

# Logiciel iMAG et relais-iMAG

Projet Traouiéro, document L3.1

Valérie BELLYNCK, Achille FALAISE, Christian BOITET, Lingxiao WANG

Ce document est le rapport de synthèse de la sous-tâche 3.1. Il est joint au rapport de synthèse L234.4 des tâches 2 à 4 à T0+24 (15/1/13).

L'objectif de cette tâche est de définir, concevoir et réaliser les logiciels nécessaires pour fournir le service iMAG dans le cadre d'une exploitation commerciale, où différents logiciels doivent partager la gestion des utilisateurs.

## Contenu

<b>RÉSUMÉ.....</b>	<b>3</b>
<b>1. INTRODUCTION.....</b>	<b>3</b>
1.1 Description initiale (dans la proposition) .....	3
1.2 Contexte .....	4
1.3 Evolution de l'objectif.....	5
1.4 But et organisation du document.....	6
<b>2. ORGANISATION DU TRAVAIL.....</b>	<b>7</b>
2.1 Décomposition en phases du volet Llogiciel iMAG.....	7
2.2 Décomposition en phases du volet Relais .....	7
2.3 Décomposition en phase du volet Contributions.....	8
2.4 L'équipe.....	8
2.5 Méthode de travail.....	8
<b>3. LOGICIEL IMAG.....</b>	<b>9</b>
3.1 Traitements linguistiques et structurels d'une page Web.....	9
3.1.1 Cahier des charges.....	9
3.1.2 Code de SECTra3.....	10
3.1.3 Adaptation à SECTra2.....	11
3.1.4 Extension de la couverture de fonctions de SECTra2 dans SECTra3.....	11
3.1.5 Perspectives sur l'export/import entre structures de BD de corpus.....	12
3.2 Externalisation de l'affichage du bandeau et de la palette.....	14
3.2.1 Transfert du bandeau.....	14
3.2.2 Transfert de la palette.....	17
3.2.3 iMAG transparentes avec Proximag.....	19
3.3 Externalisation du dorsal de traduction, SECTra-Edit.....	21
3.3.1 Cahier des charges et spécifications externes.....	21
3.3.2 Spécifications internes.....	22
3.3.3 Calendrier prévisionnel.....	22
3.4 Spécification de l'interacteur pour l'aide lexicale.....	22
3.4.1 Cahier des charges.....	22
3.4.2 Spécifications internes.....	23
3.4.3 Calendrier prévisionnel.....	25
<b>4. RELAIS IMAG.....</b>	<b>26</b>
4.1 Cahier des charges .....	26
4.2 Spécifications externes.....	27
4.3 Spécification interne.....	27
4.3.1 Architecture générale.....	28
4.3.2 Scénarios.....	29
4.3.3 Correspondance des droits des utilisateurs entre services .....	30

<b>5. DORSAL DE MODÉRATION.....</b>	<b>31</b>
5.1 Cahier de charges.....	31
5.2 Considérations techniques.....	33
5.2.1 Qu'est-ce que la programmation par aspects ?.....	33
5.2.2 Etude des parties des bases de données pour gérer les attributions.....	34
5.2.3 Délégation et asynchronisme.....	36
5.3 Spécifications internes.....	36
5.3.1 Base de données de SECTra2.....	36
5.3.2 Base de données de SECTra3.....	37
<b>6. ETAT D'AVANCEMENT.....</b>	<b>40</b>
<b>7. RÉFÉRENCES BIBLIOGRAPHIQUES.....</b>	<b>41</b>
7.1 Références bibliographiques.....	41
7.2 Références wWebographiques.....	41
<b>8. ANNEXES .....</b>	<b>42</b>
8.1 Code de la segmentation adaptée à SECTra3.....	42
8.2 Code du ré-habillage adaptée à SECTra3.....	43
8.3 Extrait de la classe Groovy.Translation qui génère le bandeau .....	44
8.4 Code velocity de l'API « à la » REST pour la liste des langues cibles d'une iMAG.....	47
8.5 Code javascript exploitant l'API qui fournit la liste des langues.....	49
8.6 Code Groovy qui adapte les interacteurs du résultat de Google.....	49

# Logiciel iMAG et relais-iMAG

Projet Traouiéro, document L3.1

Valérie BELLYNCK, Achille FALAISE, Christian BOITET, Lingxiao WANG

## RÉSUMÉ

Ce document est le rapport de synthèse de la sous-tâche 3.1. Il est joint au rapport de synthèse L234.4 des tâches 2 à 4 à T0+24 (15/1/13).

L'objectif de cette tâche est de définir, concevoir et réaliser les logiciels nécessaires pour fournir le service iMAG dans le cadre d'une exploitation commerciale, où différents logiciels doivent partager la gestion des utilisateurs.

La sous-tâche ST3.1 participe à la tâche dédiée à la modularisation des logiciels exploitables dans des services comme les passerelles iMAG, tout en préservant les droits des utilisateurs dans les différents composants. Initialement, cette sous-tâche consistait en la consolidation des logiciels spécifiques des passerelles iMAG (passerelle, bandeau et palette de contribution) et au développement du module relayant les droits entre les différents services utilisés. Une contribution supplémentaire lui a été affectée : la conception du module qui est en charge de la gestion de l'attribution des contributions et l'organisation de la modération.

Ce document est le seul document concernant cette sous-tâche, et constitue son document de synthèse. Il décrit le cahier des charges du logiciel iMAG, du relais-iMAG et de l'outillage de la collaboration, leurs spécifications, l'avancement de l'implémentation, et la suite prévue.

## 1. INTRODUCTION

### 1.1 DESCRIPTION INITIALE (DANS LA PROPOSITION)

Dans le dossier scientifique joint à la candidature du projet ayant abouti au financement du projet Traouiéro, la sous-tâche ST3.1 participant à la tâche 3 (iMAG++), était décrite comme la sous-tâche dédiée à l'opérationnalisation des outils pour la multilinguisation collaborative de pages Web et de documents. Elle était présentée comme suit dans la proposition de projet.

T3 : iMAG++ : Outils pour la multilinguisation collaborative de pages Web et de documents	
ST3.1 : t0-t6 — Logiciel-iMAG, relais-iMAG	
Objectifs	Modulariser l'organisation des passerelles iMAG
Critères d'évaluation	<ul style="list-style-type: none"> <li>▪ Adéquation au CCH (cahier des charges) et au DSE (document de spécifications externes).</li> <li>▪ Fiabilité et robustesse en usage.</li> </ul>
Responsable	GETALP
Partenaires	GETALP, Floralis (mise au courant pour valorisation)
Activité	<p>Le logiciel spécifique des passerelles iMAG est actuellement une extension du logiciel SECTra_w construit au-dessus de XWiki par CP. Huynh. On l'appelle SECTra_w/iMAG. SECTra_w/iMAG est en évolution, HT. Nguyen ayant repris et étendu ce code pour permettre la modérabilité des contributions des post-éditeurs, dans le cadre de la collaboration mentionnée plus haut avec la Métro pour l'Expo de Shanghai.</p> <p>La partie iMAG du logiciel récupère une page Web d'un site S, la fait segmenter (actuellement par Google), et l'envoie à SECTra_w, avec la segmentation obtenue. SECTra_w recherche les segments dans la mémoire de traductions dédiée à S (MT-S), appelle un ou plusieurs systèmes de TA sur</p>

	<p>les segments nouveaux ou pas encore traduits, stocke tous les segments et leurs prétraductions automatiques, et renvoie à la passerelle iMAG-S une page Web reconstruite à l'aide des meilleures prétraductions ou post-éditions disponibles dans la MT-S.</p> <p>Dans cette page, les url liées au site S sont modifiées pour que la navigation continue via l'iMAG-S dans la langue d'accès choisie, et les segments sont remplacés par des bisegments (langue source, langue d'accès) augmentés de code javascript construisant les <i>bulles d'interaction</i> qui permettent aux visiteurs de modifier interactivement la traduction d'un segment apparaissant sur la page, et/ou de passer "sans couture" sous SECTra_w pour post-éditer un ensemble de segments.</p> <ul style="list-style-type: none"> <li>▪ Il faut maintenant programmer le logiciel iMAG de façon autonome, et le modulariser le plus possible, de façon à utiliser SECTra_w comme un <i>dorsal</i> indépendant. C'est en général une bonne pratique en génie logiciel, et ici elle est rendue nécessaire par le fait qu'il faut pouvoir séparer la gestion des utilisateurs au niveau d'une passerelle iMAG-S et au niveau de l'instance de SECTra_w utilisée comme dorsal de cette passerelle.</li> <li>▪ Il faut aussi créer un <i>relais-iMAG</i>, qui, en fonction des informations sur le site S et sur les paramètres de l'accès multilingue à S sur les langues (lesquelles sont post-éditables, modérables, post-éditées, modérées, ou exclues...) et sur les systèmes de TA (lesquels sont souhaités, avec quels paramètres, avec quelles préférences entre systèmes pour chaque couple de langues, avec quels <i>diagrammes de traduction</i> pour chaque couple aussi...), appelle l'iMAG-S avec les paramètres adéquats.</li> </ul>
Livrables	<ul style="list-style-type: none"> <li>▪ Sources des programmes et documentation (sur la forge du LIG).</li> <li>▪ Site d'essai (et de comparaison avec Ariane-G5).</li> </ul>
Gestion du risque	<p>Pas de risque technique en soi. Par contre, il y a toujours le risque d'avoir sous-estimé le temps ou la compétence nécessaires pour réaliser cette sous-tâche. Les conséquences seraient que les tests seraient partiels, et il faudrait les terminer avant de certifier le bon achèvement des travaux.</p>

## 1.2 CONTEXTE

Le développement du premier logiciel iMAG fonctionnel a été fait par HUYNH Cong Phap en novembre 2008, avec une limite de temps très contraignante (3 semaines, démo pour le LIG à la fin). Il a été réalisé par extension de SECTra\_w, le système d'exploitation de corpus de traductions qu'il avait déjà développé en Xwiki. Un des intérêts visés par le choix de cette architecture toute intégrée basée sur l'exploitation d'un Wiki, était la perspective de l'intégration de la gestion des contributions et de la modération par le logiciel support.

Dans ce sens, Xwiki présente l'intérêt de pouvoir attribuer non seulement des fonctions à des rôles d'utilisateurs, mais aussi de spécifier à quels groupes d'utilisateurs ces droits peuvent s'appliquer. Dans les faits, les contenus qui ont été partagés par les codes de Xwiki sont ceux qui permettent de construire collaborativement une plateforme Web intégrant des services qui peuvent s'appuyer sur les bases de données : cela permet l'écriture collaborative des codes qui réalisent ses fonctions, mais pas de gérer la collaboration sur la construction des éléments que ces services permettent de manipuler (comme, dans le cas de SECTra, les segments des corpus de traduction). De plus, le mélange des concepts métier (gestion de corpus pour la traduction automatique) et concepts techniques (gestion de l'édition collaborative), rend très difficiles la prise en main et la maintenance par d'autres développeurs, même si un gros effort de description avait été fait pour décrire la conception

et les spécifications externes et internes dans le mémoire de doctorat de HUYNH Cong Phap.

La première version de SECTra avait été développée en septembre 2007 pour servir de support à des campagnes d'évaluation de systèmes de TA, dans le cadre du contrat TRANSAT avec FT R&D. La seconde version avait déjà été réalisée comme une extension ajoutant les fonctionnalités liées à la post-édition contributive de documents (pages Web statiques en l'occurrence), dans le cadre du contrat EOLSS de l'UNDL (25 articles d'une encyclopédie). Cette version a été nommée SECTra\_w pour que le concepteur garde bien à l'esprit que l'outil qu'il développait devait être générique. L'intégration du code des passerelles iMAG dans le code de SECTra\_w, comme preuve de concept pour montrer une possibilité d'exploitation de SECTra, a été un choix dirigé par l'urgence : une seule authentification des utilisateurs est nécessaire pour suivre et leur attribuer leurs contributions. Cependant ce choix ne résolvait pas les problèmes liés à la gestion des utilisateurs pour le jumelage du service iMAG avec l'accès à un dictionnaire issu d'une base lexicale gérée par un logiciel externe. En effet PIVAX, le logiciel fournissant l'accès au logiciel de gestion des bases lexicales, possède lui-même déjà sa propre gestion d'utilisateurs.

Le projet Traouiéro ayant pour objectif l'opérationnalisation des logiciels développés depuis de nombreuses années dans le GETALP, pour des thèses ou dans des projets, un des logiciels concernés était SECTra, dans la perspective de la création d'une entreprise capable de valoriser cette technologie. La refonte du logiciel SECTra exploité, en particulier, pour gérer les mémoires de traductions dédiées ou partagées par les instances des clients de la société, a conduit au choix d'une architecture modulaire.

L'ancienne version, SECTra2 est toujours la version exploitée à T0+24, et le transfert de ses interfaces spécialisées pour différentes exploitation du logiciel, dans les logiciels indépendants (dits « modules-vues » parce qu'ils réalisent des vues différentes des données, selon l'usage et les utilisateurs) est en cours de réalisation (voir les livrables T7o-L2.2.c-SECTra3 et T7o-L2.2.d-SECTraY).

### 1.3 EVOLUTION DE L'OBJECTIF

Ce document présente le travail fait sur la réalisation du module ayant à sa charge de fournir le service iMAG, nommé dans le titre « logiciel iMAG ».

Il fournit l'interface utilisateur des passerelles iMAG, et pour cela, exploite le logiciel SECTra, par les services auxquels il donne accès, pour toutes les fonctions métier : partage et accès (lecture et écriture) à une mémoire de traductions. Le logiciel iMAG contient :

- le module iMAG défini dans le document T7o-2.2.b §3.7, qui remplace les contenus textuels d'un site Web dédié par des objets du DOM affichant, selon le contexte de la visite de la page, (1) le texte d'origine ou (2) le texte traduit tel qu'il est défini comme la meilleure proposition de traduction de ce texte dans une langue cible et mémorisée dans une mémoire de traductions gérée comme un corpus de SECTra ou (3) le texte traduit encadré de symboles révélant sa fiabilité en terme de qualité de traduction ou encore (4) une palette d'interaction permettant au visiteur d'améliorer la qualité de la traduction affichée et d'accéder au mode avancé,
- le bandeau de paramétrage de l'accès au service iMAG pour le visiteur, et d'accès à d'autres services liés (comme la gestion des contributions pour les modérateurs...),
- le dorsal du service iMAG, ce dernier étant obtenu par externalisation à partir du système initial SECTra\_w/iMAG, appelé iMAG-édit dans le document T7o-2.2.b §3.2, qui fournit une interface pour post-éditer une liste de segments issus d'un même corpus géré par SECTra, en mesurant son temps de travail sur chaque segment.

Le logiciel iMAG doit aussi plus tard permettre d'exploiter d'autres logiciels intéressants pour assumer d'autres fonctions comme le partage et l'accès à des dictionnaires dédiés, Jibiki-Pivax.

Dans le cadre d'une exploitation commerciale, un site de commerce en ligne doit permettre à des clients de s'inscrire, de choisir une offre et d'effectuer les paiements éventuellement récurrents associés, ainsi que de paramétrer la configuration de son service. Une première version de ce site a été développée et est présentée dans le livrable T7o-L5.4.a-ValMarCom-eCom. Une des tâches de configuration consiste à définir et gérer les contributeurs et les modérateurs qui pourront intervenir en contribuant au service (post-édition des traductions, proposition de partage d'entrée de dictionnaire et négociation des définitions...).

Chacun des services exploités s'appuie sur une identification des utilisateurs et sur la mémorisation de l'attribution de leurs contributions, ce qui constitue les deux difficultés majeures à résoudre pour cette sous-tâche.

En effet, concernant le premier problème, on voudrait que :

- les utilisateurs n'aient à se connecter qu'une seule fois, sans avoir à redonner leur mot de passe en passant d'un service à l'autre (comme c'est le cas pour les services proposés par Google, Apple, Orange...);
- le service des passerelles iMAG accède aux pages des Intranets, en préservant la sécurité des sites Web des clients;
- la liste des services puisse s'étendre, et exploiter des logiciels non forcément développés par le GETALP (ou la future société qui commercialisera l'accès intégré à ces services depuis des pages Web « clientes »)
- le référencement du site Web du client ne doit pas être perturbé par l'utilisation des iMAG pour les rendre accessibles dans différentes langues.

La solution proposée peut reposer sur le partage d'un annuaire commun (comme LDAP), mais doit aussi s'adapter à ce qu'un des services ne gère pas les connexions par un tel annuaire.

Ces tâches doivent être résolues par le composant nommé « relais iMAG » dans le titre de ce document. Pour les tâches liées à la sécurité, plusieurs solutions ont été envisagées et sont présentées dans ce document, mais toutes n'ont pas été encore testées, et le choix définitif n'est pas fixé. Le cahier des charges est défini ainsi que les spécifications externes. Pour le problème du référencement, une technique de proxy a été mise en place et validée.

Concernant le deuxième problème, la gestion des attributions des contributions, il s'agit de répondre à un besoin générique. On voudrait :

- réaliser l'outillage des applications exploitées avec des programmes spécialisés dans la gestion du travail collaboratif aux quels ces tâches sont déléguées,
- minimiser l'intrusion dans les codes des logiciels existants, et le développement des codes non partagés.

Les avancées sur la résolution de ce problème sont présentées dans la troisième et dernière partie du document, dédiée au « dorsal de modération ».

#### **1.4 BUT ET ORGANISATION DU DOCUMENT**

Les objectifs de cette sous-tâche ont été raffinés et elle a été organisée en plusieurs volets : logiciel iMAG, relais iMAG et dorsal de modération, et pour la réalisation de chaque volet, des phases ont été identifiées.

En plus de l'introduction et la description de l'organisation du travail, ce document contient 3 sections, qui présentent chacune un des 3 volets de cette sous-tâche :

- volet « Logiciel »
- volet « Relais »
- volet « Dorsal de modération »

Chaque section présente, selon le cas, les études, le travail de conception (cahier des charges, spécifications externes et internes) et les réalisations effectuées pour réaliser les programmes du volet correspondant.

Enfin, une dernière partie donnera une vue globale de l'état d'avancement de la sous-tâche.

## 2. ORGANISATION DU TRAVAIL

Ce document présente ce qui a été effectué ainsi que l'organisation prévue pour la fin de la réalisation de cette sous-tâche qui a été décomposée en 3 volets:

Volet Logiciel :	Logiciel iMAG
– phase 1	Traitements linguistiques et structurels sur la page Web visitée
– phase 2	Externalisation de l'affichage du bandeau et de la palette
– phase 3	Externalisation du dorsal de traduction, SECTra-édit
– phase 4	Spécification et développement des interacteurs pour l'aide lexicale
Volet Relais :	Relais iMAG : partage ou centralisation de l'accès au services
– phase 1	Scénario et affinement du cahier des charges
– phase 2	Etudes et choix techniques
– phase 3	Réalisation
Volet Contributions :	Gestion des attributions aux contributions (dorsal de modération)
– phase 1	Scénario et affinement du cahier des charges
– phase 2	Etudes et choix techniques
– phase 3	Réalisation

### 2.1 DÉCOMPOSITION EN PHASES DU VOLET LOGICIEL IMAG

Dans la nouvelle architecture modulaire, SECTra-iMAG sera le frontal du service iMAG, et SECTra-Edit sera le dorsal de traduction du service iMAG, tous deux étant obtenus par externalisation à partir du système initial SECTra\_w/iMAG.

Ce volet a pour charge l'externalisation du code dédié aux passerelles iMAG, qui a été directement intégré dans ceux de SECTra\_w (système d'exploitation de corpus de traductions sur le Web). L'objectif de cette externalisation est que le code produit soit directement utilisable dans les futures versions de SECTra, et que l'évolution de chacun des codes soit indépendante.

La première phase doit aboutir à l'exploitation d'un logiciel unifié pour SECTra2 et SECTra3 pour réaliser les traitements linguistiques et structurels d'une page Web afin de lui fournir l'accès multilingue par passerelle iMAG.

La deuxième phase s'occupe du transfert des interfaces visibles des passerelles iMAG, c'est à dire, le bandeau et la palette, mais aussi de la possibilité de fournir une passerelle transparente pour de moteurs de recherche.

La troisième phase aura en charge le transfert des fonctionnalités d'édition des mémoires de traductions et l'interface qui les fournit, pour agir sur les mémoires de traductions du noyau SECTra3. On supposera que les problèmes d'attribution des contributions sont résolus, c'est-à-dire que le service qui reçoit les actions à réaliser est capable de les traiter. Il s'agit donc de réaliser le dorsal de traduction, SECTra-Édit.

La dernière phase consiste à réaliser l'intégration de l'interaction avec des bases lexicales, dont une base lexicale dédiée à chaque service.

### 2.2 DÉCOMPOSITION EN PHASES DU VOLET RELAIS

Le rôle du relais est de relayer les droits entre les différents services utilisés dans une architecture multiservices dans lesquels les utilisateurs peuvent avoir des rôles et des droits différents selon le service exploité.

La première phase consiste à affiner le cahier des charges, à partir du cahier des charges initial et de scénario, et aboutir aux spécifications externes.

La deuxième phase doit proposer une solution technique, et fournir des spécifications internes.

Enfin la troisième phase concerne la réalisation.

### **2.3 DÉCOMPOSITION EN PHASE DU VOLET CONTRIBUTIONS**

Le volet Contributions doit aboutir à la réalisation d'un dorsal de modération. Le cahier des charges est défini, les spécifications externes sont partiellement définies, à partir des réalisations déjà opérationnelles dans SECTra2, mais une étude technique préalable est nécessaire pour définir les spécifications internes. Cela conduit, pour ce volet, aux 3 mêmes phases que pour le volet Relais.

### **2.4 L'ÉQUIPE**

L'équipe qui se consacre à cette sous-tâche est formée de Valérie BELLYNCK, avec l'aide d'Achille FALAISE (postdoc, ATER à la rentrée 2012, développeur de SECTRa-3), de Lingxiao WANG (contrat CIFFRE de doctorat chez Lingua-et-Machina) et de Christian BOITET. D'autres personnes sont intervenues par l'intermédiaire des contraintes sur la gestion des utilisateurs provenant des composants qu'ils ont conçus et développés : Francis BRUNET-MANQUAT et Léo ZANGELMI, qui développent le site de commerce en ligne, Mathieu MANGEOT et Ying ZHANG qui travaillent sur le développement de PIVAX++.

### **2.5 MÉTHODE DE TRAVAIL**

L'opérationnalisation de logiciels passe par l'intégration dans un processus de génie logiciel s'appuyant sur la rédaction de documents. Chaque volet de cette sous-tâche conduit à la rédaction d'une série de documents, dont on vise à terme qu'ils soient produits par Doxygen à des fins

- d'homogénéité sur tout le projet Traouiéro des logiciels qu'il aura permis de consolider,
- et de cohérence de la documentation avec les fonctionnalités des versions courantes des logiciels.

Chaque composant indépendant doit aussi être déposé en tant que projet de développement sur la forge, qui fournit une publication de pages Web associées à chaque projet.

Ainsi, les documents seront consultables en ligne à l'adresse :

<http://XXX.forge.imag.fr/projects/GL> où XXX est le nom du projet du volet

La liste des documents attendus pour chaque volet est (le cas échéant) :

XXX-UG/	Manuel d'utilisateur
XXX-CCH/	Cahier des charges
XXX-CG/	Conception globale
XXX-DSE/	Document de spécification externe
XXX-GR/	Gestion des risques
XXX-PQ/	Plan qualité

Pas de DSI?

### 3. LOGICIEL IMAG

Cette section présente les phases définies pour aboutir à la réalisation de la version du logiciel iMAG adaptée à la nouvelle architecture logicielle du service iMAG, et leur état d'avancement au moment de la rédaction de ce document de synthèse du projet Traouiéro.

#### 3.1 TRAITEMENTS LINGUISTIQUES ET STRUCTURELS D'UNE PAGE WEB

Cette phase doit aboutir à exploiter les mêmes codes pour traiter les pages Web et les présenter dans les passerelles d'accès multilingue iMAG, que ce soit pour SECTra2 ou pour SECTra3. Le module de la nouvelle architecture de SECTra qui réalise le frontal des passerelles iMAG, c'est-à-dire la production de la page Web « image » de la page Web d'origine (du site client) par la transformation en accès interactif à sa vue traduite, est appelé SECTra-iMAG ou logiciel iMAG. Il a été décrit dans le document T7o-L2.2.c.

##### 3.1.1 CAHIER DES CHARGES

###### RAPPEL L2.2.c

[Le logiciel iMAG est] essentiellement d'un agrégateur de services, manipulant des pages Web et des segments fournis par d'autres services, et les présentant à l'utilisateur via une interface graphique. Cette dernière reprend les spécifications de l'interface existante.

[Le logiciel] IMAG doit en outre mémoriser les paramètres des sites élus, de [façon] à pouvoir les retrouver dans les mémoires de traductions de SECTra.

Les traitements effectués pour traiter une page Web dans SECTra2 s'appuient sur la récupération de la traduction de la page dans l'outil de Google. Ce traitement a l'avantage de récupérer en même temps la segmentation de la page, son habillage, avec le remplacement du texte de chaque segment par un objet du DOM capable de réaliser l'affichage contextuel d'une palette et de la traduction du texte qu'il contient, et la palette elle-même. La méthode était acceptable tant que Google proposait l'utilisation de l'outil Google-translator-toolkit librement, même si l'outil utilisé n'est pas exactement la trousse à outils, mais en fait, le formulaire en ligne (<http://translate.google.fr/?q=trousse+à+outil&oe=utf-8&hl=fr&sl=fr&tl=en>).

Depuis 2010, Google a modifié sa politique d'autorisation de l'utilisation de ses services, et outillé le formulaire en ligne pour qu'il soit plus difficile à utiliser, concrétisant ainsi sa volonté de ne plus l'offrir librement à une exploitation intégrée dans des programmes et des services.

En 2012, une seconde modification des services de Google a rendu indisponible les passerelles iMAG tant qu'une solution n'avait pas été trouvée. La solution mise en place pour réaliser le traitement des pages Web visitées dans SECTra2 et basé sur Google, ne peut être conservée à terme, ne serait-ce que parce qu'elle rend nous rend dépendants des évolutions des outils fournis par Google.

Par contre, l'architecture de SECTra3 contient tous les modules nécessaires au traitement complet des pages Web, sans passer par Google. Il s'agit de :

- XMLise, un outil de normalisation qui reforme le HTML, détecte l'encodage et produit de l'UTF-8.
- Proximag, un proxy qui récupère la page Web, redéfinit les liens pour assurer l'effet de passerelle de navigation par le serveur d'AXiMAG, et doit gérer les cookies pour assurer l'authentification sur les intranets de sites Web.
- SegDoc, un outil de segmentation qui sépare la structure du document (appelé squelette) de son contenu textuel, et segmente le contenu textuel de la page en unités de traduction (voir le document T7o-L3.4-SegDoc §2.5.1).
- SegWear (qui devrait plutôt s'appeler SegDress), un outil d'habillage d'un squelette avec une liste de segments (voir le document T7o-L3.4-SegDoc §2.5.2).

L'objectif du volet « logiciel iMAG » est donc de transférer l'exploitation de ces services de SECTra3 à SECTra2 pour pouvoir continuer d'utiliser SECTra2 tant que tous les autres composants n'auront pas été terminés.

D'autre part, dans le document T7o-L2.2.c, il était précisé que le logiciel iMAG devait en outre mémoriser les paramètres des sites élus, de façon à pouvoir les retrouver dans les mémoires de traductions de SECTra. Cette charge revient en réalité au relais iMAG, qui fait l'objet du volet suivant (cf. section §4).

### 3.1.2 CODE DE SECTRA3

Programmé en PHP, pour être servi par un serveur Apache, le logiciel **iMAG** réalise un « metatransformation » qui enchaîne les appels de 4 « services » (**xmlise**, **proxy**, **segdoc**, **wear**) selon le processus suivant.

Après avoir récupéré l'URL de la page à traiter, SECTra-3 la fait traiter :

1. par le service **xmlise**, programmé en PHP, qui envoie une requête HTTP paramétrée en la transmettant par la fonction `curl_exec()` de PHP, puis qui lui applique une analyse suivie de sa resérialisation normalisée par la classe « tidy ». Ainsi, la page Web à traiter est bien formée au sens de XHTML.
2. par le service **proxy**, programmé en PHP, qui charge la page Web dans un DOM, pour y remplacer les liens spécifiés dans les attributs de balises à rendre absolus, et d'autres à « proxifier ». Il s'appuie sur un code PHP opensource – sous licence BSD-OSI – `url_to_absolute.php`, pour convertir des liens relatifs (ou contenant un préfixe qui définit une base en liens absolus). Une fois les liens adaptés, la page HTML est régénérée, et l'effet de navigation via une passerelle est assuré.
3. par le service **segdoc**, intégré au flux de traitement via un petit script PHP qui transmet ses paramètres à l'application programmée en java décrite dans le document T7o-L2.3-SegDoc. Le service de segmentation extrait un squelette en remplaçant les segments qu'il a identifiés, dans la page HTML qu'il traite, par des identifiants de segments. Il est adapté pour être exploité dans le metatransformation iMAG, de sorte que :
  - les identifiants des segments qu'il utilise sont exactement les identifiants des segments dans les bases de données de SECTra3 (si un segment n'y était pas, c'est le service de segmentation qui demande sa création),
  - le chaînage des segments de la page est géré grâce aux informations mémorisées dans les contextes gérés par SECTra3 (un contexte contient l'identifiant du corpus, l'URL de la page, l'identifiant du segment précédent, et l'identifiant du segment suivant).

Voir l'annexe 8.1.

4. par le service **wear**, intégré au flux de traitement via un petit script PHP comme le service SegDoc, qui transmet ses paramètres à l'application programmée en java décrite dans le document T7o-L2.3-SegDoc. Le service de rhabillage :
  - remet le contenu associé à chaque segment à la place de la référence à son identifiant dans un squelette, c'est-à-dire demande à SECTra3 les traductions dans la langue cible pour le segment référencé, et en crée une s'il n'y en avait pas.
  - et renvoie la page HTML obtenue.

Ainsi, c'est la version de SegDoc adaptée à SECTra3 qui initialise la traduction du segment s'il n'y a pas encore de traductions de ce segment dans la langue de visite demandée (langue cible) dans la mémoire de traductions associée au site du client.

Dans le service iMAG développé pour SECTra3, la classe `Segment` contient des fonctions publiques dont celles qui sont exploitées par SegDoc : les constructeurs des classes `Segment`, `Context`, `Metadoc` et `Doc`, la méthode `addContext` de l'objet `Segment`, son constructeur, celui d'un `Contexte`, les méthodes `getContent()`, `getLang()`, et les

méthodes `addTranslationTarget()`, `getBestTranslation()` et `getSegmentTarget()` de l'objet `Segment`.

Voir l'annexe 8.2.

### 3.1.3 ADAPTATION À SECTRA2

Il s'agit d'adapter le flux de traitements mis en œuvre pour réaliser les passerelles iMAG exploitant la base de données de SECTra3 pour pouvoir l'appliquer à SECTra2, le but étant de factoriser les développements de codes futurs.

Les codes impliqués sont ceux de deux services, `SegDoc` et `Wear`. Ils exploitent des fonctions déclarées publiques de la classe `Segment` de SECTra3. Le problème est que ces fonctions sont très dépendantes du modèle conceptuel de la base de données. On ne peut donc pas écrire toutes les mêmes fonctions publiques pour SECTra2 que celles qui sont publiées par les codes de SECTra3.

Il s'agit donc de transposer seulement les fonctions nécessaires pour le traitement des iMAG dans la situation modélisée par SECTra2.

Ainsi, les fonctions à définir sont toutes celles qui agissent directement sur la base de données (c'est-à-dire qui contiennent des commandes MySQL), ainsi que les fonctions que ces fonctions utilisent.

Cependant, les concepts représentés dans les modèles conceptuels de corpus gérés par SECTra2 et par SECTra3 diffèrent suffisamment pour s'attendre à ce que la définition de ces fonctions pour SECTra2 impose la prise en compte d'informations non encore prévues dans le corpus de SECTra3, et en conséquence une évolution de son modèle conceptuel.

Le processus de développement de SECTra3 (dit « en spirale » dans le document T7o-2.2.c) n'est pas incompatible avec cette nécessité, mais dans l'état actuel, il n'est pas encore certain que les évolutions à introduire n'affectent pas tellement la structure de sa base de donnée que cela n'introduise pas à nouveau beaucoup de travaux de programmation.

### 3.1.4 EXTENSION DE LA COUVERTURE DE FONCTIONS DE SECTRA2 DANS SECTRA3

La version actuelle de SECTra3 ne considère pas de différenciation des mémoires de traductions selon les clients, pour supporter le concept d'iMAG dédiée. En effet, la base de données ne représente qu'un seul corpus multilingue capable de réaliser une mémoire de traductions, et ne permet pas de gérer son évolution dans le temps (voir §5.3.2 et Figure 18Figure 21 : structure logique d'une base de données de corpus multilingue).

Pour y introduire l'évolution temporelle (intégrant le suivi de l'attribution des contributions et la modération), une étude est en cours pour disjoindre la gestion des 2 aspects : corpus et attributions (voir §5).

Pour y introduire la gestion des corpus, la base de données de SECTra3 doit être étendue en reprenant les concepts définis pour SECTra2.

Les programmes de SECTra-iMAG devront être ensuite modifiés pour :

- agir avec et sur les données qui identifient et définissent les corpus,
- fournir les informations associant les identifiants de segments et d'utilisateurs au dorsal de modération pour qu'il renseigne la base de données définie pour gérer les attributions des contributions.

Ainsi, par exemple, la classe `SegDoc` devra être modifiée pour agir aussi :

- sur l'extension de la base de données gérant la multiplicité des corpus, pour déterminer dans quel corpus rechercher un segment (identifié alors par son corpus et son identifiant dans la MT associée à ce corpus<sup>1</sup>) ;

---

<sup>1</sup> Pas « dans ce corpus » : un même segment peut apparaître plusieurs fois dans un document, dans plusieurs documents, et dans plusieurs corpus, tout en apparaissant dans une seule mémoire de traductions.

- sur la base de données du corpus de modération, puisqu'il faudra, dans la fonction qui retrouve un segment d'après son identifiant, et fournit ses traductions, retrouver aussi leurs contributeurs.

Le travail à réaliser est identifié et délimité, mais pas commencé à la date de la rédaction de ce livrable.

### 3.1.5 PERSPECTIVES SUR L'EXPORT/IMPORT ENTRE STRUCTURES DE BD DE CORPUS

La situation actuelle oblige à faire le constat que 2 modélisations des corpus doivent cohabiter pendant le temps du transfert des fonctionnalités. Cette situation peut et vraisemblablement va perdurer, puisqu'il est peu certain que toutes les données générées par différents iMAG soient migrées d'un SECTra à l'autre en une seule fois. D'autre part, SECTra2 gère des corpus dont la base de donnée est différente. Et même, quelques iMAG non « génériques » exploitent des bases de données dont la structure est différente de celles des iMAG « génériques ». Enfin, le travail de Lingxiao Wang dans la société Lingua-et-Machina lui permettra de valider l'approche en l'appliquant à l'outil Libellex de cette entreprise.

Ainsi, la réalisation du transfert des données ne peut être résumée à un transfert de structure si les concepts diffèrent.

Il apparaît alors nécessaire de définir et de réaliser un serveur de synchronisation de corpus, appelé dans la suite, « serveur de synchro ». Pour cela, 4 phases sont identifiées : serveur de synchro « initial », puis « à la demande », puis « relais », et enfin « narratif ».

#### 3.1.5.1 Phase I : serveur de synchro initial

Il s'agit d'un script qui sait transférer (créer ou mettre à jour) des segments entre les versions 2 et 3 de SECTra. Ce script s'appuie sur les modules d'import et d'export déjà réalisés, mais il doit savoir gérer le décalage des concepts définis dans les différentes versions. On voit émerger la notion de microstructure manipulée dans Jibiki-Pivax. La méthode usuelle consiste à plonger les concepts des différents systèmes dans une organisation universelle, de façon à pouvoir identifier comment les transférer (répondre à la question : « qu'est-ce qui devient quoi, dans les 2 sens »).<sup>2</sup>

D'un point de vue pratique, il s'agit de réaliser des scripts capables de :

- demander 1 segment à SECTra2 et à SECTra3 (une fonction commune s'appliquant différemment selon la version de SECTra) et à Libellex ;
- envoyer un segment à SECTra2 et à SECTra3 (idem) ;
- entre les deux, faire la mise en correspondance ("mapping"), ce qui comprend :
  - pour les informations "communes", envoyer les infos récupérées dans les champs correspondants (« qu'est-ce qui devient quoi, dans les 2 sens »),
  - pour les informations présentes uniquement dans SECTra2, envoyer les informations qui concernent l'attribution de contributions et la gestion du traçage et de la modération, dans la BD conçue pour porter le dorsal de modération,
  - pour les informations présentes uniquement dans SECTra2, envoyer les informations qui concernent le support aux calculs de mesures d'évaluation qualitative des segments traduits, dans la BD conçue pour porter SECTra-Eval,
  - pour les informations présentes uniquement dans SECTra2, envoyer les informations qui concernent des concepts non gérés par SECTra3,
  - pour les informations présentes uniquement dans SECTra2, envoyer les informations qui restent dans une BD *ad hoc*,

<sup>2</sup> (CB) Attention, SECTra3 ne contient pas certaines entités essentielles introduites dans le thèse de Phap, comme les « segments multilingués contextualisés », les « métasegments », les « segments multilingues », les « pseudodocuments », et les « contributions » et autres annotations. Il faudra plutôt définir un SECTra4 qui sera cette « organisation universelle ». Peut-être même faudra-t-il passer à SECTra4 avant de finir SECTra3, puisque ce dernier ne contient pas les fonctionnalités « métier » (évaluation, post-édition, extraction de sous-mémoires, segmentations multiples...).

- pour les informations uniquement dans SECTra3, envoyer les informations qui concernent des concepts non gérés par SECTra2.

Les informations sur la post-édition, qui n'existent pas encore dans Sectra3, resteront (pour l'instant) dans SECTra2. Quand elles seront gérées par le dorsal de modération, les scripts de synchronisation devront être modifiés pour l'exploiter et remplacer leur gestion déportée dans SECTra2.

Pour garder les serveurs « synchronisés », il suffira d'exécuter régulièrement ce script sur tous les segments.

#### *3.1.5.2 Phase II: serveur de synchro à la demande*

Le script de synchro est modifié pour fonctionner uniquement à la demande (ou alors on créera un deuxième processus plus prioritaire).

On modifie SECTra2 et SECTra3 pour qu'ils notifient les changements au serveur de synchro. Ils ne précisent pas (à ce stade) quelle modification ils font, mais seulement qu'ils ont modifié tel segment (en donnant l'identifiant du segment dans leur propre base). C'est le serveur de synchro qui ira chercher ce qui aura été modifié sur le segment désigné, pour reporter la synchronisation à exécuter.

A partir de cette phase, plusieurs BD pourront fonctionner en parallèle sur différents serveurs (plusieurs SECTra2, plusieurs SECTra3...), ce qui sera utile pour la montée en charge, et permettra de mettre en place des versions de test des services complets.

#### *3.1.5.3 Phase III: serveur de synchro "relais"*

L'API de SECTra3 est modifiée pour qu'elle passe uniquement par le serveur de synchro, sans jamais communiquer directement avec la base. Alors toutes les requêtes agissant sur les corpus passent par une couche logique unique qui transmet différemment les ordres aux bases de données qui les porte.

#### *3.1.5.4 Phase IV: serveur de synchro "narratif"*

SECTra2 et SECTra3 ne se contentent plus de notifier quel segment ils ont modifié, mais ils précisent quelle modification ils ont fait. On a besoin d'un langage générique (au moins compatible SECTra2 et SECTra3), et dans une approche narrative.

Ce langage narratif défini pour manipuler les corpus doit aussi être exploité dans iPoCorp, l'entrepôt de bases corporales, offrant une gestion des scripts définis pour les traiter.

Par exemple, un export d'un corpus d'iPoCorp peut aboutir à des imports corrélés entre plusieurs BD de services (comme le noyau de SECTra3 et la base de données du dorsal de modération). Ils sont corrélés parce que l'import dans l'un crée des identifiants qui sont nécessaires à l'autre (les clés étrangères), donc modifie les données à importer dans un autre.

On pourrait imaginer d'aller plus loin en considérant que tout service a des données à faire évoluer dans un entrepôt à services dédié à ce type de données (iPoTruc ou iPoX). Mais les données n'étant pas autonomes, elles ne sont pas mémorisées dans un iPoTruc : les modules-vues de SECTra3 n'ont pas de BD constituant des ressources comme un corpus ou un lexique. Les ressources qui sont mémorisées et manipulées dans les iPoTruc sont autonomes.

## **3.2 EXTERNALISATION DE L’AFFICHAGE DU BANDEAU ET DE LA PALETTE**

Cette phase présente la conception et la réalisation du bandeau et de la palette d'interaction à mettre en place pour la nouvelle architecture.

La méthode choisie pour la mise en place du bandeau et de la palette a été conçue pour réaliser une intégration commune dans l'ancienne et dans la nouvelle architecture. Sa

réalisation a été le support de l'étude technique qui a permis de définir la stratégie de transfert décrite dans le document T7o-L2.2.d-SECTraY.

Ce document rappelle aussi la technique utilisée pour rendre transparente le passage par une passerelle, pour, par exemple ne pas interférer avec les paramètres du serveur client qui assurent son référencement.

### 3.2.1 TRANSFERT DU BANDEAU

Ce que nous appelons « bandeau » est un affichage permanent, dans la partie supérieure de la fenêtre du navigateur, d'une bande rendant explicite le passage de la page Web visitée par un service Web intermédiaire plutôt que par accès direct. Google-translate et Microsoft-bing présentent un tel bandeau (voir Figure 1 et Figure 2).



**Figure 1 : bandeau de Google-translate**



**Figure 2 : bandeau de Microsoft-bing**

L'intégration du service dans les pages Web d'un site Web peut être réalisée de plusieurs manières, qui sont proposées lors de l'étape de configuration pendant la souscription du service ou ultérieurement en agissant sur le paramétrage de l'offre choisie.

Lors de la visite initiale du site client dans la langue dans laquelle il a été publié (c'est-à-dire sans encore être passé par la passerelle iMAG), une liste de langues cible (les langues d'accès pour visiter le site) est proposée. Lorsqu'une langue est choisie, la page visitée est rafraîchie et remplacée par sa version résultant du passage dans le service iMAG. Pour rendre sensible le fait que la visite se fait à travers une passerelle, et n'est donc plus la visite directe, le bandeau est affiché (voir Figure 6).

Un bandeau minimal a été réalisé pour SECTra3-iMAG (voir Figure 3). Il ne permet que de spécifier l'URL de la page Web à visiter, préciser le corpus, la langue source et la langue d'accès.



**Figure 3 : bandeau de SECTra3-iMAG**

#### 3.2.1.1 Cahier des charges et spécifications externes

Il faudra, lorsque le relais et le dorsal de modération auront été développés, ajouter au bandeau réalisé pour SECTra3 les fonctions qui étaient présentes dans le bandeau réalisé pour SECTra2/iMAG, c'est à dire des liens pour se connecter, accéder à l'aide en ligne, et des liens adaptés aux droits de l'utilisateur, comme, pour un modérateur, gérer le flux de contribution qu'il doit gérer, ou pour un utilisateur identifié, visualiser les statistiques calculées sur ces contributions.

La Figure 4 montre le bandeau dans une session identifiée par un modérateur (compte de « wally ») alors que celui de la Figure 6 le montre dans une session anonyme.



**Figure 4 : bandeau de SECTra2/iMAG**

Dans SECTra2, les configurations des iMAG permettent de choisir une langue source<sup>3</sup> et un ensemble de langues cibles. La réalisation du bandeau, et en particulier du menu déroulant permettant d'accéder à la visite de la page courante dans une des langues qu'il propose, dépend de ce paramétrage.

La liste des langues cible est affichée en respectant une légende de couleur :

- les langues dans lesquelles une faible proportion des segments de la page a été post-éditée et validée par un modérateur, sont affichées en orange (langues dites « faiblement modérées »). les langues sont affichées en rouge quand il n'y a pas de tel segment (langues dites « pas modérées »),
- les langues sont affichées en vert, au contraire, si la plupart des segments ont été modérés (langues dites « fortement modérées »).

Le bandeau proposant de changer de langue de visite, ce choix ne doit plus apparaître dans le contenu de la page accédée (voir Figure 5 et Figure 6).



**Figure 5 : Page Web d'origine et accès à sa passerelle iMAG par le choix d'une langue de visite**



**Figure 6 : Accès à une page Web via son iMAG : apparition du bandeau**

Quelques options permettent d'agir sur l'affichage :

- la case à cocher « Reliability » permet d'activer ou non un effet visuel pour rendre sensible la fiabilité présumée de la traduction de chaque segment, avec une légende de couleurs différente de celle du menu du choix des langues :
  - si le segment est encadré d'accolades de couleur verte, c'est qu'il a été post-édité par un contributeur ayant un compte et connecté (ce peut être un visiteur bilingue traducteur occasionnel, un modérateur, un traducteur professionnel, agréé ou non par le site élu),
  - si le segment est encadré d'accolades de couleur orange, c'est qu'il a été post-édité par un contributeur non identifié,
  - si le segment est encadré d'accolades de couleur rouge, c'est que la traduction est le résultat d'une traduction automatique.

<sup>3</sup> Comme on intégrera la détection des langues, il faudra plutôt définir une *langue source par défaut*, et des langues source à ne pas traduire, toutes les autres étant à traduire dans la langue d'accès (cible) choisie.

- le bouton radio « Original » permet de diviser en deux la zone principale de la fenêtre, située juste sous le bandeau, verticalement, pour voir simultanément la page originale et la page traduite, alors que le bouton radio « Translation only » permet de revenir à l'affichage de la page traduite seulement, sur toute la largeur de cette zone.
- Il faudra prévoir quelques contrôles de plus pour permettre de visualiser à droite non pas la page originale, mais la page traduite par TA seulement, avec tel ou tel système, ou avec pour chaque segment la TA retenue, et/ou avec la présentation « Trace » permettant de visualiser l'effort de post-édition.

Pour aller plus loin, le paramétrage d'une iMAG doit pouvoir permettre au client de lui attribuer une présentation (message de bienvenue, feuille de style), et de choisir les interacteurs qu'il veut proposer à ses visiteurs, selon leur rôle.

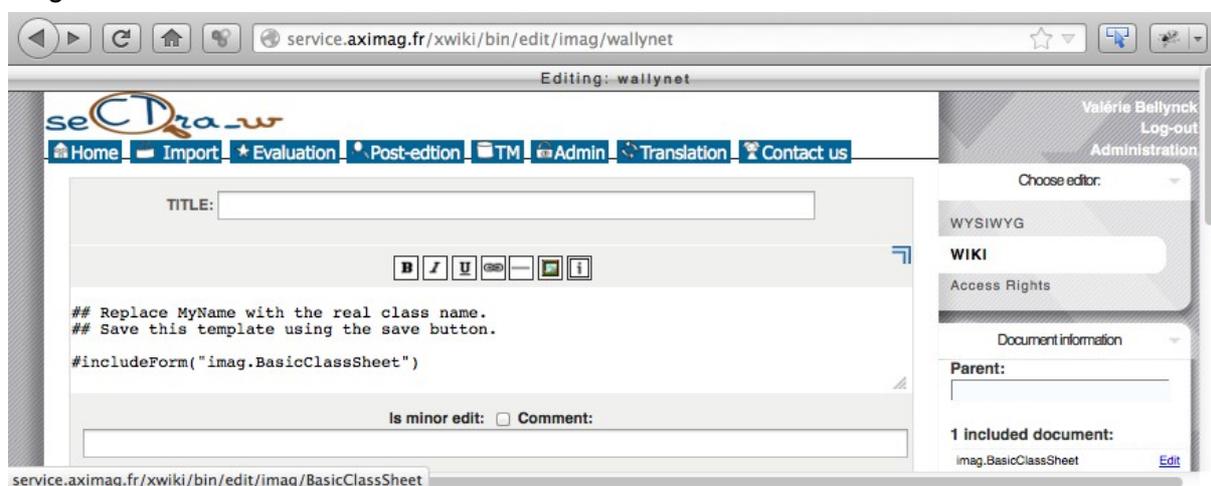
Par exemple, la feuille de style pourrait générer une présentation « à la Neon », avec le texte source écrit au-dessus du texte cible, ligne par ligne, et la palette pourrait contenir les noms des contributeurs à chaque segment, et/ou son origine (niveau de \* à \*\*\*\*\*), et/ou sa note (de 0 à 20).

Cette description constitue le cahier des charges du bandeau et ses spécifications externes. Sa réalisation est assujettie à celles du relais et du dorsal de modération.

### 3.2.1.2 Réalisations en cours

Les codes qui réalisent le bandeau dans SECTra2/iMAG sont contenus dans la classe qui programme toutes les fonctionnalités des iMAG, c'est-à-dire une seule « page » de xWiki.

Pour mémoire, les URL des sites Web xWiki définissent l'action, l'espace et la page qui sont respectivement « edit », « view » et « wallynet » dans l'URL de la Figure 7 Figure 7. Le code de la page est exécuté quand l'action est « view », comme c'est le cas dans la Figure 8 Figure 6.



**Figure 7 : « page » xWiki générant une iMAG = script en Velocity pour inclure la classe générique**

```

## Hong-Thai Nguyen
## 04/06/2010
## Presentation of creation/edition form of an iMAG
## 24/02/2012: Xan, Ying, Lingxiao: petites modifications sur le "bouton d'appel à une
iMAG-S"
## -- pas trouvé la variable contenant les préfixes supplémentaires autorisés
## 27/02/2012: (CB) Elle y est, voir la ligne : Prefix of the processing site:
#showItem("ulr_prefix", $isExisted)

<script type="text/javascript" src="$xwiki.getSkinFile("tablekit/imag_create.js")">
</script>

1 iMAG name $doc.web / $doc.name

#set($class = $doc.getObject("imag.BasicClass").xWikiClass)
#set($obj) = $doc.getObject("imag.BasicClass")

#set ($grvTrans = $xwiki.parseGroovyFromPage("Groovy.Translation"))
#set ($lang_list = $grvTrans.getAllStringLanguageNames())
#set ($smt_list = $grvTrans.getAllMTSystemsName())
##### Mise en commentaire sans doute par Hong Thai: EWS_MTSystème et
EWS_SL sont affectés plus loin. #####
# $class.get("EWS_MTSystème").setValues($smt_list)
# set($class.get(" language_code").classProperty.values = $lang_list)
# set($class.get(" EWS_SL").classProperty.values = $lang_list)

```

**Figure 8 : code Velocity de la classe générique des iMAG**

Dans xWiki, les pages peuvent être des scripts définis dans différents langages de script, dont principalement le langage de script développé par Apache, Velocity, mais aussi Groovy (ou même Python, Ruby et PHP qui n'ont pas été utilisés).

Velocity est utilisé pour définir les modèles de pages Web, et est mélangé aux directives de présentation Wiki (par exemple, si les 3 premiers caractères de la ligne sont "1. ", alors xWiki encadrera le texte qui suit sur la même ligne par les balises <h1> et </h1>). Ces codes eux-mêmes peuvent inclure d'autres contenus écrits en Velocity ou dans le langage Groovy.

Groovy est très proche de Java, mais peut s'utiliser dans xWiki sans requérir d'étape de compilation, déploiement et redémarrage du serveur. Ce sont les contenus en Groovy qui peuvent accéder à la base de données.

L'annexe §8.3 présente l'extrait du code de la classe Groovy.Translation, en langage Groovy, qui génère le bandeau. On voit que le code Groovy prépare une chaîne en insérant des valeurs de paramètres dans le code HTML du bandeau, et en prenant en compte différentes réalisations selon les droits de l'utilisateur de la session courante.

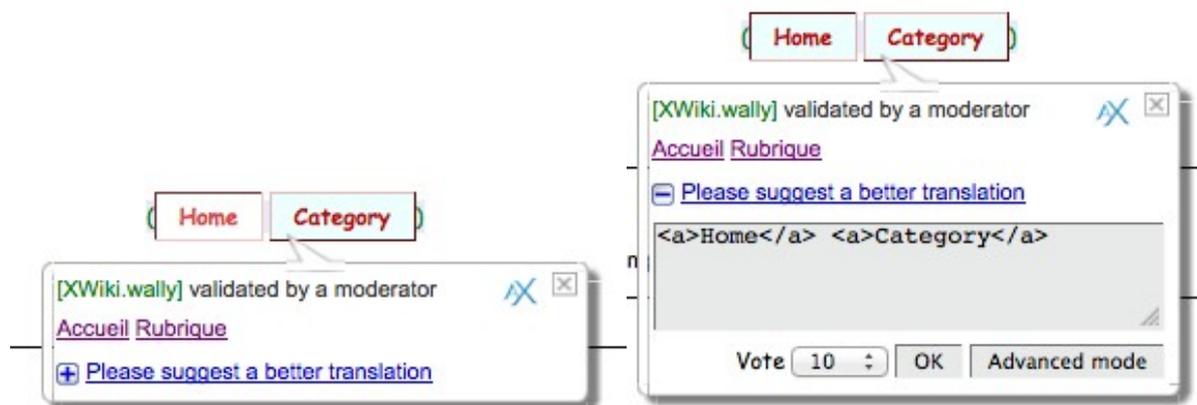
L'annexe §8.4 présente la classe Velocity qui a été définie pour renvoyer la liste des langues cible d'une iMAG. Dans ce code,

- seule la partie centrale réalise l'action visée ;
- la première partie récupère les paramètres transmis dans la requête et les informations mémorisées dans le « contexte » de la requête (le « contexte » est le concept et un objet de xWiki qui permet d'accéder aux informations sur l'URL, les informations sur le client, sur la session de l'utilisateur courant...);
- et la dernière partie gère les retours d'information pour expliquer les paramètres manquant quand ceux qui ont été transmis ne suffisent pas pour que l'API puisse répondre à la requête de façon satisfaisante.

Les sites clients peuvent intégrer l'exploitation de ce type de classe fournie par xWiki dans des codes Javascript.

### 3.2.2 TRANSFERT DE LA PALETTE

La palette de post-édition en contexte qui surgit au survol des segments de texte de SECTra2 (voir Figure 9) est la même que celle de Google-translate.



**Figure 9 : palette "bulle" iMAG de SECTra2, identique à celle de Google, en état plié et en état déplié**

Il n'y a pas encore de palette de post-édition en contexte dans SECTra3-iMAG.

### 3.2.2.1 Cahier des charges

La palette doit afficher :

- le surnom du contributeur, et l'état de validation résultant de la modération,
- le texte dans la langue source,
- un interacteur pour accéder aux possibilités de contribution
- dans le cas où l'interacteur est déplié,
  - un champ de saisie pour permettre la contribution,
  - un interacteur pour choisir une note de 0 à 20 (pour que le contributeur puisse informer sur la qualité de sa contribution),
  - un bouton pour envoyer la proposition de contribution,
  - un bouton d'accès au mode avancé (c'est-à-dire au dorsal de traduction, réalisé par SECTra-Edit),
- un bouton de fermeture de la palette.

IMPORTANT (demandé par Laurent Besacier pour des expés) : il faudra insérer, aussi bien dans les vues des segments dans SECTra que dans les palettes, un ou des contrôles permettant de « marquer » le segment courant par rapport à certains critères définissables par l'administrateur du projet en cours (par exemple, erreur de tel type, ou mauvaise segmentation, ou expression/terme difficile...).

Sa présentation doit être modifiable par les configurations d'une iMAG, en fournissant une feuille de style.

### 3.2.2.2 Codage dans SECTra2

Ce codage est déjà opérationnel, à part bien sûr ce qui suit « IMPORTANT ».

### 3.2.2.3 Externalisation du code de SECTra2

Le code présenté dans l'annexe §8.4 est un modèle pour réaliser des API « à la REST » pour toutes les fonctions de SECTra2 à externaliser.

### 3.2.2.4 Transfert dans le code de SECTra3

C'est en cours.

### 3.2.3 IMAG TRANSPARENTES AVEC PROXIMAG



### 3.3 EXTERNALISATION DU DORSAL DE TRADUCTION, SECTRA-EDIT

Pour l'implémentation de SECTra-Edit, nous avons choisi un développement de transfert : il s'agit d'extraire les interacteurs multifonctionnels implémentés dans SECTra-2, en y développant préalablement les fonctions REST qui fournissent les informations dont ils ont besoin.

#### 3.3.1 CAHIER DES CHARGES ET SPÉCIFICATIONS EXTERNES

Le cahier des charges de l'interface de post-édition des mémoires de traductions est défini dans la thèse de HUYNH Cong Phap, comme un éditeur adapté à la post-édition collaborative et contributive (Chapitre II, §3.1.2.3 et §3.2.2).

*« L'éditeur doit permettre à plusieurs contributeurs de travailler simultanément sur la même collection de données (segments, pages, ou documents). Pour cela, l'éditeur doit gérer et contrôler les modifications des contributeurs. [...]*

*SECTra\_w doit également gérer les versions de post-éditions, mais il assure que la version suivante est l'amélioration de la précédente et que la version finale est en principe la meilleure.*

*L'éditeur doit également être mis en page selon les principes de présentation des interfaces de visualisation et d'évaluation [...] : verticalité, horizontalité, localité, proactivité. [...]*

*L'éditeur a plusieurs modes de fonctionnement [ : le] Mode édition, [où] l'éditeur incite les contributeurs à travailler comme des post-éditeurs [ , le] Mode lecture, [qui] sert à visualiser une "trace" de l'effort de post-édition et à vérifier la qualité des résultats.*

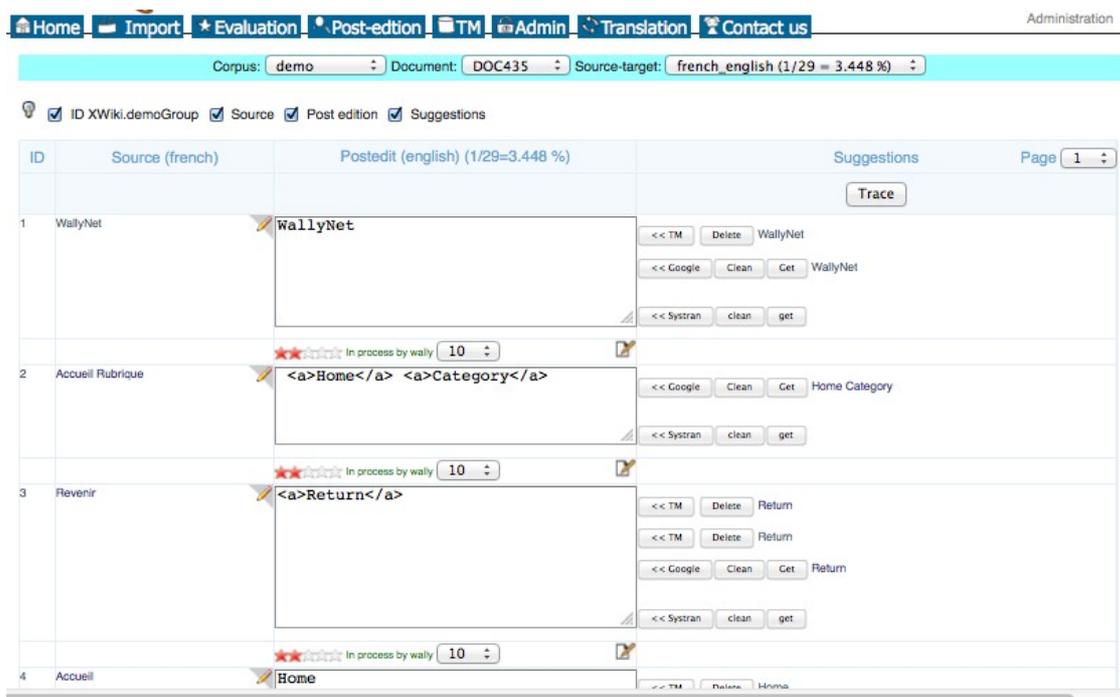
*[... L'éditeur doit intégrer des] fonctions de support à la post-édition : Appel de TA, pour chaque segment à post-éditer en une langue, il faut fournir plusieurs pré-traductions produites par plusieurs systèmes de TA ; recherche « exacte », [... pour fournir] la recherche exacte des segments source dans les MT.*

*[... L'éditeur doit intégrer des] fonctions de support des informations lexicales [... pour] fournir aux post-éditeurs des informations lexicales utiles de façon proactive.*

*[... Il doit intégrer des] fonctions de recherche et de filtrage des données : Filtrage de données [ , pour] filtrer des traductions selon plusieurs critères ; Recherche-remplacement de données [ , pour] chercher un texte en langue source ou/et en langue cible, et de remplacer le texte en langue cible par un autre texte.*

*[... Il doit intégrer des] fonctions de support au travail collaboratif : communication et échange de connaissances entre les post-éditeurs [... ;] gestion de versions de post-édition [... ;] association des métadonnées au segment.*

*[... Il doit intégrer des] fonctions de gestion d'effort et de progrès de post-édition, de construction et de visualisation des documents source et cible. »*



**Figure 12 : Dorsal de post-édition de SECTra2**

Pour réaliser SECTra-Edit, il s'agit de transférer l'interface de SECTra2 pour la post-édition collaborative, de façon à ce que les actions qu'elle transmet au noyau de SECTra2 puissent être appliquées à un autre noyau, au pris d'un minimum de recodage.

Le processus de transfert est décrit dans le document T7o-L2.2.d-SECTraY.

### 3.3.2 SPÉCIFICATIONS INTERNES

En résumé, il s'agit de générer le même code HTML, mais d'associer les requêtes HTTP gérées par le code Javascript à un code générique permettant de réaliser l'appel adapté aux fonctions fournies par le noyau visé.

### 3.3.3 CALÉNDRIER PRÉVISIONNEL

Le transfert de l'interface d'édition intensive des traductions, réalise le dorsal de traduction des iMAG et est constitué par le module SECTra-édit.

Son développement est prévu entre avril et juin 2013. Sont prévus pour cette tâche, Lingxiao Wang et Valérie Belynyck.

## 3.4 SPÉCIFICATION DE L' INTERACTEUR POUR L'AIDE LEXICALE

### 3.4.1 CAHIER DES CHARGES

Le cahier des charges de l'interface de post-édition des mémoires de traductions défini dans la thèse de HUYNH Cong Phap, (Chapitre II, §3.1.2.3) précise les fonctions attendues de l'interacteur pour l'aide lexicale, en tant que fonction de support des informations lexicales »

*« Cette fonction doit fournir aux post-éditeurs des informations lexicales utiles de façon proactive. Ces informations lexicales doivent donc pouvoir être obtenues par fusion à partir de ressources différentes, et elles doivent être relatives au domaine du segment. Cette fonction doit fournir une interface permettant de visualiser, de proposer, et de normaliser des informations lexicales. Elle doit aussi permettre de mettre à jour des informations lexicales quand un segment source est corrigé. »*

Une aide lexicale est intégrée à Google-toolkit.

### 3.4.2 SPÉCIFICATIONS INTERNES

Les spécifications internes de la communication avec le serveur de données lexicales, PIVAX, sont déduites de celles qui ont dirigées l'intégration qui avait été faite précédemment entre SECTra\_w et PIVAX par Hong-Thai Nguyen pour sa thèse.

Pour une démonstration de faisabilité, SECTra\_w faisait appel au serveur de données lexicales, PIVAX. L'extrait suivant vient de la thèse de Hong-Thai Nguyen (§2.3.2.4 et §4.1.3.3). et présente les spécifications internes qui seront adaptées à la communication entre la nouvelle architecture de SECTra et la nouvelle réalisation de PIVAX, Jibiki-Pivax.

« Pour une démonstration utilisant la connexion avec SECTra\_w, on montre ici l'exemple de la commande de récupération de résultat de traduction des articles de EOLSS. » EOLSS est un corpus issu de pages Web (.aspx), constituant 25 articles de l'EOLSS (*Encyclopedia of Life Support Systems*) qui ont été traduits collaborativement de l'anglais vers les 5 autres langues officielles de l'UNESCO (français, espagnol, russe, chinois, arabe) en utilisant SECTra\_w.

<pre>&lt;COMMANDE num_cde="4"&gt;   &lt;num_cde&gt;4&lt;/num_cde&gt;   &lt;nom_cde&gt;DOCUMENT_REV_LIST&lt;/nom_cde&gt;   &lt;url&gt;xwiki/bin/view/Corpus/PostEdit&lt;/url&gt;   &lt;method&gt;get&lt;/method&gt;   &lt;delay&gt;1000&lt;/delay&gt;   &lt;num_cde_previous_list/&gt;   &lt;PARAMETRE_SAISIE&gt;     &lt;PARAMETRE&gt;       &lt;nom_param&gt;projName&lt;/nom_param&gt;       &lt;libelle_param&gt;         Project       &lt;/libelle_param&gt;       &lt;dim_lib_parent&gt;         id_parent_a_attacher_dans_interface       &lt;/dim_lib_parent&gt;       &lt;dim_lib_lenght&gt;20&lt;/dim_lib_lenght&gt;       &lt;type_param&gt;Hidden&lt;/type_param&gt;       &lt;valeur_def_param&gt;         EOLSS       &lt;/valeur_def_param&gt;       &lt;local_valeur_list&gt;         /PROJECT_LIST/PROJECT       &lt;/local_valeur_list&gt;       &lt;dim_parent&gt;         id_parent_a_attacher_dans_interface       &lt;/dim_parent&gt;       &lt;dim_lenght&gt;20&lt;/dim_lenght&gt;     &lt;/PARAMETRE&gt;   &lt;/PARAMETRE_SAISIE&gt;</pre>	<pre>&lt;RESULTAT&gt;   &lt;item-group name="rev-projet"&gt;     &lt;item name="no" type="xpath" out="text"       expression="/html[1]/body[1]/div[1]       /tbody[1]/tr/td[1]/div[1]"/&gt;     &lt;item name="documentName" type="xpath"       out="text"       expression="/html[1]/body[1]       table[1]/tbody[1]/tr/td[2]"/&gt;     &lt;item name="documentStatus" type="xpath"       out="text"       expression="/html[1]/body[1]/       div[1]tbody[1]/tr/td[3]"/&gt;   &lt;/item-group&gt; &lt;/RESULTAT&gt; &lt;PRESENTATION type="xslt"&gt;   &lt;xslt-code&gt;   &lt;!-- XSLT code ici ... --&gt;   &lt;/xslt-code&gt; &lt;/PRESENTATION&gt; &lt;/COMMANDE&gt;</pre>
--	--

Les codes XML de gauche ci-dessus présentent la commande envoyée par SECTra\_w dans une requête HTTP à EMEU\_w<sup>4</sup>, aussi appelé « moniteur Web » (Environnement Multilingue pour EOLSS basé sur UNL) qui réalise un portail unique pour fournir l'accès sur le Web à SECTra\_w pour récupérer le résultat de traduction, PIVAX pour la partie lexicale, TRADOH+ pour le service de traduction automatique, etc. et ainsi centralise la communication entre ces environnements.

« Le code XML de droite ci-dessus, présente le résultat rendu en retour à la commande précédente. Le résultat fournit 3 expressions XPath (pour désigner le numéro de document (article dans EOLSS), le nom identifiant le document et le statut actuel de la traduction) et un

<sup>4</sup> EMEU\_w a été réalisé par NGUYEN Hong Thai dans sa thèse par extension du service Web UNLdeco conçu et réalisé par Gilles Sérasset en 2003, pour un contrat visant à multilingualiser le site B@bel de l'Unesco via UNL.

code de transformation XSLT pour définir la présentation du résultat dans l'interface d'EMEU\_w.

Cela montre l'efficacité réelle dans le contexte de gestion de système. Si le système SECTra\_w en cours de développement change sur la page de présentation, cela prend quelques minutes pour adapter le fichier XML du métalangage. Même un non-informaticien peut le faire. Ensuite, le moniteur continue à rendre le service attendu.

Dans la partie d'en-tête de commande, on décrit le chemin URL, la méthode d'envoi et le temps d'attente. Si le système accédé ne répond pas dans le temps d'attente, on considère que le système est occupé et on renvoie la requête en peu plus tard. On a construit un mécanisme de gestion de commandes à envoyer dans une file d'attente. On peut configurer cette file d'attente avec une stratégie de file (FIFO), de pile (LIFO), ou toute autre stratégie programmable. »

Le texte ci-dessous est extrait de la thèse de HUYNH Cong Phap chapitre II § 3.1.2.3.c.ii :

« L'appel de la préparation d'un minidictionnaire dans SECTra\_w et PIVAX avec les paramètres de la requête HTTP et le format du résultat HTML a la forme suivante. »

Paramètre	Nom	Valeur par défaut	Note
Protocole	Proto- cole	http	Protocole de commande
Nom_serveur	server	eolss.imag.fr	Nom de serveur
Port	port	80	Port
URL	url	/pivax ?QuickSearch.- po	Chemin sur le serveur
Mot_à_cher- cher	head- word	'test'	Mot ou suite de mots à consulter
Délai	delai	100	Temps d'attente maximal, s'il est dépassé, on passe à une autre requête
Lemma	lemma	test	Demander d'effectuer la lemmatisation si elle existe
Langue_sourc e	ls	yes	Langue source
Langue_cible	lc	"	Langue cible
Segment_ID	segid		Identificateur de segment pour que SECTra_w réattache le résultat à ce segment

Le résultat renvoyé par PIVAX est un fichier HTML sous le format simple de tableau. Sur ce fichier, on déclare les informations à extraire (par exemple en utilisant des XPath comme dans Jibiki) et à afficher sur l'interface simplifiée de PIVAX dans SECTra\_w. »

```
<html>
  <body>
    <table>
      <tr><td lang="en">test<td></tr>
      <tr><td lang="fr" cat="verb">tester<td></tr>
      <tr><td lang="fr" cat="noun">essai<td></tr>
      <tr><td lang="fr" cat="noun">test<td></tr>
    </table>
    <div>... autres informations ...</div> </body>
```

<b>&lt;/html&gt;</b>		
<b>MOT_SOURCE</b>	<b>/html/body/table/tr/td[lang="en"]</b>	<b>Récupération du mot source</b>
<b>MOT_CIBLE</b>	<b>/html/body/table/tr/td[lang="fr"]</b>	<b>Récupération de la liste des équivalents dans les langues cibles</b>

### 3.4.3 CALENDRIER PRÉVISIONNEL

La réalisation de l'interacteur pour l'aide lexicale nécessite l'opérationnalisation du moniteur Web EMEU\_w ou d'un autre composant assumant le même rôle en ce qui concerne la communication entre SECTra et Jibiki-Pivax.

Le calendrier ne contient qu'une date estimée de début du développement, les autres parties des logiciels iMAG devant être réalisés prioritairement.

Cependant cette fonctionnalité est cruciale pour la pertinence des offres commerciales d'exploitation des passerelles iMAG.

Sa réalisation est donc urgente.

## 4. RELAIS IMAG

Lorsque le problème du transfert des droits a été identifié, le rôle du relais devait être de relayer les droits entre les différents services utilisés dans une architecture multiservices dans lesquels les utilisateurs peuvent avoir des rôles et des droits différents selon le service exploité. Il s'agit de transmettre ces droits sans introduire de faille de sécurité dans les services exploités.

Non seulement les types des rôles peuvent être différents (quelqu'un peut être client dans un site commercial de souscription d'offres de services, et modérateur dans un site contributif visant à construire collaborativement de la connaissance), mais les termes utilisés pour identifier les rôles peuvent être identiques tout en ne correspondant pas à des accès identiques aux mêmes fonctions.

RAPPEL  
doc.  
scientifique

Il faut [...] créer un *relais-iMAG*, qui, en fonction des informations sur le site S et sur les paramètres de l'accès multilingue à S sur les langues (lesquelles sont post-éditables, modérables, post-éditées, modérées, ou exclues...) et sur les systèmes de TA (lesquels sont souhaités, avec quels paramètres, avec quelles préférences entre systèmes pour chaque couple de langues, avec quels *diagrammes de traduction* pour chaque couple aussi...), appelle l'iMAG-S avec les paramètres adéquats.

### 4.1 CAHIER DES CHARGES

Les logiciels principaux, initialement exploités par les passerelles iMAG, sont réimplémentés dans le cadre du projet Traouiero. Si, à la fin du projet Traouiero, la réimplémentation n'est pas terminée, les cahiers des charges et les spécifications de chaque module sont précisés. Il s'agit principalement de :

- SECTra-3, le système de gestion de corpus sur lequel s'appuie le dorsal,
- PIVAX++, le système de gestion de bases lexicales, intégrant aussi un dorsal,
- SegDoc, le segmenteur de documents et de pages Web, et SegNorm, quand il est cumulé avec une phase de normalisation,
- Prox-iMAG, le proxy dédié,
- TRADOH, le service REST d'appel de systèmes de traduction automatique,
- SANDOH, le service REST de détection pages de même langue et de même codage,
- le site de commerce en ligne permettant de souscrire aux offres et de définir le paramétrage associé.

Pour tous les modules à réimplémenter, on prévoit que, lorsqu'une identification est requise pour que l'utilisateur puisse y exploiter des fonctionnalités particulières, ces réimplémentations prennent en compte leur composition avec les autres services, et intègrent une connexion via un annuaire (LDAP).

Cependant, il faut aussi considérer l'intégration de modules externes pour lesquels l'authentification par LDAP n'est pas réalisée. De plus, les passerelles iMAG visant à offrir un service d'accès multilingue aux contenus de sites Web dédiés à des clients dans des « iMAG » (interactive Multilingual Access Gateway), certains de ces sites intègrent déjà leur propre méthode d'identification. Le même annuaire peut être partagé, mais les droits doivent être relayés d'un service à l'autre pour minimiser le nombre d'authentifications demandées à l'utilisateur.

Les ressources linguistiques constituées lors de l'exploitation du service iMAG étant dédiées, les clients demandent que les informations qu'elles contiennent soient protégées.

Le relais doit réaliser 2 fonctions différentes clairement identifiées :

- assurer une authentification centralisée, évitant au maximum qu'un utilisateur n'ait à se reconnecter, ou alors, en exprimant clairement pour quel service il doit le faire.
- spécifier et transmettre les droits des utilisateurs d'un service à l'autre.

## 4.2 SPÉCIFICATIONS EXTERNES

Le relais sert de point d'accès principal, de bus entre les services, et centralise la gestion des authentifications des utilisateurs.

Fonctions que doit implémenter le relais :

- **do(\_service, \_login, \_pass, <paramètres pour le service appelé>)** : où *\_login* et *\_pass* identifient l'utilisateur, et *\_service* est le service à appeler. Le reste des paramètres est transmis à ce dernier.

Fonctions que doivent implémenter tous les services:

- **what(\_service)** : où *\_service* est le service à appeler. Cette fonction doit rendre la liste des informations que peut demander le relais sur les paramètres d'un service donné.
- **what(\_service, \_login, \_pass,)** : où *\_login* et *\_pass* identifient l'utilisateur, et *\_service* est le service à appeler. Cette fonction doit rendre la liste des informations que peut demander le relais sur les paramètres des droits d'un utilisateur par rapport à un service.

L'exploitation d'un annuaire comme LDAP pour tous les nouveaux composants développés dans l'équipe GETALP est la méthode retenue pour l'avenir. Il n'en reste pas moins que certains composants à intégrer peuvent ne pas être reliables à un annuaire. De plus, comme les composants n'ont pas été initialement conçus pour être exploités en services Web, les réponses obtenues dépendent des droits des utilisateurs, et il faut relayer les sessions client-serveur en client-serveur-serveur, tout en préservant la sécurité des données.

La définition d'un relais (relais-iMAG au départ, relais plus général dans la suite) a donc pour objectif d'externaliser la gestion des droits des utilisateurs, par encapsulation dans un service dédié à la négociation des ces droits entre des composants logiciels et des services.

Cette phase consiste à produire les documents contenant les spécifications fonctionnelles, et les spécifications détaillées décrivant les besoins et choix effectués pour le développement du relais. Il s'agit dans un premier temps de délimiter le périmètre de l'application, en le concrétisant dans des scénarios.

Le relais sert de point d'accès principal, de bus entre les services, et centralise la gestion des authentifications des utilisateurs.

Fonction :

- **do(\_service, \_login, \_pass, <paramètres pour le service appelé>)** : où *\_login* et *\_pass* identifient l'utilisateur, et *\_service* est le service à appeler. Le reste des paramètres est transmis à ce dernier.

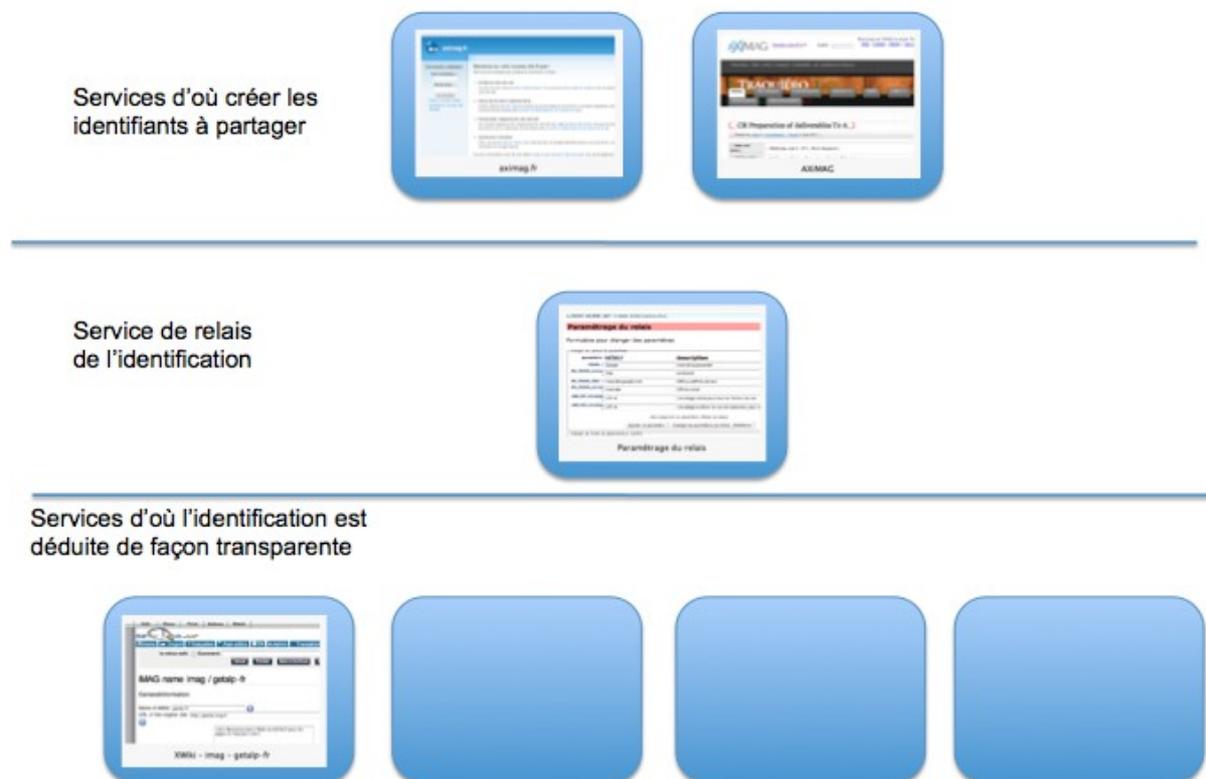
## 4.3 SPÉCIFICATION INTERNE

Avant la réalisation de cette tâche, la spécification interne du relais était dans un état très préliminaire. Le service demandé était directement appelé en lui transmettant ses paramètres, indépendamment de toute vérification de permission.

La conséquence était que, par programme, en réalité, tout était permis. Cela était suffisant dans un le cadre d'un premier cycle de développement pour commencer à tester l'intégration des autres composants. Mais, dès que l'ont veut offrir des services qui garantissent l'intégrité des données, et donc, dès que l'on est engagé vis-à-vis de clients, il n'est pas réaliste de ne rien sécuriser, et il faut mettre en place un outil pour relayer les droits.

Typiquement, pour chaque demande reçue par un service S1 en provenance d'un autre service S2, le relais négocie les permissions pour consulter ou contribuer dans S1 depuis S2, selon les paramétrages définis par le client dans l'offre commerciale qu'il a contractée.

Du point de vue de l'utilisateur, la navigation entre les différents services minimise les demandes de connexion.

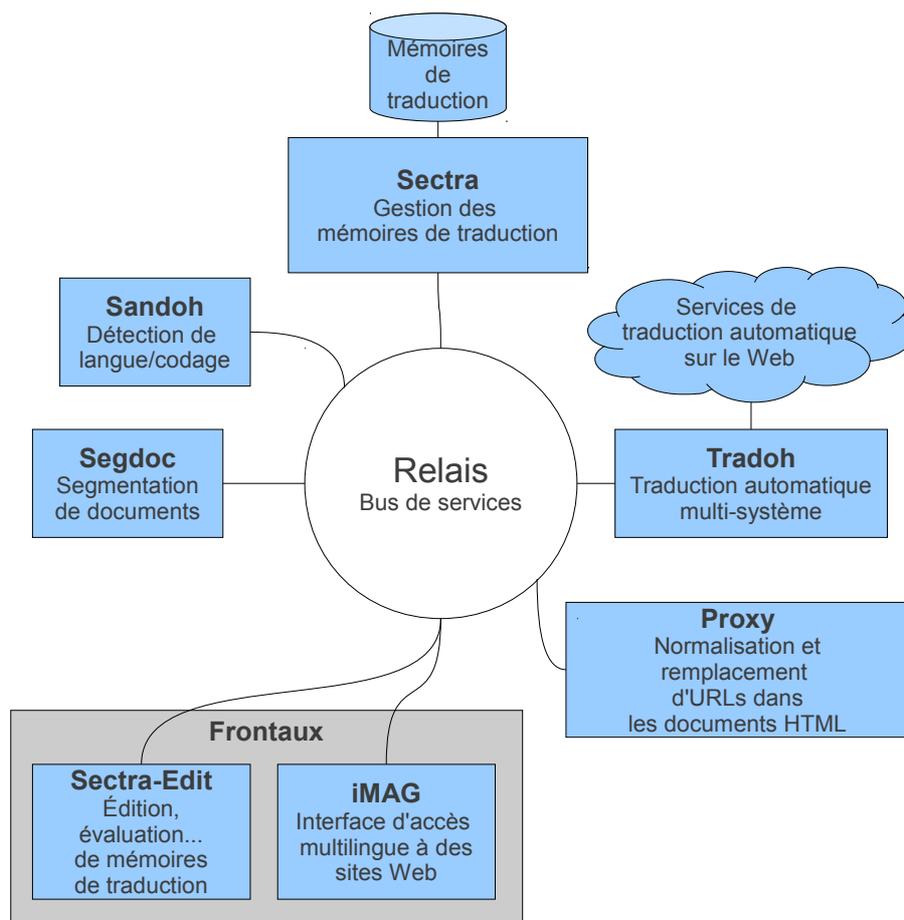


**Figure 13 : services reliés par le relais**

#### 4.3.1 ARCHITECTURE GÉNÉRALE

Dans la réimplémentation des logiciels exploités dans les passerelles iMAG, pour l'intégration des modules, nous privilégions un bus interne de service, qui correspond au relais-iMAG prévu dans la sous-tâche 3.1. Ce bus central offre ainsi un intérêt multiple :

- il permet une optimisation des communications entre modules ;
- il permet une centralisation de l'authentification et des droits des utilisateurs des différents modules, et donc une *authentification unique* des utilisateurs, qui n'ont ainsi pas besoin de créer de multiples comptes ;
- il ouvre la voie à la sous-tâche 3.2 (tableau\_blanc-iMAG, agents, tâches en boucles infinies), et sur le long terme il ouvre la voie à l'utilisation de langages de requêtes narratifs évolués.



**Figure 14: architecture des nouveaux modules.**

#### 4.3.2 SCÉNARIOS

Le scénario décrit les actions d'un utilisateur ayant besoin d'identifications multiples dans les services utilisés. Deux scénarios sont étudiés, selon le type d'utilisateur : un client qui souscrit à une offre et configure ses services, et un contributeur, employé par une société cliente, qui consulte une page Web de son entreprise par iMAG.

##### 4.3.2.1 Scénario Client

Dans ce scénario, un nouveau client a choisi une offre depuis le site de commerce en ligne et configure les services auxquels cette offre lui donne droit. Il s'agit donc du relais des informations de XXXXXXXXXXXXXXXXXXXX

##### 4.3.2.2 Scénario Contributeur

Dans ce scénario, un utilisateur francophone souhaite accéder en français à une page Web d'un site chinois élu<sup>5</sup>. Dans un premier temps, la page est prétraitée et segmentée. On vérifie ensuite si ces segments sont déjà présents dans la mémoire de traduction, et le cas échéant on récupère les meilleures traductions. Puis on traduit automatiquement les segments pour lesquels aucune traduction préexistante n'a été trouvée. On dispose alors d'une traduction pour chaque segment. Il ne reste plus alors qu'à reconstituer la page avec les segments traduits. Ce scénario est détaillé dans la Figure 15.

**Figure 15 : processus d'accès à une page Web dans une langue différente de celle du site**

<sup>5</sup> C'est à dire d'une page qui a été abonnée au service iMAG.

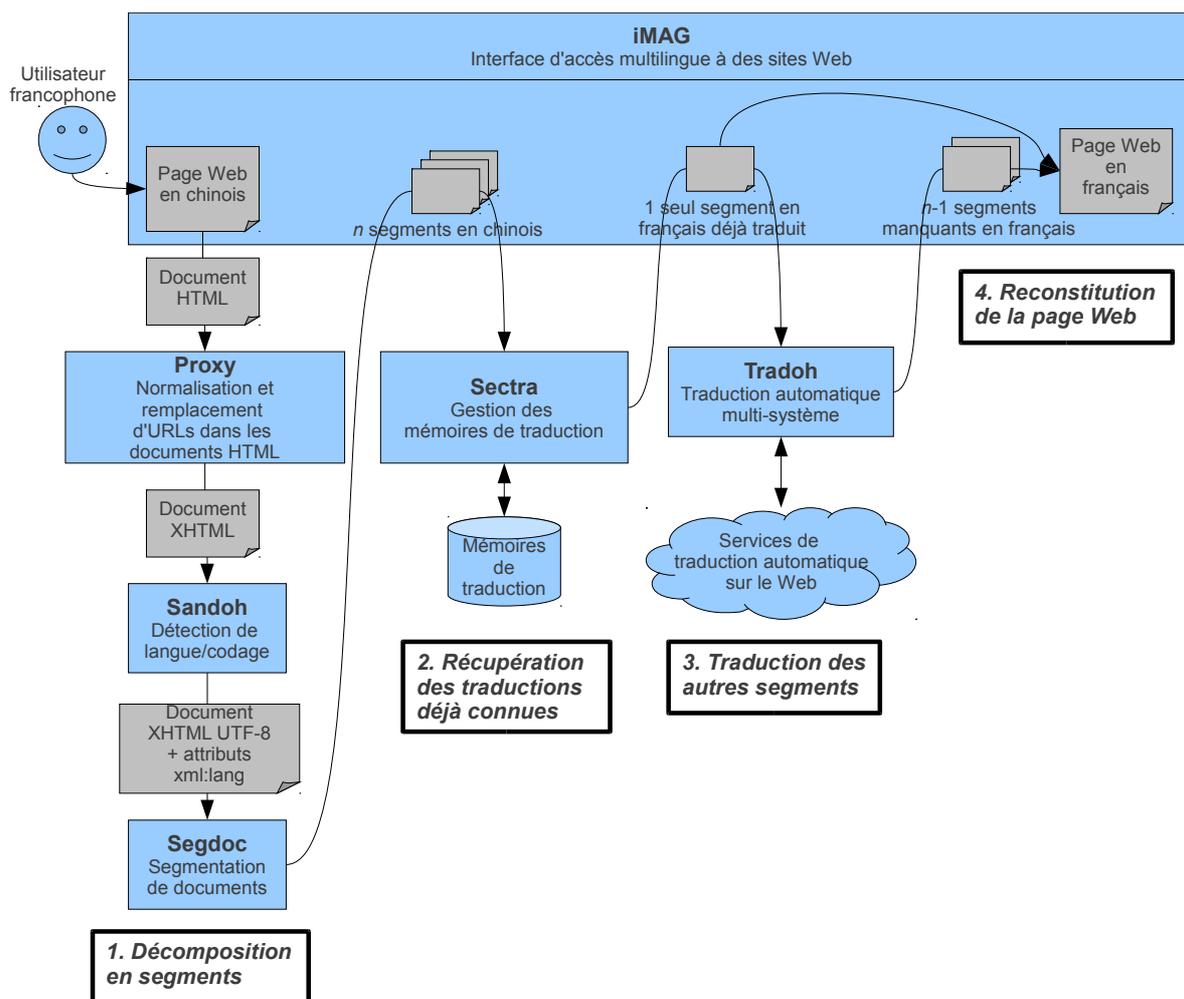


Figure 16: les 4 principales étapes du traitement d'une page Web par une iMAG.

#### 4.3.3 CORRESPONDANCE DES DROITS DES UTILISATEURS ENTRE SERVICES

Une nouvelle approche est envisagée pour gérer le transfert des droits : il s'agit de s'appuyer sur un environnement de transfert à pivot, avec les mêmes principes que pour les pivots lexicaux.

Chaque utilisateur a un identifiant unique porté par une URI. Le logiciel TopBraid Composer permet de définir des services de transfert des propriétés de chaque utilisateur d'un service à l'autre, de manière explicite, à travers ce pivot.

Le logiciel TopBraid Live permet de générer un service REST associé à chaque transfert.

L'intérêt de l'approche est que la désignation des transferts peut évoluer et que le nombre de services peut augmenter sans que cela ne remette en cause alors l'architecture globale et le protocole de communication entre les services.

## 5. DORSAL DE MODÉRATION

Ce volet a été ajouté à la sous-tâche 3.1 « relais et iMAG », de la tâche 3, « iMAG++ » pour gérer de manière uniforme et externe l'attribution des contributions dans les ressources linguistiques exploitées dans les services des passerelles iMAG.

Le but de ce composant logiciel est de fournir un outil générique pour la gestion des rôles des utilisateurs, l'attribution des contributions et la gestion de la modération, utilisable par les différents services impliqués dans les passerelles iMAG, ainsi que par d'autres applications multiservices.

Cette section présente le cahier des charges, et les considérations techniques étudiées pour aller vers des spécifications externes puis internes du composant à développer ou à adapter.

### 5.1 CAHIER DE CHARGES

Il s'agit de définir l'aspect "association des contributions aux utilisateurs" (qui n'est pas géré dans le noyau de SECTra3 alors que complètement intégré à SECTra2), en intervenant minimalement sur le modèle conceptuel du noyau SECTra3, de façon à :

- séparer les différents métiers collaborant dans l'exploitation du logiciel,
- minimiser les risques d'effet de bord,
- développer des fonctions réexploitables dans d'autres logiciels,
- étudier la généralisation à Jibiki-Pivax.

La modération et sa gestion doivent être configurables. Le gestionnaire d'un service iMAG doit pouvoir choisir parmi différents modes de modération, comme c'est le cas dans SECTra2. Ces modes sont définis en spécifiant :

- le niveau de modération, parmi 3 choix d'implication d'un modérateur pour publier les post-éditions proposées :
  1. publication immédiate de toute contribution, proposée par n'importe quel utilisateur,
  2. publication immédiate des contributions dans l'état « en progrès », si elle a été proposée par un utilisateur agréé ou par un modérateur,
  3. publication des contributions « avec modération », de sorte qu'elles sont visibles pour leurs auteurs seulement tant qu'elles n'ont pas été validées par un modérateur,
- le niveau de sécurité, parmi 4 choix plus ou moins ouverts sur l'autorisation de n'importe qui à contribuer :
  1. par n'importe quel visiteur,
  1. par n'importe quel visiteur enregistré et connecté,
  2. seulement par les utilisateurs spécifiés dans la configuration de cette iMAG,
  3. seulement par certains utilisateurs spécifiques du groupe des administrateurs qui peuvent gérer le service AXiMAG.

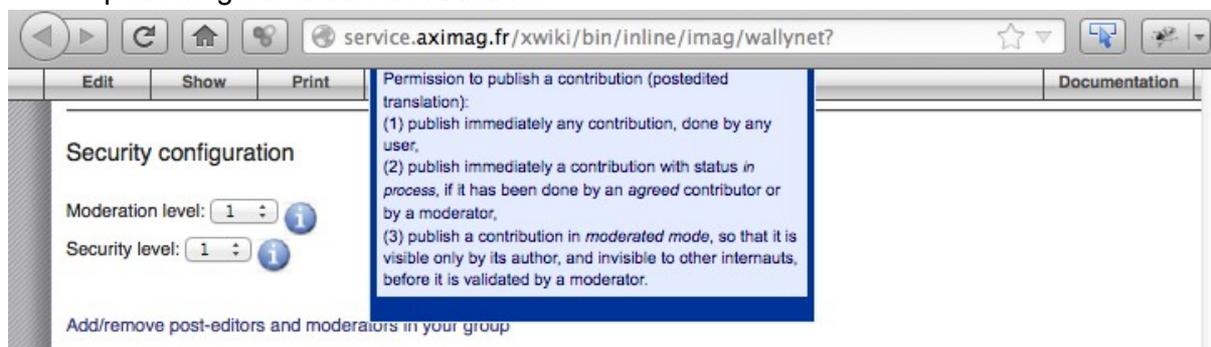


Figure 17 : configuration du niveau de modération dans SECTra2

Le gestionnaire d'un service iMAG doit aussi pouvoir attribuer des droits à des utilisateurs, Dans SECTra2, ces droits sont associés à une mémoire de traductions et non pas à une iMAG : si une mémoire de traductions est partagée par plusieurs iMAG, les droits des utilisateurs enregistrés sont les mêmes sur toutes les iMAG qui la partagent. Ces droits sont attribués pour chaque couple de langues que le service peut supporter, et non pas seulement pour les langues concernées par une iMAG, ni même par toutes les iMAG partageant la même mémoire de traductions.

UserName	Translational level	Default note	demo translator	demo moderator
Xan	★★★★★	13	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Euroloweri	★★★★★	10	<input type="checkbox"/>	<input type="checkbox"/>
KALITVIANSKI	★★★★★	10	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Wingel Zhang	★★★★★	10	<input type="checkbox"/>	<input type="checkbox"/>

**Figure 18 : gestion des droits de contributeurs enregistrés dans SECTra2**

Pour réaliser ces modes, les modules-vues (de la nouvelle architecture) exploitant les noyaux pour lesquels l'attribution des contributions doit être gérée, doivent pouvoir prendre en compte l'état d'édition des données. Par exemple, le cahier des charges de l'interface de post-édition des mémoires de traductions, défini dans la thèse de Hung Cong Phap, comme un éditeur adapté à la post-édition collaborative et contributive (Chapitre II, §3.1.2.3 et §3.2.2), précise que :

« L'éditeur doit permettre à plusieurs contributeurs de travailler simultanément sur la même collection de données (segments, pages, ou documents). Pour cela, l'éditeur doit gérer et contrôler les modifications des contributeurs. Si un segment est en cours de post-édition, le système doit changer son état par une marque spéciale pour le distinguer des autres segments, et ce segment doit être bloqué temporairement pour empêcher la post-édition par d'autres post-éditeurs.

SECTra\_w doit également gérer les versions des post-éditions, mais il assure que la version suivante est l'amélioration de la précédente et que la version finale est en principe la meilleure. »

source	translation	author	time	ver.	status	action
Read more...	Lire la suite ...	lambda	2012-11-15 16:01:22.0	3	complete	edit
Home	Accueil	lambda	2012-11-15 14:40:28.0	3	complete	edit
Atmospheric plasma, a new opportunity for surface processing	Le plasma à pression atmosphérique, une nouvelle voie pour la préparation des surfaces	Thomachot	2012-12-08 10:18:34.0	2	complete	edit
As a worldwide provider of turnkey plasma treatment solutions, equipment and services, we welcome the opportunity to work with you to	Nous fournissons des solutions clé-en-main de traitement plasma, des équipements et des services, partout dans le monde, et serons heureux de travailler avec vous pour	Thomachot	2012-11-24 22:45:51.0	2	complete	edit

**Figure 19 : dorsal de modération de SECTra-2**

Ainsi, le dorsal de modération doit-il permettre de définir le mode de modération à appliquer, et d'attribuer des droits aux utilisateurs enregistrés.

Il doit aussi publier l'état d'édition des éléments qu'il gère et permettre aux services qu'il utilise de bloquer temporairement l'édition des éléments et de la débloquer ensuite : il doit donc fournir une API.

## 5.2 CONSIDÉRATIONS TECHNIQUES

### 5.2.1 QU'EST-CE QUE LA PROGRAMMATION PAR ASPECTS ?

La programmation par aspect ou Programmation Orientée Aspect (POA) est un paradigme de programmation dans lequel les concepteurs tendent à traiter les préoccupations métier (c'est-à-dire du cœur de l'application) séparément des préoccupations transversales (en anglais *cross-cutting concerns*).

Cela consiste à déterminer et concevoir une méthode uniforme pour ajouter à un logiciel des qualités d'un certain type qui peuvent être fonctionnelles ou non fonctionnelles.

L'architecture logicielle de SECTra3 correspond exactement aux situations où la programmation par aspects est adaptée, puisque c'est au noyau que sont réservées les préoccupations liées à la gestion des corpus, et que les modules de vues fournissent des exploitations spécialisées de corpus, donc avec des métiers déjà un peu différents. La préoccupation de la gestion des utilisateurs ne fait partie d'aucun de ces métiers, ni la gestion de l'historique des contributions et de leur attribution aux utilisateurs.

Pour éviter l'entrecroisement des dépendances entre modules, la technique sur laquelle repose la programmation par aspects consiste à externaliser les aspects dans des composants développés à un autre niveau d'abstraction. Ce sont des points d'insertion (en anglais *joinpoint*) qui établissent la liaison entre l'aspect et le code métier ou un autre aspect.

« *Un point de jonction est un point dans le flot de contrôle d'un programme dans lequel un ou plusieurs aspects peuvent être appliqués.* »

D'un point de vue pratique, la programmation par aspect s'appuie sur des outils similaires aux modificateurs du langage à objets CommonLISP (*before*, *after*, *around*), complétés par des descriptions d'insertions déclaratives. Avec ces techniques et les outils associés, on peut spécifier les changements nécessaires à l'introduction d'un nouvel aspect, sans toucher aux sources du code original, dont la logique reste intacte.

Un aspect permet donc de spécifier :

- les points d'action (*pointcut*), qui définissent les points de jonction satisfaisant aux conditions d'activation de l'aspect, donc le ou les moments où l'interaction va avoir lieu,
- les greffons, c'est-à-dire les programmes (*advice*) qui seront activés avant, autour de ou après les points d'action définis.

Cette technique semble parfaitement adaptée, mais son application conduit à une génération automatique de code pour gérer les appels croisés aux points de jonction, dans le code source. Cette étape de *tissage de code* peut être réalisée de manière statique ou de manière dynamique (lors de l'exécution du logiciel). Mais quelle que soit la manière, cette étape rend difficile à analyser les programmes dans les phases de mise au point. De plus, elle ne peut pas encore être mise en œuvre avec le langage PHP, ce qui n'est pas compatible avec notre architecture hétérogène.

Nous avons choisi de ne pas utiliser directement la programmation par aspects, mais de nous en tenir à l'idée générale :

« *Un aspect est une entité logicielle qui capture une fonctionnalité transversale à une application.* »

Notre approche va consister à minimiser les incursions dans les codes du noyau, et à uniformiser les modifications et les restreindre à des appels à des codes externes. Concernant la base de données, nous allons définir une base de données dédiée aux traitements spécifiques nécessaires au métier de l'aspect, et n'ajouter que des clés externes

dans la base de données du noyau quand il le faut, même si cette migration de clé se fait pour une association 1-1 du modèle conceptuel global.

## 5.2.2 ETUDE DES PARTIES DES BASES DE DONNÉES POUR GÉRER LES ATTRIBUTIONS

### 5.2.2.1 Situation dans mediaWiki

Les différents rôles d'utilisateurs sont assez détaillés dans mediawiki.

[http://fr.wikinews.org/wiki/Wikinews:Statuts\\_des\\_utilisateurs](http://fr.wikinews.org/wiki/Wikinews:Statuts_des_utilisateurs)

Citation : « *Les statuts officiels sont ceux délivrés par la communauté et astreignant à des devoirs et des droits supplémentaires les utilisateurs qui les reçoivent. On peut en distinguer deux sortes : les statuts officiels techniques offrant de nouvelles fonctionnalités et les statuts officiels non techniques qui sont des engagements ne nécessitant pas d'outils spéciaux.* »

Selon le paramétrage de mediaWiki, la définition des rôles des utilisateurs permet d'affecter à leur groupe la permission d'accéder à certaines fonctionnalités. Par exemple, dans l'instance de MediaWiki utilisée par le site <http://fr.wikinews.org> :



L'**utilisateur anonyme** peut, selon le paramétrage de Mediawiki, modifier un contenu sans avoir de compte utilisateur. Le logiciel mémorise cependant son adresse IP qui identifie ses contributions.



L'**utilisateur enregistré** (ayant créé un compte et s'étant identifié pour visiter le site) peut contribuer et renommer des pages. Il dispose d'une page personnelle. Le logiciel utilise son pseudo pour attribuer ses contributions.



L'**administrateur** (ou **sysop** dans Mediawiki, pour *system operator*, en anglais) peut protéger/déprotéger des pages, supprimer/restaurer des pages...



Le **bureaucrate** peut changer le statut d'un utilisateur normal, jusqu'au rôle de bureaucrate, mais il ne peut diminuer ses propres droits.

Le **steward** peut modifier tous les statuts des utilisateurs.



Le **bot** (abréviation de robot) n'est pas un utilisateur réel, mais il est repéré parce qu'il fait plus d'actions par minute qu'un humain ne pourrait le faire. Ses actions n'apparaissent pas par défaut dans la liste des modifications récentes. Il faut cliquer sur 'montrer robots' pour voir ces modifications. Par défaut, les robots ne sont pas autorisés. Ils peuvent être utiles par exemple pour effectuer des corrections (fautes d'orthographe) en masse, ils peuvent être déclarés pour être autorisés le cas échéant.



Le **développeur** a accès à la base de données et peut effectuer les opérations de maintenance depuis le système du serveur. Il n'est pas pour autant **administrateur développeur**.



L'**administrateur développeur** peut, en plus de l'administrer, bloquer le site entier en cas de vandalisme majeur.



Le **vérificateur d'adresses IP** (*checkuser*) peut associer des comptes enregistrés (pseudonymes) avec des adresses IP utilisées précédemment pour des contributions anonymes.

Le **réviseur** peut valider des modifications pour qu'elles soient affichées dans des versions stables, et le **contributeur** voit ses contributions automatiquement validées. Ces rôles sont fournis par l'extension *FlaggedRevs* de Mediawiki.

Le **masqueur** de modifications peut cacher, de manière permanente, une version déterminée d'un projet Mediawiki.

L'**utilisateur banni** est un ancien utilisateur qui n'a plus le droit de créer et éditer les pages, de façon permanente. Il s'agit d'une décision exceptionnelle.

Ces rôles sont définis dans une matrice booléenne à 2 dimensions : (rôle, droit fonctionnel). Un droit fonctionnel est défini par une fonctionnalité et sa description.

Par exemple « import » caractérise la possibilité d'importer des contenus d'un coup, venant d'autres wiki ; « edit » caractérise la possibilité d'éditer des contenus non protégés. Une clé est définie par le vecteur de toutes les fonctions attribuées à un rôle, c'est à dire pour lesquelles la valeur du booléen est « vrai ».

La liste des rôles peut être étendue simplement, en ajoutant un groupe d'utilisateurs, et en renseignant sa colonne dans la matrice. De la même manière, de nouvelles fonctionnalités peuvent être prises en compte.

Quelle est l'information mémorisée dans les tables de la base de données qui ont la charge d'assurer les fonctions du métier de Mediawiki ? (c'est-à-dire de la gestion des contenus d'un ensemble de pages Web se référant mutuellement.) La gestion du multilinguisme peut-être considérée comme un autre métier, qui n'est pas le métier principal de Mediawiki.

#### 5.2.2.2 Situation dans SECTra2 via xWiki

Le choix initial de l'architecture basée sur xWiki pour développer le logiciel SECTra, était d'exploiter la richesse de la gestion des rôles d'un Wiki.

La liste des rôles a été clarifiée et est décrite dans le document de génie logiciel : <http://aximag.ligforge.imag.fr/GL/iMAG-CDC/Destination/html/tdu.html>

La Figure 20 représente les rôles des utilisateurs et les fonctionnalités auxquelles ils ont accès.



Figure 20 : rôles des utilisateurs dans SECTra2

### 5.2.2.3

### 5.2.2.4 Situation dans SECTra3

## 5.2.3 DÉLÉGATION ET ASYNCHRONISME

La délégation d'un service à un autre conduit à réaliser des tâches de manière asynchrone, ce qui impose de gérer le flux de données (jetons ou tickets de travaux). La technique proposée consiste à maîtriser la gestion des queues.

## 5.3 SPÉCIFICATIONS INTERNES

### 5.3.1 BASE DE DONNÉES DE SECTra2

D'après le cahier des charges de SECTra, définit dans la thèse de HUYNH Cong Phap, (Chapitre II, §3.1.2.6), les données sont organisées selon un modèle hiérarchique pour pouvoir unifier le traitement et la gestion des corpus. Chaque projet de post-édition a un objectif précis (évaluation, post-édition, production d'annotations, traduction de qualité...), et dispose d'une base de données lexicales (en 2 parties : générale, liée au domaine, et spécialisée, liée au projet<sup>6</sup>) et d'une mémoire de traductions.

<sup>6</sup> Il ne suffit pas de spécialiser cette 2ième base lexicale au corpus, car deux projets différents peuvent faire des choix terminologiques différents (exemple : les traductions en français de l'Unix d'IBM, AIX, sont faites différemment par IBM-France et par Bull, chacun voulant utiliser sa propre terminologie).

Chaque projet travaille sur un ou plusieurs corpus, éventuellement sur certaines parties de chaque corpus, et/ou sur certaines parties de chaque document. Par exemple, un projet pourrait s'occuper de la traduction en français des parties « claims » d'un corpus de brevets en télécommunications.

Dans les corpus de structure simple, un document est l'image en html d'un document Word ou LaTeX classique : un fichier principal (.html) et des fichiers satellites dans un répertoire associé (images, feuilles de style, etc.)

Dans les corpus de structure un peu plus complexe, comme un livre, un document peut être formé d'une hiérarchie de documents simples ou complexes, et il peut être muni de documents « compagnons ». C'est le cas de l'encyclopédie EOLSS, où une bonne partie des articles (600 sur 6000) est actuellement accompagnée de fichiers .unl contenant une segmentation et la représentation UNL de chaque segment.

Dans les corpus de structure encore plus complexe, comme les corpus de dialogues oraux interprétés (à 2 ou N langues), il y a au moins un fichier son (et le plus souvent aussi un fichier texte) par tour de parole, ainsi qu'un document principal décrivant la suite temporelle des tours de parole, et contenant des métadonnées (temps de départ, de fin, locuteur...).

Notons que certains corpus, comme ceux associés à des sites Web ou à des livres de phrases, contiennent un grand nombre de « métasegments », c'est-à-dire de générateurs de segments usuels par instanciation de variables.

Par exemple : « Mis à jour le \$date\_jj/mm/aaaa » ou « Vous êtes le \$ord\_X-ième visiteur » ou « Donnez-nous \$nb1 tasse[s] de \$boisson et \$nb2 \$pâtisserie [s] » Chaque document peut être présenté en plusieurs pages logiques, qui ne sont autres que des « pseudodocuments » présentés avec un « squelette » de type tableau.

Chaque page logique est définie par un nombre quelconque de segments. Chaque segment source a un identifiant unique qui est utilisé pour le lier avec des occurrences (des traductions automatiques, des post-éditions, une liste de versions, un minidictionnaire, un graphe UNL, etc.).

### 5.3.2 BASE DE DONNÉES DE SECTra3

La base de données de SECTra3 est centrée sur le métier « gérer des corpus multilingues à textes multilingues alignés ». Les 5 entités principales identifiées et implémentées dans le modèle conceptuel sont listées ci-dessous et leur ont été rajoutées 2 entités de SECTra2 manquantes dans SECTra3 ainsi que des précisions sur la dernière entité (Traduction) :

RAPPEL  
L2.2.b

Nouveau  
et rappel  
SECTra2

RAPPEL  
L2.2.b

Nouveau  
et rappel  
SECTra2

RAPPEL  
L2.2.b

Nouveau  
et rappel  
SECTra2

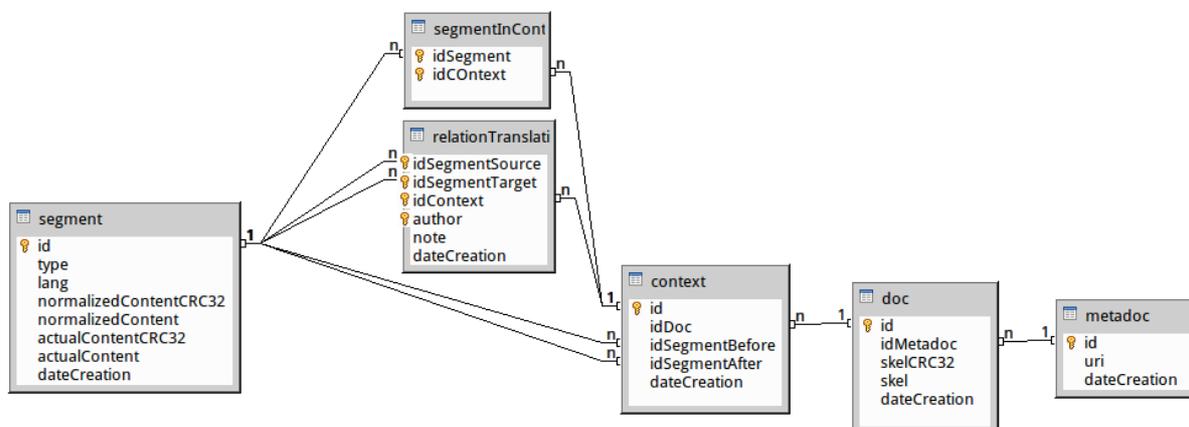
- **Méta-document** : il s'agit d'une ressource documentaire identifiée exclusivement par son *origine*, c'est à dire concrètement par une unique URI. Par définition (RFC 3986), une URI est soit une URL, c'est à dire une ressource identifiée par un chemin d'accès, soit une URN, c'est à dire une ressource identifiée par un nom. Un méta-document peut s'instancier dans plusieurs versions ou variantes : des *documents*.
- **Document** : une instance de méta-document qui se caractérise par son *contenu* et par son *type* (actuellement, uniquement HTML). Si par exemple deux fichiers correspondent au même méta-document et ont le même contenu, alors on considère qu'il s'agit de deux instances d'un même document.
- **Pseudodocument** : la représentation d'instances successives d'un document qui varie avec le temps, comme une page Web, soit parce que son contenu est modifié par ses auteurs, soit parce que certains segments sont en fait des métasegments (du genre « mis à jour le \$date »).
- **Segment** : un segment représente une unité textuelle de base pour la traduction, dans une langue donnée. Un métasegment est un segment récursif (phrases à trous ou plus exactement phrases avec variables), et est d'un type donné. Un segment contient du texte brut, et éventuellement des versions normalisées de ce texte, qui peuvent être utilisées pour rechercher des similarités entre segments.
- **Contexte** : il s'agit d'une notion complexe. Pour l'instant, cette notion est définie pour un segment comme l'intersection d'un document et des segments voisins (précédent et suivant) dans ce document. Un segment peut apparaître dans plusieurs contextes, et un contexte peut incorporer plusieurs segments (par exemple phrases à trous où le « trou » peut correspondre à plusieurs segments). La notion de contexte ne peut pas être définie pour les traductions (multiples par essence) comme pour les sources (une occurrence a un seul prédécesseur et un seul successeur).
- **Segment multilingualisé multilingue** : c'est un objet contenant un segment « source » primaire, au sens précédent, une ou plusieurs traductions automatiques pour une ou plusieurs langues, objets, comme des arbres linguistiques, des graphes UNL, des résultats d'évaluation(s), et des références aux contributions ayant produit chaque objet non primaire
- **Traduction** : dans une version antérieure, on a écrit qu'il s'agit d'une relation entre deux segments, dans un certain contexte, et avec un certain auteur.

C'est malheureusement faux car trop simpliste. Une traduction, même bonne, peut être en compétition avec plusieurs autres. Si cela se produit pour 2 ou 3 segments (source) de suite, cela n'a pas de sens de parler par exemple du contexte avant d'une traduction, car le choix d'un pronom dépend de sa référence anaphorique, et donc d'un choix précis de la traduction du segment source précédent.

Si l'on veut avoir une représentation en « réseau corporal », comme on le fait pour des « réseaux lexicaux », il faudra donc spécialiser la notion de « segment tout court », inutilisable telle quelle, en « segment source » et « segment cible » ou « traduction segmentale ».

On pourra alors considérer qu'un segment source est un nœud d'un graphe, relié Cette relation peut être notée. Des segments liés entre eux par despar une relation (directe ou indirecte) relations de traduction à des segments cibles, et que le sous-graphe obtenu forme exactementnt un *segment multilingualisé contextualisé*. Cette sorte deDu point de vue des traducteurs et des représentations « à la TMX », il faut cependant considérer un segment multilingualisé contextualisé comme un objet « super-segment » identifié comme tel, peut avoir des uneà structure complexe, qui reflète son et muni d'un historique proprehistorique.

La structure logique de la base de données de SECTra3, illustrée dans la Figure 21, ne prend en considération l'utilisateur que dans la table provenant de l'association plusieurs à plusieurs entre les segments et leurs contextes. La stratégie de programmation choisie pour développer SECTra3 prévoit des modifications, en particulier en ce qui concerne les colonnes de ses tables.

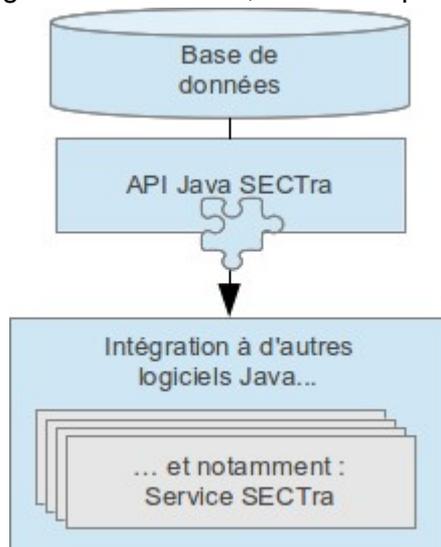


**Figure 21 : structure logique d'une base de données de corpus multilingue**

### 5.3.2.1 Architecture

Cette base de données n'est jamais accédée en dehors de SECTra lui-même. SECTra expose une couche d'abstraction objet, l'*API Java SECTra*, pour manipuler les mémoires de traductions. Cette couche présente en termes d'objets et d'héritage les concepts décrits dans la spécification fonctionnelle, et peut être intégrée à tout programme écrit en Java.

La *API Java SECTra* est utilisée pour créer une troisième couche, *Service SECTra*, qui permet d'utiliser SECTra en ligne de commande, et de la déployer en tant que service Web.



**Figure**

**Figure 22 : couches de SECTra\_3**

### 5.3.2.2 Structure de l'API Java SECTra

L'API est structurée suivant le même modèle conceptuel que la base de données.

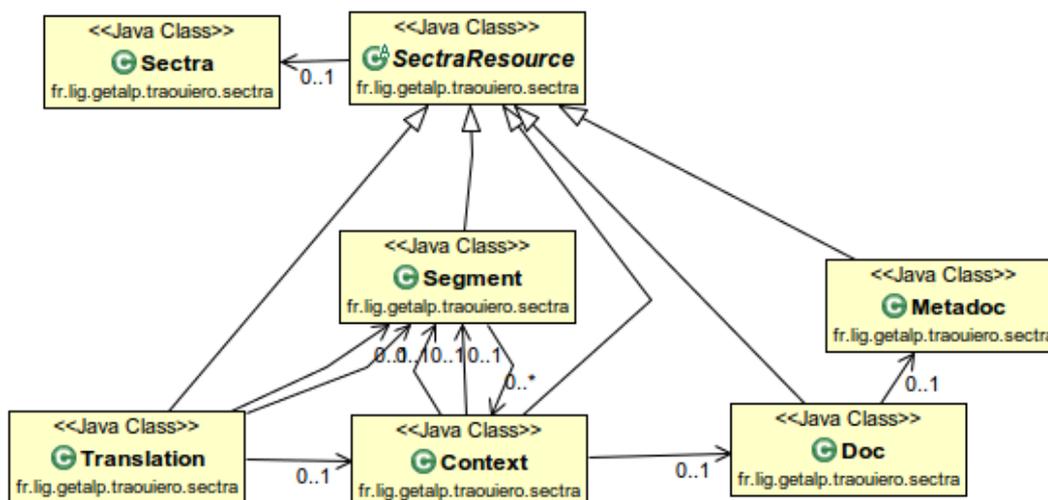


Figure 23 : diagramme UML de l'API Java

Toutes les classes correspondant à quelque chose en base de données (Segment, Translation, Context, Doc, Metadoc) dérivent de la super-classe *SectraResource*. Un corpus (classe *Sectra*) contient essentiellement un ensemble de *SectraResource*.

## 6. ETAT D'AVANCEMENT

Cette section présente une vue globale de l'état d'avancement de la sous-tâche T3.1, sur la réalisation des logiciels iMAG, du relais, et du dorsal de modération, à T0+24.

Module	Version	Cahier des charges	Spécif. externes	Spécif. internes	Implémentation		
					Version α Développement, tests unitaires	Version β Intégration, validation, débogage	Version RC <sup>7</sup> Tests de montée en charge
SECTra-Edit	1.0	100%	100%	0%	0%	0%	0%
SECTra-Test	1.0	100%	0%	0%	0%	0%	0%
iMAG	3.0	100%	100%	100%	100%	80%	0%
SECTra-Export	1.0	100%	50%	50%	20%	0%	0%
SECTra-Import	1.0	100%	20%	0%	0%	0%	0%
Proxy	1.0 (pages publiques, contenu statique)	100%	100%	100%	100%	100%	0%
	1.1 (cookies, authentification)	100%	100%	0%	0%	0%	0%
Xmlise	1.0	100%	100%	100%	80%	80%	0%

Tableau 1 : Etats d'avancement des programmes à T0+24

<sup>7</sup> RC : Release Candidate

## 7. RÉFÉRENCES

### 7.1 RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] **Boitet C., Huynh C.-P., Blanchon H. & Nguyen H.-T. (2009)** *A Web-oriented System to Manage the Translation of an Online Encyclopedia Using Classical MT and Deconversion from UNL*. Proc. RIVF 2009, Da Nang, 13-17/7/09, IEEE, P. Bellot ed., 8 p. (extended & updated from eponym CI-2008 paper)
- [2] **Huynh C.-P., Boitet C. & Blanchon H. (2008)** *SECTra\_w : an Online Collaborative System for Evaluating, Post-editing and Presenting MT Translation Corpora*. Proc. LREC-08, Marrakech, 27-31/5/08, ELRA/ELDA ed., 8 p.
- [3] **Nguyen H.-T. (2009)** *Des systèmes de TA homogènes aux systèmes de TAO hétérogènes*. Thèse, GETALP, LIG, UJF, Grenoble-1, 230 p.

### 7.2 RÉFÉRENCES WEBGRAPHIQUES

- [1] Programmation orientée aspect en Java avec AspectJ :  
<http://skebir.developpez.com/tutoriels/java/aspectj/>
- [2] Trousse à outil pour intégrer la traduction collaborative de documents, Google translator toolkit :  
<http://translate.google.com/toolkit?hl=fr>  
<http://support.google.com/translate/toolkit/answer/147809?hl=fr>
- [3] *Formulaire de traduction en ligne de Google, google-translate* :  
<http://translate.google.fr/?q=trousse+%C3%A0+outil&oe=utf-8&hl=fr&sl=fr&tl=en>
- [4] *Conversion d'URL relatives en absolues, par un script PHP* :  
<http://nadeausoftware.com/node/79>
- [5] *Conseils pour développer des site Web avec xWiki : au sujet des langages de script* :  
<http://platform.xwiki.org/xwiki/bin/view/DevGuide/Scripting>

- 
-

## 8. ANNEXES

### 8.1 CODE DE LA SEGMENTATION ADAPTÉE À SECTRA3

```

/**
 * Dans un document HTML, identifie des segments
 * et les remplace par leur n° de référence dans le corpus spécifié
 * lors de la construction de l'objet Segdoc.
 * Le résultat est un squelette de document.
 * @param docURI      URI identifiant le document
 * @param docContent  contenu du document
 * @param defaultLang langue par défaut du document
 * @param skel        un StringBufferWriter contenant le squelette et
 *                    permettant de suivre sa construction grâce à un flux
 *                    (cf. la définition de StringBufferWriter)
 * @return le squelette du document passé en entrée
 */
private String segHTML (URI docURI, String docContent, String defaultLang,
StringBufferWriter skel) {
    // Échapper les caractères utilisés comme délimiteurs de segment
    docContent = docContent.replaceAll("<"+openChar,
                                     "&#" + (int)openChar + ">");
    docContent = docContent.replaceAll(">"+closeChar,
                                     "&#" + (int)closeChar + ">");

    // Segmenter
    List<Segment> segments = new ArrayList<Segment>();
    //StringBufferWriter skel = new StringBufferWriter
    (docContent.length(), tempResultWriter);
    segHTMLv3(docContent, defaultLang, skel, segments);

    // Créer les (meta)docs
    Metadoc metadoc = new Metadoc(sectra, docURI);
    Doc myDoc = new Doc(sectra, metadoc, skel.toString());

    // Créer les contextes
    for(int i=0; i<segments.size(); ++i) {
        Context ct = new Context(sectra, myDoc,
                                i-1>=0?segments.get(i-1):null,
                                i+1<segments.size()?segments.get(i+1):null);
        segments.get(i).addContext(ct);
    }

    return skel.toString();
}
/**
 * Réalise la segmentation en "paragraphes" (càd blocs de texte+balises_faibles
 entre 2 balises fortes) d'un document HTML.
 * Appelle Segtext pour une segmentation fine (en phrases).
 */
private void segHTMLv3 (String docContent, String defaultLang, StringBufferWriter
skel, List<Segment> segments) {

    ...
    // Segmentation fine
    String newParagraphSkel = Segtext.segHTML (defaultLang,
Segtext.DEFAULT_SEGMENTATION, prunedContent[1], segments, new Properties(), this);

    skel.append(newParagraphSkel);
    // : Dans le squelette, remplacer le contenu du segment par une variable
    ...
}

```

```

public static String segHTMLWithTikal(String defaultLang, String text,
List<Segment> segments, Properties options, Segdoc segdoc) {
    // Écrire le texte à traiter dans un fichier temporaire

    // Lancer Tikal
    SRXSegmenteur srxSeg = new SRXSegmenteur();

    for(String segText : srxSeg.SRXsegment(text, defaultLang, null)) {
        // Pour chaque segment
        Segment seg = new Segment(segdoc.sectra, segdoc.type, defaultLang, segText);
        // : Nouveau segment
        segments.add(seg);
        text = text.replace(segText, segdoc.forgeSegTag(seg));
    }

    return text;
}

```

## 8.2 CODE DU RÉ-HABILLAGE ADAPTÉ À SECTRA3

```

public void wear(URI docURI, String skel, StringBufferWriter wearredSkel, String
targetLang) {
    // Créer les (meta)docs
    Metadoc metadoc = new Metadoc(sectra, docURI);
    Doc myDoc = new Doc(sectra, metadoc, skel);

    // Récupérer les segments et leur contexte
    List<Segment> segments = new ArrayList<Segment>();
    List<String> textBeforeSegment = new ArrayList<String>();
    //List<String> textAfterSegment = new ArrayList<String>();
    Matcher matcher = Pattern.compile("[^"+closeChar+"]*" + openChar
+ "SEG:([0-9]+)" + closeChar /*+ "[^"+openChar+"]*" */).matcher(skel);

    while(matcher.find()) {
        long idSeg = Long.parseLong(matcher.group(2));
        segments.add(new Segment(sectra, idSeg));
        textBeforeSegment.add(matcher.group(1));
        //textAfterSegment.add(matcher.group(3));
    }
    matcher = Pattern.compile(closeChar
+ "[^"+openChar+"]*$").matcher(skel);
    matcher.find();
    String textAfterLastSegment = matcher.group(1);

    // Remplacer les segments
    for(int i=0; i<segments.size(); ++i) {
        Segment seg = segments.get(i);
        Segment newSeg = null;
        Segment segBefore = i-1>=0 ? segments.get(i-1) : null;
        Segment segAfter = i+1<segments.size() ? segments.get(i+1) : null;
        Context ct = new Context(sectra, myDoc, segBefore, segAfter);
        // : Constuire le contexte
        //System.out.println("-"+seg.getLang()+"-"+targetLang);
        if(targetLang.equals(seg.getLang())) {
            // Le segment d'origine est déjà dans la bonne langue
            newSeg = seg;
        }
        else { // Sinon, chercher une traduction
            Translation trad = seg.getBestTranslation(targetLang, ct);
            if(trad == null) { // Pas trouvé de traduction

```

```

        newSeg = new Segment(sectra, type, targetLang,
Tools.tradoh(tradohURL, seg.getContent(), seg.getLang(), targetLang));
        seg.addTranslationTarget(newSeg, ct, "tradoh", (byte)1);
    }
    else {
        newSeg = trad.getSegmentTarget();
    }
}
wearedSkel.append(textBeforeSegment.get(i));
// : Ajouter le texte situé avant le segment
wearedSkel.append(newSeg.getContent());
//skel = skel.replaceFirst("" + openChar + "SEG:([0-9]+)"
                        + closeChar, newSeg.getContent());

}
wearedSkel.append(textAfterLastSegment);

wearedSkel.flush();
//return wearedSkel.toString();
}

```

### 8.3 EXTRAIT DE LA CLASSE GROOVY.TRANSLATION QUI GÈNÈRE LE BANDEAU

```

/**
 * Hong-Thai Nguyen
 * 8/02/2010
 * Générer la page d'entête à mettre dans un frame.
 */
String generateHeaderPage(String imagHomePath, String scriptPath, String
cssPath, String imagePath, String originalURL, String originalTitle,
Hashtable language_code, String langInterface, String url_base, String sl,
String tl, String user, boolean isAdmin, boolean isPostEditor, String
projectName, String corpusName, String HTMLSlogan, boolean showSource,
boolean isShowSeperator) {
    generateHeaderPage(imagHomePath, scriptPath, cssPath, imagePath,
originalURL, originalTitle, language_code, langInterface, url_base, sl, tl,
user, isAdmin, isPostEditor, projectName, corpusName, HTMLSlogan,
showSource, -1, isShowSeperator);
}

String generateHeaderPage(String imagHomePath, String scriptPath, String
cssPath, String imagePath, String originalURL, String originalTitle,
Hashtable language_code, String langInterface, String url_base, String sl,
String tl, String user, boolean isAdmin, boolean isPostEditor, String
projectName, String corpusName, String HTMLSlogan, boolean showSource, int
numberOfFreelancerContribution, boolean isShowSeperator) {

    String hostnamebase = "http://service.aximag.fr/";
    String langset = getLanguageSet(language_code, sl);
    String str = "";
    str = str + "<html>";
    str = str + " <head>";
    str = str + "    <meta http-equiv=\"content-type\" content=\"text/html; charset=UTF-
8\"/>";

```

```

    str = str + "    <link rel=\"stylesheet\" href=\"" + cssPath +
"imag_css_init.css" + "\"/>";
    str = str + "    <link rel=\"stylesheet\" href=\"" + cssPath +
"imag_css_color.css" + "\"/>";
    str = str + "    <script type=\"text/javascript\" src=\"" + scriptPath +
"imag_interaction.js" + "\"></script>";
    str = str + " </head>";
    str = str + " <body bgcolor=\"#ffffff\" text=\"#000000\" link=\"#0000cc\"
vlink=\"#551a8b\" alink=\"#ff0000\" dir=\"ltr\">";
    str = str + "    <div style=\"padding:6px 8px; white-space:nowrap;\">";
    str = str + "        <div style=\"float:left;\">";
    str = str + "            <a href=\"" + imagHomePath + "\" target=\"_top\">";
    str = str + "                <img src=\"" + imagePath + "imag_logo.png" + "\"
alt=\"AXiMAG\" style=\"padding:0;border:0\" width=\"150px\"/>";
    str = str + "            </a>";
    str = str + "        </div>";
    str = str + "        <div style=\"float:top; font-size:12px;text-align:center\">";
    str = str + HTMLSlogan;
    str = str + "        </div>";
    str = str + "        <div style=\"float:left;\">";
    str = str + "            <table cellpadding=\"0\" cellspacing=\"3\" border=\"0\">";
    str = str + "                <tr>";
    str = str + "                    <td style=\"font-size:12px;\">";
    if (isAdmin || isPostEditor)
        str = str + "                        <a href=\"" + originalURL + "\" title=\"Click
here to see the original page\" target=\"_blank\"> " + originalTitle + " (" +
corpusName + ")</a>";
    else
        str = str + "                            <span> " + originalTitle + "</span>";
    str = str + "                            <span>in :</span>";
    str = str + "                                generateLanguageMenu (url_base,
originalURL, language_code, sl, tl, showSource);
    str = str + "                                </td>";
    str = str + "                                <td style=\"padding:1px 8px; float:right; clear:right;
font-size:12px;\">";
    /*
    str = str + "                                <!--div id=\"imag_contribution_summary\"
style=\"display:none;\">";
    str = str + "                                <span style=\"padding-right:2px;padding-
left:6px\">Translated <span class=\"imag-number-segments\">2</span> segments by:</span>";
    str = str + "                                <span class=\"imag-target-machine-withcolor\">Google
MT(1) </span>";
    str = str + "                                <span class=\"imag-target-benevole-withcolor\">Free-
lancer(1) </span>";
    str = str + "                                <span class=\"imag-target-translator-
withcolor\">Translator(0)</span>";
    str = str + "                                <span class=\"imag-target-translation_memory-
withcolor\">Translation Memory(0)</span>";
    str = str + "                                <span class=\"imag-target-postedition-
withcolor\">Staff(0)</span>";
    str = str + "                                </div-->";
    */
    if (isShowSeperator)
        str = str + "                                <span id=\"imag_contribution_summary_mark\"
style=\"display:none;\" isShow=\"1\"/>";

```

```

else
    str = str + "                <span id=\"imag_contribution_summary_mark\"
style=\"display:none;\" isShow=\"0\"/>";
    str = str + "                </td>";
    str = str + "            </tr>";
    str = str + "        </table>";
    str = str + "    </div>";
    str = str + "    <div align=\"right\" id=\"guser\" style=\"font-size:84%;padding:0 0
4px\" width=\"100%\">";
    str = str + "        <nobr><a href=\"/xwiki/bin/view/imag/Help\"
target=\"_top\">Help</a> | <a href=\"/xwiki/bin/view/imag/Contact?xredirect=" + url_base + "\"
target=\"_top\">Contact</a> ";

//Admin
if (isAdmin) {
    str = str + "| <a href=\"#\"
onClick=\"window.open('/xwiki/bin/view/imag/imagSetting?projName=" + projectName +
"&langset=" + langset + "&lp=" + sl + "_" + tl + "&xredirect=" + url_base + "',
'Setting the post-editing right',
'height=700,width=900,scrollbars=yes,resizable=yes');\">Admin</a> ";
}
if (isPostEditor) {
    // Last change
    if (numberOfFreelancerContribution >= 0) {
        String titleLastChange = "";
        if (numberOfFreelancerContribution > 1)
            titleLastChange = numberOfFreelancerContribution + " in processing
contributions in this page";
        else
            titleLastChange = numberOfFreelancerContribution + " in processing
contribution in this page";
        str = str + "| <a title=\"" + titleLastChange + "\"
href=\"/xwiki/bin/view/imag/Log?projName=" + projectName + "&corpusName=" + corpusName
+ "&langset=" + langset + "&lp=" + sl + "_" + tl + "&xredirect=" + url_base +
"&onlyInprocess=inprocess\" target=\"_top\">Last change (<span style=\"color:orange;\">" +
numberOfFreelancerContribution + "</span></a> ";
    } else
        str = str + "| <a href=\"/xwiki/bin/view/imag/Log?projName=" + projectName +
"&corpusName=" + corpusName + "&langset=" + langset + "&lp=" + sl + "_" + tl +
"&xredirect=" + url_base + "&onlyInprocess=inprocess\" target=\"_top\">Last change</a> ";
    // Your contribution
    str = str + "| <a href=\"/xwiki/bin/view/Corpus/Statistic?projName=" + projectName
+ "&dbName=" + projectName.toLowerCase() + "&docName=" + "_ALL_" + "&user_name="
+ user + "&source_lang=" + getLanguageName(sl).toLowerCase() + "&target_lang=" +
getLanguageName(tl).toLowerCase() + "\" target=\"_top\">Your contribution</a> ";
}

// Register-Login or user_name-Logout
if (user.equals("XWiki.XWikiGuest") || user.equals("")) {
    str = str + "| <a href=\"/xwiki/bin/view/imag/Register\" target=\"_top\">Register</a>
";
    str = str + "| <a href=\"/xwiki/bin/login/XWiki/XWikiLogin?xredirect=" +
hostnamebase + url_base + "?u=" + originalURL + "&sl=" + sl + "&tl=" + tl +
"\" target=\"_top\">Log in</a> ";
} else {
    String displayUser = user;

```

```

    int pos = displayUser.indexOf("XWiki.");
    if (pos >= 0)
        displayUser = displayUser.substring("XWiki.".length(),
displayUser.length());
    str = str + "| <a href=\"/xwiki/bin/view/XWiki/" + displayUser + "\"
target=\"_top\">\" + displayUser + "</a> ";
    str = str + "| <a href=\"/xwiki/bin/logout/XWiki/XWikiLogout?xredirect=" +
hostnamebase + url_base + "?u=" + originalURL + "&sl=" + sl + "&t1=" + t1 +
"\" target=\"_top\">Log out</a> ";
}

//Copyright
str = str + "| <a href=\"/xwiki/bin/imag/Copyright\" target=\"_top\">Copyright</a>";
str = str + "</noabr>";
str = str + "    </div>";
str = str + "    </div>";
if (!sl.equals(tl))
    str = str + "    <div style=\"padding:1px 8px; float:right; clear:right; font-
size:12px;\">";
    else
        str = str + "    <div style=\"display:none; padding:1px 8px;
float:right; clear:right; font-size:12px;\">";
String isShowCheck = "";
if (isShowSeperator)
    isShowCheck = "checked=\"true\"";
str = str + "    <input type=\"checkbox\" " + isShowCheck + "
id=\"checkbox_reliability\" name=\"checkbox_reliability\" onClick=\"changeColor();\"
/>Reliability (<a href=\"#\" onclick=\"showLegend();\" title=\"Click here to see legend of
contribution type\"></a>");
str = str + "    <input type=\"radio\" id=\"ratio_translation_only\"
name=\"ratio_translation_only\" onclick=\"changeFrameTarget(1);\"
checked=\"true\"/>Translation only";
str = str + "    <input type=\"radio\" id=\"ratio_with_source\"
name=\"ratio_translation_only\" onclick=\"changeFrameTarget(2);\"/>Original";
str = str + "    </div>";
str = str + " </body>";
str = str + "</html>";
return str;
}

```

## 8.4 CODE VELOCITY DE L' API « À LA » REST POUR LA LISTE DES LANGUES CIBLES D'UNE IMAG

```

## based on Hierarchical navigation panel
## GIRARD Lionel
## 2011/12/28
##
http://extensions.xwiki.org/xwiki/bin/view/Extension/Hierarchical+Navigation+Panel
## extended to spaces
## BELLYNCK Valérie
## 2012/02/06

#set ($grvCommon=$xwiki.parseGroovyFromPage ("Groovy.Common"))
#set ($level_currentuser="#getlevel ($context.user) ")

```

```

#set ($rate_currentuser=$util.parseInt($level_currentuser.trim()))
#set ($grvTrans=$xwiki.parseGroovyFromPage("Groovy.Translation"))
##---
#set ($lang_list = $grvTrans.getAllStringLanguageNames())
#set ($helpAPI = "<br /> Project(projName - this is TM name),<br /> context
(context- this parameter is option),<br /> trace (for verbose- this
parameter is option), <br /> xpage=plain (for API-REST like).<br /> Example
: [http://service.aximag.fr/xwiki/bin/view/Test/indirect-space?
projName=wallynet&xpage=plain&trace |
http://service.aximag.fr/xwiki/bin/view/Test/indirect-space?
projName=wallynet&xpage=plain&trace]. <br /> Please contact
service.hotline@aximag.fr for more details" )

##--- on récupère les paramètres passés dans l'URL
#set ($projName=$request.projName)
#set ($isTrace = false)
#set ($isTrace = $request.trace) ## and $!
request.trace.trim().equalsIgnoreCase("true"))
##set ($trace=$request.trace)
##--- on récupère le DNS, le port et l'action
#set ($serverName=$request.getServerName())
#set ($serverPort=$request.getServerPort())
#set ($action=$context.getAction())
##----- si l'URL spécifiait bien un projet :
#if($request && $projName)
##-----
#set ($docName = "imag."+ $projName)
## test pour savoir si le nom du projet correspond bien à un document
xwiki...
#set ($doc = $xwiki.getDocument($docName))
#set ($docExist = $doc.isNew())
#set ($traceStr = 'TRACE : doc = '+$doc+' ; docExist = '+$docExist)
#if($isTrace) #warning($traceStr) #end
##----- si le nom du projet correspond bien à un doc xWiki :
#if(!$docExist)
##-----
## test pour savoir si le projet est une classe Basic d'imag...
#set ($class=$doc.getObject("imag.BasicClass").xWikiClass)
#set ($traceStr = 'TRACE : class = '+$class)
#if($isTrace) #warning($traceStr) #end
##----- si on a récupéré une classe :
#if($class)
##-----
#set ($obj=$doc.getObject("imag.BasicClass"))
#set ($simpleLink = "http://" + $serverName + ":" + $serverPort +
$doc.getValue("EWS_IMAG_URL", $obj).toString())
#set ($traceStr = 'TRACE : link = '+$simpleLink)
#if($isTrace) #warning($traceStr) #end
#set ($language_code_string=$doc.getValue("language_code", $obj).toString())
#set ($traceStr = 'TRACE : language_code_string = '+$language_code_string)
#set ($sl=$doc.getValue("EWS_SL", $obj).toString())
#set ($lsl=$grvTrans.getLanguageCode($sl).toString())
#set ($traceStr = 'TRACE : source language = '+$lsl)
#if($isTrace) #warning($traceStr) #end
##set ($langs=$grvTrans.splitLanguageStringArray($grvTrans.getLanguageCode (
$sl).toString())
#set ($langs=$grvTrans.splitLanguageStringArray($language_code_string)
#set ($langs_str = '')

```

```

foreach($lang in $langs)
    #set($langs_str = $langs_str + $lang.trim())
end
$langs_str
##set($menu= $grvTrans.generateLanguageMenu($simpleLink, "",
$language_code_string, $grvTrans.getLanguageCode($sl).toString(), true))
##$menu
##-----
##----- sinon : il n'y a pas de classe :
#else ## if($class)
    1 Config export - indirectSpace
    #set($strErr = "Not a Generic Class Sheet: ")
    #warning("$strErr$helpAPI")
    à faire au cas par cas si l'iMAG n'est pas une imag générique
end ## else($class)
##-----
##----- sinon : le nom du projet ne correspond pas à un doc xWiki :
#else ## if(!$docExist)
    1 Config export - indirectSpace
    #set($strErr = "Not such a doc in imag: ")
    #warning("$strErr$helpAPI")
end ## else(!$docExist)
##-----
##----- sinon : l'URL ne spécifiait pas de projet :
#else
    1 Config export - indirectSpace
    #set($strErr = "Please give parameters: ")
    #warning("$strErr$helpAPI")
end

```

## 8.5 CODE JAVASCRIPT EXPLOITANT L'API QUI FOURNIT LA LISTE DES LANGUES

## 8.6 CODE GROOVY QUI ADAPTE LES INTERACTEURS DU RÉSULTAT DE GOOGLE

Ce code est contenu dans l'article imag/Lib de SECTra2, qui contient aussi les fonctions qui réalisent les actions agissant sur la base de données.

```

/
*****
**
Replace all Google functions _tipon and _tipoff for showing/hiding panel(palette) by SECTra_w
functions
*****
****/
// Voici l'exemple de code HTML donné par Google (voir
sample_text_google_return_page_origine.html):
/*
<span onmouseover="_tipon(this)" onmouseout="_tipoff()"><span class="google-src-text"
    style="direction: ltr; text-align: left">Plus de 2 200 "marines" américains doivent
    arriver mardi 19 janvier en Ha•ti pour accélérer les opérations de sauvetage.</span>
    More than 2 200 U.S. marines are expected to arrive Tuesday, January 19 in Haiti to
    speed up rescue operations.</span>
<span onmouseover="_tipon(this)" onmouseout="_tipoff()"><span class="google-src-text"
    style="direction: ltr; text-align: left">Ils s&#39;ajoutent aux 1 200 soldats
    sur le terrain, qui contr™lent notamment l&#39;aéroport de Port-au-Prince.
    Puissance douce</span> They are in addition to 1 200 troops on the ground, including
    controlling the airport of Port-au-Prince. Soft Power</span>
*/

```

```

// Il faut transformer au (voir sample_text_google_return_page_nomalized.html):
/*
<span_imag_segment onmouseover="_tipon(this)" onmouseout="_tipoff()"><span class="google-src-
text"
    style="direction: ltr; text-align: left" author="Google" segmentid="123" project="imag"
docname="TEST2" sl="fr" tl="en">Plus de 2 200 "marines" américains doivent
    arriver mardi 19 janvier en Haïti pour accélérer les opérations de sauvetage.</span>
    More than 2 200 U.S. marines are expected to arrive Tuesday, January 19 in Haiti to
    speed up rescue operations.</span_imag_segment>
<span_imag_segment onmouseover="_tipon(this)" onmouseout="_tipoff()"><span class="google-src-
text"
    style="direction: ltr; text-align: left" author="XWiki.nguyenht" segmentid="435"
project="imag" docname="TEST2" sl="fr" tl="en">Ils s'ajoutent aux 1 200
    soldats sur le terrain, qui contrôlent notamment l'aéroport de
    Port-au-Prince. Puissance douce</span> They are in addition to 1 200 troops on the
    ground, including controlling the airport of Port-au-Prince. Soft Power</span_imag_segment>
*/
StringBuffer decoratePage(StringBuffer strBin, String EWS_CORPUS, String DocName, String
EWS_MTSystem, String MTGoogle_Alias, int EWS_MOD, String author, boolean isValidatedUser, String
url, String sl, String tl, Hashtable language_code, String pathCofigFile, String
isEditableDefault, String advancedModDefault, String isUpdateDefault, String
statusUpdateDefault, String tobeEditDefault, String EWS_BEGIN_TARGET_MARK, String
EWS_END_TARGET_MARK, String GroupName, String GroupMod, String GroupAdmin, String user_agent,
boolean isShowSeperator) {

    String span_imag_segment = "span_imag_segment";
    String span_imag = "span_imag";
    String span_imag_target_text = "span_imag_target_text";

    if (user_agent != null && user_agent.indexOf("msie") >= 0) { // IE, not accept non standard
tabs
        span_imag_segment = "span";
        span_imag = "span";
        span_imag_target_text = "span";
    }

    StringBuffer strB = strBin;
    //get marks envelopping a google's segment
    //TODO Ici, il faut passer au Segdoc (serveur de segmentation)
    String startSegmentMarkup = grvTrans.getProperties(pathCofigFile, "startSegmentMarkup");
    //<span onmouseover="_tipon(this)" onmouseout="_tipoff()">
    String HtmlstartSegmentMarkup = grvTrans.getProperties(pathCofigFile,
"HtmlstartSegmentMarkup"); //span
    String endSegmentMarkup = grvTrans.getProperties(pathCofigFile, "endSegmentMarkup");
    //</span>
    String endsourceMarkup = grvTrans.getProperties(pathCofigFile, "endsourceMarkup"); // class=
"google-src-text" style="direction: ltr; text-align: left;">

    //get marks envelopping a source in the segment
    String sourceMarkup = grvTrans.getProperties(pathCofigFile, "sourceMarkup"); // <span
class="google-src-text"

    // Ici, on garde cet analyse car on doit utilise le résultat, mais des qu'on a un autre
moyen, il faut y passer à SegWeb
    int i=1;
    int pos = strB.indexOf(startSegmentMarkup, 0);

    while (pos !=-1 ) {
        //Find source segment
        int pos_begin_source_tag = strB.indexOf(sourceMarkup, pos);
        int pos_begin_source_text = strB.indexOf(">", pos_begin_source_tag) + 1;
        int pos_end_source_text = grvTrans.findNextSourceEndSpan(strB, pos_begin_source_tag);
        // ArrayList segSourceResult = grvTrans.getTrueTextSourceSegment(strB, sourceMarkup,
pos_begin_source_tag); // Extraire le segment source dans <span>
        // String source_text = strB.substring(pos_begin_source_text,
Integer.valueOf(Integer.valueOf(segSourceResult.get(1)).intValue()));
        String source_text = strB.substring(pos_begin_source_text, pos_end_source_text);
        // source_text = grvTrans.cutUSGGoogle(source_text);

        String source_text_noAttributes =
grvHTMLFilter.getSourceTextWithNoAttributs(source_text);
        source_text_noAttributes = source_text_noAttributes.replaceAll("&", "&");

```

```

/*
  Il y a des segments spéciaux non-touchable !!!
  Pour la météo, c'est la date dans la page d'Agenda ou des segments dans annuaire
  // à faire: mettre une configuration
  <span class="monthview"> && <span class="dayNBview">
*/

boolean isSpecialSegment = false;
/*
  //TODO: s'il y a un segment non touchable, c'est ici

*/
String final_Sent_id = UpdateSourceTable(DocName, DocName+"_seg"+i,
source_text_noAttributes, language_code, sl);

int pos_begin_target_text = pos_end_source_text + endSegmentMarkup.length();
int pos_end_target_text = grvTrans.findNextTargetEndSpan(strB, pos_begin_target_text);
String target_text = strB.substring(pos_begin_target_text, pos_end_target_text);
String target_text_noAttributes =
grvHTMLFilter.getSourceTextWithNoAttributes(target_text);
target_text_noAttributes = target_text_noAttributes.replaceAll("&", "&");

// Attention: pour bien chercher dans la mémoire de traduction, on ne cherche que
l'information TEXTUELLE (squelette)
// car dans segment source peut contenir des balises: <a> avec
href="non_definie_dans_la_MT", <span>, <div> ... avec class="non_definie_dans_la_MT"
// Dans la mise à jour dans la mémoire de traduction, il faut supprimer les attributs
dans les balises (imag_translate.js).
// Ordre des recherche est suivant:
// - Mémoire de traduction
// - Post-édition
// - Traduction automatique par un système non-Google
// - Traduction automatique par Google

String postedition_text = "";
String postedition_author = "";
String postedition_rating = "";
String postedition_status = "";

String mt_segment_text = "";
String target_type = "machine";
String insert_target_segment = "";

// Verifier dans la table de post-édition
if (insert_target_segment.equals("")) {
  ArrayList result_postedition = getTranslation(DocName, i, (String)
language_code.get(tl).toLowerCase());
  if (result_postedition.size() > 0) {
    postedition_text = result_postedition.get(0);
    postedition_author = result_postedition.get(1);
    postedition_rating = result_postedition.get(2);
    postedition_status = result_postedition.get(3);

    if (!postedition_text.equals("")) {
      // System.out.println(postedition_text);
      postedition_text = grvHTMLFilter.decorateTargetText(postedition_text,
source_text);
      // System.out.println(postedition_text);
    }
    if (!postedition_text.equals(""))
      postedition_text = postedition_text.replaceAll("&", "&")
    if (fromXWiki.getUser(postedition_author).isUserInGroup(GroupName) ||
fromXWiki.getUser(postedition_author).isUserInGroup(GroupMod) ||
fromXWiki.getUser(postedition_author).isUserInGroup(GroupAdmin))
      target_type = "postedition";
    else
      target_type = "benevole";
  }
}

```

```

        if (!postedition_text.equals(""))
            insert_target_segment = postedition_text;
    } else {
        insert_target_segment = "";
    }
} // end if post-edition

// Chercher dans le mémoire de traduction global du site.
if (insert_target_segment.equals("")) {
    ArrayList result_translationmemory = getTranslationMemory(source_text_noAttributes,
sl, tl);
    if (result_translationmemory.size() > 0) {
        target_type = "postedition"; // Should be 'translation_memory'
        postedition_text = result_translationmemory.get(0);
        postedition_author = result_translationmemory.get(1);
        postedition_rating = result_translationmemory.get(2);
        postedition_status = result_translationmemory.get(3);

        if (!postedition_text.equals(""))
            postedition_text = grvHTMLFilter.decorateTargetText(postedition_text,
source_text);
        if (!postedition_text.equals(""))
            insert_target_segment = postedition_text.replaceAll("&", "&")
    } else { // end if translation memory
        insert_target_segment = "";
    }
}

// Vérifier la TA
// TODO: s'il existe un dico (ex. liste des noms propres), il faut intervenir à ce
niveau.
String mt_name=EWS_MTSYSTEM;
if (insert_target_segment.equals("")) {
    target_type = "machine";
    postedition_rating = 10;
    mt_segment_text = "";
    if (!EWS_MTSYSTEM.equals("Google")) {
        mt_segment_text = getMachineTranslation_text(DocName, i, (String)
language_code.get(tl).toLowerCase(), EWS_MTSYSTEM);
        if (mt_segment_text.equals("")) {
            // Use Google instead
            mt_segment_text = target_text;
            insert_target_segment = target_text;
            if (MTGoogleAlias != null && !MTGoogleAlias.equals(""))
                mt_name = MTGoogleAlias;
            else
                mt_name = "Google";
            /* very slow, call in back-end
            mt_segment_text = grvTrans.CallSysstranMTwithPost(source_text_noAttributes,
sl, tl, EWS_MTSYSTEM, DocName+"_seg"+sl+"_"+tl+"_"+i);
            if (!mt_segment_text.equals("")) {
                AddTargetTable(DocName, DocName+"_seg"+i, mt_segment_text, (String)
language_code.get(tl).toLowerCase(), EWS_MTSYSTEM);
            }
            */
        } else {
            mt_segment_text = target_text;
            insert_target_segment = target_text;
        }
        if (!mt_segment_text.equals(""))
            insert_target_segment = mt_segment_text;
        else {
            mt_segment_text = target_text;
            insert_target_segment = target_text;
        }
    } else {
        mt_segment_text = target_text;
        insert_target_segment = target_text;
    }
} // End if machine translation

```

```

        AddTargetTable(DocName, DocName+"_seg"+i, target_text_noAttributes, (String)
language_code.get(t1).toLowerCase(), "Google");

        String tobeEdit = tobeEditDefault; // Rôle à mettre quand l'utilisateur a fini la post-
édition.

        // Controle pour le niveau 3, non accepté la contribution bénévole pas encore validée.
        if (EWS_MOD >= 3 && postedition_status.equals("encours")) { // Si EWS_MOD == 3 et la
contribution non visualisée pour tous sauf l'auteur, prendre la TA.
            if (!postedition_author.equals("XWiki.XWikiGuest") &&
author.equals(postedition_author) && !postedition_text.equals("")) {
                insert_target_segment = postedition_text;
            } else {
                if (!mt_segment_text.equals(""))
                    insert_target_segment = mt_segment_text;
                else
                    insert_target_segment = target_text;
            }
        }
    }

    // Change class="google-source" End </span>
    int posTipOffEnd = -1;

    if (!pos_end_target_text.equals("</span>")) {
        posTipOffEnd = strB.indexOf("</span>", pos_end_target_text);
        if (posTipOffEnd >=0)
            strB.replace(posTipOffEnd, posTipOffEnd + "</span>".length(), "</" +
span_imag_segment + ">");
    }

    // Remplacer le texte en ligne cible
    if (isSpecialSegment)
        insert_target_segment = target_text;

    String classNameCSS = " class=\"imag-target-" + target_type;
    if (isShowSeperator)
        classNameCSS += "-withcolor\"";
    else
        classNameCSS += "\"";

    String targetTextTOReplace = "<" + span_imag + classNameCSS + ">" +
EWS_BEGIN_TARGET_MARK + "</" + span_imag + "><" + span_imag_target_text + ">" +
insert_target_segment + "</" + span_imag_target_text + "><" + span_imag + classNameCSS + ">" +
EWS_END_TARGET_MARK + "</" + span_imag + ">";
    strB.replace(pos_begin_target_text, pos_end_target_text, targetTextTOReplace);

    // Ajout l'infor au source segment
    String isEditable = isEditableDefault;
    if (postedition_status.equals("valide") || postedition_status.equals("complet") ||
postedition_status.equals("complete")) {
        if (isValidatedUser) { // Permettre seulement les traducteurs agréés et modérateurs de
changer la post-édition validée.
            isEditable = "true";
        } else {
            isEditable = "false";
        }
    }
}

String advancedMod = advancedModDefault;

String isUpdate = isUpdateDefault; // Mise à jour immédiatement sur
l'interface ?
String statusUpdate = statusUpdateDefault; // Status à afficher après la post-édition
if (isSpecialSegment) {
    isEditable = "false";
    isUpdate = "false";
}
}

```

```
String new_tag_source_repl= sourceMarkup + ' id="source' + i + ' ' + 'segmentid="' + i
+ ' " project="' + EWS_CORPUS + ' "' + ' docname="' + DocName + ' " sl="' +
grvTrans.getLanguageName(sl) + ' " tl="' + grvTrans.getLanguageName(tl) + ' " author="' +
postedition_author + ' " posteditor="' + author + ' " initMT="' + mt_name + ' " rating="' +
postedition_rating + ' " translation_type="' + target_type + ' " tobeEdit="' + tobeEdit + ' "
isEditable="' + isEditable + ' " advancedMod="' + advancedMod + ' " isUpdate="' + isUpdate + ' "
statusUpdate="' + statusUpdate + ' " currentStatus="' + postedition_status + '>';
strB.replace(pos_begin_source_tag, pos_begin_source_text, new_tag_source_repl);

if (posTipOffEnd >=0)
    strB.replace(pos, pos + startSegmentMarkup.length(), "<" + span_imag_segment + "
onmouseover=\"_tipon(this)\" onmouseout=\"_tipoff()\">");
// Segment suivant
pos = strB.indexOf(startSegmentMarkup, pos_end_target_text+1);
i++;
}

return strB;
}
```