**GUIDE D’ANNOTATION**

**CORPUS D’APPRENANTS**

**ETAPES DE TRAITEMENT**

1. **Transformation des documents .odt vers le format xml/TEI P5**

🡪 Automatique : <https://resultats.hypotheses.org/267>

1. **Contrôle de la mise en page**
   1. **Simplification de la mise en page**

Suppression de certaines marques de mise en page (tabulations, couleurs, espaces vides, etc.)

* 1. **Tableaux et figures**

Simplification des tableaux et figures par la suppression du contenu et le remplacement par les balises vides comportement un titre si présent dans le texte, par exemple :

<figure>  
 <head>Localisation du peuple naxi.</head>  
 </figure>

Figure 1 : Présentation des figures

<table>  
 <head>Tableau 1 : Phonèmes consonantiques du thaï « standard » et du dialecte du  
 sud</head>  
 <row>  
 <cell/>  
 </row>  
 </table>

Figure 2 : Présentation des tableaux

* 1. **Notes de bas de page**

Les notes de bas de page sont insérées dans le texte à l’endroit de l’appel de la note de bas de page signalé par un numéro (1, 2…) ou un signe particulier (\*). Elles sont annotées à l’aide de l’élément **<note>** dont les attributs précisent le type et l’ordre de l’appel de note (**n="1", n="2", n="\*"**) et l’emplacement (**place="foot"**).

Cette famille, n’a, semble-t-il,  
aucune parenté avec la famille sino-tibétaine**<note n="1" place="foot">**Les langues sino-tibétaines sont une famille de langues origines d’Asie. **</note>**. ». Cependant, suite à…

Figure 3 : Annotation des notes de bas de page

1. **Structuration du document**

**3.1 Structure globale d’un fichier XML/TEI**

La structure globale d’un fichier XML/TEI correspond au schéma suivant :

<TEI>

<teiHeader>métadonnées</teiHeader>

<text>données

<front>... </front>

<body>...</body>

<back>...</back>

</text>

</TEI>

**<body>** est un élément pour structurer le corps du texte proprement dit, **<front>** et **<back>** sont deux éléments pour encoder le paratexte.

* L'élément **<front>** regroupe tout ce qui précède le corps du texte <body> :
  + La page de titre,
  + Sommaire, tables

<front>  
 <titlePage rend="center">  
 <docTitle>  
 <titlePart type="main">Typologie des langues</titlePart>  
 </docTitle>  
 <byline><docAuthor>M1\_Lin\_OmH\_som</docAuthor></byline>  
 <docImprint>  
 <lb/>UFR LLASIC

<lb/>Département Sciences du langage et didactique du FLE

<lb/>Master 1 Sciences du langage

<lb/>Parcours Linguistique

<lb/>Année universitaire 2017-2018

</docImprint>  
 </titlePage>  
 </front>

Figure 4 : Présentation de la page de titre

<front>  
 <div type="table">  
 <head>Sommaire</head>  
 <p>Introduction et genèse de travail ...3 </p>  
 <p>Partie I. Étude linguistique sur le passif en français ...5 </p>  
 <p>1.1 Forme et transformation du passif en français ...5</p>  
 <p>1.2 Valeur descriptive du passif en français ...7</p>  
 <p>Partie II. Étude linguistique sur le passif en thaï ...9 </p>  
 <p>2.1 Caractéristiques de la langue thaïe ...9</p>  
 <p>2.2 Caractéristiques du passif en thaï ...10</p>  
 </div>  
 </front>

Figure 5 : Présentation du sommaire

* L'élément **<back>** regroupe tout ce qui suit le corps du texte <body>, c’est-à-dire les parties post-liminaires :
  + Sommaire, tables
  + Bibliographie, sitographie
  + Annexes

<back>  
 <div type="bibliographie">  
 <head>Bibliographie</head>  
 <listBibl>  
 <bibl>Japhug. (2015). Dans <hi rend="i">Wikipedia.</hi> Repéré le 14 janvier 2016 à <ref  
 target="https://fr.wikipedia.org/wiki/Japhug"  
 >https://fr.wikipedia.org/wiki/Japhug</ref></bibl>  
 <bibl>LACITO, CNRS, Paris III, Paris IV (UMR 7107). <hi rend="i">Langues et civilisations à  
 tradition orale.</hi> Repéré à <ref target="http://lacito.vjf.cnrs.fr/"  
 >http://lacito.vjf.cnrs.fr/</ref></bibl>  
 </listBibl>  
 </div>  
 </back>

Figure 6 : Présentation de la bibliographie

<back>  
 <div type="annexe">  
 <head>Annexe 1</head>  
 <p>Ainsi équipé de procedures d’apprentissage endogène sur corpus et de stratégies de  
 résolution des ambiguïtés qui exploitent les informations aquises, l’analyseur est capable  
 à la fois d’identifier des groupes syntaxiques et d’analyser le détail de leur  
 constitution.</p>  
 <p>Parmi celle-ci, nous avons fait le choix d’analyser la théorie de J. - C. Milner, telle  
 qu’elle est décrite dans son ouvrage “ l’introduction à une science ”.</p>  
 </div>  
 </back>

Figure 7 : Présentation de l’annexe

**3.2 Structuration interne du texte**

Si le texte est structuré en parties distinctes : introduction, partie, sous-partie, conclusion, il sera nécessaire de restituer cette structure en format XML/TEI.

Cette structuration se fera à l’aide de l’élément **<div>** (division du texte) qui contient une subdivision dans le texte suivant la structure hiérarchique (n="1" pour une partie supérieure, n="2" pour ses sous-parties, etc.). Il conviendra également noter le titre de la section/sous-section : **<head>…</head>**.

Le structure interne du texte correspond au schéma suivant (modulable selon le nombre de sections et de sous-sections) :

SOUS-PARTIES

PARTIE 1

CONCLUSION

INTRODUCTION

<body>  
 <div type="introduction">  
 <head>Introduction</head>  
 <p>Texte de l’introduction</p>  
 </div>  
 <div type="section" n="1">  
 <head>Titre de la partie 1</head>  
 <div type="section" n="2">  
 <head>Titre de la sous-partie 1</head>  
 <p>Texte de la sous-partie 1</p>  
 </div>  
 <div type="section" n="2">  
 <head>Titre de la sous-partie 2</head>  
 <p>Texte de la sous-partie 2</p>  
 </div>  
 </div>  
 <div type="section" n="1">  
 <head> Titre de la partie 2</head>  
 <p> Texte de la partie 3</p>  
 </div>  
 <div type="section" n="1">  
 <head>Titre de la partie 3</head>  
 <p>Texte de la partie 3</p>  
 </div>  
 <div type="conclusion">  
 <head>Conclusion</head>  
 <p>texte de la conclusion</p>  
 </div>  
 </body>

Figure 8 : Annotation de la structure interne du texte.

<body>  
 <div type="introduction">  
 <head>Introduction</head>  
 <p>Ici nous est proposé un texte de pétition qui, comme toute requête, vise un public et  
 tend à changer le cours des choses. …</p>  
 </div>  
 <div type="section" n="1">  
 <head>1) Structure et Enonciation</head>  
 <div type="section" n="2">  
 <head>a) Structure syntaxique du texte</head>  
 <p>Ce texte est un discours clair et, plus ou moins, concis. D’ailleurs, le titre nous  
 donne déjà un avant-goût du thème du texte. Ce dernier commence par une exclamation qui  
 vise à alerter le destinataire : « Catastrophe ! », …</p>  
 </div>  
 <div type="section" n="2">  
 <head>b) Situation énonciative</head>  
 <p>Dans le texte, la présence de l’émetteur passe par la généralisation ou par le « nous »  
 de modestie : « nous, notre ». En effet, il s’adresse à un récepteur in absentia. Ce  
 dernier est aussi inclus dans le « nous ». …</p>  
 </div>  
 </div>  
 <div type="section" n="1">  
 <head>2) Subjectivité et émotions</head>  
 <p>A priori, nous pourrons dire que la <hi rend="b">fonction référentielle</hi> du langage,  
 qui sert à décrire le monde, est exploitée dans ce texte. …</p>  
 </div>  
 <div type="section" n="1">  
 <head>3) La contribution de ces marques linguistiques à la force argumentative du  
 texte</head>  
 <p>Ce texte de pétition suit un schéma argumentatif puisque l’auteur envisage de convaincre  
 une cible. Il essaie de mobiliser le plus de monde possible et, pour ce faire, utilise les  
 moyens ou marques linguistiques dont nous avons parlé plus haut.</p>  
 </div>  
 <div type="conclusion">  
 <head>Conclusion</head>  
 <p>Agir sur la cible et la faire réagir, tel est le but ultime de ce texte. L’auteur met en  
 place bon nombre de moyens linguistiques pour solliciter la réaction du destinataire. La  
 structure est bien choisie, et la reprise lui permet d’insister sur l’essentie. …</p>  
 </div>  
 </body>

Figure 9 : Exemple de l’annotation de la structure interne du texte.

1. **Repérage et balisage des segments correspondant aux éléments susceptibles de biaiser les requêtes portant sur l’interlangue**

Certains segments de texte seront repérés manuellement, ils correspondant aux éléments susceptibles de biaiser les requêtes portant sur l’interlangue comme les citations, les traductions littérales ou les passages et exemples en langue étrangère.

**4.1 Citations**

L’annotation des citations se fera à l’aide de l’élément **<quote>.**

<p>Les proverbes sont des expressions figées, constituant « <hi rend="i">**<quote>**un tout  
 autonome**</quote>**</hi> » comme le souligne Greidanus dans son article intitulé « <hi rend="i">La  
 syntaxe du proverbe</hi> », qui se caractérisent par leur archaïsme, leur forme brève et  
 une certaine musicalité qui leur donne un caractère distingué et crée une harmonie. …</p>

Figure 10 : Exemple de l’annotation des citations.

**4.2 Passages en langue étrangère**

L’annotation des passages en langue étrangère se fera à l’aide de l’élément **<** **foreign >.**

<p>Dans l’Alphabet Phonétique International, les voyelles « rétractées » (« **<foreign>**back  
 vowels**</foreign>** » en anglais) sont représentées avec un signe diacritique droit <hi  
 rend="b">[ ̙ ]</hi> pour les distinguer des voyelles « avancées » (« **<foreign>**front  
 vowels**</foreign>** » en anglais) qui sont, quant à elles, représentées avec le signe  
 opposé, un diacritique gauche <hi rend="b">[ ̘ ]</hi>.</p>

Figure 11 : Exemple de l’annotation des passages en langue étrangère.

**4.3 Exemples**

L’annotation des exemples se fera à l’aide de l’élément **<** **eg >.**

<p>Il parait que l’ordre des mots dans les proverbes est parfois identique à l’ordre  
 « naturel » des constituants dans la phrase simple en français. En voici quelques proverbes  
 qui respectent parfaitement l’ordre syntaxique de base en français qui est Sujet – Verbe –  
 Complément :</p>  
 <p><**eg>**1. « Une hirondelle ne fait pas le printemps. »**</eg>**</p>  
 <p>**<eg>**2. « Les affaires sont les affaires. »**</eg>**</p>  
 <p>**<eg>**3. « On apprend à tout âge. »**</eg>**</p>  
 <p>**<eg>**4. « Les bons comptes font les bons amis. »**</eg>**</p>  
 <p>**<eg>**5. « Un malheur ne vient jamais seul. »**</eg>**</p>

Figure 12 : Exemple de l’annotation des exemples.

**4.4 Traductions littérales**

<seg type="traduction">…</seg>

1. **Insertion des en-têtes avec les métadonnées (conforme à la norme TEI)**

- établissement et responsables

- caractéristiques du document

Voici la liste des différents types de documents

|  |
| --- |
| autre |
| compte rendu de conférence |
| description de projet |
| dossier d'analyse |
| écrit de réflexion |
| essai argumentatif |
| fiche de lecture (un document) |
| projet de recherche |
| synthèse de lecture (plusieurs documents) |
| texte argumentatif |

- données sociobiographiques