

**NOTES ET DOCUMENTS LIMSI N° : 2011-01**

*Janvier 2011*

## Typologie des modifications dans les révisions de Wikipédia

Camille Dutrey, Houda Bouamor,  
Delphine Bernhard et Aurélien Max



## Table des matières

Introduction.....	1
1. Faible variation sémantique.....	2
1.1. Corrections de surface.....	2
1.1.1. Corrections typographiques.....	2
1.1.1.1. Espace.....	2
1.1.1.2. Ponctuation.....	2
1.1.1.3. Casse.....	3
1.1.1.4. Format de l'heure.....	3
1.1.1.5. Nombre en lettre / en chiffre.....	4
1.1.2. Corrections orthographiques.....	5
1.1.2.1. Diacritique.....	5
1.1.2.2. Caractère alphabétique.....	6
1.1.3. Corrections grammaticales.....	6
1.1.3.1. Diacritique.....	6
1.1.3.2. Caractère alphabétique.....	7
1.1.3.3. Élosion.....	8
1.2. Reformulations.....	9
1.2.1. Reformulations lexicales.....	9
1.2.1.1. Sigle.....	9
1.2.1.2. Traduction.....	9
1.2.1.3. Emprunt.....	9
1.2.1.4. Régionalisme.....	10
1.2.1.5. Anaphore.....	10
1.2.1.6. Partie du discours.....	10
1.2.2. Reformulations syntaxiques.....	11
1.2.2.1. Permutation.....	11
1.2.2.2. Voix active / passive.....	11
1.2.2.3. Type de proposition.....	11
1.2.3. Reformulations sémantiques.....	12
1.2.3.1. Couverture sémantique.....	12
1.2.3.2. Normalisation encyclopédique.....	13
1.2.3.3. Synonymie.....	14
1.2.3.4. Neutralité.....	15
1.2.3.5. Précision.....	16
2. Forte variation sémantique.....	18
2.1. Corrections factuelles.....	18

2.1.1. Antonymie.....	18
2.1.2. Conjugaison.....	19
2.1.3. Rectificatif.....	19
2.2. Vandalisme.....	20
2.2.1. Manifeste.....	20
2.2.2. Subtil.....	21
3. Limites.....	22
3.1. Identification des phénomènes.....	22
3.1.1. Rôle du contexte.....	22
3.1.2. Motivation du contributeur.....	22
3.2. Simultanéité des phénomènes.....	23
3.3. Détection des phénomènes.....	23

# INTRODUCTION

Ce document détaille une typologie des modifications locales dans les révisions de Wikipédia<sup>1</sup>, développée à partir d'une étude détaillée du corpus WiCoPaCo<sup>2</sup> (Wikipedia Correction and Paraphrase Corpus). L'objectif de cette typologie est de représenter tous les phénomènes observables dans le corpus WiCoPaCo. Nous avons organisé les phénomènes de modifications selon une dichotomie sémantique : nous avons d'une part les modifications induisant une **faible variation sémantique** entre le segment source (le texte X issu d'un article de Wikipédia avant qu'il ait subi une modification Y) et le segment cible (le même texte X après qu'il ait subi une modification Y), d'autre part les modifications induisant une **forte variation sémantique** entre le segment source et le segment cible.

Les modifications à **faible variation sémantique** se ramifient en deux grands types de modifications : les **corrections de surface** et les **reformulations**. D'une part, les corrections de surface sont exclusivement d'ordre typographique, orthographique ou grammatical et correspondent à une « mise en conformité » relative à ces trois domaines. D'autre part, les reformulations consistent à réécrire autrement un item ou un syntagme. Elles sont actives sur trois plans : le lexique, la syntaxe et la sémantique.

Toutes ces modifications peuvent porter sur différents composants de la langue ou refléter différents phénomènes, allant de la ponctuation pour les corrections typographiques à la normalisation encyclopédique pour les reformulations sémantiques.

Contrairement aux corrections et aux reformulations, les modifications qui entrent en jeu lors d'une forte variation sémantique entre le segment source et le segment cible d'un texte viennent en quelque sorte contredire l'information donnée dans le segment source. Ces modifications sont également de deux types : ce sont soit des **corrections factuelles**, c'est-à-dire des vérifications ou des rectifications portant sur des énoncés, soit du **vandalisme**. Intrinsèquement liées à la sémantique, les corrections factuelles peuvent intervenir pour corriger une information qui s'avère fautive, avec pour conséquence possible une réelle absence de lien sémantique entre le segment source et le segment cible.

Les catégories développées dans cette typologie permettent de couvrir l'ensemble des phénomènes de modifications locales présentes dans le corpus. Outre un rôle de documentation pour ce dernier, elles permettent également de présenter aux personnes intéressées le type de phénomènes qu'elles peuvent trouver dans le corpus et pour l'étude desquelles il constituerait une ressource intéressante. En effet, le corpus WiCoPaCo étant destiné à être diffusé librement, savoir que l'on peut l'utiliser par exemple pour analyser des phénomènes de vandalisme dans Wikipédia ou des erreurs orthographiques réalisées a priori par des francophones natifs peut s'avérer extrêmement intéressant.

---

1 <http://fr.wikipedia.org>

2 <http://wicopaco.limsi.fr/>

# 1. FAIBLE VARIATION SÉMANTIQUE

## 1.1. CORRECTIONS DE SURFACE

### 1.1.1. Corrections typographiques

#### 1.1.1.1. Espace

Ce type de correction consistant à insérer (**exemple 1**) ou à supprimer (**exemple 2**) une espace a pour but de normaliser le texte, principalement pour rectifier deux mots qui ne seraient pas séparés par une espace. Wikipédia est par ailleurs dotée de recommandations concernant la présence ou l'absence d'une espace selon le signe qui la suit ou la précède, notamment par rapport aux ponctuations, au signe « pour cent » et aux nombres. Cette recommandation apporte également des informations sur le type d'espace à choisir (espace, espace insécable, etc.)<sup>3</sup>.

Exemple 1

```
<modif id="316023" wp_page_id="307643" wp_before_rev_id="6036152" wp_after_rev_id="6036170"
wp_user_id="83228" wp_user_num_modif="4" wp_comment="">
<before> <m num_words="1">HistoireFief</m> d' Avouzon et château fort de Rossillon au 12ème .
Maison de l' ordre des templiers à Crozet au 13ème .</before>
<after> <m num_words="2">Histoire Fief</m> d' Avouzon et château fort de Rossillon au 12ème .
Maison de l' ordre des templiers à Crozet au 13ème .</after>
</modif>
```

Exemple 2

```
<modif id="141397" wp_page_id="56884" wp_before_rev_id="20303882" wp_after_rev_id="20303918"
wp_user_id="26677" wp_user_num_modif="631" wp_comment="Annulation des modifications 20303882 de
207 . 253 . 70 . 24 spam">
<before>* Dans la vie de tous les jours , faire des courses signifie le fait d' aller dans un magasin ( souvent
une grande <m num_words="2">sur face</m> ) pour y acheter des choses ( nourriture par
exemple ) .</before>
<after>* Dans la vie de tous les jours , faire des courses signifie le fait d' aller dans un magasin ( souvent
une grande <m num_words="1">surface</m> ) pour y acheter des choses ( nourriture par
exemple ) .</after>
</modif>
```

#### 1.1.1.2. Ponctuation

Il s'agit d'insérer (**exemple 1**), de supprimer (**exemple 2**) ou de remplacer (**exemple 3**) un caractère de ponctuation. Outre les caractères habituels, il est à noter que l'apostrophe et le trait d'union en font partie. De plus, la modification d'une ponctuation dite « faible » en ponctuation dite « forte » implique la plupart du temps une modification de la casse.

Exemple 1

```
<modif id="331768" wp_page_id="345707" wp_before_rev_id="3575537" wp_after_rev_id="3997933"
wp_user_id="3086" wp_user_num_modif="27673" wp_comment="typo lien">
<before>Ces trois parties se rejoignent pour former une épaisse masse musculaire qui se termine sous
forme d' un tendon s' insérant sur l' olécrâne de l' os ulna . Le triceps brachial est un muscle extenseur de l'
<m num_words="2">avant bras</m> sur le bras .</before>
<after>Ces trois parties se rejoignent pour former une épaisse masse musculaire qui se termine sous
forme d' un tendon s' insérant sur l' olécrâne de l' os ulna . Le triceps brachial est un muscle extenseur de l'
<m num_words="1">avant-bras</m> sur le bras .</after>
</modif>
```

<sup>3</sup> Source : <[http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Conventions\\_typographiques#Espaces](http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Conventions_typographiques#Espaces)>

## Exemple 2

```
<modif id="141295" wp_page_id="56799" wp_before_rev_id="17970513" wp_after_rev_id="18042854"
wp_user_id="260005" wp_user_num_modif="24" wp_comment="">
<before>Une tirade , farcie d' allitérations en <m num_words="1">V,</m> est celle initiale de V, le
personnage principal de V pour Vendetta . Elle est entièrement citée dans l' article sur le film tiré de B-
D .</before>
<after>Une tirade farcie d' allitérations en <m num_words="1">V</m> est celle initiale de V, le
personnage principal de V pour Vendetta . Elle est entièrement citée dans l' article sur le film tiré de B-
D .</after>
</modif>
```

## Exemple 3

```
<modif id="182862" wp_page_id="94973" wp_before_rev_id="7840972" wp_after_rev_id="8137548"
wp_user_id="4751" wp_user_num_modif="2740" wp_comment="Géographie formu">
<before>Situé en France dans massif central en Lozère <m num_words="1">.Le</m> Méjean est
ceinturé des vallées impressionnantes de la Jonte au sud , du Tarnon à l' est et du Tarn au nord et à l'
ouest .</before>
<after>Situé en France dans le massif central en Lozère <m num_words="1">, le</m> Méjean est
ceinturé des vallées impressionnantes de la Jonte au sud , du Tarnon à l' est et du Tarn au nord et à l'
ouest .</after>
</modif>
```

### 1.1.1.3. Casse

Cette modification s'applique aussi bien aux noms propres qu'aux noms communs et concerne principalement :

- les majuscules en bas-de-casse qu'il faut capitaliser (**exemple 1**) ;
- les minuscules en capitales qu'il faut mettre en bas-de-casse (**exemple 2**).

## Exemple 1

```
<modif id="407791" wp_page_id="1819909" wp_before_rev_id="20137623" wp_after_rev_id="20488825"
wp_user_id="245718" wp_user_num_modif="71" wp_comment="correction d' une partie de l'
orthographe">
<before>* Nagisa Nagase : La petite amie de Ryuji . elle se fait souvent draguer par <m
num_words="1">eikichi _</m> </before>
<after>* Nagisa Nagase : La petite amie de Ryuji . elle se fait souvent draguer par <m
num_words="1">Eikichi .</m> </after>
</modif>
```

## Exemple 2

```
<modif id="141804" wp_page_id="57076" wp_before_rev_id="11968072" wp_after_rev_id="11968147"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="Bilan matériel">
<before>Parmi les équipements publics touchés on peut citer le petit palais des sports ( entièrement
démoli et reconstruit suite à ces dommages ) , le Bikini ( salle de spectacle ) , l' École nationale supérieure
des ingénieurs de génie chimique , le lycée Galiéni , <m num_words="1">Centre</m> hospitalier "
GERARD MARCHANT " ...</before>
<after>Parmi les équipements publics touchés on peut citer le petit palais des sports ( entièrement démoli
et reconstruit suite à ces dommages ) , le Bikini ( salle de spectacle ) , l' École nationale supérieure des
ingénieurs de génie chimique , le lycée Galiéni , <m num_words="2">le centre</m> hospitalier " GERARD
MARCHANT " ...</after>
</modif>
```

### 1.1.1.4. Format de l'heure

Cette correction est peu fréquente et n'obéit *a priori* à aucune règle, si ce n'est relativement au « format » 24h (**exemple 1**). Elle peut par ailleurs être régie par les recommandations internes à Wikipédia, selon lesquelles :

« En règle générale, lorsque l'on parle d'une durée sans caractère de précision, de comparaison ou de complexité, il est préférable d'écrire cette durée en lettres (au long) : « Le vol de l'avion de la compagnie Oceanic Airlines a duré quatre heures et cinquante minutes. » En revanche, on écrit en chiffres : « Il arriva avec 32 secondes d'avance, après une course de 57 minutes 5/10 ». Concernant l'écriture des heures, il y a plusieurs cas de figure :\* l'heure pleine, sans minute, doit s'écrire avec heure : « Il est 15 heures. À 18 heures, le président fera sa déclaration. » ; \* le mot « heure » s'écrit en toutes lettres si le mot « minute » l'est également : « Le rendez-vous est fixé à 21 heures 30 minutes. » Il en est de même pour le mot « seconde » ;

\* le mot « heure » est abrégé (symbole « h ») si le mot « minute » est abrégé (symbole « min ») ou absent ; c'est en particulier souvent l'usage dans les textes techniques. Le symbole de la « seconde » est « s ». Les symboles « h », « min » et « s », comme tous symboles, doivent alors être espacés avant et après. On n'ajoute pas de zéro devant le nombre des heures ni devant celui des minutes quand ceux-ci sont inférieurs à dix :

- o 13 h 17
- o 23 h 59 min 59 s
- o 2 h 5

\* les nombres des heures se composent en lettres lorsqu'ils sont associés aux mots : « quart », « demi », « trois quarts », « midi », « minuit », ainsi que dans les textes littéraires : « Dix heures et quart », « Elle sortit de chez elle à huit heures trente. »

Source : <[http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Conventions\\_typographiques#Heure](http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Conventions_typographiques#Heure)>

#### Exemple 1

```
<modif id="406716" wp_page_id="1696582" wp_before_rev_id="9962989" wp_after_rev_id="9964928"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="Jeunes années">
<before>Adolf Hitler naît le 20 avril 1889 à <m num_words="1">6h30</m> dans l' auberge Gasthof zum
Pommer , Vorstadt Nr . 219 , à Braunau am Inn , une petite ville de Haute-Autriche près de la frontière
austro-allemande . Il est le quatrième des six enfants d' Alois Hitler et de Klara Pölzl . La plupart des enfants
meurent en bas âge ; seule sa sœur cadette Paula († 1960 ) lui survivra .</before>
<after>Adolf Hitler naît le 20 avril 1889 à <m num_words="1">18h30</m> dans l' auberge Gasthof zum
Pommer , Vorstadt Nr . 219 , à Braunau am Inn , une petite ville de Haute-Autriche près de la frontière
austro-allemande . Il est le quatrième des six enfants d' Alois Hitler et de Klara Pölzl . La plupart des enfants
meurent en bas âge ; seule sa sœur cadette Paula († 1960 ) lui survivra .</after>
</modif>
```

#### 1.1.1.5. Nombre en lettre / en chiffre

Cette modification concerne la réécriture en lettres d'un nombre écrit en chiffres (**exemple 1**) et inversement (**exemple 2**). L'écriture des nombres, ici non plus sans être régie *a priori* par des règles strictes, peut être soumise aux recommandations internes à Wikipédia, selon lesquelles :

« Il est recommandé d'écrire les nombres à un seul chiffre en toutes lettres (« au long »). Par exemple « au bout de deux ans d'existence » et non « au bout de 2 ans d'existence ». Évitez aussi, autant que possible, de commencer une phrase par un nombre écrit en chiffres. En règle générale, les nombres sont écrits de préférence « au long » lorsqu'ils indiquent des grandeurs (deux cents mètres, trois mille habitants, quinze francs...) sauf lorsque leur profusion les rendrait difficilement lisibles et pour les nombres de 70 à 99, dont le mot diffère selon les francophones ; les âges, les dates et les heures seront écrits en chiffres (exemple : « Le 28 janvier, à 18 h 15, j'aurai 24 ans et nous serons quatre amis à passer à table, la première fois depuis cinq ans. »). »<sup>4</sup>

Source : <[http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Conventions\\_typographiques#Nombres](http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Conventions_typographiques#Nombres)>

#### Exemple 1

```
<modif id="408790" wp_page_id="1993884" wp_before_rev_id="1971402" wp_after_rev_id="2165790"
wp_user_id="10905" wp_user_num_modif="4144" wp_comment="revert ( conventions typographiques )">
<before>* 1917 -Clark lance le Trutractor , un véhicule à <m num_words="1">3</m> roues , propulsé
au gaz , destiné au transport de marchandise à l' intérieur de l' entreprise-l' ancêtre du chariot de
manutention est né .</before>
<after>* 1917 -Clark lance le Trutractor , un véhicule à <m num_words="1">trois</m> roues , propulsé
au gaz , destiné au transport de marchandise à l' intérieur de l' entreprise-l' ancêtre du chariot de
manutention est né .</after>
</modif>
```

#### Exemple 2

```
<modif id="406994" wp_page_id="1719200" wp_before_rev_id="19605103" wp_after_rev_id="21077550"
wp_user_id="8325" wp_user_num_modif="5957" wp_comment="homon|Tardy|Montravel , relecture">
<before>De 1859 à 1864 il fut le <m num_words="3">quarante et unième</m> gouverneur de la
Guyane .</before>
<after>De 1859 à 1864 il fut le <m num_words="2">41 e</m> gouverneur de la Guyane .</after>
</modif>
```

---

4 Pour plus de détails : <[http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Conventions\\_concernant\\_les\\_nombres](http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Conventions_concernant_les_nombres)>



## 1.1.2. Corrections orthographiques

Les critères permettant de distinguer les corrections de type « orthographique » de corrections de type « grammatical » sont les suivants :

- les corrections de type « orthographique » visent à corriger un item incorrect pour le rendre conforme au lexique alors que les corrections de type « grammatical » portent sur un item attesté dans le lexique mais n'étant pas celui attendu par la construction syntaxique du segment qui le contient ou par le sens des items auxquels il est lié ;
- les corrections de type « orthographique », à l'inverse des corrections de type « grammatical », ne font pas intervenir le contexte de l'item sur lequel elles portent.

### 1.1.2.1. Diacritique

L'item du segment source est incorrect car il est porteur d'un diacritique ne correspondant pas à son orthographe admise - et dans ce cas il s'agit de rectifier ce dernier (**exemple 1**) - ou car un diacritique qu'il porte est absent (**exemple 2**) ou car il porte un diacritique qui n'est pas attendu (**exemple 3**). Ces trois types de modifications ont pour but de transformer une simple chaîne de caractères en une unité lexicale correcte.

Exemple 1

```
<modif id="141221" wp_page_id="56758" wp_before_rev_id="7414938" wp_after_rev_id="7414947"
wp_user_id="2739" wp_user_num_modif="20457" wp_comment="Architecture et monuments">
<before>L' <m num_words="1">èglise</m> gothique Sainte-Marie , construite au XIVe siècle , se dresse
sur la place du marché avec son célèbre autel de bois de Witt Stwosz . Chaque heure , le son d' une
trompette ( hejnal ) est lancé vers les quatre directions du haut de la plus élevée des deux tours de l'
église . Ceci rappelle à tous le joueur de trompette du XIIIe siècle qui fut transpercé d' une flèche alors qu '
il donnait l' alarme d' une invasion tatare . Le hejnal de midi est retransmis en direct à travers toute la
Pologne sur la station de radio Jedylnka .</before>
<after>L' <m num_words="1">église</m> gothique Sainte-Marie , construite au XIVe siècle , se dresse
sur la place du marché avec son célèbre autel de bois de Witt Stwosz . Chaque heure , le son d' une
trompette ( hejnal ) est lancé vers les quatre directions du haut de la plus élevée des deux tours de l'
église . Ceci rappelle à tous le joueur de trompette du XIIIe siècle qui fut transpercé d' une flèche alors qu '
il donnait l' alarme d' une invasion tatare . Le hejnal de midi est retransmis en direct à travers toute la
Pologne sur la station de radio Jedylnka .</after>
</modif>
```

Exemple 2

```
<modif id="141247" wp_page_id="56776" wp_before_rev_id="2380847" wp_after_rev_id="3363910"
wp_user_id="8194" wp_user_num_modif="143" wp_comment="bons liens vers notices satellites a creer
( topex , jason-1 )">
<before>Laltimétrie satellitale , aujourd ' hui au centre de l' activité d' océanographie spatiale , est une
technique spatiale permettant de mesurer le relief des océans , mise au point dans les années 1970 puis
1980 , et qui a vu ses capacités décuplées en termes de précision et de couverture spatio-temporelle . Ces
progrès ont été notamment obtenus grâce aux données du satellite franco-américain TOPEX / <m
num_words="1">Poseidon</m> lancé par la fusée Ariane en août 1992 . Aujourd ' hui , le successeur de
ce dernier satellite , en fin de vie , est Jason-1 ( Alcatel-Space , Cannes ) , premier satellite de ce qui devrait
devenir une véritable filière spatiale en matière de suivi océanographique à long terme .</before>
<after>Laltimétrie satellitale , aujourd ' hui au centre de l' activité d' océanographie spatiale , est une
technique spatiale permettant de mesurer le relief des océans , mise au point dans les années 1970 puis
1980 , et qui a vu ses capacités décuplées en termes de précision et de couverture spatio-temporelle . Ces
progrès ont été notamment obtenus grâce aux données du satellite franco-américain TOPEX / <m
num_words="1">Poséidon</m> lancé par la fusée Ariane en août 1992 . Aujourd ' hui , le successeur de
ce dernier satellite , en fin de vie , est Jason-1 ( Alcatel-Space , Cannes ) , premier satellite de ce qui devrait
devenir une véritable filière spatiale en matière de suivi océanographique à long terme .</after>
</modif>
```

### Exemple 3

```
<modif id="141353" wp_page_id="56859" wp_before_rev_id="7833329" wp_after_rev_id="7861939"
wp_user_id="25961" wp_user_num_modif="914" wp_comment="Ortho">
<before>En 47 av . J .-C . , les troupes de Jules César incendient la flotte d' Alexandrie ; le feu se serait
propagé aux entrepôts et aurait détruit une partie de la bibliothèque . Reconstituée , elle est détruite à
nouveau cinq ou six fois , la dernière en 642 par le général ' Amr Ibn al-' As , obéissant au calife ' Umar qui
considérerait que si les livres étaient en accord avec le Coran ils étaient superflus , et que s' ils contredisaient
le Coran , ils étaient pernicieux , comme l' avaient fait les chrétiens lors des <m
num_words="1">déstructions</m> précédentes . Ce récit est cependant remis en cause , voire assimilé
à de la propagande anti-islamique , la plupart des connaissances de l' antiquité ayant été transmises à l'
Occident par les traductions arabes telles celles d' Averroès .</before>
<after>En 47 av . J .-C . , les troupes de Jules César incendient la flotte d' Alexandrie ; le feu se serait
propagé aux entrepôts et aurait détruit une partie de la bibliothèque . Reconstituée , elle est détruite à
nouveau cinq ou six fois , la dernière en 642 par le général ' Amr Ibn al-' As , obéissant au calife ' Umar qui
considérerait que si les livres étaient en accord avec le Coran ils étaient superflus , et que s' ils contredisaient
le Coran , ils étaient pernicieux , comme l' avaient fait les chrétiens lors des <m
num_words="1">déstructions</m> précédentes . Ce récit est cependant remis en cause , voire assimilé
à de la propagande anti-islamique , la plupart des connaissances de l' antiquité ayant été transmises à l'
Occident par les traductions arabes telles celles d' Averroès .</after>
</modif>
```

#### 1.1.2.2. Caractère alphabétique

L'insertion, la suppression ou la substitution d'un ou plusieurs caractères alphabétiques est principalement observée dans deux cas de figure :

- lorsque l'item du segment source n'existe pas – en grande partie à cause de l'inversion de deux lettres, conséquence classique d'une « faute de frappe » – et qu'il faut le transformer en un item attesté dans le lexique (**exemple 1**) ;
- lorsque l'item est en fait une transcription en Alphabet Phonétique International (API) et qu'il ne respecte pas ses principes – couramment, l'utilisateur n'a pas « trouvé » comment représenter à l'écran le caractère voulu – et qu'il faut le rendre conforme à l'API (**exemple 2**).

#### Exemple 1

```
<modif id="331766" wp_page_id="345707" wp_before_rev_id="2936203" wp_after_rev_id="3575527"
wp_user_id="19112" wp_user_num_modif="5183" wp_comment="typos">
<before>Ces trois parties se <m num_words="1">rejoignent</m> pour former une épaisse masse
musculaire qui se termine sous forme d' un tendon s' insérant sur l' olécrâne de l' os ulna . Le triceps
brachial est un muscle extenseur de l' avant bras sur le bras .</before>
<after>Ces trois parties se <m num_words="1">rejoignent</m> pour former une épaisse masse
musculaire qui se termine sous forme d' un tendon s' insérant sur l' olécrâne de l' os ulna . Le triceps
brachial est un muscle extenseur de l' avant bras sur le bras .</after>
</modif>
```

#### Exemple 2

```
<modif id="408739" wp_page_id="1977538" wp_before_rev_id="21003523" wp_after_rev_id="21005840"
wp_user_id="195712" wp_user_num_modif="29" wp_comment="Alphabet et prononciation">
<before>- ec , -e et-c placés à la fin d' un mot ne se prononcent pas : frybarec ( empire ) se prononce "
frubar " ; Ghintec zint est ainsi la transcription stricte de jint . Le-r final , placé après une consonne , ne se
prononce pas ; il se prononce normalement après une voyelle : nimh Laitpanr ( grand-duc de Laipanh ) se
prononce " nif Lètpan " letpan , mais dreuc Haïder ( comte de Haïdec ) se prononce " dreu Haïder " <m
num_words="1">haïdər</m> .</before>
<after>- ec , -e et-c placés à la fin d' un mot ne se prononcent pas : frybarec ( empire ) se prononce "
frubar " ; Ghintec zint est ainsi la transcription stricte de jint . Le-r final , placé après une consonne , ne se
prononce pas ; il se prononce normalement après une voyelle : nimh Laitpanr ( grand-duc de Laipanh ) se
prononce " nif Lètpan " letpan , mais dreuc Haïder ( comte de Haïdec ) se prononce " dreu Haïder " <m
num_words="1">hajdər</m> .</after>
</modif>
```

### 1.1.3. Corrections grammaticales

#### 1.1.3.1. Diacritique

L'insertion, la suppression ou la substitution d'un diacritique est formellement semblable à la correction orthographique portant sur les diacritiques, à cela près que l'item du segment source est attesté dans le lexique : seule une connaissance du contexte proche (sémantique

comme syntaxique) permet de définir que cet item est fautif. Cela peut être la conséquence :

- d'une confusion liée à l'homophonie (**exemple 1**) ;
- d'une confusion sur la partie du discours de l'item dont il est question (**exemple 2**).

Exemple 1

```
<modif id="407546" wp_page_id="1808509" wp_before_rev_id="7455352" wp_after_rev_id="8584640"
wp_user_id="57060" wp_user_num_modif="1668" wp_comment="De la « Paix de Dieu » à la « Trêve de
Dieu » aj. wikilien + typo + ortho">
<before>Les chevaliers ont interdiction de faire la guerre , d' abord le samedi , puis du mercredi soir jusqu
' au lundi matin ( conciles d' Arles , en 1037 - 1041 ) . Enfin , outre la pénitence , le non-respect de la Trêve
peut désormais se solder par l' anathème pour le <m num_words="1">pêcheur</m> : ce dernier est
privé de sépulture chrétienne .</before>
<after>Les chevaliers ont interdiction de faire la guerre , d' abord le samedi , puis du mercredi soir jusqu
' au lundi matin ( conciles d' Arles , en 1037 - 1041 ) . Enfin , outre la pénitence , le non-respect de la Trêve
peut désormais se solder par l' anathème pour le <m num_words="1">pêcheur</m> : ce dernier est
privé de sépulture chrétienne .</after>
</modif>
```

Exemple 2

```
<modif id="258150" wp_page_id="185475" wp_before_rev_id="1877313" wp_after_rev_id="1905461"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="L' AFDX dans l' Airbus A380 typo">
<before>Pour répondre à ces demandes , Airbus <m num_words="1">à</m> choisi deux fournisseurs :
Thales ( équipements terminaux ) et Rockwell-Collins ( commutateurs et équipements
terminaux ) .</before>
<after>Pour répondre à ces demandes , Airbus <m num_words="1">a</m> choisi deux fournisseurs :
Thales ( équipements terminaux ) et Rockwell-Collins ( commutateurs et équipements
terminaux ) .</after>
</modif>
```

### 1.1.3.2. Caractère alphabétique

À l'image de la correction portant sur les diacritiques, cette correction est effectuée sur un item attesté dans le lexique mais qui s'avère incorrect lorsque l'on étudie son contexte proche. Il s'agit principalement :

- de modifier le nombre de l'item suite généralement à une faute d'accord (**exemple 1**) ;
- de modifier le genre de l'item suite généralement à une faute d'accord (**exemples 2 & 3**) ;
- d'une confusion liée à l'homophonie (**exemples 4 & 5**).

Exemple 1

```
<modif id="9" wp_page_id="7" wp_before_rev_id="669004" wp_after_rev_id="679685"
wp_user_id="4305" wp_user_num_modif="3606" wp_comment="typo + lien + modif">
<before>On nomme algèbre linéaire la branche des mathématiques qui s' occupe de l' étude des vecteurs
( ensembles ordonnés de scalaires ) , des espaces vectoriels ( ou espaces linéaires ) , des <m
num_words="1">transformation</m> linéaires et des systèmes d' équation linéaires ( théorie des
matrices ) .</before>
<after>On nomme algèbre linéaire la branche des mathématiques qui s' occupe de l' étude des vecteurs
( ensembles ordonnés de scalaires ) , des espaces vectoriels ( ou espaces linéaires ) , des <m
num_words="1">transformations</m> linéaires et des systèmes d' équation linéaires ( théorie des
matrices ) .</after>
</modif>
```

Exemple 2

```
<modif id="141524" wp_page_id="56978" wp_before_rev_id="297122" wp_after_rev_id="300811"
wp_user_id="2314" wp_user_num_modif="156" wp_comment="Ortho">
<before>La levure est un champignon unicellulaire apte à la fermentation des matières organiques ,
animales ou végétales , <m num_words="1">utilisée</m> pour la fermentation alcoolique des solutions
sucrées , ou pour faire lever la pâte du pain .</before>
<after>La levure est un champignon unicellulaire apte à la fermentation des matières organiques ,
animales ou végétales , <m num_words="1">utilisé</m> pour la fermentation alcoolique des solutions
sucrées , ou pour faire lever la pâte du pain .</after>
</modif>
```

### Exemple 3

```
<modif id="141490" wp_page_id="56967" wp_before_rev_id="20495718" wp_after_rev_id="20664158"
wp_user_id="231806" wp_user_num_modif="1142" wp_comment="orthographe">
<before>De même , il est assez fréquent que les incisives permanentes apparaissent en arrière des
incisives temporaires . Dans ce cas , on aide les dents de lait à tomber et les dents définitives peuvent
reprendre <m num_words="1">un</m> position correcte .</before>
<after>De même , il est assez fréquent que les incisives permanentes apparaissent en arrière des incisives
temporaires . Dans ce cas , on aide les dents de lait à tomber et les dents définitives peuvent reprendre <m
num_words="1">une</m> position correcte .</after>
</modif>
```

### Exemple 4

```
<modif id="408697" wp_page_id="1966685" wp_before_rev_id="7087362" wp_after_rev_id="7087376"
wp_user_id="103586" wp_user_num_modif="45" wp_comment="">
<before>Jean-Paul Corbineau est un des membres fondateurs de Tri Yann , il chante avec une <m
num_words="1">voie</m> de troubadour les plaintes du groupe , il joue également de la
guitare .</before>
<after>Jean-Paul Corbineau est un des membres fondateurs de Tri Yann , il chante avec une <m
num_words="1">voix</m> de troubadour les plaintes du groupe , il joue également de la
guitare .</after>
</modif>
```

### Exemple 5

```
<modif id="141364" wp_page_id="56859" wp_before_rev_id="10789765" wp_after_rev_id="10796021"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="Destructions de la bibliothèque">
<before>En effet de nombreux ouvrages ( Euclide , Ptolémée , Aristote , etc . ) furent traduits puis
commentés dans le monde islamique . Notons que le grand philosophe Al-Farabi était nommé le second
maître , le premier étant Aristote . C' est dire si la pensée grecque à travers les ouvrages contenus dans la
bibliothèque d' Alexandrie a nourri et a été une des bases de l' essor philosophique et mathématique qu ' a
connu le monde islamique par la suite . Il est couramment considéré que c' est grâce aux écrits d' Averroès
( Ibn Rushd ) considéré comme le Commentateur d' Aristote que la pensée philosophique grecque fit son
entrée dans l' <m num_words="1">aire</m> chrétienne .</before>
<after>En effet de nombreux ouvrages ( Euclide , Ptolémée , Aristote , etc . ) furent traduits puis
commentés dans le monde islamique . Notons que le grand philosophe Al-Farabi était nommé le second
maître , le premier étant Aristote . C' est dire si la pensée grecque à travers les ouvrages contenus dans la
bibliothèque d' Alexandrie a nourri et a été une des bases de l' essor philosophique et mathématique qu ' a
connu le monde islamique par la suite . Il est couramment considéré que c' est grâce aux écrits d' Averroès
( Ibn Rushd ) considéré comme le Commentateur d' Aristote que la pensée philosophique grecque fit son
entrée dans l' <m num_words="1">ère</m> chrétienne .</after>
</modif>
```

#### 1.1.3.3. Élision

Le phénomène d'élision (**exemples 1 & 2**), qui pourrait par ailleurs être considéré comme une succession de deux phénomènes autres (suppression d'un caractère alphabétique suivie de l'insertion d'un caractère de ponctuation), est ici typé comme une seule et même modification.

### Exemple 1

```
<modif id="316021" wp_page_id="307637" wp_before_rev_id="12730172" wp_after_rev_id="12811994"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="Dernière Version : Flight Simulator X">
<before>Le jeu <m num_words="1">ne</m> est compatible qu ' avec Windows XP ou Windows
Vista .</before>
<after>Le jeu <m num_words="1">n'</m> est compatible qu ' avec Windows XP ou Windows
Vista .</after>
</modif>
```

### Exemple 2

```
<modif id="408155" wp_page_id="1891160" wp_before_rev_id="19968798" wp_after_rev_id="19971511"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="">
<before>En 1909 , sous son initiative , un hebdomadaire et un autre journal arabophones sous la direction
<m num_words="1">de</m> Abdelaziz Thâalbi voient le jour .</before>
<after>En 1909 , sous son initiative , un hebdomadaire et un autre journal arabophones voient le jour
sous la direction <m num_words="1">d'</m> Abdelaziz Thâalbi .</after>
</modif>
```

## 1.2. REFORMULATIONS

### 1.2.1. Reformulations lexicales

#### 1.2.1.1. Sigle

L'item du segment source ou l'item du segment cible est un sigle : la modification ayant pour but de l'expliquer en le développant (**exemple 1**) ou de l'abrégé (**exemple 2**) car, par exemple, il a déjà été présenté sous sa forme étendue dans un contexte plus ou moins proche (au sein du même article).

Exemple 1

```
<modif id="128" wp_page_id="15" wp_before_rev_id="46652" wp_after_rev_id="46742"
wp_user_id="272" wp_user_num_modif="5508" wp_comment="désambig . Vienne">
<before>L'Autriche est membre de l' <m num_words="1">UE</m> et de la zone euro .</before>
<after>L'Autriche est membre de l' <m num_words="2">Union Européenne</m> et de la zone
euro .</after>
</modif>
```

Exemple 2

```
<modif id="407810" wp_page_id="1820290" wp_before_rev_id="19458952" wp_after_rev_id="19611226"
wp_user_id="16830" wp_user_num_modif="224" wp_comment="Parti libéral -&gt; PLP">
<before>Gérard-Octave Pinkers , ( Dison , -Charleroi , ) , était un homme politique belge de langue
française , membre du <m num_words="2">Parti libéral</m> , et un militant wallon</before>
<after>Gérard-Octave Pinkers , ( Dison , -Charleroi , ) , était un homme politique belge de langue
française , membre du <m num_words="1">PLP</m> , et un militant wallon</after>
</modif>
```

#### 1.2.1.2. Traduction

Il s'agit principalement de traductions de noms propres. Ces traductions peuvent aller de la langue source à la langue cible - le français - (**exemple 1**) ou l'inverse (**exemple 2**).

Exemple 1

```
<modif id="182806" wp_page_id="94894" wp_before_rev_id="1573125" wp_after_rev_id="1718453"
wp_user_id="311" wp_user_num_modif="14270" wp_comment="Catégorie : Patrimoine mondial de l'
UNESCO">
<before>La sanctifiée Élisabeth de Hongrie ( plus tard de Thuringe , aussi , comme elle passa une partie
de sa vie de 1211 à 1228 au Wartbourg comme consort de <m num_words="1">Ludwig</m>
IV ) .</before>
<after>La sanctifiée Élisabeth de Hongrie ( plus tard de Thuringe , aussi , comme elle passa une partie
de sa vie de 1211 à 1228 au Wartbourg comme consort de <m num_words="1">Louis</m>
IV ) .</after>
</modif>
```

Exemple 2

```
<modif id="182847" wp_page_id="94956" wp_before_rev_id="1733121" wp_after_rev_id="1733174"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="Le problème du temps">
<before>L' aevum serait à considérer comme distinct du temps à proprement parler , aussi bien que de l'
éternité . La différence entre laevum et le temps fait l' objet de l' article 5 de la question 10 ( <m
num_words="2">Première partie</m> ) , relative à « L' éternité de Dieu ».</before>
<after>L' aevum serait à considérer comme distinct du temps à proprement parler , aussi bien que de l'
éternité . La différence entre laevum et le temps fait l' objet de l' article 5 de la question 10 ( <m
num_words="2">Prima pars</m> ) , relative à « L' éternité de Dieu ».</after>
</modif>
```

#### 1.2.1.3. Emprunt

Il s'agit du remplacement d'un emprunt par la forme correspondante en français « standard » (**exemple 1**).

#### Exemple 1

```
<modif id="30" wp_page_id="10" wp_before_rev_id="193689" wp_after_rev_id="207083"
wp_user_id="1119" wp_user_num_modif="18610" wp_comment="implémentation anglicisme ; codage">
<before>Un algorithme est une procédure de résolution de problème énoncée sous la forme d' une série d'
opérations à effectuer . <m num_words="2">L' implémentation</m> de l' algorithme consiste en l'
écriture de ces opérations dans un langage de programmation</before>
<after>Un algorithme est une procédure de résolution de problème énoncée sous la forme d' une série d'
opérations à effectuer . <m num_words="4">La mise en œuvre</m> de l' algorithme consiste en l'
écriture de ces opérations dans un langage de programmation</after>
</modif>
```

#### 1.2.1.4. Régionalisme

Il s'agit du remplacement d'un « régionalisme » par la forme équivalente en français « standard » (**exemple 1**).

#### Exemple 1

```
<modif id="407249" wp_page_id="1778615" wp_before_rev_id="19527897" wp_after_rev_id="19528735"
wp_user_id="25450" wp_user_num_modif="4243" wp_comment="+ Portail histoire militaire">
<before>Il décède à <m num_words="1">nonante</m> ans , dans un faubourg de Hambourg ,
le .</before>
<after>Il décède à <m num_words="2">quatre-vingt dix</m> ans , dans un faubourg de Hambourg , le
.</after>
</modif>
```

#### 1.2.1.5. Anaphore

Il s'agit du remplacement d'un segment par sa représentation anaphorique (**exemples 1**) et inversement (**exemple 2**).

#### Exemple 1

```
<modif id="408678" wp_page_id="1963200" wp_before_rev_id="20807856" wp_after_rev_id="20814476"
wp_user_id="21269" wp_user_num_modif="9208" wp_comment="mise en page">
<before> <m num_words="2">Michael Sulick</m> a été nommé Directeur du National Clandestine
Service de la CIA le mercredi 12 septembre 2007 .</before>
<after> <m num_words="1">Il</m> a été nommé Directeur du National Clandestine Service de la CIA le
mercredi 12 septembre 2007 .</after>
</modif>
```

#### Exemple 2

```
<modif id="408649" wp_page_id="1957952" wp_before_rev_id="20842080" wp_after_rev_id="20842367"
wp_user_id="29599" wp_user_num_modif="734" wp_comment="Lien interne+modifs">
<before> <m num_words="2">Ce navire</m> devait servir de décor pour un film Hollywoodien , basé
sur une histoire d' officier de marine harcelé par ses coéquipiers .</before>
<after> <m num_words="4">Le Príncipe de Asturias</m> devait servir de décor pour un film
Hollywoodien , basé sur une histoire d' officier de marine harcelé par ses coéquipiers .</after>
</modif>
```

#### 1.2.1.6. Partie du discours

Cette reformulation consiste en un changement de partie du discours d'un item (**exemples 1 & 2**).

#### Exemple 1

```
<modif id="136138" wp_page_id="53258" wp_before_rev_id="2744705" wp_after_rev_id="2803251"
wp_user_id="12852" wp_user_num_modif="12144" wp_comment="tableau +liens">
<before>La bataille d' Auerstaedt opposa l' armée prussienne à une partie de l' armée française menée
par Davout le 14 octobre 1806 <m num_words="3">en parallèle de</m> la bataille d' léna .</before>
<after>La bataille d' Auerstaedt opposa l' armée prussienne à une partie de l' armée française menée par
Louis Nicolas Davout le 14 octobre 1806 <m num_words="2">, parallèlement à</m> la bataille d'
léna .</after>
</modif>
```

## Exemple 2

```
<modif id="142411" wp_page_id="57394" wp_before_rev_id="14673152" wp_after_rev_id="14868464"
wp_user_id="217715" wp_user_num_modif="484" wp_comment="Uniformisation de la mise en page">
<before>Arts et Métiers est une station du métro <m num_words="1">parisien</m> sur les lignes 3 et
11 .</before>
<after>Arts et Métiers est une station du métro <m num_words="2">de Paris</m> sur les lignes 3 et 11
, dans le 3 e arrondissement de Paris .</after>
</modif>
```

## 1.2.2. Reformulations syntaxiques

### 1.2.2.1. Permutation

Ce procédé se caractérise principalement par une modification de l'ordre des segments (**exemple 1**).

#### Exemple 1

```
<modif id="144" wp_page_id="15" wp_before_rev_id="1994029" wp_after_rev_id="1994946"
wp_user_id="18910" wp_user_num_modif="10198" wp_comment="Culture typo">
<before>La pratique religieuse y était de 35 % dans les années 1950 ( source : <m num_words="2">I
invention</m> de l' Europe d' Emmanuel Todd ) .</before>
<after>La pratique religieuse y était de 35 % dans les années 1950 ( source : <m
num_words="4">Emmanuel Todd , L' Invention</m> de l' Europe ) .</after>
</modif>
```

### 1.2.2.2. Voix active / passive

Dans ce cadre de reformulations syntaxiques, une modification de la conjugaison au niveau de la voix - passive ou active - implique une réorganisation structurelle (**exemple 1**) de la phrase.

#### Exemple 1

```
<modif id="141465" wp_page_id="56943" wp_before_rev_id="296565" wp_after_rev_id="296567"
wp_user_id="776" wp_user_num_modif="32085" wp_comment="">
<before>** Chapeau melon et bottes de cuir <m num_words="4">porté par John Steed</m>
</before>
<after>** John Steed le porte dans Chapeau melon et bottes de cuir <m num_words="2">( The
Avengers )</m> </after>
</modif>
```

### 1.2.2.3. Type de proposition

Cette reformulation correspond à une réorganisation structurelle de la phrase (**exemples 1 & 2**), potentiellement enrichie ou appauvrie d'un « mot-outil » comme une conjonction, lorsque cela est nécessaire à la conformité syntaxique du nouveau segment.

#### Exemple 1

Passage d'une proposition subordonnée circonstancielle à une proposition subordonnée relative.

```
<modif id="407636" wp_page_id="1815514" wp_before_rev_id="19482140" wp_after_rev_id="19482150"
wp_user_id="216654" wp_user_num_modif="154" wp_comment="Controverse">
<before>En 2004 , trois membres de l' école de Sagesse de Ramtha produisirent un film se proclamant
être un documentaire scientifique portant sur la mécanique quantique et ses impliquassions sur la nature
de la vie dans lequel apparait Mme . Hampton en train , sois disant , de " channeller " Ramtha , appelé What
the Bleep Do We Know ! ?. Le film a sévèrement été critiqué de la part de la communauté scientifique qui le
classifie comme étant un infomercial pseudo-scientifique <m num_words="2">en exposant</m>
grossièrement certains des principes pointus de la mécanique quantique . " The minds boggle " . The
Guardian Unlimited '</before>
<after>En 2004 , trois membres de l' école de Sagesse de Ramtha produisirent un film se proclamant être
un documentaire scientifique portant sur la mécanique quantique et ses impliquassions sur la nature de la
vie dans lequel apparait Mme . Hampton en train , sois disant , de " channeller " Ramtha , appelé What the
Bleep Do We Know ! ?. Le film a sévèrement été critiqué de la part de la communauté scientifique qui le
classifie comme étant un infomercial pseudo-scientifique <m num_words="2">qui expose</m>
grossièrement certains des principes pointus de la mécanique quantique . " The minds boggle " . The
Guardian Unlimited '</after>
</modif>
```

## Exemple 2

Passage d'une proposition subordonnée relative à une proposition indépendante.

```
<modif id="406708" wp_page_id="1696582" wp_before_rev_id="9832569" wp_after_rev_id="9832583" wp_user_id="38460" wp_user_num_modif="4852" wp_comment="La diplomatie hitlérienne pendant la guerre">
<before>Cependant , Hitler échoue en ce concerne l' Espagne franquiste . Comptant sur la reconnaissance du Caudillo qui a gagné la guerre civile espagnole en grande partie grâce au soutien de Hitler , ce dernier le rencontre à Hendaye le 23 octobre 1940 . Hitler espère notamment pouvoir obtenir l' autorisation de Franco pour conquérir Gibraltar et couper les voies de communications anglaises en Méditerranée . Les contreparties exigées par Franco ( notamment des compensations territoriales en Afrique du Nord française ) , dont le pays est par ailleurs ruiné , sont irréalisables pour Hitler , qui souhaite ménager quelque peu le régime de Vichy pour l' amener sur la voie de la collaboration <m num_words="1">, qui</m> sera d' ailleurs officialisée le lendemain par Pétain lors de sont entrevue avec le Führer à Montoire .</before>
<after>Cependant , Hitler échoue en ce concerne l' Espagne franquiste . Comptant sur la reconnaissance du Caudillo qui a gagné la guerre civile espagnole en grande partie grâce au soutien de Hitler , ce dernier le rencontre à Hendaye le 23 octobre 1940 . Hitler espère notamment pouvoir obtenir l' autorisation de Franco pour conquérir Gibraltar et couper les voies de communications anglaises en Méditerranée . Les contreparties exigées par Franco ( notamment des compensations territoriales en Afrique du Nord française ) , dont le pays est par ailleurs ruiné , sont irréalisables pour Hitler , qui souhaite ménager quelque peu le régime de Vichy pour l' amener sur la voie de la collaboration <m num_words="1">. Celle-ci</m> sera d' ailleurs officialisée le lendemain par Pétain lors de sont entrevue avec le Führer à Montoire .</after>
</modif>
```

## 1.2.3. Reformulations sémantiques

### 1.2.3.1. Couverture sémantique

Une reformulation du segment source impliquant un élargissement ou une réduction de la « couverture sémantique » se caractérise par un phénomène d'hyponymie (**exemples 1 & 2**) ou d'hyponymie et méronymie (**exemples 3 & 4**).

#### Exemple 1

```
<modif id="408792" wp_page_id="1993884" wp_before_rev_id="11790535" wp_after_rev_id="11790600" wp_user_id="162423" wp_user_num_modif="104" wp_comment="Description technique">
<before>Un support arrière de charge ou dossier est installé quand la charge est plus haute que le haut du porte-charge ; c' est une extension semblable à un ratelier boulonné ou soudée au <m num_words="1">transporteur</m> pour empêcher la charge de se déplacer en arrière</before>
<after>Un support arrière de charge ou dossier est installé quand la charge est plus haute que le haut du porte-charge ; c' est une extension semblable à un ratelier boulonné ou soudée au <m num_words="1">porte-charge</m> pour empêcher la charge de se déplacer en arrière</after>
</modif>
```

#### Exemple 2

```
<modif id="408154" wp_page_id="1891160" wp_before_rev_id="19968798" wp_after_rev_id="19971511" wp_user_id="0" wp_user_num_modif="1096911" wp_comment="">
<before>L' engagement politique de Ali Bach Hamba prend forme à travers son expérience dans le journalisme . Doté d' un vrai talent pour l' expression et l' analyse , il fonde le <m num_words="1">journal</m> francophone « Le Tunisien » en 1907 et il devient le chef de sa rédaction politique .</before>
<after>L' engagement politique de Bach Hamba prend forme à travers son expérience dans le journalisme . Doté d' un talent pour l' expression et l' analyse , il fonde le <m num_words="1">quotidien</m> francophone Le Tunisien en 1907 et devient le chef de sa rédaction politique .</after>
</modif>
```



### Exemple 3

```
<modif id="141329" wp_page_id="56859" wp_before_rev_id="619501" wp_after_rev_id="628974"
wp_user_id="4980" wp_user_num_modif="2281" wp_comment="Destructions de la bibliothèque">
<before>Le fait que beaucoup d' écrits d' Aristote soient parvenus jusqu ' à nous suggère en tout cas que
si ordres de destruction il y a eu , il s' est trouvé sur place quelques volontaires pour sauver
clandestinement cette partie des documents , et que ces volontaires avaient droit à la neutralité
bienveillante des troupes d' occupation-à supposer qu ' ils n' en aient pas fait partie . La tradition raconte d'
ailleurs en effet que les <m num_words="1">manuscrits</m> d' Aristote échappèrent en partie à la
destruction .</before>
<after>Le fait que beaucoup d' écrits d' Aristote soient parvenus jusqu ' à nous suggère en tout cas que si
ordres de destruction il y a eu , il s' est trouvé sur place quelques volontaires pour sauver clandestinement
cette partie des documents , et que ces volontaires avaient droit à la neutralité bienveillante des troupes d'
occupation-à supposer qu ' ils n' en aient pas fait partie . La tradition raconte d' ailleurs en effet que les <m
num_words="1">textes</m> d' Aristote échappèrent en partie à la destruction .</after>
</modif>
```

### Exemple 4

```
<modif id="142228" wp_page_id="57225" wp_before_rev_id="7458919" wp_after_rev_id="8510987"
wp_user_id="91974" wp_user_num_modif="5690" wp_comment="Ajout de la nationalité">
<before>Elle semble avoir bénéficié d' une mesure de clémence , puisque la famille regagnera <m
num_words="2">l' Angleterre</m> en 1959 , cinq ans après le drame .</before>
<after>Elle semble avoir bénéficié d' une mesure de clémence , puisque la famille regagnera <m
num_words="2">le Royaume-Uni</m> en 1959 , cinq ans après le drame .</after>
</modif>
```

### 1.2.3.2. Normalisation encyclopédique

Remplacement du segment source par une expression équivalente plus formelle (**exemples 1 & 2**). Ce phénomène peut également concerner la conjugaison, lorsque l'on passe d'un temps relevant traditionnellement de l'oral à un temps plus formel (**exemple 3**).

#### Exemple 1

```
<modif id="182822" wp_page_id="94894" wp_before_rev_id="19700081" wp_after_rev_id="20240199"
wp_user_id="271994" wp_user_num_modif="1621" wp_comment="corr nom">
<before>Le château a été rénové à travers son existence , et plusieurs parties anciennes ont été
recouvertes ou transformées par des constructions ultérieures et des additions . De 1952 à 1966 , le
gouvernement de <m num_words="5">l' Allemagne de l' Est</m> le restaura comme il était au ,
notamment la pièce de Luther avec ses planchers et murs en panneaux de bois originaux .</before>
<after>Le château a été rénové à travers son existence , et plusieurs parties anciennes ont été
recouvertes ou transformées par des constructions ultérieures et des additions . De 1952 à 1966 , le
gouvernement de <m num_words="4">la République démocratique allemande</m> le restaura
comme il était au , notamment la pièce de Luther avec ses planchers et murs en panneaux de bois
originaux .</after>
</modif>
```

#### Exemple 2

```
<modif id="408442" wp_page_id="1924410" wp_before_rev_id="16853928" wp_after_rev_id="16858893"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="">
<before>La construction des Forts-de-Lévis a débuté en 1865 sous le régime britannique en prévision d'
une éventuelle attaque américaine . À cette époque , la <m num_words="3">guerre civile
américaine</m> fait rage aux États-Unis . La Grande-Bretagne craint que l' armée de l' Union ( États du
Nord ) envahisse le Canada pour récupérer le terrain qui pourrait être perdu aux Confédérés ( États du Sud )
lors de la guerre civile .</before>
<after>La construction des Forts-de-Lévis a débuté en 1865 sous le régime britannique en prévision d' une
éventuelle attaque américaine . À cette époque , la <m num_words="3">Guerre de Sécession</m> fait
rage aux États-Unis . La Grande-Bretagne craint que l' armée de l' Union ( États du Nord ) envahisse le
Canada pour récupérer le terrain qui pourrait être perdu aux Confédérés ( États du Sud ) lors de la guerre
civile .</after>
</modif>
```

### Exemple 3

```
<modif id="54" wp_page_id="10" wp_before_rev_id="7772394" wp_after_rev_id="7772472"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="">
<before>Le savant arabophone andalou Averroès ( 1126 -1198 ) évoque une méthode de raisonnement où
la thèse s' affine étape par étape ( itérativement ) jusqu ' à une certaine convergence , et ceci
conformément au déroulement d' un algorithme . A la même époque , au douzième siècle , le moine
Adelard of Bath introduit en langue latine le terme de " algorismus " ( par référence au nom de Al-
Khuwarizmi ) . Ce mot <m num_words="2">va donner </m> algorithme " en français en
1554 .</before>
<after>Le savant arabophone andalou Averroès ( 1126 -1198 ) évoque une méthode de raisonnement où
la thèse s' affine étape par étape ( itérativement ) jusqu ' à une certaine convergence , et ceci
conformément au déroulement d' un algorithme . A la même époque , au douzième siècle , le moine
Adelard of Bath a introduit le terme latin de algorismus ( par référence au nom de Al-Khuwarizmi ) . Ce mot
<m num_words="1">donnera</m> algorithme en français en 1554 .</after>
</modif>
```

#### 1.2.3.3. Synonymie

Il s'agit de remplacer un item ou une expression par un item ou une expression synonymique (**exemples 1 & 2**).

Certaines reformulations synonymiques correspondent à un glissement de la langue générale vers une langue de spécialité (**exemple 3**) et inversement (**exemple 4**).

#### Exemple 1

```
<modif id="408685" wp_page_id="1963959" wp_before_rev_id="20817134" wp_after_rev_id="20829846"
wp_user_id="211351" wp_user_num_modif="366" wp_comment="">
<before>Il est le maître direct de don Julian Osorio , et <m num_words="2">le disciple</m> de don
Rosendo .</before>
<after>Il est le maître direct de don Julian Osorio , et <m num_words="2">l' élève</m> de don
Rosendo .</after>
</modif>
```

#### Exemple 2

```
<modif id="408656" wp_page_id="1959732" wp_before_rev_id="20757675" wp_after_rev_id="20757719"
wp_user_id="226605" wp_user_num_modif="780" wp_comment="La mort de Corridoni">
<before>Il est décoré de la médaille d' argent de la valeur militaire , décoration que Benito Mussolini <m
num_words="2">fera convertir</m> en médaille d' or en 1925 .</before>
<after>Il est décoré de la médaille d' argent de la valeur militaire , décoration que Benito Mussolini <m
num_words="1">transformera</m> en médaille d' or en 1925 .</after>
</modif>
```

#### Exemple 3

Dans le segment cible, « paradigme » est à prendre dans son acception épistémologique.

```
<modif id="142053" wp_page_id="57162" wp_before_rev_id="5182159" wp_after_rev_id="6778768"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="">
<before>* <m num_words="3">à la branche</m> de la psychologie sociale qui étudie les processus
mentaux de perception , mémorisation et raisonnement impliqués dans l' interaction sociale au niveau des
groupes ou des foules ;</before>
<after>* <m num_words="1">Paradigme</m> de la psychologie sociale qui étudie les processus
mentaux de perception , mémorisation et raisonnement impliqués dans l' interaction sociale au niveau des
groupes ou des foules ;</after>
</modif>
```

#### Exemple 4

```
<modif id="407851" wp_page_id="1830844" wp_before_rev_id="20691183" wp_after_rev_id="20691225"
wp_user_id="287861" wp_user_num_modif="81" wp_comment="">
<before>Archaeopteris appartient au groupe fossile des progymnospermesBECK , C . B . , et WIGHT , D .
C . 1988 . Progymnosperms . In C . B . Beck Origin and evolution of gymnosperms . Columbia University
Press , New York . 57-61 . et est considéré par de nombreux scientifiques comme le premier arbre
moderneMEYER-BERTHAUD , B . , SCHECKLER , S . E . , et WENDT , J . 1999 . http : // www . nature . com /
nature / journal / v398 / n6729 / abs / 398700a0 . html Archaeopteris is the earliest known modern tree .
Nature 398 : 700-701 . . Le genre Archaeopteris possède plus de caractéristiques communes avec les
plantes à graines que toute autre <m num_words="1">ptéridophyte</m> connue et les analyses
cladistiques récentes le placent en groupe-frère des plantes à graines .</before>
<after>Archaeopteris appartient au groupe fossile des progymnospermesBECK , C . B . , et WIGHT , D . C .
1988 . Progymnosperms . In C . B . Beck Origin and evolution of gymnosperms . Columbia University Press ,
New York . 57-61 . et est considéré par de nombreux scientifiques comme le premier arbre moderneMEYER-
BERTHAUD , B . , SCHECKLER , S . E . , et WENDT , J . 1999 . http : // www . nature . com / nature / journal /
v398 / n6729 / abs / 398700a0 . html Archaeopteris is the earliest known modern tree . Nature 398 : 700-
701 . . Le genre Archaeopteris possède plus de caractéristiques communes avec les plantes à graines que
toute autre <m num_words="2">plante fossile</m> connue et les analyses cladistiques récentes le
placent en groupe-frère des plantes à graines .</after>
</modif>
```

#### 1.2.3.4. Neutralité

Wikipédia comprend un certain nombre d'essais et de recommandations à ce propos, présentant notamment une liste de termes ou de tournures à éviter<sup>5</sup> : soit à cause de leur trop grande connotation, soit car ils impliquent une prise de position et violent le principe de neutralité d'une encyclopédie.

La neutralité (ou au contraire la partialité) d'un terme ou d'une expression peut prendre plusieurs formes et est à l'origine de nombreuses reformulations se manifestant principalement par les biais suivants :

- suppression d'un item (**exemples 1 & 2**) ;
- paraphrasage privilégiant une tournure neutre (**exemple 3**) et inversement (**exemple 4**) ;
- substitution d'un item du segment source par un item non connoté (**exemple 5 et 6**).

#### Exemple 1

```
<modif id="408782" wp_page_id="1987705" wp_before_rev_id="21113913" wp_after_rev_id="21131263"
wp_user_id="24387" wp_user_num_modif="14" wp_comment="">
<before> <m num_words="2">Très controversée</m> dans les milieux universitaires , la Théorie de la
continuité paléolithique dont il est l' auteur , suggère une différenciation linguistique entre les langues d'
Europe non pas postérieure , mais antérieure au néolithique . Il va jusqu' à remettre implicitement en doute
l' existence , ou en tout cas l' importance , d' une famille de langues indo-européennes en déclarant
notamment :</before>
<after> <m num_words="1">Controversée</m> dans les milieux universitaires , la Théorie de la
continuité paléolithique dont il est l' auteur , suggère une différenciation linguistique entre les langues d'
Europe non pas postérieure , mais antérieure au néolithique . Il va jusqu' à remettre implicitement en doute
l' existence , ou en tout cas l' importance , d' une famille de langues indo-européennes en déclarant
notamment :</after>
</modif>
```

#### Exemple 2

```
<modif id="406692" wp_page_id="1696582" wp_before_rev_id="9558605" wp_after_rev_id="9558611"
wp_user_id="51163" wp_user_num_modif="908" wp_comment="">
<before>*« Dans le nouveau Reich , il ne devra plus y avoir de place pour la cruauté envers les bêtes . »
( Extraits d' un discours d' Adolph Hitler , ces propos <m num_words="2">sympathiques inspirant</m>
l' imposante loi du 24 novembre sur la protection des animaux : " Tierschutzgesetz " )</before>
<after>*« Dans le nouveau Reich , il ne devra plus y avoir de place pour la cruauté envers les bêtes . »
( Extraits d' un discours d' Adolph Hitler , ces propos <m num_words="1">inspirant</m> l' imposante loi
du 24 novembre sur la protection des animaux : " Tierschutzgesetz " )</after>
</modif>
```

5 Recommandations de Wikipédia sur le style encyclopédique : <[http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Style\\_encyclop%C3%A9dique](http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Style_encyclop%C3%A9dique)>

Page d'essai de Wikipédia sur les termes à utiliser avec précaution : <[http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Termes\\_%C3%A0\\_utiliser\\_avec\\_pr%C3%A9caution](http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Termes_%C3%A0_utiliser_avec_pr%C3%A9caution)>

### Exemple 3

```
<modif id="408700" wp_page_id="1966685" wp_before_rev_id="15250025" wp_after_rev_id="15250313"
wp_user_id="196940" wp_user_num_modif="1124" wp_comment="Révocation des modifications par
KurtGL ; retour à la version de 193 . 67 . 23 . 131 ( changement non neutre )">
<before>Tri Yann est l' un des groupes de musique celtique les plus connus en France ; il a participé à la
diffusion d' airs traditionnels comme La jument de Michao ( <m num_words="4">souvent appelé à
tort</m> Le loup , le renard et la belette ) , Les Filles des Forges ou Dans les prisons de
Nantes .</before>
<after>Tri Yann est l' un des groupes de musique celtique les plus connus en France ; il a participé à la
diffusion d' airs traditionnels comme La jument de Michao ( <m num_words="5">aussi connu sous le
nom</m> Le loup , le renard et la belette ) , Les Filles des Forges ou Dans les prisons de Nantes .</after>
</modif>
```

### Exemple 4

```
<modif id="408629" wp_page_id="1951496" wp_before_rev_id="20704308" wp_after_rev_id="20704374"
wp_user_id="276939" wp_user_num_modif="659" wp_comment="Porte-drapeau français">
<before> <m num_words="5">En ce qui concerne les</m> Jeux Olympiques d' été , Marie-José Percec
et David Douillet sont les seuls athlètes français a avoir obtenu un titre olympique individuel en tant que
porte-drapeau .</before>
<after> <m num_words="6">Il est à noter qu ' aux</m> Jeux Olympiques d' été , Marie-José Percec et
David Douillet sont les seuls athlètes français a avoir obtenu un titre olympique individuel en tant que
porte-drapeau .</after>
</modif>
```

### Exemple 5

```
<modif id="407967" wp_page_id="1862414" wp_before_rev_id="19560273" wp_after_rev_id="19571692"
wp_user_id="185546" wp_user_num_modif="796" wp_comment="wikif">
<before>Túpac Katari est un <m num_words="1">héros</m> de la révolte Indienne en Bolivie face à l'
occupation néo-colonialiste espagnole .</before>
<after>Túpac Katari est un <m num_words="1">protagoniste</m> de la révolte Indienne en Bolivie
face à l' occupation néo-colonialiste espagnole .</after>
</modif>
```

### Exemple 6

```
<modif id="407096" wp_page_id="1763454" wp_before_rev_id="19320568" wp_after_rev_id="19888889"
wp_user_id="160840" wp_user_num_modif="490" wp_comment="Fonctions Reformulation">
<before>* Principalement d' aider les comités de filiale à régler les questions et les problèmes liés à l'
activité de <m num_words="1">prosélytisme</m> .</before>
<after>* Principalement d' aider les comités de filiale à régler les questions et les problèmes liés à l'
activité de <m num_words="2">prédication publique</m> .</after>
</modif>
```

## 1.2.3.5. Précision

Les procédés permettant d'affiner le sens d'une phrase sont les suivants :

- insertion d'un item (**exemple 1**) ;
- remplacement du segment source par un segment cible au sens plus précis (**exemples 2 & 3**) ;
- paraphrasage (**exemples 4 & 5**).

### Exemple 1

```
<modif id="407127" wp_page_id="1766137" wp_before_rev_id="18629976" wp_after_rev_id="18630042"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="">
<before> <m num_words="1">Club</m> de football corporatif , basé à Montpellier , vainqueur de la
Coupe de France des clubs corporatifs en 1974 .</before>
<after> <m num_words="2">Ancien club</m> de football corporatif , basé à Montpellier , vainqueur de
la Coupe de France des clubs corporatifs en 1974 .</after>
</modif>
```

### Exemple 2

```
<modif id="408614" wp_page_id="1950526" wp_before_rev_id="20647367" wp_after_rev_id="20647824"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="">
<before>* Alors que le mini-jeu inclus dans Final Fantasy VII ne propose qu ' une piste jouable , qui se
divise deux fois en deux au cours de la descente , la version pour téléphone portable propose <m
num_words="1">plusieurs</m> pistes différentes .</before>
<after>* Alors que le mini-jeu inclus dans Final Fantasy VII ne propose qu ' une piste jouable , qui se divise
deux fois en deux au cours de la descente , la version pour téléphone portable propose <m
num_words="1">trois</m> pistes différentes . Seule l' une est disponible au début du jeu . Pour débloquent
l' accès aux deux suivantes , il faut terminer les précédentes dans un temps limité et / ou en accumulant un
certain nombre de points .</after>
</modif>
```

### Exemple 3

```
<modif id="408615" wp_page_id="1950801" wp_before_rev_id="20615318" wp_after_rev_id="20625857"
wp_user_id="5452" wp_user_num_modif="5075" wp_comment="">
<before>Jules Benoit dit Jules de Gastyne est un écrivain de langue française , né à Sanxay ( Vienne ) en
1847 , mort <m num_words="1">en</m> 1920 .</before>
<after>Jules Benoit dit Jules de Gastyne est un écrivain de langue française , né à Sanxay ( Vienne ) en
1847 , mort <m num_words="3">le 14 juin</m> 1920 à Paris .</after>
</modif>
```

### Exemple 4

```
<modif id="135974" wp_page_id="53090" wp_before_rev_id="271672" wp_after_rev_id="408988"
wp_user_id="1175" wp_user_num_modif="2271" wp_comment="rhétorique">
<before>L' antiphrase est <m num_words="4">un mode d' expression</m> consistant à dire le
contraire de ce que l' on pense .</before>
<after>L' antiphrase est <m num_words="4">une figure de rhétorique</m> consistant à dire le
contraire de ce que l' on pense .</after>
</modif>
```

### Exemple 5

```
<modif id="135936" wp_page_id="53083" wp_before_rev_id="19979505" wp_after_rev_id="19979969"
wp_user_id="40842" wp_user_num_modif="116" wp_comment="Origine">
<before>Par la suite , la poésie moderne ira plus loin , comme dans ce vers de Nuit rhénane d' Apollinaire
<m num_words="6">, qui paraît presque sans structure rythmique</m> :</before>
<after>Par la suite , la poésie moderne ira plus loin , comme dans ce vers de Nuit rhénane d' Apollinaire
<m num_words="6">dont la césure est comme masquée</m> :</after>
</modif>
```

## 2. FORTE VARIATION SÉMANTIQUE

### 2.1. CORRECTIONS FACTUELLES

Même s'il existe une forte différence sémantique entre le segment source et le segment cible, ils sont tout de même sémantiquement liés dans les cas d'antonymie et de modifications portant sur la conjugaison. Ce n'est pas le cas des rectificatifs pour lesquels le lien sémantique entre le segment source et le segment cible n'est pas aisé à établir voire est inexistant.

La plupart de ces modifications interviennent pour corriger le fond de l'article dont il est question, pour rectifier une information erronée, trop imprécise ou peut être inintéressante dans le cadre d'une encyclopédie.

#### 2.1.1. Antonymie

Il s'agit de remplacer un item ou une expression par un item ou une expression antonymique (**exemples 1 & 2**).

Exemple 1

```
<modif id="141865" wp_page_id="57079" wp_before_rev_id="6966477" wp_after_rev_id="7201482"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="Les principes">
<before>La catalyse hétérogène se produit quand le catalyseur et les réactifs ne sont pas sous la même
phase . L' immense majorité des cas de catalyse hétérogène fait intervenir un catalyseur sous forme solide ,
les réactifs étant alors gazeux ou liquide . Précisons que les réactifs peuvent se trouver à l' état liquide pour
certains et en même temps gazeux pour d' autres , c' est le cas pour une hydrogénation d' un composé
organique en solution ( phase liquide ) sur un catalyseur solide ( phase <m
num_words="1">liquide</m> ) avec de l' hydrogène ( phase gazeuse ) . L' avantage de ce type de
catalyse est la grande facilité de séparation du catalyseur des produits et réactifs puisqu' une simple
filtration suffit .</before>
<after>La catalyse hétérogène se produit quand le catalyseur et les réactifs ne sont pas sous la même
phase . L' immense majorité des cas de catalyse hétérogène fait intervenir un catalyseur sous forme solide ,
les réactifs étant alors gazeux ou liquide . Précisons que les réactifs peuvent se trouver à l' état liquide pour
certains et en même temps gazeux pour d' autres , c' est le cas pour une hydrogénation d' un composé
organique en solution ( phase liquide ) sur un catalyseur solide ( phase <m num_words="1">solide</m> )
avec de l' hydrogène ( phase gazeuse ) . L' avantage de ce type de catalyse est la grande facilité de
séparation du catalyseur des produits et réactifs puisqu' une simple filtration suffit .</after>
</modif>
```

Exemple 2

```
<modif id="142167" wp_page_id="57218" wp_before_rev_id="1485175" wp_after_rev_id="1485191"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="Généalogie">
<before>En revanche , l' un de ses arrière-grand-oncles , Alphonse Chodron de Courcel ( 1835 -1919 ) ,
frère de Georges , fut créé baron héréditaire par lettres patentes du 6 mars 1867 , et était par ailleurs
docteur en droit , ambassadeur de France auprès de l' Empereur allemand ( 1881 -1886 ) , représentant de
la France à la Conférence africaine de Berlin ( 1884 -1885 ) , ambassadeur à Londres ( 1894 -1898 ) ,
sénateur de Seine-et-Oise , président du Conseil d' administration du chemin de fer Paris-Orléans et
membre de l' Académie des sciences morales et politiques . Cet arrière-grand-oncle a transmis la qualité de
noble à sa seule descendance , <m num_words="6">à laquelle n' appartient pas madame</m> Chirac
.</before>
<after>En revanche , l' un de ses arrière-grand-oncles , Alphonse Chodron de Courcel ( 1835 -1919 ) , frère
de Georges , fut créé baron héréditaire par lettres patentes du 6 mars 1867 , et était par ailleurs docteur en
droit , ambassadeur de France auprès de l' Empereur allemand ( 1881 -1886 ) , représentant de la France à
la Conférence africaine de Berlin ( 1884 -1885 ) , ambassadeur à Londres ( 1894 -1898 ) , sénateur de
Seine-et-Oise , président du Conseil d' administration du chemin de fer Paris-Orléans et membre de l'
Académie des sciences morales et politiques . Cet arrière-grand-oncle a transmis la qualité de noble à sa
descendance , <m num_words="2">dont Madame</m> Chirac .</after>
</modif>
```

## 2.1.2. Conjugaison

Une modification de la conjugaison (mode, temps, personne, etc.) implique un changement de sens entre le segment source et le segment cible et dans de nombreux cas une contradiction avec le sens du segment source (**exemples 1 & 2**).

Exemple 1

```
<modif id="102" wp_page_id="12" wp_before_rev_id="2846042" wp_after_rev_id="2846064"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="Histoire">
<before>Incapable de convaincre Londres de retirer ses soldats , l' IRA annonce un cessez-le-feu ( 1994 ) .
Le Sinn Féin commence à entreprendre des négociations de paix avec la Grande-Bretagne . La paix est
cependant précaire ( reprise des attentats , 1996-1997 ) . Un accord de paix du « Vendredi Saint » ( avril
1998 ) est conclu à Belfast et il est approuvé par référendum en Irlande du Nord et en République d'
Irlande . Ce plan prévoit la création d' une nouvelle Assemblée réunissant les protestants et les catholiques
et la coopération entre l' Irlande du Nord et la République d' Irlande . Le nouveau gouvernement dirigé par
un protestant est mis en place en Irlande du Nord ( 1999 ) . Toutefois , les tensions entre les protestants et
les catholiques au sujet du fonctionnement des institutions politiques et le refus de l' IRA de désarmer <m
num_words="1">font</m> planer des incertitudes sur l' avenir de la paix .</before>
<after>Toutefois , les tensions entre les protestants et les catholiques au sujet du fonctionnement des
institutions politiques et le refus de l' IRA de désarmer <m num_words="3">ont longtemps fait</m>
planer des incertitudes sur l' avenir de la paix . Mais le jeudi 28 juillet 2005 marqua un tournant décisif .
Dans un communiqué , l' IRA ordonna à tous ses membres de déposer les armes . Ce qui mit fin
officiellement à une lutte armée devenue de plus en plus impopulaire en Irlande .</after>
</modif>
```

Exemple 2

```
<modif id="408763" wp_page_id="1978847" wp_before_rev_id="20776908" wp_after_rev_id="20802905"
wp_user_id="184925" wp_user_num_modif="96" wp_comment="">
<before>Créé en 1882 , laâzoguen fût le village du chef lieu de canton de la commune du Haut-Sébaou .
laâzoguen est une localité riche par sa culture et son histoire . En kabyle , son nom signifie « les sourds » ,
noms qui <m num_words="1">auraient</m> été donné aux habitants du village par les Français au
moment de la colonisation . De plus , le nom d' laâzoguen est intimement lié à tout ce qui se fait de mieux
dans la gastronomie nationale . À ce titre on dit souvent qu' à Azazga , entre deux restaurants , il y a
souvent un autre restaurant et tous rivalisent dans l' art de mitonner les plats .</before>
<after>Créé en 1882 , laâzoguen fût le village de la commune du Haut-Sébaou . laâzoguen est une localité
riche par sa culture et son histoire . En kabyle , son nom signifie « les sourds » , noms qui <m
num_words="1">avait</m> été donné aux habitants du village par les Français .</after>
</modif>
```

## 2.1.3. Rectificatif

Même s'il est difficile à identifier au niveau de la modification, le lien sémantique entre segment source et segment cible, bien que ténu, existe dans certains cas (**exemples 1, 2 & 3**) mais peut être totalement inexistant pour d'autres (**exemples 4, 5 & 6**).

Exemple 1

```
<modif id="135939" wp_page_id="53083" wp_before_rev_id="19979969" wp_after_rev_id="19980017"
wp_user_id="40842" wp_user_num_modif="116" wp_comment="Origine">
<before>Par la suite , la poésie moderne ira plus loin , comme dans ce vers de Nuit rhénane d' Apollinaire
<m num_words="6">dont la césure est comme masquée</m> :</before>
<after>Par la suite , la poésie moderne ira plus loin , comme dans ce vers de Nuit rhénane d' Apollinaire
<m num_words="6">qui résiste à l' analyse classique</m> :</after>
</modif>
```

Exemple 2

```
<modif id="408129" wp_page_id="1884962" wp_before_rev_id="19877776" wp_after_rev_id="20803643"
wp_user_id="38453" wp_user_num_modif="5524" wp_comment="lien">
<before>* Lucius Cornelius Merula , <m num_words="6">prêtre et politicien du Ier siècle</m> av .
J .-C . .</before>
<after>* Lucius Cornelius Merula , <m num_words="8">consul suffect de la République romaine en
87</m> av . J .-C . .</after>
</modif>
```

### Exemple 3

```
<modif id="407983" wp_page_id="1865892" wp_before_rev_id="19623779" wp_after_rev_id="20732373"
wp_user_id="35756" wp_user_num_modif="5663" wp_comment="">
<before>Voyage dans les quatre principales îles des mers d' Afrique est une relation écrite par Jean-
Baptiste Bory de Saint-Vincent au début du XIXe siècle pour décrire le voyage qu ' il fit jusque dans les
Mascareignes <m num_words="5">après avoir pris part à</m> l' expédition Baudin partie du Havre
pour la Nouvelle-Hollande en octobre 1800 .</before>
<after>Voyage dans les quatre principales îles des mers d' Afrique est une relation écrite par Jean-Baptiste
Bory de Saint-Vincent au début du XIXe siècle pour décrire le voyage qu ' il fit jusque dans les Mascareignes
<m num_words="4">dans le cadre de</m> l' expédition Baudin partie du Havre pour la Nouvelle-
Hollande en octobre 1800 .</after>
</modif>
```

### Exemple 4

```
<modif id="407965" wp_page_id="1862366" wp_before_rev_id="19561037" wp_after_rev_id="19592819"
wp_user_id="121111" wp_user_num_modif="197" wp_comment="">
<before>De lui , sur le Forum Maximum <m num_words="2">bruxellois des</m> cartographes , ou
Nedermerckt comme on disait alors , les plans de là Maison du Cygne , de la Petite Boucherie , de la partie
arrière de l' Hôtel de Ville .</before>
<after>De lui , sur le Forum Maximum <m num_words="4">comme disent pompeusement les</m>
cartographes , ou Nedermerckt comme on disait alors , les plans de la Maison du Cygne , de la Petite
Boucherie , de la partie arrière de l' Hôtel de Ville .</after>
</modif>
```

### Exemple 5

```
<modif id="278312" wp_page_id="222050" wp_before_rev_id="15106988" wp_after_rev_id="16773551"
wp_user_id="240666" wp_user_num_modif="1" wp_comment="">
<before>Des groupuscules d' idéologie identitaire et nationaliste utilisent un graphisme simplifié proche
de ce symbole ( un rond barré d' une croix régulière — sans jambage allongé — ) qui représente pour eux
<m num_words="3">l' Occident chrétien</m> .</before>
<after>Des groupuscules d' idéologie identitaire et nationaliste utilisent un graphisme simplifié proche de
ce symbole ( un rond barré d' une croix régulière — sans jambage allongé — ) qui représente pour eux <m
num_words="6">la supériorité de la race celto-germanique</m> .</after>
</modif>
```

### Exemple 6

```
<modif id="408080" wp_page_id="1877262" wp_before_rev_id="19762636" wp_after_rev_id="19772185"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="">
<before>Pierre Jean Louis Ovide Doublet est un homme d' <m num_words="1">église</m> né à Orléans
en 1749 et mort à La Valette ( Malte ) en 1824 .</before>
<after>Pierre Jean Louis Ovide Doublet est un homme d' <m num_words="1">état</m> né à Orléans en
1749 et mort à La Valette ( Malte ) en 1824 .</after>
</modif>
```

## 2.2. VANDALISME

### 2.2.1. Manifeste

Il s'agit de l'insertion d'une chaîne de caractères au sein du segment source, produisant un non-mot au sein du segment cible (**exemples 1 & 2**).

#### Exemple 1

```
<modif id="171" wp_page_id="15" wp_before_rev_id="16368514" wp_after_rev_id="16378313"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="Histoire">
<before>L' Autriche a été occupée <m num_words="1">par</m> les Romains , les Huns , les Lombards ,
les Ostrogoths , les Bavares et les Francs . Les Babenberg dominèrent l' Autriche entre le Xe s et XIIIe
siècle . Les Habsbourg leur succédèrent et restèrent au pouvoir jusqu ' en 1918 , d' abord sous lenom de
Habsbourg jusqu ' en 1780 , puis sous le nom de Habsbourg-Lorraine .</before>
<after>L' Autriche a été occupée <m num_words="2">psh ! ! ar</m> les Romains , les Huns , les
Lombards , les Ostrogoths , les Bavares et les Francs . Les Babenberg dominèrent l' Autriche entre
le</after>
</modif>
```



## Exemple 2

```
<modif id="182964" wp_page_id="95126" wp_before_rev_id="12498990" wp_after_rev_id="12499022"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="">
<before>La zone ouest autour du Jourdain et de la partie supérieure de l' Euphrate donna naissance aux
premiers campements agraires connus il y a 11 000 ans . Les campements les plus anciens connus à ce jour
sont à Iraq ed-Dubb ( Jordanie ) et Tell Aswad ( Syrie ) , suivi de près par Jéricho . Les plus anciennes villes ,
états , et <m num_words="1">écrits</m> connues apparurent plus tard en Mésopotamie ( dont l'
étymologie signifie « entre les fleuves » , faisant référence aux terres comprises entre l' Euphrate et le
Tigre ) à l' est . Ces découvertes ont permis de surnommer la région « le berceau de la
civilisation . »</before>
<after>La zone ouest autour du Jourdain et de la partie supérieure de l' Euphrate donna naissance aux
premiers campements agraires connus il y a 11 000 ans . Les campements les plus anciens connus à ce jour
sont à Iraq ed-Dubb ( Jordanie ) et Tell Aswad ( Syrie ) , suivi de près par Jéricho . Les plus anciennes villes ,
états , et <m num_words="1">écrits</m> connues apparurent plus tard en Mésopotamie
( dont l' étymologie signifie « entre les fleuves » , faisant référence aux terres comprises entre l' Euphrate et
le Tigre ) à l' est . Ces découvertes ont permis de surnommer la région « le berceau de la
civilisation . »</after>
</modif>
```

## 2.2.2. Subtil

Il s'agit de l'insertion d'une chaîne de caractères au sein du segment source, produisant une unité lexicale mais sans aucun sens par rapport au fond de l'article (**exemples 1 & 2**). Ce phénomène se caractérise la plupart du temps par une reformulation sémantique prenant la forme du remplacement d'un item par un autre. Le repérage de ce type de révision fait appel à des acquis extra-linguistiques pour juger si le segment cible est cohérent et admissible par rapport à l'information traitée dans l'article.

## Exemple 1

```
<modif id="142027" wp_page_id="57147" wp_before_rev_id="12268463" wp_after_rev_id="12310310"
wp_user_id="0" wp_user_num_modif="1096911" wp_comment="En France">
<before>Devant la Cour de <m num_words="1">Cassation</m> la représentation des parties est
assurée par des officiers ministériels dits " Avocats au Conseil d' État et à la Cour de Cassation " dits aussi "
Avocats aux Conseils " qui font partie d' une organisation indépendante de celle des autres avocats et qui
disposent d' un privilège de représentation devant ces deux juridictions .</before>
<after>Devant la Cour de <m num_words="1">Castration</m> la représentation des parties est
assurée par des officiers ministériels dits " Avocats au Conseil d' État et à la Cour de Cassation " dits aussi "
Avocats aux Conseils " qui font partie d' une organisation indépendante de celle des autres avocats et qui
disposent d' un privilège de représentation devant ces deux juridictions .</after>
</modif>
```

## Exemple 2

```
<modif id="406827" wp_page_id="1696582" wp_before_rev_id="16643769" wp_after_rev_id="16741459"
wp_user_id="141929" wp_user_num_modif="78" wp_comment="">
<before>Adolf Hitler ( 20 avril 1889 à Braunau am Inn , Autriche - 30 avril 1945 à Berlin ) était un homme
politique allemand , <m num_words="1">Chancelier</m> du Reich , puis Président ( titre qu ' il délaissa
pour celui de Führer ) . Porté par le parti qu ' il créa en 1921 , le NSDAP ( Nationalsozialistische Deutsche
Arbeiterpartei ) , il fonda le Troisième Reich et fut à l' origine de la seconde guerre mondiale .</before>
<after>Adolf Hitler ( 20 avril 1889 à Braunau am Inn , Autriche - 30 avril 1945 à Berlin ) était un homme
politique allemand , <m num_words="1">boulanger</m> du Reich , puis Président ( titre qu ' il délaissa
pour celui de Führer ) . Porté par le parti qu ' il créa en 1921 , le NSDAP ( Nationalsozialistische Deutsche
Arbeiterpartei ) , il fonda le Troisième Reich et fut à l' origine de la seconde guerre mondiale .</after>
</modif>
```

# 3. LIMITES

## 3.1. IDENTIFICATION DES PHÉNOMÈNES

### 3.1.1. Rôle du contexte

Certains phénomènes de reformulation *a priori* relativement simples à identifier - et à catégoriser - demandent en fait une analyse plus fine nécessitant la connaissance d'un contexte élargi (**exemple 1 & 2**).

Exemple 1

En partant du principe que son contexte est neutre, cette modification est à catégoriser dans « reformulations sémantiques > neutralité ». En admettant que l'article porte sur une commune identifiée comme - par exemple - une station de sports d'hiver, le fait de dire qu'elle ne se trouve « qu'à » 500 mètres d'altitude n'est pas une marque de partialité mais est tout à fait justifié compte tenu du contexte.

```
<modif id="408767" wp_page_id="1978847" wp_before_rev_id="20920031" wp_after_rev_id="20920093" wp_user_id="184925" wp_user_num_modif="96" wp_comment="">
<before>D' une superficie de 78 . 657 km² , la commune <m num_words="2">se trouve</m> à 500 mètres d' altitude , entourée de montagnes .</before>
<after>D' une superficie de 78 . 657 km² , la commune <m num_words="3">n' est qu ' </m> à 500 mètres d' altitude , entourée de montagnes .</after>
</modif>
```

Exemple 2

Selon le contexte, la segment source « Platon, Hippodamos ou Aristote » peut être une liste incomplète des philosophes concernés et le segment cible « les philosophes » une reprise anaphorique de philosophes bien identifiés et listés plus haut dans l'article ou une approximation car ces derniers ne sont pas bien identifiés.

```
<modif id="351315" wp_page_id="394748" wp_before_rev_id="12612357" wp_after_rev_id="12613142" wp_user_id="124750" wp_user_num_modif="1717" wp_comment="+">
<before>la République de Platon et les cités existantes . Organisation de l' espace , organisation sociale et organisation politique rationnelles sont les axes selon lesquels <m num_words="4">Platon , Hippodamos ou Aristote</m> pensent la cité idéale .</before>
<after>la République de Platon et les cités existantes . Organisation de l' espace , organisation sociale et organisation politique rationnelles sont les axes selon lesquels <m num_words="2">les philosophes</m> pensent la cité idéale , à laquelle les architectes et les urbanistes réfléchissent déjà sur le terrain . Son plan est géométrique , ses proportions mathématiques .</after>
</modif>
```

### 3.1.2. Motivation du contributeur

L'identification de certaines révisions peut paraître incertaine dans la mesure où seule une connaissance des motivations des contributeurs les ayant effectuées permettrait de trancher (**exemple 1**).

Exemple 1

Si « novelliste » correspond à un emprunt mal interprété du terme anglais et faux-ami « novelist », qui signifie « romancier », la révision serait dans ce cas à classer dans « reformulations lexicales > emprunt » ; si « novelliste » correspond à une variante orthographique de « novéliste », terme français signifiant « auteur de nouvelles », alors la révision est à classer dans « révisions ».

```
<modif id="408576" wp_page_id="1942355" wp_before_rev_id="20486796" wp_after_rev_id="20493923" wp_user_id="19339" wp_user_num_modif="14450" wp_comment="">
<before>Sa fille Nayantara Sahgal , qui vint occuper la maison de sa mère à Dehradun , est une <m num_words="1">novelliste</m> connue .</before>
<after>Sa fille Nayantara Sahgal , qui vint occuper la maison de sa mère à Dehradun , est une <m num_words="1">romanière</m> connue .</after>
</modif>
```

## 3.2. SIMULTANÉITÉ DES PHÉNOMÈNES

L'on peut rencontrer certains problèmes de catégorisation, relativement à la typologie en vigueur et quelle qu'elle soit, lorsqu'on observe des phénomènes de simultanété, de superposition dans les révisions.

## 3.3. DÉTECTION DES PHÉNOMÈNES

Il est difficile de rendre compte de certains phénomènes de réécriture dans la mesure où il sont mal identifiés dans le corpus, ne sont pas inclus dans le balisage des révisions. Cela peut rendre hésitant le choix de la catégorie dans laquelle placer la révision dont il est question (**exemple 1**) ou nuire à son identification (**exemple 2**), notamment dans le cas de « reformulations lexicales » et plus précisément d'« anaphore » (**exemple 3**) comme dans le cas de « reformulations syntaxiques » et plus précisément de « permutation » (**exemples 4 & 5**).

Dans certain cas, il semble très difficile de percevoir la modification (**exemples 6, 7 & 8**). Il s'agit de problèmes relevant de l'algorithme d'alignement et sortant donc du cadre de cette typologie.

### Exemple 1

Si l'on ne prend en compte que le segment compris dans l'élément <m> alors ce qui est interprété comme la suppression d'une partie de ce segment est à catégoriser dans « reformulations sémantiques > neutralité » ; si l'on prend en compte le contexte hors <m> l'insertion de l'item « improprement » modifie l'interprétation de cette modification que l'on catégorisera plutôt dans « reformulations sémantiques > précision ».

```
<modif id="182843" wp_page_id="94940" wp_before_rev_id="18914954" wp_after_rev_id="21221698"
wp_user_id="188656" wp_user_num_modif="219" wp_comment="">
<before> <m num_words="2">Traditionnellement , l'</m> emplacement de Charybde a été associé
au détroit de Messine , au large de la côte de la Sicile , en face d' une roche appelée Scylla . Un tourbillon y
est causé par la confluence de courants , mais il est rarement dangereux . Pour cette raison , une thèse
récente suggère plutôt une origine au voisinage de la Grèce , sur sa côte nord-ouest , près de l' île
Leucade .</before>
<after> <m num_words="1">L'</m> emplacement de Charybde a été improprement associé au détroit
de Messine , au large de la côte de la Sicile , en face d' une roche appelée Scylla . Un tourbillon y est causé
par la confluence de courants , mais il est rarement dangereux . Pour cette raison , une thèse récente
suggère plutôt une origine au voisinage de la Grèce , sur sa côte nord-ouest , près de l' île Leucade , tandis
que le Bosphore semble avoir été sa plus fréquente localisation dans les texte antérieurs à la chute de
Constantinople .</after>
</modif>
```

### Exemple 2

```
<modif id="407103" wp_page_id="1763622" wp_before_rev_id="18617746" wp_after_rev_id="18617816"
wp_user_id="105140" wp_user_num_modif="649" wp_comment="">
<before> Claude François Paparel , <m num_words="1">mort</m> en 1725 , condamné à résider dans l'
île Sainte-Marguerite . Il est le fils de Jean-François Paparel ( 1686 -1761 ) , sr de la Salle , premier commis
des greffes du conseil d' État , lui-même fils de Claude Paparel , receveur du grenier à sel et de Madeleine
Chariot . Les Paparel , famille du Lyonnais portent : D' azur , à 3 tours d' or ; au chef du mesme chargé d' un
lion passant de gueules .</before>
<after> Claude François Paparel , <m num_words="1">né</m> en 1659 et mort le 18 mai 1725 ,
condamné à résider dans l' île Sainte-Marguerite . Il est le fils de Jean-François Paparel ( 1686 -1761 ) , sr de
la Salle , premier commis des greffes du conseil d' État , lui-même fils de Claude Paparel , receveur du
grenier à sel et de Madeleine Chariot . Les Paparel , famille du Lyonnais portent : D' azur , à 3 tours d' or ; au
chef du mesme chargé d' un lion passant de gueules .</after>
</modif>
```

### Exemple 3

```
<modif id="407488" wp_page_id="1798943" wp_before_rev_id="19035269" wp_after_rev_id="19038011"
wp_user_id="35756" wp_user_num_modif="5663" wp_comment="">
<before> La dernière cargaison est exportée en 1942 , en direction de la France métropolitaine . Tout le
monde pense le <m num_words="2">Bourbon pointu</m> éteint à jamais .</before>
<after> La dernière cargaison est exportée en 1942 , en direction de la France métropolitaine . One ne
compte que 200 kilos . Par la suite , le Bourbon pointu disparaît des registres . Tout le monde le
<m num_words="1">pense</m> éteint à jamais .</after>
</modif>
```

#### Exemple 4

```
<modif id="406861" wp_page_id="1697174" wp_before_rev_id="20446491" wp_after_rev_id="20446990"
wp_user_id="285010" wp_user_num_modif="4" wp_comment="Carrière politique">
<before> <m num_words="6">Il est à noter que Desmons</m> ne fut jamais maire de Brignon car il
était opposé au cumul des mandats .</before>
<after> <m num_words="6">Opposé au cumul des mandats , il</m> ne fut jamais maire de
Brignon .</after>
</modif>
```

#### Exemple 5

```
<modif id="351357" wp_page_id="394902" wp_before_rev_id="5400706" wp_after_rev_id="5401289"
wp_user_id="56644" wp_user_num_modif="5112" wp_comment="">
<before>Christine Sèvres <m num_words="4">est une chanteuse française</m> décédée le 1 er
novembre 1981 . Elle était mariée avec l' auteur-compositeur-interprète Jean Ferrat .</before>
<after>Christine Sèvres <m num_words="4">, née en 1931 et</m> décédée le 1 er novembre 1981 ,
est une chanteuse française . Elle était mariée avec l' auteur-compositeur-interprète Jean
Ferrat .</after>
</modif>
```

#### Exemple 6

```
<modif id="182896" wp_page_id="95010" wp_before_rev_id="676327" wp_after_rev_id="784515"
wp_user_id="2322" wp_user_num_modif="1015" wp_comment="ébauche">
<before>Frédéric IV de Hohenzollern , comte de Zollern après Frédéric III , sont sorties plusieurs branches
souveraines . Elles régnèrent sur de petites principautés en Souabe . Une de ces lignées , celle des
Hohenzollern-Sigmaringen , subsiste aujourd ' hui . Au milieu du XIXe siècle , son chef céda volontairement
sa souveraineté sur Sigmaringen à la Prusse . Mais l' un de ses fils , Charles , fut élu en 1866 prince de
Roumanie puis proclamé roi en 1881 . L' actuelle maison royale de Roumanie , qui a régné jusqu ' en 1947 ,
est <m num_words="5">donc un rameau des Hohenzollern-Sigmaringen</m> .</before>
<after>Charles Ier , roi de Roumanie , est <m num_words="6">né 1839 et mort en
1914</m> .</after>
</modif>
```

#### Exemple 7

```
<modif id="182791" wp_page_id="94888" wp_before_rev_id="9677436" wp_after_rev_id="9677751"
wp_user_id="58456" wp_user_num_modif="17114" wp_comment="Genèse et bases théoriques Bossuet">
<before>* Mais les guerres de religion voient s' affronter les nobles du parti protestant au roi catholique .
Les fils de Henri II sont contestés et leur autorité reste fragile . Henri IV renforce la situation et fait cesser la
guerre civile par l' édit de Nantes en 1598 . Selon Pierre Joxe , l' édit de Nantes contribue à renforcer l'
autorité royale . Son effet , plus que de fonder la tolérance envers les protestants , fut d' aboutir à une
véritable religion royale . Henri IV <m num_words="6">fut assassiné par Ravallac en
1610</m> .</before>
<after>* Mais les guerres de religion voient s' affronter les nobles du parti protestant au roi catholique .
Les fils de Henri II sont contestés et leur autorité reste fragile . Henri IV renforce la situation et fait cesser la
guerre civile par l' édit de Nantes en 1598 . Henri IV fut assassiné par Ravallac en 1610 . Selon Pierre Joxe ,
l' édit de Nantes contribue à renforcer l' autorité royale . Son effet , plus que de fonder la tolérance envers
les protestants , fut d' aboutir à une véritable religion royale . A partir de la conversion d' Henri IV <m
num_words="11">, la fidélité au roi l' a emporté sur l' appartenance religieuse</m> .</after>
</modif>
```

#### Exemple 8

```
<modif id="407071" wp_page_id="1750731" wp_before_rev_id="18536691" wp_after_rev_id="18536696"
wp_user_id="168749" wp_user_num_modif="1433" wp_comment="">
<before>La SODECI ( Société <m num_words="2">mère : SAUR-Société</m> d' Aménagement
Urbain et Rural ) , est une société d' économie mixte chargée d' assurer l' approvisionnement
en eau potable de la Côte d' Ivoire . La SODECI offre à la Côte d' Ivoire un service public de l' eau
potable moderne , capable de faire face dans de bonnes conditions à la forte croissance de la demande et
ce , tout en africanisant le plus rapidement possible le management de la société .</before>
<after>La Société <m num_words="2">de distribution</m> d' eau de la Côte d' Ivoire</after>
</modif>
```



**Notes et Documents LIMSI N° : 2011-01**

**Auteurs (Authors) :** Camille Dutrey, Houda Bouamor, Delphine Bernhard et Aurélien Max

**Titre :** Typologie des modifications dans les révisions de Wikipédia

**Title :** A typology of modifications in Wikipedia's revision history.

**Nombre de pages (Number of pages) :** 26

**Résumé :** Ce document détaille une typologie des modifications locales dans les révisions de Wikipédia, développée à partir d'une étude détaillée du corpus WiCoPaCo (Wikipedia Correction and Paraphrase Corpus). L'objectif de cette typologie est de représenter tous les phénomènes observables dans le corpus WiCoPaCo. Nous avons organisé les phénomènes de modifications selon une dichotomie sémantique : nous avons d'une part les modifications induisant une faible variation sémantique entre le segment source et le segment cible, d'autre part les modifications induisant une forte variation sémantique entre le segment source et le segment cible.

**Mots clés :** Wikipédia, révisions, correction orthographique, paraphrase, vandalisme

**Abstract :** This document outlines a typology of local modifications in Wikipedia's revision history, based on a detailed study of the WiCoPaCo corpus (Wikipedia Correction and Paraphrase Corpus). The purpose of this typology is to represent all observable phenomena in the WiCoPaCo corpus. We organized the modification phenomena according to a semantic dichotomy: on the one hand, there are modifications inducing a weak semantic variation between the source segment and the target segment, on the other hand, modifications inducing a strong semantic variation between the source segment and the target segment.

**Key words :** Wikipedia, revisions, spelling correction, paraphrase, vandalism