
**Annotations des marques de la
subjectivité langagière:
discours rapporté, passages entre
guillemets et lexique des affects**

**Manuel de codage
Octobre 2015**

Auteur : Magdalena Augustyn

Coordination et contact : Agnès Tutin (Agnès.Tutin@u-grenoble3.fr), Isabelle Rousset (Isabelle.Rousset@u-grenoble3.fr), Francis Grossmann (Francis.Grossmann@u-grenoble3.fr)

Annotateurs : Gwendoline Bloquet, Mohand Mahrazi, Magdalena Augustyn, Mériam Haddara

Mise aux normes : Magdalena Augustyn, David Graceffa, Liliya Tsvetanova

Collaborateurs : Francis Grossmann, Fanny Rinck, Vannina Goossens, Georges Antoniadis

Le projet a été réalisé dans le cadre du **Projet Plan Pluri-Formation** « *Développement et exploitation de ressources linguistiques pour la didactique du français à l'aide d'outils de TAL. Étude des marqueurs linguistiques de la subjectivité* », piloté par Georges Antoniadis et Francis Grossmann (2003-2007, LIDILEM - Université Stendhal Grenoble 3).

La mise aux normes XML/TEI du corpus a été réalisée dans le cadre d'un financement du **Consortium Corpus Écrits** (IR Corpus) (2013-2014, LIDILEM - Université Stendhal Grenoble 3).

Remerciements : Nous remercions ici chaleureusement Laurent Romary, pour son aide sur les la TEI et la norme MAF, ainsi que Bertrand Gaiffe. Merci aussi à Nicole Dufournaud et à Emmanuelle Morlock pour leurs conseils sur l'annotation du discours rapporté.

Table des matières

1. LE CORPUS : NATURE ET COMPOSITION	4
2. INFORMATIONS GENERALES	5
2.1 LES EN-TETES	5
2.2 ENCODAGE DU CORPS DU TEXTE	5
2.2.1 ÉLÉMENTS MARQUANT LA DIVISION DU TEXTE	6
2.2.2 AUTRES ELEMENTS	6
2.3 ANNOTATION DES MARQUES TYPOGRAPHIQUES	7
2.3.1 LES ITALIQUES	7
3. ANNOTATION DU DISCOURS RAPPORTE	7
3.1 SEGMENTS INTRODUCTEURS DE DISCOURS DIRECT OU INDIRECT	9
3.1.1 IDENTIFICATION DES SEGMENTS INTRODUCTEURS	9
3.1.2 CATEGORISATION DES SEGMENTS INTRODUCTEURS	10
3.2 SEQUENCES DU DISCOURS RAPPORTE (DD, DI, DIL)	11
3.2.1 DISCOURS DIRECT	12
3.2.2 DISCOURS INDIRECT	13
3.2.3 DISCOURS INDIRECT LIBRE	14
3.3 DISCOURS NARRATIVISE ET MODALISATION DU DISCOURS SECOND	14
3.3.1 DISCOURS NARRATIVISE	14
3.3.2 MODALISATION DU DISCOURS SECOND	15
3.4 FORMES MIXTES DU DISCOURS RAPPORTE	15
3.5 AUTRES ELEMENTS D'ANNOTATION	16
3.5.1 LES DIALOGUES	16
3.5.2 LES PASSAGES RICHES (PR)	17
4. ANNOTATION DES PASSAGES ENTRE GUILLEMETS	18
4.1 EMPLOIS AUTONYMIQUES : <Q ANA="AUTONYMIE"> « ... » </Q>	18
4.2 EMPLOIS CITATIONNELS	19
4.2.1 EMPLOIS CITATIONNELS AUTONOMES : <Q ANA="CITA_A"> « ... » </Q>	19
4.2.2 EMPLOIS CITATIONNELS NON AUTONOMES : <Q ANA="CITA_NA"> «...» </Q>	20
4.3 LA FONCTION DE DENOMINATION : <Q ANA="DENOMINATION"> "... " </Q>	20
4.4 LA FONCTION DE MODALISATION : <Q ANA="MODALISATION"> «...» </Q>	21
4.5 LE PROVERBE : <Q ANA="PROVERBE"> « ... » </Q>	22
4.6 <Q ANA="AUTRE"> « ... » </Q>	22
4.7 FONCTIONS MULTIPLES	22
4.7.1 LA MODALISATION DANS LES CAS D'EMPLOIS MULTIPLES	22
4.7.1.1 Dénomination et modalisation	23

4.7.1.2	Modalisation et citation non autonome	23
4.7.2	DENOMINATION ET CITATION NON AUTONOME	23
4.7.3	CITATION ET AUTONYMIE	24
5.	ANNOTATION DU LEXIQUE DES AFFECTS : MOTS SIMPLES ET COLLOCATIONS	25
5.1	ANNOTATION DES MOTS SIMPLES	25
5.1.1	PROPRIETES DU LEXIQUE DES AFFECTS	26
5.2	ANNOTATION DES COLLOCATIONS	32
5.2.1	PROPRIETES DU LEXIQUE DES AFFECTS	34
5.2.2	PROPRIETES DES COLLOCATIFS	34
5.2.2.1	Fonctions lexicales : <f name="FL"><symbol value=" "/>	34
5.2.2.2	Type sémantique / glose sémantique : <f name="FLsemanticType"> <symbol value=" "/>	36
BIBLIOGRAPHIE		37

1. Le corpus : nature et composition

Le corpus annoté (775454 mots) comporte un ensemble de textes variés, principalement des écrits libres de droits puisque les corpus élaborés dans le projet devaient être utilisables librement pour les usagers. Cela restreint le corpus à des œuvres plutôt anciennes. Un des objectifs étant didactique, nous avons sélectionné pour les œuvres littéraires – qui constituent l’essentiel du corpus – des romans largement utilisés dans le cadre scolaire. Le corpus annoté, dont la composition détaillée apparaît dans le Tableau 1, comporte ainsi des ouvrages classiques au programme des collèges comme *Les lettres de mon moulin* d’Alphonse Daudet, ou *La petite Fadette* de George Sand. De manière générale, l’annotation suit les recommandations de la Text Encoding Initiative (P5). Pour l’annotation du lexique des émotions, la norme MAF (Morpho-syntactic annotation framework ; ISO/FDIS 24611) dans sa sérialisation TEI, a été adoptée.

Texte	Type	Nombre de mots	Annotation du lexique des affects et des collocations	Annotation du DR et des PEG
<i>La petite Fadette</i> (George Sand)	Littéraire	74456	sand_petitefadette_emo.xml	-
<i>Les Lettres de mon moulin</i> (Alphonse Daudet)	Littéraire	46950	daudet_moulin_emo.xml	daudet_moulin_dr_peg.xml
<i>Colomba</i> (Prosper Mérimée)	Littéraire	52619	merimee_colomba_emo.xml	-
<i>Le petit chose</i> (Alphonse Daudet)	Littéraire	83561	daudet_petitchose_emo.xml	-
<i>Le mystère de la chambre jaune</i> (Gaston Leroux)	Littéraire	86 271	leroux_chambrejaune_emo.xml	-
<i>Contes</i> (Perrault)	Littéraire	21684	perrault_contes_emo.xml	-
<i>Les contes du lundi</i> (Alphonse Daudet)	Littéraire	68088	daudet_lundi_emo.xml	daudet_lundi_dr_peg.xml
<i>L’île mystérieuse</i> (Jules Verne)	Littéraire	199426	verne_ilemysterieuse_emo.xml	verne_ilemysterieuse_dr_peg.xml
<i>Le droit à la paresse</i> (Jules Lafargue)	Essai polémique	12269	lafargue_droitlaparesse_emo.xml	lafargue_droitlaparesse_dr_peg.xml
2 articles de la revue LIDIL	Écrits scientifiques	10074	benarabah_emo.xml trimaille_emo.xml	benarabah_dr_peg.xml trimaille_dr_peg.xml
<i>La place de la LSF dans l’intégration scolaire des enfants sourds</i> (Agnès Millet)	Écrits scientifiques	13642	millet_emo.xml	millet_dr_peg.xml
<i>Une vie</i> (Guy Maupassant)	Littéraire	75414	-	maupassant_vie_dr_peg.xml
TOTAL		744454		

Tableau 1 : La composition du corpus

2. Informations générales

2.1 Les en-têtes

Les en-têtes TEI des documents indiquent la provenance du document, le type d'annotation et la disponibilité du document. Tous les textes annotés sont disponibles selon la licence CC suivante :

- Disponible gratuitement sous licence Creative Commons :
<http://creativecommons.org/licenses/by-nc-sa/3.0/fr/>
- Cette licence implique la libre reproduction et diffusion de l'œuvre au public, notamment par téléchargement, mais est soumise aux conditions suivantes :
 - Attribution — Vous devez créditer l'Oeuvre, intégrer un lien vers la licence et indiquer si des modifications ont été effectuées à l'Oeuvre. Vous devez indiquer ces informations par tous les moyens possibles mais vous ne pouvez pas suggérer que l'Offrant vous soutient ou soutient la façon dont vous avez utilisé son Oeuvre.
 - Pas d'Utilisation Commerciale — Vous n'êtes pas autorisé à faire un usage commercial de cette Oeuvre, tout ou partie du matériel la composant.
 - Partage dans les Mêmes Conditions — Dans le cas où vous effectuez un remix, que vous transformez, ou créez à partir du matériel composant l'Oeuvre originale, vous devez diffuser l'Oeuvre modifiée dans les mêmes conditions, c'est à dire avec la même licence avec laquelle l'Oeuvre originale a été diffusée.

2.2 Encodage du corps du texte

Les documents comportent les éléments suivants :

- l'élément `<text>` regroupe à la fois le corps du texte et tous les éléments externes ;
- l'élément `<front>` (texte liminaire) contient toutes les informations qui apparaissent avant le corps du texte : page de titre (titre, auteur, maison d'édition, date, etc.), dédicaces, préfaces, sommaires¹, nom et affiliation de l'auteur (dans les articles scientifiques), etc. ;
- l'élément `<body>` (corps du texte) marque le début et la fin du corps du texte proprement dit ;
- l'élément `<back>` (texte annexe) contient les informations supplémentaires : appendice, références des notes de bas de page, etc.

¹ Les titres du sommaire sont considérés comme des items d'une liste et sont notés de la façon suivante :

```
<div type="sommaire">
  <head>Sommaire</head>
  <list>
    <item>Préface (dans le livre papier mais pas ici)</item>
    <item>Avant propos (idem)</item>
    <item>...</item>
  </list>
</div>
```

2.2.1 Éléments marquant la division du texte

Chaque document est composé d'éléments de structure balisés.

L'élément `<div>` marque les divisions du texte et contient au moins un élément d'un niveau de structuration inférieur (un autre élément `<div>` de niveau inférieur ou un paragraphe `<p>`). Les éléments `<div>` sont numérotés selon le niveau de structuration : `<div1>` (division du premier niveau), `<div2>` (division du niveau inférieur correspondant aux sous-parties), `<div3>`, etc.

L'élément `<div>` peut contenir des attributs :

- L'attribut `@type` de l'élément `<div>` indique le type de la subdivision (conte, nouvelle, chapitre, partie, section..., nommées selon le type du texte) ;
- L'attribut `@n` indique le numéro de la subdivision.

Les éléments `<div>` peuvent être suivis par un élément `<head>` contenant le titre de la division.

Exemple : Structuration du texte *Les Lettres de mon moulin* (Alphonse Daudet)

```
<div1 type="nouvelle" n="13">
  <head>Ballades en prose</head>
  <p>En ouvrant ma porte ce matin, il y avait autour de mon
moulin un grand tapis de gelée blanche. L'herbe luisait et craquait
comme du verre; toute la colline grelottait... Pour un jour ma chère
Provence s'était déguisée en pays du Nord; et c'est parmi les pins
frangés de givre, les touffes de lavandes épanouies en bouquets de
cristal, que j'ai écrit ces deux ballades d'une fantaisie un peu
germanique, pendant que la gelée m'envoyait ses étincelles blanches
et que là-haut, dans le ciel clair, de grands triangles de cigognes
venues du pays d'Henri Heine, descendaient vers la Camargue en criant
: "Il fait froid... froid... froid." </p>

  <div2 type="section" n="1">
  <head>I</head>
  <head>La mort du dauphin</head>
  <p>...</p>
```

L'élément `<p>` marque les paragraphes. Dans l'annotation du lexique des affects, les paragraphes sont numérotés, l'élément `<p>` contient l'attribut `@xml:id` avec le numéro du paragraphe :

```
<p xml:id="15">Or, au moment même où tant de catastrophes
s'accomplissaient sur terre et sur mer, un drame, non moins
saisissant, se jouait dans les airs bouleversés. </p>
```

2.2.2 Autres éléments

Certains éléments annotés sont propres au type ou à la structure de certains textes :

- **Les notes de bas de pages** : les appels de notes dans le corps du texte sont mis entre balises, par exemple `<ref target="" #_ftn1'' />[1]</ref>`, les références apparaissent dans `<back>` de la manière suivante :

```
<div type="note">
  <note>
    <anchor xml:id="_ftn1"/>
    <ref>1.</ref>Descartes, Les Passions de l'âme.</note>
  </div>
```

- **Les tableaux** : les tableaux sont supprimés et seule leur présence est mentionnée de la manière suivante :

```
<table><head>Tableau 2: Résultats (%) pour la variable(k)</head><row><cell>
</cell></row></table>
```

Les tableaux contiennent au moins une balise `<row>` et une balise `<cell>` qui sont vides, mais qui sont nécessaires pour la bonne formation du document.

- **Les dédicaces dans les nouvelles** (*Les Lettres de mon moulin*, Daudet : les dédicaces sont mises entre balises `<div type="dedication">...</div>`).

2.3 Annotation des marques typographiques

2.3.1 Les italiques

L'élément `<hi>` est employé pour les différents types de mises en évidence typographique, il nous permet d'annoter des séquences du discours rapporté qui sont marquées par les italiques (et non les guillemets).

Par exemple :

- (1) [...] et quand dom Balaguère, la figure rayonnante, se tourne vers l'assistance `<seg type="intro_DD" ana="V">` en criant `</seg>` de toutes ses forces `<quote type="DD" ana="CA_AVANT">`: `<hi rend="italic">Ite missa est</hi>`, `</quote>` il n'y a qu'une voix dans la chapelle pour lui `<seg type="intro_DD" ana="V">`répondre`</seg>` un `<quote type="DD" ana="CA_AVANT">` `<hi rend="italic">Deo gratias</hi>``</quote>` si joyeux, si entraînant, qu'on se croirait déjà à table au premier toast du réveillon. (daudet_moulin_dr_peg.xml)
- (2) Nos bergers provençaux `<seg type="intro_DD" ana="V">`appellent`</seg>` cette manœuvre: `<quote type="DD" ana="CA_AVANT">` `<hi rend="italic">vira la bano au gisclé</hi>` `</quote>`- tourner la corne au vent. (daudet_moulin_dr_peg.xml)

3. Annotation du discours rapporté

Le premier volet du projet concerne l'annotation du discours rapporté avec comme objectif l'identification des éléments constitutifs à des fins de description linguistique / pédagogiques. Le choix des phénomènes du discours rapporté a été principalement déterminé par la perspective d'une exploitation didactique et donc au besoin de recourir à des catégories stables définies à partir d'indices formels.

Ainsi, les trois formes canoniques du discours rapporté sont annotées : le discours direct, le discours indirect et le discours indirect libre. Leur définition (la distinction selon le rapport entre le discours citant et le discours cité, et en fonction du marquage du discours cité) est fondée sur un modèle consensuel (par exemple Rosier, 1999). La mise à l'épreuve de ce classement dans le corpus a nécessité une prise en compte de formes mixtes. D'autres phénomènes qui sont aux confins du discours rapporté sont également pris en compte : modalisations en discours second et discours narrativisé.

Les étapes et les informations associées à l'annotation (au format XML/TEI P5²) sont les suivantes :

- Repérage et description des phénomènes du discours rapporté (DR) :
 - Délimitation et catégorisation de différentes formes du DR : discours direct (DD), discours indirect (DI), discours indirect libre (DIL), discours narrativisé (DN), modalisation du discours second (MDS) et repérage des segments des dialogues
 - Pour le discours direct (uniquement) : la catégorisation des types d'introducteurs : syntagme préposé, postposé, double segment introducteur, incise, sans syntagme introducteur
 - Délimitation et catégorisation des segments introducteurs (pour le discours direct et indirect) : verbe simple, syntagme verbal, syntagme nominal ...

Discours rapporté	Segments introducteurs
<p>- Les phénomènes de discours rapporté :</p> <pre><quote type="DD" ana="CA_AVANT" "CA_APRES" "INCIS E" "SANS"> ... </quote></pre> <pre><quote type="DI" "DIL" ... </quote></pre> <p>- Le discours narrativisé et modalisation du discours second :</p> <pre><seg type="DN" "MDS"></pre>	<pre><seg type="intro_DD" "intro_DI" "CA" ana="V" "N" "SV" "SN" "SP" "SA" "P"> ... </seg></pre>

² <http://www.tei-c.org/Guidelines/P5/>

<pre>... </seg></pre> <p>- Les dialogues :</p> <pre><seg type="dialogue"> ... </seg></pre> <p>- Les passages riches :</p> <pre><seg type="PR"> ... </seg></pre>	
---	--

Tableau 2 : Annotation XML/TEI du discours rapporté

3.1 Segments introducteurs de discours direct ou indirect

Dans le cas du discours direct et indirect, le **segment introducteur** est balisé avec l'élément **<seg>** et les attributs suivants :

- l'attribut désignant le type d'introducteur : **@type="intro_DD" | "intro_DI"**
- l'attribut apportant une information grammaticale : **@ana="V" | "N" | "SV" | "SN"...**

3.1.1 Identification des segments introducteurs

Dans l'identification des segments introducteurs, l'intérêt s'est porté davantage sur l'annotation des formes fondamentales. Ainsi, ont été annotés les introducteurs déclaratifs (renvoyant explicitement à la parole) de type : *dire, crier, balbutier, voix, réponse, conseil, paroles*, à l'exclusion des verbes introducteurs non déclaratifs (ou les « verbes gestes », Rosier, 2008), par exemple *respirer* dans l'exemple suivant :

(3) Alors seulement elle respire : <quote type="DD" ana="SANS">"Il a été bien poli, cet officier."</quote>
(daudet_lundi_dr_peg.xml)

Une limite intra-phrastique a été imposée dans l'identification des segments introducteurs. De cette manière, les segments potentiellement introducteurs mais se trouvant dans une phrase antérieure ou postérieure n'ont pas été pris en compte. Par exemple :

(4) Les Briseville insistèrent. <quote type="DD" ana="SANS">"Comment ? si vite ? Restez donc encore un peu."</quote> (maupassant_vie_dr_peg.xml)

En ce qui concerne les limites grammaticales des segments introducteurs, il est à noter que, dans le cas des verbes, les verbes modaux ne sont pas compris dans les balises. Par exemple :

- (5) [...] c'est bien Mistral dont Montaigne aurait pu `<seg type="intro_DD" ana="V">dire</seg>`: `<quote type="DD" ana="CA_AVANT">Souvienne-vous de celui à qui ... </quote>`.
(daudet_moulin_dr_peg.xml)

3.1.2 Catégorisation des segments introducteurs

Les segments introducteurs ont été catégorisés selon le type du discours rapporté qu'ils introduisent (l'attribut `@type="intro_DD"|"intro_DI"`) et selon la catégorie grammaticale du segment (l'attribut `@ana="V"|"N"|"SV"|"SN"...`).

Type de segment	Élément et attributs	Exemple
Verbe	<code><seg type="intro_DD" ana="V">...</seg></code>	Pourquoi donc, Monsieur Spilett ? <code><seg type="intro_DD" ana="V">Répondit</seg></code> Harbert.
	<code><seg type="intro_DI" ana="V">...</seg></code>	Le Camarguais <code><seg type="intro_DI" ana="V">racontait</seg></code> qu'il venait de Nîmes [...]
Syntagme verbal	<code><seg type="intro_DD" ana="SV">...</seg></code>	[...] il lui <code><seg type="intro_DD" ana="SV">vint d'un air goguenard</seg></code> : - Et ta femme, à toi, rémouleur ? Pour quelle paroisse tient-elle?
	<code><seg type="intro_DI" ana="SV">...</seg></code>	Cependant, le cours d'eau s'élargissait sensiblement, et Cyrus Smith <code><seg type="intro_DI" ana="SV">était porté à croire</seg></code> qu'il aurait bientôt atteint son embouchure.
Syntagme nominal	<code><seg type="intro_DD" ana="SN">...</seg></code>	Puis, tout à coup, vers le soir, <code><seg type="intro_DD" ana="SN">un grand cri</seg></code> : " Les voilà! " [...]
Syntagme adverbial	<code><seg type="intro_DD" ana="SA">...</seg></code>	C'était si bien imité que je ne pus m'empêcher de rire. Mais lui, <code><seg type="intro_DD" ana="SA">très froidement</seg></code> : - Vous croyez que je plaisante... regardez mes yeux.
Syntagme prépositionnel	<code><seg type="intro_DD" ana="SP">...</seg></code>	puis, <code><seg type="intro_DD" ana="SP">d'une voix stridente</seg></code> , comme pour haranguer un banquet de deux cents couverts:[...]

Proposition	<code><seg type="intro_DD" ana="P">...</seg></code>	<code><seg type="intro_DD" ana="P">Tout le mémoire était écrit dans le français bizarre que parlait M. de Sieboldt</seg> : " Si j'aurais des actionnaires..., si je réunirais des fonds... "</code>
-------------	---	---

Tableau 3 : Annotation des segments introducteurs

3.2 Séquences du discours rapporté (DD, DI, DIL)

L'élément `<quote>` couvre les différents types du discours rapporté : discours direct, discours indirect et discours indirect libre, qu'il y ait ou non passage entre guillemets. Selon la TEI P5 : « quote (citation) contient une expression ou un passage que le narrateur ou l'auteur attribue à une origine extérieure au texte. »

Cet élément contient les attributs suivants :

- L'attribut `@type` indiquant le type du discours rapporté :
`@type="DD"|"DI"|"DIL"`
- L'attribut `@ana` indiquant la place du segment introducteur par rapport à la séquence du discours rapporté (uniquement dans le cas du discours direct)
`@ana="CA_AVANT" (préposé)|"CA_APRES"(postposé)|"INCISE"|"SANS"`.

REMARQUE 1

L'élément `<quote>` englobe le tour de parole même dans le cas d'incises. Il est alors nécessaire de séparer le discours citant à l'intérieur des **séquences discontinues du discours rapporté**. Afin de permettre de ne pas tenir compte, lors d'une recherche, du discours citant qui ne se situe pas sur le même plan énonciatif, l'élément `<seg>` avec comme attribut `@type="exquote"` est utilisé pour marquer le discours citant.

(6) `<quote type="DD" ana="INCISE">-Oui, Nab, <seg type="exquote"><seg type="intro_DD" ana="V">répondit</seg> Gédéon Spilett,</seg> et nous le retrouverons ! </quote>`
(verne_ilemysterieuse_dr_peg.xml)

Le même procédé est utilisé dans le cas des séquences de discours direct annotées par `<quote type="AUTRE">` avec deux segments introducteurs, dont le deuxième le plus souvent en incise :

(7) Ici le patron `<seg type="intro_DD" ana="V">s'interrompt</seg>` :
`<quote type="DD" ana="AUTRE">- Attention, Nardi ! <seg type="exquote"><seg type="intro_DD" ana="V">cria</seg>-t-il</seg>, le feu s'éteint. </quote>`
(daudet_moulin_dr_peg.xml)

REMARQUE 2

Dans le cas des **verbes discontinus** (inversion, insertion d'un adverbe) la balise `<seg type="CA"> ... </seg>` est utilisée, par exemple :

- (8) `<quote type="DD" ana="INCISE">"Précisément, <seg type="exquote"><seg type="intro_DD" ana="V">fit <seg type="CA">-il</seg>observer</seg></seg>, ces gentlemen ont pris terre sur la côte méridionale, et, s'ils ont suivi le littoral, il est à craindre qu'ils n'aient découvert le petit port, auquel cas je ne donnerais pas un demi-dollar de notre Bonaventure. "</quote>`
(verne_ilemysterieuse_dr_peg.xml)
- (9) `Il <seg type="intro_DI" ana="SV">faisait, <seg type="CA">d'ailleurs, </seg>observer</seg><quote type="DI">qu'Harbert et Gédéon Spilett étaient devenus des archers habiles [...]. </quote>`
(verne_ilemysterieuse_dr_peg.xml)

3.2.1 Discours direct

Le discours cité est accompagné d'indices typographiques spécifiques (guillemets, italique), il a une autonomie énonciative et syntaxique par rapport au discours citant. Il sert à restituer les paroles citées de manière « fidèle ».

La présence (ou non) et la position du segment introducteur est signalé par l'attribut @ana :

`<quote type="DD" ana="CA_AVANT">... </quote>` : balise du discours direct avec le discours citant apparaissant avant.

- (10) `Harbert le fit observer à Pencroff, <seg type="intro_DD" ana="V">en ajoutant</seg> : <quote type="DD" ana="CA_AVANT">"Cela prouverait qu'il existe dans l'ouest une côte assez rapprochée."</quote>`
(verne_ilemysterieuse_dr_peg.xml)

`<quote type="DD" ana="CA_APRES">... </quote>` : balise du discours direct avec le discours citant apparaissant après.

- (11) `<quote type="DD" ana="CA_APRES">- Comme une carpe ? </quote> <seg type="intro_DD" ana="V">s'écria</seg> Harbert, très surpris de la proposition.`
(verne_ilemysterieuse_dr_peg.xml)

`<quote type="DD" ana="INCISE">...</quote>` : balise du discours direct avec un segment introducteur en **incise** (lorsque le discours citant ne termine pas le passage cité).

- (12) `<quote type="DD" ana="INCISE">"C'était une frégate anglaise, monsieur, <seg type="exquote"> <seg type="intro_DD" ana="V">s'écria</seg> le capitaine Nemo, redevenu un instant le prince Dakkar,</seg> une frégate anglaise, vous entendez bien ! Elle m'attaquait ! J'étais resserré dans une baie étroite et peu profonde ! ... il me fallait passer, et... j'ai passé !"</quote>`
(verne_ilemysterieuse_dr_peg.xml)
- (13) `<quote type="DD" ana="INCISE"> -Il est même fâcheux, <seg type="exquote"><seg type="intro_DD" ana="V"> fit observer</seg>le marin</seg>, que nous ayons oublié de prendre cette précaution lors de notre premier voyage à l'île Tabor. </quote>` (verne_ilemysterieuse_dr_peg.xml)

`<quote type="DD" ana="AUTRE">... </quote>` : l'élément ayant pour attribut `@ana="AUTRE"` est utilisé lorsqu'il y a deux segments introducteurs.

- (14) Ici le patron `<seg type="intro_DD" ana="V">s'interrompt</seg>` : `<quote type="DD" ana="AUTRE">- Attention, Nardi ! <seg type="exquote"> <seg type="intro_DD" ana="V">cria</seg>-t-il</seg>, le feu s'éteint.</quote>` (daudet_moulin_dr_peg.xml)

`<quote type="DD" ana="SANS">... </quote>` : l'élément ayant pour attribut `@ana="SANS"` est utilisé lorsqu'il n'y a pas de segment introducteur du discours rapporté.

- (15) Un aide de camp couvert de boue force la consigne, franchit le perron d'un saut : `<quote type="DD" ana="SANS">Maréchal! Maréchal !...</quote>`
(daudet_lundi_dr_peg.xml)

3.2.2 Discours indirect

Le discours cité est reformulé d'une manière plus ou moins fidèle et il est présenté généralement sous la forme d'une complétive (16), d'une infinitive (17) ou d'un groupe nominal (18).

Le segment introducteur `<seg type="intro_DI" ana="...">` se trouvant ici toujours avant la séquence du discours rapporté, l'élément `<quote>` ne nécessite pas la présence de l'attribut `@ana` précisant sa position comme c'est le cas dans le discours direct.

- (16) Le marin s'occupait de ce travail, quand Harbert lui `<seg type="intro_DI" ana="V">demanda</seg>` `<quote type="DI">s'il avait des allumettes.</quote>`
(verne_ilemysterieuse_dr_peg.xml)
- (17) Tout d'abord, le reporter `<seg type="intro_DI" ana="V">dit</seg>` au marin `<quote type="DI">de l'attendre`

en cet endroit même, où il le rejoindrait</quote> [...] ((verne_ilemysterieuse_dr_peg.xml)

- (18) Un médecin consulté dix ans auparavant, parce qu'elle éprouvait des étouffements, <seg type="intro_DI" ana="V"> avait parlé</seg><quote type="DI"> d'hypertrophie</quote>. (maupassant_vie_dr_peg.xml)

3.2.3 Discours indirect libre

L'annotation du discours indirect libre permet de signaler les passages sans le marquage du discours cité mais avec, le plus souvent, une modalité décalée qui peut porter sur la forme de la phrase (par exemple le passage à l'interrogative), le changement de temps grammatical ou autres phénomènes.

- (19) Alors l'inquiétude des colons fut grande. <quote type="DIL">Que s'était-il passé ? Ayrton n'était-il donc plus au corral, ou, s'il s'y trouvait encore, n'avait-il plus la liberté de ses mouvements ? Devait-on aller au corral par cette nuit obscure ?</quote> (verne_ilemysterieuse_dr_peg.xml)
- (20) Une joie délirante, un attendrissement infini devant la splendeur des choses noya son cœur qui défaillait. <quote type="DIL"> C'était son soleil ! son aurore ! le commencement de sa vie ! le lever de ses espérances ! </quote> (maupassant_vie_dr_peg.xml)
- (21) Ma femme pourrait bien, mais <quote type="DIL">elle ne veut pas: elle prétend qu'on trouve dans les faits divers des choses qui ne sont pas convenables. </quote> (daudet_moulin_dr_peg.xml)

3.3 Discours narrativisé et modalisation du discours second

D'autres phénomènes qui sont aux confins du discours rapporté sont également pris en compte. Il s'agit des phénomènes qui dépassent le cadre strict du discours rapporté, comme les *modalisations du discours second* et le *discours narrativisé*. Comme il ne s'agit pas des formes canoniques du discours rapporté (ce qui est le cas du discours direct, indirect et indirect libre), ce type de séquences a été annoté non pas par l'élément <quote> mais par un élément <seg> avec l'attribut @type spécifiant s'il s'agit du discours narrativisé (DN) ou de la modalisation du discours second (MDS).

3.3.1 Discours narrativisé

Le discours narrativisé permet de mentionner qu'un discours a eu lieu et consiste à représenter le contenu des propos de manière condensée et de les intégrer entièrement

dans le récit (le lecteur n'a pas accès directement aux propos qui ont été prononcés, mais il prend connaissance de leur contenu de manière condensée, résumée).

- (22) **<seg type="DN">**Et il leur fit connaître le résultat de ses observations.**</seg>** (verne_ilemysterieuse_dr_peg.xml)
- (23) Alors, d'une chose à l'autre, **<seg type="DN">**M. Hamel se mit à nous parler de la langue française, disant que c'était la plus belle langue du monde, la plus claire, la plus solide ; qu'il fallait la garder entre nous et ne jamais l'oublier, parce que, quand un peuple tombe esclave, tant qu'il tient bien sa langue, c'est comme s'il tenait la clef de sa prison... **</seg>** (daudet_lundi_dr_peg.xml)
- (24) **<seg type="DN">**Ils parlèrent d'eux, de leurs habitudes, de leurs goûts, sur ce ton plus bas, intime, dont on fait les confidences.**</seg>** **<seg type="DN">**Il se disait déjà dégoûté du monde, las de sa vie futile ;**</seg>** c'était toujours la même chose ; on n'y rencontrait rien de vrai, rien de sincère. (maupassant_vie_dr_peg.xml)

3.3.2 Modalisation du discours second

Il s'agit d'une modalisation que le locuteur opère sur le discours d'un autre dans son propre énoncé. Le discours d'un autre est rendu visible à l'aide des marques explicites, telles que *selon X, d'après X, paraît-il*, etc.

- (25) **<seg type="MDS">** **<seg type="MDS_CA">**D'après son observation,**</seg>** il n'était pas douteux, pour lui, que la balle, entrée par devant, ne fût sortie par derrière.**</seg>** (verne_ilemysterieuse_dr_peg.xml)
- (26) **<seg type="MDS">****<seg type="MDS_CA">**Il paraît**</seg>**qu'au bas de la côte elle avait trouvé la Sorgue grossie par la pluie d'orage, et qu'en voulant passer à toute force, elle avait risqué de se noyer**</seg>**. (daudet_moulin_dr_peg.xml)

L'élément **<seg type="MDS_CA">** ...**</seg>** permet d'isoler une séquence qui introduit le discours d'un autre, comme *il paraît* dans l'exemple ci-dessous.

3.4 Formes mixtes du discours rapporté

Dans le corpus, certains exemples ne relèvent pas d'un type du discours rapporté. En particulier, leurs caractéristiques dépassent les définitions classiques des différents types du discours rapporté et constituent ce que l'on appelle les « formes mixtes » (Cf. Rosier 1999), par exemple :

- Une séquence du discours direct (« îlot citationnel ») faisant partie d'une complétive ou d'une infinitive introduite par un segment introducteur (cette forme mixte entre DI et DD est annotée en tant que DD avec, simultanément, la fonction « citation non autonome » des guillemets : <q ana="cita_na">):

(27) Comme on lui <seg type="intro_DD" ana="V">avait recommandé</seg> de <quote type="DD" ana="CA_AVANT"><q ana="cita_na">"prendre du mouvement"</q></quote>, elle s'acharnait à marcher. (maupassant_vie_dr_peg.xml)

- Discours narrativisé (DN) avec des passages entre guillemets (fonction « citation non autonome» des guillemets : <q ana="cita_na">):

(28) <seg type="DN">et elle parlait de <q ana="cita_na">"son"</q> hypertrophie à tout propos</seg> et si souvent qu'il semblait que cette affection lui fût spéciale, lui appartenant comme une chose unique sur laquelle les autres n'avaient aucun droit. (maupassant_vie_dr_peg.xml)

- Discours indirect libre (DIL) contenant une séquence du discours direct avec un segment introducteur (le passage entre guillemets du discours direct est annoté simultanément en tant que « citation autonome» : <q ana="cita_a">) est annoté de la manière suivante :

(29) <quote type="DIL">Non ! Ce n'était pas possible ! Le télégramme <seg type="intro_DD" ana="V">avait<seg type="CA">bien</seg> dit</seg></quote> : <quote type="DD" ana="CA_AVANT"><q ana="cita_a">"venez au corral en toute hâte." </q> </quote> (verne_ilemysterieuse_dr_peg.xml)

- Discours indirect libre avec un segment introducteur (cette forme mixte entre le discours indirect libre et le discours indirect est annotée uniquement en tant que discours indirect libre) :

(30) <quote type="DIL">Que faire ?</quote> Se demandait le pauvre Nab. (verne_ilemysterieuse_dr_peg.xml)

3.5 Autres éléments d'annotation

3.5.1 Les dialogues

L'élément <seg type="dialogue"> est utilisé pour baliser les cas où l'on a au moins deux tours de parole. Il contient systématiquement les balises du discours rapporté (<quote>) et des passages entre guillemets (<q>, cf. plus loin).

(31) Tout bouffant de colère et rouge comme un coq, le maréchal paraît à la fenêtre, sa queue de billard à la main :

<seg type="dialogue"><quote type="DD" ana="SANS">“ Qu'est-ce qu'il y a ?... Qu'est-ce que c'est ?... Il n'y a donc pas de factionnaire par ici ?
 - Mais, maréchal...
 - C'est bon... Tout à l'heure... qu'on attende mes ordres, nom d... D... ! ”**</quote></seg>** Et la fenêtre se referme avec violence. (daudet_lundi_dr_peg.xml)

(32) **<seg type="dialogue">**“**<quote type="DD" ana="INCISE">**Bon !
<seg type="exquote"> **<seg type="intro_DD" ana="V">**S'écria
</seg>Pencroff**</seg>**, ce volcan qui voudrait faire des siennes ! Mais qu'il essaye ! Il trouvera son maître !
 ...**</quote>**
 -**<quote type="DD" ana="CA_APRES">**Qui donc ?**</quote>** **<seg type="intro_DD" ana="V">**Demanda**</seg>** Nab.
 -**<quote type="DD" ana="SANS">**Notre génie, Nab, notre génie, qui lui bâillonnera son cratère, s'il fait seulement mine de l'ouvrir ! **</quote>**”**</seg>**
 (verne_ilemysterieuse_dr_peg.xml)

3.5.2 Les passages riches (PR)

L'élément **<seg type="PR">...</seg>** est utilisé pour marquer les passages riches, c'est-à-dire des passages regroupant des phénomènes diversifiés du discours rapporté. La taille du passage riche ne devra pas excéder une page.

(33) **<seg type="PR">****<quote type="DD" ana="INCISE">**- Pauvre de moi!
<seg type="exquote">**<seg type="intro_DD" ana="V">**disait**</seg>**-il.**</seg>** Maintenant, je n'ai plus qu'à mourir... Le moulin est déshonoré. **</quote>**
 Et il sanglotait à fendre l'âme, appelant son moulin par toutes sortes de noms, lui parlant comme à une personne véritable.
 À ce moment les ânes arrivent sur la plate-forme, et nous
<seg type="intro_DD" ana="V">nous mettons **<seg type="CA">**tous **</seg>** à crier **</seg>**bien fort comme au beau temps des meuniers :
<quote type="DD" ana="CA_AVANT">- Ohé! du moulin Loo Ohé! maître Cornille !**</quote>** Et voilà les sacs qui s'entassent devant la porte et le beau grain roux qui se répand par terre, de tous côtés...
 Maître Cornille ouvrait de grands yeux. Il avait pris du blé dans le creux de sa vieille main et il **<seg type="intro_DD" ana="V">**disait**</seg>**, riant et pleurant à la fois :
<quote type="DD" ana="CA_AVANT">- C'est du blé!... Seigneur Dieu!... Du bon blé !... Laissez-moi que je le regarde. **</quote>**
 Puis, se tournant vers nous :
<quote type="DD" ana="SANS">- Ah ! je savais bien que vous me reviendriez... tous ces minotiers sont des voleurs. **</quote>**

Nous voulions l'emporter en triomphe au village :
 <quote type="DD" ana="SANS">- Non, non, mes enfants; il
 faut avant tout que j'aie donné à manger à mon
 moulin... Pensez donc! il y a si longtemps qu'il ne s'est
 rien mis sous la dent! </quote></seg>
 (daudet_moulin_dr_peg.xml)

4. Annotation des passages entre guillemets

L'élément **<q>** est employé pour tous les types de passages entre guillemets. L'attribut **@ana="autonymie"|"dénomination" |...** indique la fonction/le type du passage entre guillemets :

Les principales fonctions dégagées pour les passages entre guillemets sont les suivantes :

- la fonction purement autonymique
- la fonction citationnelle (citation autonome et non autonome)
- la fonction de dénomination
- la fonction de modalisation
- le proverbe

Plusieurs valeurs sont utilisées pour le même attribut **@ana** lorsqu'on peut associer aux passages entre guillemets plusieurs fonctions simultanément (le plus souvent la fonction est double). Dans ces cas, les valeurs apparaissent séparées par un espace (pour le détail, voir plus loin : 4.6 « Fonctions multiples »). Des explications sur les annotations effectuées sont données dans (Rinck & Tutin 2007) et (Rinck *et al.* 2009).

4.1 Emplois autonymiques : **<q ana="autonymie"> « ... » </q>**

La fonction est purement autonymique lorsque les guillemets signalent une rupture quant au statut du signe : les guillemets ne servent qu'à introduire le signe linguistique pour lui-même, le signe perdant ici sa fonction référentielle.

(34) Ce **<q ana="autonymie">"nous"</q>** engageait le reporter, mais le reporter n'était pas homme à reculer, et quand le projet lui fut communiqué, il l'approuva sans réserve. (verne_ilemysterieuse_dr_peg.xml)

La fonction autonymique est aussi souvent introduite par l'italique et non les guillemets, cette marque typographique peut cependant apparaître également entre guillemets, par exemple « *analyse* peut être considéré comme un nom concret ».

(35) Des termes comme **<q ana="autonymie"><hi rend="italic">mic</hi>"</q>** [microphone], **<q ana="autonymie"><hi rend="italic">freestyle</hi>"</q>** [prestation réalisée en direct], **<q ana="autonymie"><hi rend="italic">Dj</hi>"</q>** [**<hi rend="italic">Disc**

Jockey</hi>], qui constituent la base d'une sorte de technolecte du rap correspondent bien à l'emprunt du mot avec "l'introduction de la chose". (trimaille_dr_peg.xml)

Les emplois purement autonymiques sont parfois introduits par des termes métalinguistiques, comme « le mot "impossible" » dans l'exemple suivant :

(36) Décidément, Pencroff avait rayé le mot **<q ana="autonymie">"impossible"</q>** du dictionnaire de l'île Lincoln. (verne_ilemysterieuse_dr_peg.xml)

4.2 Emplois citationnels

On considère qu'il y a emploi citationnel lorsque les guillemets signalent que le segment qu'ils entourent est à comprendre comme un énoncé produit par un autre locuteur (individuel ou collectif) que celui du discours citant. Le locuteur à la source de cet énoncé doit donc être explicitement mis en scène, i.e. il doit être clairement identifiable par des indices linguistiques (nom propre, syntagme nominal), même si ce n'est pas dans la même phrase.

En se basant sur des critères énonciatifs et syntaxiques, on distinguera la *citation autonome* et la *citation non autonome*.

4.2.1 Emplois citationnels autonomes : **<q ana="cita_a"> « ... » </q>**

Les emplois citationnels sont autonomes s'il y a :

- une *autonomie énonciative* du passage entre guillemets (on est dans le « je-ici-maintenant » de l'énoncé cité ; par exemple « je » renvoie au locuteur du discours cité, et non à celui du discours citant) ;
- et/ou une *autonomie syntaxique* du passage entre guillemets par rapport au discours citant : généralement une phrase, mais aussi une phrase averbale ou une réponse elliptique. Ce type est souvent marqué par l'usage de : « ... », mais ce n'est pas toujours le cas, car le verbe introducteur peut être en incise ou après le passage entre guillemets, et il peut ne pas y en avoir.

(37) Alors l'homme proposa ses services pour des promenades en mer, répétant son nom coup sur coup afin de le faire bien entrer dans les mémoires : **<q ana="cita_a">« Lastique, Joséphin Lastique.</q>** (maupassant_vie_dr_peg.xml)

(38) Le Corse placide eut un éclair de fierté dans l'œil. **<q ana="cita_a">"Oui, madame, c'était un célèbre, celui-là. Il a mis à bas six gendarmes.[...]"</q>** (maupassant_vie_dr_peg.xml)

4.2.2 Emplois citationnels non autonomes : `<q ana="cita_na"> «...» </q>`

Les emplois citationnels sont non autonomes lorsqu'ils sont intégrés énonciativement et syntaxiquement dans le discours enchâssant. Il s'agit par exemple :

- Des formes du discours indirect introduit par verbe + *que*

(39) Il allait sur les boulevards, de groupe en groupe, pérorait au milieu de la foule silencieuse, plein d'espoir, de bonnes nouvelles, sûr du succès, malgré tout, vous répétant vingt fois de suite que `<q ana="cita_na">` « les cuirassiers blancs de Bismarck avaient été écrasés jusqu'au dernier... » `</q>` (daudet_lundi_dr_peg.xml)

- Des îlots citationnels : la citation, une suite de mots entre guillemets plus petite que la phrase, est intégrée syntaxiquement au discours citant.

(40) Comme on lui avait recommandé de `<q ana="cita_na">`"prendre du mouvement" `</q>`, elle s'acharnait à marcher. (maupassant_vie_dr_peg.xml)

(41) Il était donc probable que Cyrus Smith résoudrait aisément cette question `<q ana="cita_na">`"de continent ou d'île" `</q>`, à laquelle il donnait, non sans raison, le pas sur toutes les autres. (verne_ilemysterieuse_dr_peg.xml)

- Des formes mixtes de type : *selon X, pour X, ou d'après X ...* : bien que l'on soit ici à la frontière entre citation autonome et non autonome. En l'absence de critères énonciatifs et/ou syntaxiques permettant de le décider, ces cas sont considérés comme des citations non autonomes.

(42) Comme la cuisinière n'en finissait pas d'enlever ses vêtements, la veuve Dentu se trouva là juste à point, venue soudain, ainsi que le prêtre, comme s'ils avaient `<q ana="cita_na">`"senti la mort" `</q>`, selon le mot des domestiques. (maupassant_vie_dr_peg.xml)

4.3 La fonction de dénomination : `<q ana="dénomination"> "... " </q>`

Les guillemets ont une fonction de dénomination lorsque leur usage signale un acte de dénomination, l'assignation d'une étiquette dénominative. L'accent est donc mis à la fois sur le signifiant et le signifié. Il peut s'agir :

- d'introduire une catégorisation, un concept, une nouvelle notion ou un néologisme :

(43) La maîtrise de ce code `<q ana="dénomination">`"high" `</q>` apparaît valorisée, explicitement par le fond et implicitement par la forme. (trimaille_dr_peg.xml)

- de signaler un titre (titre d'œuvres, noms de rues, titres honorifiques ou statutaires) :

(44) Le titre de l'album de Freeman **<q ana="dénomination">"L'palais de justice"</q>** par sa position d'exergue semble assez symptomatique de ce phénomène. (trimaille_dr_peg.xml)

Le segment guillemeté peut être précédé de « intitulé », ou « dit » ou « appelé ». Cela peut conduire alors à des cas d'emplois multiples (cf. 4.6 : dénomination + citation, dénomination + modalisation).

4.4 La fonction de modalisation : **<q ana="modalisation"> «...» </q>**

Les guillemets ont une fonction de modalisation quand ils signalent une distance de l'auteur vis-à-vis du terme : l'auteur montre qu'il ne prend pas complètement en charge ce terme. On peut trouver plusieurs cas de figure :

- soit le terme est inapproprié par rapport à ce qu'il désigne :
- soit le terme n'est pas tout à fait satisfaisant vis-à-vis du lecteur (le terme est trop savant, ou au contraire trop familier) :
- soit l'auteur se refuse à prendre en charge le terme car une telle prise en charge est contraire à (ou éloignée de) ses convictions, son point de vue.

Cette fonction est parfois accompagnée des gloses méta-énonciatives (*si je puis dire, comme on dit, passez-moi l'expression...*).

(45) On peut imaginer la réflexion d'un adolescent confronté à l'incompréhension d'un message émanant de ses idoles, censées avoir partagé les mêmes expériences, les mêmes **<q ana="modalisation">"galères"</q>**. (trimaille_dr_peg.xml)

(46) Ensuite, les rappeurs s'affirment également comme détenteurs d'une expertise langagière vis-à-vis de ceux qui mettent systématiquement en équation rap et **<q ana="modalisation">" langage de la rue "</q>**, et qui considèrent cette variété comme un code à la fois bas et restreint : en un mot une **<q ana="modalisation">" sous-langue "</q>** impropre à véhiculer autre chose que des propos incivils, et donc inapte à communiquer hors des cités. (trimaille_dr_peg.xml)

Certains emplois sont proches d'une citation de propos sans locuteur identifié, où l'auteur montre que ce sont d'autres (qui restent indéfinis) qui prennent en charge le segment mis entre guillemets :

(47) Le langage des rappeurs comme celui des jeunes en mal d'insertion est souvent considéré comme **<q ana="modalisation">" vulgaire ou incivil "</q>** (trimaille_dr_peg.xml)

- (48) [...]c'est aussi un moyen d'attirer l'attention de ceux qui le qualifient de **<q ana="modalisation">** " vulgaire **</q>**.
(trimaille_dr_peg.xml)

Dans le cadre de ce projet, la citation est restreinte aux cas où la source est identifiable, de ce fait, les cas cités ci-dessus sont annotés uniquement comme modalisation.

4.5 Le proverbe : **<q ana="proverbe">** « ... » **</q>**

Le proverbe est considéré comme un cas particulier. Il se rapproche d'une citation avec une source collective qui ne peut être identifiée. C'est le cas des proverbes et des adages par exemple, qui appartiennent à tout le monde et à personne, ou encore du discours de la doxa, de l'opinion commune. Comme la source n'est pas réellement identifiable, contrairement aux citations, et que c'est une catégorie un peu trop particulière de modalisation, ils sont intégrés dans une catégorie spécifique.

- (49) **<q ana="proverbe">** "Il n'y a pire sourd que celui qui ne veut entendre" **</q>** (millet_dr_peg.xml)

4.6 **<q ana="autre">** « ... » **</q>**

Quelques cas à part, ne faisant pas partie des fonctions précédentes, ont été annotés avec l'attribut **@ana="autre"**, par exemple :

- (50) Il y avait à l'arrière, devant le rouf, un grand baquet plein d'eau de pluie où l'équipage venait boire, et je me rappelle que, la dernière gorgée finie, chacun de ces pauvres diables secouait son gobelet avec un **<q ana="autre">** " Ah ! " **</q>** de satisfaction, une expression de bien-être à la fois comique et attendrissante.
(daudet_moulin_dr_peg.xml)
- (51) Elle baissa la tête d'un mouvement très lent qui peut-être voulait dire **<q ana="autre">** " oui " **</q>**. Et le prêtre qui jetait encore de l'eau bénite leur en envoya quelques gouttes sur les doigts. (maupassant_vie_dr_peg.xml)
- (52) Elle faisait **<q ana="autre">** " non " **</q>** de la tête, incapable de parler davantage. (maupassant_vie_dr_peg.xml)

4.7 Fonctions multiples

Dans les cas d'emplois multiples (ou fonctions multiples), une valeur s'ajoute à une autre. Les valeurs (généralement limitées à deux) sont co-présentes.

4.7.1 La modalisation dans les cas d'emplois multiples

La valeur de modalisation semble très souvent transparaître dans l'usage des guillemets, de manière plus ou moins explicite. Quand la fonction est clairement perçue, et que la valeur de modalisation semble être également présente, on ne retiendra la valeur *modalisation* que lorsque :

- le passage entre guillemets s'accompagne de gloses méta-énonciatives, signalant la modalisation, par exemple : *si je puis dire, si vous voulez, passez-moi l'expression, en quelque sorte, si l'on veut, n'ayons pas peur des mots, ...*
- et/ou lorsque le terme entre guillemets peut être considéré comme une marque lexicale de subjectivité, c'est-à-dire que son usage est marqué d'un point de vue affectif ou évaluatif.

4.7.1.1 Dénomination et modalisation

Dénomination et modalisation peuvent se superposer, par exemple dans le cas d'un néologisme, un terme étranger, un terme qui est marqué sociolinguistiquement (terme qui fait jeune, branché, etc.) ou un terme trop expert. D'une part, il y a fonction d'assignation d'une étiquette, mais d'autre part, les guillemets signifient que l'auteur émet quelques réserves quant à l'emploi du terme.

(53) Il avait une de ces belles têtes **<q ana="dénomination_modalisation">**numismatiques**</q>**, qui semblent faites pour être frappées en médailles, les yeux ardents, la bouche sérieuse, la physionomie d'un savant de l'école militante. (verne_ilemysterieuse_dr_peg.xml)

4.7.1.2 Modalisation et citation non autonome

Il faut que la valeur modale du passage entre guillemets corresponde à l'expression des sentiments ou de l'évaluation de la part du locuteur du discours citant.

(54) Puis il parla du pays, qu'il déclarait très **<q ana="cita_na_modalisation">**pittoresque**</q>** [...] (maupassant_vie_dr_peg.xml)

4.7.2 Dénomination et citation non autonome

Ce cas est particulièrement fréquent dans le texte scientifique, quand l'auteur utilise un concept et y adhère (dénomination) tout en imputant sa paternité à un autre auteur (citation).

(55) C'est en m'appuyant sur la distinction dégagée par John J. Gumperz (1982) entre **<q ana="cita_na_dénomination">**" we code **</q>** et **<q ana="cita_na_dénomination">**" they code **</q>**, que j'ai étudié les productions de groupes marseillais [...] (trimaille_dr_peg.xml)

On rencontre également fréquemment cette double valeur dans le texte littéraire : l'expression entre guillemets est attribuée au personnage (citation) et à sa vision du monde, sa « conceptualisation » (dénomination) :

(56) Il arrivait le plus souvent vers quatre heures de l'après-midi, rejoignait petite mère dans **<q ana="cita_na_dénomination">** son allée **</q>** et lui offrait le bras pour faire **<q ana="cita_na_dénomination">** son exercice **</q>**. (maupassant_vie_dr_peg.xml)

4.7.3 Citation et autonymie

La double valeur peut apparaître lorsque le passage entre guillemets correspond à un passage cité qui, du point de vue du statut du signe, a une valeur autonymique (entendue ici comme absence de valeur référentielle).

(57) Elle s'aperçut qu'il lui disait **<q ana="autonymie_cita_na">** tu **</q>** maintenant [...] (maupassant_vie_dr_peg.xml)

5. Annotation du lexique des affects : mots simples et collocations

Pour le lexique des affects, le standard TEI-P5 ne s'est pas révélé parfaitement adapté, du fait de la complexité de l'annotation des collocations, dont certains éléments sont reliés à plusieurs autres mots. Par exemple, dans *avoir une peur terrible*, deux collocations sont superposées : *avoir peur* et *peur terrible*. Pour faciliter ces traitements, nous avons opté pour la norme Language resource management — Morpho-syntactic annotation framework (MAF) (ISO/FDIS 24611), dans une sérialisation TEI.

L'annotation concerne les mots simples d'affect (par exemple, *plaisir*, *apprécier*, *admirer*) mais aussi les collocations du type *avoir peur* ou *frissonner de plaisir*.

L'annotation est ici en partie déportée, c'est-à-dire que le repérage des segments annotés est dissocié des informations sur ces éléments. Les éléments lexicaux (mots simples et collocatifs) sont signalés dans le texte par l'élément `<w>` avec le numéro du paragraphe et le numéro du token.

Pour les choix linguistiques effectués, le lecteur pourra se reporter à (Augustyn et al. 2008) et (Augustyn & Tutin, 2009).

5.1 Annotation des mots simples

L'annotation des mots simples se présente dans le texte de la manière suivante :

```
(58) <p xml:id="1041">- Adieu les stylets, dit <w  
xml:id="p1041_t1">gaiement</w> Colomba; maintenant j'ai un  
éventail, pour vous en donner sur les doigts quand vous  
direz du mal de mon pays. </p> (merimee_colomba_emo.xml)
```

À la suite de chaque paragraphe l'élément `<div type="wordForm">` associe l'annotation aux segments indiqués dans le texte (avec le numéro du paragraphe).

```
<div type="wordForm">  
  <spanGrp type="wordForm">  
    <span target="#p1041_t1" ana="#fs1" corresp="#entry1"/>  
  </spanGrp>  
</div>
```

L'élément suivant `<div type="lexicon">` comprend les différentes propriétés du mot d'affect. Deux ensembles de propriétés sont annotés :

- `<form>` contenant la forme lemmatisée `<orth>` et la catégorie grammaticale `<pos>`
- `<sense>` contenant une série de propriétés (`<f>`) sémantiques et syntaxiques.

Exemple : Propriétés du mot d'affect *gaiement*

```
<div type="lexicon">
  <entry xml:id="entry1">
    <form>
      <orth>gaiement</orth>
      <pos>ADV</pos>
    </form>
    <sense>
      <fs xml:id="fs1" corresp="#entry1">
        <f name="semanticType">
          <symbol value="affect"/>
        </f>
        <f name="semanticDomain">
          <symbol value="gaieté"/>
        </f>
        <f name="register">
          <symbol value="courant"/>
        </f>
        <f name="intensity">
          <symbol value="moyen"/>
        </f>
        <f name="polarity">
          <symbol value="positif"/>
        </f>
      </fs>
    </sense>
  </entry>
</div>
```

5.1.1 Propriétés du lexique des affects³

La plupart des propriétés annotées sont communes à toutes les catégories grammaticales des mots d'affect. Les propriétés suivantes sont annotées :

Type sémantique : <f name="semanticType"><symbol value=" " /></f>

Cette propriété, avec dans la plupart des cas la valeur "affect", permet d'indiquer l'appartenance au lexique des affects.

Dans le cas des noms (uniquement), les valeurs sont différentes et correspondent au traitement de la polysémie. Plusieurs configurations sont possibles :

- 1) **Affect** – lorsque le nom dénote uniquement un affect. Certains critères déterminent cet emploi :
 - a. Le nom se combine généralement avec *éprouver*, *ressentir* ;
 - b. Il a toujours un agent humain ; par métonymie, il peut porter sur une partie du corps ;
 - c. Il exprime un mouvement intérieur, « mental » (par opposition aux pures sensations, ou aux attitudes) ;
 - d. Il a une cause et/ou un objet ;

³ Augustyn M. et al. (2008) présente les détails de l'annotation avant la mise aux normes XML/TEI.

e. Il est souvent introduit par un déterminant partitif quand le nom n'est pas modifié ;

2) **Affect+...** - Un nom dénote un affect couplé à une dimension sémantique particulière, par exemple :

affect+manif – l'affect est manifesté physiquement, le nom d'affect est souvent complément de cause du verbe ou du nom (*pleurer de joie, un cri d'indignation*) :

(59) Les yeux de Colomba <w xml:id="p287_t2"> brillaient </w> d'une <w xml:id="p287_t3"> **joie**</w> <w xml:id="p287_t4"> maligne</w> qu'elle n'y avait point encore remarquée. (merimee_colomba_emo.xml)

→ <f name="semanticType"><symbol value="affect+manif"/>

affect+attitude – lorsque le nom d'affect est introduit par *avec, sans* ou dans des expressions telles que *un air de, un ton de ...*, il désigne une attitude qui accompagne une autre action. Au plan syntaxique, c'est souvent un complément de manière.

(60) Après une heure de travail, Pencroff était en nage, et il jeta les morceaux de bois avec <w xml:id="p596_t1"> **dépit**</w>. (verne_ilemysterieuse_emo.xml)

→ <f name="semanticType"><symbol value="affect+attitude"/>

affect+relation - Le nom exprime à la fois un affect et une relation institutionnalisée entre plusieurs personnes (*amitié, inimitié, ...*) :

(61) Peu à peu, quand il les vit honnêtes, énergiques, liés les uns aux autres par une <w xml:id="p4834_t1"> **amitié**</w> <w xml:id="p4834_t2"> fraternelle</w>, il s'intéressa à leurs efforts. (verne_ilemysterieuse_emo.xml)

→ <f name="semanticType"><symbol value="affect+relation"/>

affect+manif_attitude – l'affect s'accompagne d'une manifestation physique, d'un comportement ou d'une attitude mais il est difficile de décider si l'on a affaire à un type ou l'autre :

(62) Pencroff ne <w xml:id="p278_t1"> cacha</w> point son <w xml:id="p278_t2"> **désappointement**</w> très <w xml:id="p278_t3"> vif</w>. (verne_ilemysterieuse_emo.xml)

→ <f name="semanticType"><symbol value="affect+manif_attitude"/>

affect+sensation - dans certains contextes, les noms d'affect s'apparentent davantage à des sensations (*plaisir, jouissance, douleur, souffrance*) : aspects physiques et « moraux » sont indissociables :

(63) D'habitude, sitôt la forge éteinte, le soleil couché, il s'asseyait sur un banc devant sa porte pour <w xml:id="p251_t1"> savourer</w> cette <w xml:id="p251_t2"> bonne</w> <w xml:id="p251_t3"> **lassitude**</w> que donne le poids du travail et de la chaude journée [...]
(daudet_lundi_emo.xml)

→ <f name="semanticType"><symbol value="affect+sensation"/>

3) **Non affect+...** - Un nom ne dénote pas un affect dans un contexte donné mais il constitue une acception polysémique de l'emploi « affect », par exemple :

non affect+cause_objet - Le nom correspond à la cause ou à l'objet (dans le cas de noms d'affect interpersonnels comme *amour* ou *passion*, mais pas *tendresse* ou *haine*). Il est généralement comptable. Il peut renvoyer à des éléments humains (*ses amours étaient tous bruns*). Ces acceptions des noms sont souvent introduites par le verbe *être*.

(64) Il ne reculerait devant aucune tâche, et, chasseur passionné, il ferait un métier de ce qui, jusqu'alors, n'avait été pour lui qu'un <w xml:id="p931_t1"> **plaisir**</w>. (verne_ilemysterieuse_emo.xml)

→ <f name="semanticType"><symbol value="non affect+cause_objet"/>

non affect+état – Le nom (par exemple *stupeur*, *solitude*, *accablement*) décrit davantage la manière d'être, stable ou variable :

(65) Et qui sait ce que deviendrait le dernier vivant de nous, après une longue <w xml:id="p2952_t1"> **solitude**</w> sur cette île ? (verne_ilemysterieuse_emo.xml)

→ <f name="semanticType"><symbol value="non affect+état"/>

non affect+qualité – Le nom (par exemple *bonté*, *charité*, *timidité*, *méfiance*) décrit davantage une propriété/un trait de caractère qu'un affect éprouvé :

(66) Toujours souriant et doux, M. Viot répondit que sa <w xml:id="p270_t1"> **bienveillance**</w> m'était acquise et qu'il m'aiderait volontiers de ses conseils; (daudet_petitchose_emo.xml)

→ <f name="semanticType"><symbol value="non affect+qualité"/>

non affect+qualité_causative - Le nom décrit une propriété/un caractère d'une chose ou d'un événement qui cause le ressenti d'un affect particulier : *la tristesse d'un film*) :

(67) Une banquise des mers glaciales n'eût pas été plus capricieusement dressée dans sa sublime <w xml:id="p3365_t1">**horreur**</w> (verne_ilemysterieuse_emo.xml)

→ <f name="semanticType"><symbol value="non affect+qualité_causative"/>

non affect+verbalisation – des verbalisations qui peuvent véhiculer l'affect du même nom :

(68) Monsieur della Rebbia, dit-il, je suis chargé de bien des compliments pour vous par vos amis anglais: miss Nevil fait mille <w xml:id="p526_t1"> **amitiés**</w> à mademoiselle votre sœur. (merimee_colomba_emo.xml)

→ <f name="semanticType"><symbol value="non affect+verbalisation"/>

non affect+action - des noms comme *hostilité* peuvent avoir un sens de nom d'action (nom extensif, sensible aux phases, au temps) :

(69) Mais les deux factions étaient privées de leurs chefs, et les Corses, disciplinés dans leurs <w xml:id="p822_t5"> **fureurs**</w>, en viennent rarement aux mains dans l'absence des principaux auteurs de leurs guerres intestines. (merimee_colomba_emo.xml)

→ <f name="semanticType"><symbol value="non affect+action"/>

Domaine sémantique : <f name="semanticDomain"><symbol value="..."></f>

L'attribut <f name="semanticDomain"> comprend le nom du domaine tel qu'établi dans la typologie notionnelle élaborée au préalable⁴. Il s'agit d'étiquettes conventionnelles regroupant une même « famille » sémantique des mots d'affect. Par exemple, le domaine « peur » sera commun aux mots suivants : *peur, effroi, horreur, terreur, effrayer, affoler, terrible, effrayant, peureux, lâche, épouvantable, etc.*

Il est à noter que certains mots se trouvent dans plusieurs domaines, du fait de la polysémie qui leur est propre. *Horreur*, par exemple, se trouve à la fois dans le domaine « peur » (ex : *film d'horreur*) et dans le domaine « dégoût » (ex : *cette attitude lui fait horreur*), l'appartenance à l'un ou l'autre domaine étant déterminée par le contexte.

⁴ La liste complète des domaines est : abatement, admiration, adoration, affection, agacement, bienveillance, bonheur, convoitise, déception, dégoût, désir, doute, émoi, émotion, enthousiasme, étonnement, fierté, gaieté, gêne, goût, hargne, honte, hostilité, indifférence, inquiétude, insatisfaction, jalousie, lassitude, méfiance, mélancolie, mépris, nervosité, passion, peur, pitié, plaisir, reconnaissance, regret, respect, souffrance, soulagement.

Certains mots d'affect appartiennent simultanément à deux domaines sémantiques. Par exemple le verbe *s'émerveiller* appartient au domaine « étonnement » et « admiration » simultanément. Cette double appartenance est codée de la manière suivante :

```
<f name="semanticDomain">
  <symbol value="étonnement+admiration"/>
</f>
```

Type syntaxique : <f name="syntacticType"> <symbol value="agentif|causatif|autre"/></f>

Les verbes et les adjectifs sont classés selon l'opposition *agentif/causatif* fondée principalement sur les propriétés sémantiques, par exemple :

- le type "agentif" correspond à un sujet qui est expérimenteur de l'affect, par exemple *aimer, s'étonner, regretter, content, enthousiaste* ;
- le type "causatif" correspond à un sujet qui est cause ou objet de l'affect, par exemple *rassurer, décevoir, fascinant, étonnant, ennuyeux* ;
- la valeur "autre" apparaît dans le cas de la forme participiale du verbe, par exemple : *les débats passionnés*.

Pronominal : <f name="pronominal"> <symbol value="-|+"></f>

La description des verbes comporte une information sur la forme pronominale de certains d'entre eux.

Registre :

```
<f name="register"><symbol value="courant|familier|littéraire|vieilli|vieux|psycho"/></f>
```

Cette propriété permet de préciser le niveau de langue (courant, familier, littéraire, etc.) ou l'appartenance à une langue de spécialité. Les valeurs sont basées sur les informations des dictionnaires.

Intensité : <f name="intensity"> <symbol value="moyen|haut"/></f>

Cet élément permet de spécifier le degré d'intensité (haut ou moyen) de l'affect. Il s'agit d'une propriété inhérente au nom d'affect, les propriétés externes (la combinatoire ou la détermination) n'étant pas prises en compte lors de l'attribution de ce trait.

Lorsque le degré d'intensité est indéterminé la valeur reste vide.

Polarité : <f name="polarity"> <symbol value="positif|négatif|neutre"/></f>

La polarité caractérise les mots d'affect selon l'opposition « positif » vs « négatif », selon que l'affect est agréable (par exemple *joie, amour*) ou désagréable (*tristesse, épouvanter*). Certains mots d'affect ont la valeur polarité "neutre", dans le cas où il n'est pas possible d'attribuer la valeur positive ou négative, par exemple : *éprouver, émouvoir, goût, étonner*.

Le tableau ci-dessous résume les différentes propriétés annotées en fonction de la catégorie syntaxique des mots d'affects :

Catégorie	Propriétés annotées
Verbe : <pos>V</pos> + la forme participiale : <pos>VPPADJ</pos>	Type sémantique : <f name="semanticType"><symbol value="..."/></f> Domaine sémantique : <f name="semanticDomain"><symbol value="..."/></f> Type syntaxique : <f name="syntacticType"> <symbol value="..."/></f> Pronominal : <f name="pronominal"> <symbol value="..."/></f> Registre : <f name="register"><symbol value="..."/></f> Intensité : <f name="intensity"> <symbol value="..."/></f> Polarité : <f name="polarity"> <symbol value="..."/></f>
Nom : <pos>N</pos>	Type sémantique : <f name="semanticType"><symbol value="..."/></f> Domaine sémantique : <f name="semanticDomain"><symbol value="..."/></f> Registre : <f name="register"><symbol value="..."/></f> Intensité : <f name="intensity"> <symbol value="..."/></f> Polarité : <f name="polarity"> <symbol value="..."/></f>
Adjectif : <pos>A</pos>	Type sémantique : <f name="semanticType"><symbol value="..."/></f> Domaine sémantique : <f name="semanticDomain"><symbol value="..."/></f> Type syntaxique : <f name="syntacticType"> <symbol value="..."/></f> Registre : <f name="register"><symbol value="..."/></f> Intensité : <f name="intensity"> <symbol value="..."/></f> Polarité : <f name="polarity"> <symbol value="..."/></f>
Adverbe : <pos>ADV</pos>	Type sémantique : <f name="semanticType"><symbol value="..."/></f> Domaine sémantique : <f name="semanticDomain"><symbol value="..."/></f> Registre :

	<pre> <f name="register"><symbol value="..."/></f> Intensité : <f name="intensity"> <symbol value="..."/></f> Polarité : <f name="polarity"> <symbol value="..."/></f> </pre>
--	---

Tableau 4 : Propriétés annotées selon la catégorie grammaticale du mot d'affect

5.2 Annotation des collocations

Les principes linguistiques de l'annotation des collocations sont décrits dans (Augustyn & Tutin 2009).

L'annotation des collocations se présente dans le texte de la manière suivante, chaque mot de la collocation étant annoté, ici *avoir + plaisir* :

```

(70) <p xml:id="73">-Votre père! Par ma foi, c'était un brave!
J'<w xml:id="p73_t1">aurais</w> du <w
xml:id="p73_t2">plaisir</w> à le revoir, et je le
reconnaitrais, j'en suis sûr. </p>
(merimee_colomba_emo.xml)

```

À la suite de chaque paragraphe, l'élément `<div type="wordForm">` associe l'annotation aux segments indiqués dans le texte où apparaît la collocation. L'élément `<link type="collocation" target="#span2 #span1"/>` permet d'établir un lien entre deux éléments (la base et le collocatif) et d'en établir la nature (collocation).

```

<div type="wordForm">
  <spanGrp type="wordForm">
    <span xml:id="span2" target="#p73_t2" ana="#fs2"
corresp="#entry2"/>
    <span xml:id="span1" target="#p73_t1" ana="#fs1"
corresp="#entry1"/>
  </spanGrp>
  <link type="collocation" target="#span2 #span1"/>
</div>

```

L'élément `<div type="lexicon">` comprend les différentes propriétés du mot d'affect et de son collocatif. Dans les deux cas, deux ensembles de propriétés sont annotés :

- **<form>** contenant la forme lemmatisée `<orth>` et la catégorie grammaticale `<pos>`. Les catégories grammaticales sont notées de la manière suivante :
 - Pour les mots d'affects : V (verbe), N (nom), A (adjectif), ADV (adverbe), VPPADJ (verbe participé passé) ;
 - Pour les collocatifs : N (nom), A (adjectif), ADV (adverbe), vt (verbe transitif), vi (verbe intransitif), vti (verbe transitif indirect), vpronti (verbe pronominal transitif indirect), vpront (verbe pronominal transitif), locadv (locution adverbiale), prep (préposition), locvi (locution verbale indirecte), locvti (locution verbale transitive indirecte), locadj (locution adjectivale), locprep (locution prépositionnelle), locvt (locution verbale transitive).

- **<sense>** contenant une série de propriétés (<f>) sémantiques et syntaxiques (détaillées plus loin, 5.2.1 et 5.2.2), ainsi que l'information sur la nature du composant de la collocation (la base ou le collocatif).

Exemple : Propriétés du mot d'affect *plaisir*

```

<div type="lexicon">
  <entry xml:id="entry2">
    <form>
      <orth>plaisir</orth>
      <pos>N</pos>
    </form>
    <sense>
      <fs xml:id="fs2" corresp="#entry2">
        <f name="semanticType">
          <symbol value="affect"/>
        </f>
        <f name="semanticDomain">
          <symbol value="plaisir"/>
        </f>
        <f name="register">
          <symbol value="courant"/>
        </f>
        <f name="intensity">
          <symbol value="moyen"/>
        </f>
        <f name="polarity">
          <symbol value="positif"/>
        </f>
        <f name="typecollocation">
          <symbol value="base"/>
        </f>
      </fs>
    </sense>
  </entry>

```

Exemple : Propriétés du collocatif *avoir* du mot d'affect *plaisir*

```

<entry xml:id="entry1">
  <form>
    <orth>avoir</orth>
    <pos>vt</pos>
  </form>
  <sense>
    <fs xml:id="fs1" corresp="#entry1">
      <f name="FLsemanticType">
        <symbol value="éprouver"/>
      </f>
      <f name="FL">
        <symbol value="oper1"/>
      </f>
      <f name="typecollocation">
        <symbol value="collocatif"/>
      </f>
    </fs>
  </sense>

```

```
</entry>  
</div>
```

5.2.1 Propriétés du lexique des affects

Les propriétés du lexique des affects sont annotées de la même manière que dans le cas de l'annotation des mots simples (la partie 5.1.1.).

5.2.2 Propriétés des collocatifs

Le traitement et la délimitation des collocations sont fondés sur certains principes.

Le premier point concerne les **phénomènes de « mise en facteur »** d'éléments de la collocation. Il s'agit de deux cas de figure :

- une base utilisée par plusieurs collocatifs (*avoir une peur terrible = avoir peur + peur terrible*)
- un collocatif portant sur plusieurs bases (p.ex. *il pleurait de rage et de désespoir*).

Les deux phénomènes sont traités de la même manière que les collocations « simples » : les liens entre les bases et leurs collocatifs sont signalés en début de chaque annotation (élément `<div type="wordForm">`).

Une autre question se pose pour la **délimitation des collocatifs complexes** :

- dans le cas des **collocatifs associés aux formes composées** (formes pronominales, temps composés, formes passives), l'élément du collocatif contient uniquement la forme lexicale pleine, à l'exclusion des mots purement grammaticaux (pronom réfléchi, auxiliaire). Les pronoms réfléchis sont intégrés dans le collocatif uniquement dans le cas des verbes intrinsèquement pronominaux comme *se pâmer (de joie)* ;
- dans le cas des **collocatifs incluant plusieurs lexies** (par exemple « *ne pas se tenir* » dans *ne pas se tenir de joie*), lorsque tous les éléments apparaissent obligatoires ils sont intégrés dans l'élément du collocatif.

L'annotation des collocations est fondée sur le modèle des Fonctions Lexicales (Mel'čuk I., Clas A. et Polguère A., 1995, Mel'čuk I. et al., 1984, 1988, 1992, 2000) qui est adapté au codage syntaxique et sémantique des collocations. Ainsi, deux principales propriétés du collocatif sont annotées : **fonction lexicale** et **glose sémantique**.

5.2.2.1 Fonctions lexicales : `<f name="FL"><symbol value=""/>`

Les Fonctions Lexicales permettent de modéliser et d'encoder les liens lexicaux entre les éléments de la collocation (la base et le collocatif). Il s'agit des fonctions lexicales syntagmatiques (standard), qui permettent d'encoder des liens de combinatoire

La présentation d'une fonction lexicale permet de dénoter :

- Le sens général
- La configuration syntaxique d'emploi

Une fonction lexicale s'applique à une unité lexicale (la base) et décrit une valeur d'une autre unité lexicale (le collocatif) en combinaison avec la première. Par exemple, la fonction Magn qui dénote les modificateurs d'intensité permet de modéliser la relation de nombreuses collocations, par exemple :

Magn (*gai*) = *comme un pinson*

(71) Il était toujours rouge comme une tomate, tantôt <w xml:id="p37_t1">gai</w> <w xml:id="p37_t2">comme un pinson</w>, et tantôt sérieux comme un pape. (leroux_chambrejaune_emo.xml)

Magn (*désespoir*) = *sans bornes*

(72) C'est vrai... son <w xml:id="p85_t1">désespoir</w> doit être <w xml:id="p85_t2">sans bornes</w>! (leroux_chambrejaune_emo.xml)

Magn (*dégoûter*) = *profondément*

(73) J'en étais <w xml:id="p764_t2">profondément</w> <w xml:id="p764_t3">dégoûté</w>. (leroux_chambrejaune_emo.xml)

La fonction lexicale Oper_n dénote les verbes supports : Oper₁ si le premier actant de la base, l'agent, est sujet du verbe support et Oper₂ si le deuxième actant de la base est le sujet du verbe support :

Oper₁ (*envie*) = *avoir*

(74) Mais je vis bien que le jeune homme n'avait jamais <w xml:id="p218_t1">eu</w> moins <w xml:id="p218_t2">envie</w> de rire [...] (leroux_chambrejaune_emo.xml)

Oper₂ (*amour*) = *obtenir*

(75) Enfin, en récompense de tant d'exploits, le pêcheur d'anchois <w xml:id="p645_t1">obtient</w> l'<w xml:id="p645_t2">amour</w> d'Estérelle, et il est nommé consul par les habitants de Cassis. (daudet_moulin_emo.xml)

Les différentes fonctions lexicales peuvent se combiner en fonctions lexicales complexes permettant ainsi de préciser certaines nuances sémantiques, par exemple :

- la fonction Magn peut se combiner avec la fonction Oper_n afin de dénoter un verbe support à valeur intensive, par exemple :

Magn+Oper₁ (*envie*) = *mourir*

(76) Miss Nevil avait oui parler des improvisatrices corses et
 <w xml:id="p206_t1">mourait</w> d'<w xml:id="p206_t2">
 envie</w> d'en entendre une. (merimee_colomba_emo.xml)

- la fonction Oper_n peut se combiner également avec les fonctions dénotant la valeur aspectuelle, comme Incep (valeur inchoative) :

incepoper1 (*colère*) = *entrer*

(77) Le soir, quand je contai mon aventure à l'ami Jacques, il
 <w xml:id="p1158_t1">entra</w> dans une belle <w
 xml:id="p1158_t3">colère</w>. (daudet_petitchose_emo.xml)

- La fonction Func_n qui désigne un verbe support d'un nom qui fonctionne comme son sujet grammatical se combine par exemple avec la fonction lexicale causative liqu :

liquifunc0 (*angoisse*) = *échapper*

(78) Pour <w xml:id="p130_t1"> échapper</w> à cette <w
 xml:id="p130_t2"> angoisse</w>, il se mit à boire.
 (daudet_lundi_emo.xml)

L'attribution des fonctions lexicales a été fondée sur les ouvrages et dictionnaires de référence de la lexicologie explicative et combinatoire, comme le Dictionnaire Explicatif et Combinatoire (Mel'čuk I. et al., 1984, 1988, 1992, 2000) et la version électronique Dicouèbe⁵ : <http://olst.ling.umontreal.ca/dicouebe/>.

5.2.2.2 Type sémantique / glose sémantique : <f name="FLsemanticType"> <symbol value=" "/>

L'annotation de gloses sémantiques devrait permettre une facilité d'accès au sens de la collocation à l'aide d'un système de paraphrasage simplifié. Le système est basé sur des étiquettes simples comme *commencer*, *intense*, *positif*, *négatif*, opérateur causatif *faire*, par exemple :

COLLOCATION (base + collocatif)	GLOSE SÉMANTIQUE	FONCTION LEXICALE
<i>se mettre + (en) colère</i>	/commencer_éprouver/	IncepOper1 ; IncepPred
<i>passion + fermenter</i>	/augmenter/	IncepPredPlus
<i>mettre le comble + (à la) stupéfaction</i>	/faire augmenter/	CausPredPlus
<i>transporté + (de) joie</i>	/affecter/	Func1
<i>remord + ronger</i>	/affecter vivement/	Magn+Fact1; Magn+Func1
<i>anxiété + inexprimable</i>	/intense/	Magn
<i>fureur + sourde</i>	/peu intense/	AntiMagn
<i>abattre + peine</i>	/faire cesser/	causfinfunc0

⁵ La documentation disponible sur : <http://olst.ling.umontreal.ca/dicouebe/DiCoDOC.pdf>

Tableau 5. Exemple de correspondances glose/Fonction Lexicale⁶

Le problème des étiquettes sémantiques s'est posé dans le cas des relations plus spécifiques :

- fonctions non standard - la fonction lexicale n'est pas systématisée parce que la relation qui relie les éléments de la collocation n'est pas productive, par exemple : *liens d'amitié, crainte enfantine / puérile, amour maternel, amitié fraternelle.*
- fonctions standard difficilement paraphrasables - même si la fonction lexicale est annotée, il est difficile dans de trouver une glose simple, comme dans :
NonOper1 : *soyez sans (inquiétude, crainte)*
A1Real2 : *stupéfait d'admiration*

Dans ces deux cas, la collocation est signalée sans traitement spécifique associé (statistiquement, ces cas sont relativement peu nombreux, le système d'étiquettes sémantiques est dans l'ensemble très couvrant).

Bibliographie

- Augustyn M., & Tutin A. (2009). *Constitution d'un corpus annoté autour du lexique des émotions: collocations et fonctions lexicales*. In Beck D., Gerdes K., Milićević J. and Polguère A. Actes de la Quatrième conférence internationale sur la Théorie Sens-Texte (MTT' 09), OLST (pp. 25-35). Disponible à l'adresse : <http://meaningtext.net/mtt2009/03-MAugustynATutin.pdf>
- Augustyn M., Ben Hamou S., Bloquet G., Goossens V., Rinck F., Loiseau M. (2008). *Constitution de ressources pédagogiques numériques : le lexique des affects*. In *Autour de la langue et du langage : perspective pluridisciplinaire*, Presses Universitaires de Grenoble, Grenoble.
- Mel'čuk I. et al. (1984, 1988, 1992, 2000). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I, II, III, IV*. Montréal : Les Presses de l'Université de Montréal.
- Mel'čuk I., Clas A. et Polguère A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain : Duculot.
- Rinck F., Grossmann F., Boch F. (2009). *Observer le rôle des guillemets dans un corpus : une voie d'accès à l'énonciation ?*. Colloque énonciation et texte au cœur de la grammaire (11-13 mars 2009). Toulouse.
- Rinck F. & Tutin A. (2007). *Annoter la polyphonie dans les textes : le cas des passages entre guillemets*. *Corpus*, 6:79-100.

⁶ Liste complète des gloses sémantiques utilisées : à_cause_de, affecter, affecter vivement, agréable, améliorer, assouvir, augmentation, augmenter, avoir_lieu, bon, caché, causer, cesser, combattre, contenu, continuation, continuer, contrôler, court, croissant, début, débutant, déclinant, dégradation, dégrader, désagréable, diminuer, diminution, disséminer, dissimuler, durer, élément, éprouver, éprouver vivement, exagérer, excès, excessif, exubérant, faire augmenter, faire cesser, faire croître, faire diminuer, faire naître, faire naître à nouveau, fermenter, figure, fin, gagner, général, incontrôlé, intense, intense et bien, intense et négatif, intense et réprovable, intensément, intensité, irrégulier, justifié, léger, long, manifestant, manifestation, manifestation_physique, manifeste, manifester, manifester_physiquement, mauvais, maximum, naître, obtenir, réalisation, réaliser, renaître, se disséminer, subir, suppression, supprimer, vaincre.

- Rosier L. (1999). *Le discours rapporté. Histoire, théories, pratiques*. Paris, Bruxelles : De Boeck-Duculot.
- Rosier L. (2008). *Le discours rapporté en français*. Paris : Ophrys.