HAL
open science

Détection hors contexte des émotions à partir du contenu linguistique d'énoncés oraux : le système EmoLogus

Marc Le Tallec, Jean-Yves Antoine, Jeanne Villaneau, A. Savary, A. Syssau

To cite this version:

Marc Le Tallec, Jean-Yves Antoine, Jeanne Villaneau, A. Savary, A. Syssau. Détection hors contexte des émotions à partir du contenu linguistique d'énoncés oraux : le système EmoLogus. Taln 2010, 2010, Montreal, Canada. pp.167. hal-00537177

HAL Id: hal-00537177

https://hal.science/hal-00537177v1

Submitted on 17 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Détection hors contexte des émotions à partir du contenu linguistique d'énoncés oraux : le système EmoLogus

Marc Le Tallec[1], Jeanne Villaneau[2], Jean-Yves Antoine[1], Agata Savary[1], Arielle Syssau-Vaccarella[3]

(1) Université François Rabelais Tours – LI
(2) Université Européenne de Bretagne – VALORIA
(3) Université Montpellier 3

Résumé Le projet EmotiRob, soutenu par l'agence nationale de la recherche, s'est donné pour objectif de détecter des émotions dans un contexte d'application original : la réalisation d'un robot compagnon émotionnel pour des enfants fragilisés. Nous présentons dans cet article le système qui caractérise l'émotion induite par le contenu linguistique des propos de l'enfant. Il se base sur un principe de compositionnalité des émotions, avec une valeur émotionnelle fixe attribuée aux mots lexicaux, tandis que les verbes et les adjectifs agissent comme des fonctions dont le résultat dépend de la valeur émotionnelle de leurs arguments. L'article présente la méthode de calcul utilisée, ainsi que la norme lexicale émotionnelle correspondante. Une analyse quantitative et qualitative des premières expérimentations présente les différences entre les sorties du module de détection et l'annotation d'experts, montrant des résultats satisfaisants, avec la bonne détection de la valence émotionnelle dans plus de 90% des cas.

Abstract The ANR Emotirob project aims at detecting emotions in an original application context: realizing an emotional companion robot for weakened children. This paper presents a system which aims at characterizing emotions by only considering linguistic content. It is based on the assumption that emotions can be compound: simple lexical words have an intrinsic emotional value, while verbal and adjectival predicates act as a function on the emotional values of their arguments. The paper describes the algorithm of compositional computation of the emotion and the lexical emotional norm used by this algorithm. A quantitative and qualitative analysis of the differences between system outputs and expert annotations is given, which shows satisfactory results, with a good detection of emotional valency in 90.0% of the test utterances.

Mots-clés : Emotion ; valence émotionnelle ; norme lexicale émotionnelle ; robot compagnon ; compréhension de parole

Keywords: Emotion; Emotional valency; Emotional lexical standard; companion robot; spoken language understanding

1 Introduction

Un nouveau champ d'étude en robotique est celui des robots compagnons susceptibles d'exécuter des tâches complexes et doués d'un système d'interaction enrichi avec les êtres vivants. Le projet EmotiRob, soutenu par l'Agence nationale de la Recherche, fait partie de ce domaine de recherche : il s'est fixé pour objectif de concevoir un robot peluche compagnon, autonome et réactif, capable d'interagir émotionnellement avec des enfants fragilisés afin de leur apporter un peu de réconfort. Des expérimentations antérieures (Saint-Aimé *et al.*, 2007) ont montré l'apport de ces robots dans ce type de situation.

Dans ce projet le robot peut exprimer des émotions à l'aide de mouvements faciaux. Pour produire une interaction émotionnelle pertinente, le robot doit tout d'abord reconnaître l'état émotionnel de l'enfant. Cette détection peut se faire à partir d'indices visuels (analyse de visage) ou par analyse des productions orales (indices prosodiques et/ou contenu linguistique des énoncés). Le système EmoLogus que nous avons développé se concentre précisément sur la détection des émotions à partir du contenu propositionnel. Cet article présente le système ainsi que ses premiers résultats.

2 Détection des émotions

Il n'existe pas de réel consensus sur la définition et la représentation des émotions. Disons simplement qu'il s'agit d'un état cognitif complexe influencé par le contexte à court-terme (contexte et historique de l'interaction) comme à long terme (vécu personnel et socioculturel) et dont la perception varie de manière sensible d'une personne à l'autre. Afin d'objectiver la caractérisation des émotions, deux grandes approches ont été étudiées par le passé. La première établit une **catégorisation nominale** des émotions en classes appelées *modalités émotionnelles*. (Ekman, 1999; Cowie, Cornelius, 2003): outre l'état émotionnel neutre, on distingue ainsi la colère, la joie, le dégout, la peur, la surprise et la tristesse. La seconde approche réalise une **catégorisation ordinale** dans un espace multidimensionnel. Parmi les échelles de valeurs retenues, on trouve le degré d'excitation ou la valence émotionnelle (émotion positive vs. négative). Quelle que soit l'approche suivie, les travaux sur l'émotion aboutissent à deux constats (Forbes-Riley, Litman, 2004; Devillers *et al.*, 2005; Lee, Narayanan, 2005; Callejas, Lopez-Cozar, 2008) :

1. En interaction réelle, les tours de parole ne portent majoritairement aucune émotion perceptible : plus de 80% des énoncés peuvent ainsi être qualifiés de neutres
2. Toutes les expériences d'annotation présentent un faible accord inter-annotateurs, avec des valeurs de Kappa (Landis, Koch, 1977) comprises entre 0,32 et 0,55. Au final, une annotation de référence ne peut être obtenue que par vote majoritaire entre plusieurs experts.

La plupart des recherches concernant la détection des émotions en situation de dialogue s'appuient sur des indices prosodiques. Notre objectif est d'étudier ce que peut apporter une détection linguistique. Cet axe ayant été très peu exploré, nous nous limitons dans un premier temps à une caractérisation des émotions en valence (positive, nulle ou négative) associée à un degré d'intensité.

3 Détection des émotions avec EmoLogus : principe de compositionnalité

Une première approche envisageable pour la détection linguistique des émotions consiste à attribuer une valence émotionnelle à chaque mot et à calculer la valence émotionnelle globale d'un énoncé comme la somme (éventuellement normalisée) des valences lexicales de ses termes. Cette première approche, qui ne considère pas la structure de l'énoncé, sera utilisée comme baseline par la suite.

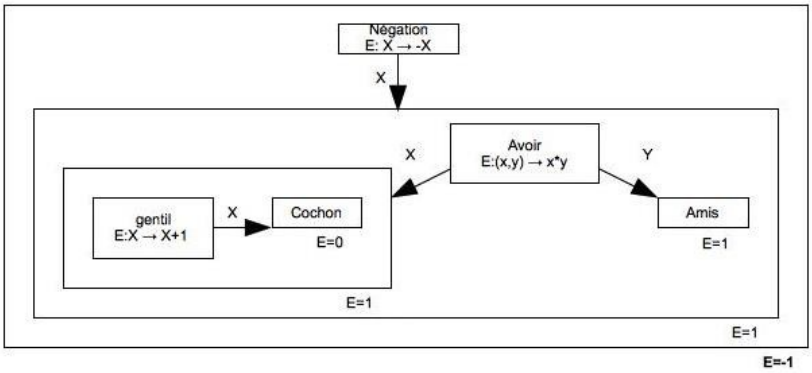

Figure 1: Exemple de calcul d'émotion *sur la phrase « le gentil cochon n'a pas d'amis »*

À l'opposé, le système EmoLogus repose sur le principe que l'émotion contenue dans un énoncé est compositionnelle : elle dépend à la fois de l'ensemble des valeurs émotionnelles des mots et de la structure sémantique qui décrit précisément les relations des mots entre eux. En pratique, les mots non prédicatifs du vocabulaire possèdent une valeur émotionnelle intrinsèque (valence), tandis que les mots prédicatifs (essentiellement les verbes ou les adjectifs) sont associés à des fonctions qui donnent une émotion à partir des valences émotionnelles des arguments du prédicat. Pour illustrer ce mécanisme, considérons la phrase de la figure 1 et la formule logique qui la représente. Le calcul commence par la prise en compte de la valence émotionnelle des mots *cochon* (E=0) et *amis* (E=1), qui ne sont que de simples arguments de la formule. L'adjectif *petit* et le verbe *avoir* fonctionnent comme des prédicats. *Gentil* ayant comme définition E = X X +1, le groupe nominal *petit cochon* va avoir pour valeur E =1. L'application successive de ces prédicats nous permet d'arriver au résultat final pour l'énoncé E= -1. On trouve dans notre lexique une dizaine de fonctions émotionnelles, comme celles présentées en exemple ci dessous :

- Décalage positif E : x ➛ x+1 (exemple : mignon)
- Décalage négatif E : x ➛ x-1 (exemple : énervé)
- Emotion dépendante des deux arguments E : (x,y) ➛ x * y (exemple : perdre)
- Emotion inverse à l'objet E : (x,y) ➛ -y (exemple : casser)

Le principe de compositionnalité sur lequel repose notre système est assez simple. Il nécessite toutefois :

1. une compréhension robuste de la parole spontanée pour fournir la structure prédicative correcte,
2. une base lexicale émotionnelle propre sur laquelle se basera le calcul des émotions.

Ces deux éléments font l'objet des deux paragraphes suivants.

4 Normes émotionnelles et définition des prédicats

La première ressource nécessaire à EmoLogus est la valeur émotionnelle de chacun des lexèmes du vocabulaire. La psychologie expérimentale a de longue date défini des normes lexicales émotionnelles associant à un mot sa valence.

A notre connaissance, il n'existe que deux études de ce type spécifiques aux enfants : (Vasa *et al*, 2006) pour l'anglais et (Syssau et Monnier, 2009) pour le français. Ces études ont montré que si les normes émotionnelles sont stables chez l'adulte et l'adolescent, elles varient fortement au cours des premières étapes du développement cognitif de l'enfant. La norme émotionnelle de Syssau et Monnier repose sur une catégorisation ordinale en trois modalités (négatif, neutre, positif). Nous avons complété ces travaux par l'évaluation de 80 mots absents de la norme et présents dans notre lexique qui couvre globalement le

vocabulaire d'un enfant de 7 ans. Cette nouvelle norme a été établie pour deux âges différents (5 et 7 ans) par expérimentation dans 4 écoles primaires.

La baseline repose entièrement sur cette norme lexicale. Par contre, dans EmoLogus, la norme lexicale associée aux mots prédicatifs est remplacée par un prédicat émotionnel, comme nous l'avons présenté au paragraphe précédent. Une seconde norme, portant sur ces prédicats, a été obtenue suivant une procédure d'annotation réalisée par 5 experts adultes. Chacun d'entre eux a proposé au plus deux définitions pour chaque prédicat, avant qu'une procédure de consensus ne définisse la fonction retenue à chaque fois. Il est intéressant de noter qu'il a finalement été possible d'arriver à un accord total. La notion de fonction prédicative émotionnelle semble donc paradoxalement être plus stable que celle de valence émotionnelle : les sujets semblent arriver plus facilement à un accord sur l'émotion portée par un prédicat (comme par exemple *tuer*) que sur des mots simples (*banane*, *école*), où le vécu personnel de chacun entre plus facilement en jeu.

5 Expérimentations et résultats : étude quantitative

A l'heure actuelle, EmoLogus est capable de caractériser l'émotion contenue dans un énoncé hors-contexte : nous n'avons pas encore développé de stratégie pour modéliser l'influence de l'émotion courante sur l'appréhension de chaque nouvel énoncé. Le comportement hors contexte du système a été évalué sur le corpus Brassens (Le Tallec *et al*., 2009) comportant 173 énoncés enfantins annotés en émotions.

Pour réaliser une annotation hors-contexte des énoncés, ces derniers ont été présentés dans un ordre aléatoire à 5 annotateurs adultes qui devaient leur attribuer une modalité parmi 5, de -2 (très négatif) à +2 (très positif). Cette annotation combinait donc valence émotionnelle (positif/neutre/négatif) et intensité de l'émotion portée. La valeur référence de chaque phrase a été déterminée par un vote majoritaire sur les décisions des experts. Les résultats de cette annotation ont été présentés dans (Le Tallec *et al*., 2009). Le tableau 1 compare la précision d'annotation du système EmoLogus et de la baseline sur le corpus de test.

	EmoLogus (avec intensité)	Baseline (avec intensité)	EmoLogus (sans intensité)	Baseline (sans intensité)
Précision	90,00%	68,80%	94,00%	72,83%

Tableau 1: Précision d'annotation du système EmoLogus et de la baseline.

EmoLogus présente des résultats encourageants avec un taux de bonne réponse à 90%, bien supérieur à ce que donnerait une procédure basique de calcul des émotions.

EmoLogus :

	-2	**-1**	**0**	**1**	**2**
-2	4	2	0	0	0
-1	2	18	0	0	0
0	1	5	116	2	0
1	0	0	3	16	1
2	0	0	0	1	2

Baseline :

	-2	***-1***	***0***	***1***	***2***
-2	4	0	0	1	0
-1	3	12	7	0	0
0	0	6	90	4	0
1	0	4	18	11	1
2	0	0	7	3	2

Tableau 2: Matrice de confusion des erreurs.

Le tableau 2 présente la matrice de confusion des erreurs pour les deux systèmes testés. Pour chaque case, on trouve l'annotation humaine en abscisse et celles des systèmes en ordonné. L'erreur la plus grave que

peut commettre un système est de détecter une valence émotionnelle opposée à celle attendue. Ce type d'erreur n'est jamais observé avec EmoLogus, à la différence de la baseline. La majorité des erreurs (47%) consiste à attribuer une émotion neutre à un énoncé annoté comme positif ou négatif. On pourrait qualifier cette situation de problème de rappel, si le neutre ne constituait pas un état émotif propre. Enfin, l'insertion d'une émotion inexistante aux yeux des experts ne concerne que 18% des erreurs d'EmoLogus.

Une part significative des erreurs observées ne concerne par ailleurs que l'estimation de l'intensité émotionnelle, la valence étant correctement détectée (exemple : positif vs. très positif). Si on se limite à la détection de la valence émotionnelle, la précision du système EmoLogus atteint alors 94%.

L'analyse qualitative des erreurs montre que notre système rencontre des difficultés à modéliser ce que nous pourrions qualifier d'isotopie émotionnelle. Prenons l'exemple du verbe *être enfermé*. En toute logique, nos experts ont associé à ce prédicat la fonction $E : x \rightarrow -x$. Si le sujet est associé à une valence positive (exemple : *princesse*), son enfermement est perçu négativement. Si on adopte cette fonction, un sujet non porteur d'émotion, ou mal identifié (E=0) entrainera donc le calcul d'une émotion neutre. Pourtant, il semble que les annotateurs ressentent une légère émotion négative dans ces situations. Il semble donc nécessaire d'attribuer une valeur négative par défaut à ces prédicats lorsque la valeur émotionnelle de leur argument est inconnue ou neutre. C'est l'émotion portée par les arguments qui modifie le comportement émotionnel du prédicat, un peu comme, en sémantique, les sèmes isotopiques sont activés en contexte.

Un autre problème, circonscrit principalement à certains adjectifs, résulte de ce qu'on pourrait qualifier de polysémie émotionnelle. Prenons le cas de l'adjectif *petit*. Globalement cet adjectif a tendance à changer l'argument qu'il qualifie pour le rendre plus positif (exemple : *le petit loup*) : $x \rightarrow x+1$. D'autres comportement émotionnels doivent toutefois être considérés. Par exemple, nos annotateurs ne perçoivent pas de décalage vers le positif sur un exemple comme *la petite maison*. Il semble donc que certains prédicats doivent se voir attribuer plusieurs comportements émotionnels qui seront à désambigüiser dans le contexte local de l'énoncé.

Il reste enfin à aborder la question de l'influence du contexte général du discours sur la perception des émotions, qui n'a pas été évaluée dans le cadre de ces premières expérimentations.

6 Conclusion et perspectives

Les premières évaluations du système de détection des émotions EmoLogus sont encourageants. Les résultats obtenus semblent en effet montrer qu'il est possible de détecter la bonne émotion dans un énoncé dans 90% des cas. Un des points les plus importants est que le système n'annote jamais un énoncé avec une émotion inverse à celle attendue. Nous avons modifié certains prédicats afin de répondre aux problèmes rencontrés lors de nos expérimentations, en particulier dans les cas liés à une émotion sur un objet neutre émotionnellement.

Nous travaillons désormais sur la détection des émotions en contexte (Liscombe *et al*., 2005). Outre les problèmes classiques de résolution des anaphores, cette détection soulève deux problèmes.

1. La dynamique des émotions dans un récit, qui définit comment la valeur émotionnelle d'un énoncé dépend de celle des énoncés précédents.

2. Le suivi de la valeur émotionnelle des « personnages » : si « mon petit frère est méchant », il le reste probablement dans les énoncés suivants.

Par ailleurs, nous étudions actuellement l'influence des erreurs de la reconnaissance de la parole sur le comportement du système. Enfin, nous prévoyons de coupler notre module de détection linguistique des émotions avec un système s'appuyant sur la prosodie du signal de parole (Chetouani, 2009).

Remerciements

Ce projet est financé par **l'Agence Nationale de la Recherche** (projet PSIROB06_174281).

Références

BASSANO D., LABRELL F., CHAMPAUD C., *et al.* (2005). Le DLPF, un nouvel outil pour l'évaluation du développement du langage de production en français. *Enfance*, 2(5):171–208.

CALLEJAS Z., LOPEZ-COZAR R. (2008). Influence of contextual information in emotion annotation for spoken dialogue systems. *Speech Communication.* 50. 416-433.

CHETOUANI M., (2009). *Mutlisensory Signal Processing for Emotion Recognition* .Workshop on "Current Challenges and Future Perspectives of Emotional Humanoid Robotics", IEEE International Conference on Robotics and Automation (ICRA'09).

COWIE R., CORNELIUS R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication.* 40. 5-32.

DEVILLERS L., VIDRASCU L., VASILESCU I. (2005). Emotion detection in task-oriented spoken dialogs. *Journal of Neural Networks*. 18(4).

EKMAN P. (1999) Patterns of emotions: New Analysis of Anxiety and Emotion . Plenum Press.

EL MAAROUF I., LE TALLEC M., VILLANEAU J. (2009), Ontologies naturelles et coercion : formalisation de connaissances à partir d'observations en corpus. J*ournées de linguistiques de corpus,* Lorient, France.

FORBES-RILEY K., LITMAN L. (2004). Predicting emotion in spoken dialogue from multiple knowledge sources. Actes de *HLT/NAACL'2004*. 161-176.

LANDIS J., AND KOCH G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* 33. 159-174.

LEE C.M., NARAYANAN S. (2005). Towards detecting emotions in spoken dialogs. *IEEE Transactions On Speech and Audio Processing.* 13(3). 293-303.9

LE TALLEC M., VILLANEAU J., ANTOINE J-Y., SAVARY A., ARIELLE SYSSAU-VACCARELLA A. (2009). Détection des émotions à partir du contenu linguistique d'énoncés oraux : application à un robot compagnon pour enfants fragilisés. Actes de *TALN 2009*.

LISCOMBE J., RICCARDI G., HAKKANI-TÜR D. (2005). Using context to improve emotion detection in spoken dialogue. Actes *INTERSPEECH 2005*.

SAINT-AIMÉ S., LE-PÉVÉDIC B., DUHAUT D., SHIBATA T. (2007). Emotirob : Companion Robot Project. Actes de *RO-MAN 2007*.

SYSSAU A., MONNIER C. (2009). Children's emotional norms for six hundred French words. *Behavior, Research, and Methods,* 41, 213-219.

VASA R. A., CARLINO A. R., LONDON K., MIN C. (2006). Valence ratings of emotional and non-emotional words in children. *Personality and Individual Differences,* 41, 1169-1180.

VILLANEAU J., ANTOINE J-Y. (2009). Deeper spoken language understanding for man-machine dialogue on broader application domains: a logical alternative to concept spotting. Workshop *SRSL09* in *EACL 2009*.