

Composition multilingue de sentiments

Stefanos Petrakis¹ Manfred Klenner¹ Étienne Ailloud¹ Angela Fahrni¹
(1) Institute of Computational Linguistics, University of Zurich, Switzerland
{petrakis, klenner, ailloud}@ifi.uzh.ch, angela.fahrni@swissonline.ch

Résumé. Nous présentons ici PolArt, un outil multilingue pour l’analyse de sentiments qui aborde la composition des sentiments en appliquant des transducteurs en cascade. La compositionnalité est assurée au moyen de polarités préalables extraites d’un lexique et des règles de composition appliquées de manière incrémentielle.

Abstract. We introduce PolArt, a multilingual tool for sentiment detection that copes with sentiment composition through the application of cascaded transducers. Compositionality is enabled by prior polarities taken from a polarity lexicon and the compositional rules applied incrementally.

Mots-clés : Analyse de sentiments.

Keywords: Sentiment Detection.

Moyens techniques : Accès à Internet : <http://www.cl.uzh.ch/kitt/polart/>.

1 Introduction

L’analyse de sentiments a pour objectif d’identifier les polarités (positives ou négatives) de portions d’un texte—mots, syntagmes et phrases. La polarité d’un syntagme est obtenue en combinant les polarités des mots qui le composent, et ainsi de suite : L’analyse de sentiments est donc à traiter de manière compositionnelle. Par exemple, la phrase « This is an unrealistic hope » est négative, car un adjectif négatif jouxtant un nom positif produit un syntagme nominal négatif.

PolArt fait usage de lexiques de sentiments, spécifiques à chacune des langues-cibles. Pour la version anglaise nous utilisons le lexique de subjectivité (T.Wilson & P.Hoffmann, 2005), pour la version allemande un lexique manuellement extrait et annoté à partir de GermaNet (<http://www.sfs.uni-tuebingen.de/GermaNet/>). La version française utilise un lexique automatiquement induit à partir du lexique allemand, en se basant sur le dictionnaire allemand-français disponible à l’adresse <http://dict.leo.org/>. Le lexique de sentiments allemand contient 8500 mots étiquetés avec une polarité et comme la version française est basée sur la version allemande et que nous conservons les synonymes, le lexique français en contient plus de 8500 (par ex., l’allemand ‘Angst’ induit les polarités pour : ‘anxiété’, ‘peur’, ...).

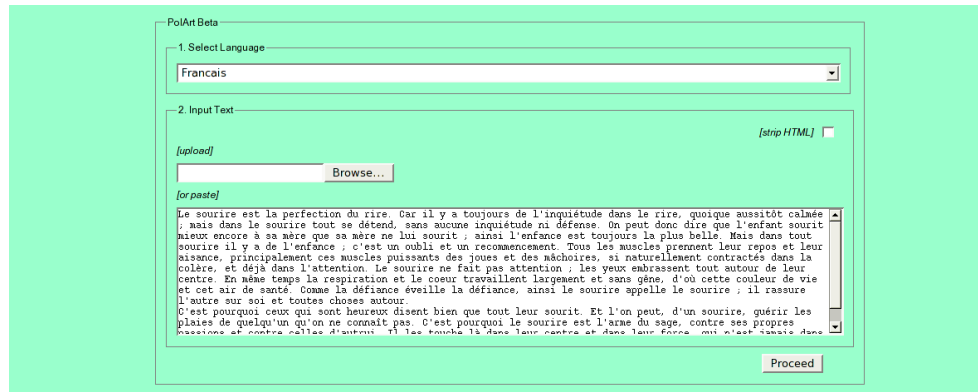


FIG. 1 – Outil de visualisation : entrée

2 PolArt—l’outil de visualisation

La figure 1 montre le panneau de saisie de PolArt, notre outil d’analyse de sentiments. L’utilisateur peut directement insérer du texte dans la zone de saisie ou bien décider de charger un fichier contenant le(s) texte(s) à analyser. Le système travaille sur du texte brut, mais une option permet néanmoins d’analyser du texte avec balises HTML. Enfin, l’utilisateur doit sélectionner la langue-cible appropriée, pour l’instant à choisir parmi anglais, allemand et français.

La figure 2 montre un exemple de sortie produite à partir d’un texte français. Ici, une exploration par cibles est possible. Pour illustrer ceci, nous avons choisi une liste arbitraire de mots-cibles (fenêtre de gauche). Le choix de ces mots-cibles est spécifique à l’application désirée. En cliquant sur un élément de la liste (par ex. ‘défiance’), les occurrences positives, négatives et non évaluées de la cible sont rendues accessibles (par déroulement dans la même fenêtre). Un clic sur une valeur de polarité (par ex. négative) fait apparaître tous les syntagmes évaluant la cible selon cette valeur et leur fréquence s’affiche dans la fenêtre en haut à droite (fenêtre des syntagmes polarisés). En cliquant sur le syntagme (par ex. ‘la défiance éveille la défiance’), on fait apparaître le contexte discursif dans la fenêtre en bas à droite (fenêtre contextuelle); les syntagmes interprétés par le système sont mis en surbrillance avec des couleurs différenciées.

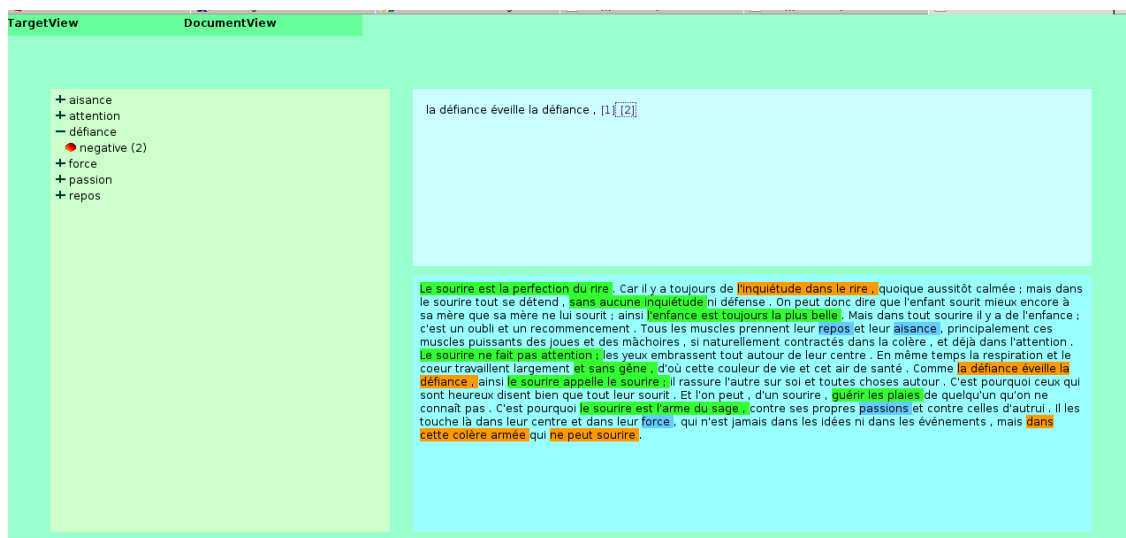


FIG. 2 – Outil de visualisation : sortie

3 Composition de la polarité

Notre approche de l'analyse de sentiments est conceptuellement simple et pragmatique ; elle se contente d'un petit nombre de règles de composition élaborées à la main et d'un schéma incrémentiel en cascade. Passer à un domaine différent ne nécessitera que l'adaptation du lexique, pas des règles. Les données de départ pour cette cascade sont fournies par la sortie de l'étiqueteur-chunker (H.Schmid, 1994), disponible pour l'anglais, le français et l'allemand. En partant par ex. de la phrase « He doesn't fail to verify his excellent idea », la cascade se présente comme suit (les indices indiquent l'itération, \rightarrow la réécriture ; la polarité préalable des lexèmes est donnée par leur exposant : '+' signifie positif, '*' dénote un échangeur de polarité, 'idea' est neutre) :

his excellent ⁺ idea	\rightarrow	POS ₁
to verify ⁺ POS ₁	\rightarrow	POS ₂
fail* POS ₂	\rightarrow	NEG ₃
doesn't* NEG ₃	\rightarrow	POS ₄

Au final la polarité est négative au niveau de la phrase. Nous avons élaboré un métalangage pour faciliter l'adaptation des règles (actuellement 80 pour l'anglais, 70 pour l'allemand et 10 pour le français), ce qui est nécessaire pour décrire les configurations spécifiques à chaque langue (notamment la position de l'adjectif dans le syntagme nominal, etc.). Notre évaluation de PolArt porte sur des textes anglais et allemands. La version française est actuellement en cours d'évaluation. Les données anglaises comprennent 248 syntagmes issus de 35 textes du groupe « I battle depression » de l'Experience Project (<http://www.experienceproject.com/>). Pour l'allemand ce sont 1209 syntagmes issus de romans d'Eduard von Keyserling. Cette évaluation donne 90,38% de précision pour les syntagmes positifs et 80,48% pour les négatifs.

4 Conclusion

Nous avons présenté PolArt, un outil robuste et trilingue d'analyse de sentiments. Une version de démonstration est accessible en ligne à <http://www.cl.uzh.ch/kitt/polart/>. Nous avons montré qu'une simple cascade de réécriture de polarités fournit un schéma simple et efficace pour une composition de polarité robuste. Nos règles de réécriture sont indépendantes du domaine, mais spécifiques à une langue (de même que les lexiques de polarité). Les versions anglaise et allemande sont stables et déjà évaluées ; la version française est en développement.

Remerciements

Ce travail de recherche est supporté par le projet 10015–122546/1 du Fonds National Suisse.

Références

- H.SCHMID (1994). Probabilistic part-of-speech tagging using decision trees. In *Inten. Conf. on New Methods in Language Processing 1994*.
- T.WILSON J. & P.HOFFMANN (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *HLT/EMNLP 2005*, Vancouver, CA.