



**HAL**  
open science

# Utilité, utilisabilité et amélioration du prototype EXXELANT

Emilie Charles

► **To cite this version:**

Emilie Charles. Utilité, utilisabilité et amélioration du prototype EXXELANT. Linguistique. 2011.  
dumas-00631025

**HAL Id: dumas-00631025**

**<https://dumas.ccsd.cnrs.fr/dumas-00631025>**

Submitted on 11 Oct 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Utilité, Utilisabilité et Amélioration du prototype **EXXELANT**

Nom : **CHARLES**  
Prénom : **Emilie**

UFR **LLASIC**

---

Mémoire de master **2 professionnel** - 20 crédits

Parcours : **Modélisation et Traitements Automatiques en Industrie de la Langue**

Sous la direction de **Claude Ponton et Virginie Zampa**

Année universitaire 2010-2011





# Utilité, Utilisabilité et Amélioration du Prototype **EXXELANT**

Nom : **CHARLES**  
Prénom : **Emilie**

UFR **LLASIC**

---

Mémoire de master **2 professionnel** - 20 crédits

Spécialité : **Modélisation et Traitements Automatiques en Industrie de la Langue**

Sous la direction de **Claude Ponton et Virginie Zampa**

Année universitaire 2010-2011

## Remerciements

“The only moral it is possible to draw from this story is that one should never throw the letter Q in a privet bush, but unfortunately there are times when it is unavoidable.”<sup>1</sup>

Je tiens tout d’abord à remercier Virginie, (je fais fi de cet injuste classement par ordre alphabétique et t’accorde pour une fois la première place), et Claude de m’avoir fait confiance en me proposant ce stage. Je les remercie également de leur patience et de m’avoir laissé faire la morte sans même me gronder. Mais surtout, merci pour les réunions hebdomadaires annulées, reportées, décalées qui m’ont permis de profiter de Morphée quelques heures de plus ; pour les pauses clope, les pauses café, les pauses café-clope, sans lesquelles ma productivité s’en serait trouvée grandement diminuée ; enfin, pour les relectures et corrections rapides et complètes, même en période de vacances. Je remercie Thomas, car il faut bien envisager l’éventualité que je n’ai pas l’occasion de le faire dans les remerciements d’une thèse. Merci donc pour l’accueil au coin de ton bureau, le prêt de mouchoirs (je m’engage à te les rendre au centuple), ToolBook, Dédé, Léo l’Escargot, les vannes, les conseils et les encouragements. Je n’oublie pas mes mamans, Nada et Sabrina, qui ont bien pris soin de leur fille, ni mes compagnes de galère, Léo, Audrey et Elodie avec qui j’ai pu ronchonner de concert. Un grand merci à celles qui nous ont sûrement le plus supportés, mon caractère de Grognoirs et moi, Anthéa, Julie et Lucie, avec qui je devrais sûrement être gentille pendant une petite centaine d’années en compensation. Merci à ‘Rnaud pour le bout de code sur TreeTagger, merci à Lain pour les impressions, Mum pour la titine et à tous ceux que j’oublie. Et enfin, merci à l’agence CIC d’Albertville de m’avoir fourni un poste et un bureau qui m’ont permis de boucler la rédaction de ce mémoire à temps.

En résumé: “So long and thanks for all the fish.”<sup>2</sup>

---

<sup>1</sup> Douglas Adams, *The Restaurant at the End of the Universe*, 1980, p237.

<sup>2</sup> Douglas Adams, *The Hitchhiker’s Guide to the Galaxy*, 1979.

# Sommaire

<b>REMERCIEMENTS.....</b>	<b>4</b>
<b>INTRODUCTION.....</b>	<b>6</b>
<b>1. PRESENTATION D'EXXELANT .....</b>	<b>7</b>
1.1. LE PROJET IDILL.....	8
1.2. L'EXEMPLIER EXXELANT.....	8
<i>Généralités.....</i>	<i>8</i>
<i>La base de données Idill.....</i>	<i>11</i>
<i>L'interface de recherche d'EXXELANT.....</i>	<i>13</i>
1.3. TRAVAIL A REALISER.....	17
<i>Authentification et refonte de l'interface.....</i>	<i>17</i>
<i>Export d'un texte ou d'une sélection de phrases.....</i>	<i>18</i>
<i>Enrichissement du corpus et modification du formulaire de recherche.....</i>	<i>18</i>
<i>Réflexion sur l'usage des traces.....</i>	<i>19</i>
<b>2. AMELIORATION DU PROTOTYPE.....</b>	<b>20</b>
2.1. REFONTE DE L'INTERFACE.....	21
2.2. EXPORT DE PHRASES ET DE TEXTE.....	23
2.3. DEVELOPPEMENT DU SYSTEME DE CONNEXION.....	25
<i>Fonctions administrateur.....</i>	<i>26</i>
<i>Fonctions enseignant.....</i>	<i>30</i>
<i>Modification de mot de passe.....</i>	<i>32</i>
2.4. ENRICHISSEMENT DU CORPUS.....	33
<i>Création de corpus et ajout de texte.....</i>	<i>34</i>
<i>Étiquetage morphosyntaxique automatique des textes.....</i>	<i>35</i>
<i>Balisateur manuel des erreurs.....</i>	<i>37</i>
<i>Modification du formulaire de recherche.....</i>	<i>39</i>
<i>Perspectives d'évolution.....</i>	<i>40</i>
<b>3. DE LA NECESSITE DES TRACES DANS L'OUTIL PEDAGOGIQUE MULTIMEDIA.....</b>	<b>42</b>
3.1. TRAÇAGE DE L'UTILISATION.....	43
3.2. A QUI SONT DESTINEES LES TRACES ?.....	44
<i>Le concepteur-développeur.....</i>	<i>44</i>
<i>L'enseignant.....</i>	<i>45</i>
<i>L'apprenant.....</i>	<i>45</i>
3.3. COMMENT REPRESENTER LES TRACES ?.....	46
<i>Représentation des traces à destination d'un concepteur-développeur.....</i>	<i>48</i>
<i>Représentation des traces à destination d'un enseignant.....</i>	<i>49</i>
<i>Représentation des traces à destination d'un apprenant.....</i>	<i>54</i>
3.4. LE CAS D'EXXELANT.....	56
3.5. CONCLUSIONS SUR LES TRACES D'UTILISATION.....	57
<b>CONCLUSION.....</b>	<b>58</b>
<b>BIBLIOGRAPHIE.....</b>	<b>59</b>
<b>TABLE DES ANNEXES.....</b>	<b>62</b>
<b>TABLE DES FIGURES.....</b>	<b>69</b>

## Introduction

Dans le cadre de ma deuxième année de master, j'ai effectué un stage professionnel de quatre mois, du 1<sup>er</sup> mars 2011 au 1<sup>er</sup> juillet 2011 au sein du laboratoire LIDILEM de l'université Stendhal Grenoble 3. Mon encadrement a été assuré conjointement par M. Claude Ponton et Mlle Virginie Zampa, maîtres de Conférences à l'université Stendhal et membres du LIDILEM.

Le laboratoire de LInguistique et DIidactique des Langues Étrangères et Maternelles (LIDILEM) est une équipe d'accueil rattachée à l'UFR LLASIC de l'université Stendhal Grenoble 3. Créé en 1987, il est constitué d'une cinquantaine de membres permanents et d'une centaine de jeunes chercheurs. Les domaines de recherche identifiés au sein du laboratoire sont nombreux : acquisition du langage, morphosyntaxe, sociolinguistique, didactique des langues, traitement automatique de la langue etc. Cette diversité en fait un pôle de recherche interdisciplinaire axé à la fois sur le langage en lui-même et sur l'apprentissage, l'enseignement et la transmission des langues.

Les trois principaux axes de recherche du LIDILEM sont :

- Descriptions linguistiques, Traitement Automatique des Langues, Corpus
- Sociolinguistique et acquisition du langage
- Didactique des langues, recherches en ingénierie éducative

Au cours de ce stage, j'ai eu à étudier l'utilité et l'utilisabilité d'un prototype d'outil pédagogique multimédia pour l'apprentissage du français et à y apporter des améliorations significatives. Un travail de recherche concernant les traces d'interaction dans un tel outil a également été réalisé.

**1.**

**Présentation d'EXXELANT**



## 1.1. Le projet Idill

Le réseau européen Kaleidoscope<sup>3</sup> regroupe des chercheurs de différentes disciplines à travers l'Europe. Son but est, grâce à cette collaboration, l'évolution de l'apprentissage assisté par les technologies (Technology-Enhanced Learning).

Le projet IDIIL<sup>4</sup> regroupe des chercheurs en linguistique, en didactique des langues et en traitement automatique des langues. Ces chercheurs proviennent du CENTre de Traitement Automatique du Langage (CENTAL<sup>5</sup>) et du CECL<sup>6</sup> (Centre for English Corpus Linguistics), deux pôles de recherche de l'Université Catholique de Louvain (Belgique), ainsi que du LIDILEM. Cédrick Fairon, Sylviane Granger, Georges Antoniadis, Olivier Kraif, Claude Ponton et Virginie Zampa font partie de ce projet. L'idée initiale est que « des données annotées linguistiquement couplées à des algorithmes puissants, rendent possible l'automatisation d'un grand nombre de processus utiles pour l'apprentissage des langues assisté par les technologies : lustrage de textes, détection d'erreurs et rétroaction, analyse discursive, évaluation des réponses de l'apprenant, génération d'exercices etc. »<sup>7</sup>.

## 1.2. L'exemplier EXXELANT

### *Généralités*

EXXELANT (EXample eXtractor Engine for LANguage Teaching) est un exemplier présenté sous forme de concordancier développé dans le cadre du projet IDILL. Cette application permet, grâce à une interface web développée en PHP liée à une base de données MySQL, d'effectuer des recherches sur un corpus de textes d'apprenants en Français Langue Étrangère (le corpus FRIDA-BIS, dérivé du corpus FRIDA (French Interlanguage DAtabase) (Granger 2001)) et d'en tirer des extraits correspondant aux critères spécifiés.

Les corpus d'apprenants, en lien avec des outils informatiques performants, ont un rôle primordial à jouer dans l'apprentissage des langues assisté par ordinateur, notamment

---

<sup>4</sup> <http://www.noe-kaleidoscope.org/group/idill/>

<sup>5</sup> <http://www.uclouvain.be/cental>

<sup>6</sup> <http://www.uclouvain.be/cecl.html>

<sup>7</sup> « Linguistically annotated data coupled with powerful algorithms make it possible to automate a wide range of processes that are particularly relevant for TEL: glossing of texts, error detection and feedback, discourse analysis, rating of learners' answers, exercise generation, etc. »

Source : <http://www.noe-kaleidoscope.org/group/idill/Home/>

si les erreurs sont annotées. L'étude de corpus d'apprenants peut permettre par exemple d'identifier des types d'erreurs récurrents et ainsi de concevoir des outils plus spécifiques, axés sur les difficultés de la langue en prenant en compte la langue maternelle de l'apprenant. Ces corpus peuvent également être utilisés par les enseignants comme matériau pédagogique pour la création d'activités, par les apprenants pour développer leurs capacités d'autoévaluation, ou encore par les chercheurs, pour penser des outils de détection automatique d'erreurs en fonction de motifs d'usage de la langue déterminés par une analyse empirique des données.

Le corpus FRIDA a été collecté dans le cadre du projet européen FreeText par Sylviane Granger, de l'Université Catholique de Louvain. Il rassemble des productions d'apprenants en Français Langue Étrangère, de langues maternelles et de niveaux différents. Une annotation et une proposition de correction manuelles des erreurs ont été effectuées sur ce corpus. Le domaine de l'erreur, la catégorie de l'erreur, la catégorie du mot erroné ainsi que la correction sont annotés à l'aide de triplets de balises.

```
<G><GEN><DEI>#Toutes$Tous</DEI></GEN></G> les choses concernant les  
employés seront <G><GEN><VSP>#décidées$décidés</VSP></GEN></G>
```

Figure 1 - Exemple d'annotation des erreurs dans le corpus FRIDA

Dans cet exemple, la première erreur est une erreur de grammaire (G), portant sur le genre (GEN) d'un adjectif indéfini (DEI) (voir Annexe 3 pour la signification des trois niveaux de balisage). La correction est annoncée par le caractère #, la forme erronée par le caractère \$.

Le corpus FRIDA a été récupéré dans le cadre du projet IDILL pour créer le corpus FRIDA-BIS. Afin de permettre une exploitation optimale des données, grâce notamment au traitement automatique de la langue (TAL), le corpus FRIDA a subi plusieurs modifications. Il a été nettoyé et son balisage standardisé au format XML. En effet, le système de balisage des erreurs présent dans le corpus FRIDA ne permettait pas de gérer correctement la présence de plusieurs erreurs sur un même élément. Ainsi il a été abandonné au profit d'un balisage XML plus propre, permettant un codage plus clair de plusieurs erreurs. Des métadonnées sur les textes ont été ajoutées, telles que la longueur des textes exprimée en nombre de mots et la densité d'erreur. Enfin, lemme et catégorie grammaticale des items des textes ont été identifiés de façon automatique grâce à

TreeTagger<sup>8</sup>, un outil permettant l'étiquetage morphosyntaxique et la lemmatisation de formes. Cet étiquetage a été enrichi, (informations sur le genre, le nombre etc.), à l'aide de scripts utilisant le dictionnaire gratuit de l'Association des Bibliophiles Universels<sup>9</sup> (ABU).

Le corpus FRIDA-BIS est constitué de 764 productions d'apprenant comportant 9466 phrases, 20474 erreurs et 179642 mots. En reprenant l'exemple donné précédemment, nous obtenons avec le nouveau balisage:

```
<ERR IDE="7" DOMI="G" CER1="GEN" CGRI="DEI">
<INI><tok id="t30" orth="Tous" base="tout" ctag="pro"
msd="indef">Tous</tok></INI>
<COR><tok id="t31" orth="Toutes" base="unknown" ctag="nom"
msd="">Toutes</tok></COR>
</ERR>
<tok id="t32" orth="les" base="le" ctag="det" msd="artic">les</tok>
<tok id="t33" orth="choses" base="chose" ctag="nom" msd="femin">choses</tok>
<tok id="t34" orth="concernant" base="concerner" ctag="ver"
msd="ppres">concernant</tok>
<tok id="t35" orth="les" base="le" ctag="det" msd="artic">les</tok>
<tok id="t36" orth="employés" base="employé" ctag="nom"
msd="mascu">employés</tok>
<tok id="t37" orth="seront" base="être" ctag="ver" msd="ifutu
pluri">seront</tok>
```

Figure 2 - Exemple d'annotation des textes dans le corpus FRIDA-BIS

Tous les mots, ou tokens, sont entourés par les balises `<tok></tok>`, qui ont pour attributs *orth*, correspondant à la forme, *base* au lemme, *ctag* à la catégorie grammaticale et *msd* aux traits morphosyntaxiques. Ces attributs découlent de l'étiquetage effectué grâce à TreeTagger et au dictionnaire d'ABU. Chaque erreur est encadrée par les balises `<ERR></ERR>`. Les informations concernant ces erreurs sont spécifiées dans les attributs de la balise `<ERR>`. Elles sont codées par le triplet domaine d'erreur, catégorie d'erreur et catégorie grammaticale du mot, respectivement *dom*, *cer* et *cgr*. Les erreurs peuvent

<sup>8</sup> <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

<sup>9</sup> <http://abu.cnam.fr/>

comporter jusqu'à quatre triplets les définissant. Le mot erroné est entouré par les balises <INI></INI> et la correction par les balises <COR></COR>.

Toutes les informations balisées ont été insérées dans une base de données; l'interface d'EXXELANT permet d'effectuer des requêtes sur cette base.

### ***La base de données Idill***

La base de données idill comporte six tables : *textes*, *phrases*, *tokens*, *erreurs*, *triplet*, *corrections*. Elle contient les textes du corpus FRIDA-BIS, découpés en tokens, et toutes les informations présentes dans le balisage

Les textes comportent plusieurs phrases. Chaque phrase est composée de tokens. Les champs *idtokd* et *idtokf* de la table *phrases* contiennent respectivement l'identifiant du premier token et du dernier token de la phrase. La table *tokens* contient tous les mots de tous les textes. L'affichage des phrases et des textes se fait grâce à cette table. Les tokens peuvent comporter une erreur ou non. Tous les tokens possèdent un identifiant d'erreur. Si le mot n'est pas erroné, cet identifiant est nul. Plusieurs tokens peuvent être associés à la même erreur. De la même façon, plusieurs corrections peuvent s'appliquer à la même erreur. La table *triplet* contient les informations sur les erreurs. Chaque erreur est définie par un triplet d'informations.

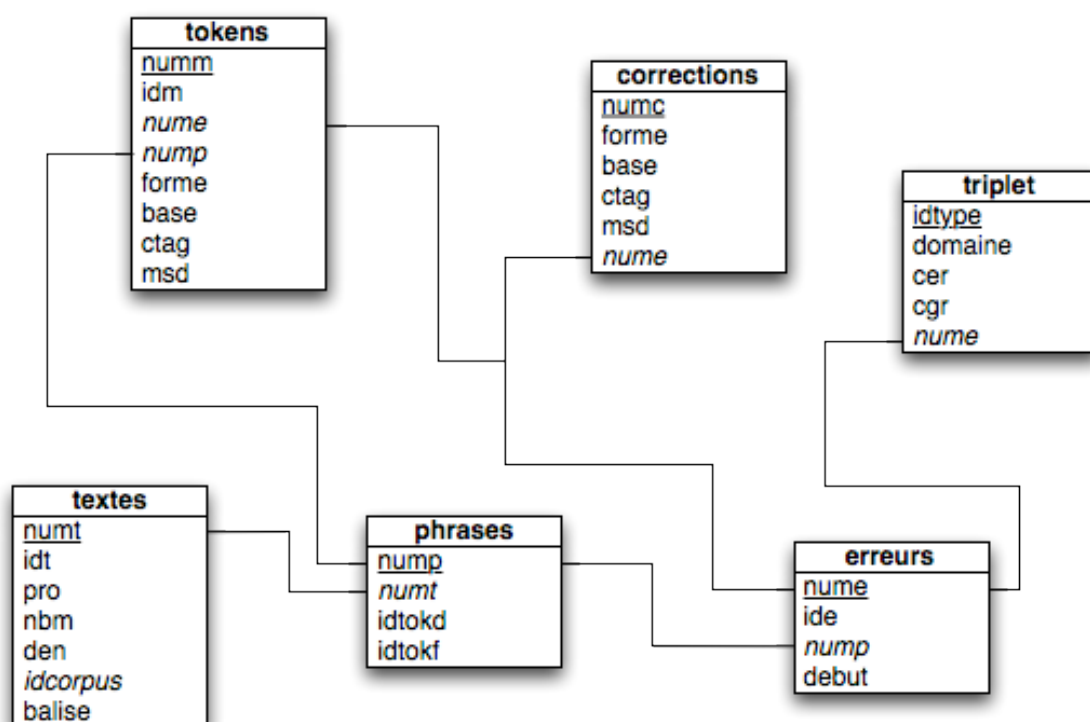


Figure 3 - Modèle de la base de données Idill avant le stage

## Traitement des données du formulaire

Les critères de recherche sont spécifiés dans un formulaire et récupérés en POST dans la page *requete.php*<sup>10</sup>. Ils sont stockés en variables de session pour un éventuel affichage ultérieur, notamment en cas de modification ou d'affinage de la requête. Une série de tests permet d'élaborer la requête complexe.

La construction de la requête complexe se fait en plusieurs temps. Les sous-requêtes correspondant aux quatre types de contraintes sont créées dynamiquement, puis sont imbriquées. Les identifiants des phrases satisfaisant aux critères de la recherche sont renvoyés.

```
//Contraintes sur le texte
$CT= "SELECT phrases.nump AS IDPT FROM phrases JOIN textes ON
phrases.numt=textes.numt WHERE 1" ;

//Provenance
if ($pro != "tout" && $pro != "") {
$CT.= " AND pro='".$pro."' " ;
}
//Nombre de mots (longueur)
if ($nbm != "tout" && $nbm != "") {
if ($nbm == "court") {
$CT.= " AND nbm BETWEEN 28 AND 111" ;
}
if ($nbm == "moyenne") {
$CT.= " AND nbm BETWEEN 111 AND 147" ;
}
if ($nbm == "moyenne+") {
$CT.= " AND nbm BETWEEN 147 AND 274" ;
}
if ($nbm == "longs") {
$CT.= " AND nbm BETWEEN 274 AND 3804" ;
}
}

//Densité
if ($den != "tout" && $den != "") {
if ($den == "faible") {
$CT.= " AND den BETWEEN 0 AND 12" ;
}
if ($den == "moyenne") {
$CT.= " AND den BETWEEN 12 AND 17" ;
}
if ($den == "moyenne+") {
$CT.= " AND den BETWEEN 17 AND 24" ;
}
if ($den == "elevee") {
$CT.= " AND den BETWEEN 24 AND 69" ;
}
}
```

Figure 4 - Création de la sous-requête correspondant aux critères sur les textes

---

<sup>10</sup> Voir annexe 1.

Si l'utilisateur spécifie comme provenance *English*, comme densité *toutes* et comme longueur *moyenne*, la requête sera donc de cette forme :

```
SELECT phrases.nump AS IDPT FROM phrases
JOIN textes ON phrases.numt=textes.numt WHERE 1
AND pro='english' AND nbm BETWEEN 12 AND 17
```

Figure 5 - Sous-requête correspondant aux critères sur les textes

## L'interface de recherche d'EXXELANT

L'interface d'EXXELANT est composée de trois pages principales. La page *idill.php* comporte le formulaire permettant de définir les critères de la recherche. La page *recherche.php* traite les données envoyées par le formulaire et effectue la requête. Enfin, la page *resultat.php* affiche les résultats obtenus.

Voici le formulaire de recherche d'EXXELANT avant le stage:

Figure 6 - Interface de recherche d'EXXELANT avant le stage

Un système de listes déroulantes et de champs de saisie de texte permet de spécifier les critères de la recherche. L'utilisateur peut spécifier jusqu'à 23 critères différents. Ces critères correspondent à quatre types de contraintes : contraintes sur les textes, contraintes sur la cible, contraintes sur le contexte gauche et contraintes sur le contexte droit.

L'utilisateur peut choisir la langue maternelle des apprenants à l'origine des productions, la densité d'erreur des textes et leur longueur, exprimée en nombre de mots. Le choix de la langue peut permettre de travailler sur des erreurs typiques de cette langue. Par exemple, pour des apprenants anglophones, la place de l'adjectif qualificatif pose souvent problème, l'adjectif étant toujours devant le nom en anglais alors que sa position fluctue en français, ou encore pour des apprenants arabophones la confusion des sons « i » et « é », les deux sonorités étant pour eux très proches. La densité d'erreur donne une indication sur le niveau de l'apprenant. De plus, en choisissant des textes de faible densité, il est possible de voir si les erreurs présentes sont des erreurs qu'un natif pourrait commettre ou s'il s'agit d'erreurs persistantes. La longueur des textes peut informer sur la maîtrise de la langue par l'apprenant : plus un texte est long, plus l'apprenant maîtrise la langue.

#### Sélection du corpus

[Nouvelle requête](#)

Provenance	Densité d'erreurs	Longueur des textes (nombre de mots)
toutes	toutes : 0 à 68	tous : 29 à 3803

**Recherche d'expression**

Figure 7 - Définition des contraintes sur les textes

Pour la cible, l'utilisateur peut choisir de rechercher une expression fautive ou non. Si la recherche porte sur une erreur, il peut alors choisir le domaine d'erreur et éventuellement la catégorie d'erreur. La page *tabb\_err.htm*, accessible depuis la page de recherche, présente les croisements possibles des deux premiers niveaux d'erreur et le troisième niveau d'erreur. Des champs de saisie permettent de rechercher la forme ou le lemme de l'expression recherchée et des listes déroulantes de définir la catégorie grammaticale et les traits grammaticaux de l'expression. Enfin, l'utilisateur peut choisir de rechercher une correction d'erreur et dispose pour cela des mêmes critères que pour la recherche d'une expression. L'utilisateur peut rechercher une expression non-fautive, afin de présenter des usages corrects. Au contraire, il peut ne rechercher que des expressions erronées, dans le but de créer des exercices de détection d'erreur. Ou encore, il peut choisir une expression erronée et une correction associée pour obtenir des résultats sur une faute précise, l'emploi de l'auxiliaire *être* au lieu de l'auxiliaire *avoir* par exemple.

Figure 8 - Définition des critères sur la cible

Les contextes droit et gauche peuvent aussi être spécifiés, avec les mêmes critères que pour une expression ou une correction, auxquels s’ajoute la possibilité de définir un intervalle, correspondant au nombre de mots entre les contextes et l’expression recherchée. Dans le cadre d’une étude de l’emploi de verbes à particules prépositionnelles par exemple, il est utile de pouvoir spécifier le verbe à rechercher et dans le contexte droit la préposition employée. Cela peut permettre de faire des comparaisons pour souligner la différence de sens selon la préposition employée (entre *parler de quelqu’un* et *parler à quelqu’un* par exemple).

Pour qu’une recherche soit possible, au moins un des critères parmi ceux-ci doit être renseigné : domaine d’erreur, forme, lemme ou catégorie de l’expression, forme, lemme ou catégorie de la correction. Ce choix a été fait pour ne pas permettre une diffusion du corpus dans son ensemble, celui-ci n’étant pas libre.

Les valeurs des listes déroulantes sont des listes fermées. Les listes déroulantes comprenant les traits et les catégories d’erreur sont mises à jour dynamiquement, grâce à un script Ajax, en fonction de la catégorie grammaticale ou du domaine d’erreur sélectionné<sup>11</sup>.

Figure 9 - Listes déroulantes dynamiques

<sup>11</sup> Voir annexe 2.



## Affichage des résultats

L'affichage des résultats se fait à l'aide des identifiants des phrases sélectionnés par la requête complexe. Grâce à ceux-ci, il est possible d'obtenir les informations sur les tokens composant la phrase, y compris l'identifiant d'erreur. Si celui-ci n'est pas nul, la correction est récupérée. Ces opérations sont réalisées à l'aide de requêtes SQL intermédiaires. Une série de tests permet ensuite d'afficher les phrases dans un tableau, selon 5 colonnes : numéro, texte (contient l'identifiant du texte), contexte gauche, cible et contexte droit.

**Résultats** ([Rappel de la requête en cours](#))

[Nouvelle requête](#) | [Modifier la requête](#) | [Affiner la requête](#) | [Traces](#) | [Afficher la sélection](#)

(Sélectionner les phrases)

Numéro	Texte	Contexte gauche	Cible	Contexte droit	Choix
1	<a href="#">2254</a>	<i>j' étais dans un <u>jardin</u> très grand et <u>occupé</u> par les arbres et les fleurs , tout à coup un tout petit</i>	<b>chat</b>	<i>a <u>existé</u> dans ce <u>jardin</u> entre les <u>arbres</u> .</i>	<input type="checkbox"/>
2	<a href="#">2277</a>		<b>chat</b>	<i>. Mais , toutefois , il a aperçu son</i>	<input type="checkbox"/>
3	<a href="#">2277</a>		<b>chat</b>	<i>est entré . Il a ouvert la porte et le</i>	<input type="checkbox"/>

Figure 10 - Affichage des résultats par phrase

Survoler un mot affiche son lemme, sa catégorie grammaticale et éventuellement les traits qui lui sont associés. Survoler un mot erroné, identifié par un soulignement, affiche son lemme, sa catégorie grammaticale, éventuellement ses traits associés, ainsi que sa correction.

*fleurs , tout à coup un tout petit*  
Lemme: fleur, Cat: nom, Tags: femin

Figure 11 - Survol d'un mot juste

*j jardin très grand et occupé par les arb*  
Lemme: unknown, Cat: adj, Corrigé: jardin

Figure 12 - Survol d'un mot erroné

Cliquer sur l'identifiant numérique du texte, en bleu, ouvre une fenêtre dans laquelle s'affiche la production complète de l'apprenant contenant la phrase. L'identifiant du texte, la provenance de la production, le nombre de mots du texte et la densité d'erreur sont également affichés. La phrase sélectionnée est affichée en gras dans le texte.

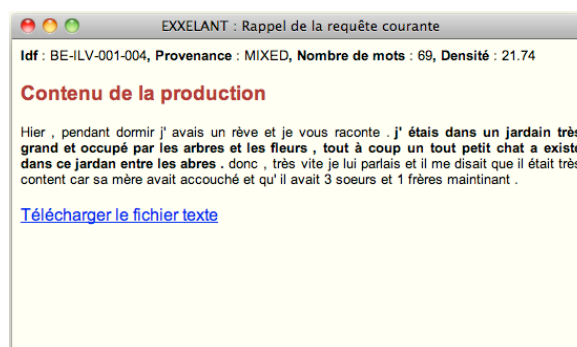


Figure 13 - Fenêtre surgissante contenant le texte sélectionné

## Modification et affinage d'une requête

Il est possible, une fois les résultats affichés, de modifier les critères de recherche ou d'affiner la requête. Dans le cas d'une modification, les critères de la recherche précédente sont affichés dans le formulaire de recherche et modifiables. Dans le cas d'un affinage, les critères sont affichés dans le formulaire mais ne sont pas modifiables. Il est seulement possible d'ajouter des critères de recherche.

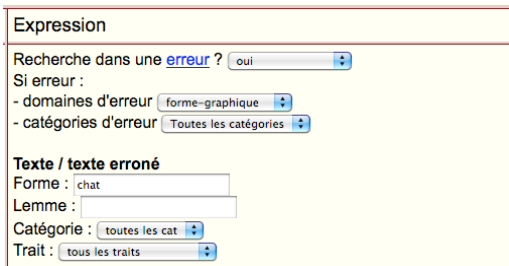


Figure 14 - Modification d'une requête



Figure 15 - Affinage d'une requête

### 1.3. Travail à réaliser

Dans le cadre du projet IDILL, le travail autour d'EXXELANT s'est essentiellement focalisé sur les aspects de double balisage (erreurs et morphosyntaxe) ainsi que sur les aspects requêtage. L'utilité d'un tel outil est indéniable. Cependant, pour pouvoir être utilisable concrètement par un enseignant de Français Langue Étrangère (FLE), EXXELANT nécessitait des améliorations et une adaptation au contexte d'utilisation pédagogique.

#### *Authentification et refonte de l'interface*

L'une des perspectives d'évolution d'EXXELANT est la création d'un environnement d'apprentissage autour de l'exemplier. Cet environnement permettrait à des enseignants de générer des activités à destination des apprenants à partir des résultats de l'application et également d'ajouter de nouvelles productions d'apprenants. Dans cette optique, un système d'authentification à l'aide d'identifiants personnels sera mis en place. Le système d'authentification permettra de relier un enseignant aux activités qu'il aura créées ou aux textes qu'il aura ajoutés. La génération d'activité ne sera pas développée dans le cadre de ce stage.

De ces améliorations découlent trois niveaux d'utilisateurs, avec des droits spécifiques.

- L'administrateur : pourra créer, modifier ou supprimer des comptes enseignants. Il aura la responsabilité de créer de nouveaux corpus et d'y ajouter les textes bruts, sur demande des enseignants. Les textes subiront lors de l'ajout un balisage morphosyntaxique automatique à l'aide de TreeTagger et du dictionnaire d'ABU.
- Les enseignants : auront, en plus de l'accès à l'interface de recherche, la possibilité de créer, modifier ou supprimer des comptes étudiants, organisés en groupes. Ils auront accès à une interface de balisage des erreurs des textes bruts ajoutés à la base suite à leur demande.
- Les étudiants : n'auront accès pour l'instant qu'à l'interface de recherche.

L'interface graphique sera retravaillée en lien avec ces améliorations.

### ***Export d'un texte ou d'une sélection de phrases***

Une recherche peut renvoyer un grand nombre de correspondances. Ainsi, suite à une requête, un enseignant peut se retrouver face à plusieurs dizaines voire même centaines de résultats difficilement exploitables en classe. Il semble donc nécessaire de permettre à l'utilisateur de faire une sélection de phrases et de pouvoir exporter cette dernière pour une utilisation ultérieure.

Un utilisateur peut consulter le texte comprenant l'un des résultats. De la même façon, il semble important qu'il puisse exporter ce texte, pour une éventuelle étude de la phrase dans son contexte par exemple.

### ***Enrichissement du corpus et modification du formulaire de recherche***

Actuellement EXXELANT ne permet qu'une interrogation du corpus FRIDA-bis. Pour permettre une utilisation personnalisée de l'outil et un enrichissement de la base de données, la possibilité sera donnée aux enseignants d'ajouter de nouveaux textes issus de leur propre collecte en classe. L'étiquetage morphosyntaxique des textes sera automatisé. L'annotation des erreurs sera quant à elle effectuée manuellement grâce à une interface adaptée aux compétences informatiques de l'enseignant.

L'interface de recherche subira des modifications en lien avec cette amélioration. Il sera possible d'effectuer une recherche dans tous les corpus, dans un corpus choisi ou dans les corpus d'un utilisateur en particulier, et d'effectuer des requêtes sur des textes ne possédant que l'étiquetage morphosyntaxique.

### ***Réflexion sur l'usage des traces***

Un outil pédagogique informatique doit constamment être remis en cause. De la même façon qu'un enseignant adapte le contenu de son cours ou sa façon d'expliquer en fonction de ses apprenants, l'outil doit apprendre de ses utilisateurs et évoluer en corrélation avec ce qu'il peut observer de leurs interactions. D'autre part, les notes obtenues aux évaluations par exemple constituent une rétroaction à la fois pour l'enseignant, lui permettant de contrôler les connaissances acquises, et pour l'apprenant, lui permettant de se rendre compte de son niveau de compétence. L'analyse des traces de ces interactions pour effectuer ce travail est largement répandu, notamment dans les plateformes d'apprentissage en ligne (Dokeos<sup>12</sup>, Moodle<sup>13</sup>, Claroline<sup>14</sup>, Ganesha<sup>15</sup>, etc.). Une étude des différents types de traces, de leur utilité et enfin de l'usage qui peut en être fait, en général et dans le cas précis d'EXXELANT, devra être réalisée.

---

<sup>12</sup> <http://www.dokeos.com/fr>

<sup>13</sup> <http://moodle.org/>

<sup>14</sup> <http://www.claroline.net/index.php?lang=fr>

<sup>15</sup> <http://www.ganesha.fr/>

**2.**

**Amélioration du prototype**

Le prototype EXXELANT ne comportait qu'une interface de recherche et une interface d'affichage des résultats. Afin d'en faire un outil plus utile et utilisable dans un réel contexte pédagogique, des fonctionnalités supplémentaires ont été développées.

## **2.1. Refonte de l'interface**

Dans le cadre du projet IDILL, aucun travail spécifique n'a été mené sur l'interface. Celle-ci ainsi que la charte graphique sont donc celles du projet de 2006. Un travail de remodelage, de modernisation et surtout d'adaptation aux fonctionnalités nouvelles et au public était nécessaire. L'interface et le design sont des éléments importants dans la conception d'un outil pédagogique informatique. Un tel outil se doit d'être à la fois simple d'utilisation pour des utilisateurs à littératie informatique limitée et attrayant.

La nouvelle interface a été optimisée pour le navigateur Mozilla Firefox, dans un souci de portabilité. Ce navigateur est en effet multiplateformes, et en conformité avec les normes du W3C. L'accent a été mis sur le dynamisme, grâce à JavaScript et Ajax, pour rendre la manipulation de l'outil simple et attractive et limiter les rechargements de page.

La charte graphique a été complètement retravaillée. L'interface actuelle est composée d'une division centrale aux bords arrondis servant de cadre. Cette division a comme couleur dominante le bleu marine, auquel s'ajoute en dégradé une couleur spécifique au statut de l'utilisateur connecté : vert pour un administrateur, jaune pour un enseignant, bleu pour un apprenant. Les textes sont affichés en noir dans une division de contenu, à fond blanc, dans un souci de clarté.

**Exxelant**  
Connexion professeur | Documentation

[Nouveau groupe](#) | [Voir groupes](#) | [Nouvelle recherche](#) | [Mes corpus](#) | [Modifier mon mot de passe](#)

Rechercher...

Sélection du corpus

Utilisateur	Corpus	Annotation des erreurs	Provenance	Densité d'erreur	Longueur (nombre de mots)
FRIDA	Tous	Sans importance	toutes	toutes : 0 à 68	tous : 29 à 3803

Recherche d'expression

Contexte gauche	Cible	Contexte droit
Rechercher dans une erreur? <span style="float: right;">Sans importance</span> Si erreur: Domaine d'erreur: <span style="float: right;">Tous les domaines</span> Catégorie d'erreur: <span style="float: right;">Toutes les catégories</span> Texte/Texte erroné Forme: <input type="text"/> Lemme: <input type="text"/> Catégorie: <span style="float: right;">Toutes les catégories</span> Trait: <span style="float: right;">Tous les traits</span> Intervalle (nombre de mots): <input type="text"/>	Rechercher dans une erreur? <span style="float: right;">Sans importance</span> Si erreur: Domaine d'erreur: <span style="float: right;">Tous les domaines</span> Catégorie d'erreur: <span style="float: right;">Toutes les catégories</span> Texte/Texte erroné Forme: <input type="text"/> Lemme: <input type="text"/> Catégorie: <span style="float: right;">Toutes les catégories</span> Trait: <span style="float: right;">Tous les traits</span> Correction Forme: <input type="text"/> Lemme: <input type="text"/> Catégorie: <span style="float: right;">Toutes les catégories</span> Trait: <span style="float: right;">Tous les traits</span>	Rechercher dans une erreur? <span style="float: right;">Sans importance</span> Si erreur: Domaine d'erreur: <span style="float: right;">Tous les domaines</span> Catégorie d'erreur: <span style="float: right;">Toutes les catégories</span> Texte/Texte erroné Forme: <input type="text"/> Lemme: <input type="text"/> Catégorie: <span style="float: right;">Toutes les catégories</span> Trait: <span style="float: right;">Tous les traits</span> Intervalle (nombre de mots): <input type="text"/>

Figure 16 - Nouvelle interface de recherche d'EXXELANT

Enfin, pour faciliter la navigation dans la liste des résultats, qui peut s'avérer longue, une flèche blanche sur le côté droit du cadre a été ajoutée. Cette flèche permet de remonter en haut de la page.

407 résultats

Numéro	Texte	Contexte gauche	Cible	Contexte droit	Choix
1	<a href="#">1991</a>	L' ouvrier peut	<b>être</b>	motiver par des primes , des indemnités supplémentaires , des possibilités à promouvoir , ... Le dirigeant doit donner une promotion de temps à temps aux ouvriers qui ont bien presté .	<input type="checkbox"/>
2	<a href="#">2002</a>	L' Etat a besoin de recette afin d'	<b>être</b>	capable d' obtenir une gestion acceptable .	<input type="checkbox"/>
3	<a href="#">2004</a>	Premièrement les services publics come la santé et l' enseignement doivent	<b>être</b>	financiés .	<input type="checkbox"/>
4	<a href="#">2007</a>	Le revenu des contribuables peut	<b>être</b>	divisé en 4 catégories : Les revenus mobiliers ( les intérêts , les dividendes ... ) , immobiliers ( la location d' une maison ... ) , professionnels ( le salaire ... ) et divers .	<input type="checkbox"/>
5	<a href="#">2015</a>	Les ressources de l' argent peuvent	<b>être</b>	des obligations , des comptes d' épargne , des terrains , etc. Aussi le travail est un de ses ressources .	<input type="checkbox"/>
6	<a href="#">2021</a>	Le but d' une telle opération est souvent de réaliser des synergies , de réduire les coûts et d'	<b>être</b>	présent sur le marché international .	<input type="checkbox"/>
7	<a href="#">2023</a>	Les conséquences d' une promotion peuvent	<b>être</b>	l' expatriation et l' augmentation du salaire .	<input type="checkbox"/>
8	<a href="#">2024</a>	Il est nécessaire de faire la distinction entre l' absentéisme externe (	<b>être</b>	absent ) <sub>1</sub> et l' absentéisme interne ( être distrait <sub>1</sub> arriver en retard ) .	<input type="checkbox"/>
9	<a href="#">2024</a>	Il est nécessaire de faire la distinction entre l' absentéisme externe ( être absent ) <sub>1</sub> et l' absentéisme interne (	<b>être</b>	distrait <sub>1</sub> arriver en retard ) .	<input type="checkbox"/>

Figure 17 - Nouvel affichage des résultats

## 2.2. Export de phrases et de texte

Le corpus FRIDA-BIS est un corpus conséquent comprenant 179 642 tokens. Ainsi, une recherche peut renvoyer de nombreux résultats selon les critères spécifiés. Il n'était possible ni de sauvegarder, ni de sélectionner des résultats, faisant du tri de ceux-ci un exercice long et fastidieux, et limitant l'utilisation de l'application à une utilisation en ligne. Pour permettre d'exploiter les résultats hors ligne et la création d'exempliers, la fonctionnalité de sélection et d'export de celle-ci a été ajoutée. Un utilisateur peut aussi être intéressé par l'étude d'un résultat en contexte, c'est-à-dire au sein du texte complet. La fonctionnalité d'export d'un texte sélectionné a donc été ajoutée. Il n'est pas possible de choisir plusieurs textes et de les exporter en une seule fois. Le corpus FRIDA-BIS n'étant pas un corpus libre, il ne doit pas être possible pour l'utilisateur d'obtenir l'ensemble du corpus. Pour la sélection de phrases ou la sélection d'un texte, l'export se fait dans un fichier au format texte brut, exploitable sur tous les systèmes d'exploitation.

Les exports de phrases ou d'un texte sont disponibles depuis la page de résultats. Des cases à cocher, affichées dynamiquement en même temps que les résultats, permettent d'effectuer une sélection de phrases. Chaque case à cocher a pour valeur l'identifiant de la phrase à laquelle elle se rapporte. Une fois la sélection effectuée, l'utilisateur peut l'afficher en cliquant sur le bouton *Sélectionner les phrases*. Les identifiants des phrases sélectionnées sont récupérés par la méthode POST dans la page *selection.php*. Les phrases sont affichées à l'écran et écrites dans un fichier texte possédant un nom aléatoire, pour éviter les éventuelles collisions de fichiers entre plusieurs utilisateurs. Il est possible d'imprimer la sélection ou de télécharger le fichier texte.





Figure 18 - Affichage de la sélection de phrases dans l'application

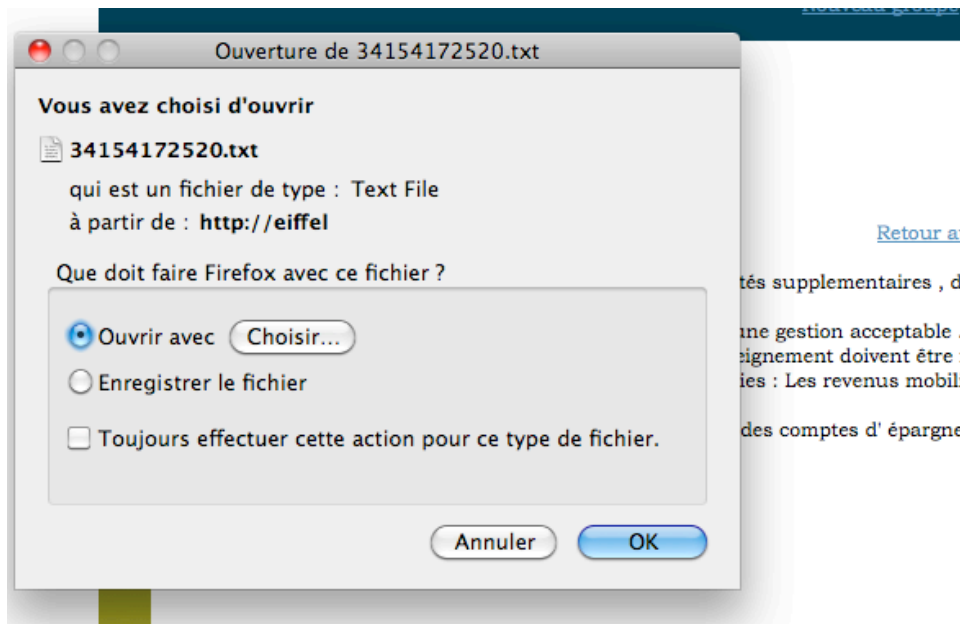


Figure 19 - Téléchargement de la sélection de phrase

Pour sélectionner un texte, il suffit de cliquer sur l'identifiant du texte au début d'une ligne de résultat. Le texte choisi est affiché dans une fenêtre surgissante. Au moment de l'affichage, le texte est également écrit dans un fichier texte qui possède lui aussi un nom aléatoire. L'utilisateur a les mêmes possibilités que pour la sélection de phrases, soient l'impression du texte ou le téléchargement du fichier texte. Pour que le fichier texte soit plus lisible, un saut de ligne a été ajouté après chaque phrase.

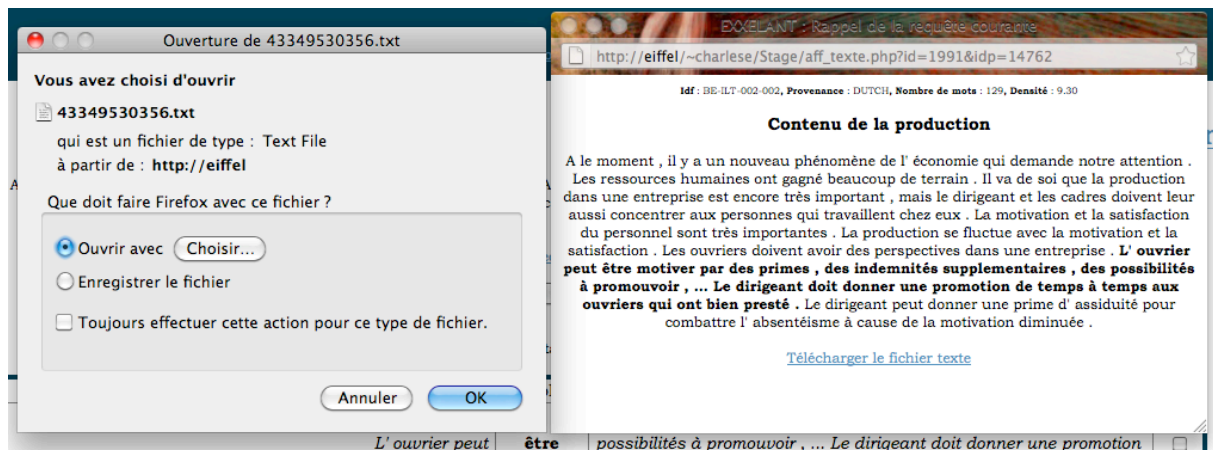


Figure 20 - Téléchargement d'un texte

Le téléchargement de fichier n'étant pas automatique en PHP, il a été forcé par des headers.

### 2.3. Développement du système de connexion

Le prototype EXXELANT ne comportait qu'une interface de recherche et une interface d'affichage des résultats. Aucun système de connexion n'avait été mis en place. Les nouvelles fonctionnalités de l'application ainsi que l'objectif d'en faire une véritable plateforme pédagogique permettant de générer et d'effectuer des activités impliquent la sauvegarde de données spécifiques aux utilisateurs, et donc l'identification de ceux-ci grâce à un nom d'utilisateur et un mot de passe. Cette identification permet également de limiter l'utilisation de l'outil à un cadre pédagogique bien défini.

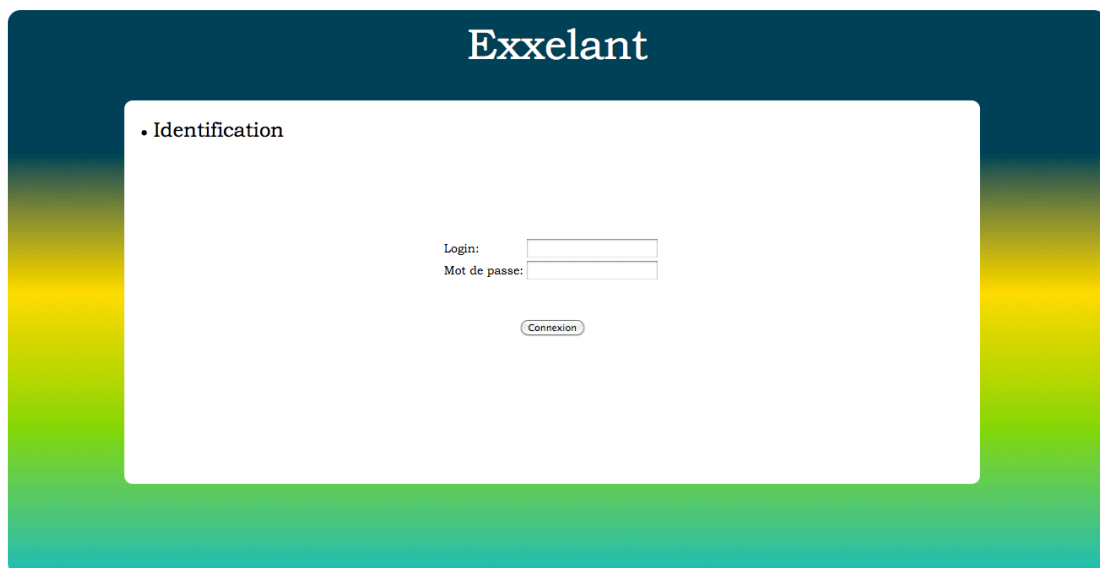


Figure 21 – Page d'accueil de l'application avec formulaire de connexion

Au cours de la conception des nouvelles fonctionnalités, trois niveaux d'utilisateurs avec des droits différents ont été identifiés : administrateur, enseignant et apprenant. Pour permettre l'identification des utilisateurs, une table a été ajoutée à la base de données existante<sup>16</sup>, la table *utilisateur*, dans laquelle sont stockés nom, prénom, adresse email, login, mot de passe, et statut.

Les utilisateurs des trois niveaux peuvent modifier leur mot de passe. La modification de mot de passe mise à part, un apprenant n'a accès qu'à l'interface de recherche.

### ***Fonctions administrateur***

L'administrateur possède un rôle de modérateur. Cette application est prévue pour être utilisée dans un cadre pédagogique défini. L'administrateur possède donc un rôle de gestion des comptes enseignants ; c'est à lui que revient leur création. L'administrateur a aussi un rôle de modérateur en ce qui concerne l'enrichissement du corpus. L'ajout de texte à la base est réalisé par un administrateur sur demande d'un enseignant, ce qui permet d'effectuer une vérification du contenu des textes avant l'ajout. Un compte administrateur est créé depuis l'interface de gestion de la base de données. Les porteurs du projet sont chargés de la création de comptes de niveau administrateur.

Lorsque l'administrateur se connecte, il accède à la liste des utilisateurs de niveau enseignant. Dans le cas où aucun compte enseignant n'est enregistré, un message le spécifie et un lien vers l'interface d'ajout d'enseignant est affiché. Cette interface est accessible depuis le menu également.

---

<sup>16</sup> Voir le nouveau modèle de base de données en annexe 4.

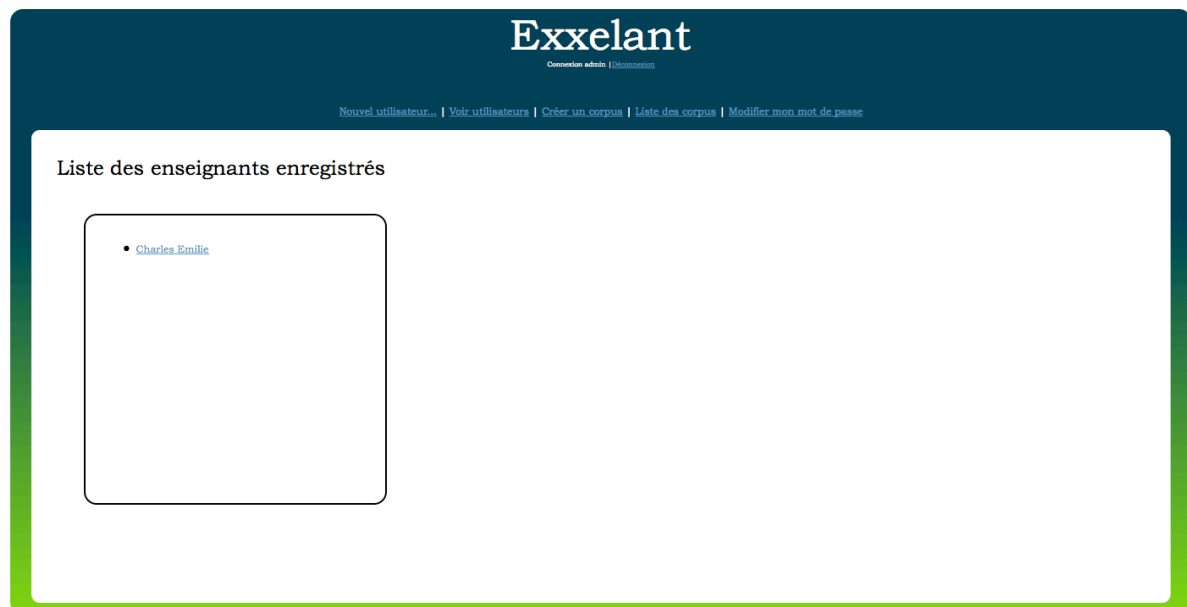


Figure 22 - Page d'accueil administrateur

Pour créer un nouveau compte enseignant, un formulaire comprenant les champs Nom, Prénom et Adresse e-mail doit être renseigné ; login et mot de passe sont générés automatiquement par la suite. Une fois le formulaire soumis, le contenu des champs est vérifié en PHP. Le remplissage des champs est d'abord contrôlé ; aucun ne doit être vide, tous les champs étant obligatoires. Ensuite, la formation correcte de l'adresse e-mail est vérifiée à l'aide d'une expression régulière et son existence dans la base de données grâce à une requête SQL. Si le formulaire est correctement renseigné, les données sont traitées pour générer un login et un mot de passe.

La génération de login s'effectue à partir des nom et prénom. Pour pouvoir le générer, ces deux chaînes doivent d'abord être nettoyées, pour éviter la présence de caractères spéciaux dans le login. Accents et cédilles sont supprimés, ainsi que les tirets, et les chaînes sont transformées en minuscules. Les sept premiers caractères du nom sont récupérés, ou le nom entier s'il fait moins de sept caractères. Le login est généré en concaténant cette chaîne à la première lettre du prénom. Une requête SQL vers la base de données permet de vérifier s'il existe ou non. S'il existe, la lettre suivante du prénom est concaténée à la chaîne. Ce traitement est inséré au sein d'une boucle « tant que ». Ce système de génération permet donc les homonymes. La probabilité qu'après avoir parcouru toutes les lettres du prénom aucun login ne soit disponible est considérée comme faible.

Le mot de passe est généré automatiquement, par un choix aléatoire de 8 caractères parmi 58 caractères, comprenant majuscules, minuscules, chiffres et caractères spéciaux.

Les données du nouvel enseignant sont alors insérées dans la table utilisateur de la base de données. Celles-ci sont affichées immédiatement après l'ajout et imprimables dans un format adapté. L'administrateur a alors trois choix : imprimer, ajouter un nouvel enseignant ou voir la liste des enseignants.



Figure 23 - Confirmation de création d'un nouveau compte enseignant

La modification d'un compte est accessible depuis la liste des enseignants enregistrés. Cliquer sur le nom d'un enseignant affiche une division contenant ses informations et la liste des actions disponibles, grâce à un script AJAX. Nom, prénom, adresse e-mail, login et mot de passe sont affichés ; l'administrateur peut imprimer ou modifier ces informations, ou supprimer le compte enseignant. En cliquant sur modifier, un formulaire comprenant des champs textes préremplis par les informations actuelles sur l'enseignant remplace l'affichage des informations. Seuls le nom, le prénom et l'adresse e-mail peuvent être modifiés, login et mot de passe étant générés automatiquement. Lors de la validation, une boîte de dialogue s'affiche, demandant confirmation à l'administrateur en précisant que les données de connexion sont susceptibles d'être modifiées. En effet, si les nom et prénom sont modifiés, le login sera régénéré en fonction de ces nouvelles données.

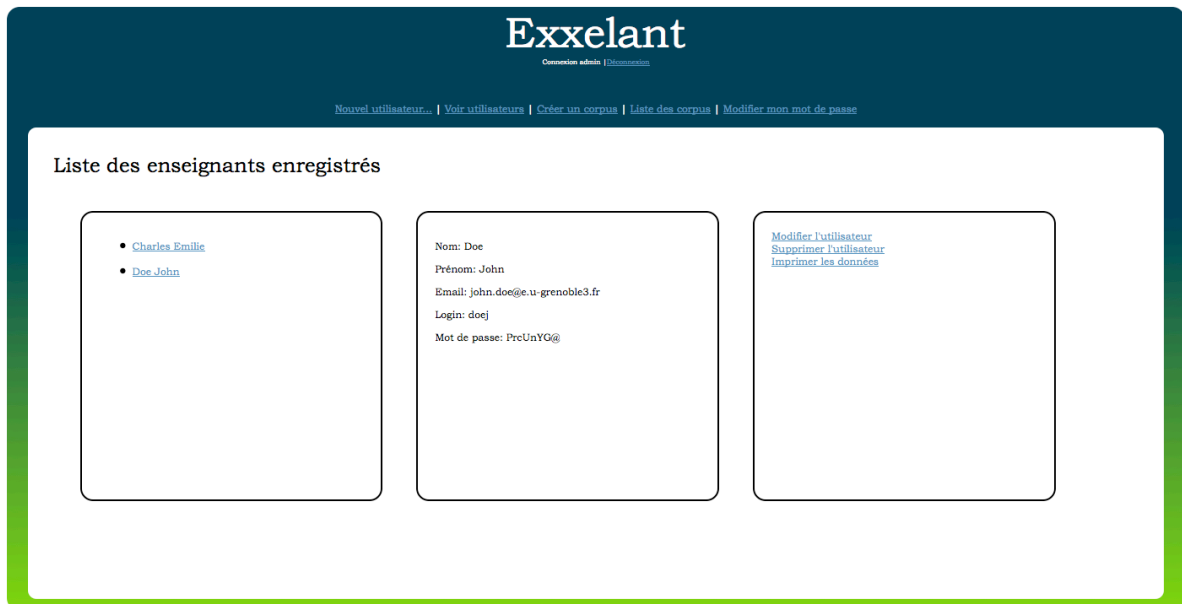


Figure 24 - Affichage des informations d'un enseignant

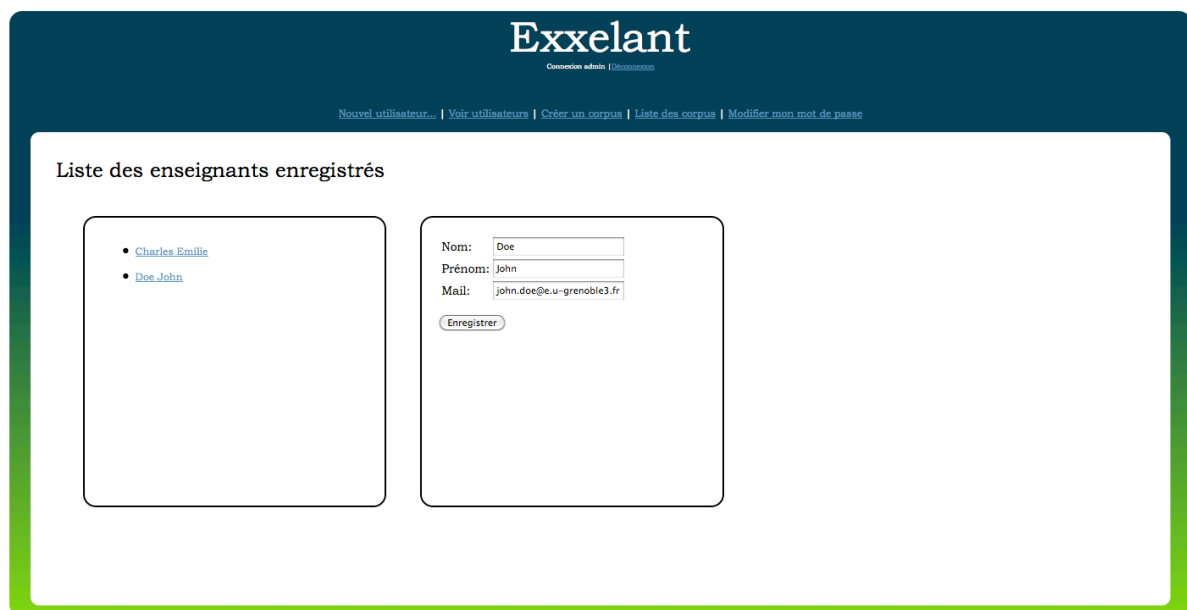


Figure 25 - Modification des informations d'un enseignant

Si l'administrateur choisit de supprimer un compte, ce ne sont pas seulement ses informations qui seront supprimées. La suppression d'un compte enseignant implique la suppression de ses groupes classes. Les comptes des membres d'un groupe sont également détruits si les membres ne font partie d'aucun autre groupe. De même que pour la modification d'un compte, une boîte de dialogue demande confirmation à l'utilisateur. La suppression d'un compte n'est pas réversible.

## ***Fonctions enseignant***

Un enseignant a la possibilité de créer des groupes et des comptes apprenants associés à ces groupes. Après authentification, un enseignant accède à la liste de ses groupes. Si aucun groupe n'a été créé, un message le signalant est affiché, et l'enseignant est incité à créer un nouveau groupe. La création de groupes est disponible depuis le menu également. Il n'est pas possible de créer un compte apprenant hors d'un groupe.

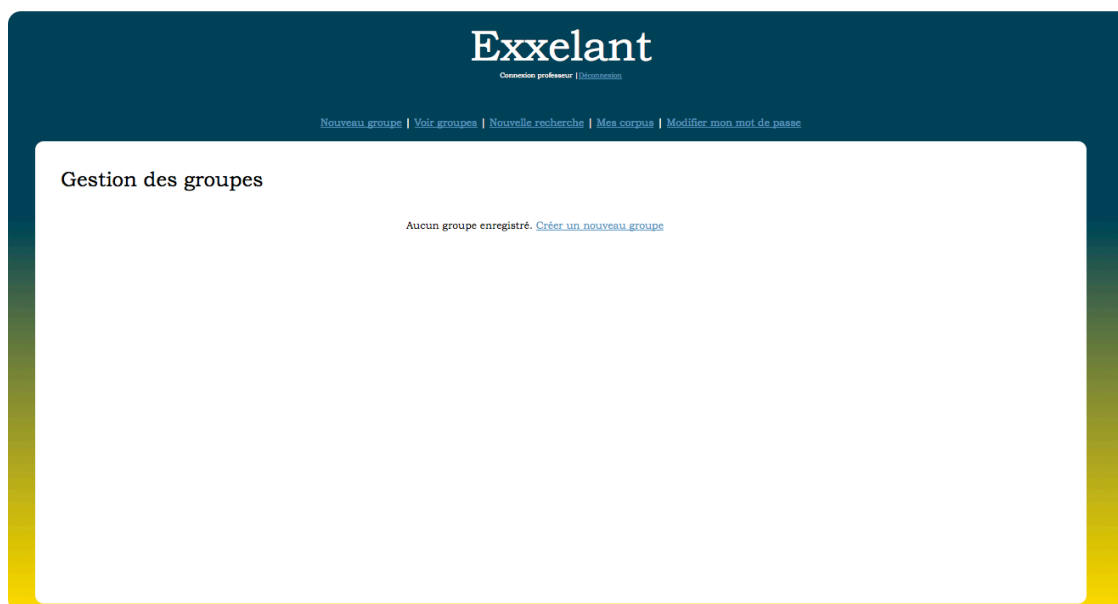


Figure 26 - Page d'accueil enseignant

La gestion de groupe d'apprenants a entraîné l'ajout de la table *groupe* à la base de données. Cette table comprend le nom du groupe et l'identifiant de l'enseignant l'ayant créé. Pour permettre à un apprenant d'appartenir à plusieurs groupes, la table *groupe\_ut* a également été ajoutée. Elle comprend l'identifiant de groupe et l'identifiant de l'apprenant.

Pour créer un groupe, l'enseignant doit simplement le nommer. Si l'enseignant a déjà créé un groupe portant ce nom, un message d'erreur est affiché. Une fois le groupe créé, un formulaire permettant d'ajouter un nouveau membre au groupe s'affiche. Les mêmes informations que lors de la création d'un compte enseignant par l'administrateur sont requises. La vérification du contenu des champs ainsi que la génération des login et mot de passe sont identiques. Après l'ajout, les informations sur l'apprenant sont affichées, avec la possibilité de les imprimer, d'ajouter un nouvel apprenant ou de consulter la liste des groupes.

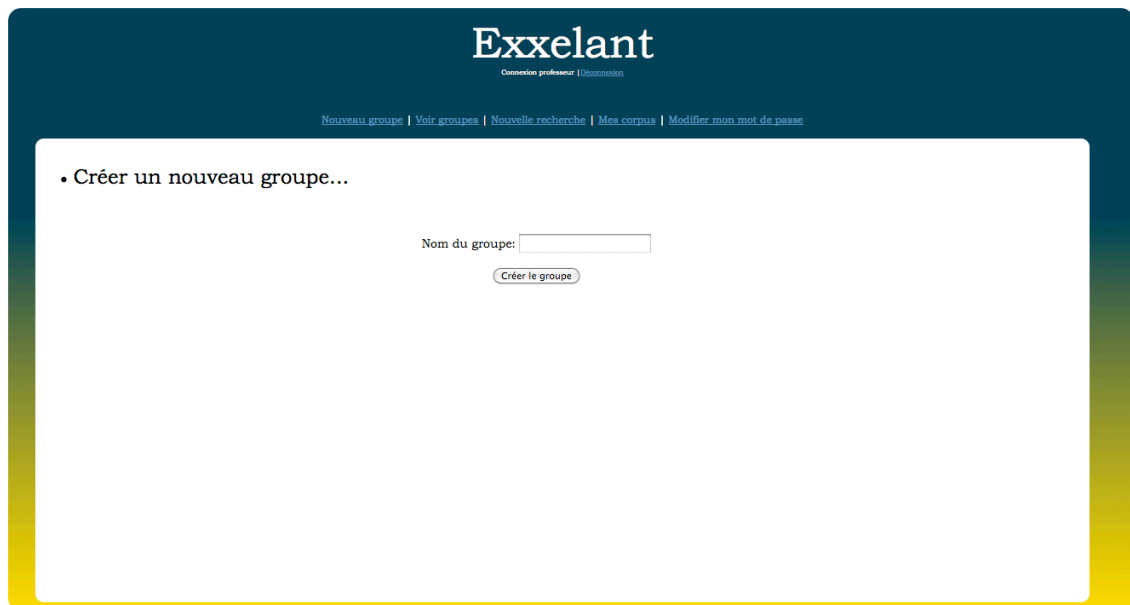


Figure 27 - Création d'un groupe classe

La modification d'un groupe ou d'un compte apprenant est disponible depuis la liste des groupes. En cliquant sur le nom d'un groupe, l'enseignant affiche la liste des membres du groupe et les actions possibles sur le groupe : modifier le nom du groupe, ajouter un membre, supprimer le groupe et imprimer la liste des membres. La modification du nom du groupe est effectuée grâce à un formulaire comprenant un champ texte. Ajouter un membre permet d'afficher le formulaire d'ajout d'étudiant, pour lequel les mêmes informations et les mêmes vérifications que précédemment sont effectuées. Supprimer un groupe supprime également les comptes de ses membres, si ceux-ci ne font partie d'aucun autre groupe.



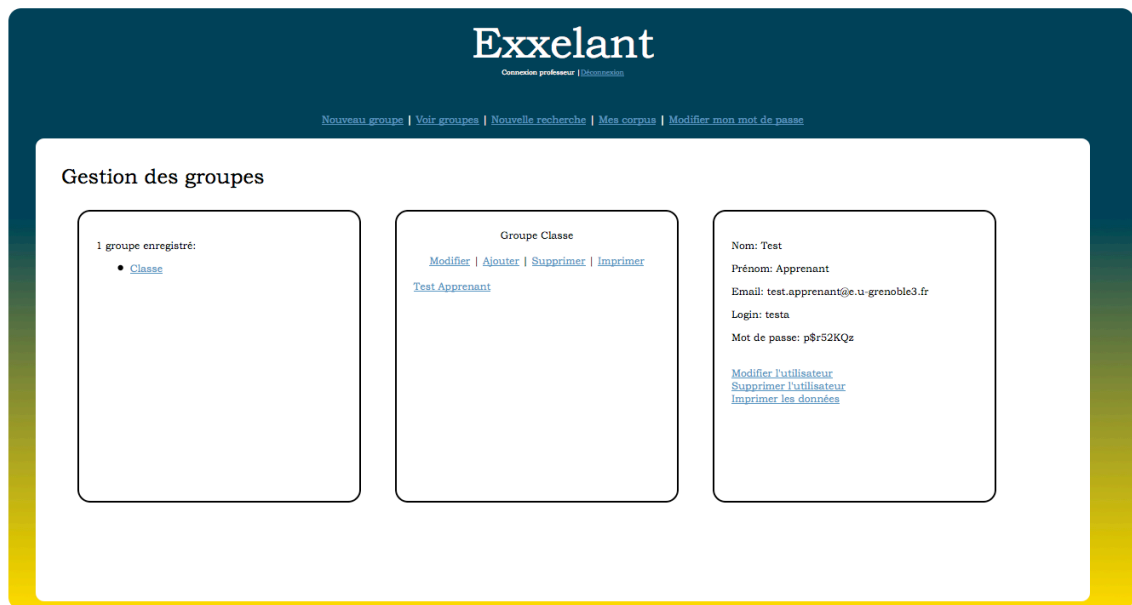


Figure 28 - Affichage des informations sur un groupe

Il est également possible d'imprimer ou de modifier les informations d'un membre, ou de supprimer le compte, en cliquant sur le nom de l'apprenant.

### ***Modification de mot de passe***

Le mot de passe étant généré automatiquement, il peut être difficile à retenir. Ainsi la possibilité de modifier le mot de passe a été mise en place. Une fois connecté, un utilisateur peut modifier son mot de passe. Un formulaire, requérant l'ancien mot de passe, le nouveau mot de passe et la confirmation de ce nouveau mot de passe, permet d'effectuer cette modification. Après soumission du formulaire, une vérification du contenu des champs est effectuée en PHP. Le remplissage des champs ainsi que l'identité des deux nouveaux mots de passe sont vérifiés en PHP. La conformité de l'ancien mot de passe est vérifiée grâce à une requête SQL.

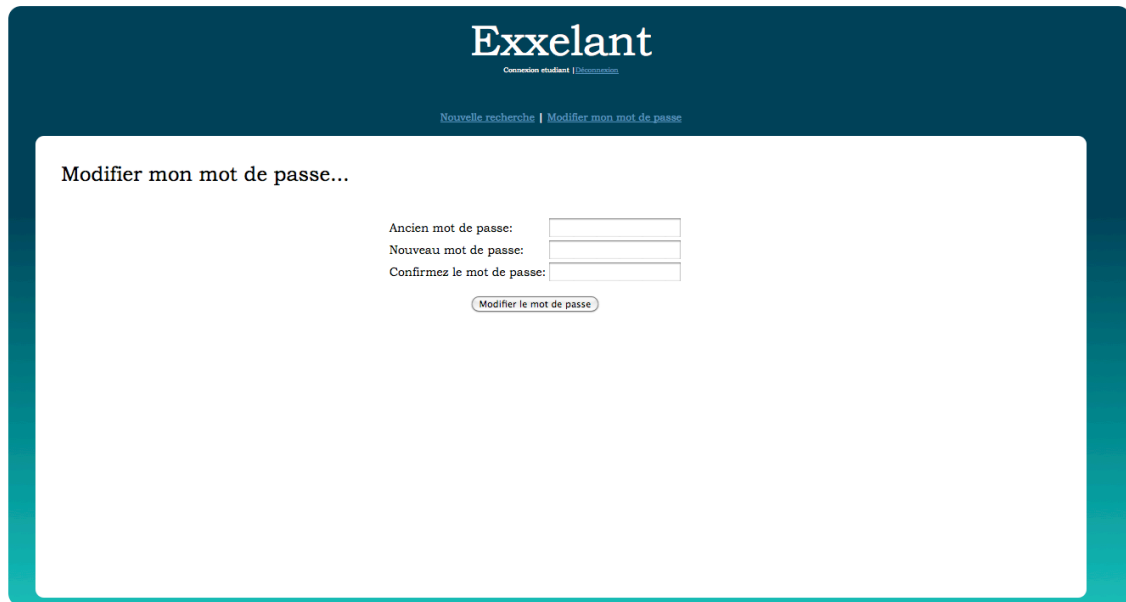


Figure 29 - Interface de modification du mot de passe

## 2.4. Enrichissement du corpus

Jusque-là, EXXELANT ne permettait que d’explorer le corpus prédéfini FRIDA-BIS. Afin d’enrichir la base de données, et donc l’application, un outil collaboratif mettant à contribution les enseignants a été développé.

Les enseignants peuvent demander l’ajout de leurs propres textes à la base. Au delà de l’enrichissement certain au niveau de la diversité des langues maternelles des apprenants ou des types d’erreur, cela permet aussi à un enseignant d’exploiter son propre corpus et à un apprenant de travailler sur ses propres productions.

La création de nouveaux corpus a entraîné plusieurs modifications<sup>17</sup>. La table *corpus* a été ajoutée à la base de données. Elle comprend le nom du corpus et l’identifiant de l’enseignant à l’origine de ce corpus. De plus, pour qu’un corpus soit exploitable, il doit au moins posséder une annotation morphosyntaxique des items du texte. Un système d’étiquetage morphosyntaxique automatique en ligne a donc été développé. L’annotation des erreurs quant à elle ne peut se faire automatiquement. Ainsi une interface de balisage manuel pour les enseignants a été conçue. Parmi les textes ajoutés, certains comporteront donc le seul étiquetage morphosyntaxique, d’autres seront complétés par le balisage des erreurs. La table *textes* a été modifiée pour accueillir l’identifiant du corpus auquel un texte

---

<sup>17</sup> Voir le nouveau modèle de base de données en annexe 4.

appartient, et le champ *balise* a été ajouté, permettant d'identifier si le texte bénéficie d'une double annotation ou non.

### ***Création de corpus et ajout de texte***

La création de corpus et l'ajout de texte se font sur demande d'un enseignant. Comme il semblait nécessaire que les textes soient validés avant d'être insérés dans la base, ces opérations sont effectuées par un administrateur (informé par mail suite à une demande de création/ajout de texte de la part d'un enseignant). Un enseignant souhaitant créer un corpus doit préciser son nom et le nom du corpus. Les textes à ajouter doivent être fournis au format texte brut, et la langue maternelle de l'auteur du texte doit être précisée. Notons que la configuration du serveur sur lequel l'application est installée ne permettant pas l'envoi automatique d'un e-mail à l'administrateur, un simple *mailto* est utilisé.

En cliquant sur *Créer un corpus* un champ texte permet de nommer le corpus et une liste déroulante de choisir parmi les enseignants enregistrés l'auteur du corpus. Une fois ces informations spécifiées, l'administrateur accède à l'interface d'ajout de texte. Deux possibilités s'offrent à lui : ajouter un texte en le copiant dans la zone de texte ou utiliser le champ *Parcourir* pour ajouter un fichier. Seuls des fichiers au format texte brut peuvent être téléchargés. La possibilité de copier des textes dans la zone de texte permet de ne copier qu'une partie d'un texte ou encore de copier des textes d'un document qui n'est pas au format texte brut, par exemple depuis un document PDF ou Word.

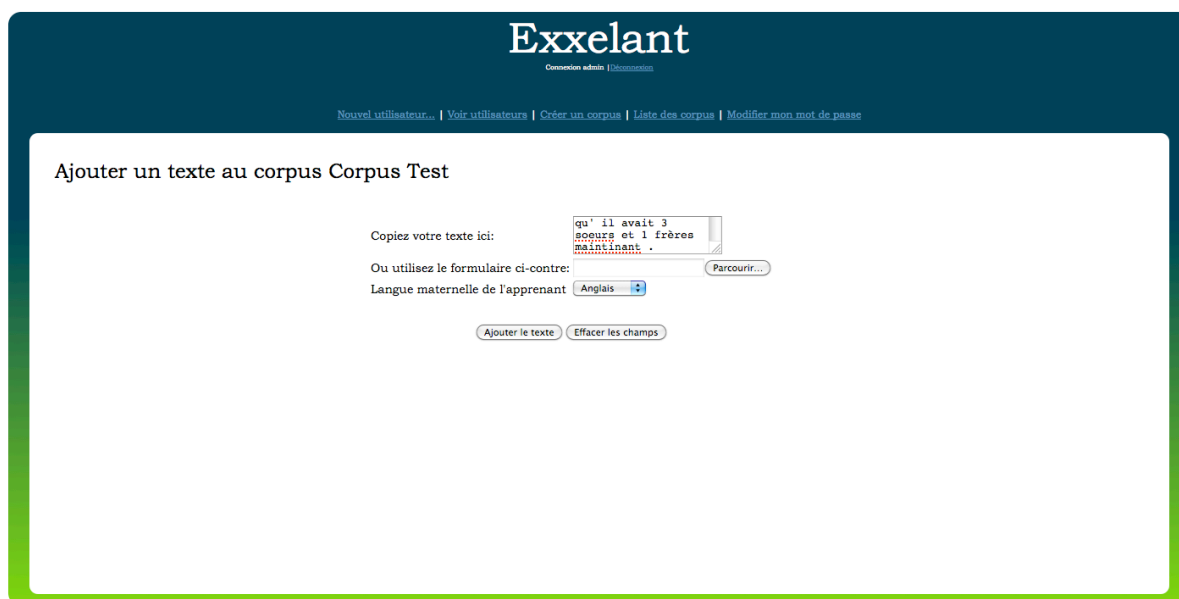


Figure 30 - Interface d'ajout de texte

La langue maternelle de l'apprenant ayant produit le texte est choisie grâce à une liste déroulante, dans un souci de normalisation. Si la langue n'apparaît pas dans la liste, l'administrateur peut choisir l'entrée *Autre*. Une fois le formulaire soumis, la conformité du formulaire est vérifiée en PHP et JavaScript. L'utilisateur doit avoir utilisé soit la zone de texte, soit le champ *Parcourir* et la langue maternelle doit être précisée. Si le formulaire est conforme, la langue maternelle ainsi que l'identifiant du corpus auquel appartient le texte sont insérés dans la base. Le texte subit ensuite un balisage morpho-syntaxique à l'aide de TreeTagger, enrichi par des scripts utilisant le dictionnaire gratuit d'ABU.

### ***Étiquetage morphosyntaxique automatique des textes***

L'étiquetage morphosyntaxique est la condition *sine qua non* pour qu'une recherche puisse être effectuée sur un texte. L'étiquetage automatique a été ajouté au sein de la plateforme et ne demande aucune intervention de la part de l'administrateur

Pour permettre l'étiquetage automatique des textes ajoutés, TreeTagger a été installé sur le serveur. TreeTagger est un outil qui prend en entrée une chaîne de caractères et renvoie pour chaque item de la chaîne son lemme et sa catégorie grammaticale. L'étiquetage des traits grammaticaux de TreeTagger n'étant pas complet, le dictionnaire gratuit de l'Association des Bibliophiles Universels est utilisé pour enrichir les données. Ce dictionnaire se présente sous la forme d'une liste de mots dans un document texte. Pour une utilisation en ligne optimale, le dictionnaire a été inséré dans une table de la base, ce qui permet une recherche rapide des traits correspondant à l'item, en fonction du lemme et de la catégorie identifiés.

Claude	NAM	Claude
se	PRO:PER	se
fait	VER:pres	faire
masser	VER:infi	masser
le	DET:ART	le
dos	NOM	dos
pendant	PRP	pendant
la	DET:ART	le
manif	NOM	manif
.	SENT	.

Figure 31 - Exemple de sortie TreeTagger

Le texte récupéré soit grâce au champ *Parcourir*, soit grâce à la zone de texte, est écrit dans un fichier au format texte brut, dont le nom est généré aléatoirement pour éviter une fois encore les collisions entre les utilisateurs. TreeTagger est exécuté sur le serveur sur le fichier texte. La sortie du programme est traitée ligne par ligne et formatée à l'aide d'une expression régulière, pour obtenir pour chaque mot son lemme, sa catégorie grammaticale et éventuellement ses traits. Les catégories sont normalisées pour correspondre aux catégories existantes dans la base de données. Si l'item n'est pas reconnu par TreeTagger, le lemme *unknown* ainsi qu'une catégorie grammaticale déterminée en fonction du contexte lui sont attribués.

Les données renvoyées par le balisage de TreeTagger sont alors enrichies grâce au dictionnaire d'ABU. Une requête prenant en paramètres la forme, le lemme et la catégorie renvoyés par TreeTagger permet alors de récupérer les traits correspondants dans le dictionnaire. Un traitement spécifique est mis en place pour les verbes. TreeTagger fait un travail de désambiguïsation lors de son traitement. Lorsqu'il s'agit d'un verbe, il renvoie en plus de la catégorie, le temps identifié. Pour ne pas perdre cette désambiguïsation, le temps des verbes dans le dictionnaire a été séparé de la liste des traits et inséré dans un champ spécifique. Ainsi, pour un verbe, la requête sur le dictionnaire prend en paramètres la forme, le lemme, la catégorie et le temps. Les informations découlant du balisage ainsi que le mot sont ensuite insérés dans la table *tokens*, item par item. Les items possédant le lemme « unknown » n'ont pas de traits grammaticaux associés.

Les résultats d'une recherche sont affichés par phrases. Il est donc nécessaire de les identifier. Ainsi, la valeur de chaque item est testée ; s'il s'agit d'une ponctuation de fin de phrase, une nouvelle entrée est ajoutée dans la table *phrases*. Un système de compteur permet d'identifier le premier et le dernier item de chaque phrase, et de stocker leurs identifiants dans la table *phrases*.

Un autre compteur permet de compter le nombre de mots d'un texte. A chaque tour de boucle, celui-ci est incrémenté. A la sortie de la boucle, une requête permet de mettre à jour les informations sur le texte, en ajoutant le nombre de mots et un identifiant au texte. L'identifiant est créé en concaténant la chaîne « *txt* » à l'identifiant numérique présent dans la base de données.

### ***Balilage manuel des erreurs***

Le corpus de base possède un étiquetage morphosyntaxique et un balilage des erreurs. Chaque erreur possède un triplet d'information la définissant, ainsi qu'une correction associée. Une telle annotation permet dans le cadre de l'application d'effectuer une recherche sur un type d'erreur ou sur une correction. Dans l'optique de faire d'EXXELANT un outil de génération d'activités, cette annotation est primordiale, puisqu'elle permet de travailler sur des erreurs authentiques et déjà corrigées. Seulement, l'annotation des erreurs ne peut s'effectuer de façon automatique.

L'annotation des erreurs sur le corpus FRIDA-BIS a été réalisée manuellement, à l'aide d'un balilage XML. Dans le cas des nouveaux corpus, ce sont les enseignants qui ont la possibilité d'annoter les textes de leurs corpus. Ces enseignants en Français Langue Étrangère ne sont pas spécialistes en informatique ; il fallait donc prévoir une interface d'annotation des erreurs simple et facile d'utilisation. La possibilité d'utiliser un éditeur XML annexe tel que Morphon a été envisagée. Cette solution avait pour inconvénient de nécessiter l'installation d'un logiciel et de demander des compétences informatiques spécifiques à l'utilisateur. Le choix s'est alors porté sur le développement d'une interface d'annotation au sein du prototype. Le balilage XML a également été abandonné. Les textes comportant déjà l'étiquetage morphosyntaxique et étant déjà découpés en items, le balilage XML et la DTD associée permettant une insertion correcte dans la base de données n'étaient plus utiles.

L'interface d'annotation des erreurs a été développée en PHP, JavaScript et Ajax. Elle se veut dynamique et facile à prendre en main. L'enseignant y accède depuis la liste de ses corpus. Cliquer sur le nom d'un corpus affiche les textes de celui-ci. Cliquer sur l'identifiant d'un texte affiche le lien vers l'interface d'annotation. Seuls les textes ne comportant pas l'annotation des erreurs sont cliquables.

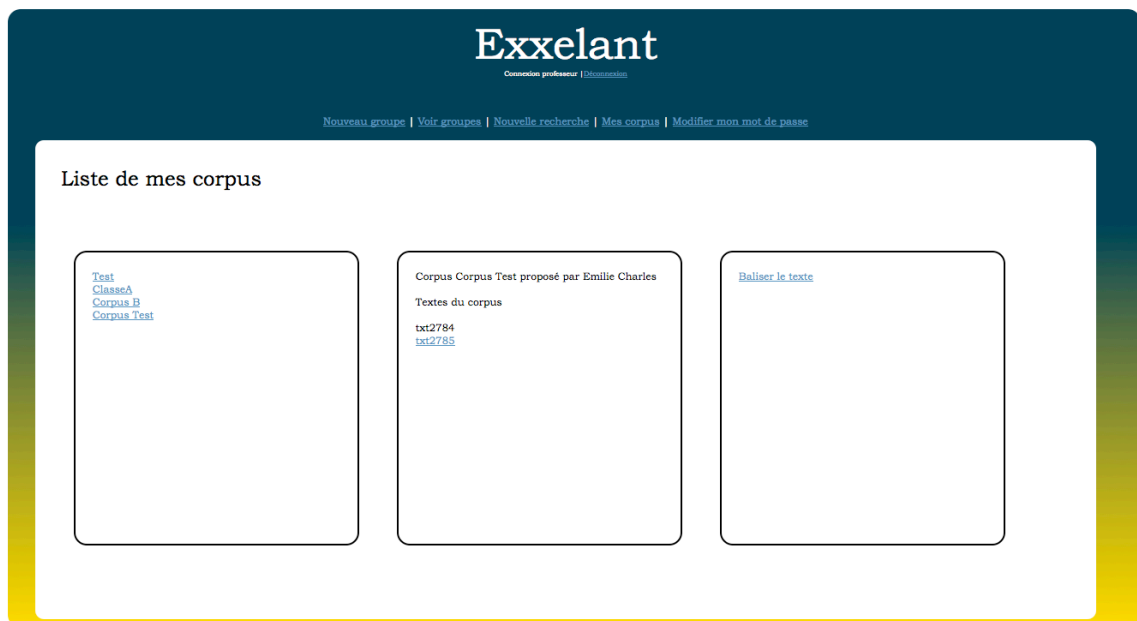


Figure 32 - Liste des corpus d'un enseignant

Le texte à annoter est reconstitué à l'écran à partir de ses items. Chaque item est entouré de balises HTML `<span>`, faisant de chaque item un objet et permettant ainsi l'utilisation de fonctions JavaScript.

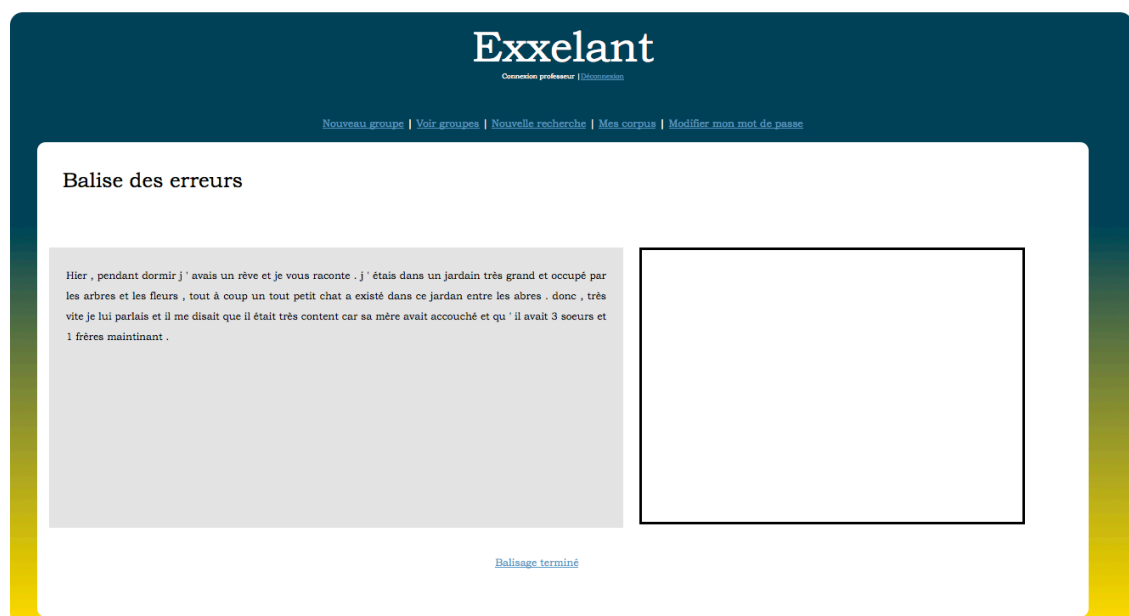


Figure 33 - Interface de balisage des erreurs

Au clic sur un mot, un formulaire apparaît à droite et le mot cliqué est coloré en rouge et souligné, indiquant qu'il est en cours d'annotation. Le formulaire sert à renseigner les informations sur l'erreur : domaine d'erreur, catégorie d'erreur, catégorie grammaticale et proposition de correction. Une fois le formulaire rempli, l'utilisateur doit cliquer sur Valider ces informations. Le formulaire est vérifié et une entrée est créée dans un tableau

stocké en variable de session. Cette entrée contient d'une part les informations sur l'erreur et le numéro de la phrase contenant l'item, et d'autre part l'identifiant de l'item. Une fois annoté, le mot est coloré en vert. Il est possible d'annoter jusqu'à quatre erreurs sur le même mot. Cliquer sur un mot annoté permet d'afficher le formulaire, (si le maximum de quatre fautes n'est pas atteint), et la liste des annotations existantes, que l'enseignant peut supprimer.

Toutes ces actions sont disponibles sur des groupes d'items. La possibilité d'annoter plusieurs items est utile par exemple dans le cas d'une inversion de mots (« *on les peut diviser* » au lieu de « *on peut les diviser* ») ou encore d'une graphie erronée utilisant plusieurs mots au lieu d'un seul (« *pour quoi* » au lieu de « *pourquoi* »).

L'annotation des erreurs doit être effectuée en une seule fois, les erreurs possédant un identifiant indiquant leur position au sein du texte. Ainsi, tout le texte doit être annoté pour permettre une insertion dans l'ordre du texte, réalisée en fonction des identifiants des items annotés. Une fois toutes les erreurs insérées dans la base, la densité d'erreur est calculée en fonction du nombre d'erreurs et du nombre de mots du texte.

### ***Modification du formulaire de recherche***

Avec la possibilité d'enrichir la base de données par de nouveaux corpus, le prototype EXXELANT ne permet plus d'interroger qu'un seul corpus défini. Il était donc nécessaire d'ajouter de nouveaux critères à la recherche.

Trois critères ont été ajoutés : *utilisateur*, *corpus* et *annotation des erreurs*. *Utilisateur* permet de choisir le corpus en fonction de la personne qui l'a ajouté à la base. *Corpus* précise le sous-corpus à sélectionner. Par exemple si *utilisateur* est précisé avec la valeur « toto », corpus ne contiendra que les sous-corpus introduits par « toto ». Si *utilisateur* n'est pas spécifié la liste déroulante de corpus contient tous les sous-corpus de chacun des utilisateurs. Enfin, il est possible de choisir de rechercher dans des textes comportant l'*annotation des erreurs* ou non.

Rechercher...

Sélection du corpus					
Utilisateur	Corpus	Annotation des erreurs	Provenance	Densité d'erreur	Longueur (nombre de mots)
FRIDA	Tous	Sans importance	toutes	toutes : 0 à 68	tous : 29 à 3803

Figure 34 - Critères concernant le corpus à interroger



### ***Perspectives d'évolution***

L'utilité et l'utilisabilité du prototype EXXELANT ont été améliorées grâce à l'ajout de ces nouvelles fonctionnalités. Celles-ci constituent les premiers pas vers le développement d'une véritable plate-forme d'apprentissage en ligne. Les perspectives d'évolution de l'application sont nombreuses. En effet, la base de textes est riche, par son étiquetage morphosyntaxique et son annotation des erreurs. Une telle base permettrait de mettre en place une génération d'activités au sein de l'application.

#### **Génération d'activités**

La création d'une activité pédagogique multimédia demande certaines compétences en informatique, qui ne sont pas forcément acquises chez les enseignants non-spécialistes. Il existe de nombreux outils aidant à la constitution de ces activités, sous forme de logiciels ou encore de LMS. À terme, EXXELANT pourrait intégrer une interface adaptée de génération d'activités, s'appuyant sur les textes de la base. En effet, ces textes bénéficient d'un étiquetage morphosyntaxique et pour certains d'une annotation des erreurs avec leurs corrections, autant de données utilisables dans la réalisation d'exercices et qui faciliteraient leur correction. Il est facile d'imaginer un exercice où l'apprenant devra identifier par exemple la catégorie grammaticale de mots, ou encore corriger les formes erronées. La génération d'activité à partir de ces textes permettrait de focaliser les exercices sur des erreurs typiques, pouvant prendre en paramètre la langue maternelle de l'apprenant. Elle permettrait également à des enseignants de créer des exercices à partir de leurs propres corpus, et ainsi de donner aux apprenants la possibilité de travailler sur leurs productions.

La mise en place d'une génération d'activité devra entraîner la mise en place d'un système de correction, pouvant être automatisé grâce aux annotations des textes, et d'un système d'évaluation de l'apprenant.

#### **Communication entre les utilisateurs**

EXXELANT, grâce à ses améliorations, permet désormais de gérer une communauté d'enseignants et d'apprenants. Il peut être intéressant de fournir aux membres de cette communauté un moyen de communiquer entre eux. Le développement d'un forum permettrait d'organiser des activités collaboratives entre les apprenants. De plus, le prototype étant conçu pour une utilisation à distance, il semble important que les apprenants puissent contacter leur tuteur depuis l'application même, grâce à une messagerie interne par exemple.

Une interface spécifique pourrait aussi être conçue pour faciliter la communication entre enseignants et administrateurs dans le cas d'une création de corpus. Il pourrait s'agir d'un formulaire donnant la possibilité de télécharger les textes du corpus en précisant les informations nécessaires sur les textes, que l'administrateur pourrait ensuite consulter, valider et ajouter depuis son compte.

### **Traces d'usage**

La conception d'un outil pédagogique multimédia doit se faire en dialogue constant entre spécialistes en didactique et spécialistes en informatique. Une fois l'outil développé, cependant, le travail de conception doit constamment être remis en cause en s'appuyant sur une analyse de l'usage de l'outil. Cette analyse est indispensable pour améliorer de façon cohérente une application. Pour permettre cette analyse, un traçage de l'activité des utilisateurs devra être mis en place dans le même temps que ces améliorations. Ce traçage permet également de fournir des rétroactions tant aux enseignants qu'aux apprenants.

**3.**

**De la nécessité des traces dans l'outil pédagogique  
multimédia**

L'apprentissage assisté par ordinateur est un outil stratégique pour l'amélioration de la formation à distance. Un enseignant, dans un contexte de formation en présentiel, est capable de déceler les failles dans ses séquences pédagogiques au vu du parcours et des évaluations et de fournir des rétroactions pertinentes aux apprenants. De la même façon, il est indispensable de fournir des outils pédagogiques multimédia qui d'une part peuvent être adaptés en fonction des résultats de l'utilisation de ceux-ci, et d'autre part donnent aux apprenants des retours sur leur apprentissage. Contrairement au présentiel, ceci ne peut être effectué directement. Un traçage de l'utilisation de l'outil doit donc être mis en place.

### **3.1. Traçage de l'utilisation**

Tracer l'utilisation d'une application consiste de façon générale à enregistrer plusieurs types de données résultant d'interactions avec l'outil :

- les données de connexion, comprenant le nom d'utilisateur, l'heure de connexion, le numéro de la session, etc.
- les informations fournies par les clics souris (les pages visitées, les heures de consultation associées, etc.)
- les données produites par l'utilisateur, pouvant comporter les résultats obtenus lors d'une évaluation ou encore les messages postés sur un forum.

Ces enregistrements constituent des traces, qui reflètent les interactions de l'utilisateur et permettent de reconstituer son parcours au sein d'une application. Notons cependant que ces traces n'offrent pas une vue omnisciente de l'activité. En effet, ces données ne fournissent pas d'information sur le contexte d'utilisation. Pour pallier ce manque, plusieurs procédés peuvent être mis en place.

Avouris *et al.* (2005) proposent l'utilisation de médias audio et vidéo pour compléter les données des traces d'interaction. Héraud *et al.* (2005) choisissent plutôt d'employer un observateur humain. Ils ajoutent également un traçage de l'activité de l'utilisateur sur son poste, hors de l'application. L'usage de tels procédés, en lien avec les traces d'interaction, rend possible une reconstitution relativement complète de l'activité effective de l'utilisateur et offre des perspectives d'analyses plus exhaustives, limitant ainsi d'éventuelles conclusions erronées. Cependant, une telle observation de l'apprenant présente des inconvénients. Que ce soit avec les enregistrements audio ou vidéo, ou en présence d'un tuteur, les apprenants ont conscience de l'observation et peuvent se sentir

listés ou espionnés. Ce sentiment rend l'utilisation de l'application moins spontanée et par conséquent, altère l'authenticité des traces d'interaction. De plus, la mise en place d'une observation limite l'unité d'espace et éventuellement de temps, ce qui entre en contradiction avec le concept de formation à distance. Il est envisageable d'organiser ponctuellement des sessions d'observation, seulement l'adaptation de l'outil se ferait alors également de façon ponctuelle et non pas en continu. Du fait de ces inconvénients, nous ne traiterons ici que des traces d'interaction.

### **3.2. A qui sont destinées les traces ?**

Les traces d'interaction constituent une masse de données importante, utilisable par les différents acteurs impliqués dans un outil pédagogique multimédia, concepteur-développeur, enseignant ou apprenant, avec pour chacun un intérêt spécifique.

#### ***Le concepteur-développeur***

Un concepteur-développeur peut profiter de ces données dans une optique d'adaptation et d'amélioration, pédagogique ou technique. Lors de la conception, un scénario pédagogique prédictif est construit. Il décrit le déroulement prévu de la situation d'apprentissage et définit les activités pédagogiques, la temporalité, les modalités d'évaluation et les objectifs didactiques visés. Grâce aux traces, il est possible de retracer un scénario effectif d'utilisation. La comparaison des scénarii prédictif et effectif peut permettre de déceler des anomalies dans la réalisation des séquences d'apprentissage. Ces anomalies doivent être analysées et peuvent amener une reconception de l'outil, tant au niveau technique que pédagogique.

Prenons comme exemple d'anomalie révélée par les traces un échec systématique à une activité. Le concepteur peut émettre plusieurs hypothèses pour expliquer ce phénomène. Il peut s'agir d'un problème dans la conception pédagogique : un manque de clarté dans les consignes, une scénarisation pédagogique mal conçue, des objectifs didactiques trop élevés ou encore que les connaissances requises ont mal été délivrées à l'apprenant. Le problème peut aussi se situer au niveau du parcours de l'apprenant, par exemple si les séquences de cours nécessaires à la réussite de cette activité n'ont pas été consultées. Il convient ici de s'interroger sur la clarté des indications quant au déroulement à suivre, au vu du nombre massif d'échecs, et donc de se pencher également sur la scénarisation de la séquence. Enfin, il peut s'agir d'un problème d'ordre technique : erreur

dans la correction automatique, page inexistante, problème d'affichage des ressources multimédia, etc. Le concepteur-développeur se doit d'envisager toutes les hypothèses, de les confirmer ou de les infirmer en utilisant entre autres les traces, et d'effectuer ensuite une reconception de l'outil afin de l'améliorer et de l'adapter.

### ***L'enseignant***

Les traces sont essentielles pour les enseignants pour un suivi optimum de leurs apprenants. Elles permettent d'obtenir des données statistiques sur les performances d'une classe ou d'un apprenant en particulier. Grâce à ce retour, l'enseignant peut identifier les points forts et les points faibles des apprenants et contrôler la compréhension et les compétences acquises. Il lui est possible d'identifier également au cas par cas les apprenants en difficulté qui nécessitent donc une attention particulière.

Au delà de la rétroaction fournie sur les apprenants, c'est aussi une rétroaction sur son enseignement. De bons résultats en moyenne à une activité portant sur une séquence de cours en présentiel auraient tendance à signifier une bonne transmission des connaissances et inversement. En somme, les traces, grâce aux rétroactions qu'elles permettent de construire, donnent à un enseignant les informations nécessaires à l'optimisation de son cours. Il peut planifier et adapter le contenu de ses séquences, insister sur les points comportant des lacunes et fournir une aide supplémentaire aux apprenants en difficulté.

### ***L'apprenant***

Dans un contexte d'enseignement en présentiel, un apprenant reçoit des rétroactions sur son activité par son enseignant, sous forme de remarques, de notes ou encore d'appréciations. Ces rétroactions aident l'apprenant à comprendre ses erreurs et à prendre conscience de son niveau de compétence. Il est difficile de fournir des rétroactions aussi complètes, personnalisées et contextualisées de façon automatique dans un outil de formation à distance. Cependant, les traces d'interaction permettent d'offrir aux apprenants les informations nécessaires pour s'auto-évaluer.

Grâce à celles-ci, les apprenants peuvent visualiser leur parcours, savoir où ils en sont et avoir une vue d'ensemble de leur activité et de leur progression. Ils peuvent également analyser leur performance à l'aide de leurs résultats aux évaluations par exemple, et identifier leurs propres points faibles. En organisant une réflexion sur leurs erreurs, les apprenants peuvent eux-mêmes déterminer les points à renforcer et planifier en

fonction de ceux-ci les activités à travailler pendant les sessions suivantes et s'entraîner sur les exercices sur lesquels ils ont eu des difficultés. Donner aux apprenant l'accès à ces données, c'est stimuler leur sens de l'auto évaluation ; ils sont ainsi plus impliqués dans leur processus d'apprentissage.

Il est évident que pour effectuer ce travail de réflexion et d'auto-évaluation, les traces ne peuvent être présentées sous leur forme brute. Elles doivent être représentées de façon adaptée aux apprenants. Il en est de même en ce qui concerne les concepteurs et les enseignants.

### **3.3. Comment représenter les traces ?**

Pour qu'un concepteur, un enseignant ou un apprenant puisse profiter des traces d'interaction, il ne suffit pas d'en faire la collecte. Les données récoltées doivent être représentées de façon à être compréhensibles et analysables. Nous avons vu précédemment que concepteur, enseignant et apprenant peuvent utiliser les mêmes données avec des objectifs différents. Ainsi, la représentation des traces doit prendre en compte non seulement leurs compétences, mais aussi leurs objectifs.

Le traçage de l'activité d'un utilisateur est un procédé mis en place au sein de la plupart des plates-formes pédagogiques actuelles. Celles-ci enregistrant de nombreuses informations, mais proposent bien souvent un simple affichage tabulaire. Prenons l'exemple de la plate-forme Moodle<sup>18</sup>. Voici l'interface de consultation des traces accessible à un enseignant :

---

<sup>18</sup> <http://moodle.org/>

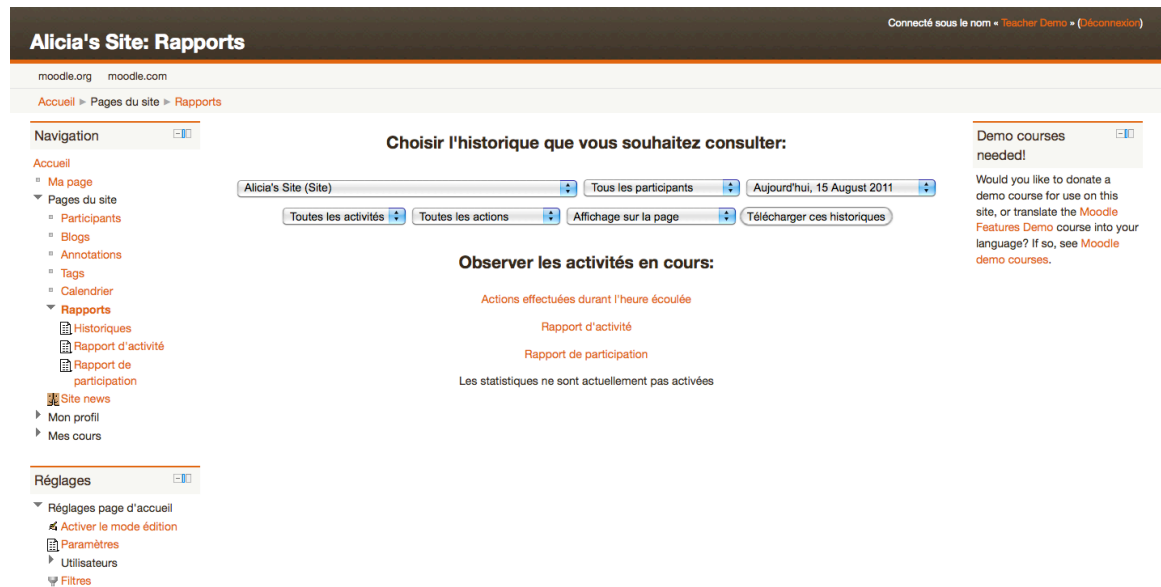


Figure 35 - Interface de consultation de traces de la plate-forme Moodle

Des listes déroulantes permettent de trier et d'affiner les résultats selon cinq critères : la séquence de cours considérée, l'apprenant, la date, l'activité de la séquence et l'action réalisée. Il est possible d'afficher les résultats directement sur la page ou de les télécharger.

L'affichage dans la page se fait sous forme de tableau, une structure pratique pour afficher de nombreux résultats, mais qui n'offre pas de vue d'ensemble.

Cours	Heure	Adresse IP	Nom complet du cours	Action	Information
Hotness home	Mon 15 August 2011, 07:50	78.251.117.65	Teacher Demo	course report log	Alicia's Site
Features Demo	Mon 15 August 2011, 07:50	66.249.71.231	Guest User	folder view	A directory
Caracteristicas	Mon 15 August 2011, 07:50	95.120.242.77	Guest User	assignment view	Tarea de subida avanzada de archivos
Film Studies	Mon 15 August 2011, 07:50	209.74.201.47	Admin User	course view	Higher Education Film Studies Module
WATER HOME	Mon 15 August 2011, 07:50	66.249.71.231	Guest User	forum view discussion	Greetings
Japanese demo	Mon 15 August 2011, 07:50	119.235.237.92	Guest User	assignment view	オンラインテキスト
Features Demo	Mon 15 August 2011, 07:50	66.249.71.231	Guest User	assignment view	Online text
Μάθημα Επιδειξης Δυνατοτήτων Moodle	Mon 15 August 2011, 07:50	67.195.111.31	Guest User	glossary view	Οριζόμενο από τον Εκπαιδευτή Γλωσσάρι
WATER HOME	Mon 15 August 2011, 07:50	66.249.71.231	Guest User	forum view discussion	Hey there !
무들기능 익힘터	Mon 15 August 2011,	66.249.71.231	Guest User	folder view	자료 참고

Figure 36 - Affichage des traces d'interaction dans la plate-forme Moodle

Analyser un tel tableau constitue une démarche longue et fastidieuse, impliquant l'utilisation d'outils externes à la plate-forme pour être efficace. Pouvoir spécifier des



critères est intéressant, mais cette représentation ne semble pas adaptée à un enseignant, car trop technique et peu claire.

Cet exemple souligne la nécessité d'une représentation des traces qui tient compte du destinataire pour une analyse optimale.

### ***Représentation des traces à destination d'un concepteur-développeur***

Les traces d'interaction sont utiles à un concepteur pour déceler les anomalies dans l'utilisation d'une application par rapport à un scénario prédictif défini lors de la conception pédagogique. Le concepteur peut reconstituer manuellement un scénario effectif grâce aux données des traces, mais cette reconstitution demande du temps et du travail et peut rendre décourageante l'idée de reconcevoir l'outil régulièrement en fonction des résultats de l'analyse des traces. Ainsi serait-il intéressant d'automatiser ce processus de reconstitution pour fournir au concepteur une représentation directement analysable.

Une comparaison ne peut s'effectuer qu'entre objets de même type. La représentation des traces doit donc s'appuyer sur la forme du scénario prédictif. Iksal et Choquet (2005) proposent l'usage d'IMS Learning Design, un langage de modélisation pédagogique largement utilisé en ingénierie pédagogique. IMS Learning Design utilise le langage XML pour décrire un scénario d'apprentissage en termes de rôles pris par les apprenants, d'activités et d'environnement. Les traces collectées sont formatées sur le même modèle pour permettre de faire une comparaison.

Un scénario prédictif peut aussi être représenté graphiquement.

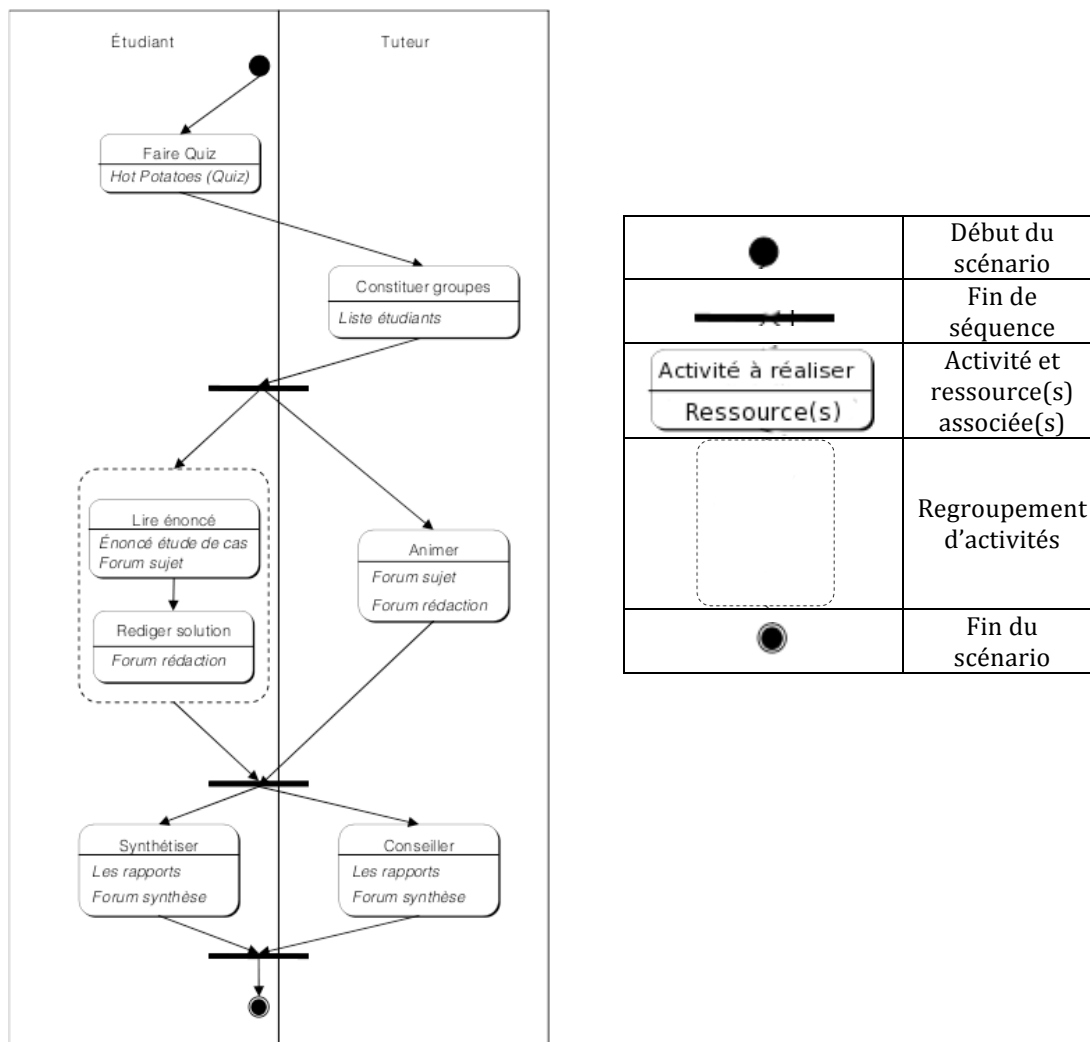


Figure 37 - Représentation graphique du scénario pédagogique <sup>19</sup>

La représentation graphique est également possible et pourrait même découler d'un formatage XML des données. Cette représentation a l'avantage de demander moins de compétences techniques pour être analysée.

### ***Représentation des traces à destination d'un enseignant***

Dans la représentation des traces à destination d'un enseignant, il est indispensable de prendre en compte ses compétences. Un enseignant n'est ni un analyste, ni un informaticien. Les traces d'interaction doivent donc lui être présentées sous une forme facilement compréhensible et analysable.

Les traces peuvent fournir à l'enseignant des informations sur l'ensemble de la classe et sur un apprenant en particulier. Il a besoin d'avoir en premier lieu une vue d'ensemble des résultats. Une représentation tabulaire telle que celle utilisée dans la plate-

<sup>19</sup> Source : <http://jeanphilippepernin.net/wiki/>

forme Moodle, permet de faire un tri dans les données grâce aux listes déroulantes et d'afficher de nombreux résultats. Elle ne permet pas cependant d'ordonner les résultats selon un critère choisi (nom de l'apprenant, date, etc.), opération qui pourrait être réalisée en utilisant par exemple un système d'entêtes dynamiques. De plus, la quantité de données est telle qu'il est difficile de repérer une tendance. Une représentation graphique permet au contraire d'avoir une vue de l'activité des apprenants à grande échelle. C'est grâce à cette vue que l'enseignant peut déterminer quels points sont à étudier plus en détail.

Les notes obtenues aux évaluations permettent d'établir des statistiques et notamment un pourcentage de réussite aux activités. Ces statistiques peuvent être représentées sous la forme d'un histogramme comme suit :

### **Pourcentage de réussite de tous les apprenants sur chaque activité**

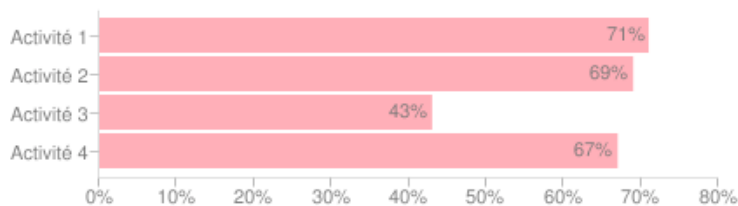


Figure 38 - Histogramme représentant le pourcentage de réussite total à chaque activité

Au premier coup d'œil, l'enseignant remarque que l'activité 3 présente un pourcentage de réussite largement inférieur aux autres activités. Cela peut venir d'un apprenant en particulier, qui a échoué à cette activité et baisse le pourcentage de réussite. Il serait bon que l'enseignant ait accès au détail des résultats obtenus à l'activité 3 pour parfaire son diagnostic.

### **Pourcentage de réussite à l'activité 3 pour tous les apprenants**

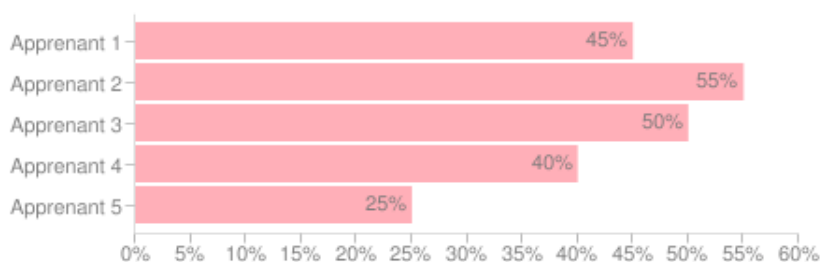


Figure 39 – Histogramme représentant le pourcentage de réussite total à l'activité 3

Le diagramme fait ressortir des taux de réussite moyens pour tous les apprenants. Il est possible de supposer que cette activité est difficile pour les apprenants. Notons que l'apprenant 5 a obtenu un pourcentage de réussite relativement inférieur aux quatre autres

apprenants. Il pourrait s'agir d'un apprenant en difficulté. Pour confirmer cette hypothèse, l'enseignant pourrait avoir à sa disposition un diagramme offrant le détail des résultats des apprenants à toutes les activités.

### **Pourcentage de réussite de chaque apprenant à chaque activité**

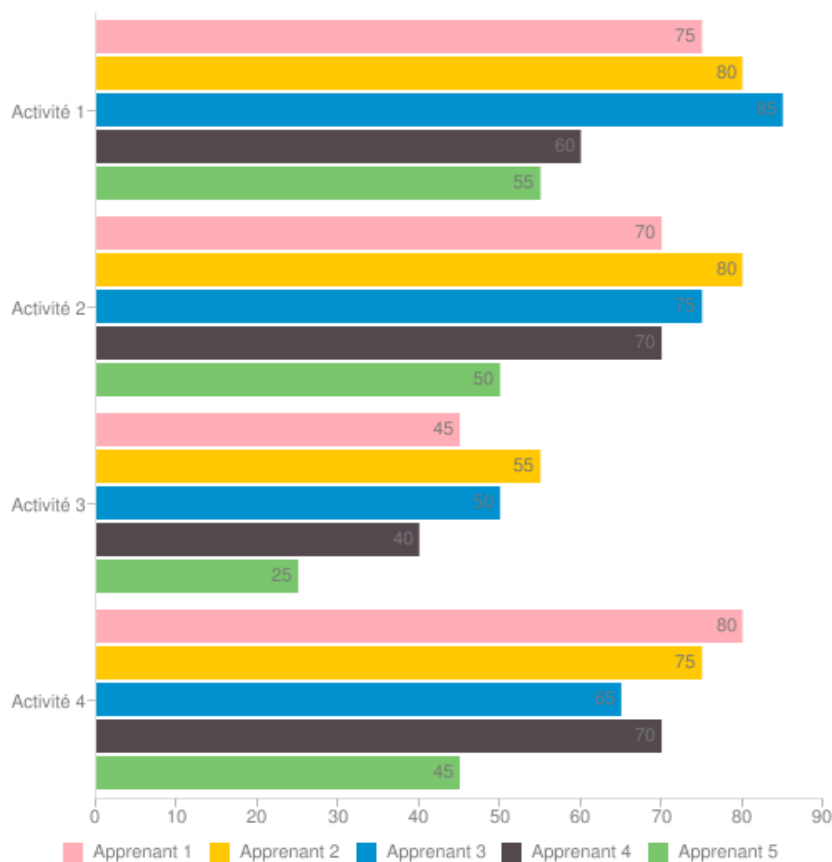


Figure 40 - Histogramme représentant le pourcentage de réussite de chaque apprenant à chaque activité

Les difficultés de l'apprenant 5 sont mises en valeur à l'aide de cet histogramme. Il permet également de situer le niveau des autres apprenants. Une fois un apprenant en difficulté identifié, l'enseignant peut profiter d'un accès aux données de cet apprenant, notamment ses résultats aux activités, les réponses fournies et son parcours au sein de l'application.

Les résultats peuvent être affichés de la même façon que précédemment, à l'aide d'un histogramme.

## Pourcentage de réussite aux activités pour l'apprenant 5

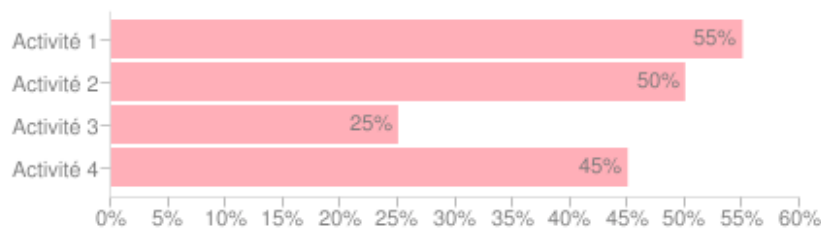


Figure 41 – Histogramme représentant le pourcentage de réussite de l'apprenant 5 à chaque activité

En ce qui concerne le parcours, plutôt que de limiter son affichage à un tableau récapitulatif ordonné chronologiquement, il est possible ici encore d'imaginer une représentation graphique qui soit facile à analyser. En s'appuyant sur le scénario pédagogique, un parcours « correct » peut être défini. Prenons un exemple simple :

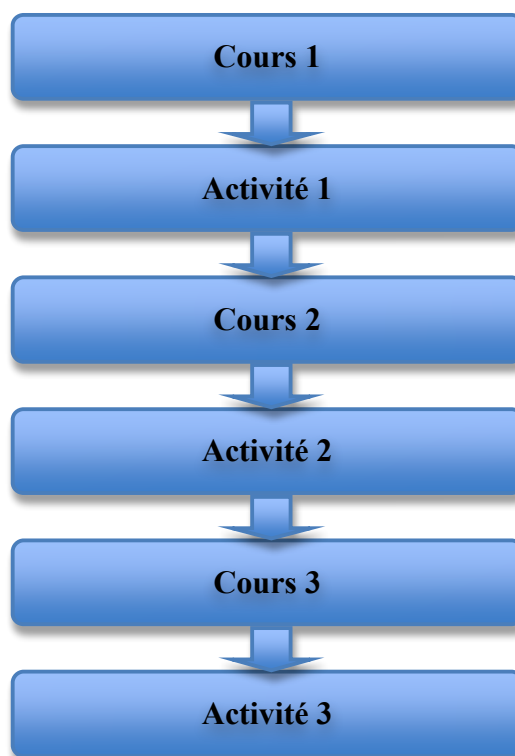


Figure 42 - Représentation graphique du parcours à suivre

À ce parcours peuvent s'ajouter des digressions autorisées, comme l'accès au forum ou à des activités et cours déjà consultés. Lors de la représentation graphique, un code couleur peut être ajouté pour plus de clarté : vert pour un parcours autorisé, bleu pour une digression, rouge pour un parcours incorrect. Le pourcentage de réussite peut également être ajouté à côté de l'activité correspondante. Considérons maintenant le parcours de l'apprenant 5, supposé en difficulté, représenté ainsi :

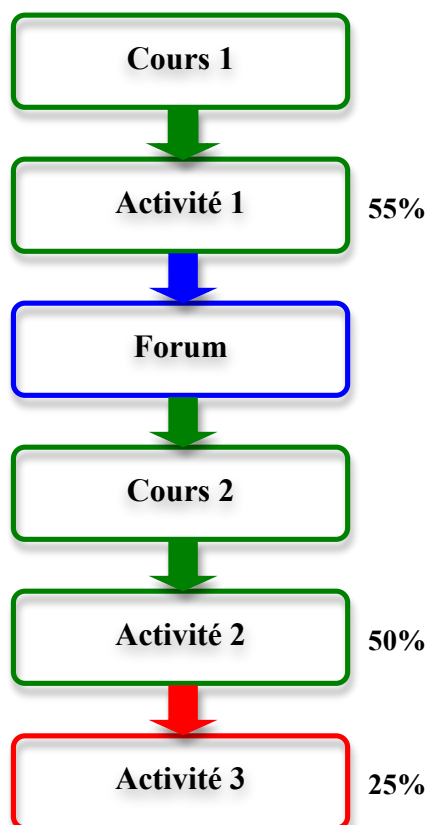


Figure 43 - Représentation graphique du parcours de l'apprenant 5

L'enseignant voit rapidement que le parcours « correct » n'a pas été suivi. Cela n'explique qu'en partie les mauvais résultats obtenus à l'activité 3. En effet, au vu de ses résultats globaux, comparés aux résultats des autres apprenants, il semble qu'il ait des difficultés malgré tout.

L'emploi de représentations graphiques pourrait simplifier la détection des informations pertinentes. Elle ne peut cependant pas être utilisée seule, puisque l'enseignant peut souhaiter avoir accès à des informations précises, comme les réponses à un exercice, qui peuvent difficilement prendre une forme graphique. De plus, la notion de pourcentage de réussite peut ne pas être pertinente, si l'exercice ne comporte que peu de questions par exemple. La représentation graphique peut alors venir en complément ou en remplacement de certains tableaux de données seulement.

Tous les histogrammes présentés dans cette partie ont été réalisés à l'aide de Chartify<sup>20</sup>, un plugin jQuery<sup>21</sup> qui prend en entrée un simple tableau HTML. L'intégration de tels graphiques ne relève donc pas d'une prouesse technique et peut être mise en place facilement dans une application web.

<sup>20</sup> <http://www.daytona.se/chartify>

<sup>21</sup> <http://jquery.com/>

## ***Représentation des traces à destination d'un apprenant***

Les données des traces représentent pour l'apprenant l'historique de son activité. Celui-ci lui permet de voir sa progression dans l'application et de développer son sens de l'auto-évaluation : identification de ses points faibles, orientation vers l'activité pertinente pour combler ses lacunes ou encore retravailler les exercices qui lui ont posé problème. En somme, les traces permettent de construire un carnet de bord. Tout comme pour un enseignant, la représentation graphique des traces peut faciliter l'analyse de l'apprenant. Elle apporte également un côté ludique, permettant de l'intéresser. Les informations principales nécessaires à l'apprenant sont son avancement dans l'application, son parcours, sa progression et ses résultats.

Pour indiquer l'avancement, plutôt que de seulement signifier le nombre d'étapes restantes ou le pourcentage réalisé, celui-ci peut être présenté sous la forme d'un diagramme circulaire.

## **Avancement**

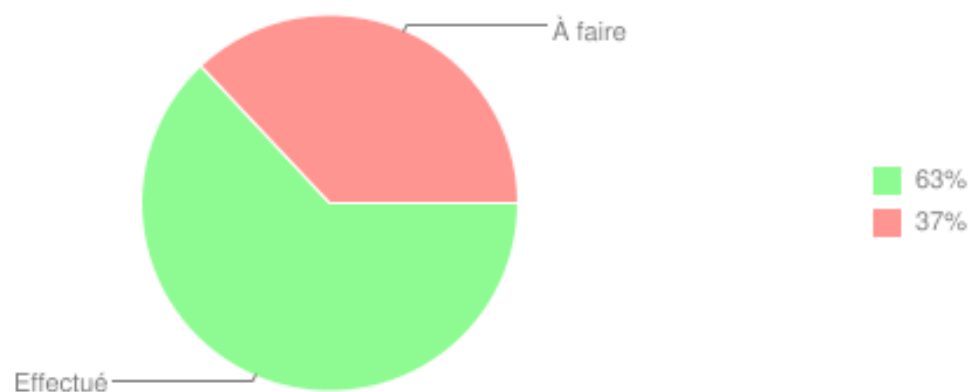


Figure 44 - Diagramme circulaire représentant l'avancement. Réalisé à l'aide de Chartify

Le pourcentage accompli est mieux visualisé et voir le diagramme évoluer est encourageant. Selon les apprenants, s'il s'agit d'un public jeune par exemple, il peut être question d'une représentation encore plus imagée, comme un potager dont les légumes poussent au fur et à mesure des activités réalisées, ou encore un puzzle qui se complète peu à peu.

Fournir à l'apprenant son parcours lui permet de reprendre où il s'est arrêté et de voir en détail ce qu'il a effectué. La même représentation que pour les enseignants peut être utilisée. Dans un souci de guidage, des avertissements peuvent être ajoutés si la

séquence de cours nécessaires à une activité n'a pas été consultée et l'activité suivante à réaliser selon le parcours « correct » peut être précisée.

La progression de l'apprenant peut s'exprimer à travers ses résultats successifs aux exercices. Celle-ci peut être représentée sous la forme d'une courbe.

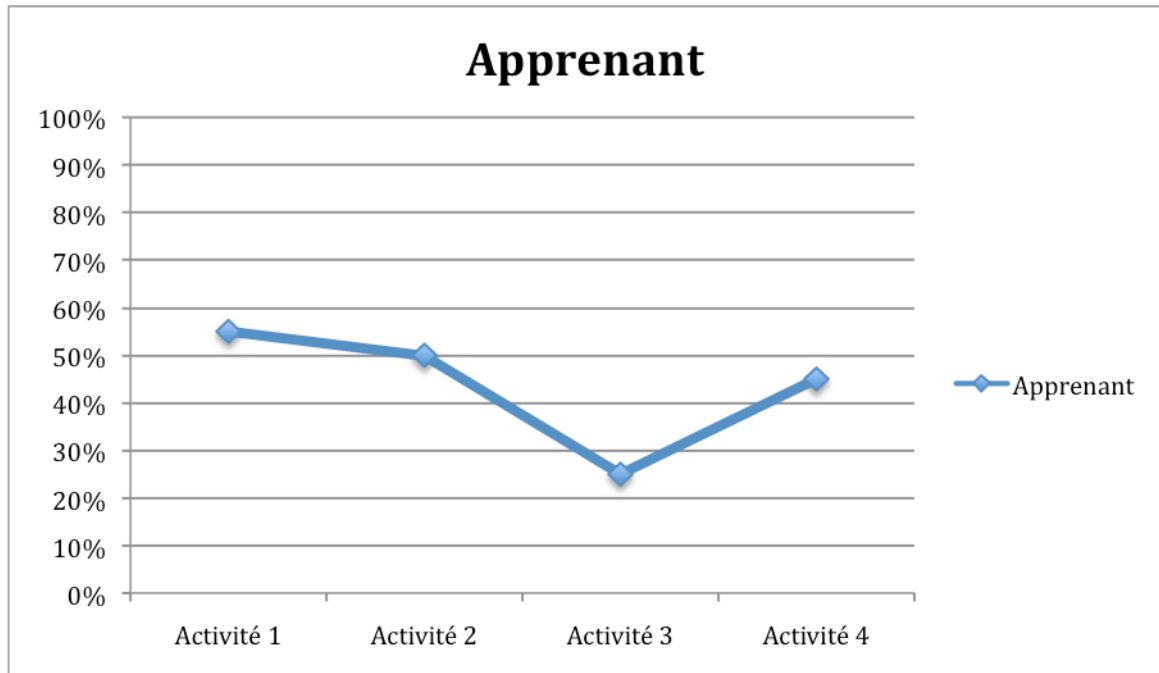


Figure 45 - Courbe de progression d'un apprenant en fonction de ses résultats aux activités

L'apprenant a aussi besoin d'avoir une vue d'ensemble de ses résultats. L'historique présenté pour l'enseignant peut tenir ce rôle.

### **Pourcentage de réussite aux activités pour l'apprenant 5**

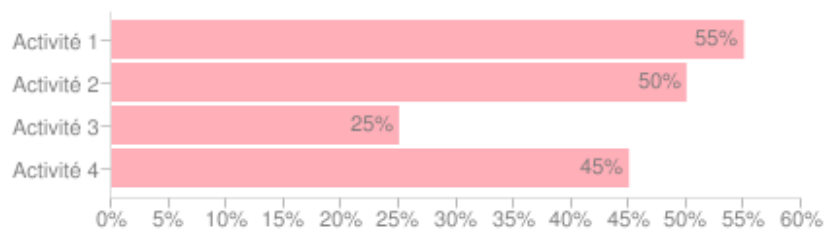


Figure 46 - Histogramme représentant le pourcentage de réussite de l'apprenant 5 à chaque activité



### **3.4. Le cas d'EXXELANT**

Bien que l'utilisation d'EXXELANT ne soit pas actuellement définie par un scénario pédagogique et que le prototype ne comprenne ni séquence de cours, ni exercice, un traçage peut tout de même être mis en place. Celui-ci pourrait permettre au concepteur-développeur de détecter des anomalies techniques dans l'application, au niveau de l'ajout de textes ou de l'interface de recherche par exemple, ou d'informer un enseignant sur les interrogations effectuées par ses apprenants.

Au niveau du concepteur-développeur, il peut être intéressant de lui fournir des informations sur les requêtes effectuées sur la base de données, notamment les insertions (ajout de texte, insertion du balisage manuel des erreurs). Par exemple, calculer le temps d'exécution de ces requêtes permettrait d'identifier les points à optimiser dans la conception technique. La structure des données insérées pourrait également être contrôlée. Il s'agirait alors de définir un modèle de cette structure et d'effectuer une comparaison. Dans le cas du balisage d'une erreur, les informations nécessaires sont l'identifiant du token erroné, l'identifiant de la phrase contenant le token, la correction, le domaine d'erreur, la catégorie d'erreur ainsi que la catégorie grammaticale du token. Si les données insérées ne correspondent pas à ce modèle, il existe alors une anomalie. La représentation de ces traces pourrait prendre la forme de statistiques pour le temps d'exécution, pouvant inclure en paramètre la longueur des textes. En ce qui concerne le contrôle de la structure des données, il n'est pas nécessaire de fournir au concepteur le détail de toutes les informations insérées ; seules les anomalies détectées pourraient être notifiées sous la forme d'avertissements.

Tracer les recherches effectuées par un apprenant et leurs éventuels affinages peut être utile à l'enseignant. Fournir seulement le détail des critères pour chaque recherche ne semble pas être pertinent. Ici encore, des statistiques pourraient être calculées, permettant d'identifier des tendances dans l'utilisation de l'interface d'interrogation, des critères prédominants dans l'ensemble des recherches effectuées (recherches portant plutôt sur erreur ou inversement, nombre de critères spécifiés, domaine d'erreur récurrent, etc.), pouvant éventuellement être des indications sur les lacunes de l'apprenant.

Un traçage de l'utilisation présenterait un intérêt certain pour le prototype EXXELANT mais représente une masse de travail importante. Il n'a donc pas été

développé ni défini plus précisément dans le cadre de ce stage, la courte durée de celui-ci ne le permettant pas.

### **3.5. Conclusions sur les traces d'utilisation**

Le traçage de l'utilisation d'un outil pédagogique multimédia permet d'adapter l'outil aux utilisateurs et d'apporter une nouvelle dimension pédagogique, notamment sous la forme de rétroactions, profitable aux enseignants et aux apprenants. Pour qu'il soit efficace, il doit être pensé en fonction des acteurs concernés. Pour être plus efficace encore, la conception technique doit prendre en compte ce traçage. Inclure des métadonnées pour décrire les domaines de compétences mis en jeu ou encore les thèmes abordés dans les activités par exemple permettrait d'obtenir des statistiques plus précises regroupées en termes d'objectifs didactiques.

La pertinence des traces passe également par une représentation adaptée, afin d'optimiser leur analyse. Limiter les traitements à effectuer par l'utilisateur en fournissant des données formatées de façon automatique représente un gain de temps et d'efficacité dans l'analyse. Par ailleurs, l'avancement des nouvelles technologies amène de plus en plus d'outils libres et gratuits permettant cette automatisation (bibliothèques JavaScript, programmes développés en PHP, etc.) dont il faut savoir tirer partie.

Les exemples présentés précédemment sont volontairement simples, pour permettre une compréhension globale de la notion de traces. Il existe cependant des outils bien plus complexes de représentation et d'exploitation des traces : l'approche MUSERTE (Modéliser les Utilisateurs et les Tâches pour Tracer l'Expérience) présentée par Mille et Prié (2006), l'outil ColAT (Collaboration Analysis Tool) d'Avouris et al. (2005), le système DIAS (Discussion Interaction Analysis System) de Bratitsis et Dimitracopoulou (2005), le métalangage UTL (Usage Tracking Language) proposé par Iksal et Choquet (2005) ou encore le système GISMO (Graphical Interactive Student Monitoring System) de Mazza et Milani (2005).

L'utilité des traces est indéniable. Il faut cependant garder en tête qu'elles ne sont que des indications et ne pas tirer de conclusions hâtives de ce qu'elles peuvent mettre en lumière. Il s'agit de mener une investigation grâce à ces données, en conservant un regard critique.

## Conclusion

L'utilisabilité du prototype EXXELANT a été grandement améliorée grâce aux développements réalisés au cours de ce stage. La modernisation et la dynamisation de l'interface en font désormais un outil plus simple et plus agréable à utiliser et la fonctionnalité d'export de phrases et de texte permet aux enseignants une exploitation plus aisée des résultats d'une recherche. EXXELANT est d'autre part devenu une véritable plateforme collaborative en permettant aux enseignants d'ajouter leurs propres textes. Ceci permettra non seulement un enrichissement certain de la base de textes, mais aussi de nouvelles perspectives didactiques, puisqu'il sera possible de travailler sur des productions personnelles. Ces avancées ne sont que les premiers pas vers le développement d'une plateforme pédagogique. Il reste encore du chemin à parcourir, à commencer par la mise en place d'une génération d'activités à partir des textes des corpus. Dans cette optique, un traçage de l'usage du prototype en l'état actuel permettrait d'avoir un retour sur sa viabilité, technique notamment, et d'aller plus loin dans le développement.

Ce stage de quatre mois au sein du LIDILEM m'a permis de mettre en œuvre les compétences informatiques et didactiques acquises au cours de ma formation de master. Bien qu'intégré au sein d'un laboratoire de recherche et non pas d'une entreprise, ce stage a représenté pour moi une réelle expérience professionnelle. Mes encadrants se sont positionnés en tant que commanditaires, exprimant des demandes en me laissant carte blanche pour y répondre. Des réunions hebdomadaires m'ont permis tout au long de mon stage d'exposer mes propositions et de poser mes questions afin de cerner au mieux leurs attentes. La conception technique des améliorations de l'outil m'a permis de consolider mes connaissances du langage PHP, de développer ma maîtrise d'Ajax et JavaScript et de savoir gérer une base de données complexe. J'ai également du effectuer un vrai travail de conception pédagogique, m'efforçant de produire une application simple et intuitive dans son utilisation, adaptée la littératie informatique limitée du public de destination. J'ai aussi effectué mes premiers pas dans le domaine de la recherche en étudiant l'utilité des traces dans un outil de formation multimédia.

## Bibliographie

Antoniadis G., Ponton C., Zampa V. (2010). Exxelant et Mirto : Deux exemples d'environnement d'ALAO intégrant des outils TAL. In *Multilinguisme et traitement des langues naturelles*. PUQ. Montréal. Canada.

Avouris N., Komis V., Fiotakis G., Margaritis M., Voyiatzaki E. (2005). Logging of fingertip actions is not enough for analysis of learning activities. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Barre V., El-Kechaï H., Choquet C. (2005). Re-engineering of collaborative e-learning systems : evaluation of system, collaboration and acquired knowledge qualities. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Bratitsis T. & Dimitracopoulou A. (2005). Data Recording and Usage Interaction Analysis in Asynchronous Discussions : The D.I.A.S. System. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Duval P., Merceron A., Scholl M. (2005). Recording and Analyzing usage of Learning Systems : experience and reflection with an online course on Java Programming. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Feng M. & Heffernan N. T. (2005). Informing Teachers Live about Student Learning: Reporting in Assistentment System. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Gilbert-Darras F., Delozanne E., Le Calvez F., Merceron A., Labat J. M., Vandebrouck F. (2005). Towards a Design Pattern Language to Track Student's Problem-Solving Abilities. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Granger, S., Kraif, O., Ponton, C., Antoniadis, G., Zampa, V. (2007). Integrating learner corpora and natural language processing : A crucial step towards reconciling technological sophistication and pedagogical effectiveness. *RecALL*, 19(3): 252-268.

Heraud J. M., Marty J. C., France L., Carron T. (2005). Helping the Interpretation of Web Logs : Application to Learning Scenario Improvement. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Hulshof C. D. (2005). Usage Analysis for Studying Scientific Discovery Learning Processes: Results from Three Different Learning Environments. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Iksal S., & Choquet C. (2005). Usage Analysis Driven by Models in a Pedagogical Context. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Kumar A. (2005). Usage Analysis in Tutors for C++ Programming. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Laflaquière J., Settouti L. S., Prié Y., Mille A. (2007). Traces et inscriptions de connaissances. *18<sup>ème</sup> Journées Francophones d'Ingénierie des Connaissances*, Grenoble, France, 2007.

Loghin G. C. (2006). Aide à la compréhension du comportement de l'utilisateur par la transformation des traces collectées. *Ières Rencontres Jeunes Chercheurs en EIAH, RJC-EIAH'2006*, pages 88 à 95.

Le Calvez F., Giroire H., Tisseau G., Duma J. (2005). Analysis of tracks from student's use of a system to teach a problem solving method. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Mazza R. & Milani C. (2005). Exploring Usage Analysis in Learning Systems: Gaining Insights From Visualisations. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Mbala A., Reffay C., Anyouzoa A. G. N. (2005). Supporting Distributed Collaborative Learning with Usage Analysis Based Systems. In Workshop on Usage Analysis in

Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Mille A. & Prié, Y. (2006). Une théorie de la trace informatique pour faciliter l'adaptation dans la confrontation logique d'utilisation/logique de conception. *Transition*, 12-12. Disponible sur <http://liris.cnrs.fr/publis/?id=3002>

Muehlenbrock M. (2005). Automatic Action Analysis in an Interactive Learning Environment. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Nicaud J. F., Chaachoua H., Bittar M., Bouhineau D. (2005). Student's modelling with a lattice of conceptions in the domain of linear equations and inequations. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Sander E., Nicaud J. F., Chachoua, Croset M. C. (2005). From usage analysis to automatic diagnosis : The case of the learning of algebra. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

Settouti L. S. (2006). Systèmes à base de trace pour l'apprentissage humain. *Ières Rencontres Jeunes Chercheurs en EIAH, RJC-EIAH'2006*, pages 131 à 139.

Stefanov K. & Stefanova E. (2005). Analysis of the Usage of the Virtuoso System. In Workshop on Usage Analysis in Learning systems, *The 12<sup>th</sup> International Conference on Artificial Intelligence in Education AIED*, Amsterdam, 2005.

## **Table des annexes**

ANNEXE 1 DONNEES ENVOYEEES PAR LE FORMULAIRE DE RECHERCHE .....	63
ANNEXE 2 TABLE DES CATEGORIES GRAMMATICALES ET DES TRAITES ASSOCIES .....	66
ANNEXE 3 TABLE DES DOMAINES D'ERREUR ET CATEGORIE D'ERREUR ASSOCIEES .....	67
ANNEXE 4 NOUVEAU MODELE DE BASE DE DONNEES .....	68

## Annexe 1

### Données envoyées par le formulaire de recherche

Voir	Données	Libellé	Type	Valeur	Facultatif
A	\$_POST['gp-ling']	Provenance	Chaîne	tout, english, dutch, mixed	Oui
B	\$_POST['den-err']	Densité d'erreurs	Chaîne	tout, faible, moyenne, moyenne+, elevee	Oui
C	\$_POST['t-textes']	Longueur des textes (nombre de mots)	Chaîne	tout, court, moyenne, moyenne+, longs	Oui
D	\$_POST['cgforme']	Contexte gauche Forme	Chaîne	Définie par l'utilisateur	Oui
E	\$_POST['cglemme']	Contexte gauche Lemme	Chaîne	Définie par l'utilisateur	Oui
F	\$_POST['cgcat']	Contexte gauche Catégorie	Chaîne	tout, abv, adj, adv, con, det, int, nom, npr, num, pct, pre, pro, sym, ver	Oui
G	\$_POST['cgtrait']	Contexte gauche Trait	Chaîne	Cf Table 2	Oui
H	\$_POST['cgint']	Contexte gauche Intervalle	Entier	Définie par l'utilisateur	Oui
I	\$_POST['erounon']	Recherche dans une erreur	Chaîne	tout, oui, non	Oui
J	\$_POST['dom-err']	Domaines d'erreur	Chaîne	tout, F, M, G, L, X, R, Q, Y, Z	1, 2
K	\$_POST['cat-err']	Catégories d'erreur	Chaîne	Cf Table 3	Oui
L	\$_POST['forme']	Forme	Chaîne	Définie par l'utilisateur	3, 4
M	\$_POST['lemme']	Lemme	Chaîne	Définie par l'utilisateur	5, 6
N	\$_POST['cate']	Catégorie	Chaîne	tout, abv, adj, adv, con, det, int, nom, npr, num, pct, pre, pro, sym, ver	7, 8
O	\$_POST['trait']	Trait	Chaîne	Cf Table 2	Oui
P	\$_POST['cforme']	Correction Forme	Chaîne	Définie par l'utilisateur	9, 10
Q	\$_POST['clemme']	Correction Lemme	Chaîne	Définie par l'utilisateur	11, 12
R	\$_POST['ccat']	Correction Catégorie	Chaîne	tout, abv, adj, adv, con, det, int, nom, npr, num, pct, pre, pro, sym, ver	13, 14
S	\$_POST['ctrait']	Correction Trait	Chaîne	Cf Table 2	Oui
T	\$_POST['cdforme']	Contexte droit Forme	Chaîne	Définie par l'utilisateur	Oui
U	\$_POST['cdlemme']	Contexte droit Lemme	Chaîne	Définie par l'utilisateur	Oui
V	\$_POST['cdcat']	Contexte droit Catégorie	Chaîne	tout, abv, adj, adv, con, det, int, nom, npr, num, pct, pre, pro, sym, ver	Oui
W	\$_POST['cdtrait']	Contexte droit Trait	Chaîne	Cf Table 2	Oui
X	\$_POST['cdint']	Contexte droit Intervalle	Entier	Définie par l'utilisateur	Oui



- A : Choix de la provenance du corpus interrogé.
- B : Choix de la densité d'erreurs dans les textes du corpus interrogé.
- C : Choix de la longueur des textes du corpus interrogé.
- D : Définition d'une forme présente dans le contexte gauche.
- E : Définition d'un lemme présent dans le contexte gauche.
- F : Définition de la catégorie grammaticale présente dans le contexte gauche.
- G : Définition du trait grammatical présent dans le contexte gauche.
- H : Définition de l'intervalle entre le contexte gauche spécifié et l'expression recherchée, en nombre de mots.
- I : Choisir de rechercher ou non un texte erroné.
- J : Choix du domaine d'erreur à rechercher.
- K : Choix de la catégorie d'erreur à rechercher.
- L : Choix de la forme à rechercher.
- M : Choix du lemme à rechercher.
- N : Choix de la catégorie grammaticale à rechercher.
- O : Choix du trait grammatical à rechercher.
- P : Choix de la forme de la correction.
- Q : Choix du lemme de la correction.
- R : Choix de la catégorie grammaticale de la correction.
- S : Choix du trait grammatical de la correction.
- T : Définition d'une forme présente dans le contexte droit.
- U : Définition d'un lemme présent dans le contexte droit.
- V : Définition de la catégorie grammaticale présente dans le contexte droit.
- W : Définition du trait grammatical présent dans le contexte droit.
- X : Définition de l'intervalle entre le contexte droit spécifié et l'expression recherchée, en nombre de mots.



## Annexe 2

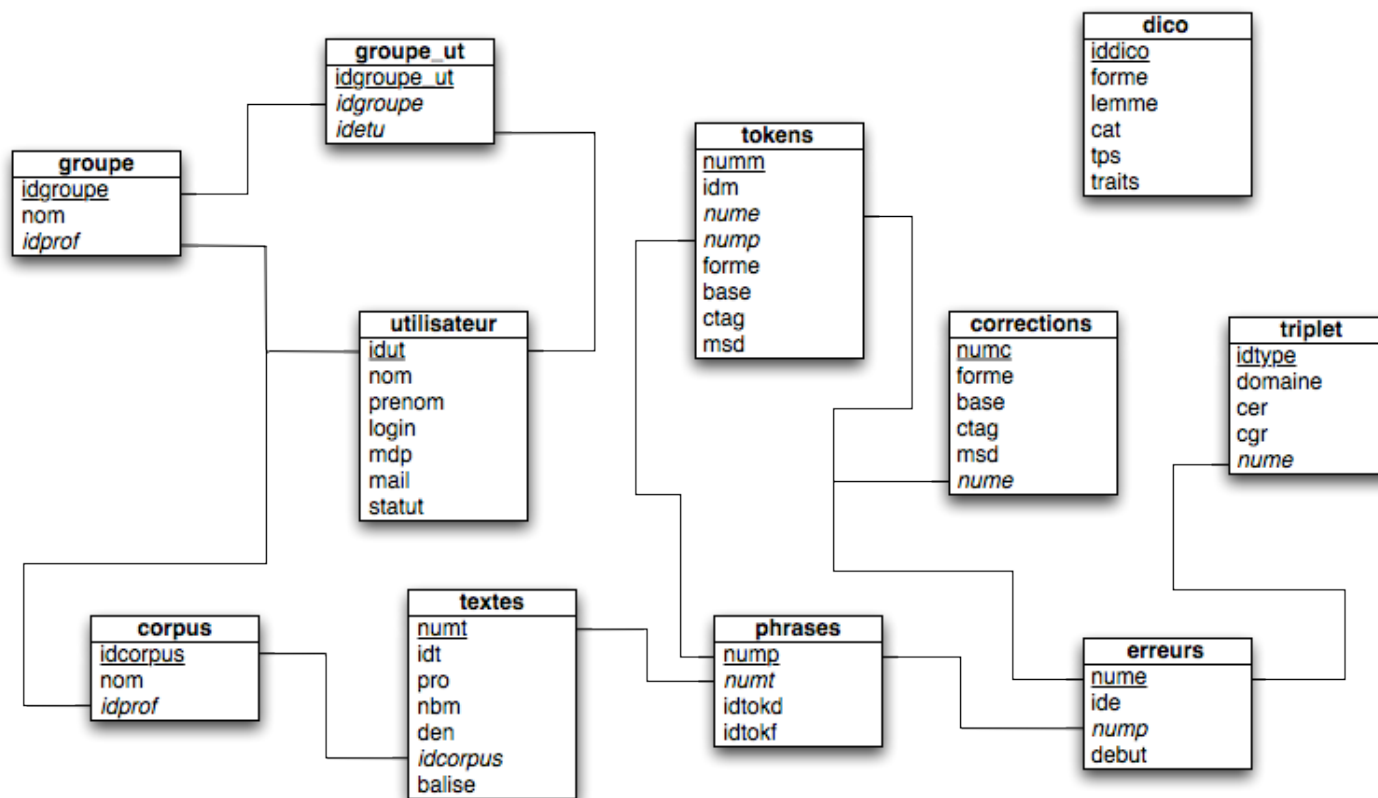
### Table des catégories grammaticales et des traits associés

Catégorie	Traits associés
Abv (Abréviation)	
Adj (Adjectif)	Tout, singu, pluri, invnb, femin, mascu, invge, defin, indef, propr, inter, demon, cardi, ordin
Adv (Adverbe)	
Con (Conjonction)	Tout, coord, subor
Det (Déterminant)	Tout, singu, pluri, femin, mascu, contr, defin, indef, artic
Int (Interjection)	
Nom (Nom)	Tout, singu, pluri, femin, mascu, invnb, invge
Npr (Nom propre)	
Num (Nombre)	
Pct (Ponctuation)	
Pre (Préposition)	
Pro (Pronom)	Tout, singu, pluri, femin, mascu, defin, indef, propr, posse, refle, relat, inter, demon
Sym (Symbole)	
Ver (Verbe)	Tout, infin, ipres, iimpa, ifutu, ipsim, ppass, ppres, cpres, spres, simpa, imper, pers1, pers2, pers3, singu, pluri

**Annexe 3**  
**Table des domaines d'erreur et catégorie d'erreur associées**

Domaine d'erreur	Catégories d'erreur associées
F (Forme-graphique)	AGL (agglutination), MAJ (majuscule), DIA (diacritique), HOM (homonymie), GRA (graphe)
M (Morphologie)	MDP (dérivation-préfixation), MDS (dérivation-suffixation), MFL (flexion), MFC (flexion-confusion), MCO (composition)
G (Grammaire)	CLA (classe), AUX (auxiliaire), GEN (genre), MOD (mode), NBR (nombre), PER (personne), TPS (temps), VOI (voix), EUF (euphonie)
L (Lexique)	SIG (signification), CPA (complément de l'adjectif), CPD (complément de l'adverbe), CPV (complément du verbe), CPN (complément du nom), FIG (figement)
X (Syntaxe)	MAN (manque), COH (cohésion), ORD (ordre), RED (redondance)
R (Registre)	RLE (lexique), RSY (syntaxe)
Q (Ponctuation)	CON (confusion), TRO (redondance), OUB (oubli)
Y (Style)	CLR (clarté), LOU (lourdeur)
Z (Faute de frappe)	

## Annexe 4 Nouveau modèle de base de données



## Table des figures

FIGURE 1 - EXEMPLE D'ANNOTATION DES ERREURS DANS LE CORPUS FRIDA .....	9
FIGURE 2 - EXEMPLE D'ANNOTATION DES TEXTES DANS LE CORPUS FRIDA-BIS .....	10
FIGURE 3 - SCHEMA DE LA BASE DE DONNEES IDILL AVANT LE STAGE.....	11
FIGURE 4 - CREATION DE LA SOUS-REQUETE CORRESPONDANT AUX CRITERES SUR LES TEXTES.....	12
FIGURE 6 - INTERFACE DE RECHERCHE D'EXCELANT AVANT LE STAGE.....	13
FIGURE 5 - SOUS-REQUETE CORRESPONDANT AUX CRITERES SUR LES TEXTES .....	13
FIGURE 7 - DEFINITION DES CONTRAINTES SUR LES TEXTES .....	14
FIGURE 8 - DEFINITION DES CRITERES SUR LA CIBLE .....	15
FIGURE 9 - LISTES DEROUANTES DYNAMIQUES.....	15
FIGURE 10 - AFFICHAGE DES RESULTATS PAR PHRASE .....	16
FIGURE 11 - COURVOL D'UN MOT JUSTE .....	16
FIGURE 12 - COURVOL D'UN MOT ERRONE.....	16
FIGURE 13 - MENURE SURGISSANTE CONTENANT LE TEXTE SELECTIONNE .....	16
FIGURE 14 - MODIFICATION D'UNE REQUETE .....	17
FIGURE 15 - AFFINAGE D'UNE REQUETE.....	17
FIGURE 16 - NOUVELLE INTERFACE DE RECHERCHE D'EXCELANT.....	22
FIGURE 17 - NOUVEL AFFICHAGE DES RESULTATS.....	22
FIGURE 18 - AFFICHAGE DE LA SELECTION DE PHRASES DANS L'APPLICATION .....	24
FIGURE 19 - TELECHARGEMENT DE LA SELECTION DE PHRASE .....	24
FIGURE 20 - TELECHARGEMENT D'UN TEXTE .....	25
FIGURE 21 - INDEX DE L'APPLICATION AVEC FORMULAIRE DE CONNEXION .....	25
FIGURE 22 - PAGE D'ACCUEIL ADMINISTRATEUR .....	27
FIGURE 23 - CONFIRMATION DE CREATION D'UN NOUVEAU COMPTE ENSEIGNANT .....	28
FIGURE 24 - AFFICHAGE DES INFORMATIONS D'UN ENSEIGNANT .....	29
FIGURE 25 - MODIFICATION DES INFORMATIONS D'UN ENSEIGNANT .....	29
FIGURE 26 - PAGE D'ACCUEIL ENSEIGNANT .....	30
FIGURE 27 - CREATION D'UN GROUPE CLASSE .....	31
FIGURE 28 - AFFICHAGE DES INFORMATIONS SUR UN GROUPE .....	32
FIGURE 29 - INTERFACE DE MODIFICATION DU MOT DE PASSE.....	33
FIGURE 30 - INTERFACE D'AJOUT DE TEXTE .....	34
FIGURE 31 - EXEMPLE DE SORTIE TREE TAGGER.....	35
FIGURE 32 - LISTE DES CORPUS D'UN ENSEIGNANT .....	37
FIGURE 33 - INTERFACE DE BALISAGE DES ERREURS .....	38
FIGURE 34 - CRITERES CONCERNANT LE CORPUS A INTERROGER .....	39
FIGURE 35 - INTERFACE DE CONSULTATION DE TRACES DE LA PLATE-FORME MOODLE.....	46
FIGURE 36 - AFFICHAGE DES TRACES D'INTERACTION DANS LA PLATE-FORME MOODLE .....	46
FIGURE 37 - REPRESENTATION GRAPHIQUE DU SCENARIO PEDAGOGIQUE .....	48
FIGURE 38 - HISTOGRAMME REPRESENTANT LE POURCENTAGE DE REUSSITE TOTAL A CHAQUE ACTIVITE .....	49
FIGURE 39 - HISTOGRAMME REPRESENTANT LE POURCENTAGE DE REUSSITE TOTAL A L'ACTIVITE 3 .....	49
FIGURE 40 - HISTOGRAMME REPRESENTANT LE POURCENTAGE DE REUSSITE DE CHAQUE APPRENANT A CHAQUE ACTIVITE.....	50
FIGURE 41 - HISTOGRAMME REPRESENTANT LE POURCENTAGE DE REUSSITE DE L'APPRENANT 5 A CHAQUE ACTIVITE.....	51
FIGURE 42 - REPRESENTATION GRAPHIQUE DU PARCOURS A SUIVRE.....	51
FIGURE 43 - REPRESENTATION GRAPHIQUE DU PARCOURS DE L'APPRENANT 5 .....	52
FIGURE 44 - DIAGRAMME CIRCULAIRE REPRESENTANT L'AVANCEMENT. REALISE A L'AIDE DE CHARTIFY.....	53
FIGURE 45 - REPRESENTATION GRAPHIQUE DU PARCOURS SUIVI A DESTINATION D'UN APPRENANT .....	54
FIGURE 46 - COURBE DE PROGRESSION D'UN APPRENANT EN FONCTION DE SES RESULTATS AUX ACTIVITES ...	54
FIGURE 47 - HISTOGRAMME REPRESENTANT LE POURCENTAGE DE REUSSITE DE L'APPRENANT 5 A CHAQUE ACTIVITE.....	54

**MOTS-CLÉS** : ALAO, ingénierie pédagogique, FLE, corpus d'apprenants, traces, base de données, concordancier

## **RÉSUMÉ**

EXXELANT (EXample eXtractor Engine for LANguage Teaching) est un outil développé dans le cadre du projet Idill du réseau d'excellence européen Kaleidoscope. Cette application permet d'effectuer des recherches sur un corpus de textes d'apprenants en Français Langue Étrangère et d'en tirer des extraits correspondant aux critères spécifiés. Ce mémoire porte sur les améliorations apportées à ce prototype afin de développer son utilisabilité : système de connexion, export d'une sélection de phrases ou d'un texte, enrichissement du corpus en permettant la création de nouveaux corpus et l'ajout de textes à la base de données par les enseignants et interface de balisage manuel des erreurs. La dernière partie porte sur la nécessité des traces d'interaction dans un outil pédagogique multimédia. Elle propose une présentation des types de traces, de leur utilité ainsi que de la façon de les représenter.

**KEYWORDS** : CALL, instructional design, French as a foreign language, learner corpora, tracks, database, concordancer

## **ABSTRACT**

EXXELANT (EXample eXtractor Engine for LANguage Teaching) is a tool developed within the framework of the European network of excellence Kaleidoscope's Idill project. This application allows to browse a French-language-learner corpus and extract examples corresponding to the specified criteria. This dissertation is about the improvements carried out to the prototype aiming to increase its usability: login system, selected sentences or text export, corpus enrichment by allowing teachers to create new corpora and add texts to the database and interface for manual error markup. The last part deals with the necessity of usage tracking in a multimedia educational tool. It offers a presentation of track types and their usefulness along with the way to represent them.

**DECLARATION**

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

NOM : ... CHARLES ..... PRENOM : ... Guilie .....

DATE : ... 29/03/2011 .....