Détection des émotions à partir du contenu linguistique d'énoncés oraux : application à un robot compagnon pour enfants fragilisés

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

Marc.letallec@univ-tours.fr

Résumé – Le projet ANR Emotirob aborde la question de la détection des émotions sous un cadre original : concevoir un robot compagnon émotionnel pour enfants fragilisés. Notre approche consiste à combiner détection linguistique et prosodie. Nos expériences montrent qu'un sujet humain peut estimer de manière fiable la valence émotionnelle d'un énoncé à partir de son contenu propositionnel. Nous avons donc développé un premier modèle de détection linguistique qui repose sur le principe de compositionnalité des émotions : les mots simples ont une valence émotionnelle donnée et les prédicats modifient la valence de leurs arguments. Après une description succincte du système logique de compréhension dont les sorties sont utilisées pour le calcul global de l'émotion, cet article présente la construction d'une norme émotionnelle lexicale de référence, ainsi que d'une ontologie de classes émotionnelles de prédicats, pour des enfants de 5 et 7 ans.

Abstract – Project ANR Emotirob aims at detecting emotions from an original point of view: realizing an emotional companion robot for weakened children. In our approach, linguistic detection and prosodie are combined. Our experiments show that human beings can estimate the emotional value of an utterance from its propositional content in a reliable way. So we have implemented a first model of linguistic detection, based on the principle that emotions can be compound: lexical words have an emotional value while predicates can modify emotional values of their arguments. This paper presents a short description of the logical understanding system, the outputs of which are used for the final emotional value calculus. Then, the creation of a lexical emotional reference standard is presented with an ontology of emotional predicate classes for children, aged between 5 and 7.

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. Projet EmotiRob : robot compagnon émotionnel

Des expériences menées au Japon (Wada et al., 2004) et poursuivies par nos soins au Centre de Kerpape ont montré le réconfort que pouvait apporter un robot compagnon (*Paro* : www.paro.jp) à des personnes âgées ou malades. Le projet EmotiRob vise à créer un robot compagnon peluche destiné à de jeunes enfants (5/7 ans) fragilisés en hospitalisation longue. Il vise à corriger la principale faiblesse relevée lors de ces expériences: l'absence d'expressivité du robot. Tenue dans les bras, la peluche réagit par des émotions faciales et des sons. Capable d'exprimer des émotions élémentaires, le robot doit «comprendre» les propos de l'enfant et déterminer son état émotionnel, pour apporter la réponse la plus appropriée. Cet article se concentre sur la détection par le robot peluche des émotions de l'enfant.

2. Détection des émotions : définitions et état de l'art

Très étudiées en psychologie cognitive, les émotions n'ont retenu l'attention du TAL que dans les années 70 (synthèse de parole puis animation 3D). Plus récemment, le développement des interfaces affectives a remis en avant cette question. En DHM, l'idée est d'identifier l'état émotionnel de l'utilisateur pour adapter au mieux la réponse du système : par exemple, transférer un appel vers un opérateur humain en cas d'énervement. En règle générale, la détection de l'état émotionnel revient à caractériser sa valence et sa modalité.

2.1. Emotions : valence et modalité

Il n'existe toujours pas de consensus sur la définition d'émotion ni sur sa caractérisation. Il faut souligner ici la difficulté d'appréhender l'émotion, état cognitif complexe influencé par le contexte à court-terme (contexte et historique de l'interaction) comme à long terme (vécu personnel et culturel). Deux approches principales sont toutefois envisagées pour caractériser les émotions. La première établit une **catégorisation nominale** en classes souvent appelées *modalités émotionnelles*. Ces classes varient suivant les auteurs, mais on retrouve toujours sept modalités principales (Ekman, 1999; Cowie, Cornelius, 2003): ce sont 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 (positif/négatif). Cette dernière retient l'attention des psycholinguistes qui ont montré qu'elle peut biaiser les résultats expérimentaux.

Deux constats peuvent être relevées des travaux sur la catégorisation en émotions (Forbes-Riley *et al*., 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. La perception des émotions varie de manière sensible d'une personne à l'autre. Toutes les expériences d'annotation présentent des accords inter-annotateurs faibles 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 multiples experts.

2.2. Détection des émotions sur des corpus oraux ou multimédias.

Au vu de cette variabilité inter-annotateur, on saisit la difficulté que représente la détection automatique des émotions. Les méthodes de détection actuelles (Ververidis, Kotropoulos, 2006 ; Schuller *et al.,* 2007) reposent sur des classifieurs qui travaillent sur des indices paralinguistiques (acoustique, prosodie). Hormis quelques marques de disfluences, les indices linguistiques sont rarement considérés. Au mieux, les mots de l'énoncé ne sont considérés que comme marqueurs lexicaux isolés. L'intégration d'indices linguistiques améliore pourtant toujours les performances (Schuller *et al.,* 2007). Celles-ci demeurent toutefois perfectibles : une pertinence de 75% et une F-mesure de 0,6 constituent déjà des performances de pointe.

3. Détection linguistique des émotions : validation d'un principe

Ainsi, la détection linguistique des émotions constitue une voie peu explorée. L'objectif de ce projet est d'étudier si l'étude de la structure linguistique des énoncés ne pourrait pas conduire à une amélioration des performances. Nous nous limitons pour l'instant à la détection d'une valence émotionnelle positive/négative, suivant une échelle centrée de 5 valeurs [-2 ; +2].

Nous avons cherché à vérifier expérimentalement si une telle catégorisation linguistique était réalisable par unêtre humain dans le contexte applicatif qui nous intéresse. Aucun robot réactif émotionnel n'étant à ce jour disponible, nous avons procédé au recueil d'un corpus pilote. Collecté par le laboratoire AdiCore, ce corpus se compose d'une vingtaine d'historiettes imaginées et racontées par les enfants de CE1. Les transcriptions ont été annotées en valence émotionnelle sans écoute audio, par 9 annotateurs représentant des groupes adulte, adolescent et enfantin (5 à 9ans). Afin d'étudier l'effet du contexte (Callejas, Lopez-Cozar, 2008), deux annotations ont été réalisées : dans un premier temps, les énoncés ont été présentés dans un ordre aléatoire, ensuite ils ont été proposés en contexte en suivant la progression du récit.

Une première étude statistique a consisté à calculer le coefficient *alpha* de Cronbach (Cronbach, 1951), qui teste si les annotateurs ont le même objectif d'annotation. Dans notre cas, sa valeur est de 0,90. Elle démontre la cohérence de l'annotation par rapport à l'échelle de valeurs choisie : c'est bien la même réalité (la valence émotionnelle) qu'ont annotée les sujets.

L'analyse des annotations montre que les adultes sont pondérés dans leur catégorisation: 67% des énoncés sont considérés comme neutres. A l'opposé, les catégorisations enfantines sont très marquées : les classes extrêmes représentent 55% des annotations. L'influence de l'âge sur la perception des émotions est donc manifeste. Bien connue en psychologie, elle semble ignorée en TALN. Ce résultat se retrouve dans la mesure de l'accord inter-annotateur.

Nos classes d'annotation n'étant pas indépendantes (échelle multivaluée), nous utilisons un Kappa pondéré : une même annotation contribue pour 1 à l'accord global, un écart d'une échelle (exemple : neutre = 0 et positif = 1) pour 0,5, tandis qu'un écart plus grand représente un désaccord. L'effet de l'âge sur l'annotation est clair : le Kappa commun à l'ensemble des sujets est faible, mais les accords générationnels sont bons. Chez les adultes, on observe un Kappa moyen de 0,86 hors contexte et 0,84 en contexte. Ici, les désaccords concernent exclusivement l'intensité de l'émotion : un énoncé jugé positif n'est jamais jugé négatif par un autre adulte. Le contexte semble peu influer sur la catégorisation. Ces analyses se retrouvent chez les adolescents, où l'accord est de 0,76 hors et en contexte. Chez les enfants, on observe au contraire une forte variation de catégorisation. Le Kappa est de 0,49 hors contexte et 0,38

en contexte. Des désaccords extrêmes existent, du type -2 / +2. Ce résultat s'explique par le fait que l'âge des sujets variait de 5 à 9 ans. Des études psycholinguistiques sur des normes lexicales émotionnelles (Syssau, Monnier, 2009) ont en effet montré une bascule émotionnelle marquée à 7 ans chez les filles et 9 ans chez les garçons.

Au final, ces résultats montrent qu'un accord générationnel est possible, mais qu'il est important de développer des systèmes adaptés à des tranches d'âge assez fines.

4. Modèle de détection linguistique des émotions

Il apparait ainsi qu'un humain peut détecter de manière assez fiable l'émotion portée par un énoncé en ne considérant que son contenu linguistique. D'où notre recherche d'une détection s'appuyant fortement sur ce contenu linguistique. Plus précisément : (i) un modèle de détection linguistique considère l'énoncé dans sa dimension structurale et non pas comme un simple sac de marqueurs lexicaux, (ii) en parallèle, des indices acoustiques et prosodiques seront considérés par un module de détection paralinguistique des émotions non décrit dans cet article. À l'opposé d'une intégration précoce des indices paralinguistiques et linguistiques, nous privilégions une fusion des sorties des modules paralinguistique et linguistique.

Notre modèle de détection linguistique s'appuie sur un principe de compositionnalité de l'émotion. Nous postulons en effet que la valence émotionnelle d'un énoncé dépend de la valence des mots qui le composent et de leurs relations sémantiques. On distingue : (i) les **lexèmes** émotionnellement **élémentaires**, qui portent une valence donnée ; ces valences sont décrites par une norme lexicale émotionnelle présentée au paragraphe 5, (ii) les **lexèmes prédicatifs**, dont le comportement émotionnel est de modifier la valence de leurs arguments. Le paragraphe 6 présente la méthodologie suivie pour leur construction. La caractérisation des relations entre les prédicats et arguments suppose qu'on ait établi la structure sémantique de l'énoncé. C'est la tâche du système LOGUS présenté au paragraphe 7.

5. Norme lexicale émotionnelle pour enfants de 5 à 7 ans

À la base de notre modèle émotionnel, nous avons besoin de connaître la valence émotionnelle qu'associent les enfants à chaque mot de leur lexique. C'est l'objet des normes lexicales émotionnelles, utilisées de longue date en psychologie expérimentale. Les normes lexicales émotionnelles compilent les évaluations subjectives d'une population de juges à propos d'une ou plusieurs caractéristiques émotionnelles des mots. Certaines normes concernent des caractéristiques émotionnelles originales comme l'évaluation de la durée de l'émotion évoquée par le mot (Niedenthal et *al*., 2004 ; Zammuner, 1998) ou encore la dominance, c'est-à-dire le fait d'être sous le contrôle de l'émotion ou de contrôler l'émotion évoquée par le mot (Bradley, Lang, 1999). Mais dans toutes les normes lexicales émotionnelles, deux caractéristiques sont systématiquement évaluées : la valence et l'intensité.

Ces deux caractéristiques sont majoritairement évaluées par une population adulte sur des échelles de jugement nominales (positif, neutre, négatif) ou ordinales (i.e., -5 très négatif à +5 très positif). A notre connaissance, seules deux normes récentes compilent les évaluations faites par de jeunes enfants : celle de Vasa, Carlino, London et Min (2006) pour l'anglais, et

celle de Syssau et Monnier (2009) en langue française (enfants âgés de 5, 7 et 9 ans). Les échelles de réponse utilisées ici sont les mêmes que celles utilisées avec les adultes avec de légères modifications. Le nombre de modalités est réduit (3 pour l'étude de Syssau et Monnier) et chaque modalité de réponse est à associée à un dessin représentant un visage souriant, triste ou neutre. L'examen des résultats montrent que dès 5 ans, les enfants sont capables de juger avec un accord inter juge conséquent la valence émotionnelle des mots.

Dans le cadre du projet EmotiRob, nous complétons la norme de Syssau et Monnier par l'évaluation de la valence émotionnelle de 80 nouveaux mots par des enfants de 5 et 7 ans. Dans la norme originelle, les mots sont classés par âge d'acquisition et sont essentiellement des noms et des adjectifs. Nous ajoutons ici de nouveaux mots extraits du lexique compilé par (Bassano *et al.*, 2005) : cette extension présente l'intérêt de proposer aux enfants des mots caractéristiques de leur âge (e.g., *gronder*, *écrabouiller*, *rigolo*) et surtout des verbes, qui n'ont encore jamais été évalués par des enfants du point de vue de la valence émotionnelle.

Pour les enfants de 5 ans, les mots ajoutés sont divisés en 2 listes de 40 mots évaluées dans deux sessions différentes. Pour ceux de 7 ans, la liste est évaluée en une seule session. A chaque âge, deux ordres aléatoires de présentation des mots sont définis, chaque ordre étant présenté à la moitié des participants. Ces expérimentations ont été réalisées dans 4 écoles à Lorient et à Blois. Pour plus de renseignement sur le protocole expérimental, on se référera à (Syssau, Monnier, 2009). Ces expérimentations prendront fin en avril 2009. Nous disposerons à cette date d'une norme lexicale de référence pour le français et présentant une large couverture par rapport aux productions enfantines attendues.

6. Classes de prédicats émotionnels

Une étude pilote a été réalisée pour tester la validité de notre principe de combinaison prédicat-argument au niveau émotionnel. Nous avons extrait du corpus présenté au paragraphe 3 un ensemble de n-uplets prédicats-arguments (e.g., *la sorcière est enfermée dans une prison*) qui ont été annotés en valence émotionnelle par 4 sujets adultes. L'accord inter-annotateur obtenu, avec le Kappa de 0,52, est très acceptable en classification émotionnelle.

Dans notre modèle, les prédicats sont décrits par leur *comportement émotionnel*, i.e. leur influence sur la valence de leurs arguments. Ainsi, le prédicat *énervé* décale la valence de son sujet vers le côté négatif, tandis que *méchant* impose la valence négative à l'énoncé, quel que soit l'argument. Nous avons recensé huit classes de comportement émotionnel (tableau 1). Une fois stabilisée, cette catégorisation sera validée avec une cohorte d'annotateurs.

Classe	Exemple	Calcul de la valence résultante
Prédicats **conservant** la valence d'un argument	*donner*	$\mathrm{Val}(x) = v \Rightarrow \mathrm{Val}(pred(x)) = v$
Prédicats **inversant** la valence d'un argument	*tuer*	$\mathrm{Val}(x) = v \Rightarrow \mathrm{Val}(pred(x)) = -v$
Prédicats **renforçant** la valence d'un argument	*vrai*	$\mathrm{Val}(x) = v \Rightarrow \mathrm{Val}(pred(x)) = 2 * v$
Prédicats **réduisant** la valence d'un argument	*petit*	$\mathrm{Val}(x) = v \Rightarrow \mathrm{Val}(pred(x)) = v / 2$
Prédicats **décalant positivement** la valence	*aimable*	$\mathrm{Val}(x) = v \Rightarrow \mathrm{Val}(pred(x)) = v + 1$
Prédicats **décalant négativement** la valence	*énervé*	$\mathrm{Val}(x) = v \Rightarrow \mathrm{Val}(pred(x)) = v - 1$
Prédicats **imposant** une valence **positive**	*gentil*	$\forall x\ \mathrm{Val}(pred(x)) = 1 \mid 2$
Prédicats **imposant** une valence **négative**	*méchant*	$\forall x\ \mathrm{Val}(pred(x)) = -1 \mid -2$

Tableau 1 – Classes prédicatives de comportement émotionnel

Notons enfin que le principe de combinaison prédicats-arguments peut résoudre le problème de la polysémie (Le Pesant, Mathieu-Colas,1998), et donc également celui de la détermination de l'émotion associée à un mot polysémique.

7. Compréhension de la parole : adaptation du système LOGUS

Notre modèle de détection linguistique des émotions suppose que soient caractérisées les relations prédicatives internes à l'énoncé. Les productions de l'enfant ne sont pas restreintes alors que la plupart des systèmes de compréhension de parole actuels sont destinés à des applications très finalisées : la tâche peut donc, a priori, sembler hors de portée. Néanmoins, le public concerné (très jeunes enfants) permet de restreindre la taille du vocabulaire. Par ailleurs, le robot émotionnel ne vise pas une compréhension parfaite : on peut espérer que la prosodie supplée la compréhension dans le cas de propos peu ou mal compris.

Pour réaliser cette tâche, nous utilisons le système LOGUS fondé sur une approche logique permettant une compréhension robuste sur des tâches complexes (Villaneau et al., 2004). LOGUS n'utilise pas de patrons sémantiques prédéfinis : les associations possibles entre concepts reposent sur une connaissance sémantique qui autorise ou non leur regroupement. Le système étant originellement dédié au renseignement touristique, la construction d'une ontologie du domaine couvert par le projet EmotiRob a été effectuée. Le système peut désormais analyser la plupart des énoncés de notre corpus de développement. Nous travaillons actuellement à l'implémentation du module de calcul de valence émotionnelle.

8. Conclusion

Nos études expérimentales suggèrent que le contenu linguistique d'un énoncé peut porter une valeur émotionnelle mesurable. Nous avons mis en place les éléments nécessaires à sa détection automatique. Il convient encore de comparer différents modèles de calcul global et de vérifier nos hypothèses sur un corpus représentatif. Ce module linguistique de détection émotionnelle pourra alors être couplé avec une analyse prosodique.

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'evaluation du developpement du langage de production en francais. *Enfance*, 2(5):171–208.

Bradley, M. M., & Lang, P. J. (1999). Fearfulness and affective evaluations of pictures. Motivation and Emotion, 23, 1-13.CALLEJAS Z., LOPEZ-COZAR R. (2008) Influence of contextual information in emotion annotation for spoken dialogue systems. *Speech Communication.* 50. 416-433.

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

CRONBACH L. J. (1951). Describing Coefficient alpha and the internal structure of tests. *Psychometrika. 16(3)*, 297-334.

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.

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

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.

LE PESANT D., MATHIEU-COLAS M. (1998). Introduction aux classes d'objets. *Langages* (131), Larousse, Paris, France. 6-33.

NIEDENTHAL P. M., AUXIETTE C., NUGIER A., DALLE N., BONIN P., FAYOL M. (2004). A prototype analysis of the French category "émotion". *Cognition and Emotion,* 18. 289-312.

SCHULLER B. *et al.* (2007) The Relevance of Feature Type for the Automatic Classification of Emotional User States. *Proc. Interspeech'2007*, Anvers, Belgique. 2253-2256.

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

VERVERIDIS D., KOTROPOULOS C. (2006) Emotional Speech Recognition: Resources, features and methods, *Speech communication*, 48(9). 1162-1181.

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., RIDOUX O. ANTOINE J-Y. (2004) Logus : compréhension de l'oral spontané *Revue d'Intelligence Artificielle*, 18(5-6). 709-742

WADA K., SHIBATA T., SAITO T., TANIE K.. (2004) Effects of robot-assistance activity for elderly people and nurse at day service center. 11TH IEEE, 92:1780–1788.

ZAMMUNER V.L. (1998). Concepts of emotion: Emotioness and dimensional rating of italian emotion words. *Cognition and emotion, 12, (2*), 243-272.