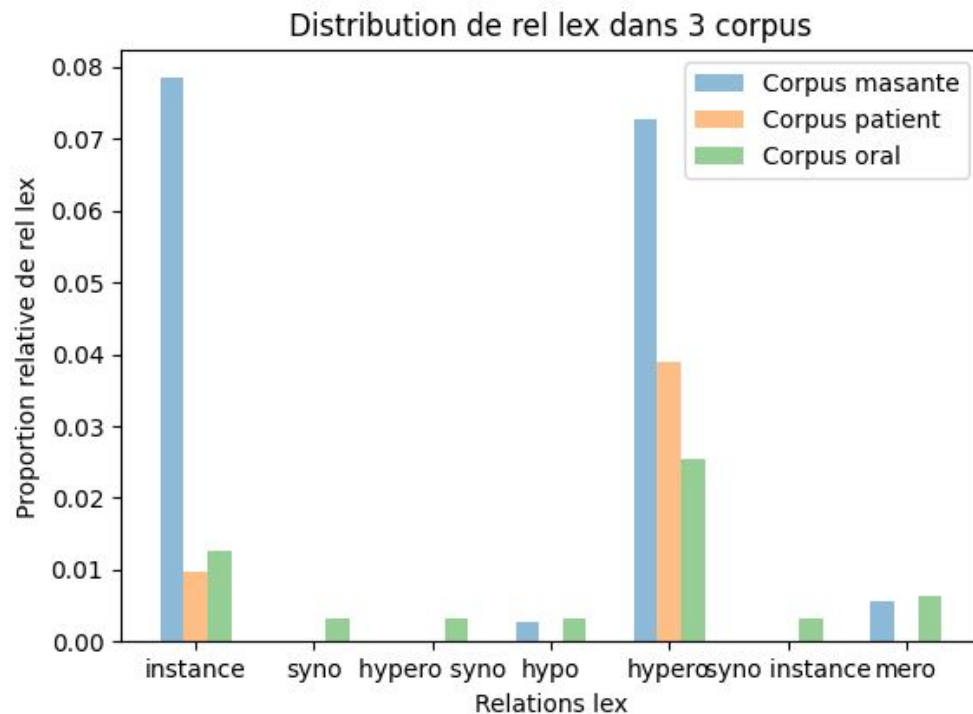
## Corpus oral :

<NP1>les jeunes délinquants</NP1> <MR type="pres">c'est</MR> <MRE>par exemple</MRE> <DH>euh</DH> <P2 rel\_pragm="explic">ils payent pas leurs euh leurs euh billets</P2>





## Les relations lexicales, nature et répartition :



---

# CONCLUSION

# Conclusion



- La modalité et le contexte d'émission du discours (oral, écrit, spontané, professionnel, etc...) semble influencer la réalisation des exemplifications.
- Le corpus écrit par des patients sur les douleurs de dos présente des similarités avec l'oral.
- Les corpus oraux et non professionnels ont moins de diversité dans leur MRE que le forum médecin.
- Le corpus médecin présente un fort taux d'hyponymie ce qui peut être lié avec le contexte d'explication pour des patients



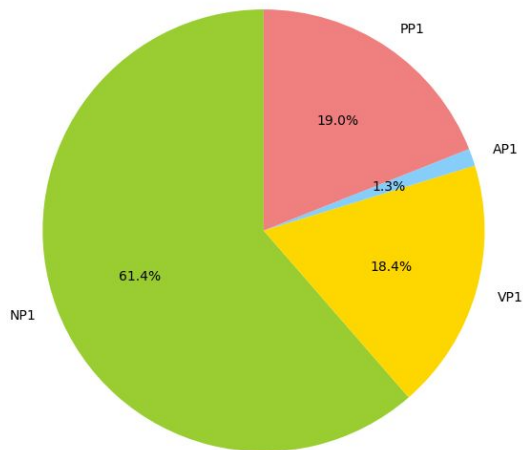
L'analyse des fréquences des marqueurs d'exemplification peut fournir des indications sur le profil des locuteurs.

---

**Pour aller plus loin : la distribution  
des catégories syntaxiques**

# Pourcentages de nos catégories

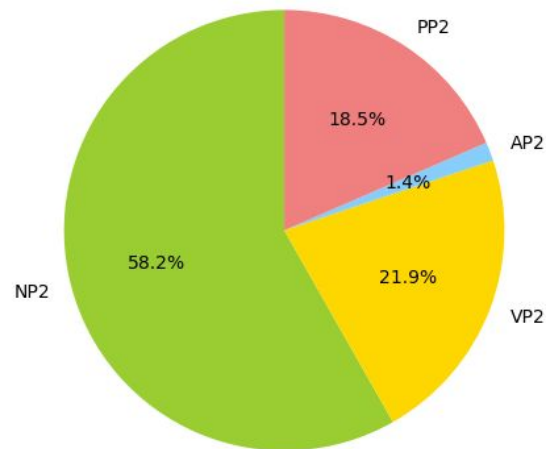
Répartition des catégories de modificateurs pour les segments reformulés oraux



## Exemple : SEGMENT 1

<NP1>un très très bon réseau d'amis</NP1>

Répartition des catégories de modificateurs pour les reformulations orales

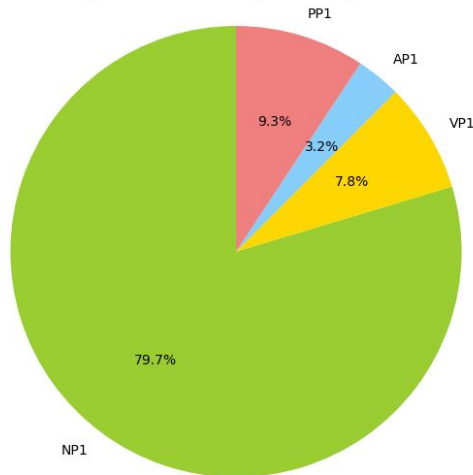


## Exemple : SEGMENT 2

<NP2 rel\_pragm="explic">pas mal de tâches administratives</NP2>

# Pourcentages de nos catégories

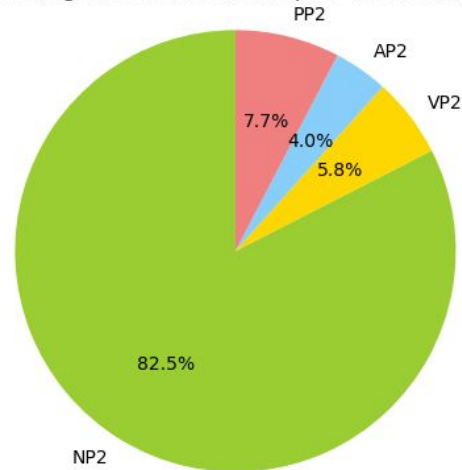
Répartition des catégories de modificateurs pour les segments reformulés écrits



Exemple : SEGMENT 1

<NP1>les douleurs que vous décrivez dans le ventre</NP1>

Répartition des catégories de modificateurs pour les reformulations écrites

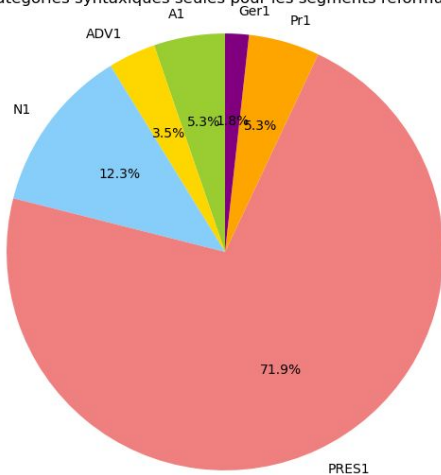


Exemple : SEGMENT 2

<NP2 rel\_lex="hypo(verrue/lésion)"  
rel\_pragm="def">une lésion d'origine virale qui peut  
donc donner naissance à plusieurs verrues-filles</NP2>

# Pourcentages de nos catégories

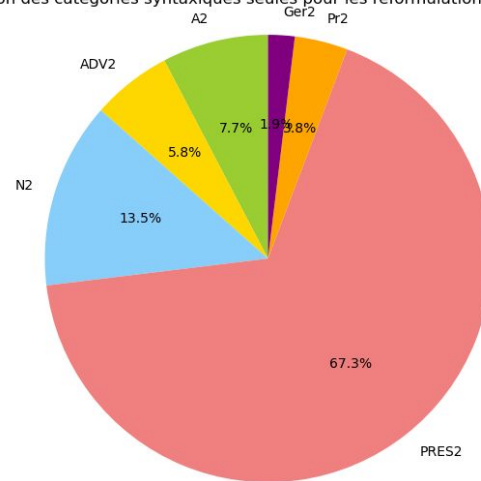
Répartition des catégories syntaxiques seules pour les segments reformulés du corpus oral



## Exemple : SEGMENT 1

<PRES1>c'est quand même pas sclérosant</PRES1>

Répartition des catégories syntaxiques seules pour les reformulations du corpus oral

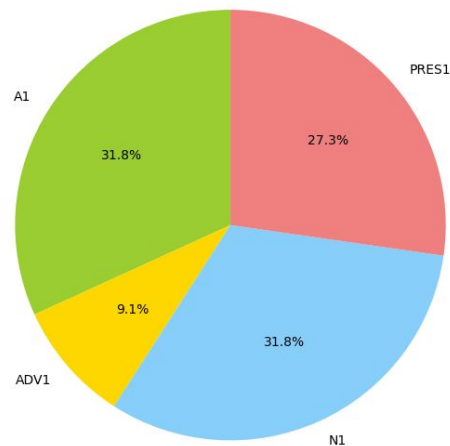


## Exemple : SEGMENT 2

<PRES2 rel\_lex="mero(humain/homme)"  
modif\_morph="deriv(humain/homme)" rel\_pragm="explic">c'est  
les les hommes qui se battent entre eux</PRES2>

# Pourcentages de nos catégories

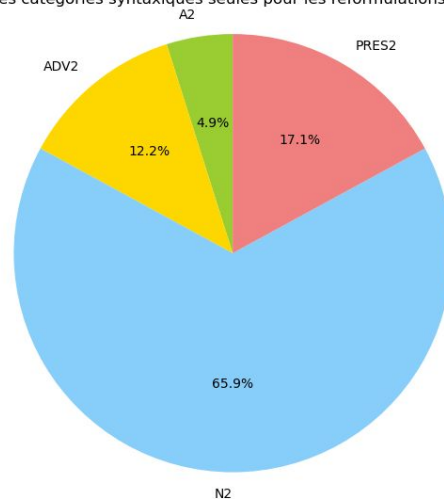
Répartition des catégories syntaxiques seules pour les segments reformulés du corpus écrit



Exemple : SEGMENT 1

<N1>antibiotique</N1> - <A1>neurologiques</A1>

Répartition des catégories syntaxiques seules pour les reformulations du corpus écrit



Exemple : SEGMENT 2

<N2 rel\_lex="hypero(glandes salivaires/parotide, sous-maxillaire)" rel\_pragm="prec">parotide et sous-maxillaire</N2>



---

**Que peut-on en tirer ?**



# Les catégories syntaxiques : que nous disent-elles ?

On observe une constante pour les catégories syntaxiques avec modifieurs, les reformulations sont le plus souvent employées avec un groupe nominal NP.

En revanche, pour les catégories syntaxiques seules, c'est beaucoup plus diversifié. La présence équivalente du nom et de l'adjectif dans le premier segment de l'écrit tient au vocabulaire scientifique.

- **A l'écrit** la catégorie du segment 2 augmente et reste la même que dans le segment 1, moins de catégories interviennent car c'est plus normé.
- **A l'oral** la catégorie du segment 2 diminue car il y a plus de catégories syntaxiques qui interviennent, à l'oral nous utilisons plus de variations.

Par exemple la catégorie PRES, pour présentateurs est très fortement utilisée à l'oral cela est lié à la tournure spontanée.

Nous pouvons voir que lorsque le segment 1 est un PRES à l'écrit alors le segment 2 sera plutôt un NOM, la catégorie N2 double de pourcentage, à l'écrit les structures syntaxiques sont plus rigides et donc mieux respectées. (Ex : C'est mon ami) et non pas (C'est ... il est mon ami).

Nous comprenons mieux pourquoi la modalité orale contient un plus grand panel de catégories syntaxiques.

Si nous lions cette piste aux exemplifications et autres marqueurs nous pouvons donc en conclure que les reformulations sont bien porteuses de sens, et ne sont pas faites au hasard. Le lien sous-jacent entre les 2 segments tient d'un objectif précis dans l'intention du locuteur.

- ★ Les reformulations sont parfois même utilisées à des fins commerciales ou dans le but de convaincre un interlocuteur pour leur témoigner de l'intérêt en reprenant des segments dans la phrase source créant ainsi une écoute active.



# REMERCIEMENTS

MERCI POUR VOTRE ATTENTION notamment VOTRE PRÉSENCE !