



**HAL**  
open science

## Le “ contexte pédagogique ” comme pré-requis à la notion de facette de texte

Mathieu Loiseau

► **To cite this version:**

Mathieu Loiseau. Le “ contexte pédagogique ” comme pré-requis à la notion de facette de texte. 1st International Conference on Linguistic and Intercultural Education (CLIE 2008), 2008, Alba Iulia, Roumanie. pp.27 - 29. hal-01586328

**HAL Id: hal-01586328**

**<https://hal.science/hal-01586328>**

Submitted on 20 Jun 2018

**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.

# LE « CONTEXTE PÉDAGOGIQUE » COMME PRÉ-REQUIS À LA NOTION DE FACETTE DE TEXTE

**Mathieu Loiseau**

Laboratoire Lidilem - Université Stendhal Grenoble 3 (France)

**Résumé :** Notre travail s'inscrit dans la perspective de création d'une base de textes indexés pédagogiquement pour l'enseignement des langues. Au cours de notre travail nous avons été amenés à consulter les nombreux travaux sur la description de ressources pédagogiques (Learning Object Metadata (LOM), Shareable Content Object Reference Model, Educational Modelling Language...). Cependant, aucune solution proposée ne semble viable pour la description de ressources brutes, a fortiori pour la description de textes en vue de leur utilisation dans l'enseignement des langues. En effet, les standards, quand ils s'intéressent à l'indexation de ressources (comme LOM, the Gateway to Educational Metadata ou Education Network of Australia Metadata) considèrent les propriétés dites pédagogiques comme intrinsèques aux objets. Or, nous avons pu montrer, à travers certains exemples et surtout par une consultation de plus d'une centaine d'enseignants, que les propriétés pédagogiques des textes dépendent en réalité étroitement du contexte dans lequel le texte va être utilisé : le contexte pédagogique. Dans notre cas, il regroupe toutes les informations relatives à l'exploitation du texte : les objectifs didactiques, bien sûr, mais aussi l'ensemble des caractéristiques du public (niveau, âge, centres d'intérêt, langue 1, etc.) et de l'institution (filière / diplôme, contraintes matérielles, etc.)

Pour prendre en compte cette dynamique des propriétés pédagogiques des textes, nous définissons la notion de facette d'un texte de la manière suivante :

« Une facette de texte est une propriété définie en vue de l'exploitation pédagogique de ces textes, assortie au minimum d'un mécanisme, informatisable ou non, permettant de calculer une valeur de cette propriété pour tout texte en fonction d'un contexte pédagogique donné. »

**Mots-clés :** ALAO, TAL, Objet pédagogique, Indexation pédagogique

## 1. Corpus pour l'enseignement des langues et indexation pédagogique

Avec l'avènement de l'approche communicative, la plus utilisée par les enseignants de langues (cf. (Levy, 1997) et <sup>5</sup>), le texte authentique est au centre des préoccupations des enseignants.

Alors que les corpus sont nombreux, aucun n'est dédié à la recherche de textes pour l'enseignement des langues. Les mécanismes de recherches reposent sur de traditionnelles recherches par mots-clés. Il est à noter que les enseignants s'adaptent à des outils informatiques développés à d'autres fins, comme le montre le Data Driven Learning de Tim Johns<sup>1</sup>. Malgré tout, certains des travers de l'Apprentissage des langues assisté par ordinateur (ALAO) mentionnés dans (Antoniadis *et al.*, 2004a) résument bien la situation des corpus pour l'enseignement des langues : quand un enseignant cherche un texte dans un corpus, il n'existe pas de système lui permettant de formuler sa requête en termes relevant de sa problématique, utilisant des concepts pédagogiques.

---

<sup>1</sup> <http://www.eisu2.bham.ac.uk/johnstf/timconc.htm>

## Indexation pédagogique pour l'enseignement des langues

Ce sont ces considérations et la perspective d'intégration à la plateforme Multi-apprentissages Interactifs par des Recherches sur des Textes et l'Oral (MIRTO)<sup>2</sup> (Antoniadis *et al.*, 2007), qui nous ont mené à un projet de création d'une base de textes destinée aux enseignants de langues. Ce projet doit donner lieu à l'implémentation d'un prototype (actuellement en cours de développement).

Une telle base de texte doit permettre les cas d'utilisations basiques suivants :

- recherche de textes;
- ajout de textes.

Comme en témoigne la définition de « langage documentaire » de Lefèvre (2000), les utilisateurs, ici les enseignants de langues, sont au centre du processus d'indexation de la base :

« Langage artificiel qui fournit une représentation formalisée et univoque des documents d'un corpus *et des questions qui intéressent un groupe d'utilisateurs*, afin de permettre le repérage simple des documents du corpus *qui répondent aux questions de ces utilisateurs*. »

L'indexation pédagogique sera donc une :

Indexation effectuée selon un système décrivant des objets en fonction de critères pédagogiques (relevant de la problématique de la didactique).

Le travail présenté ici relève donc de l'indexation pédagogique de textes pour l'enseignement des langues. Les objets considérés sont des textes, dont on veut que les usagers (des enseignants de langues) puissent les retrouver en fonction de questions relevant de leur problématique, i.e. celle de la didactique des langues.

Nous commencerons par détailler les besoins des utilisateurs, avant de les confronter aux standards de description de ressources pédagogiques existants pour pouvoir introduire la notion de facette d'un texte et présenter un modèle dans lequel elle s'inscrit.

## **2. Évaluation des besoins des utilisateurs**

Pour définir des critères pertinents pour l'indexation pédagogique de textes pour l'enseignement des langues, plutôt que de privilégier une méthode d'enseignement par rapport à une autre, nous nous sommes orientés vers une démarche empirique visant à obtenir un aperçu des pratiques des enseignants. L'étude s'est articulée en trois parties :

- une étude préliminaire qualitative fondée sur huit entretiens avec des enseignants de langue (Loiseau, 2003);
- un questionnaire court visant à avoir un aperçu du rapport des enseignants au texte authentique, au classement et à la recherche de textes;
- un questionnaire long visant à préciser les informations recueillies pour le premier questionnaire et à isoler des critères de recherche de textes.

---

<sup>2</sup> Plateforme utilisant des techniques du TAL pour la génération automatique d'activités à partir de textes tout venants.

## Questionnaires

L'étude préliminaire a soulevé un certain nombre de questions auxquelles nous avons essayé de répondre via le premier questionnaire.

Ce dernier existait aussi bien sous forme papier que sous forme électronique. Il a été rempli par 133 enseignants, principalement sous sa forme électronique, notamment suite à une note sur [cafepedagogique.net](http://www.cafepedagogique.net)<sup>3</sup>. Cela signifie que si nous espérons initialement toucher aussi bien des enseignants « technophobes » que des enseignants maîtrisant l'outil informatique, les réponses aux questionnaires concernent plus particulièrement cette dernière catégorie et ne permet pas de tirer de conclusions quant à l'influence de la *littératie informatique*<sup>4</sup> sur les pratiques des enseignants.

Les objectifs du premier questionnaire étaient simples : en premier lieu la validation d'une hypothèse qui semble évidente - mais lourde de conséquences - formulée suite à notre étude préliminaire, selon laquelle à un texte donné peuvent correspondre plusieurs utilisations. Les autres questions visaient à fournir des pistes de réflexion sur les pratiques des enseignants, sur leur utilisation des textes authentiques, la manière dont ils recherchent ou trouvent les documents qu'ils utilisent en classe et la manière dont ils classent ces textes.

112 des 115 enseignants ayant répondu à la question<sup>5</sup> déclarent « pouvoir utiliser un même texte dans plusieurs contextes différents ». Ceci n'est pas qu'un souhait, 106 des 112 déclarent l'avoir déjà fait. Ils confirment ainsi l'hypothèse que nous avons formulée concernant l'importance du contexte d'utilisation de la ressource, reste à définir son rôle dans le choix de ressources pédagogiques.

Outre la confirmation de cette hypothèse, nous avons pu tirer les conclusions suivantes de ce premier questionnaire :

- les enseignants privilégient l'utilisation de textes authentiques;
- ils ont recours aux textes construits pour en maîtriser le contenu linguistique (structures grammaticales, vocabulaire), en particulier pour les groupes de débutants;
- le type d'activité et le public semblent être des critères de recherche;
- nous n'avons pas été en mesure de tirer de conclusions sur les recueils des enseignants.

Ces résultats bien que répondant à certaines de nos questions ne nous permettaient pas d'isoler ces critères de recherche ou de classement. Nous avons donc élaboré un second questionnaire, beaucoup plus long que le précédent, afin de tenter d'isoler des critères.

Ce second questionnaire était beaucoup plus précis que le premier, mais aussi beaucoup plus fastidieux à remplir : jusqu'à 77 questions, dont 36 questions ouvertes<sup>6</sup> et certaines questions répétées plusieurs fois mais à considérer en fonction d'un contexte différent. Ce questionnaire permettait donc aux enseignants de revenir en

---

<sup>3</sup> <http://www.cafepedagogique.net>

<sup>4</sup> Traduction du terme anglais de *computer literacy* retrouvée notamment chez (Mangenot, 2003) et (Reffay & Chanier, 2005)

<sup>5</sup> L'intégralité des réponses à ce questionnaire se trouvent sur la page <http://mathieu.loiseau.free.fr/bdtp/q1/stats.php>

<sup>6</sup> <http://mathieu.loiseau.free.fr/bdtp/login.php>

arrière et de ne pas répondre à toutes les questions. Chaque question a reçu, en moyenne, 50 réponses.

### Résultats

Nous avons pu, à travers la description des recueils de textes des enseignants, isoler un certain nombre de critères de recherche dont les plus fréquemment utilisés sont : thème, contenu<sup>7</sup> ou objectif<sup>8</sup> linguistique et niveau.

Le reste du questionnaire était dédié, outre à délimiter les attentes des enseignants, à confronter aux pratiques déclarées des enseignants l'hypothèse selon laquelle certains critères avaient une influence les uns sur les autres. Nous avons entre autre pu montrer que le type d'activité avait un effet sur :

- la longueur du texte; [F(5,143)=3,362; p<,01]
- le nombre d'éléments représentatifs de la notion sur laquelle porte l'activité (qu'elle soit grammaticale ou lexicale); [F(4,127)=4,739; p<,005]
- la quantité de structures inconnues (autres que l'objet de la leçon); [ $\chi^2=32,177$ ; dl=10; p<,001]
- la quantité de vocabulaire inconnu (autre que l'objet de la leçon). [ $\chi^2=28,949$ ; dl=10; p<,005]

Les enseignants déclarent que le niveau des apprenants influence la quantité de structures inconnues tolérée dans un texte, tout comme leur L1 quand celle-ci est prise en compte lors du choix du texte.

### Conclusions

Un texte est utilisable dans plusieurs contextes différents. Par exemple, la quantité de structures et de vocabulaire inconnus varie en fonction du type d'activité. Ces « inconnaitances » relèvent du niveau des apprenants. Nous pouvons donc dire qu'en fonction du type d'activité, un texte peut correspondre à divers publics différents. Cet exemple illustre à notre avis le fait que certaines caractéristiques « pédagogiques » d'un texte ne sont pas des caractéristiques figées. Elles dépendent de la combinaison de caractéristiques intrinsèques du texte, qui relèvent de l'objet lui-même (comme son contenu) et de ce que Recker et Wiley (2001) appellent les données extrinsèques, « qui décrivent le contexte dans lequel la ressource est utilisée<sup>9</sup> ».

Notre prototype ne couvrira pas tous les besoins des enseignants et ceux qui seront couverts ne le seront que partiellement. Notre système se doit d'être évolutif et de permettre l'ajout de traits supplémentaires.

Enfin, le système ne pourra pas, pour chaque requête, ne fournir qu'un texte qui soit le plus conforme à cette requête. Cela ne posera pas de problème de retrouver un texte à partir de son titre et de son auteur. Par contre, pour d'autres types de requête certains critères seront trop difficile à formaliser.

---

<sup>7</sup> du texte

<sup>8</sup> de l'activité à laquelle le texte est destiné

<sup>9</sup> *Intrinsic properties are those that are derivable by simply having the resource at hand [...], whereas extrinsic properties describe the context in which the resource is used.* (traduction de l'auteur)

Par exemple le « thème » du texte : il n'existe pas d'outil TAL qui permette d'extraire les « thèmes » du texte, une annotation manuelle poserait des problèmes de cohérence de l'annotation et d'exhaustivité. Même si l'annotation des thèmes d'un texte était cohérente et exhaustive, il faudrait pouvoir l'associer aux requêtes des enseignants. Ce genre de caractéristiques demandera donc un certain nombre d'approximations que ce soit pour l'indexation ou la formulation / interprétation de la requête. D'autres caractéristiques, comme « l'intérêt que le texte pourra susciter chez les apprenants » sont, à notre connaissance, extrêmement difficiles à modéliser à l'heure actuelle. Elles interviennent malgré tout dans le choix des enseignants. Le système ne saurait se substituer à ce processus de choix, il peut cependant le faciliter, notamment en proposant une aide à la décision : il peut fournir un sous-ensemble de textes candidats mettant en exergue certains des éléments (caractéristiques intrinsèques) utilisés dans le choix du texte.

### **3. Inadéquation des standards de description de ressources pédagogiques avec les besoins des utilisateurs**

Les standards de description de ressources pédagogiques n'apportent pas de réponse aux besoins des enseignants cités plus haut.

#### Standards étudiés

Dans le cadre de cette étude de l'existants, nous nous sommes intéressés à plusieurs standards. Nous avons étudié d'une part le Learning Object Metadata (LOM)<sup>10</sup> et certains de ses profils d'application ou d'autres projets connexes : CanCore<sup>11</sup> (profil d'application), Sharable Content Object Reference Model (SCORM), Instructional Management Systems (IMS) metadata et ARIADNE qui ont participé à l'élaboration de LOM ou le réutilisent. D'autre part, nous avons consulté des standards proches de la Dublin Core Metadata Initiative (DCMI) : le Gateway to Educational Metadata (GEM) et educational network of Australia (edna). Il est à noter que le LOM et la DCMI ont un accord d'interopérabilité (DCMI, --).

Comme cette analyse des standards n'est pas l'objectif de notre article, nous nous contenterons ici de détailler nos remarques concernant le LOM, mais elles sont aussi représentatives de la situation pour la plupart des standards.

#### LOM

Les spécifications du LOM (IEEE LTSC WG12, 2002) proposent plus de soixante-dix descripteurs, pour chacun desquels sont donnés une définition, une cardinalité, un type de valeurs parfois assorti d'un vocabulaire qui, malgré les exemples fournis, peut-être flou et ne prendre véritablement un sens qu'au sein d'une communauté d'utilisateurs. Nous pouvons citer comme exemple le descripteur 5.3, *Interactivity Level*<sup>12</sup> qui peut prendre une valeur parmi *very low, low, medium, high, very high*.

---

<sup>10</sup> <http://ltsc.ieee.org/wg12/>

<sup>11</sup> <http://www.cancore.ca/fr/>

<sup>12</sup> Catégorie 5 (Educational)

Chacun des éléments est non obligatoire et peut être répété afin de décrire un « objet pédagogique » :

« N'importe quelle entité, numérisée ou non, qui pourrait être utilisée pour l'apprentissage, l'enseignement ou la formation.<sup>13</sup> » (IEEE LTSC WG12, 2002)

Le descripteur *1.8, Aggregation Level*<sup>14</sup>, illustre cette définition en instituant quatre niveaux d'agrégation allant des ressources brutes au cursus complet, qui doivent donc être décrits avec les mêmes éléments. Pernin (2004) fait justement remarquer que

« la volonté d'intégrer au sein d'un même modèle des entités de niveau conceptuellement très différent : les ressources nécessaires à la mise en place d'activités pédagogiques et les activités elles-mêmes »

est l'une des causes des ambiguïtés et imprécisions du modèle.

### Des propriétés pédagogiques intrinsèques aux objets

Tous ces standards de métadonnées reposent sur un même système, à savoir le recensement des propriétés d'un objet. L'essence même de ces métadonnées induit une fixité des propriétés, notamment des propriétés pédagogiques qui, si elle peut convenir pour des objets déjà exploités pédagogiquement, nous semble inadaptée pour une indexation pédagogique de ressources brutes, *a fortiori* de textes pour l'enseignement des langues.

Les descripteurs de la catégorie 5 (*Educational*) du LOM désignent certaines des propriétés « pédagogiques » des textes.

Prenons quelques éléments : *5.10, Description*, qui doit contenir des « commentaires sur la manière dont l'objet doit être utilisé » ; *5.7, Typical Age Range*, l'âge de l'utilisateur final de l'objet ; et *5.8, Difficulty*. Nous avons montré qu'un texte pouvait avoir plusieurs utilisations. Chaque utilisation différente *5.10, Description* affectera non seulement la difficulté de l'objet mais aussi le public auquel il est destiné. Le LOM permet plusieurs groupes 5 (*Educational*). Dans notre cas, une description exhaustive de la ressource exigerait que toutes les utilisations possibles de chaque texte soient indiquées, chacune sous la forme d'un n-uplet *Educational*.

Dans ces conditions, la description de chaque texte devient trop fastidieuse pour imaginer les indexer pédagogiquement (pour l'enseignement des langues) par ce biais, que ce soit via le LOM ou les autres standards de description de ressources pédagogiques.

## **4. Contexte pédagogique et facette**

D'après la définition que nous en avons donnée, il nous semble très difficile d'arriver à une indexation pédagogique de textes pour l'enseignement des langues en considérant toutes les propriétés pédagogiques comme intrinsèques à l'objet. En effet, les « questions » des utilisateurs relèvent de la recherche de texte pour une utilisation donnée. Or, comme nous l'avons dit, l'utilisation d'un texte en est une propriété

---

<sup>13</sup> *any entity -digital or non-digital- that may be used for learning, education or training* (Traduction de l'auteur)

<sup>14</sup> Catégorie 1 (General)

extrinsèque : un texte peut être utilisé dans plusieurs contextes radicalement différents. La prise en charge de ces propriétés extrinsèques, potentiellement à valeurs multiples, à l'aide de métadonnées nous paraît très fastidieuse pour ne pas dire irréalisable. Il nous faut donc définir notre propre modèle.

Devant la difficulté à traiter toutes les propriétés des textes, nous avons regroupé les propriétés extrinsèques, les propriétés relevant de la contextualisation du texte dans l'enseignement des langues, sous le terme de « contexte pédagogique ». Il regroupe l'ensemble des caractéristiques du public (niveau, âge, centres d'intérêt, etc.), de l'institution (filière / diplôme, contraintes matérielles) ainsi que les « objectifs didactiques ». Comme nous allons le voir le contexte pédagogique n'est pas un et indivisible, mais la somme de propriétés et de mécanismes qui permettent de prendre en compte ses différents aspects. Notre travail consistant à mettre en relation les propriétés, aussi bien intrinsèques qu'extrinsèques pour chaque texte. Pour y parvenir nous proposons la notion de facette.

### Facettes de texte

La notion de facette doit nous permettre d'avoir recours à des mécanismes capables de combiner des caractéristiques intrinsèques et d'en calculer d'autres en fonction d'un contexte pédagogique donné.

Une *facette de texte* est une propriété définie en vue de l'exploitation pédagogique de ces textes, assortie au minimum d'un mécanisme, informatisable ou non, permettant de calculer une valeur de cette propriété pour tout texte en fonction d'un contexte pédagogique donné.

D'après cette définition, l'exploitation pédagogique ultérieure des textes considérés préside à la création des facettes. Toute caractéristique de texte pouvant être utile à un enseignant de langues pour la recherche ou le choix d'un texte pourra donc être une facette de ce texte. Pour calculer certaines facettes, nous aurons parfois recours à des propriétés intrinsèques des textes, qui en tant que telles n'ont pas nécessairement d'intérêt didactique. C'est donc au niveau de la facette que se crée la valeur ajoutée pédagogique.

Dans cette définition, nous différencions la facette, qui se situe à un niveau conceptuel, de sa valeur, que nous appellerons dorénavant *valeur* (du texte selon une facette pour un contexte pédagogique) pour éviter toute confusion.

Prenons l'exemple d'une facette que nous désignerons comme  $F_{e^tRep}$ . Cette facette représente le nombre d'éléments représentatifs d'une notion. Pour donner une valeur à cette facette nous avons besoin de la notion concernée, qui relève des objectifs pédagogiques de l'enseignant. En effet, une telle facette permettra à l'enseignant d'effectuer son choix uniquement parmi les textes qui contiennent en nombre suffisant la notion qu'il désire travailler avec ses apprenants.



« 19. ¿ Qué opinion tiene usted de la revolución bolivariana de la republica bolivariana de Venezuela ?

Me parece un proyecto político de gran interés que merece ser apoyado y más aún debido al grado de desinformación y manipulación que flota sobre él aquí en España.

34. Lo siento señora, pero como resumen de su entrevista, lo único que veo es que le molesta que sean otros y no los suyos, los que tienen poder. Su comentario sobre las elecciones la retrata...

Si hay que jugar, a lo que sea, habrá que jugar en igualdad de condiciones. Pero siempre hay quien está dispuesto a hacer lo que sea y a costa de quien sea para jugar con ventaja. »

<http://www.elmundo.es/encuentros/invitados/2004/10/1241/>

### Figure 1 - Exemple de texte

Pour le texte de la Figure 1, si l'enseignant s'intéresse au participe passé (éléments soulignés) la vue vaudra 3. Alors que la vue de Figure 1 selon la facette  $F_{e^{tRep}}$  pour le contexte pédagogique *haber\* que + Inf* (expression du devoir en espagnol) vaudra 2 (éléments encadrés). Ce qui nous mène directement à plusieurs remarques :

- la notion concernée relève de ce que nous avons appelé le contexte pédagogique ;
- la vue de la facette pour un texte donné dépend du contexte pédagogique ;
- un texte peut avoir plusieurs vues pour une même facette (une vue par contexte pédagogique).

Une facette n'est pas tenue de prendre en compte tout ce que nous avons appelé le contexte pédagogique, notamment car cela n'a pas forcément de sens. Dans le cas de  $F_{e^{tRep}}$  par exemple, le fait que les apprenants soient au collège ou en formation continue n'influence pas le nombre d'éléments représentatifs de la notion.

La facette va donc plus loin que les métadonnées par sa dynamicité, sa valeur n'est pas figée mais est calculée en fonction d'autres propriétés. Cependant, cela n'exclue en rien une facette dont la valeur est indépendante du contexte, que cette valeur soit calculée ou entrée au préalable par l'utilisateur. La facette que nous appellerons  $F_{nbmots}$  dénombrant le nombre de mots d'un texte, peut être calculée automatiquement<sup>15</sup> et aura la même valeur quel que soit le contexte pédagogique.  $F_{nbmots}$  peut être utilisée par l'enseignant pour rechercher un texte à utiliser parmi ceux qui tombent dans le cadre d'une contrainte de taille qu'il aura définie. De la même manière la valeur facette  $F_{auteur}$ , concernant l'auteur d'un texte sera entrée manuellement, pour chaque texte, par un utilisateur humain. Elle pourra ensuite être utilisée pour rechercher un texte parmi ceux d'un auteur donné.  $F_{auteur}$  a une valeur pédagogique évidente en cours de littérature, mais pas seulement. Elle peut aussi avoir une influence pour le choix d'un texte dans un autre contexte :

« Roald Dahl, [...] all his short stories are packed with these verbs [...] for emotion and gestures [...], that in french [require] a whole phrase [...]» (Témoignage issu de (Loiseau, 2003))

<sup>15</sup> En utilisant un tokenizer et un compteur

<sup>16</sup> Roald Dahl, toutes ses nouvelles sont remplies de ces verbes d'émotion et de gestes qui s'expriment en français avec des phrases entières. (Traduction de l'auteur)

## Séquence d'utilisation

Afin d'expliciter un peu plus notre modèle, nous allons utiliser cette notion de facette dans un exemple d'utilisation de notre prototype.

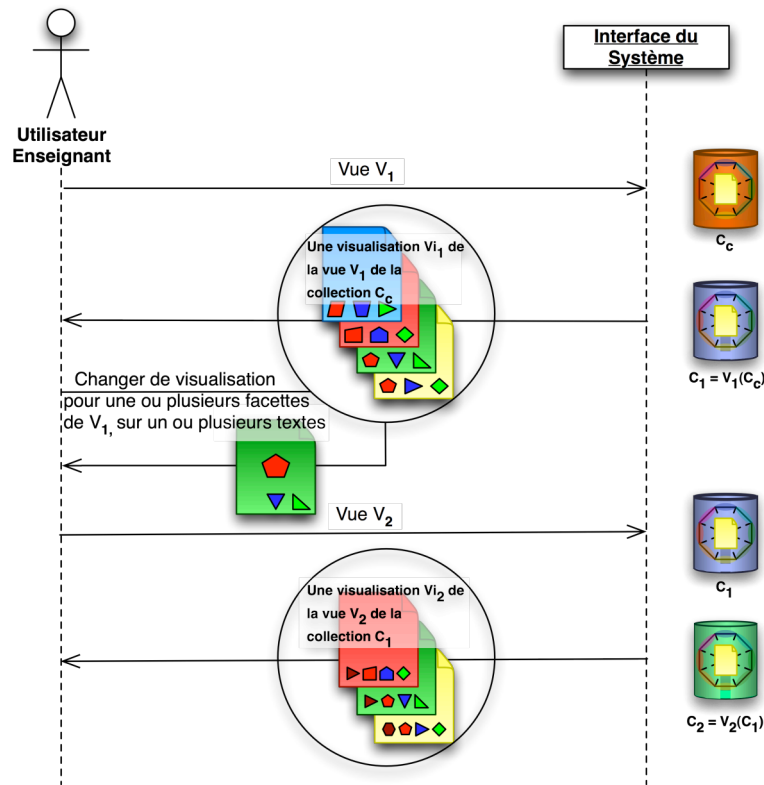


Figure 2 - Exemple d'interaction d'un utilisateur avec le système

Dans la Figure 2, nous plaçons un utilisateur-enseignant en situation de rechercher un texte dans notre système. Le système accède aux textes via la notion de facette. Imaginons que l'enseignant recherche un texte pour un exercice lacunaire d'approfondissement d'une notion grammaticale. D'après les réponses à notre questionnaire que nous décontextualisons complètement ici pour les besoins de l'exemple, un texte pour ce type d'exercice devra contenir entre 5 et 15 occurrences de la notion sur laquelle les étudiants vont travailler. L'enseignant qui voudrait par exemple travailler sur le prétérit en anglais demandera les textes de la collection complète dont la vue pour la facette  $F_{\text{etRep}}$  vaudra entre 5 et 15 pour le contexte pédagogique prétérit. Les textes vérifiant cette contrainte se retrouvent dans C<sub>1</sub> dans la figure. Ils sont ensuite présentés à l'utilisateur en mettant en exergue les propriétés intrinsèques utilisées pour calculer la facette, selon ce que nous appelons une visualisation. Pour l'aider à choisir parmi les textes l'utilisateur peut, selon la vue, se voir proposer plusieurs visualisations. Par exemple si l'utilisateur vient d'introduire le prétérit à ses apprenants, qu'ils n'ont pas encore été confrontés aux verbes irréguliers, il ne sera pas nécessaire de voir les verbes au prétérit en contexte pour exclure les

textes contenant des verbes irréguliers<sup>17</sup>. Dans la Figure 3, la visualisation « liste » ne permet pas de choisir un texte, mais permet d'en exclure.

L'enseignant peut malgré cette première recherche se voir proposer trop de textes par le système. Son exercice s'adressant à des apprenants encore débutants, pour un tel exercice, il ne veut pas que le texte soit trop long. Il peut donc affiner sa recherche, en recherchant parmi les textes de  $C_1$ , ceux dont la longueur est inférieure à 200 mots, utilisant ainsi  $F_{nbmots}$ . La collection de textes ainsi obtenue est  $C_2$ . L'utilisateur peut une fois de plus choisir entre les deux visualisations de la Figure 3, puisque le nombre de mots n'est associé à aucune visualisation autre que la valeur de la facette. Pour choisir le texte de son exercice, l'utilisateur aura recours à la visualisation « texte » qui lui permet de voir les éléments au prétérit en contexte.

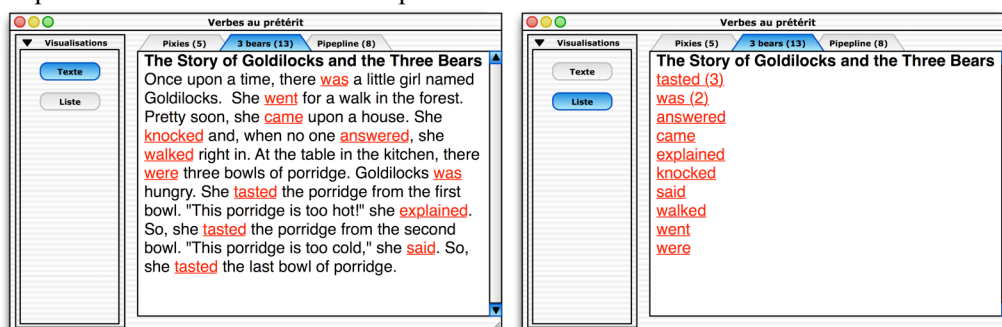


Figure 3 - Deux visualisations d'un même texte pour la vue de « The Story of Goldilocks and the Three Bears » selon la facette  $F_{e^{tRep}}$  pour le contexte pédagogique prétérit

### Un premier pas vers l'indexation pédagogique de textes

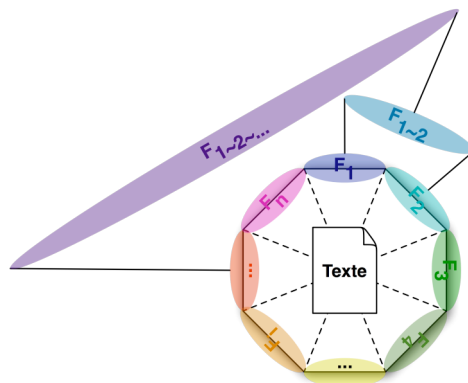
Les facettes que nous avons présentées restent d'un niveau de valeur ajoutée pédagogique relativement bas selon l'axe de la conceptualisation. Il ne s'agit pas d'associer une activité à un texte, mais seulement d'en compter certaines structures. Cependant,  $F_{e^{tRep}}$ , qui est implantable avec des outils disponibles<sup>18</sup> constitue une première étape vers une conceptualisation plus élevée. En effet, la notion de facette permet un mécanisme que nous n'avons pas évoqué jusqu'ici : la combinaison de facettes.

Prenons les deux facettes  $F_{nbmots}$  et  $F_{e^{tRep}}$ . Si nous nous référons aux résultats de notre second questionnaire, nous pouvons proposer une approximation de facette évaluant l'adéquation d'un texte à un contexte pédagogique. Ce dernier dépend de deux paramètres: le type d'activité à laquelle le texte est destiné et la notion concernée par cette activité ( $F_{activiT_{No}}$ ). Étant donné l'effet du type d'activité sur la longueur du texte et sur le nombre d'éléments représentatifs (cf. §2), il suffira d'associer à chaque type d'activité des valeurs pallier et des tolérances pour le nombre d'éléments représentatifs de la notion ( $F_{e^{tRep}}$ ) et pour la longueur du texte ( $F_{nbmots}$ ). Nos données concernent des pratiques déclarées et non des pratiques attestées. Elles manquent donc de précision pour que la facette ainsi constituée soit réellement performante, mais nous pensons

<sup>17</sup> D'après les résultats de notre second questionnaire, les enseignants tolèrent peu de structures inconnues dans les exercices structuraux.

<sup>18</sup> analyseur morphologique et compteur

qu'elles peuvent constituer une première approximation à confronter aux usages des enseignants (lors de leur utilisation du prototype par exemple) et à affiner ou revoir, le cas échéant. Cette facette pourrait ensuite être améliorée en y ajoutant une prise en compte du niveau des apprenants et ainsi de suite.



**Figure 4 - Combinaison des facettes  $F_1$  et  $F_2$  pour créer la facette  $F_{1\sim 2}$**

Le processus itératif que nous venons de décrire nous semble une piste pour la création de facettes de plus en plus élaborées et de plus en plus en adéquation avec les besoins des enseignants. Dans la Figure 4, nous avons pris soin de faire coexister ces facettes « combinées » avec celles qu'elles utilisent, afin de bien montrer que ce sont des facettes bien différentes. En effet, la facette composée ne couvre pas nécessairement tous les besoins de chacune des facettes qui la composent et ne doit pas les remplacer. Permettre aux utilisateurs de chercher un texte en fonction de l'activité dans laquelle ils vont l'utiliser ne doit pas leur ôter la possibilité de rechercher un texte en fonction de sa longueur.

Enfin ce processus doit s'appuyer sur les pratiques des enseignants et en collaboration étroite avec des didacticiens, afin de ne pas reproduire les erreurs constatées par (Antoniadis et al., 2004b) dans beaucoup de systèmes d'ALAO : la résolution d'un problème didactique comme s'il s'agissait d'un problème informatique. Comme le fait remarquer Morand (2007), « Il n'est pas souhaitable de laisser l'informatique aux mains des seuls informaticiens : ce n'est pas au shérif de faire la loi. »

## Références

- Antoniadis, G., Échinard, S., Kraif, O., Lebarbé, T., Loiseau, M., Ponton, C. (2004a). NLP-based scripting for CALL activities. *Proceedings of Coling '04 Workshop on eLearning for Computational Linguistics and Computational Linguistics for eLearning*. Genève.
- Antoniadis, G., Échinard, S., Kraif, O., Lebarbé, T., Loiseau, M., Ponton, C. (2004b). CALL: from current problems to NLP solutions. *Proceedings of EUROCALL*. Vienne.
- Antoniadis, G., Ponton, C., Zampa, V. (2007). De la nécessité du TAL dans les EIAH de langues : les cas EXXELANT et MIRTO. *Actes de la conférence EIAH 2007*. Lausanne.
- DCMI (Dublin Core Metadata Initiative) (----). *Memorandum of Understanding between the Dublin Core Metadata Initiative and the IEEE Learning Technology Standards Committee*, accédé le 26 août 2008 depuis : <http://dublincore.org/documents/dcmi-ieee-mou/>.

- IEEE LTSC WG12 (2002). *Final 1484.12.1 LOM Draft Standard Document*, accédé le 26 août 2008 depuis :  
[http://ltsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf).
- Lefèvre, P. (2000). *La recherche d'informations : du texte intégral au thesaurus*. Paris: Hermès.
- Levy, M. (1997). *Computer-Assisted Language Learning - context and conceptualization*. Oxford: Oxford University Press.
- Loiseau, M. (2003). *Vers la création d'une base de données de ressources textuelles indexée pédagogiquement pour l'enseignement des langues*. Mémoire de DEA, accédé le 26 août 2008 depuis :  
[http://mathieu.loiseau.free.fr/bdtp/fichiers/DEASCL\\_BDTIP-2003.pdf](http://mathieu.loiseau.free.fr/bdtp/fichiers/DEASCL_BDTIP-2003.pdf).
- Mangenot, F. (2003). Présentation d'Eurocall 2003 "New Literacies in Language Learning & Teaching". *ALSIC* 6(2), , accédé le 26 août 2008 depuis :  
[http://alsic.u-strasbg.fr/Num11/mangenot/alsic\\_n11-poi1.htm](http://alsic.u-strasbg.fr/Num11/mangenot/alsic_n11-poi1.htm).
- Morand, B. (2007). Le logiciel, sujet et objet de la norme. *Droit et société* 1(65), 41-51.
- Pernin, J-P. (2004). À propos des objets pédagogiques. *Entre technique et pédagogie : la création de contenus multimédia pour l'enseignement et la formation*. Neuchâtel, accédé le 26 août 2008 depuis :  
[http://www-clips.imag.fr/arcade/User/jean-philippe.pernin/recherche/download/Article\\_Pernin\\_Neuchatel07Nov03.pdf](http://www-clips.imag.fr/arcade/User/jean-philippe.pernin/recherche/download/Article_Pernin_Neuchatel07Nov03.pdf)
- Reffay, C. & Chanier, T. (2005) *Approche minimaliste de la formation d'enseignants du supérieur au rôle de concepteur-tuteur de cours en ligne*, accédé le 26 août 2008 depuis :  
<http://archive-edutice.ccsd.cnrs.fr/docs/00/03/94/49/PDF/ReffayChanierprepublie2005.pdf>
- Recker, M. & Wiley, D. (2001) A Non-authoritative Educational Metadata Ontology for Filtering and Recommending Learning Objects, *Interactive Learning Environments* 9(3), 255-271.