



2 Juin 2015 - 24 Juillet 2015

Contribution à la détection de systèmes de valeurs et de catégories de consommateurs en analyse d'opinion

Maria BORITCHEV

Sous la direction de
Patrick SAINT-DIZIER

Résumé

L'analyse de l'opinion d'un groupe de consommateurs sur un produit donné s'effectue, dans un premier temps, par détection des critères évaluatifs propres au groupe considéré. Cette évaluation permet de subdiviser le groupe en catégories de consommateurs en fonction du degré d'importance que ceux-ci attachent à ces critères. Au-delà de cette première analyse, il est intéressant de se pencher sur la justification de ces résultats, en procédant à une analyse sémantique de l'argumentation utilisée dans l'expression de l'opinion. Ainsi, il devient possible de préciser les systèmes de valeurs des consommateurs, en les contextualisant.

Table des matières

1	Corpus	4
1.1	Démarche de travail	4
1.2	Analyse linguistique et conceptuelle	6
2	Analyse lexicale et syntaxique	7
2.1	La théorie \bar{X}	7
2.2	Patrons et grammaires	9
2.3	TextCoop	10
3	Analyse sémantique et pragmatique	10
3.1	Arguments et argumentation	10
3.2	Modélisation linguistique du système d'argumentation	11
	Références	15
	Annexes	16
	Annexe 1 : catégories de films proposées par IMDb (données du 11 juillet 2015) .	16
	Annexe 2 : extraits du corpus : critiques de <i>The Sound of Music (La mélodie du bonheur)</i>	17
	Annexe 3 : exemple d'exécution de TextCoop	19

Introduction

La détection de systèmes de valeurs et des catégories de consommateurs en analyse d'opinion est un problème complexe dont les solutions présentent un intérêt majeur. En effet, il s'agit d'étendre la connaissance que l'on peut avoir d'un domaine et surtout de ses utilisateurs, permettant ainsi d'améliorer la satisfaction de ces derniers. Par conséquent, les résultats concernant ce champ de recherche en analyse d'opinion sont très prisés par les professionnels des domaines de la consommation.

Depuis 2008, Patrick Saint-Dizier, membre de l'équipe Argumentation, Décision, Raisonnement, Incertitude et Apprentissage (ADRIA) de l'Institut de Recherche en Informatique de Toulouse (IRIT) et ses collaborateurs développent TextCoop, un modèle et outil d'analyse de discours modulable, à l'origine de plusieurs applications. L'une d'entre elles est l'analyse d'opinion de consommateurs francophones sur des produits tels que des restaurants ou des hôtels, les textes d'opinion considérés dans ces domaines étant relativement simples d'un point de vue lexical et argumentatif. Une ébauche a ensuite été faite concernant le domaine politique, qui s'est avéré extrêmement délicat à traiter, mettant en jeu des structures de discours complexes comme l'ironie ou la métaphore et nécessitant un grand nombre de références externes au texte étudié pour pouvoir être interprétées.

Ce document constitue le rapport de mon stage de troisième année de licence d'informatique à l'École Normale Supérieure de Lyon, réalisé au sein de l'IRIT et encadré par Patrick Saint-Dizier. L'objectif de ce stage est de détecter des systèmes de valeurs et des catégories de consommateurs en analyse d'opinion de critiques amateurs anglophones de cinéma, en adaptant le modèle et outil TextCoop au cadre considéré.

Ce rapport présente dans un premier temps la démarche de constitution du corpus d'étude que nous avons effectuée durant le stage ; puis, après définition de quelques outils d'analyse lexicale existants, nous déterminons et utilisons des systèmes de valeurs propres aux auteurs du corpus précédemment constitué. Enfin, nous établissons des modèles d'analyse sémantique, qui pourraient permettre de catégoriser les consommateurs.

1 Corpus

L'analyse d'opinion se modélise sur la base de données empiriques. En Traitement Automatique des Langues Naturels (TALN), la recherche se construit à partir de corpus composés de textes existants et non de l'intuition que l'on peut avoir de la langue, ce qui permet de ne pas caractériser uniquement les constructions des éléments de discours du chercheur. La démarche adaptée au travail avec TextCoop nécessite deux corpus : un corpus de développement et un corpus de test, qui n'a pas été utilisé ici. Le corpus de développement est analysé manuellement et étudié en profondeur, afin de définir un modèle et d'adapter l'outil TextCoop de manière à obtenir une reconnaissance la plus fidèle possible pour ces textes bien construits. Le corpus test permet, lui, de tester la pertinence de l'adaptation effectuée, sur des textes en quantité généralement plus importante et de qualité moindre, en comparant les résultats obtenus par analyse manuelle avec ceux obtenus par l'outil.

Le corpus sur lequel s'appuie le travail effectué a été constitué au cours du stage. Patrick Saint-Dizier et ses collaborateurs ont travaillé sur des opinions de consommateurs francophones dans les domaines de l'hôtellerie et de la restauration (relativement simple), puis de la politique (relativement compliqué). J'ai donc choisi de concentrer mon étude sur un corpus anglophone, afin d'analyser les structures linguistiques d'une langue différente et d'obtenir une plus grande diversité d'utilisateurs. Les critiques de cinéma constituent un domaine dont la difficulté se situe entre celles du domaine de l'hôtellerie et de la politique : il y a des structures linguistiques complexes, mais elles ne nécessitent pas forcément un grand nombre de références ou de connaissances externes pour être comprises (voir critères de constitution du corpus ci-dessous). Les critiques composant le corpus ont été sélectionnées sur le site internet IMDb (Internet Movie Database) selon une démarche descendante : choix des catégories de films, puis des films, puis des textes, puis des fragments de textes effectivement utilisables.

1.1 Démarche de travail

Internet Movie Database (IMDb : <http://www.imdb.com/>) est un site internet anglophone répertoriant un grand nombre d'informations sur la production multimédia mondiale. Il s'agit de la plus grande base de données collaborative du monde dans ce domaine avec des informations sur 3 250 000 œuvres visuelles. IMDb offre à ses utilisateurs la possibilité de noter et de commenter les films répertoriés, créant ainsi une base de données de critiques de films faites par des amateurs. La catégorisation proposée par le site est relativement fine, allant des comédies (107 893 films répertoriés) aux films de guerre (7 422 films répertoriés).

Voir Annexe 1 pour une catégorisation complète.

Quatre catégories ont été sélectionnées pour constituer le corpus :

- (a) Films familiaux
- (b) Films de science-fiction
- (c) Films d'animation
- (d) Musicals

En effet, on prend pour hypothèse que les catégories (a) et (d) sont sujettes à controverse : les films sont discutés et les avis des spectateurs sont tranchés sur chaque film de ces catégories, soit le consommateur adhère au genre, soit non.

Les catégories (b) et (c) ont supposément des critères évaluatifs proches (aspects visuels, écriture), car ces films dépaysent et demandent à l'utilisateur de se laisser convaincre par l'univers non réaliste (dessiné, futuriste) qui lui est présenté.

La sélection des films au sein des catégories utilise les fonctionnalités du site. Tout d'abord, les films sont triés par nombre de votes; ensuite, le choix se fait selon la valeur des notes moyennes, le but étant d'obtenir un échantillon suffisamment diversifié. Les moyennes ne doivent pas être supérieures à 8/10 : trop de bonnes notes implique une faible quantité de critiques négatives; les moyennes ne doivent pas être inférieures à 4/10 : trop de notes basses implique une faible quantité de critiques positives. Les films choisis doivent donc avoir des notes se rapprochant le plus de 8 (par valeurs inférieures), de 6, et de 4 (par valeurs supérieures).

De plus, les critères suivants ont été pris en compte :

- Pas de films de grands studios d'animation (ex : Disney, Pixar, Studio Ghibli).
- Pas de films de réalisateurs connus du grand public (ex : Tim Burton, Christopher Nolan).
- Pas de suites.
- Pas de reprises (remakes).
- Pas d'adaptations de livres connus.

Ces critères ont pour but d'éliminer un maximum de films « à arguments d'autorité », afin d'éviter des textes d'opinion trop référencés nécessitant des connaissances externes pour pouvoir être analysés. D'autres contraintes pourraient bien sûr être considérées.

L'arrêt de la recherche des films au sein d'une catégorie donnée se fait une fois les trois films (à 8, 6 et 4 de moyenne) trouvés, ou lorsque l'on atteint un nombre de votants pour lequel le nombre de critiques disponibles est insuffisant pour construire un échantillon significatif (nombre de critiques inférieur à 50).

On obtient ainsi la sélection de films suivante :

Catégorie	Titre du film	Année	Moyenne	Nombre votants
Famille	How the Grinch Stole Christmas	2000	6.0/10	124 151
Famille	The Iron Giant	1999	8.0/10	113 251
Famille	Super Mario Bros.	1993	4.0/10	34 784
S-F	District 9	2009	8.0/10	482 092
S-F	Iron Sky	2012	6.0/10	74 421
S-F	Plan 9 from Outer Space	1959	4.0/10	28 333
Animation	Shark Tale	2004	6.0/10	118 151
Animation	Pink Floyd The Wall	1982	8.0/10	54 847
Animation	The Adventures of Rocky & Bullwinkle	2000	4.1/10	15 739
Musicals	The Sound of Music	1965	8.0/10	130 385
Musicals	Attac of the Killer Tomatoes!	1978	4.6/10	8 978
Musicals	A Chorus Line	2014	6.0/10	7 070

Pour les films d'animation : arrêt de la recherche après avoir atteint 8 348 votants, 11 critiques. Pour les musicals : arrêt de la recherche après avoir atteint 4 583 votants, 44 critiques.

Lors de la lecture des critiques, IMDb dispose d'une fonctionnalité permettant de filtrer celles contenant des références explicites à l'intrigue du film. Le choix des critiques se fait donc en activant cette fonctionnalité et en triant par nombre de votants ; en effet, les utilisateurs peuvent voter pour les critiques des autres, les qualifiant d'utiles ou non. Ensuite, il s'agit d'obtenir une certaine diversité dans la sélection : des notes variées, des opinions divergentes. Les critiques trop courtes présentent un intérêt argumentatif limité ; les trop longues font souvent référence à des arguments externes imprévisibles.

Le corpus sur lequel a été effectué le travail présenté dans ce rapport est finalement de taille très modeste mais suffisante pour apprendre à travailler dessus. Vis-à-vis de TextCoop, il ne s'agit donc que du corpus moteur, dont la validation demande un travail technique très complexe. Une suite à l'étude menée nécessiterait d'établir un corpus bien plus important, et donc d'y consacrer un temps bien plus conséquent. Cela n'a pas été fait, car le but n'était pas la création d'un logiciel.

Au moment du choix et de la lecture plus précise des critiques, des obstacles imprévus ont surgi. Outre de simples problèmes d'orthographe, la difficulté principale était une subjectivité propre à chaque auteur (ex : "I saw this movie with my children"). Ainsi, peu de critiques pouvaient être sélectionnées, et parmi celles qui pouvaient l'être, tout ne pouvait pas être analysé. Il a donc fallu trier les textes comme leur contenu.

1.2 Analyse linguistique et conceptuelle

L'analyse manuelle du corpus se fait dans un premier temps par annotation des textes. En effet, dans la plupart des textes d'opinion, le discours s'articule autour de *paires attribut(a)/valeurs(e.e)* : à un attribut (une caractéristique du produit) sont associées des expressions évaluatives (approbation, désapprobation, nuances...). Ces données lexicales doivent faire l'objet d'une recherche minutieuse car elles sont propres à chaque domaine.

Exemple 1. *"Every aspect, from the cinematography_a to the acting_a, the sets_a to the costumes_a, the music_a, choreography_a, script_a, is top notch_{e.e}."*

L'analyse manuelle du corpus permet de mettre en place des catégories de systèmes de valeurs ; par analyse lexicale, on trouve les *attributs* et les *expressions évaluatives* majoritaires dans les textes.

Les attributs peuvent être réalisés linguistiquement de manières différentes. On les regroupe donc en catégories sémantiques, afin de pouvoir faire des observations significatives. On attribue de plus à chaque expression évaluative une note entre 1 (évaluation très négative) et 5 (évaluation très positive), permettant de quantifier la qualification faite de l'attribut : dans l'exemple précédent, "top notch" correspond à une note de 5.

Attribut	Expression(s) de l'attribut	Expressions évaluatives(note)
film	movie	triumph in all departments(5), family friendly(4), family-oriented(4), amiable(4)
script	story, humor, romance, suspense, action, storyline	sweet(4), genuine(4), exciting(5), odd(2)
acting	performance, voice	brilliant(5), good(4), incredible(5), too staged(2), distressingly one-note(1)
sets	costumes, location photography, cinematography, design, camera work, scenery	beautiful(4), breathtaking(5)
music	songs	nice(4), playful(4), fun(4), simplistic(2)

Il existe des expressions évaluatives généralistes, applicables à tous les attributs : "top notch"(5), "wonderful"(5). L'attribut "film" n'a pas le même statut que les quatre autres. Il correspond à l'impression générale du spectateur et sert souvent à donner la recommandation, conclusion du texte d'opinion. Il est réalisé de la même manière au niveau de l'analyse lexicale, mais est à traiter de manière différente lors d'une analyse argumentative.

Il ne s'agit pas, ici, de fournir un relevé exhaustif d'éléments de langage. On peut cependant observer, relativement à la taille modeste du corpus, un déséquilibre entre la quantité d'expressions évaluatives positives (note > 3) et négatives (note < 3). En effet, il semblerait qu'une manière commune pour un spectateur d'exprimer sa non-satisfaction par un film est d'énumérer tout ce qu'il aurait souhaité y retrouver, en appliquant la négation.

Exemple 2. *"I don't think the lyrics come close to some of the Oscar Hammerstein classics, the score is not among Rodgers' best, for me at least. I particularly don't like the opening[...]."*

2 Analyse lexicale et syntaxique

L'annotation manuelle du corpus permet de mener à bien l'analyse lexicale, en construisant le lexique à partir des données relevées. On cherche à présent à effectuer une analyse syntaxique, repérer les constructions au sein du discours. Il existe plusieurs manières d'y parvenir. Par choix, le travail présenté ci-dessous s'appuie sur ce qui a été fait dans TextCoop. La théorie \bar{X} permet de mieux comprendre l'articulation des attributs et des expressions évaluatives, par ce qu'elle a de générique, régulier et systématique, ce qui rend son utilisation pratique dans le cadre d'une modélisation.

2.1 La théorie \bar{X}

Définition 1. *Les catégories lexicales principales sont les noms, les verbes, les adjectifs et les prépositions, notées respectivement N, V, A et P.*

Définition 2. *Un syntagme est un groupe de mots formant une unité autonome par son sens, sa fonction et sa structure.*

Exemple 3. *"the joyful melody" et "quite loud" sont des syntagmes.*

Définition 3. Le noyau d'un syntagme est le mot principal donnant toutes les contraintes devant être rencontrées dans le syntagme.

Définition 4. Un syntagme nominal (resp. verbal, adjectival, prépositionnel) est un syntagme dont le noyau est un nom (resp. verbe, adjectif, préposition). Le syntagme est appelé projection du noyau.

Exemple 4. Dans la phrase "The melody was quite loud before silence fell.", "quite loud" est un syntagme adjectival.

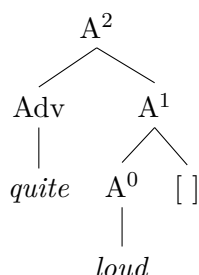
Proposée par Noam Chomsky en 1970 [Haegeman, 1991], la théorie \bar{X} permet de mettre en évidence les similitudes structurelles des syntagmes. X et Y sont des variables pouvant être évaluées à N, V, A ou P.

De manière générale, on distingue deux niveaux de projection du noyau (noté X^0) : celui du syntagme, noté X^2 , est considéré lié au noyau par un niveau intermédiaire, noté X^1 [Sells, 1985]. Le format général d'un syntagme dans la théorie \bar{X} peut ainsi être résumé par les règles suivantes [Haegeman, 1991] :

$$\left\{ \begin{array}{l} X^2 \rightarrow Spec, X^1 \\ X^2 \rightarrow Y^2, X^1 \\ X^1 \rightarrow X^0, Y^2 \\ X^1 \rightarrow X^1, Y^2 \end{array} \right.$$

où *Spec* désigne le spécificateur (voir tableau ci-dessous) et la virgule marque la non-rigidité de l'ordre des éléments (X^2 peut se dériver en *Spec*, X^1 comme en X^1 , *Spec*). A l'aide de ces règles, il est possible de représenter les syntagmes par des arbres de dérivation.

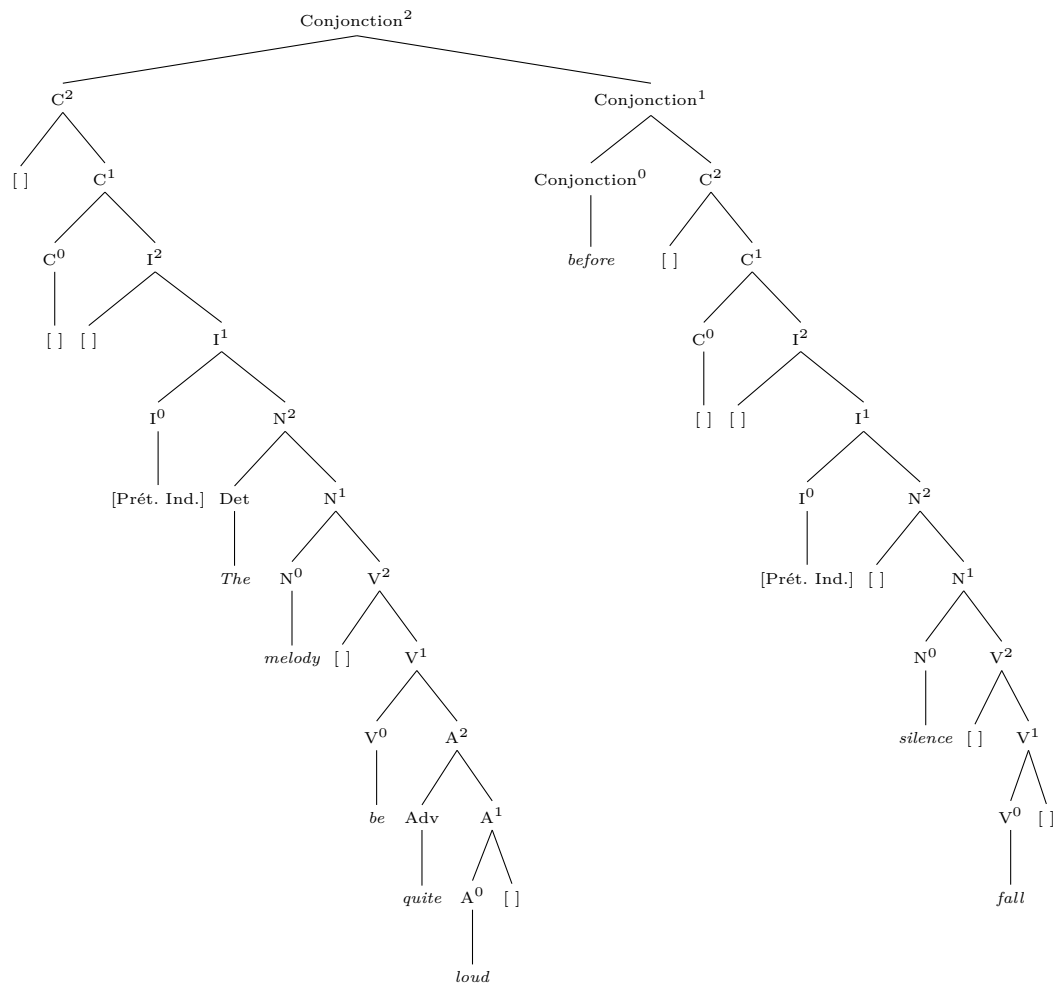
Exemple 5. "quite loud" est ainsi représenté par l'arbre suivant :



Afin de pouvoir représenter la structure syntaxique des phrases, il est nécessaire d'introduire des catégories non-lexicales. On définit donc l'inflexion (I) et le complémentariseur (C). Ce sont des catégories permettant d'articuler les syntagmes au sein des phrases. La conjonction permet la juxtaposition des syntagmes; le complémentariseur introduit les propositions (en anglais, C^0 correspond à "whether", "that", "for", "it"), et l'inflexion est responsable de la conjugaison (morphologie et aspect) des verbes au sein des syntagmes modaux. On peut ainsi résumer l'utilisation de la théorie \bar{X} à l'aide du tableau suivant :

X	Spec	Y ²
N	Det	C ²
A	Adv	P ²
V	Aux/I	N ² /P ² /A ²
P	Adv	N ²
I	N ²	V ²
C	Pronom interrogatif et relatif	I ²

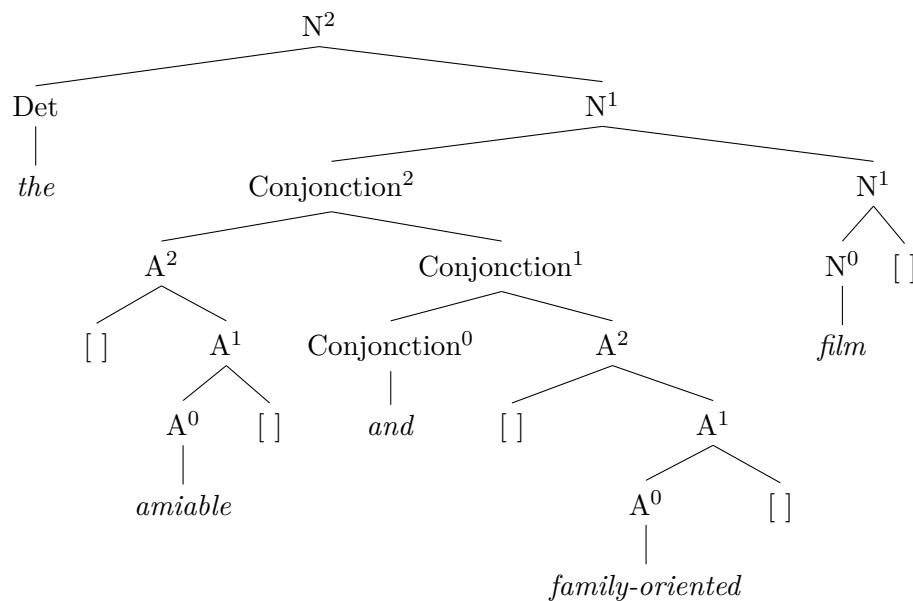
Exemple 6. La phrase "*The melody was quite loud before silence fell.*" est représentée par l'arbre suivant :



2.2 Patrons et grammaires

TextCoop annoté les textes en les parcourant, à l'aide de grammaires qui lui sont connues (grammaires formalisées de la langue anglaise). Une fois un outil de représentation des phrases mis en place, on peut tenter de l'abstraire afin d'obtenir des patrons de lecture propres au corpus. On pourra alors les implémenter. Pour cela, on procède par lecture des feuilles des arbres \bar{X} .

Exemple 7. "the amiable and family-oriented film" est représenté par l'arbre ci-dessous :



et correspond au patron suivant :
[valeur, ..., attribut].

2.3 TextCoop

TextCoop est un environnement dont le principe repose sur la programmation logique. Il sert à décrire et à implémenter les structures de discours (voir ci-dessus) et les contraintes y étant liées. Il utilise pour cela entre autres le langage Dislog (Discourse in Logic), créé pour décrire les règles de grammaire et les données lexicales. Dislog est une extension du formalisme des grammaires à clauses définies (DCG) [Saint-Dizier, 1994] à l'analyse de discours, et permet d'y intégrer des connaissances et des déductions.

De plus, Dislog possède sa propre stratégie d'analyse de discours, qui n'est pas la même que celle de Prolog, avec un moteur spécifique et une analyse dédiés au discours. Par conséquent, une fois les patrons définis, il ne faut pas spécifier la stratégie, puisque l'ordre de l'exécution est forcé. TextCoop présente donc un moteur universel, muni de règles génériques, relativement simples et basées sur la langue, suivies de règles génériques ou spécifiques propres au domaine considéré.

3 Analyse sémantique et pragmatique

Une fois des catégories de systèmes de valeurs internes aux critiques déterminées, on s'interroge sur les auteurs. On aimerait pouvoir préciser ces catégories en profilant les utilisateurs en fonction de leur manière d'argumenter, identifiant les attributs importants pour chacun et tirant d'éventuelles conclusions sur leurs systèmes de valeurs. Pour cela, après avoir étudié des modèles existants pour l'argumentation au sein de dialogues [Amgoud, 2009], nous avons mis en place un modèle simplifié correspondant au problème posé dans le contexte du corpus considéré.

3.1 Arguments et argumentation

Définition 5. Un argument est une proposition introduite par un connecteur, destinée à appuyer, attaquer ou nuancer un propos.

L'étude de l'argumentation est une tâche particulièrement complexe du fait de la grande diversité des structures de discours, de la forme que celles-ci prennent dans la langue et surtout des connaissances externes potentiellement nécessaires à leur identification. En effet, le contexte est essentiel pour pouvoir affirmer qu'un groupe de mots constitue un argument et il l'est plus encore pour pouvoir établir des liens entre les arguments déjà reconnus.

Afin de pouvoir mener à bien l'étude de l'argumentation du corpus, nous aimerions établir un modèle à trois niveaux : tout d'abord on modélise au sein de chaque texte critique les affirmations et les arguments par des unités attribut-valeur-argument ; ensuite, il devient possible d'articuler les unités et ainsi de visualiser l'argumentation au sein d'un texte ; enfin, on peut émettre des hypothèses concernant les systèmes de valeurs propres aux utilisateurs.

3.2 Modélisation linguistique du système d'argumentation

Les constructions linguistiques des systèmes d'argumentation peuvent être extrêmement compliquées et les attributs comme les expressions évaluatives ne sont pas nécessairement explicites. On peut modéliser au sein de chaque texte critique les affirmations et les arguments par des unités attribut-valeur-argument. Il peut y avoir permutation des constituants (Exemple 9), présence de coordinations (Exemple 10), et l'adjacence n'est pas nécessaire (Exemple 11). Nous distinguons trois schémas principaux pour ces unités (+ étant un ou plus, * étant aucun ou plus) :

Définition 6. *1^{er} schéma* : l'argument sert à préciser le propos en l'appuyant ou en le nuancant, il n'est donc pas nécessaire à l'affirmation : [attribut, expression évaluative, argument*].

Exemple 8. "*stunning_{e.e} visuals_a*" se modélise par [visuals, stunning, \emptyset].

Définition 7. *2nd schéma* : l'expression évaluative se déduit à partir de l'attribut et d'un ou plusieurs arguments : [attribut, \emptyset , argument⁺].

Exemple 9. "*the film_a brings nostalgic thoughts and reminds one of Fellini*" se modélise par [film, good, [brings nostalgic thoughts, reminds one of Fellini]], "good" étant l'expression évaluative déduite.

Définition 8. *3^{ème} schéma* : l'attribut se déduit à partir d'une expression évaluative et d'un ou plusieurs arguments : [\emptyset , expression évaluative, argument⁺].

Exemple 10. "*Too staged_{e.e} but still impressive*." se modélise par [performance, too staged, [still impressive]], "performance" étant l'attribut déduit.

Si plusieurs attributs ou plusieurs expressions évaluatives sont présents dans une même affirmation, une solution est de dupliquer la structure des unités autant de fois que nécessaire.

Un texte argumentatif est composé d'une recommandation et d'une séquence d'unités attribut-valeur-argument. La recommandation est une unité attribut-valeur, de la forme [film, expression évaluative, \emptyset]. Les unités au sein d'une critique sont liées à la recommandation par une relation d'attaque ou de support. De la même manière, les arguments sont liés aux unités auxquelles ils appartiennent. On définit ainsi le modèle qui suit, induit à partir des lectures sur le cadre de Dung pour les systèmes d'argumentation abstraits [Amgoud, 2009].

Définition 9. Soit \mathcal{A} l'ensemble des arguments du texte, \mathcal{U} l'ensemble des unités attribut-valeur (séparés des arguments). On définit $\mathcal{R}_1 \subset \mathcal{A} \times \mathcal{U} \times \{\text{attaque, support}\}$, relation entre arguments et unités attribut-valeur.

Il est parfois difficile d'identifier un argument comme supportant ou attaquant. Il s'agit du *problème de connaissance* : pour identifier le positionnement d'un argument dans le texte d'opinion, il faut des connaissances externes, des données lexicales et sémantiques, ainsi que des schémas argumentatifs adaptés. [Saint-Dizier, 2015].

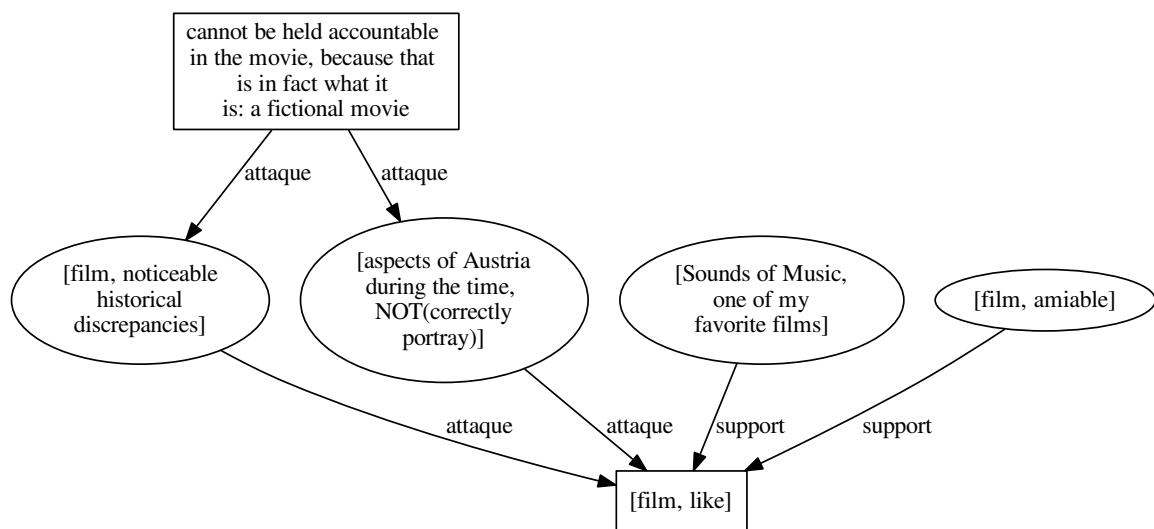
Exemple 11. Dans l'extrait "*While the historical discrepancies throughout the film are noticeable, and it does not correctly portray aspects of Austria during that time – that cannot be held accountable in the movie, because that is in fact what it is: a fictional movie.*", on a $\mathcal{R}_1 = \{("cannot\ be\ held\ accountable\ in\ the\ movie,\ because\ that\ is\ in\ fact\ what\ it\ is:\ a\ fictional\ movie", [film, noticeable\ historical\ discrepancies], \mathbf{attaque}), ("cannot\ be\ held\ accountable\ in\ the\ movie,\ because\ that\ is\ in\ fact\ what\ it\ is:\ a\ fictional\ movie", [aspects\ of\ Austria\ during\ that\ time, NOT(correctly\ portray)], \mathbf{attaque})\}$

Définition 10. Soit ρ la recommandation d'un texte argumentatif. On définit $\mathcal{R}_2 \subset \mathcal{U} \times \{\rho\} \times \{attaque, support\}$, relation entre les unités attribut-valeur et la recommandation.

Il est désormais possible de représenter les textes par des graphes orientés étiquetés.

Exemple 12. "*The Sound of Music has been, and probably always will be, one of my favorite films. While the historical discrepancies throughout the film are noticeable, and it does not correctly portray aspects of Austria during that time – that cannot be held accountable in the movie, because that is in fact what it is: a fictional movie. One cannot help but love the amiable and family-oriented film.[...]*"

est représenté par le graphe orienté étiqueté suivant :



On s'intéresse désormais non seulement aux intérêts et à la manière d'argumenter de l'utilisateur, mais aussi à la hiérarchie de ses intérêts : quel est le critère qui compte le plus ?

Afin de commencer à répondre à cette question, on représente chaque texte par un tableau contenant les unités attribut-valeur-argument classées par attribut, la recommandation étant une unité attribut-valeur-argument d'attribut "film" et d'argument vide. On calcule ensuite la moyenne obtenue pour chaque attribut ($m(attribut)$), puis la moyenne de ces moyennes (M), hors attribut "film". Un argument compte pour + ou - 1 par rapport à la note attribuée à l'unité à laquelle il se rapporte.

Une fois le calcul effectué, il devient possible de comparer les résultats.

- Si $M = m(\text{film})$, alors tous les attributs sont d'importance égale pour le spectateur considéré.
- Si $M > m(\text{film})$, alors les attributs de moyenne faible sont d'importance plus grande que ceux de moyenne élevée pour le spectateur considéré.
- Si $M < m(\text{film})$, alors les attributs de moyenne élevée sont d'importance plus grande que ceux de moyenne faible pour le spectateur considéré.

Exemple 13. *"The Sound of Music has been, and probably always will be, one of my favorite films. While the historical discrepancies throughout the film are noticeable, and it does not correctly portray aspects of Austria during the time – that cannot be held accountable in the movie, because that is in fact what it is: a fictional movie. One cannot help but love the amiable and family-oriented film. Julie Andrews has a wonderful voice, and portrays the role of Maria as the optimistic nun who become nanny and then mother to the Von Trapp family well. Plummer is not particularly my favorite character, but he also portrays the captain in a decent manner. Wise is an experienced director and his experience is seen through the popularity and wonderful clips and action of the film. The songs of the movie are playful and fun to sing along with, and easily get stuck in your head. Overall I like the film and would definitely consider it well deserving of all its awards and nominations."*

est représenté par le tableau suivant :

Attribut	Unité	Note	m
film	[film, like, \emptyset]	4	4
	[Sounds of Music, one of my favorite films, \emptyset]	4	
	[film, amiable, \emptyset]	4	
	[film, family-oriented, \emptyset]	4	
	[film, popularity of, \emptyset]	4	
script	[\emptyset ,noticeable historical discrepancies,[cannot be held accountable in the movie, because that is in fact what it is: a fictional movie]]	2+1	3.7
	[aspects of Austria during that time, NOT(correctly portray), [cannot be held accountable in the movie, because that is in fact what it is: a fictional movie]]	2+1	
	[action, wonderful, \emptyset]	5	
	[clips, wonderful, \emptyset]	5	
sets	[clips, wonderful, \emptyset]	5	5
acting	[portrays the role, well, \emptyset]	4	3.5
	[portrays the captain, in a decent manner, \emptyset]	3	
music	[voice, wonderful, \emptyset]	5	4.3
	[songs, playful, \emptyset]	4	
	[songs, fun, \emptyset]	4	

$$M = \frac{3.7 + 5 + 3.5 + 4.3}{4} \approx 4.1 > 4 = m(\text{film})$$

Pour l'utilisateur ayant écrit cette critique, le jeu d'acteur ("acting") est un critère plus important que les aspects visuels ("sets").

Ces résultats ne sont qu'une première approximation des systèmes de valeurs propres à chaque spectateur. En effet, on pourrait penser à un modèle plus fin appliquant des pondérations aux différents critères et les classant ainsi par degré d'importance pour l'utilisateur.

Conclusion

Une fois le court corpus de développement constitué, des questions concrètes ont pu être formulées concernant les systèmes de valeurs et les catégories de spectateurs non-professionnels. L'analyse lexicale et argumentative du corpus a permis d'adapter le code de TextCoop au problème posé (voir Annexe 3), tandis qu'un travail sur les théories et les modèles existants a abouti à une modélisation plus formelle des systèmes de valeurs des utilisateurs. L'utilisation de cette modélisation suggère des directions de recherches futures, notamment concernant l'importance relative des critères pour chaque spectateur. Il serait de plus intéressant de poursuivre le travail avec TextCoop et un corpus de test, afin d'obtenir des résultats permettant de juger de manière concrète la pertinence de l'adaptation effectuée. Enfin, la question des catégories de consommateurs reste ouverte, puisque la taille du corpus et la quantité d'informations disponibles concernant les profils des utilisateurs n'étaient pas entièrement suffisantes pour permettre d'ébaucher un début de réponse.

Références

- [Amgoud, 2009] L.Amgoud, *Contribution to argumentation theory and its applications*, Mémoire d'Habilitation à Diriger des Recherches, Université Paul Sabatier, 2009.
- [Garcia Villalba, Saint-Dizier, 2013] M.Garcia Villalba, P.Saint-Dizier, *Argument Extraction in Opinion Analysis: Identifying the Reasons behind Consumer Evaluations*, Knowledge Discovery, Transfer, and Management in the Information Age, IGI Global, 2013.
- [Haegeman, 1991] L.Haegeman, *Introduction to Government & Binding Theory*, Basil Blackwell, 1991.
- [Plantin, 1996] C.Plantin, *L'argumentation*, Editions du Seuil, 1996.
- [Saint-Dizier, 1989] P.Saint-Dizier, *An Introduction to Programming in Prolog*, Springer-Verlag, 1989.
- [Saint-Dizier, 1994] P.Saint-Dizier, *Advanced Logic Programming for Language Processing*, Academic Press, 1994.
- [Saint-Dizier, 2015] P.Saint-Dizier, *An Exploration of the Relatedness Problem between Arguments: Combining the Generative Lexicon with Lexical Inference*, European Conference on Argumentation, Lisbonne, 2015.
- [Sells, 1985] P.Sells, *Lectures on Contemporary Syntactic Theories*, CSLI Lecture Notes, 1985.
- [Walton, Reed, Macagno, 2008] D.Walton, C.Reed, F.Macagno, *Argumentation Schemes*, Cambridge University Press, 2008.

Annexes

Annexe 1 : catégories de films proposées par IMDb (données du 11 juillet 2015)

- Drames (165 729 films répertoriés)
- Comédies (107 893 films répertoriés)
- Films romantiques (38 066 films répertoriés)
- Films d'action (34 918 films répertoriés)
- Thrillers (32 599 films répertoriés)
- Films policiers (27 581 films répertoriés)
- Films familiaux (22 106 films répertoriés)
- Films d'aventure (20 168 films répertoriés)
- Films d'horreur (20 068 films répertoriés)
- Films musicaux (16 140 films répertoriés)
- Films à énigmes (14 042 films répertoriés)
- Documentaires (12 471 films répertoriés)
- Films de science-fiction (12 304 films répertoriés)
- Films de fantasy (12 241 films répertoriés)
- Films d'animation (11 862 films répertoriés)
- Télé-réalités (11 646 films répertoriés)
- Musicals (10 577 films répertoriés)
- Débats télévisés (8 361 films répertoriés)
- Westerns (8 263 films répertoriés)
- Films de sport (7 721 films répertoriés)
- Films historiques (7 438 films répertoriés)
- Films de guerre (7 422 films répertoriés)
- Biographies (6 239 films répertoriés)
- Jeux télévisés (4 720 films répertoriés)
- Actualités (4 313 films répertoriés)
- Films Noirs (617 films répertoriés)
- Films pour adultes (34 films répertoriés)

Annexe 2 : extraits du corpus : critiques de *The Sound of Music* (*La mélodie du bonheur*)

Synopsis : L'action se déroule peu avant le déclenchement de la Seconde Guerre mondiale en Autriche, au moment de l'Anschluss. Maria est une jeune femme qui se prépare à devenir religieuse dans une abbaye bénédictine située à Salzbourg. Son couvent l'envoie en tant que gouvernante dans une famille de sept enfants, celle du capitaine Georg Ritter von Trapp, veuf. Au début, les enfants se montrent plutôt hostiles. Elle tient bon, tout en leur enseignant le chant, et bientôt gagne leur confiance.

Le capitaine fréquente une baronne dans le but de trouver une mère pour ses enfants. L'arrivée de la gouvernante l'amène à revoir ses sentiments vis-à-vis de la baronne. Celle-ci ne voit pas cette situation d'un bon oeil et incite Maria à quitter la famille. De retour au couvent, après quelques jours de prière, l'abbesse lui fera comprendre que l'amour terrestre ne peut être éteint par la prière. Les leçons de chant de Maria, quant à elles, ont tellement porté leurs fruits que l'oncle a inscrit la famille Trapp à un concert de chant qui se tient à Salzbourg. Pendant ce temps, les nazis prennent le pouvoir en Autriche.

De retour de voyage de noces avec Maria, le capitaine apprend qu'il doit reprendre le service dans la marine, mais refuse de participer à la guerre. La famille tente de fuir le soir avant le concert, mais est surprise par les nazis, qui acceptent de les laisser chanter. À la fin du spectacle, bien que surveillés, les von Trapp parviennent à se cacher dans l'abbaye et à s'enfuir vers la Suisse.

CRITIQUE 1 :

This film is a triumph in all departments. Every aspect, from the cinematography to the acting, the sets to the costumes, the music, choreography, script, is top notch. While the film is family friendly and has a sweet story, it is constantly amazing the way people attack it as saccharine and sugary. Ernest Lehman inserted many moments of wit, humor, romance and poignancy. The art directors purposefully chose muted settings and colors.

Each of the actors bent over backwards to provide a brilliant performance. Andrews is already down in history for the performance of a lifetime (and a voice to match), but Plummer is not to be forgotten. Not only is he regal and handsome, but his decision to play the Captain as a complex, sophisticated man with a sly dose of sarcasm was wonderful. His steely, stern persona is eventually melted down by the irrepressible Andrews to great effect. Every supporting performance is also delivered with the right amount of appeal, humor or menace as called for in the script.

Baroness Schraeder's songs, wisely, were cut, further separating her from all the glee around her, so that she could whip out such zingers as "Why didn't you tell me...to bring along my harmonica?" or when she's told that Andrews may not make a great nun, "If you need anything, I'd be happy to help you." The character is given a polished and integral position in the film and virtually every line of her dialogue is a howler. Though Wood was lovely in her role as the Mother Abbess, it was Parker who should have gotten an Oscar nod...and won!

This type of film-making is gone. The location photography, the simplicity of story and design, the sheer good-spiritedness of it all...they just can't do this anymore. Thankfully, there's this flawless gem to turn to when one just want to feel good. But saccharine? No...compare this to other beloved musicals with their garish colors and sugary story lines ("Seven Brides...", "Singin' in the Rain", "...Molly Brown", "The Music Man", to name just a few). They are all highly enjoyable, but are hardly less sweet than this! Just one word...Nazis!! Though virtually everyone knows the outcome, there is still genuine suspense at the climax of "The Sound of Music". The film has it all.

CRITIQUE 2 :

Sometimes there's a cultural institution about which you just don't see the big fuss, and for me, it's this film. I got dragged to it at least twice as an 8-year-old, didn't really like it then, and still don't get it. I don't think the lyrics come close to some of the Oscar Hammerstein classics, the score is not among Rodgers' best, for me at least. I particularly don't like the opening, the booming, over-orchestrated title song. I much prefer Julie Andrews overall in "Mary Poppins," and vocally, by far, in "My Fair Lady." If Christopher Plummer hated being in this movie, I think it shows. Having said all that, millions of people – including my late mother, who rarely went to any movies, yet saw this at least three times, I think – adore the film. So I guess I'm in the minority, but still, I have no desire to ever see this again.

CRITIQUE 3 :

This movie is the kind that makes people jump up and shout at the screen. Me, too, things like : "Shut up, already!" and "Look behind the tombstone, you silly Nazis!". I guess when it comes to "The Sound Of Music," you either have it, or you've had it. Oh, the film starts out strong, on one of the biggest shots in movies, and it has some beautiful camera work to go along with some nice if somewhat simplistic songs by Rodgers and Hammerstein. But if Julie Andrews' other best-known musical, from 1964, involved a spoonful of sugar, this 1965 release feels more like a truckful of treacle, especially in the second half when the story gets ropier, the songs dodgier, and the acting too staged by half.

Except for her singing where she shines, Andrews gives a distressingly one-note performance as a postulant nun, Maria, sent on sabbatical to look after the seven rambunctious children of a stern naval captain in Salzburg, Austria. In time, a remarkably condensed time, she has converted the children into a happy singing group and Captain von Trapp into a sappy velveteen Romeo. But will they find freedom as one big happy family, or else be crushed by the Nazi jackboot? I guess that's supposed to be the drama in this piece, but it's pushed together so haphazardly and drawn out so much over the course of three hours that it never feels in the least real, with or without the fruity singing. Maria sings about bluebirds and sleigh bells and whiskers on kittens, the kids perform ornate puppet shows and sing their way to bed before a party of admiring guests, and even Capt. von Trapp joins in, singing "Edelweiss" to an audience of Nazis so solemnly he puts them all to sleep so his family can escape.

There are good numbers, namely the opening one with Andrews performing the title song on a mountain top in a nicely low-key hands-in-pockets way, and the best song in the score which follows that, "How Do You Solve A Problem Like Maria?" well-staged in an abbey courtyard as a dialogue amongst several nuns. One, Sister Sophia, brings an Audrey Hepburn-like quality to her singing. With director Robert Wise's steady hand and cinematographer Ted McCord's nice attention to the stained-glass and groined-ceiling decor, the sequence really sticks with you and is enjoyable to see over again. But Wise makes some odd decisions with other numbers that follow, especially the icky "Edelweiss" and "16 Going On 17," a nice song undone by campy staging. "Climb Ev'ry Mountain," a showstopper on stage, is performed by the Mother Abbess with a terribly trilly overdub and in such dark shadow you'd think she was a mob informant for the FBI.

Finally, the movie needed to end with the wedding, which while saccharine, is at least a pretty high note in keeping with what came before. The Nazi chase that follows is just silly, an attempt to add some movie drama that comes off as sour bubble gum. Lots and lots of people don't agree with me and have made "The Sound Of Music" a cultural touchstone.

CRITIQUE 4 :

I first watched this movie when I was about 5 or 6 years old and didn't think anything of it but now that I am 14 years old and am old enough to appreciate and understand the storyline I am in love with everything about it : the breathtaking scenery of Austria, the exciting storyline and most of all, the classic songs by Rodgers and Hammerstein.

Has Julie Andrews ever been captured more beautifully? As a huge fan of hers I love the way she captivates the audience with her natural beauty and sparkle. She absolutely lights up the screen as Maria von Trapp and is fantastic in her role. She truly deserved the best actress Oscar for her performance. Her voice is incredible, one of the best I have ever heard and the last note of the "do re mi" sequence proves that. Her acting is just as good as her singing and I believe that this picture shows that she is not only a brilliant stage actress but is also a legend on the screen.

Christopher Plummer gives a slight darkness to the film as the stern Captain von Trapp which is great because it gives variety to the characters in the film and shows that not all of them are sweet and sugary. Eleanor Parker (Baroness Shoerader), Richard Haydn (Max), Peggy Wood (Mother Abbess) and all 7 children form a great supporting cast and add to the success of the movie. The Sound of Music is truly a classic movie and will be around for many more years to come. If you haven't seen this movie then I strongly recommend you do because I guarantee you that you will love it. It is unique in that it does not rely on drugs, violence, sex or bad language for its success and is just simply a brilliant masterpiece suitable for all ages. Exquisite!!

Annexe 3 : exemple d'exécution de TextCoop

Texte soumis : *"This film is amiable. Every aspect, from the cinematography to the acting, the sets to the costumes, the music, choreography, script, is wonderful. The film is wonderful and has a sweet story."*

Sortie de TextCoop : this < opinionattr = " film " val = " 4 " > film is amiable < /opinion> . every aspect , from the < opinionattr = " cinematography " val = " 5 " > cinematography to the < opinionattr = " acting " val = " 5 " > acting , the < opinionattr = " sets " val = " 5 " > sets to the < opinionattr = " costumes " val = " 5 " > costumes , the < opinionattr = " music " val = " 5 " > music , < opinionattr = " choreography " val = " 5 " > choreography , < opinionattr = " script " val = " 5 " > script , is wonderful < /opinion> < /opinion> < /opinion> < /opinion> < /opinion> . the < opinionattr = " film " val = " 5 " > film is wonderful < /opinion> and has a < opinionattr = " story " val = " 4 " > sweet story < /opinion> .