

Vers une grammaire de l'évaluation des objets culturels

Dominique Legallois

CRISCO (CNRS – FRE 2 805) – Université de Caen Basse-Normandie

Dominique.Legallois@unicaen.fr

Stéphane Ferrari

GREYC (CNRS – UMR 6 072) – Université de Caen Basse-Normandie

Stephane.Ferrari@info.unicaen.fr

Résumé :

Cette étude traite de l'identification des formes linguistiques destinées à l'expression de l'évaluation des objets culturels. À partir d'un corpus, nous dégagons trois niveaux coordonnés et complémentaires qui, une fois formalisés et implémentés, constituent *une grammaire de l'évaluation* à partir de laquelle le phénomène peut être abordé (au moins partiellement) automatiquement. Ces trois niveaux sont : le niveau expérientiel, le niveau lexico-grammatical, le niveau énonciatif. Nous présentons une expérimentation informatique montrant la faisabilité de l'implémentation, et ouvrant la voie à un type d'analyse discursive automatisée traitant du phénomène de l'évaluation.

Mots-clés : expression de l'évaluation, grammaire locale, expérimentation sur corpus.

Abstract :

This paper focuses on the identification of the linguistics objects used to express evaluation. On the basis of a corpus study, we draw three coordinated and complementary levels of rules which constitute a grammar of evaluation once formalized and implemented. These levels are the experiential, the lexico-grammatical and the enunciative ones. We present a computer experimentation proving the feasibility of an implementation, and leading to further developments for an automatic discourse analyzer taking the evaluation phenomenon into account.

Keywords : formulation of evaluation, local grammar, corpus experimentation.

Introduction

Les sites *amazon.fr* et *fnac.fr* offrent la possibilité pour les lecteurs de déposer leurs avis sur les livres lus, afin de les recommander ou de les déconseiller aux consommateurs internautes. Quelques-unes de ces critiques, réunies dans un corpus restreint à 51 092 mots (représentant approximativement 400 critiques, essentiellement de romans, mais aussi de

BD, de poésie et d'essais) constituent un objet d'observation et d'analyse précieux et complexe pour un travail d'identification des formes linguistiques destinées à l'expression de l'évaluation dans les discours. *Évaluation* est ici entendu dans une acception large, d'ailleurs assez difficile à circonscrire : appréciation des qualités esthétiques, pratiques d'un objet, mais aussi réaction affective, comportementale, voire somatique face aux qualités des choses, des personnes et des événements, ou encore jugement d'ordre moral ou éthique. Notre objectif à terme, est l'élaboration d'une *grammaire de l'évaluation* des objets culturels (livres, films, théâtre, etc.) implémentable à des fins d'extraction automatique, pour l'analyse de comparaison entre textes : devant la complexité et la permanence du phénomène évaluatif, nous nous attendons à des mises en discours fortement différentes de son expression selon les objets évalués et les stratégies évaluatives.

Notre approche est donc fondée sur l'usage discursif¹, plutôt que sur la description lexicologique entreprise, par exemple, par certains travaux sur le lexique des sentiments (cf. le numéro 105 de *Langue Française*, ou Mathieu 2000) ; elle diffère également des analyses thématiques (Rastier 1995), fondée sur l'étude lexicométrique, en privilégiant les propriétés lexico-grammaticales et énonciatives d'expressions routinières. En effet, notre corpus comprend nombre de textes largement stéréotypés, dans lesquelles les expressions dédiées à l'évaluation sont souvent préconstruites, prédonnées : le langage évaluatif portant sur les livres est en partie formulaire.

Nous reconnaissons à l'évaluation une portée considérable sur des champs linguistiques entiers : au niveau lexical², un grand nombre de lexèmes, quelle que soit leur catégorie, sont par nature évaluatifs : *tragédie, succès, splendide, échouer, réussir, heureusement*, etc. ; au niveau des constructions, le système comparatif et superlatif constitue une expression grammaticalisée de l'évaluation ; au niveau énonciatif, la pragmatique linguistique inspirée par Ducrot et Anscombe place la fonction évaluative au cœur même du dispositif argumentatif : *c'est un bon livre* ne constitue pas d'emblée un énoncé informatif, c'est d'abord un énoncé argumentatif fondé sur une évaluation³ ; les marqueurs enclosifs⁴ (*une espèce de/un vrai/un véritable N*) sont employés dans des énoncés évaluatifs ; enfin, un certain type de métaphores nominales, parmi les plus répandues et acceptant les modifications enclosives, est essentiellement évaluatif (*Sophie est une vipère*). Mais ces champs ne seront pas étudiés ici pour eux-mêmes ; notre point de départ reste les discours avérés et leur complexité, et l'évaluation sera considérée comme phénomène textuel plutôt que phénomène inhérent à la langue elle-même.

Nous présentons dans ce qui suit, les caractéristiques des niveaux fonctionnels dégagés par l'observation et l'étude du corpus ; nous donnons ensuite, avant d'exposer la méthode d'implémentation employée, l'exemple d'une analyse d'un texte qui illustre l'enchevêtrement des niveaux à la base de notre *grammaire*.

Analyse du corpus

À l'issue de l'examen de notre corpus, nous avons considéré trois niveaux fonctionnels complémentaires et interactifs pour la constitution de la *grammaire de l'évaluation* :

1. L'analyse linguistique de l'acte d'évaluation a fait récemment l'objet d'un certain nombre d'études « corpus driven approach » dans le domaine anglo-saxon : par exemple, l'ouvrage collectif sous la responsabilité de S. Hunston et G. Thompson (2000) dont les contributions montrent des traitements sémantiques et grammaticaux possibles de l'évaluation ; et, de façon encore plus substantielle, le travail de J. Martin et P. White (2005) sur la notion d'*appraisal*, dans la perspective de la grammaire fonctionnelle systémique.
2. C. Kerbrat-Orecchioni 1997.
3. On pourra lire O. Galatanu (2002) pour l'analyse des valeurs intrinsèques au lexique et leurs effets en discours.
4. Cf. Legallois 2002.

- Niveau des cadres expérientiels ;
- Niveau des séquences lexico-grammaticales ;
- Niveau des configurations énonciatives.

Ces trois niveaux correspondent aux méta-fonctions que distinguent Halliday (1996) : fonction idéationnelle (pour nous, cadre expérientiel), fonction textuelle (niveau lexico-grammatical), fonction interpersonnelle (niveau énonciatif).

I Les cadres expérientiels

Le premier niveau identifie les aspects de l'objet évalué. Une analyse de l'évaluation d'un livre est vite confrontée à un problème inhérent à la constitution de l'objet même : on peut évaluer différents aspects ou *qualia* ; par exemple, le contenu, le style, la satisfaction ou la déception par rapport à des attentes, etc. L'évaluation peut porter également sur l'auteur du livre, sur l'histoire. Autrement dit, la forme de l'expression d'un jugement est naturellement configurée par rapport à ce que nous avons nommé des cadres expérientiels. Quelques exemples de cadres :

L'emprise du livre sur le lecteur : On ne peut plus le lâcher, jusqu'à la fin/Comme beaucoup d'entre vous, je suis tombée sous le charme de la douceur du récit de Philip Roth.

Les attentes satisfaites ou non du lecteur : Je reste de loin sur ma faim/Je m'attendais à mieux de K. DICK/J'ai été surprise par le style de ce livre/Vivement la suite !

L'effort investi pour sa lecture : Lisez le livre, il en vaut la peine/Le livre se lit facilement et rapidement/Il faut s'accrocher au début

Son impact affectif sur le lecteur : On pleure un peu, on rit, on s'émeut!...

Sa valeur axiologique : L'Aliéniste est avant tout un EXCELLENT roman.

La prescription ou la proscription du livre (recommander un livre est une façon indirecte mais implacable de l'évaluer positivement) : A conseiller pour ceux qui aiment les thrillers.

Ces cadres, même s'ils sont identifiés à partir d'un corpus précis, sont suffisamment généraux pour être appliqués à l'évaluation d'autres objets culturels ; en effet, l'observation d'avis portant sur des CD musicaux, des jeux vidéos ou des films permet de constater la présence de cadres identiques. Ce phénomène s'explique ainsi : l'évaluation porte rarement sur les propriétés intrinsèques de l'œuvre, mais sur les rapports que les sujets ont avec cette œuvre. De ce fait, les aspects jugés par la critique livresque sont facilement transposables à d'autres objets : efforts, impacts affectifs, prescriptions, attentes, mais aussi style, effets hédoniques (par ex. *passer un agréable moment : Voici le plus beau recueil de lettres au collègue de pataphysique. Un réel moment de bonheur de découvrir ce monde inexploré* (à propos de *Je voudrais pas crever* de B. Vian)), etc. sont autant de cadres communs à l'expérience des objets culturels.

II Séquences lexico-grammaticales

Le second niveau est celui des séquences lexico-grammaticales ; c'est ainsi que nous proposons une articulation du phénomène phraséologique à l'analyse de l'évaluation. À condition de ne pas voir dans la phraséologie un ensemble de formes radicalement figées, il est possible de concevoir des séquences lexico-grammaticales récurrentes, bien que polymorphes, dédiées ici à l'évaluation. Autrement dit, notre tâche a été de recenser les expressions « préfabriquées », de la simple collocation (par ex. *conseiller vivement*) aux configurations plus larges. Par exemple :

on n'a jamais aussi bien rendu l'amour réciproque/Aucun livre de ma connaissance n'a jamais si bien démontré [...] les dégâts [...] que peuvent occasionner la vie

ce « pattern » [*ne jamais (aus)si bien + verbe de représentation/explication*] est ici considéré comme une construction relativement ouverte, mais constituant malgré tout une unité prédonnée, directement disponible dans la compétence linguistique du locuteur. Les séquences lexico-grammaticales ont en partie été repérées grâce au logiciel « Collocates⁵ » qui permet d'identifier les n-grams du corpus ; nous procédons à une vérification afin de nous assurer que les répétitions collocatives sont porteuses d'évaluation ou en sont des indices.

Parmi ces séquences, certaines sont entièrement dédiées à un cadre expérientiel, d'autres sont beaucoup plus indépendantes et peuvent s'actualiser dans plusieurs cadres. Nous donnons quelques exemples parmi les dizaines répertoriées (à noter que l'*évalué* renvoie à l'objet évalué, l'*évaluatème* à la valeur accordée à l'évalué, le *siège* à la personne qui « expérimente » l'évalué – le *siège* peut être ou non l'évaluateur) :

[à lire absolument] : cette séquence figée, employée 16 fois dans le corpus, s'actualise dans le cadre « prescription », comme la collocation [[Évaluateur [*conseiller vivement*]] [Évalué]]

[siège] [*ne pas pouvoir lâcher avant/jusque*] : cette séquence (11 occurrences) s'actualise dans le cadre « emprise », et connaît plusieurs réalisations :

Pas question de lâcher le bouquin avant la fin.

Je n'ai pas pu le lâcher avant de l'avoir terminé.

On ne peut plus le lâcher, jusqu'à la fin.

On ne parvient à lâcher le roman qu'à la dernière page.

(enfin/voilà/voici) un [évalué] qui [évaluatème] : il s'agit d'une construction à phrase averbale particulièrement récurrente dans le corpus (22 fois). Cette séquence s'actualise dans plusieurs cadres possibles : *un livre qui donne à rêver* (cadre « hédonique ») ; *un livre qui fait réfléchir* (cadre « valeur intellectuelle ») ; *un roman qui tire le lecteur entre notamment l'humour, l'amour, les rejets, les situations grotesques* (cadre « emprise »).

Det ([enclosure]) [évaluatème] : cette séquence s'actualise principalement dans le cadre « valeur » : *Dix petits nègres est un vrai petit bijou* ; *un vrai petit Jules Vernes ou Barjavel* ; la présence de l'enclosure ici, est un indice imparable de la fonction évaluative du terme subséquent. Ainsi, *Jules Vernes/Barjavel* sont-ils étiquetés *évaluatèmes*.

Nous recensons ainsi près d'une trentaine de séquences évaluatives ou introductrices d'évaluation dont les rôles thématiques sont étiquetés non pas à partir de catégories générales (par ex. agent, bénéficiaire, etc.), mais à partir de rôles propres à l'expression de l'évaluation. Ces séquences sont de dimensions et de natures hétérogènes : du syntagme récurrent à la phrase figée. Là encore, une projection sur d'autres textes (projection qui n'est pas encore systématisée à l'heure actuelle) permet de voir des constructions fort apparentées sémantiquement et grammaticalement ; par exemple, au sujet de l'audition du requiem de Mozart :

Cette interprétation du requiem k626 est un véritable feu d'artifice. J'en suis resté scotché sur mon fauteuil. Bravo ! (amazon.fr)

Ou à propos du jeu vidéo *Morrowind* :

Ce jeu est tout simplement magnifique : si vous avez une x-box, *Morrowind* est incontournable. Les graphismes sont superbes et l'ambiance vous immerge totalement dans l'univers. Les quêtes sont très variées et le joueur ne s'ennuie jamais : il y a toujours quelque chose à faire !!! Je suis resté scotché sur ce jeu pendant toute une semaine et je suis même pas au 1/4 du jeu ! Je le recommande même à ceux qui ne sont pas spécialement fan du genre : vous ne serez pas déçu ! (amazon.fr)

Ainsi, dans la perspective d'une implémentation rendant compte de l'évaluation de tout objet culturel, il est important d'assigner aux deux séquences *ne pas pouvoir lâcher/rester*

5. Conçu par Michael Barlow.

scotcher une catégorie subsumant les diverses réalisations. C'est par ce travail de généralisation que pourra être établie une systématité valant pour l'ensemble des objets culturels.

III Configurations énonciatives

Le niveau énonciatif est fondamental pour une analyse générale du discours évaluatif de l'objet culturel⁶. Les évaluations, en tant qu'acte de discours, doivent être mesurées selon leur force illocutoire. C'est à ce niveau que s'articulent et se construisent les stratégies argumentatives : il s'agit, pour le locuteur, de se mettre en scène pour faire partager son avis : premier plan, engagement, retrait, prise en charge faible, etc. Cette mise en scène, dans notre corpus, est relativement normée dans la mesure où le genre est lui-même partiellement stéréotypé ; mais là encore, la formalisation du niveau énonciatif devra permettre toute projection vers d'autres objets afin d'élaborer des points de comparaisons et de différences.

Ainsi, par exemple :

Les marqueurs restreignant au seul énonciateur la validation de l'énoncé : *À mon goût, à mon avis, selon moi.*

Les marqueurs délimitant le public intéressé : *une mine d'informations pour tous ceux qui s'intéressent à la psychologie en général.*

Les verbes d'attitude propositionnelle (impliquant la modalité épistémique): *Je crois que Philip Roth a atteint le sommet avec Opération Shylock.*

Les tournures concessives : *Ce bouquin est certes intéressant au début, mais il devient très vite rébarbatif.*

Les adverbess intensifs (marquant explicitement le degré d'engagement de l'énonciateur): *Vraiment, véritablement, absolument, impérativement, totalement, etc.*

Pronoms personnels (l'évaluateur peut s'effacer devant l'expérimentateur, attribuer le jugement à une instance collective, projeter une évaluation du destinataire, etc.): *Plus vous avancerez dans la lecture, plus vous serez dégoûtés par ce simili d'érudition prétentieux et bourré de fautes !*

Les interjections : *Vraiment, beurk...*

Ce niveau est le plus complexe des trois à formaliser dans la mesure où les formes sont extrêmement hétérogènes, de dimensions parfois larges, dépassant le simple énoncé. La « stratégie » consiste en fait à s'appuyer le plus possible sur les séquences lexico-grammaticales, qui constituent à notre avis, le niveau intermédiaire entre niveau des cadres expérientiels et niveau des configurations énonciatives. Nous voudrions illustrer ce phénomène par un exemple.

IV Exemple d'un traitement d'un avis

L'exemple tiré du corpus est le suivant (à propos de *Le sang du temps* de Maxime Chattam) :

je suis déçu par ce livre, on regrette la fameuse trilogie. Malgré cela on se laisse quand même entraîner dans notre lecture mais pas jusqu'à l'envoûtement.

On voit ici les divers niveaux enchevêtrés dont nous donnons les éléments dans un tableau pour faciliter la lecture :

6. Cf. Charaudeau 1988.

FORMES	CADRES EXPERIENTIELS	SÉQUENCES LEXICO-GRAMMATICALES	CONFIGURATIONS ÉNONCIATIVES
je suis déçu par ce livre	Attente (déçue)	[Évalué] [<i>décevoir</i>] [Évaluateur]	Je = Évaluateur Énonciateur
on regrette la fameuse trilogie	Attente (déçue)	[Évaluateur] [<i>regretter</i>] [Évaluer]	On = Évaluateur collectif
la fameuse trilogie	Renommée	[adj. Évaluatème] [Évalué]	Évaluateur collectif
on se laisse quand même entraîner dans notre lecture	Emprise	[Siège de l'expérience] [<i>se laisser entraîner/</i> <i>envoûter/prendre</i>] [<i>dans/par</i> Évalué]	On = Évaluateur collectif
Malgré		Connecteur argumentatif	concession
quand même		Connecteur argumentatif	concession
mais pas jusqu'à l'envoûtement	Emprise	[<i>jusqu'à</i> Évaluatème]	Force de l'évaluation

L'illustration par ce simple exemple montre la difficulté mais aussi l'intérêt de la tâche. Dans le même texte, figurent plusieurs cadres expérimentiels ; chaque cadre est cependant marqué par des indices lexicaux et grammaticaux propres qui suffisent à identifier la nature du cadre. De même, les valeurs énonciatives accordées aux pronoms sont inférées d'une part, de la forme même du pronom (je – on), et, d'autre part, de la combinaison entre le pronom et le verbe, voire de la valeur intrinsèque de l'adjectif (*fameux*, dans cet emploi, est un adjectif médiatif indicateur d'un jugement collectif). On notera que la notion d'*évaluateur collectif* ne renvoie pas à une instance énonciative, ni à une source évaluative effective : il s'agit d'une construction – d'une stratégie – de la part du locuteur afin de minimiser son engagement personnel, mais aussi de l'inscrire dans une participation collective certes factice mais efficiente.

L'étiquetage des séquences lexico-grammaticales s'appuie sur les données d'un analyseur syntaxique (*Tree Tagger*) ; les données de *Tree Tagger* doivent pourtant être reconfigurée dans un format plus « sémantique » dans lequel figurent, par exemple, les rôles thématiques propres à l'évaluation, ou encore la fonction de connecteur à portée énonciative.

Ni le tableau, ni l'analyse implémentée ne donnent directement l'interprétation globale de la nature évaluative ; plutôt, ils fournissent les indices discursifs généralisés nécessaires à l'interprétation « humaine », et conduisent à une factorisation des données utiles au balisage de parcours interprétatifs.

Expérimentation informatique

Afin d'expérimenter le modèle sur corpus, nous utilisons *LinguaStream*⁷, une plateforme de TAL qui permet notamment l'utilisation dans une même chaîne de traitements de différents formalismes (Widlöcher & Bilhaut 2005, Enjalbert 2005, chap. 10). L'objectif est pour nous de réaliser un outil informatique facilitant l'observation des régularités lexico-grammaticales précédentes, tant sur le corpus d'étude original que sur de nouvelles données. Nous visons à plus long terme la possibilité d'apprécier de manière semi-automatisée la variation de l'expression de l'évaluation selon les textes.

Une expérimentation comme celle que nous proposons ici suppose de reformuler l'ensemble de nos hypothèses précédentes, à caractère plutôt descriptif, en un modèle opératoire, à caractère prescriptif, comme montré par Ferrari *et al.* (2005). Les formalismes mis

7. <http://www.linguastream.org/>

à disposition dans LinguaStream laissent une grande liberté dans l'expression du modèle opératoire, qui peut être mis en œuvre tant à l'aide d'automates de type expressions régulières que de grammaires de type Prolog. Nous avons tiré parti de cette offre, certains types d'analyse étant mieux adaptés à la mise en œuvre des patrons lexico-grammaticaux, d'autres à la « remontée » d'informations sémantiques depuis un lexique jusqu'à des éléments textuels. Cependant, ni les formalismes exploités ni les composants développés ne sont le reflet direct de l'analyse précédente, car les trois niveaux de la grammaire proposée sont en réalité dispersés dans de multiples composants, réalisés à l'aide de formalismes différents, d'une part, et certains cohabitent quelquefois au sein d'un même composant, d'autre part. Enfin, le corpus d'origine a été préalablement transcodé en XML, selon les méthodes préconisées par Habert *et al.* (1998). Il contient désormais des informations sur les éléments logiques des avis, selon leur disponibilité : titre, date, lecteur diffusant l'avis, titre et auteur du livre visé... L'extrait de la figure 1 permet d'apprécier ces différents éléments.

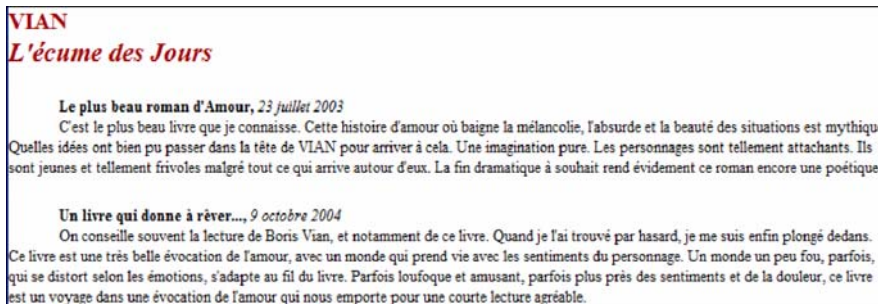


Figure 1 : extrait du corpus en version structurée XML. (L'affichage est réalisé à l'aide d'une feuille de style CSS qui différencie visuellement les éléments de structure XML. Les avis y sont regroupés par œuvre, tels que collectés.)

Chaîne de traitements pour observer l'expression de l'évaluation

Parmi les trois niveaux de la grammaire précédente, nous proposons une première mise en œuvre particulièrement adaptée pour les niveaux I et II, c'est-à-dire le niveau expérientiel et le niveau lexico-grammatical. Nous donnons quelques pistes pour permettre une prise en considération du niveau III de l'énonciation.

La chaîne LinguaStream de la figure 2 montre les différents composants utilisés pour l'expérimentation. Chaque boîte y représente un composant ou une ressource, les flèches entre les boîtes représentent la transmission d'information entre composants. La première colonne de composants consiste en quelque sorte en une préparation du corpus aux analyses suivantes. La première boîte représente une ressource : la version XML du corpus (CCL pour corpus de Critiques de Livres). La boîte suivante, CCL XML Marker, est un composant LinguaStream permettant de sélectionner les éléments XML pertinents d'une ressource structurée pour les analyses ultérieures, et, le cas échéant, de les typer ; dans notre cas, nous concentrons les analyses sur le titre et le corps des avis, les informations concernant par exemple les dates et les auteurs des avis seront ignorées des analyses menées ultérieurement. Les deux boîtes suivantes représentent une segmentation en mots (Tokenizer) et la catégorisation grammaticale à l'aide du *tree tagger* (Schmid 1994). À l'issue de cette première colonne de composants, la chaîne d'analyse se poursuit avec la transmission de deux informations en parallèles : une version du document d'origine enrichi au fur et à mesure d'ancres permettant d'y repérer les différents éléments analysés, et les résultats des analyses, transmis en parallèle et codés dans un fichier indépendant lors d'une sauvegarde. Cette première colonne de composants influence la qualité des résultats des composants dédiés à la mise en œuvre de notre modèle, dans la mesure où ils exploitent une partie des informations qui y ont été produites.

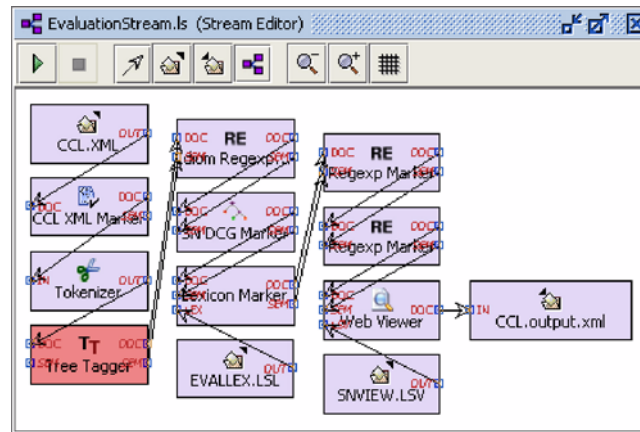


Figure 2 : chaîne de composants d'analyse dans LinguaStream.

La deuxième colonne de composants est celle qui représente le plus la partie des analyses qui concerne notre modèle. La boîte RE – Idiom Regexp exploite des automates pour une amorce de l'analyse des formes lexico-grammaticales, fondée sur la présence de certains mots dans un certain ordre, avec vérification de la catégorie grammaticale si besoin. Ainsi, la structure *Aucune comparaison avec [comparant]* s'y traduit par la règle déclarative suivante :

```
<idiom> privatif () % [0-2] {lemma : comparaison} </idiom>
/sem {synt : SPpost_avec, sem : aucuncomp, eval : idiom}
```

et une règle intitulée « privatif » exploitée par la précédente et disponible pour d'autres :

```
(« pas » « de » | « plus » « de » | « guère » « de » | {lemma : aucun} | {lemma : nul})
```

La première règle permet de marquer comme élément *idiom* un mot dont le lemme est « comparaison » et qui est précédé d'un privatif, un ou deux mots supplémentaires pouvant s'intercaler. L'information qui est associée à l'élément découvert est une structure de trait renseignant sur la nature de l'élément repéré et/ou précisant quelle analyse mener ensuite pour compléter le patron : *eval : idiom* permet de caractériser ici un type de résultat de l'analyse de l'évaluation, *sem : aucuncomp* précise quel patron a été employé, *synt : SPpost_avec* sera utilisé par un composant ultérieur pour associer le syntagme prépositionnel suivant l'expression repérée. La deuxième règle a pour objectif de généraliser le patron initialement observé, afin de permettre une certaine variabilité lexicale lors de la confrontation à de nouveaux corpus.

La deuxième boîte SN DCG Marker représente un composant d'analyse de syntagmes nominaux. Il s'agit d'une grammaire Prolog⁸ dans laquelle nous avons injecté une partie de l'information lexicale liée à notre modèle. Les clauses suivantes illustrent la notion pour les noms présents dans notre lexique :

```
nom (lem : L..E) --> ls_lookupToken (_,tag : nom..lemma : L,_) , ls_lexicon (eval, E, lemma).
nom (lem : L) --> ls_token (_,tag : nom..lemma : L).
```

La première clause, appliquée en priorité lors de l'appel du prédicat *nom ()*, permet de récupérer le lemme issu des premières analyses (*tree tagger*), sans consommer le mot analysé,

8. Composant réalisé en collaboration avec T. Charnois, GREYC – CNRS UMR 6072. Travail en cours. Plus précisément, les clauses exploitent le formalisme GULP, proposé par Covington (1994), pour permettre la manipulation en Prolog des structures de traits.

et de combiner à cette information celle présente dans un lexique de formes lemmatisées. Si le mot analysé n'est pas dans le lexique, la deuxième clause s'appliquera alors, se contentant de récupérer le lemme du nom.

Dans cette phase, toute information lexico-sémantique susceptible de concerner l'expression de l'évaluation est exploitée. Pour le niveau I, on retrouve par exemple le cadre expérientiel de la valeur axiologique, avec des adjectifs comme *bon*, *mauvais*, *superbe*, *extraordinaire*... Pour le niveau III, on retrouve des informations sur les adverbes intensifs, les interjections... Pour le niveau II, l'analyse effectuée directement le repérage des structures pré-construites comme *Det ([enclosure]) vrai/véritable/pur [évaluatème]*. Il s'agit à cette étape de l'analyse de fournir aux modules suivants toute information lexicalisée susceptible d'être exploitée pour un niveau ou pour un autre de la grammaire.

Les deux dernières boîtes de la deuxième colonne de composants représentent un complément d'analyse lexicale permettant de compléter l'information précédente notamment pour la catégorie verbale, qui n'est pas actuellement exploitée par le module d'analyse des syntagmes. Ce découpage temporaire des premiers modules reste quelque peu artificiel, il est destiné à terme à être remplacé par l'utilisation d'un analyseur syntaxique robuste et d'une analyse lexicale unique.

L'exploitation de tous les résultats précédents se fait par les composants représentés par les deux premières boîtes de la troisième colonne, qui consistent en un filtrage des informations précédemment associées aux syntagmes pour ne conserver que celles en rapport avec notre étude. En effet, sur le corpus de critiques de livres, conserver l'ensemble des informations de cette analyse conduit à un fichier de 30 Mo pour l'affichage dans un navigateur (présence de nombreuses divisions HTML cachées contenant les informations associées aux syntagmes). Mais ces derniers composants dédiés ont pour objectif essentiel de limiter la taille du fichier destiné à l'affichage ; rien n'empêche de stocker par ailleurs la totalité des informations pour une autre exploitation. Les autres composants représentés dans la troisième et la dernière colonne consistent en une préparation à l'affichage, conduisant aux exemples présentés dans la section suivante.

Exemples de résultats

Les multiples analyses précédentes permettent d'associer et de combiner différentes informations à différentes unités textuelles. Le résultat consiste la plupart du temps en l'association d'une structure de traits comme celles de la figure 3 à une unité textuelle particulière.

meurtres? Comme l'a dit précédemment Fabcorbille, «pourtant, la solution de ce petits nègres est [+][un vrai petit bijou]: ie viens [+][à peine] de le terminer que de autre [+][de ses ro	g/40 [X] de partage	qui la est de loin mon préféré. Tellement proche d'une t et angoisse [+][une vrai perle] [+][de sueur froide]... Déposé par policière su	g/56 [X]
<p>Chef-d'oe Agatha nou rebondissements. chef]-d'oeuvre de Steve Hode L'affaire nouvelles</p> <p>[+][Bonne Steve Hode parcours d'un ass discutées. Il faut le du mal à adhérer à</p>	<p>det: un</p> <p>nom: lem: bijou</p> <p>sn: adj: niveau: 2</p> <p>adj.Ant: qualif: enclosure</p> <p>sapost: adj: lem: petit</p>	<p>compl</p> <p>02 ts nai s e</p>	<p>det: une</p> <p>nom: lem: perle</p> <p>sn: lem: vrai</p> <p>adj.Ant: adj: niveau: 2</p> <p>qualif: enclosure</p>

Figure 3 : 2 exemples de structures de traits et tolérance à l'agrammaticalité.

Dans ces exemple, le groupe nominal « un vrai petit bijou » est repéré comme pertinent pour notre analyse, c'est-à-dire exprimant l'évaluation ou en rapport direct avec un autre

élément l'exprimant. En consultant la structure de traits associés, on remarque la présence d'un adjectif marquant une enclosure, issu du niveau II de notre grammaire. Les analyses menées permettent de repérer aussi cette structure lorsque l'accord est incorrect, comme dans le deuxième extrait de la figure 3, « une vrai perle ». En effet, les règles Prolog d'analyse locale n'exploitent que la catégorie grammaticale principale, sans vérifier ni tenir compte de l'accord en genre et en nombre. Le filtrage des informations permet d'obtenir des résultats comme celui de la figure 4, où un des verbes associé au cadre expérientiel de l'emprise est suivi de groupes prépositionnels pour lesquels l'information syntaxico-sémantique est conservée.

The image shows a text analysis interface. The main text is: "d'imagination pour se laisser emporter [...] [dans le rêve merveilleux] [...] [de Tolkien]. Cet ouvrage est de plus d Jackson est largement inspiré". Below this, there are several lines of text: "[+] [le plus beau livre", "Beaucoup sont ceux qui", "tu le prologue pour la premier", "devient palpitant. Ce livre vou", "quand on les a fini on en veut". To the right of the text, there is a box containing the following information: "eval/486 [X]", "cadre: emprise", "synt: SPdans", "niveau: I", and "21 juin 2003". Below this box, there is a paragraph of text: "e cet ouvrage et qui ont été découragés par la difficulté à co", "é mais j'ai continué car j'en avais eu [...] [des bons échos]. P", "apporte [...] [dans un monde] a la fois similaire et si différent d", "dit qu'on aurait du le lire moins vite."

Figure 4 : filtrage pour affichage.

Les résultats obtenus à l'heure actuelle permettent essentiellement de valider les motifs initialement proposés et d'envisager leur exploitation sur un autre corpus. Les structures de traits qui sont construites à ce stade d'avancement de nos travaux et la nature des éléments sur lesquels elles portent n'ont pas de fait un caractère définitif. Il reste notamment à mener une réflexion sur le type d'information que nous désirons y faire figurer, cela en rapport avec un éventuel cadre applicatif particulier. Dans leur état actuel, il est toutefois déjà possible de remonter par exemple sur des unités telles la phrase ou des éléments de structure logique (paragraphe, section et avis dans ce corpus) une information quantifiée indiquant combien d'éléments de chaque niveau de notre grammaire ont été employés au sein de telles unités, ou encore quels cadres expérientiels.

Conclusion générale et perspectives

Nous avons proposé une grammaire locale de l'évaluation s'articulant sur trois niveaux : expérientiel, lexico-grammatical et énonciatif. Une première expérimentation a montré la faisabilité de la mise en œuvre pour les deux premiers niveaux de cette grammaire. Cette expérimentation exploite une chaîne de traitements fondés sur des formalismes différents ; cette chaîne utilise de multiples composants linguistiques déjà développés par ailleurs et pour d'autres besoins. Le dernier niveau, celui de l'énonciation, présente cependant un degré de difficulté supérieur. Il nécessite selon nous la manipulation d'unités discursives diverses pour lesquels les formalismes actuellement exploités (automates, grammaire Prolog) ne sont pas les mieux adaptés. Aussi, nous envisageons un nouveau composant consacré à ce niveau, qui permette l'expression de contraintes sur des unités variées, sans tenir compte nécessairement de l'ordre entre ces unités ni de l'ordre dans leur traitement, en s'appuyant sur un formalisme tel celui proposé par Widlöcher (2006).

Nous visons une implémentation possédant plusieurs champs d'applications possibles. Ainsi, nous voudrions examiner les éléments généraux communs à l'évaluation d'objets culturels différents, afin d'extraire les éléments constitutifs du genre. Les différences spécifiques de chaque objet devront bien sûr être théorisées (par exemple, une critique cinéma peut porter sur le jeu des acteurs, aspect qui n'a pas sa contrepartie dans la

critique livresque). Une observation des différences de modalités évaluatives selon les genres des livres devra être menée (intuitivement, une critique d'un roman paraît différente d'une critique d'un essai ou d'une BD).

La projection du système pourra permettre également de mesurer sur corpus proximités et différences entre critiques « amateurs » et critiques professionnelles (par exemple, au niveau des cadres expérientiels) dans l'espoir d'analyser les modes d'institutionnalisation de la critique. D'autres types de textes devront également être pris en compte ; nous pensons à la publicité des objets culturels (encart publicitaire pour tel livre dans tel quotidien) : on perçoit des liens évidents entre le langage formulaire des internautes et ceux des publicitaires, par exemple dans la construction averbale très récurrente : *un livre qui + évaluatème*.

Enfin, l'implémentation d'une grammaire de l'évaluation peut apporter un outil appréciable pour procéder à une comparaison entre types de textes différents, mais qui partagent tous la mise en discours des valeurs individuelles et collectives propres à une société : discours épидictiques⁹, politiques et idéologiques. Construire en discours l'évaluation, c'est toujours construire sa propre subjectivité (et son ethos) pour l'orienter dans le champ des valeurs sociales.

Bibliographie

- CHARAUDEAU P. (1988), « La critique cinématographique : faire voir et faire parler », in *La presse : produit, production, réception*, Didier érudition (Langages Discours et Sociétés), p. 47-70.
- COVINGTON M. A. (1994), GULP 3.1: An Extension of Prolog for Unification-Based Grammar. *Research Report AI – 1994 – 06*, The University of Georgia, Artificial Intelligence Center, Athens, Georgia, USA.
- DOMINICY M. & FREDERIC M. (éds.) (2001), *La mise en scène des valeurs : la rhétorique de l'éloge et du blâme*, Lausanne, Delachaux et Niestlé.
- ENJALBERT P. (dir.) (2005), *Sémantique et traitement automatique du langage naturel*, Hermès Sciences, Traité IC2.
- FERRARI S., BILHAUT F., WIDLÖCHER A. & LAIGNELET M. (2005), « Une plate-forme logicielle et une démarche pour la validation de ressources linguistiques sur corpus : application à l'évaluation de la détection automatique de cadres temporels », in *Actes des 4^{es} Journées de la Linguistique de Corpus*, G. WILLIAMS (éd.), à paraître aux Presses Universitaires de Rennes.
- GALATANU O. (2002), « Le concept de modalité : les valeurs dans la langue et dans le discours », in *Les valeurs : séminaire Le lien social*, Nantes, 11 et 12 juin 2001/organisé par le CALD-GRASP ; coord. scientifique et présentation Olga Galatanu, Maison des Sciences de l'Homme Ange Guépin.
- GROSS M. (1995), « Une grammaire locale de l'expression des sentiments », *Langue Française*, 105, p. 70-87.
- HABERT B., FABRE C. & ISSAC F. (1998), *De l'écrit au numérique : constituer, documenter, normaliser un corpus électronique*, Paris, InterEditions.
- HALLIDAY M. A. K. (1996), *An introduction to functional grammar*, Sydney, Arnold.
- HUNSTON S. & THOMPSON G. (éds) (2000), *Evaluation in Text. Authorial Stance and the Construction of Discourse*, Oxford, Oxford University Press.
- KERBRAT-ORECCHIONI C. (1997), *L'énonciation : de la subjectivité dans le langage*, Paris, A. Colin.
- LEGALLOIS D. (2002), « Incidence énonciative des adjectifs vrai et véritable en antéposition nominale », *Langue Française*, 136.
- MARTIN J. & WHITE P. (2005), *The Language of Evaluation: Appraisal in English*, Palgrave Macmillan Hardcover.
- MATHIEU Y.Y. (2000), *Les verbes de sentiment : de l'analyse linguistique au traitement automatique*, Paris, CNRS Éditions.
- RASTIER F. (dir.) (1995), *L'analyse thématique des données textuelles : l'exemple des sentiments*, Paris, Didier érudition.

9. Dominicy & Frédéric 2001.

SCHMID H. (1994), « Probabilistic Part-of-Speech Tagging Using Decision Trees », *International Conference on New Methods in Language Processing*, Manchester, UK.

WIDLÖCHER A. (2006), « Analyse par contraintes de l'organisation du discours », in *Actes de la Conférence Traitement Automatique du Langage Naturel (TALN 2006)*, Louvain, Presses universitaires de Louvain, Belgique, p. 367-376.

WIDLÖCHER A. & BILHAUT F. (2005), « La plate-forme LinguaStream : un outil d'exploration linguistique sur corpus », in *Actes de la 12^e Conférence Traitement Automatique du Langage Naturel (TALN)*, M. Jardino (éd.), ATALA LIMSI, Dourdan, France, p. 517-522.