

Déterminer automatiquement la valence affective de phrases : Amélioration de l'approche lexicale.

Yves Bestgen

Fonds national de la Recherche scientifique, Université catholique de Louvain
PSOR, Place du Cardinal Mercier, 10 — B-1348 louvain-la-Neuve — Belgique

Abstract

An approach simple, but effective, to determine automatically the emotional valence of a sentence consists in using the valence of the words which it contains. The problem is that the effectiveness of this procedure depends on the number of words contained in the sentence whose valence is known. To bypass this limitation, we describe several procedures allowing to determine automatically the emotional valence of words on the basis of their use in corpora. An experiment based on 702 sentences evaluated by judges confirms the effectiveness of these procedures.

Résumé

Une approche simple, mais efficace, pour déterminer automatiquement la valence affective d'une phrase consiste à se baser sur la valence des mots qu'elle contient. Le problème est que l'efficacité de cette procédure dépend du nombre de mots contenus dans la phrase dont la valence est connue. Pour dépasser cette limitation, nous décrivons plusieurs procédures permettant de déterminer automatiquement la valence affective de mots sur la base de leur emploi dans des corpus. Une expérience réalisée à partir de 702 phrases évaluées par des juges confirme l'efficacité de ces procédures.

Mots-clés : valence affective, procédure automatique, analyse sémantique latente.

1. Introduction

L'analyse de données textuelles, qu'elle s'effectue dans le champ de l'analyse de contenu ou dans celui du traitement automatique des langues naturelles, s'est, depuis toujours, intéressée prioritairement à décrire et à résumer le contenu des textes analysés. Depuis quelques années, l'attention se porte également sur l'attitude de l'auteur d'un texte. Le client qui a envoyé un courriel au service après-vente d'une firme est-il particulièrement mécontent? Les avis donnés dans les forums ou dans les "blogs" à propos de produits ou de films sont-ils positifs ou négatifs? Quelle est l'image d'un parti politique ou d'un dirigeant dans la presse? Toutes ces questions, qui portent sur la manière dont un élément est présenté et évalué dans un texte, sont particulièrement complexes pour les techniques classiques d'extraction d'information (Das et Chen, 2001 ; Wilks, 1997). Elles présentent pourtant des applications nombreuses comme, par exemple, la possibilité de transmettre aux membres les plus chevronnés d'un service clientèle les messages dont le ton émotionnel est le plus marqué.

Récemment, différents auteurs ont proposé de répondre à ces questions au moyen de procédures automatiques permettant de déterminer le caractère positif ou négatif, favorable ou défavorable d'un mot, d'une phrase ou d'un texte. Si certaines recherches tentent d'appliquer les méthodes classiques en linguistique computationnelle à l'évaluation du caractère subjectif

d'énoncés (Pang et al., 2002 ; Riloff et al., 2003 ; Wiebe et al., 2004 ; Yu et Hatzivassiloglou, 2003), d'autres s'inspirent de techniques déjà bien éprouvées en analyses de contenu. Dans ce champ, l'approche classique consiste à se baser sur l'orientation sémantique ou valence des mots. Il s'agit d'une des composantes de la signification connotative d'un mot reflétant "*un jugement d'appréciation, ou de dépréciation, porté sur l'objet dénoté*" (Kerbrat-Orecchioni, 1977, p.110). Dès 1965, Heise a proposé de constituer un dictionnaire de valence en demandant à des juges d'évaluer sur la dimension agréable-désagréable un échantillon des mots les plus fréquents d'une langue. Depuis lors, des dictionnaires pour différentes langues ont été constitués (Hogenraad et al., 1995 ; Whissell et al., 1986). Ces dictionnaires ont été employés pour évaluer la valence affective de textes, mais aussi d'unités plus petites comme des phrases. La procédure proposée par Heise (1965) est très simple. On dresse d'abord la liste des mots différents et de leur fréquence dans l'unité textuelle. Cette liste est comparée au dictionnaire contenant les mots dont on connaît la valence. A chaque fois qu'un mot se trouve dans les deux listes, on affecte la valeur indiquée dans le dictionnaire au mot du texte. Enfin, on calcule la moyenne des valeurs connues.

Malgré le caractère rudimentaire de cette technique, des arguments en faveur de sa validité ont pu être apportés (Anderson et al., 1982 ; Bestgen, 1994 ; Das et Chen, 2001). On a ainsi montré que les valeurs qu'elle attribue à des textes, des paragraphes ou des phrases permettent de prédire significativement la valence affective rapportée par des lecteurs. Très récemment, Bestgen, Fairon et Kevers (2004) ont confirmé ce résultat, tout en soulignant une limitation importante de l'approche. Il apparaît en effet que l'intensité affective d'une phrase est d'autant mieux prédite qu'on prend en compte la valence affective d'un plus grand nombre des mots qui la composent. Ces auteurs ont constitué un corpus de 702 phrases, extraits d'articles de journaux, qu'ils ont présentées à des juges afin qu'ils indiquent si elles évoquaient une idée plutôt désagréable, neutre ou agréable. Pour estimer la valence des phrases sur la base des mots qui les composent, ils ont employé un dictionnaire de 3000 mots (Hogenraad et al., 1995). La corrélation entre les évaluations des phrases et celles obtenues à partir du dictionnaire est de 0.39, une valeur statistiquement très significative ($p < 0.0001$). Afin de déterminer l'efficacité d'un dictionnaire qui prend en compte l'ensemble des mots inclus dans les phrases, ils ont demandé à deux juges de décider si chaque mot présent dans les phrases, mais absent du dictionnaire, était positif, négatif ou neutre. La corrélation entre la valence des phrases et celle obtenue sur la base de ce dictionnaire exhaustif est de 0.56.

Ce résultat est encore plus problématique qu'il n'y paraît à première vue. En effet, quel est l'intérêt d'une technique automatique s'il est nécessaire de demander à des juges de compléter le dictionnaire afin que celui-ci couvre l'ensemble du lexique employé dans le matériel à évaluer? Il y a toutefois une manière de résoudre ce problème : estimer la valence des mots qui ne se trouvent pas dans le dictionnaire à partir de l'analyse d'un grand corpus. La thèse sous-jacente est qu'il devrait être possible de déterminer la valence affective d'un terme en se basant sur les mots auxquels ce terme est fréquemment associé dans le corpus en question. Cette thèse trouve sa justification dans la manière dont un mot acquiert sa valence affective (Osgood et al., 1957). Celle-ci peut résulter des expériences individuelles avec l'objet dénoté, mais aussi des associations fréquentes avec d'autres mots dont la valence est connue.

L'objectif principal de la présente recherche est d'évaluer l'efficacité de deux techniques qui ont été proposées pour estimer la valence de mots sur la base de leur emploi dans des corpus. Dans la suite de ce rapport, les deux techniques sont décrites et leur efficacité, tant pour estimer la valence de mots que celle de phrases, est évaluée au travers d'une expérience.

2. Déterminer automatiquement la valence de mots

A ma connaissance, les premiers chercheurs ayant proposé une procédure automatique pour déterminer la valence de mots sont Hatzivassiloglou et McKeown (1997). Leur algorithme a pour objectif de déterminer l'orientation sémantique d'adjectifs à partir de l'analyse de leurs cooccurrences avec des conjonctions. La limitation principale de cet algorithme est qu'il a été développé spécifiquement pour les adjectifs et que la question de son application à d'autres catégories grammaticales n'a pas encore été résolue (Turney et Littman, 2002).

Turney et Littman (2002, 2003) et Bestgen (2002) ont proposé indépendamment des techniques plus générales puisqu'elles permettent d'estimer la valence de n'importe quel terme (adjectif, nom, verbe, mais aussi nom de personne ou de société) sur la base de sa proximité sémantique avec d'autres mots dont la valence est connue.

2.1 L'indice PR-ASL

La technique proposée par Turney et Littman se base sur la proximité sémantique entre le mot cible et 14 points de repère : 7 à valence positive (*good, excellent, ...*) et 7 à valence négative (*bad, poor, ...*). Un mot est d'autant plus positif qu'il est plus proche des points de repère positifs et plus éloigné des points de repère négatifs. Cette technique emploie donc un dictionnaire de valence composé de seulement 14 mots et l'étend automatiquement en déterminant la proximité sémantique entre ceux-ci et n'importe quel autre mot.

Pour mesurer la proximité sémantique sur la base d'un corpus de quelques millions de mots, Turney et Littman ont recours à l'analyse sémantique latente (ASL), une technique mathématique qui vise à extraire un espace sémantique de très grande dimension à partir de l'analyse statistique de l'ensemble des cooccurrences dans un corpus de textes (Deerwester et al., 1990; Landauer et al., 1998). Le point de départ de l'analyse est un tableau lexical qui contient le nombre d'occurrences de chaque mot dans chaque segment de textes. Ce tableau fait l'objet d'une décomposition en valeurs singulières qui en extrait les dimensions orthogonales les plus importantes. Dans cet espace, le sens de chaque mot est représenté par un vecteur. Pour mesurer la similarité sémantique entre deux mots, ou dans le cas présent entre un mot et un point de repère, on calcule le cosinus entre les vecteurs qui les représentent. Plus un mot est sémantiquement proche d'un point de repère, plus les deux vecteurs pointent dans la même direction et donc plus leur cosinus se rapproche de 1. Pour estimer la valence d'un mot, on calcule donc les cosinus entre ce mot et chaque point de repère positif ainsi que les cosinus entre ce mot et chaque point de repère négatif. La différence entre ces deux ensembles de cosinus indique la valence du mot.

Turney et Littman ont évalué l'efficacité de leur technique en comparant l'orientation prédite à celle définie dans le *General Inquirer Lexicon* (Stone et al., 1966) qui contient une liste de 3596 mots étiquetés comme positifs ou négatifs. Calculée sur la base de corpus de 10 000 000 de mots, PR-ASL étiquette correctement 65% des mots.

2.2 L'indice DI-ASL

La technique proposée par Bestgen (2002) est très similaire à celle de Turney et Littman. La principale différence entre ces deux techniques est que PR-ASL emploie comme points de repère quelques mots sélectionnés a priori alors que DI-ASL sélectionne un ensemble spécifique de points de repère pour chaque mot à évaluer parmi plusieurs milliers de mots dont la valence est connue. Plus précisément, DI-ASL s'appuie sur un dictionnaire de 3000 mots dont la valence a été évaluée par une trentaine de juges (Hogenraad et al., 1995). La

valence inconnue d'un mot correspond à la valence moyenne (non pondérée par la proximité) de ses 30 plus proches voisins dont la valence est connue. Ici aussi, les plus proches voisins sont identifiés sur la base d'une analyse sémantique latente et correspondent aux 30 mots ayant le plus grand cosinus avec le mot cible. Pour évaluer cet indice, Bestgen (2002) a comparé les valeurs prédites pour 60 mots du dictionnaire aux valeurs réelles et a obtenu des corrélations comprises entre 0.56 et 0.70. Il a également montré que prendre les 30 plus proches voisins donnait lieu à une meilleure estimation que la prise en compte de 5 voisins.

3. Expérience

L'objectif principal de la présente recherche est d'évaluer l'efficacité de ces deux techniques pour estimer la valence de phrases. Deux autres objectifs subsidiaires seront également poursuivis. Tout d'abord, les deux techniques s'appuient sur l'analyse sémantique latente pour estimer les similarités sémantiques dans des corpus de quelques millions de mots. Il ne s'agit toutefois pas de la seule procédure possible. Des indices computationnellement plus simples peuvent être employés. Le deuxième objectif vise donc à déterminer si l'emploi de l'ASL permet réellement d'accroître la qualité des estimations de valence par rapport à ces autres indices. Enfin, le type de corpus employé pour estimer les valences pourrait également affecter la précision de ces estimations. Tout particulièrement, on peut penser que plus un corpus contient des textes similaires à ceux qui devront ultérieurement être évalués, meilleur sera la prédiction. Tester cette hypothèse est notre troisième objectif.

3.1 Méthode

3.1.1 Tests pour évaluer l'efficacité des indices

Pour comparer l'efficacité des indices, deux tests ont été employés. Le premier est une épreuve de validation visant à vérifier qu'il est possible de déterminer la valence affective de mots en se basant sur les mots qui leur sont associés dans des textes. Pour ce test, j'ai employé le dictionnaire composé de 3 000 mots évalués sur la dimension agréable-désagréable (Hogenraad et al., 1995) qui sert également à l'estimation des valences des mots. Un minimum de 30 juges (étudiants dans des établissements d'enseignement supérieur ou dans des écoles techniques) ont procédé à l'évaluation des termes sur une échelle à 7 points allant de *Très désagréable* (1) à *Très agréable* (7). A titre d'exemple, la liste suivante donne les valeurs attribuées à quelques mots extraits aléatoirement de ce dictionnaire : détresse = 1.4, impassible = 2.6, ambigu = 3.2, outil = 4.3, revenir = 5.0, admiratif = 5.7, doux = 6.0.

Le second test est le plus important. Il vise à déterminer si les valences affectives obtenues grâce aux différents indices permettent de prédire efficacement la valence affective de phrases. Pour ce test, le matériel est composé de 702 phrases parues dans le quotidien national belge *Le Soir* en 1995. Ces phrases ont été évaluées par 10 juges. Leur tâche était d'indiquer, sur une échelle à 7 points, dans quelle mesure le contenu de chacune d'elles évoquait une idée plutôt désagréable, neutre ou agréable. L'accord interjuges, mesuré au moyen du coefficient alpha de Cronbach, est de 0.93, indiquant que la tâche a été bien comprise par les juges et que leurs évaluations sont très similaires. Une présentation détaillée de la procédure employée pour constituer ce matériel est donnée dans Bestgen et al. (2004).

3.1.2 Indices pour déterminer la valence de mots

Les deux indices principaux sont dérivés des techniques d'estimation décrites ci-dessus. Ils se basent tous les deux sur l'analyse sémantique latente pour déterminer la proximité entre des mots et des points de repère dont la valence est connue.

- PR-ASL : Il s'agit de la procédure proposée par Turney et Littman et décrite ci-dessus. La valence d'un mot correspond à la somme des cosinus entre ce mot et les points de repère positifs dont on soustrait la somme des cosinus entre ce mot et les points de repère négatifs. La valence d'une phrase correspond à la moyenne de ces différences de cosinus.
- DI-ASL : C'est la seconde technique présentée ci-dessus. Pour chaque mot dont la valence est inconnue, on identifie les 30 plus proches voisins dont on connaît la valence et on affecte la valence moyenne de ceux-ci à ce mot. Comme dans la procédure originale d'estimation sur la base d'un dictionnaire, la valence d'une phrase correspond à la moyenne des valences des mots qui composent la phrase.

Deux autres indices, qui ne nécessitent pas le recours à une analyse sémantique latente, ont été testés. Ils trouvent leur origine dans les travaux de Turney et Littman (2003). Ces auteurs ont proposé d'estimer la proximité sémantique entre un mot et un point de repère sur la base, non d'une analyse sémantique latente, mais d'une analyse des simples cooccurrences entre ces deux termes. Le problème est que cette approche nécessite un très grand corpus de textes pour être efficace. Turney et Littman ont employé dans leurs analyses un corpus composé de l'ensemble des textes anglais disponibles sur Internet, soit quelque 100 milliards de mots¹. L'origine de cette limitation se trouve dans la nécessité d'observer des cooccurrences entre le mot et le point de repère pour obtenir un indice de proximité non nul. Lorsque la valence est estimée sur la base de quelques points de repère et d'un corpus de quelques millions de mots, cette situation se produit trop fréquemment pour permettre une estimation efficace. Un tel problème a beaucoup moins de chance de survenir lorsque plusieurs milliers de points de repère peuvent être employés. Nous avons donc testé deux indices qui se basent sur le dictionnaire (30 plus proches voisins) et sur les simples cooccurrences dans un corpus : la Pointwise Mutual Information (DI-PMI) préconisée par Turney et Littman et le classique cosinus (DI-COS). Ces deux indices peuvent être calculés, dans le cas de données binaires, à partir de quatre valeurs (Manning et Schütze, 1999, pp. 179 et 298-300) :

m_1 = le nombre de segments dans lesquels le mot_1 est présent,

m_2 = le nombre de segments dans lesquels le mot_2 est présent,

m_{12} = le nombre de segments dans lesquels le mot_1 et le mot_2 sont simultanément présents,

N = le nombre total de segments.

$$COS = \frac{m_{12}}{\sqrt{m_1 \times m_2}} \qquad PMI = \log_2 \left(\frac{\frac{m_{12}}{N}}{\frac{m_1}{N} \times \frac{m_2}{N}} \right)$$

¹ Turney et Littman ont employé pour cette analyse le moteur de recherche Altavista.

3.1.3 Corpus employés pour calculer les proximités.

Chacune des techniques d'estimation de la valence s'appuie sur un corpus pour estimer les proximités entre les mots à évaluer et les points de repère, que ces derniers soient issus d'une brève liste ou d'un dictionnaire de valence. Afin de prendre en compte la similarité entre le corpus et les phrases à évaluer, trois corpus ont été constitués :

- Corpus *Soir95*. Il comprend l'ensemble des articles dont sont extraites les phrases cibles, c'est-à-dire ceux parus pendant les 4 premiers mois de 1995. Il s'agit donc du corpus le plus spécifique au matériel à évaluer.
- Corpus *Soir97*. Un corpus comparable a été construit à partir des quatre premiers mois du *Soir* 1997. Ce corpus est donc issu de la même source, mais d'une période ultérieure.
- Corpus *Littéraire*. Un corpus de textes littéraires a été construit à partir de romans, nouvelles et contes disponibles sur le Web (principalement dans les bases littéraires ABU et Frantext). Il contient approximativement 5 300 000 mots. Ce corpus est composé de textes très différents de ceux dont sont issues les phrases à évaluer.

Ces corpus ont été subdivisés en segments de 125 mots. Tous les mots d'un segment devaient provenir du même texte. Tous les segments de moins de 125 mots (articles de petites tailles et le dernier segment incomplet d'un texte) ont été supprimés. Ce découpage a produit 40 635 segments pour le corpus *Littéraire* et plus de 50 000 pour les deux autres corpus. Afin de pouvoir comparer des corpus de types différents, mais de mêmes tailles, seuls les 40 635 premiers segments des corpus *Soir95* et *Soir97* ont été pris en compte.

Pour pouvoir comparer les valeurs obtenues à partir des quatre indices et des trois corpus, les mêmes critères pour la sélection des mots ont été employés pour la construction des tableaux lexicaux : lemmatisation par le logiciel TreeTagger (Schmidt, 1994), suppression de mots outils et suppression des mots de fréquence totale inférieure à 10^2 . A chaque fois, les 300 premiers vecteurs propres ont été conservés.

3.2 Analyses et résultats

3.2.1 Analyse 1 : Estimation de la valence de mots

Le premier test est a priori le plus logique à effectuer. Il vise à déterminer si les valences prédites par les indices correspondent à celles perçues par des juges. Pour les trois corpus, les quatre indices ont été employés pour estimer la valence de tous les mots présents dans le dictionnaire et dans chacun des corpus. Ces estimations ont été comparées aux valeurs données dans le dictionnaire au moyen du coefficient de corrélation de Pearson.

Il est important de noter que ce test présente une limitation qui en réduit la portée lorsque l'on compare PR-ASL aux trois autres indices. En effet, et contrairement à PR-ASL, ceux-ci s'appuient tous sur ce même dictionnaire d'évaluation pour estimer la valence des mots. Bien évidemment, lors de la recherche des plus proches voisins d'un mot, celui-ci n'est jamais pris en compte. On peut néanmoins penser que cette double utilisation du dictionnaire leur procure un avantage. Ce problème ne contamine toutefois pas la comparaison entre ceux-ci.

² Des analyses complémentaires ont été réalisées en changeant le seuil minimal de fréquence pour les mots (plus grand ou égal à 5 et plus grand ou égal à 2). Les résultats ne sont pas modifiés.

Le Tableau 1 indique très clairement que PR-ASL fonctionne moins bien que les trois autres mesures. Comme expliqué ci-dessus, il est préférable d'attendre la seconde analyse pour conclure à propos de l'efficacité de cet indice. Pour ce qui concerne les trois autres indices, les différences sont moins importantes. Néanmoins, DI-COS et DI-ASL sont les plus efficaces. Lorsqu'on compare les trois corpus, on observe que c'est le corpus *Littéraire* qui donne lieu aux meilleurs résultats.

Corpus	N	PR-ASL	DI-ASL	DI-COS	DI-PMI
Soir95	2603	0.36	0.53	0.54	0.47
Soir97	2596	0.36	0.54	0.52	0.44
Littéraire	2666	0.39	0.60	0.63	0.55

Tableau 1 : *Corrélations entre les valences estimées et les évaluations des juges pour tous les mots présents dans chaque corpus et dans le dictionnaire (N indique le nombre de mots pris en compte).*

3.2.2 Analyse 2 : Estimation de la valence de phrases

Le test le plus important porte sur la possibilité d'estimer la valence des phrases en se basant sur les mots qui les composent. Pour chaque corpus et pour chaque indice, on a estimé la valence de tous les mots catégorisés comme des noms, des verbes, des adjectifs ou des adverbes par TreeTagger (Schmidt, 1994). Comme pour la procédure classique, la valence d'une phrase est estimée par la moyenne de toutes les valences connues. Ces estimations ont été comparées aux évaluations des juges au moyen du coefficient de corrélation de Pearson. Deux niveaux de référence pour jauger l'efficacité des indices sont donnés par le résultat de l'analyse effectuée sur la base du dictionnaire original et du dictionnaire étendu. Bestgen et al. (2004) ont obtenu une corrélation de 0.39 avec le dictionnaire original. Sur la base du dictionnaire étendu, cette corrélation est de 0.56.

Corpus	PR-ASL	DI-ASL	DI-COS	DI-PMI
Soir95	0.43	0.54	0.54	0.50
Soir97	0.42	0.51	0.54	0.50
Littéraire	0.30	0.44	0.47	0.45

Tableau 2 : *Corrélations entre les valences des phrases estimées par les quatre indices sur la base des trois corpus et les évaluations faites par des juges*

On observe que la technique proposée par Turney et Littman fait à peine mieux que l'approche classique. Cette performance est toutefois appréciable puisque PR-ASL ne se base que sur 14 mots dont la valence est déterminée a priori, par des experts humains donc, alors que l'approche classique s'appuie sur un dictionnaire de 3 000 mots évalués par des juges.

Les trois indices qui étendent le dictionnaire original obtiennent des résultats nettement supérieurs et presque équivalents à ceux obtenus par l'entremise du dictionnaire étendu. Le résultat problématique pour l'approche lexicale, rapporté dans l'introduction, ne remet donc plus en cause cette approche.

DI-COS semble légèrement plus efficace que DI-ASL et DI-PMI. Si l'analyse sémantique latente ne donne pas lieu à des performances nettement moins bonnes que les indices basés sur les simples cooccurrences, force est de constater qu'elle n'améliore pas ces performances.

Enfin, la comparaison des corpus montre que les plus efficaces sont ceux qui sont composés à partir de textes du même genre que celui dont est issu le matériel à évaluer. On notera toutefois que les deux corpus extraits du Soir fonctionnent aussi bien l'un que l'autre alors que l'un est spécifique au matériel de test, puisqu'il inclut les articles dont sont issues les phrases à évaluer et que le second est postérieur de deux ans.

Conclusion

L'objectif principal de cette recherche était d'évaluer différents indices permettant d'estimer sur la base de corpus la valence de mots et d'utiliser ces valences apprises pour décider si des phrases évoquaient pour des lecteurs une idée agréable, neutre ou désagréable. L'expérience rapportée ci-dessus montre qu'obtenir par ce genre de techniques automatiques la valence de mots est aussi efficace que de demander à des juges d'évaluer ces mots. Ceci est démontré par l'indice de Turney et Littman qui ne s'appuie que sur quatorze mots dont la valence a été déterminée par des experts pour évaluer n'importe quel autre mot. Or, l'expérience montre que cet indice est aussi efficace qu'une approche basée sur un dictionnaire composé de 3000 mots très fréquents évalués par des juges. Les trois autres indices démontrent également l'efficacité des techniques automatiques puisqu'ils permettent une estimation aussi efficace que celle obtenue au moyen du dictionnaire de 3000 mots rendu exhaustif par le recours à des juges supplémentaires.

La présente recherche montre aussi que le recours à l'analyse sémantique latente pour estimer la valence de termes n'est pas indispensable et qu'il est important de prendre en compte le type d'énoncés à évaluer pour choisir le corpus qui servira à l'estimation des valences. Ce dernier résultat mérite une attention particulière. La première analyse montre que la meilleure approximation des valences données par des juges est obtenue au moyen du corpus littéraire. Par contre, pour évaluer la valence de phrases extraites d'articles de journaux, ce même corpus littéraire est nettement moins efficace que des corpus composés d'articles de journaux. Il aurait été intéressant de disposer d'un matériel de test composé de phrases extraites de textes littéraires afin de déterminer si le corpus littéraire donne lieu avec ce matériel à la meilleure estimation de la valence par rapport aux corpus d'articles. En l'absence d'un tel contrôle, il semble préférable de rester prudent dans l'interprétation de cette observation.

Plus généralement, on peut penser que, dans la présente expérience, nous sommes proches de l'efficacité maximale de l'approche lexicale pour évaluer des phrases. En effet, les techniques automatiques sont aussi efficaces que l'approche classique basée sur un dictionnaire exhaustif. Or, la corrélation entre les valences prédites et les valences obtenues de juges est juste supérieure à 0.50. Si on souhaite aller au-delà, il est probablement indispensable de prendre en compte d'autres facteurs linguistiques que les mots qui composent une phrase et les candidats ne manquent pas. En effet, le caractère "simpliste" de cette approche basée sur les mots considérés individuellement a fait l'objet de plusieurs critiques (Bestgen, 1994 ; Pang et al., 2002 ; Polanyi et Zaenen, 2003) qui plaident pour la combinaison d'informations lexicales

et d'analyses linguistiques plus complexes. Par exemple, Polanyi et Zaenen (2003) soulignent la nécessité de prendre en compte les négations, mais aussi les opérateurs modaux. On notera toutefois que ces critiques envers l'approche lexicale ne rejettent pas celle-ci, mais soulignent la nécessité de la compléter. Disposer des indices lexicaux les plus performants possible est donc un prérequis pour suivre ces nouvelles pistes.

Références

- Anderson C.W. and McMaster G.E. (1982). Computer assisted modeling of affective tone in written documents. *Computers and the Humanities*, 16 : 1-9.
- Bestgen Y. (1994). Can emotional valence in stories be determined from words ? *Cognition and Emotion*, 8 : 21-36.
- Bestgen Y. (2002). Détermination de la valence affective de termes dans de grands corpus de textes. *Actes du Colloque International sur la Fouille de Texte*, Nancy, INRIA : 81-94.
- Bestgen Y., Fairon C. and Kevers L. (2004). Un baromètre affectif effectif. *Actes des 7^{es} Journées internationales d'Analyse statistique des Données Textuelles*, Louvain-la-Neuve, Presses universitaires de Louvain : 182-191,.
- Das S. and Chen M. (2001). Yahoo! for Amazon : Opinion extraction from small talk on the web. Working Paper (under review), Santa Clara University.
- Deerwester S., Dumais S.T., Furnas G.W., Landauer T.K. and Harshman R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, 41 : 391-407.
- Hatzivassiloglou V. and McKeown K.R. (1997). Predicting the semantic orientation of adjectives. *Proceedings of the 35th Meeting of the Association for Computational Linguistics* : 174-181.
- Heise D.R. (1965). Semantic differential profiles for 1000 most frequent english words. *Psychological Monographs*, 79 : 1-31.
- Hogenraad R., Bestgen Y. and Nysten J.L. (1995). Terrorist Rhetoric : Texture and Architecture. In Nissan et Schmidt (Eds.), *From Information to Knowledge*, Intellect : 48-59.
- Kerbrat-Orecchioni C. (1977). *La connotation*, Lyon, Presses universitaires de Lyon.
- Manning C. and Schütze H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MIT Press.
- Osgood C., Suci G. and Tannenbaum P. (1957). *The measurement of meaning*, University of Illinois.
- Pang B., Lee L. and Vaithyanathan V. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in natural language processing* : 79-86.
- Polanyi L. and Zaenen A. (2003). Shifting attitudes. In L. Lagerwerf, W. Spooren, & L. Degand (Eds.), *Determination of information and tenor in texts : Multidisciplinary approaches to discourse 2003*, Münster, Nodus Publikationen : 61-69.
- Riloff E., Wiebe J. and Wilson T. (2003). Learning subjective nouns using extraction pattern bootstrapping. In *Proceedings of the 7th Conference on Natural Language Learning* : 25-32.
- Schmidt H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. Version électronique disponible sur [<http://www.ims.uni-stuttgart.de/Tools/DecisionTreeTagger.html>].
- Stone P. J., Dunphy D. C., Smith M. S. and Ogilvie D. M. (1966). *The General Inquirer : A Computer Approach to Content Analysis*. MIT Press, Cambridge, MA.
- Turney P. and Littman M. (2002). Unsupervised learning of semantic orientation from a hundred-billion-word corpus. Technical Report, National Research Council Canada.
- Turney P. and Littman M. (2003). Measuring Praise and Criticism : Inference of Semantic Orientation from Association. *ACM Transactions on Information Systems (TOIS)*, 21 : 315-346
- Wiebe J., Wilson T., Bruce R., Bell M. and Martin M. (2004). Learning Subjective Language. *Computational Linguistics*, 30 : 277-308.

- Wilks Y. (1997). Information Extraction as a Core Language Technology, in M. T. Paziensa (Ed.) *Information Extraction*, Springer : 1-9.
- Yu H. and Hatzivassiloglou V. (2003). Towards answering opinion questions : Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* : 129-136.

Note à propos de l'auteur

Yves Bestgen est chercheur qualifié du Fonds national de la recherche scientifique (FNRS).