

## **LEXIQUE DES AFFECTS : CONSTITUTION DE RESSOURCES PEDAGOGIQUES NUMERIQUES**

AUGUSTYN Magdalena

LIDILEM (Laboratoire de Linguistique et Didactique des Langues Etrangères et Maternelles),  
Université Stendhal Grenoble 3 (Grenoble), France

magdalena.augustyn@bvra.e.u-grenoble3.fr

BEN HAMOU Sabrina

LIDILEM, Université Stendhal Grenoble 3 (Grenoble), France

sabrina\_ben\_hamou@yahoo.fr

BLOQUET Gwendoline

LIDILEM, Université Stendhal Grenoble 3 (Grenoble), France

bloquetgwen@yahoo.fr

GOOSSENS Vannina

LIDILEM, Université Stendhal Grenoble 3 (Grenoble), France

vannina.goossens@u-grenoble3.fr

RINCK Fanny

LIDILEM, Université Stendhal Grenoble 3 (Grenoble), France

fanny.rinck@u-grenoble3.fr

LOISEAU Mathieu

LIDILEM, Université Stendhal Grenoble 3 (Grenoble), France

mathieu.loiseau@u-grenoble3.fr

**Résumé :** Le travail présenté s'inscrit dans le cadre d'un projet de développement et d'exploitation de ressources textuelles pour la didactique du français à l'aide d'outils de traitement automatique des langues. Il se centre sur le lexique des affects pour rendre compte de la modélisation informatique des ressources textuelles, et de la description du lexique et son annotation sémantique en corpus, sur laquelle se base l'extraction d'exemples et la génération d'exercices

**Mots-clés :** lexique des affects, corpus, annotation sémantique, ALAO

Le travail présenté s'inscrit dans un projet PPF<sup>1</sup> qui vise la constitution et l'exploitation de ressources linguistiques pour l'enseignement/apprentissage de phénomènes énonciatifs entrant dans l'expression du point de vue et de la subjectivité dans les textes. D'abord les affects, où le ressenti de l'auteur et/ou de ses personnages est caractérisé lexicalement en tant que « tristesse », « joie », etc., ensuite la part non affective de la subjectivité du locuteur, telle qu'elle se manifeste à travers la modalisation (notamment l'évaluation), le discours rapporté et la polyphonie. Le projet réunit linguistes, talistes et didacticiens et prend comme point de départ l'idée qu'une approche en corpus de ces phénomènes doit favoriser leur enseignement/apprentissage. Dans une perspective d'aide à la compréhension et à la rédaction de textes, il s'agit de recourir à des données langagières attestées, pour susciter une réflexion métalinguistique des apprenants basée sur le repérage, l'analyse et la manipulation de ces phénomènes, leurs co-occurrences et leurs effets en termes d'expression de la subjectivité.

---

<sup>1</sup> Projet Plan pluri-formation "Développement et exploitation de ressources linguistiques pour la didactique du français à l'aide d'outils de TAL. Etude des marqueurs linguistiques de la subjectivité", 2003-2006.

Nous nous centrerons ici sur le lexique des affects pour traiter du travail préalable à l'exploitation pédagogique<sup>2</sup> : cette dernière requiert d'une part une modélisation informatique des ressources textuelles, pour qu'elles soient utilisables en tant que ressources pédagogiques, par l'enseignant qui peut y puiser des exemples comme par l'apprenant par le biais d'exercices générés à partir de ces textes. D'autre part, l'objet d'enseignement, en l'occurrence le lexique des affects, implique de procéder à une description linguistique du phénomène, qui se poursuit par une annotation dans les textes de la manière dont il s'y présente. Cette annotation est exploitée alors dans l'extraction d'exemples et dans la génération d'exercices.

## 1. Modélisation informatique des ressources

A terme, les ressources constituées dans le cadre du projet PPF seront mises à disposition des enseignants et des apprenants par le biais de la plateforme d'apprentissage des langues MIRTO. Cette plateforme permet, à travers l'exploitation modulaire des ressources, d'éviter les principaux écueils des logiciels d'apprentissage des langues assisté par ordinateur (ALAO), parmi lesquels la pauvreté des informations de sens associées aux données langagières (Antoniadis et al., 2004, 1) et le manque de diversité des exemples disponibles dans le cas d'une exploitation par phrases.

Pour arriver à cette modularité, l'architecture de MIRTO dissocie les données (corpus) des traitements (Antoniadis et al., 2004, 2).

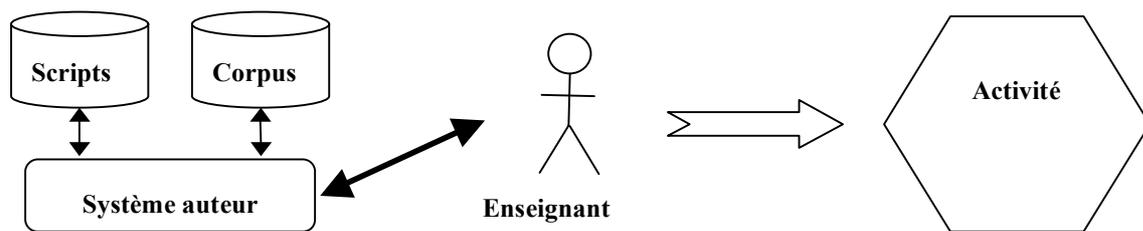


Figure 1 - Génération d'une activité

Dans cette architecture un script est un ensemble de traitements de bas niveaux (fonctions), qui permet, à la donnée d'un texte et de paramètres reflétant le contexte didactique, de générer automatiquement une activité. Le nombre de scripts n'est pas figé, c'est en fonction des objectifs pédagogiques qu'ils sont créés.

<sup>2</sup> Sur des exemples d'exercices et de séances, cf. Rinck et Lebarbé (à paraître).

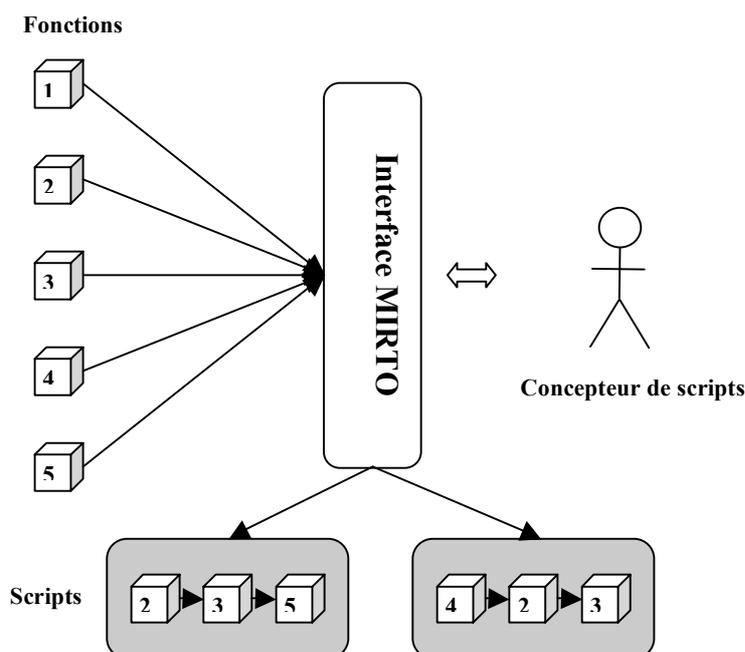


Figure 2 - Création de scripts à partir de fonctions pour un objectif pédagogique donné

Chaque texte, ou ensemble de phrases recueillies dans le corpus, peut donc être présenté à l'utilisateur (enseignant ou apprenant) sous plusieurs formes différentes en fonction des scripts employés (par exemple sous forme d'exercice lacunaire, d'un texte permettant d'accéder à la définition de certains mots, d'un texte dans lequel certains phénomènes énonciatifs sont mis en relief ou même une combinaison des trois).

Dans le cadre du projet, les ressources pédagogiques exploitables ont la particularité de représenter des phénomènes énonciatifs pour l'expression de la subjectivité. Mis à part certains traitements automatiques, pour les cas notamment de l'annotation morphologique et structurelle du document, le repérage de ces phénomènes énonciatifs requiert l'intervention d'un annotateur humain. Cela est d'autant plus vrai dans le cadre pédagogique et didactique où les erreurs d'extraction seraient contraires aux objectifs initiaux. Dans la suite de cette communication, nous présenterons la typologie mise en place pour l'annotation du lexique des affects.

## 2. Le lexique des affects

### 2.1. Description et organisation du lexique

Le lexique des affects a donné lieu à de nombreuses recherches (Mathieu, 2000, Balibar-Mrabti, 1995, etc.) sur lesquelles nous nous sommes appuyés pour réaliser des travaux de description linguistique (Tutin et al., à paraître, Grossmann et Tutin, 2005, Goossens, 2005) et des travaux à visée didactique (par exemple Grossmann et Boch, 2003, Boch et al., 2005, etc.) préalables à l'élaboration de notre typologie. Dans ce projet nous avons privilégié une typologie notionnelle afin de favoriser l'accès sémantique aux éléments lexicaux par les apprenants. De même, nous avons fait le choix de nous limiter à deux niveaux hiérarchiques, domaine (il s'agit d'étiquettes conventionnelles regroupant une même « famille » sémantique d'affects, ex : *joie*) et sous-domaine (ex : *bonheur, plaisir*, etc. pour le domaine de la joie) afin d'en faciliter l'exploitation didactique. Sous chaque type et sous-type, nous avons procédé à l'élaboration de listes du lexique des affects (une liste des noms, une des verbes, etc.).

Etant donné la difficulté de mettre au jour des tests formels pour classer sémantiquement les unités lexicales, nous avons recouru à la « méthode des juges » qui consiste en un accord interannotateurs portant sur les traits attribués aux unités lexicales (marques d'usage, polarité, intensité, etc.).

## 2.2 L'annotation du lexique des affects

L'annotation a été réalisée en deux étapes et sous la forme de balises XML, à l'aide d'outils informatiques : Nooj pour le repérage et le filtrage des unités lexicales appartenant aux affects ; Intex pour l'insertion automatique des traits, vérifiés et désambiguïsés manuellement en fonction du contexte. C'est à cette étape que nous traitons la polysémie du lexique des affects, en nous basant sur des critères linguistiques dépendant du contexte (pluralisation, détermination...). Ceux-ci nous permettent de décider si le sens de l'unité lexicale renvoie à l'affect en lui même (*X éprouve de la joie*) ou à une acception polysémique régulière (Cruse, 1986, Pustejovsky, 1991, 1995, Nunberg & Zaenen, 1997, etc.) comme une qualité causant l'affect (*un livre d'une grande tristesse*), des actes dénotant l'affect (*faire des tendresses*), etc. ou irrégulière.

## Références bibliographiques

- ANTONIADIS Georges, ECHINARD Sandra, KRAIF Olivier, LEBARBE Thomas, LOISEAU Mathieu, PONTON Claude (2004). CALL: from current problems to NLP solutions in *Actes de EUROCALL*, Vienne, Autriche, 1-4 septembre 2004
- ANTONIADIS Georges, ECHINARD Sandra, KRAIF Olivier, LEBARBE Thomas, LOISEAU Mathieu, PONTON Claude (2004) NLP-based scripting for CALL activities, in *Actes du workshop de Coling '04 sur eLearning for Computational Linguistics and Computational Linguistics for eLearning*, Genève, Suisse, août 2004
- BALIBAR-MRABTI, Antoinette (ed.) (2005). Grammaire des sentiments, *Langue Française*, 105.
- BOCH, Françoise, CAVALLA, Cristelle & GROSSMANN, Francis (2005). Apprendre à intégrer le lexique des sentiments au cours de l'écriture de textes narratifs. Colloque international *Lexique et production verbale : vers une meilleure intégration des apprentissages lexicaux*, 9, 10 mars. Lyon : Institut National de Recherche Pédagogique.
- CRUSE, D. Alan (1986). *Lexical Semantics*. Cambridge : University Press.
- GOOSSENS, Vannina (2005). Les noms de sentiment : esquisse de typologie sémantique fondée sur les collocations verbales, *Lidil*, 32 :103-121.
- GROSSMANN, Francis & BOCH, Françoise (2003). Production de textes et apprentissage lexical : l'exemple du lexique de l'émotion et des sentiments, *Repères*, 28 : 117-135.
- GROSSMANN, Francis & TUTIN, Agnès (eds) (2005). Sémantique des noms et adjectifs d'émotion, *Lidil*, 32.
- MATHIEU, Yvette Yannick (2000). Les verbes de sentiment : de l'analyse linguistique au traitement automatique. Paris : CNRS.
- NUNBERG, Geoffrey & ZAENEN, Annie (1997). La polysémie systématique dans la description lexicale, *Langue Française*, 113 : 12-23.
- PUSTEJOVSKY, James (1991). The Generative Lexicon, *Computational Linguistics*, 17 (4) : 409-441.
- PUSTEJOVSKY, James (1995). *The Generative Lexicon*. Cambridge : MIT Press.
- RINCK Fanny, LEBARBE Thomas, (à paraître) Constitution et exploitation pédagogiques de ressources linguistiques pour un enseignement/apprentissage du discours rapporté, in *Actes des journées Linguistique de Corpus*, Lorient, France, septembre 2005.

TUTIN, Agnès, NOVAKOVA, Iva, GROSSMANN, Francis & CAVALLA, Cristelle (2005). Esquisse de typologie des noms d'affect à partir de leurs propriétés combinatoires, *Langue Française*. Paris : Larousse.

### ***Logiciels***

SILBERZTEIN, Max (1992). *Intex*, <http://intex.univ-fcomte.fr>, (nd). Consulté en septembre 2005.

SILBERZTEIN, Max (2002). *Nooj*, <http://perso.wanadoo.fr/rosavram/pages/noojpag.html>, (nd). Consulté en novembre 2005.