

**Marc Van Campenhoudt**

Centre de recherche Termisti

Institut supérieur de traducteurs et interprètes

Haute École de Bruxelles

[mvc@termisti.be](mailto:mvc@termisti.be)

[www.termisti.refer.org](http://www.termisti.refer.org)

## **L'INITIATION AUX LANGAGES DE BALISAGE : UN APPRENTISSAGE DEVENU INCONTOURNABLE POUR MAÎTRISER LA TRADUCTION ASSISTÉE PAR ORDINATEUR**

Communication présentée dans le cadre du  
*VIII<sup>e</sup> Séminaire de traduction scientifique et technique*  
Lisbonne, 14 novembre 2005  
Institut franco-portugais,  
Union latine et  
Fundação para a ciência e a tecnologia

### **RÉSUMÉ**

Le langage de balisage XML (*Extensible Markup Language*) permet de rendre fidèlement compte du contenu et de la structure d'un document dans un format texte non propriétaire. Grâce au codage Unicode, il garantit une représentation universelle des caractères au sein de documents multilingues. Un fichier XML peut être produit immédiatement dans un simple éditeur de texte, à l'aide d'un éditeur spécifique ou via l'outil d'exportation de tout autre logiciel. Grâce au langage XSL, un fichier XML peut être converti aisément vers une multitude de formats informatiques. XML est donc l'outil privilégié d'échanges de données linguistiques entre ordinateurs.

Le langage XML a servi à élaborer de nombreuses normes d'échange dans des domaines qui touchent directement à l'ingénierie linguistique et, plus particulièrement, au monde de la traduction assistée par ordinateur. Dans cette communication, après un exposé des fondements de XML, l'auteur présente brièvement les principaux standards utilisables en bureautique, linguistique de corpus, terminologie, traduction assistée, localisation et sous-titrage. Il se focalise ensuite sur les nouvelles exigences de formation que pose l'usage généralisé de XML dans ces sphères d'activités. À travers l'expérience du DESS en traduction et industries de la langue de l'Institut supérieur de traducteurs et interprètes (Bruxelles), il examine successivement les objectifs fondamentaux à atteindre en termes de compétences professionnelles, les contenus à envisager et les moyens didactiques à mettre en œuvre.

**Mots clés** : XML, traduction assistée par ordinateur, formation.

## 1 XML, UN MÉTALANGAGE

### 1.1 Caractéristiques et avantages

Le métalangage XML (*Extensible Markup Language* ou *langage de balisage extensible*) est un standard du *World Wide Web Consortium* (W3C)<sup>1</sup>. Libre de tout droit, il offre un format universel non propriétaire et ne nécessite aucun logiciel dédié. Grâce au codage Unicode, il garantit une représentation universelle des caractères au sein de documents multilingues, ce qui explique son intérêt évident pour les industries de la langue.

Comme HTML (*HyperText Markup Language*), utilisé pour la mise en page des sites Internet, XML est un langage de balisage. Il permet d'identifier des éléments textuels en les isolant au moyen de balises ouvrantes et fermantes sur le modèle <balise>contenu</balise>. En dépit de nombreuses analogies, il convient toutefois de ne pas confondre ces deux langages. En effet, HTML sert surtout à préciser une mise en page et ne décrit guère le contenu et la structure des documents diffusés. En outre, les noms de balises HTML sont fixés et les tolérances syntaxiques sont nombreuses.

Nous parlons de XML comme d'un métalangage, car il permet de décrire très fidèlement un document structuré, à savoir de décrire l'ossature structurelle d'un document et de préciser chacune de ses catégories de contenu. Il est ainsi aisé d'y sélectionner des données particulières : les citations dans un texte, les didascalies dans une pièce de théâtre, les définitions dans un dictionnaire...

Un fichier d'extension \*.XML est sauvegardé en texte brut, ce qui garantit une pérennité de l'archivage<sup>2</sup> et une indépendance totale vis à vis de l'industrie logicielle. Outre que ceci facilite un échange entre systèmes informatiques concurrents, cela permet de produire des données de très grande qualité sans devoir investir dans une solution logicielle coûteuse, parce que dédiée à un marché professionnel restreint.

### 1.2 Un formalisme accessible

Codées dans un fichier texte, les données XML sont directement lisibles par toute personne alphabétisée. La logique des balises qui encadrent et structurent les informations est facilement accessible à un diplômé de lettres, confronté très tôt dans son cursus à la rigueur de la linguistique contemporaine. Par ailleurs, la notion de texte balisé est maîtrisée par un nombre croissant d'individus initiés à la rédaction de pages web balisées à l'aide du langage HTML.

Les règles d'écriture des balises XML sont peu nombreuses, mais très strictes. Divers gratuits, à commencer par les navigateurs pour Internet sont capables de vérifier si un fichier XML est, selon la terminologie, « bien formé », c'est-à-dire s'il respecte bien les règles d'écriture. Quoique non indispensables, des logiciels spécifiques, comme *XML Spy* et *Oxygen*, permettent d'aisément sélectionner les balises tout en se conformant aux règles (exemple 1).

---

1. [www.w3.org/XML](http://www.w3.org/XML).

2. Imprimé sur un papier de qualité, un document XML pourra encore être lu et scanné dans plusieurs siècles, chose qui est loin d'être garantie pour un support de type électronique.

```

<?xml version="1.0" encoding="iso-8859-1"?>
<!-- Travail pour le cours de Terminologie de M. Van Campenhoudt (DESS TIL, 2003-2004) -->
<!-- Arias, Laurent -->
<!-- Date de création : 2004-03-31 -->
<!DOCTYPE cesDoc SYSTEM "cesDoc.dtd">
<cesDoc version="4.3" type="text">
  <cesHeader version="2.0">
    <fileDesc>
      <titleStmnt>
        <h.title>Guide de l'utilisateur Adobe Audition 1.5 pour Windows</h.title>
      </titleStmnt>
      <publicationStmnt>
        <distributor>Adobe Systems Incorporated</distributor>
        <pubAddress>Adobe Systems Incorporated, 345 Park Avenue, San Jose, CA 95110-2704, États-Unis d'Amérique
      </pubAddress>
        <availability status="restricted">Droits réservés en vertu de la législation américaine sur les droits d'auteurs
      </availability>
      <pubDate value="ISO8601">2004</pubDate>
    </publicationStmnt>
    <sourceDesc>
      <biblStruct>
        <monogr>
          <h.title>Guide de l'utilisateur Adobe Audition 1.5 pour Windows</h.title>
          <h.author>Adobe Systems Incorporated</h.author>
          <imprint>
            <pubPlace>San Jose, Californie, États-Unis d'Amérique</pubPlace>
            <publisher>Adobe Systems Incorporated</publisher>
            <pubDate value="ISO8601">2004</pubDate>
          </imprint>
        </monogr>
      </biblStruct>
    </sourceDesc>
  </fileDesc>
  <profileDesc>
    <langUsage>
      <language id="fr" iso639="fr">Français</language>
      <language id="en" iso639="en">Anglais</language>
    </langUsage>
  </profileDesc>
</cesHeader>
</cesDoc>

```

### Exemple 1 : texte balisé au format XML (logiciel XML Spy)

Au-delà de l'écriture, des formalismes nommés *description du type de document (DTD)* ou *schéma XML* permettent de contraindre la structure d'un document XML, laquelle est toujours arborescente. Simples à maîtriser, ces formalismes garantissent tout à la fois une grande liberté de création des balises et une très grande rigueur de pensée (exemple 2). Ainsi, à l'usage, on s'aperçoit qu'un document XML permet une créativité beaucoup plus importante que nombre de structures de bases de données, souvent trop restrictives face à l'univers à décrire, voire mal pensées au vu de la complexité des relations à établir entre les catégories de données.

```

1 <?xml version="1.0" encoding="iso-8859-1"?>
2 <!ELEMENT lexique (concept_A10.14+)>
3 <!ELEMENT concept_A10.14 (domaine_A4, organisme_A10.2.2.10, auteur_A10.2.2.1, date_A.10.2.1.2, langue_A.10.7+)>
4 <!ELEMENT langue_A.10.7 (definitionGrp, termeGrp+)>
5 <!ELEMENT definitionGrp (definition_A.5.1, source_A10.19+)>
6 <!ELEMENT termeGrp (entree_A.1, grammaire_A2.2, contexteGrp+, note_usage_A.2.3.1?)>
7 <!ELEMENT contexteGrp (contexte_A.5.3, source_A10.19)>
8 <!ELEMENT domaine_A4 (#PCDATA)>
9 <!ELEMENT organisme_A10.2.2.10 (#PCDATA)>
10 <!ELEMENT auteur_A10.2.2.1 (#PCDATA)>
11 <!ELEMENT date_A.10.2.1.2 (#PCDATA)>
12 <!ELEMENT definition_A.5.1 (#PCDATA)>
13 <!ELEMENT entree_A.1 (#PCDATA)>
14 <!ELEMENT grammaire_A2.2 (#PCDATA)>
15 <!ELEMENT contexte_A.5.3 (#PCDATA)>
16 <!ELEMENT source_A10.19 (#PCDATA)>
17 <!ELEMENT note_usage_A.2.3.1 (#PCDATA)>
18 <!ATTLIST concept_A10.14
19     id ID #REQUIRED
20 >
21 <!ATTLIST langue_A.10.7
22     nom (CA | CS | DA | DE | EL | EN | ES | ET | FI | FR | GA | HU | IT | LA | LT | LV | MT | NL | NO | PL | PT | RO | RU | SK | SL | SV | TR | UK) #RE
23 >
24 <!ATTLIST domaine_A4
25     lenoch (AD | AG | AR | AS | AT | AU | BA | BZ | CE | CH | CO | DE | DI | DO | EC | ED | EL | EN | FI | GO | HI | IC | IN | JU | LA | MA | ME | MG | M
26 >
27 <!ATTLIST organisme_A10.2.2.10
28     nom (ISTI_DESS) #REQUIRED
29 >
30 <!ATTLIST date_A.10.2.1.2
31     norme (ISO8601) #REQUIRED
32 >
33

```

**Exemple 2** : DTD décrivant un modèle de données terminologique (logiciel Oxygen)

### 1.3 Le mécanisme des feuilles de style

Le langage XSLT (*eXtended Stylesheet Language Transformations*) est une recommandation du consortium W3C<sup>3</sup> qui permet de convertir un fichier XML vers un autre fichier XML ou vers bien d'autres formats, comme, par exemple, HTML, RTF, *OpenDocument* ou encore PDF. Le mécanisme de transformation est particulièrement puissant sur la toile : une feuille de style (fichier d'extension \*.XSL) attachée à un fichier XML permet à un navigateur récent de présenter automatiquement ledit fichier XML au format HTML.

On perçoit ainsi qu'un fichier XML ne doit donc pas être considéré comme un simple fichier intermédiaire conçu uniquement pour procéder à des échanges. XML est tout destiné pour servir de format source original, parfaitement documenté, les formats dotés de moins de potentiel descriptif n'apparaissant plus que comme des « sorties » liées à des applications logicielles particulières, moins universelles.

3. [www.w3.org/TR/xslt](http://www.w3.org/TR/xslt).

```

1 <?xml version="1.0" encoding="iso-8859-1"?>
2 <!--
3     Auteur: Marc Van Campenhoudt
4     Fichier: dessisti2html2005.xsl
5     Date: 29/12/2005
6     But : convertit en HTML les fichiers XML conformes à dessisti2005.dtd
7 -->
8 <xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
9   <xsl:output method="html" indent="yes" encoding="iso-8859-1"/>
10  <xsl:template match="/lexique">
11    <html>
12      <head>
13        <title>DESS en traduction et industries de la langue : travail d'examen</title>
14        <meta name="description" content="fiche terminologique conçue dans le cadre du DESS
15          en traduction et industries de la langue de l'ISTI (dessisti2005.dtd). Fiche
16          transformée de XML en HTML à l'aide de la feuille de style
17          dessisti2html2005.xsl"/>
18        <meta name="keywords" content="terminologie ISTI DESS XML HTML"/>
19        <meta name="author" content="Centre de recherche Termisti"/>
20        <meta name="generator" content="transformation XSL"/>
21      </head>
22      <body>
23        <center>
24          <h2>Questions approfondies de terminologie et de terminotique</h2>
25          <h3>Travail d'examen</h3>
26        </center>
27        <xsl:apply-templates select="concept_A10.14"/>
28      </body>
29    </html>
30  </xsl:template>
31  <xsl:template match="concept_A10.14">
32    <hr/>
33    <table border="0" cellpadding="4" cellspacing="4">
34      <tr valign="top" bgcolor="#FFFF99">
35        <+>

```

**Exemple 3** : feuille de style convertissant une fiche terminologique au format HTML (logiciel *Oxygen*)

## 2 LES APPLICATIONS XML DANS LE DOMAINE DE LA TRADUCTION ASSISTÉE

L'univers XML est devenu un élément incontournable de la traduction assistée par ordinateur (TAO). Dans les pages qui suivent, nous examinerons son rôle dans différentes matières : la bureautique, les corpus de textes, les bases de données lexicales et terminologiques, les mémoires de traduction, la localisation et le sous-titrage.

### 2.1 La bureautique

La version 2.0 de la suite bureautique *OpenOffice*, sortie en 2005, utilise désormais le format ODF (*Open Document Format for Office Applications*). Il s'agit d'un format XML adopté comme standard ouvert par Oasis, l'*Organization for the Advancement of Structured Information Standards*<sup>4</sup>. Ce consortium regroupe des acteurs majeurs de l'univers informatique non inféodé à la société Microsoft.

<sup>4</sup> www.oasis-open.org.

En 2004, ODF, qui est utilisé par nombre de logiciels propres à l'univers d'Unix et de Linux, tels *Star Office* (Sun), *KOffice* (KDE), *AbiWord* (Gnome), a été évalué positivement par la Commission européenne comme base d'un format d'échange non lié à une technologie particulière. À l'invitation du même organisme, il a été soumis à l'Iso, pour approbation comme norme internationale (Iso/IEC Dis 26 300 2005).

Microsoft n'est pas en reste, puisque le format de base pour *Microsoft Office 12*, annoncé pour 2006, utilisera *Microsoft Office Open XML*, un format de base annoncé comme gratuit, libre de droit et documenté.

Toutes ces informations tendent à confirmer que XML est désormais appelé à devenir un format natif pour nombre d'applications logicielles, ce qui ne sera pas sans avantages pour leurs utilisateurs.

## 2.2 Les corpus textuels

Depuis les années quatre-vingts, un mouvement de fond a conduit les organismes soucieux de la conservation du patrimoine écrit et les institutions de recherche exploitant de larges données textuelles à définir des formats d'échange balisés permettant d'identifier la structure et le contenu des documents. Ce souci a donné naissance à la *TEI (Text Encoding Initiative)*<sup>5</sup>, une norme de balisage d'abord conçue en SGML<sup>6</sup> et qui a ensuite évolué vers XML. Plutôt « littéraire », au sens large, cette norme couvre un large spectre d'applications de l'écrit et propose des passerelles intéressantes avec la suite *OpenOffice*.

Plus spécifique à l'ingénierie linguistique, mais largement compatible avec la TEI, la norme *XCES (Corpus Encoding Standard for XML)*<sup>7</sup> constitue une évolution de la norme SGML de *EAGLES (Expert Advisory Group on Language Engineering Standards)*, un consortium financé par la Commission européenne (DGXIII) dans les années quatre-vingt-dix.

## 2.3 Bases de données lexicales et terminologiques

La norme Iso 16 642 (2003) *TMF (Terminological Markup Framework)*<sup>8</sup> adopte l'approche conceptuelle monosémique prônée par la théorie terminologique classique. Elle précise un métamodèle de structure et de balisage pour tout format d'échange de données, dont elle est en quelque sorte la matrice. De ce fait, TMF peut constituer un format d'échange à part entière dès lors que l'on se fonde sur les catégories de données énumérées dans la norme Iso 12 620 (1999).

Le format *TBX (TermBase eXchange)*<sup>9</sup> est un bon exemple de standard conforme à cette norme Iso 12 620 TMF. Évolution du projet européen *Salt*<sup>10</sup>, auquel était associé notre centre de recherche, il est proposé par le groupe *Open Standards for Container/Content Allowing Re-use (Oscar)* de la *Localization Industry Standards Association (LISA)*.

---

5. [www.tei-c.org](http://www.tei-c.org).

6. SGML (*Standard Generalized Markup Language*) est l'ancêtre de XML.

7. [www.cs.vassar.edu/XCES](http://www.cs.vassar.edu/XCES).

8. [www.loria.fr/projets/TMF](http://www.loria.fr/projets/TMF).

9. [www.lisa.org/standards/tbx](http://www.lisa.org/standards/tbx).

10. [www.loria.fr/projets/SALT/](http://www.loria.fr/projets/SALT/).

Davantage orienté vers le traitement automatique des langues naturelles, particulièrement la traduction automatique, le format Olif (*Open Lexicon Interchange Format*)<sup>11</sup> est une évolution du projet européen Otelo (*Open Translation Environment for Localization*). Il permet l'échange de ressources lexicales et terminologiques selon une approche ouverte à la polysémie.

## 2.4 Mémoires de traduction

Les mémoires de traduction sont des bases de données constituées d'alignements de paires de segments de textes en langues source et cible identifiés comme équivalents (exemple 4). Les logiciels de traduction assistée par ordinateur (TAO) les exploitent pour repérer les passages qui ont déjà été rencontrés précédemment dans la langue source et en dupliquer la traduction dans la langue cible.

Source (English)	Target (French)
PART 1	Chapitre 1
FUNDAMENTAL RULES	Règles fondamentales
1 SAFETY	1 Sécurité
1.1 Helping Those in Danger	1.1 Aider ceux qui sont en danger
A boat or competitor shall give all possible help to any person or vessel in danger.	Un voilier ou un concurrent doit apporter toute l'aide possible à toute personne ou navire en danger.
1.2 Life-saving Equipment and Personal Buoyancy	1.2 Equipement de sauvetage et flottabilité personnelle
A boat shall carry adequate life-saving equipment for all persons on board, including one item ready for immediate use, unless her class rules make some other provision.	Un voilier doit avoir à bord un équipement de sauvetage approprié pour toutes les personnes embarquées, y compris un dispositif prêt pour usage immédiat, à moins que ses règles de classe ne prévoient quelque autre disposition.
Each competitor is individually responsible for wearing personal buoyancy adequate for the conditions.	Chaque concurrent est personnellement responsable du port d'un système de flottabilité personnelle approprié aux conditions.
2 FAIR SAILING	2 Navigation loyale
A boat and her owner shall compete in compliance with recognized principles of sportsmanship and fair play.	Un voilier et son propriétaire doivent concourir selon les principes reconnus de sportivité et de jeu loyal.
A boat may be penalized under this rule only if it is clearly established that these principles have been violated.	Un voilier peut être pénalisé selon cette règle seulement s'il est clairement établi que ces principes ont été bafoués.
3 ACCEPTANCE OF THE RULES	3 Acceptation des règles
By participating in a race conducted under these racing rules, each competitor and boat owner agrees	En participant à une course régie selon les présentes règles de course, chaque concurrent et propriétaire de voilier accepte

Exemple 4 : segments alignés anglais-français (logiciel WinAlign)

Proposé par le groupe Oscar (Lisa), le format ouvert TMX (*Translation Memory Exchange*)<sup>12</sup> permet de stocker et d'échanger une mémoire de traduction quel que soit le logiciel de TAO utilisé. Des logiciels comme *Trados 7*, *SDLX 2004*, *WorldServer* ou *Ambassador* sont certifiés conformes à TMX par l'association Lisa. Une norme complémentaire, SRX (*Segmentation Rules eXchange*)<sup>13</sup>, est également proposée, avec pour objectif de spécifier davantage les règles de segmentation qui ont servi à créer une mémoire de traduction balisée selon TMX.

11. [www.olif.net](http://www.olif.net).

12. [www.lisa.org/standards/tmx](http://www.lisa.org/standards/tmx).

13. [www.lisa.org/standards/srx](http://www.lisa.org/standards/srx).

Utilisable en localisation, TMX est aujourd'hui concurrencé par XLIFF (*XML Localization Interchange File Format*)<sup>14</sup>, le successeur du format *OpenTag*. Proposé par le consortium Oasis (cf. 2.1), il permet de localiser directement un produit logiciel sans avoir à se soucier du format natif. XLIFF distingue, en effet, le texte à traduire des lignes de programmation et inclut des métadonnées sur la version du travail de localisation.

On notera finalement que le groupe Oscar a en chantier une norme nommée GMX (*Gilt Metrics eXchange*)<sup>15</sup> qui a pour ambition de permettre de quantifier un projet de traduction de manière standardisée. Ce projet de norme devrait envisager le comptage des mots et caractères, la complexité de la tâche et la qualité finale requise et pourrait s'appliquer à des fichiers XLIFF.

## 2.5 Sous-titrage

Une recommandation du W3C, Smil (*Synchronized Multimedia Integration Language*)<sup>16</sup>, permet de synchroniser du son, de l'image et du texte sur une page web. Comme l'a montré Laurent Arias (2004), ce langage permet de sous-titrer un document vidéo dans plusieurs langues pour un public allophone, de retranscrire le contenu d'un document sonore pour des malentendants ou encore d'adjoindre du son à des images ou à des textes en vue de les décrire aux malvoyants. Smil a débouché sur la norme *Timed Text*<sup>17</sup>, destinée à préciser un standard de minutage indépendant des logiciels.

## 3 XML DANS LA FORMATION DU TRADUCTEUR SPÉCIALISÉ EN TAO

### 3.1 Le cas du DESS en traduction et industries de la langue

Avec plusieurs collègues de l'Institut supérieur de traducteurs et interprètes (Isti, Bruxelles), nous avons lancé en 1998 un diplôme d'études supérieures spécialisées (DESS) en traduction et industries de la langue<sup>18</sup> qui avait pour ambition de former des traducteurs diplômés aux méthodes de la TAO dans un large spectre d'activités traductives. Quelque sept années plus tard, nous observons que, comme nous l'avions pressenti, l'informatique réunit désormais toutes les activités auxquelles nous entendions initier nos étudiants : sous-titrage<sup>19</sup> et surtitrage, localisation de logiciels, localisation de sites internet, traductique et terminologie. Il n'est donc pas étonnant de constater l'émergence de formats XML adaptés à chacune de ces activités.

---

14. [www.xliff.org](http://www.xliff.org).

15. [www.lisa.org/standards/gmx](http://www.lisa.org/standards/gmx).

16. [www.w3.org/AudioVideo](http://www.w3.org/AudioVideo).

17. [www.w3.org/AudioVideo/TT](http://www.w3.org/AudioVideo/TT).

18. [www.heb.be/isti/desstil.htm](http://www.heb.be/isti/desstil.htm).

19. Observons que l'un des tout premiers logiciels de sous-titrage sur PC a été créé à l'initiative de notre collègue Corinne Imhauser, enseignante au DESS.



### 3.2 Viser des capacités précises, au-delà de la simple information

Dès la création de ce programme de spécialisation, ses concepteurs avaient souligné l'importance primordiale d'une formation à la problématique du balisage des documents structurés. Celle-ci se déroule dans le cadre d'un cours spécifique, intitulé *Conception et structuration de documents électroniques*. Les étudiants y sont systématiquement formés à XML, ce qui leur permet d'envisager ensuite les différents standards propres aux industries de la langue dans le cadre des cours spécialisés : traductique (TEI, XCES, TMX), localisation (ODF, TMX, XLIFF, SRX, GMX), terminologie (TEI, XCES, Olif, TMF, TBX) et sous-titrage (Smil, TT).

L'objectif de formation est que le diplômé soit à même de :

- comprendre un document XML ;
- rédiger une DTD ;
- vérifier si un document est bien formé et s'il est valide ;
- opérer une transformation XSL ;
- modifier une feuille de style XSL produisant du HTML ou du XML ;
- choisir le standard XML le plus adéquat ;
- vérifier si un logiciel de TAO exporte un fichier XML conforme au standard retenu.

### 3.3 Un exemple de progression suivie : le cours de terminologie

Le langage XML est devenu absolument incontournable en terminologie et permet des réalisations de très grande ampleur, comme nous l'avons montré très tôt dans le cadre du projet européen Dhydro<sup>20</sup> (Descotte *et al.* 2001, Van Campenhoudt 2002). Titulaire du cours intitulé *Questions approfondies de terminologie et de terminotique*, nous proposons à nos étudiants de DESS un parcours supervisé qui correspond aux différentes étapes d'un travail de terminographie assistée par ordinateur et nécessite la maîtrise de XML.

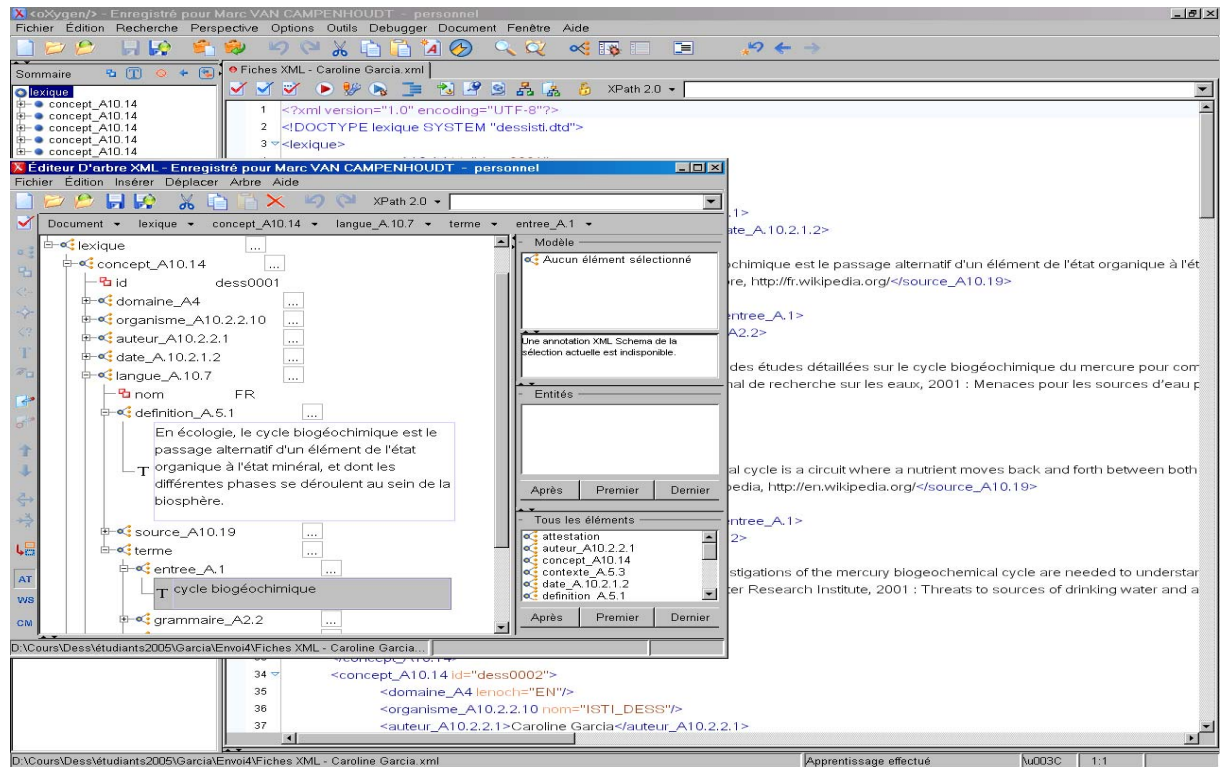
Les étudiants doivent tout d'abord réunir un vaste corpus textuel spécialisé dans deux langues et le baliser conformément à la norme XCES (*cf.* exemple 1). Ils doivent ensuite y opérer une recherche de candidats-termes à l'aide de concordanciers et rédiger des fiches d'attestation. Cette recherche initiale, pratiquée dans une perspective de description de l'usage, est ensuite suivie de la rédaction de véritables fiches terminologiques conformes à une DTD spécifique. Cette DTD, reproduite dans l'exemple 2, répond aux exigences structurelles de la norme Iso 16 642 et utilise les catégories de données de la norme Iso 12 620. Elle rend obligatoire l'usage des champs « définition », « contexte » et « source », trop souvent négligés dans les terminographies contemporaines, y compris dans certaines grandes bases de données de référence.

Les fiches terminologiques proprement dites doivent être rédigées immédiatement en XML. Cette contrainte s'explique par notre souhait de voir nos étudiants mener une véritable activité de balisage qui les forme à la rigueur de XML. Les logiciels utilisés (*XML Spy* ou *Oxygen*) leur permettent de vérifier si leurs fichiers sont bien formés (conformes à la syntaxe XML) et s'ils sont valides (conformes à la DTD). L'évolution des interfaces rend ces outils – d'un coût modique – très agréables à utiliser. Par exemple, *Oxygen* propose un éditeur d'arbre qui autorise l'encodage des données dans des graphes arborescents très rigoureux (exemple 5). Ceci permet d'éviter de commettre certaines

---

20. [www.loria.fr/projets/MLIS/DHYDRO](http://www.loria.fr/projets/MLIS/DHYDRO).

aberrations structurelles paradoxalement autorisées par de célèbres logiciels dédiés à la terminographie.



### Exemple 5 : encodage dans une structure arborescente à l'aide du logiciel Oxygen

La dernière étape est celle de la transformation des fiches XML au format HTML (exemple 6) à l'aide d'une feuille de style XSL (présentée partiellement dans l'exemple 3). Dans le cadre du cours nous ne demandons pas à l'étudiant de concevoir lui-même une telle feuille de style, mais d'en comprendre le contenu et de pouvoir y modifier, si nécessaire, les balises HTML.

La procédure suivie ne doit pas être perçue comme un refus de former nos étudiants aux logiciels de gestion terminologique, déjà enseignés au deuxième cycle. En effet, nous ne souhaitons pas remettre en cause la commercialisation de ces outils professionnels, nécessaires pour traiter un grand flux de données dans le cadre d'un format de fiche particulier. À nos yeux, la pratique intensive de XML doit conduire l'étudiant de DESS à expérimenter de nombreux modèles de données et à se poser la question de leur pertinence. Il ne fait aucun doute que cette compétence lui permettra de regarder d'un œil critique les solutions commerciales qui lui seront proposées ultérieurement dans un cadre professionnel, sachant qu'elles ne peuvent toutes convenir à des besoins particuliers.

Exemple de fiche terminologique convertie en HTML depuis XML - Mozilla Firefox

Fichier Edition Affichage Aller à Marque-pages Outils ?

http://www.termisti.refer.org/exemplefiche.htm

<b>concept n°</b> : dess0001	<b>organisme</b> : ISTI_DESS	<b>date</b> : 2005.01.27
<b>domaine</b> : TR	<b>auteur</b> : E. Xemple	

**FR**

**définition** : Alarme sonore et visuelle indiquant que le navire dérive au-delà du rayon programmé autour de la position de mouillage du bateau.

source : BAYO IMPORT, 1992a : GPS 50-Manuel d'utilisation, France : Bayo Import, p. 44  
source : RAYTHEON MARINE COMPANY, 1989 : Raystar 920 GPS Navigator-Instruction Manual, Japan : Mamtech Co., Ltd., p. 55

**entrée** : alarme de mouillage

**grammaire** : s.f.

**note d'usage** : terme le plus utilisé

**contexte** : Alarme de mouillage : Cette alarme permet de surveiller la position du bateau au mouillage. Si le bateau dérive au-delà de la distance programmée, une alarme sera mise en indiquant un dérapage possible de l'ancre.

source : E.T.N.A. SE, sd-b : RAYSTAR 590 GPS NAVIGATOR, Octeville Sur Mer : E.T.N.A. , p. 25

**contexte** : Une alarme de mouillage est indispensable si l'on ne compte pas assurer une veille en permanence à bord d'un yacht ancré momentanément sur un fond délicat .

source : GRÉE (A.), 1996 : Traité de navigation, Paris, Gallimard, p. 19

**entrée** : alarme de dérapage

**grammaire** : s.f.

**note d'usage** :

**contexte** : Quand les conditions d'alarme sont détectées, ces symboles apparaissent à l'écran, [quelle que] soit la fonction sur laquelle est votre appareil. Quand les différents signaux sont allumés, la signification des indicateurs est la suivante : ANC - Indique que l'alarme de veille mouillage a été activée.

source : E.T.N.A. SE, sd-b : RAYSTAR 590 GPS NAVIGATOR, Octeville Sur Mer : E.T.N.A. , p. 40

**EN**

**définition** : An audible and visual alarm giving notice of the drifting of the vessel outside the preset anchor alarm radius. The alarm range is preset as the radius of circle centered at the anchoring position

source : A.M.R.C. N.V., sd : PRELIMINARY GXL 1100 GPS-Instruction Manual, UK : Raytheon Marine Sales & Service Company, p. 43  
source : C.I.M.A., 1991 : GPS 50 Personal Navigator-Owner's Manual, Taiwan : Garmin, p. A.1  
source : KODEN ELECTRONICS CO., LTD., 1992 : Service Manual-GPS Navigator KGP-930, Japan : Kodan Electronics Co., Ltd., p. 3-13

**entrée** : anchor watch alarm

Terminé

### Exemple 6 : fiche terminologique convertie au format HTML

#### EN GUISE DE CONCLUSION

Il importe de rappeler que XML n'est qu'un outil de représentation des données : il est indispensable d'être correctement formé aux méthodes de la traduction pour bien traduire un document structuré balisé en XML. Il reste que les grands bureaux de traduction demandent de plus en plus à leurs sous-traitants - traducteurs indépendants - de disposer d'outils logiciels d'aide à la traduction aptes à traiter des données linguistiques au format XML. Par ailleurs, nous constatons qu'un véritable débouché existe dans ces grands bureaux pour les diplômés de notre DESS, capables de tutoyer les formats informatiques et de coordonner de vastes chantiers de traduction ou de localisation. Une véritable filière de recrutement voit ainsi le jour, nos plus anciens diplômés devenant dans ces entreprises, les maîtres de stage de nos nouveaux étudiants.

Certes, il restera toujours des segments du marché de la traduction au sein desquels il sera possible de travailler sans avoir à se soucier de XML. Mais à l'heure de la globalisation et de la compression des coûts, il importe de souligner que de réelles possibilités de carrière demeurent ouvertes au traducteur européen qui accepte de s'initier aux formalismes de l'univers de la linguistique computationnelle et du génie logiciel.

## BIBLIOGRAPHIE

Les notes du texte renvoient directement aux sites Internet des normes XML citées. Ces sites ont nourri la présente communication.

Arias (L.), 2004 : *Le langage SMIL et la localisation de vidéos en flux continu*, mémoire de DESS, ISTI : Bruxelles.

Descotte (S.), Husson (J.-L.), Romary (L.), Van Campenhoudt (M.), Viscogliosi (N.), 2001 : « Specialized lexicography by means of a conceptual data base: establishing the format for a multilingual marine dictionary », dans Vainio (J.), éd., *Maritime Terminology: Dictionaries and Education, Proceedings of the Second Conference on Maritime Terminology, 11-12 May 2000, Turku, Finland*, Turku : University of Turku, p. 63-81 (Publications from the Centre for Maritime Studies, University of Turku, A36).

Iso 12 620, 1999 : *Aides informatiques en terminologie – Catégories de données*, Genève : Organisation internationale de normalisation.

Iso 16 642, 2003 : *Computers Applications in Terminology - Terminological Markup Framework*, Genève, Organisation internationale de normalisation, [www.loria.fr/projets/TMF](http://www.loria.fr/projets/TMF).

Iso/IEC Dis 26 300, 2005 : *Format de document ouvert pour applications de bureau (document ouvert)*, v1.0, Genève : Organisation internationale de normalisation.

Popescu-Belis (A.), 2005 : *L'impact d'XML pour les technologies multilingues Cours XML pour le DESS en TAO et le DEA en TIM, Semestre d'été 2005*, [www.issco.unige.ch/staff/andrei/xml/index.html](http://www.issco.unige.ch/staff/andrei/xml/index.html), dernière mise à jour : juin 2005.

Van Campenhoudt (M.), 2002 : « Lexicographie vs terminographie : quelques implications théoriques du projet Dhydro », communication dans le cadre du séminaire « Terminologie et nouvelles technologies de l'information et de la communication » (1<sup>er</sup> décembre 2000), dans Zinglé (H.) dir., *Travaux du Lilla*, Université de Nice-Sophia Antipolis, 2002, n° 4, p. 91-103.

---