

# Extraction automatique de modificateurs de valence affective dans un texte

Etude exploratoire appliquée au cas de l'adverbe

Noémi Boubel

*Université catholique de Louvain*

## 1. Introduction et problématique générale

Notre recherche s'insère dans le domaine de la fouille d'opinion, domaine devenu particulièrement important en traitement automatique du langage depuis une dizaine d'années. Les premières tâches de fouille d'opinion (Turney 2002) cherchaient à classer des textes présentant l'avis d'un auteur sur un produit (des critiques de films, par exemple) selon qu'ils exprimaient un jugement positif ou négatif. Les nombreux travaux suivants ont progressivement élargi le champ d'application sur la base d'objectifs applicatifs et de besoins industriels précis (Pang et Lee 2008). Dans les tâches classiques de fouille d'opinion, l'attribution d'une polarité (ou d'une *valence affective*) à un texte repose essentiellement sur le nombre de termes positifs ou négatifs qu'il contient. Le contexte dans lequel apparaissent ces termes n'est pas pris en compte. Pourtant, il peut avoir un impact important (lors d'une négation par exemple). Pour pallier ce problème, Zaenen et Polanyi (2004) postulent l'existence d'éléments contextuels (*contextual valence shifters*) qui modifient la valeur initiale d'un terme de trois façons possibles : en l'intensifiant (*intensifiers*), l'atténuant (*diminishers*) ou l'inversant (*negations*). Ces notions ne sont pas développées dans les théories francophones, c'est pourquoi nous utiliserons ici les termes traduits de *modificateurs*, d'*inverseurs*, d'*atténuateurs* et d'*intensifieurs*. Récemment, certains travaux (Kennedy et Inkpen 2006 ; Musat et Trausan-Matu 2010 ; Wilson et al. 2004 ; Ding et al. 2008 ; Klenner et al. 2009) utilisent ces concepts afin d'améliorer des tâches de classification. Pour cela, ils établissent intuitivement des listes de modificateurs les plus courants. Il n'existe pas, à notre connaissance,

d'étude systématique de ces phénomènes. Dans une autre optique, certaines théories linguistiques, développées sous l'angle de la subjectivité, cherchent à modéliser de façon rigoureuse le langage évaluatif (Martin et White 2005 ; Kerbrat-Orecchioni 1997 ; Charaudeau 1992), et quelques travaux intègrent ces théories dans des tâches de fouille d'opinion (Wiebe et al. 2005 ; Daille et al. 2011). Néanmoins, ces théories s'avèrent difficiles, de par leur finesse, à adapter à un traitement automatique.

Nous nous proposons d'étudier plus en détail ces phénomènes de modification de valence. Dans cette optique, nous avons précédemment développé une procédure automatique (Boubel et Bestgen 2011) qui permet de repérer, sur la base de traitements statistiques, les contextes lexico-syntaxiques qui peuvent avoir un impact sur la polarité initiale d'un terme donné. L'objectif à terme est d'obtenir automatiquement une liste de modificateurs et leur impact. Pour ce faire, la présente étude cherche à analyser quelques résultats statistiques obtenus et voir si l'on peut en déduire automatiquement l'impact d'un modificateur.

Nous présentons tout d'abord la méthodologie d'extraction utilisée ainsi que l'hypothèse de recherche sous-jacente. Nous détaillons ensuite les résultats de la procédure appliquée uniquement aux adverbes et dégageons trois types d'adverbes qui répondent à des caractéristiques statistiques communes. La question est de savoir si ces propriétés statistiques communes sont le reflet d'un comportement sémantique similaire en contexte. Pour cela, nous étudions systématiquement le rôle de ces adverbes dans le corpus de travail. Nous concluons enfin sur l'apport de cette étude sur la méthodologie que nous voulons développer et améliorer.

## **2. Méthodologie**

L'idée de départ est de s'intéresser à la distribution et au contexte de termes dont on connaît : (1) la polarité initiale, (2) la polarité du texte (ou portion de texte) dans lequel ils apparaissent. Ainsi, nous nous proposons de comparer la polarité d'un mot avec la polarité de la phrase dans laquelle il apparaît, d'identifier son contexte lexico-syntaxique (dans un premier temps, ses dépendances syntaxiques directes), et d'en conclure le rôle de ce contexte dans l'ensemble de l'énoncé. Nous pouvons faire les hypothèses suivantes : (1) lorsqu'un terme positif apparaît dans un texte négatif, le contexte auquel est associé ce terme positif devrait normalement inverser ou atténuer sa valeur, et inversement pour un terme négatif dans un énoncé positif ; (2) par contre, si un terme positif est trouvé dans

un texte positif (ou un terme négatif dans un texte négatif), son contexte aura éventuellement un effet d'intensification sur la valeur du terme ou n'aura pas d'effet particulier.

Pour appliquer cette méthodologie (Boubel et Bestgen 2011), nous utilisons un corpus composé d'extraits de critiques de films notés de 1 à 5 (de très mauvais à très bon) provenant du site *Allociné*<sup>1</sup>. Nous utilisons ensuite un lexique de termes classés selon leur valeur positive et négative, lexique créé automatiquement grâce à la méthode de Vincze et Bestgen (2011). Cette méthode se base sur un lexique constitué manuellement de 3252 mots évalués selon leur polarité (Hogenraad et al. 1995), et sur le corpus de critiques à notre disposition. Le lexique obtenu a alors une plus grande couverture lexicale et est adapté au domaine des critiques de cinéma (il met en évidence par exemple que *artificiel* et *détaché* sont souvent utilisés pour qualifier un mauvais film). En revanche, il peut comporter des erreurs (*très* classé comme positif). Enfin, grâce à un analyseur syntaxique automatique (*Xerox Incremental Parser*, Aït-Mokhtar et al. 2002), les relations de dépendance syntaxique dans lesquelles apparaissent les mots positifs et négatifs (du lexique) sont extraites. Nous obtenons donc, après ces différents traitements, une liste contenant : (1) les relations de dépendance incluant un terme fortement positif ou négatif, (2) la polarité du terme en question, (3) la note de la critique dans laquelle la relation a été extraite. Ces données sont ensuite traitées d'un point de vue statistique et chaque relation de dépendance associée à un terme positif ou négatif est analysée au moyen du test du chi-carré, qui évalue s'il y a indépendance entre la note de la critique et la fréquence de la relation. Nous retenons les relations pour lesquelles le résultat du test est significatif (seuil de 0,05) et en calculons les résidus ajustés. De cette façon, nous mettons en évidence des relations de dépendance syntaxique dont la note de la critique et la valence du terme associé ont un effet sur leur distribution dans le corpus. Nous parlerons alors d'une relation surreprésentée ou sous-représentée dans une note particulière (de 1 à 5) avec un terme d'une certaine polarité (positif ou négatif). Nous restreignons ici notre recherche à l'étude des adverbes et étudions pour cela uniquement la relation de dépendance impliquant un adverbe à un adjectif positif ou négatif.

---

<sup>1</sup><http://www.allocine.fr/>

### 3. Description systématique des adverbes extraits et de leurs propriétés statistiques

Note :	1/5	2/5	3/5		4/5		5/5
SuR avec Adj. Positif	<i>particu- lièrement pas tellement</i>	<b>peu pas</b>	<b>moins plutôt assez trop</b>	<b>un peu parfois toujours fort</b>	<i>particulièrement très</i>		<b>plus à la fois absolument profondément</b>
SuR avec Adj. Négatif	<b>complète- ment fausse- ment aussi de plus propre- ment franche- ment</b>	<b>lourdement vite souvent totalement assez franche- ment bien aussi</b>	<b>un peu parfois trop pas moins assez</b>		<b>jamais</b>		<b>jamais à la fois presque</b>
SoR avec Adj. Positif			<b>profondément particulièrement</b>		<b>moins trop peu</b>	<b>plutôt parfois un peu</b>	<b>moins pas plutôt assez</b> <b>un peu trop très parfois</b>
SoR avec Adj. Négatif	<b>parfois un peu jamais trop presque</b>	<b>jamais parfois un peu moins</b>	<b>complètement</b>		<b>trop assez aussi franchement vite souvent totalement lourdement plutôt bien de plus</b>		<b>un peu assez trop aussi</b>

Tableau 1. Distribution statistique des adverbes potentiellement modificateurs

Notre méthodologie a été appliquée aux adverbes associés à des adjectifs positifs ou négatifs qui apparaissent plus de 20 fois dans le corpus. Nous obtenons ainsi une liste de 30 adverbes surreprésentés (SuR) ou sous-représentés (SoR). Le tableau 1 nous montre la distribution de ces adverbes selon les deux critères utilisés par le traitement statistique : la cotation de la critique de 1 à 5 (en colonne), et la valence de l'adjectif auquel l'adverbe est associé (en ligne). Sur la base de ce tableau, nous avons repéré trois comportements statistiques relativement similaires (un poids plus important est donné à la surreprésentation

qu'à la sous-représentation). Ils sont représentés par différentes couleurs : les adverbes en jaune seront étudiés en 3.1, ceux en rouge en 3.2, ceux en vert en 3.3, et enfin les adverbes particuliers (en noir) en 3.4. Ces sections passent en revue le rôle sémantique de chaque adverbe dans le corpus afin d'en dégager éventuellement des points communs.

### 3.1. *Surreprésentation dans les critiques notées 3/5*

Nous remarquons tout d'abord la présence de toute une série d'adverbes (en jaune) surreprésentés dans les critiques d'opinion mitigée (3/5) (souvent sous-représentés dans les autres catégories) : *un peu, trop, assez, parfois, moins, plutôt, toujours, fort, et pas* (traité dans la section 3.2 au vu de ses propriétés statistiques particulières).

A première vue, ces adverbes semblent être utilisés pour apporter une nuance et apparaissent souvent dans des structures rhétoriques, en particulier dans des mécanismes de comparaison, ou d'opposition-concession, comme le montrent les exemples ci-dessous.

- (1) Le cadre sublime rachète un scénario *un peu* fumeux qui travaille dans les rouages d'une intrigue *plutôt* maigre. [3/5]
- (2) Ce conte philosophique et humaniste n'a rien de déplaisant, mais souffre d'un développement *assez* convenu, d'une dimension symbolique *un peu* grossière et d'une mise en scène qui frôle l'exercice de style. [3/5]
- (3) La poésie trash du réalisateur ne faiblit pas, contrairement au rythme de son récit *parfois* répétitif. [3/5]

En outre, quelques-uns de ces adverbes sont intégrés dans des constructions spécifiques qui répondent aussi aux caractéristiques mises en avant. C'est le cas en particulier de *moins* utilisé comme comparatif, superlatif, ou dans la construction d'opposition *plus...plus...moins* (4), et de *assez* dans la construction *assez...pour* (5).

- (4) On l'eût aimé *moins* glacé, *plus* fiévreux, *plus* emporté. [3/5]
- (5) C'est bien nous que Natacha Cagnard voulait transformer en souris de laboratoire, mais son film n'est pas *assez* labyrinthique *pour* qu'on se laisse prendre au piège. [3/5]

Ils apparaissent également souvent répétés (6) ou combinés (7) et s'inscrivent dans des syntagmes ou phrases complexes qui accumulent les termes positifs ou négatifs. Ils servent ainsi souvent à articuler des propos portant sur diverses

caractéristiques du film (rythme, réalisation, jeu des acteurs, musique, etc.). Il s'avère enfin que la grande majorité des relations extraites sont pertinentes (bonne extraction de la relation et de la valence de l'adjectif) et que les structures extraites traitent bien des aspects positifs ou négatifs du film.

- (6) Tout est *trop* beau, *trop* lisse, *trop* coloré. [3/5]
- (7) Le trait se fait *parfois un peu trop* pittoresque, *trop* caricatural. [3/5]

Deux cas particuliers, *toujours* et *fort*, se détachent des utilisations mises en avant ci-dessus. Ils servent en effet globalement d'intensifieurs dans les notes 3/5. Néanmoins, leurs propriétés statistiques se justifient bien, car ils sont souvent utilisés dans des structures rhétoriques plus larges (8 et 9) et moins dans des critiques tranchées (1/5 ou 5/5). Le cas de *assez* se démarque également, surreprésenté dans les notes 3/5, mais aussi dans les notes négatives avec les adjectifs négatifs. Cela peut s'expliquer par le fait que, dans certains cas, il semble appuyer le propos du locuteur (et non l'atténuer) (10).

- (8) Parfois maladroit et didactique, mais *toujours* sincère. [3/5]
- (9) Si Duplicity est *fort* agréable, il n'apporte strictement rien au genre. [3/5]
- (10) Le scénario ne se fait jamais chair, et rend finalement l'entreprise *assez* artificielle. [2/5]

L'analyse dans le corpus des premiers adverbes mis en évidence par notre méthodologie reflète des situations relativement similaires et homogènes. Nous pouvons donc conclure que les adverbes surreprésentés dans les notes 3/5 indiquent essentiellement la présence de structures rhétoriques plus complexes (opposition-concession, comparaison) qui mettent souvent en relation en même temps des termes positifs et négatifs.

### 3.2. *Adverbes surreprésentés avec un adjectif dont la polarité ne coïncide pas avec la polarité de la critique*

Nous relevons 6 adverbes (en rouge) surreprésentés lorsqu'ils sont associés à un adjectif positif dans une critique négative, ou inversement : *pas*, *jamais*, *presque*, *peu*, *particulièrement*, *tellement*. Nous pouvons leur supposer, de par cette caractéristique, une tendance à atténuer ou inverser la polarité de l'adjectif. Cela ne semble pas être adapté à *particulièrement* et *tellement*, cas que nous aborderons dans la section 3.4. En revanche, les adverbes *pas*, *jamais*, *presque*, *peu* montrent un comportement moins étonnant et partagent certains points communs. En effet, ces adverbes portent souvent directement sur l'adjectif qui

leur est associé, et apparaissent moins dans des structures rhétoriques complexes (concession, opposition, ...), contrairement aux adverbes rencontrés dans le cas précédent.

L'adverbe *jamais* (11) illustre particulièrement bien cette tendance : il inverse la polarité des adjectifs négatifs et également celle des adjectifs positifs (bien que cela soit beaucoup moins fréquent). Une intensification supplémentaire pourrait même être décelée. L'adverbe *peu* (12), quant à lui, remplit de façon homogène la fonction d'atténuation de la polarité de l'adjectif.

- (11) Vive, réjouissante, naturelle, légère, bien vue, *jamais* caricaturale, (...). [5/5]  
 (12) C'est absurde, *peu* crédible, inintéressant (...). [2/5]

L'adverbe *presque* présente un fonctionnement relativement diversifié et complexe à définir. Toutefois, lui attribuer une faculté d'inversion ou d'atténuation n'est pas choquant. Par exemple, dans (13), on peut considérer que l'expression *presque éprouvante* est positive, polarité inverse de la valeur initialement négative de l'adjectif *éprouvant*. D'un autre point de vue, on peut considérer que *presque* nuance le crescendo de la phrase entière.

- (13) Le film, par son dépouillement même, la rigueur de sa forme, atteint une intensité étonnante, *presque* éprouvante. [5/5]

Enfin, l'adverbe *pas*, présente des propriétés statistiques appuyant l'hypothèse qu'il inverse effectivement la polarité des adjectifs positifs et négatifs. Toutefois, il faut rester prudent sur ces conclusions car l'extraction porte uniquement sur un cas particulier (*pas* associé à un adjectif) et les adjectifs associés à *pas* dans le corpus sont parfois classés de façon non pertinente (erreurs dues à la constitution automatique du lexique). Enfin, il est important de noter la surreprésentation de *pas* dans les notes 3, contrairement aux autres adverbes de cette section. Nous remarquons effectivement une utilisation courante de *pas* dans des structures rhétoriques plus larges (à l'image des adverbes de la première section) (14).

- (14) Si le graphisme et le dessin ne sont *pas* sublimes, le film ménage de jolis moments de suspense. [3/5]

En conclusion, la surreprésentation de relations adverbes-adjectifs dans des critiques dont la polarité est opposée à celle de l'adjectif semble significative d'une inversion ou d'une atténuation directe de l'adjectif (une sous-représentation dans les critiques négatives avec des adjectifs négatifs ou dans les critiques positives avec des adjectifs positifs appuie cette hypothèse). Nous devons tout de même rester prudent car les cas relevés s'avèrent avoir souvent un

fonctionnement complexe, et deux incohérences apparaissent : *tellement* et *particulièrement*. Notons également l'impossibilité, sur la base des résultats statistiques, de distinguer les cas d'inversion des cas d'atténuation.

### 3.3. *Adverbes surreprésentés avec des adjectifs dont la polarité est la même que celle de la critique*

La dernière caractéristique statistique qui se détache est une surreprésentation des adverbes associés à des adjectifs négatifs dans des critiques négatives ou à des adjectifs positifs dans des critiques positives (et parfois une sous-représentation dans les cas inverses). Les 14 adverbes qui répondent à ces caractéristiques semblent tous apporter une intensification à la valeur de l'adjectif ou à l'énoncé. Ces adverbes ne sont pourtant pas tous utilisés de la même façon. Certains sont préférentiellement associés à des adjectifs d'une polarité particulière (3.3.1), d'autres sont inclus dans des structures lexico-syntaxiques plus larges (3.3.2), d'autres encore, plus particuliers et plus rares, peuvent être eux-mêmes considérés comme positifs ou négatifs (3.3.3).

#### 3.3.1. *Adverbes préférentiellement associés à des adjectifs d'une certaine polarité*

La plupart de ces adverbes ont un comportement intensifiant relativement similaire quelle que soit la polarité de l'adjectif auquel ils sont associés, et sont simplement plus fréquemment utilisés avec une polarité particulière. En effet, les adverbes *bien*, *franchement*, *souvent*, *totalemment*, *vite* (15), *complètement* et *proprement* (16) apparaissent essentiellement avec les adjectifs négatifs. Les adverbes *profondément* et *absolument* (17), de leur côté, sont plus souvent utilisés dans un contexte positif. Certains cas d'intensification peuvent être plus discutés que d'autres mais globalement la valeur de l'énoncé semble être appuyée par les adverbes cités ci-dessus.

(15) (...) le film s'avérant *vite* incohérent, mal joué et répétitif (...) [1/5]

(16) (...) une vulgarité, une absence de timing comique et un manque d'imagination *proprement* assommants. [1/5]

(17) (...) car le film est *absolument* jubilatoire. [5/5]

#### 3.3.2. *Adverbes utilisés dans des structures intensifiantes plus complexes*

Les adverbes *aussi* (18) et *plus* (19) sont utilisés également pour leur effet d'intensification mais s'inscrivent dans des structures plus complexes, principalement dans des structures comparatives ou superlatives.



- (18) The Killer Inside Me est *aussi* creux *que* déplaisant. [1/5]  
 (19) (...) la comédie *la plus* brillante *qu'on* ait savourée depuis longtemps (...). [5/5]

Notons également l'extraction de la relation de dépendance *de plus* associé à un adjectif. Cette analyse syntaxique peut ne pas sembler adaptée aux cas extraits et être considérée comme une erreur de l'analyseur automatique, mais elle met tout de même en évidence des structures très spécifiques et typiquement intensifiantes (20 à 22).

- (20) *Quoi de plus* triste *qu'une* comédie qui n'arrache pas un sourire. [2/5]  
 (21) Sans doute ce que le cinéma israélien nous a envoyé *de plus* vaseux depuis longtemps. [1/5]  
 (22) *Rien de plus* pathétique à supporter *qu'une* comédie romantique ratée (...). [2/5]

### 3.3.3. Adverbes porteurs d'une certaine polarité

Les adverbes *lourdement* et *faussement*, tous deux surreprésentés avec des adjectifs négatifs dans des énoncés négatifs se démarquent un peu des adverbes vus précédemment. En effet, plusieurs interprétations sont souvent possibles. En ce qui concerne *lourdement*, il est plausible, au vu des attestations dans le corpus, de considérer qu'il intensifie la valeur d'adjectifs négatifs, mais également qu'il a une valeur négative qui lui est propre (23). Le cas de *faussement* s'avère encore plus complexe. Nous pouvons en effet considérer qu'il nie la réalité de l'adjectif auquel il est associé, tout en ayant de toute façon un impact négatif. Ainsi, il aura un effet d'inversion sur un adjectif positif (24) ; le rôle qu'il joue en compagnie d'un adjectif négatif est plus difficile à définir (25).

- (23) De bons comédiens ne parviennent pas à sauver ce thriller *lourdement* mélodramatique. [2/5]  
 (24) Une biographie fanfaronne et *faussement* intelligente (...). [1/5]  
 (25) Bonitzer n'a rien d'autre à offrir qu'un marivaudage *faussement* détaché, *faussement* névrotique et *faussement* cauchemardesque. [1/5]

En résumé, un adverbe surreprésenté lorsqu'il est associé à un adjectif dont la polarité est la même que celle de la critique est clairement un intensifieur. En effet les adverbes extraits remplissent tous cette fonction, bien que certains aient des fonctionnements plus complexes que d'autres. Cette propriété est, de plus, la plus productive des trois propriétés statistiques dégagées.

### 3.4. *Cas particulier et cas problématiques*

Un premier adverbe, *à la fois*, présente des propriétés statistiques uniques : une surreprésentation dans les critiques positives qu'il soit associé à un adjectif positif ou négatif. Cette caractéristique s'explique par le fait que la fonction inhérente de cet adverbe est d'associer plusieurs adjectifs. D'après les attestations dans le corpus, *à la fois* (26) met en effet en relation plusieurs termes, pouvant être ou non de même polarité et est utilisé plus souvent dans des critiques positives. Néanmoins, comme *à la fois* est le seul adverbe répondant à ces propriétés statistiques particulières, il est difficile, d'en tirer une règle plus générale.

(26) Tout *à la fois* glacial, effrayant, malin, réjouissant, efficace. [4/5]

Nous rencontrons enfin trois adverbes réellement problématiques dont les propriétés statistiques sont rares et s'expliquent quelque fois difficilement : *particulièrement*, *très* et *tellement*.

Ainsi, *particulièrement* est surreprésenté avec les adjectifs positifs aussi bien dans les critiques positives que négatives. La surreprésentation avec un adjectif positif dans un contexte positif suggère un effet d'intensification, comme nous l'avons constaté dans 3.3. Cet effet est bien confirmé par les attestations du corpus (27). En revanche, la surreprésentation avec un adjectif positif dans un contexte négatif suggère une atténuation ou une inversion (section 3.2), effet qui n'est pas confirmé par les attestations de *particulièrement*. Il s'avère que cette propriété statistique provient certainement d'une mauvaise classification de certains adjectifs associés à *particulièrement* dans le lexique de valence.

(27) Et le final, (...), est *particulièrement* savoureux. [4/5]

Les propriétés statistiques de *tellement* (surreprésenté avec les adjectifs positifs dans les critiques négatives) suggèrent un effet d'atténuation ou d'inversion, effet qui ne se confirme pourtant pas dans le corpus. En effet, les attestations de *tellement* (28) montrent globalement un impact intensifiant clair de l'adverbe, la plupart du temps utilisé dans des structures comparatives.

(28) Un film *tellement* anodin *qu'on* a du mal à croire qu'il soit signé par les réalisateurs des géniaux "Mary à tout prix" et "Deux en un".  
[2/5]

Enfin, l'adverbe *très* présente un résultat illogique : il est en même temps surreprésenté et sous-représenté dans les notes positives avec des adjectifs

positifs (surreprésenté dans les notes 4, et sous-représenté dans les notes 5). Dans le corpus, *très* joue pourtant clairement un rôle d'intensifieur.

En conclusion, seuls 3 des 30 adverbess extraits posent réellement problème ici. Cela provient essentiellement de quelques problèmes méthodologiques (notamment du bruit généré par le traitement automatique lors de la création du lexique et de l'analyse syntaxique).

#### 4. Conclusion et perspectives

Nous avons cherché dans cet article à évaluer une procédure automatique qui vise à établir une liste de modifieurs et leur impact (à partir de l'étude d'adverbes). Il s'avère que, d'une part, les adverbess extraits sont effectivement des modifieurs de valence pertinents. D'autre part, la méthodologie développée nous a permis d'identifier, sur la base de propriétés statistiques communes, trois types d'adverbes modifiant la valeur d'adjectifs positifs ou négatifs: (1) des adverbess intégrés dans des structures rhétoriques complexes et qui mettent souvent en relation plusieurs termes polarisés, (2) des adverbess ayant un rôle clair d'intensification sur l'adjectif qui lui est associé, (3) des adverbess atténuant ou inversant directement la valeur d'un adjectif. Cette dernière catégorie, plus difficile à appréhender que les deux premières, mérite peut-être un autre type de traitement. Nous obtenons donc une classification qui se distingue quelque peu de celle de Zaenen et Polanyi (2004). En effet, la différence entre atténuateur et inverseur ne ressort pas dans notre étude.

Il est raisonnable de penser que ces conclusions peuvent être valables pour d'autres structures que les adverbess. Nous pouvons alors envisager de traiter tout type de dépendances syntaxiques et de finaliser la procédure d'extraction automatique en produisant directement une liste de trois types de modifieurs sur la base des règles établies ici. Nous obtiendrons ainsi vraisemblablement une liste plus importante que celles déjà existantes, même si l'exhaustivité recherchée initialement ne sera pas atteinte pour l'instant. Il nous faut souligner, ensuite, que les effets d'inversion sont difficiles à appréhender de cette manière car ils reposent sur des structures souvent plus complexes (*être à mille lieues de ...*). Une nouvelle stratégie nécessiterait d'être développée pour repérer ces éléments particuliers.

Enfin, ce travail se situe à mi-chemin entre le domaine du traitement automatique de la langue et l'étude linguistique. Nous avons analysé systématiquement le rôle

de 30 adverbes dans un corpus et cherché à simplifier les conclusions dans le but de les intégrer à une procédure automatique. Dans le prolongement de cette question, nous pouvons nous interroger sur la pertinence de la distinction entre atténuation, inversion et intensification. Est-elle réellement adaptée à la fois aux stratégies concrètes d'un locuteur et à une intégration dans un système automatique ? Dans notre étude, nous pouvons entrevoir l'existence de deux comportements quelque peu différents. Certains adverbes semblent avoir un effet plus global sur l'énoncé (d'autant plus s'ils mettent en relation plusieurs termes porteurs de valence). D'autres ont un impact plus direct sur le terme auquel ils sont associés. Autrement dit, on ne peut pas toujours déterminer la valence de chaque terme en contexte séparément. Il serait intéressant d'approfondir cette question.

## Références

- Aït-Mokhtar, S., Chanod, J.-P. et Roux, C. (2002) 'Robustness Beyond Shallowness: Incremental Deep Parsing'. *Natural Language Engineering* 8, 121-144.
- Boubel, N. et Bestgen Y. (2011) Une procédure pour identifier les modificateurs de la valence affective d'un mot dans des textes. Dans *Actes de TALN11*, 2:137-142. Montpellier.
- Charaudeau, P. (1992) *Grammaire du sens et de l'expression*. Collection Hachette éducation. Paris: Hachette.
- Daille, B., Dubreil, E., Monceaux, L. et Vernier, M. (2011) 'Annotating opinion—evaluation of blogs: the Blogoscopy corpus'. *Language Resources and Evaluation*: 1–29.
- Ding, X., Liu, B. et Yu, P. (2008) A holistic lexicon-based approach to opinion mining. Dans *Proceedings of the international conference on Web search and web data mining*, 231–240.
- Hogenraad, R., Bestgen, Y. et Nysten, J.-L. (1995) Terrorist Rhetoric : Texture and Architecture. Dans *From Information to Knowledge*, 48-59. Intellect Book. Nissan et Schmidt.
- Kennedy, A. et Inkpen, D. (2006) 'Sentiment classification of movie reviews using contextual valence shifters'. *Computational Intelligence* 22 (2): 110–125.
- Kerbrat-Orecchioni, C. (1997) *L'énonciation : de la subjectivité dans le langage*. Paris : Colin.
- Klenner, M., Petrakis, S. et Fahrni, A. (2009) Robust compositional polarity classification. Dans *Proceedings of RANLP*.
- Martin, J.R. et White, P.R.R. (2005) *The language of evaluation: Appraisal in English*. Palgrave Macmillan.
- Musat, C. et Trausan-Matu, S. (2010) 'The Impact of Valence Shifters on Mining Implicit Economic Opinions'. *Artificial Intelligence: Methodology, Systems, and Applications*: 131–140.
- Pang, B. et Lee, L. (2008) 'Opinion mining and sentiment analysis'. *Foundations and*

*Trends in Information Retrieval* 2 (1-2): 1–135.

Turney, P. (2002) Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Dans *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*.

Vincze, N. et Bestgen, Y. (2011) Identification de mots germes pour la construction d'un lexique de valence au moyen d'une procédure supervisée. Dans *Actes de TALN11*, 1:223-234. Montpellier.

Wiebe, J., Wilson, T. et Cardie, C. (2005) 'Annotating expressions of opinions and emotions in language'. *Language Resources and Evaluation* 39 (2): 165–210.

Wilson, T., Wiebe, J. et Hwa, R. (2004) Just how mad are you? finding strong and weak opinion clauses. Dans *Proceedings of the 19th national conference on Artificial intelligence*, 761–767. AAAI'04. San Jose, California: AAAI Press.

Zaenen, A. et Polanyi, L. (2004) Contextual valence shifters. Dans *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text*, 106–111.