



**HAL**  
open science

## Intégration et évaluation de mécanismes de productions humoristiques dans un agent conversationnel virtuel

Matthieu Riou, Bassam Jabaian, Stéphane Huet, Thierry Chaminade, Fabrice Lefèvre

### ► To cite this version:

Matthieu Riou, Bassam Jabaian, Stéphane Huet, Thierry Chaminade, Fabrice Lefèvre. Intégration et évaluation de mécanismes de productions humoristiques dans un agent conversationnel virtuel. Workshop sur les Affects, Compagnons Artificiels et Interactions (WACAI), 2018, île de Porquerolles, France. hal-02021367

**HAL Id: hal-02021367**

**<https://hal.science/hal-02021367v1>**

Submitted on 26 Feb 2019

**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.

# Intégration et évaluation de mécanismes de productions humoristiques dans un agent conversationnel virtuel

Matthieu Riou  
CERI-LIA, Université  
d'Avignon  
Avignon, France  
matthieu.riou@alumni.univ-  
avignon.fr

Bassam Jabaian  
CERI-LIA, Université  
d'Avignon  
Avignon, France  
bassam.jabaian@univ-  
avignon.fr

Stéphane Huet  
CERI-LIA, Université  
d'Avignon  
Avignon, France  
stephane.huet@univ-  
avignon.fr

Thierry Chaminade  
INT, Université d'Aix Marseille  
Marseille, France  
thierry.chaminade@univ-  
amu.fr

Fabrice Lefèvre  
CERI-LIA, Université  
d'Avignon  
Avignon, France  
fabrice.lefevre@univ-  
avignon.fr

## ABSTRACT

Dans cet article, nous présentons nos travaux en cours autour des agents conversationnels oraux virtuels et de leur adaptation en ligne aux utilisateurs. En particulier plusieurs possibilités d'introduction de productions de traits humoristiques dans les systèmes de dialogue oral sont étudiées, afin de rendre plus naturelles les interactions entre un agent et les utilisateurs. Enfin, nous discutons comment les sciences cognitives, et en particulier les neurosciences, peuvent aider à évaluer de manière plus complète les systèmes et les évolutions proposés en les comparant, à la fois objectivement et subjectivement, aux situations d'interactions sociales entre humains.

## Keywords

Agent conversationnel virtuel, réseaux de neurones, interactions sociales, productions humoristiques

## 1. INTRODUCTION

Les agents conversationnels virtuels oraux, aussi appelés systèmes de dialogue oral, sont principalement utilisés pour des tâches spécifiques comme les services d'assistance technique ou les systèmes de réservations (de vols, hôtels, restaurants, etc.). Ces systèmes ont pendant longtemps reposé sur des systèmes experts basés sur des règles ou des patrons lexicaux décrits manuellement [14] pour générer leurs réponses. Depuis quelque temps, ils ont été remplacés ou améliorés avec des approches stochastiques [16]. Ces nouveaux systèmes n'améliorent pas considérablement les performances

objectives, mais permettent d'obtenir une plus grande variabilité dans leurs réponses, évitant ainsi des répétitions peu naturelles, et réduisent le recours à l'expertise humaine préalable. En contrepartie ils requièrent une grande quantité de données pour être entraînés.

Pour améliorer l'expérience et l'engagement de l'utilisateur, certains systèmes tentent d'intégrer des compétences sociales, comme l'expression d'émotions [11]. Dans nos travaux actuels, nous nous intéressons à l'insertion de productions de traits humoristiques dans un système de dialogue oral. Cet article étudie les possibilités offertes par les approches stochastiques pour la génération de réponses humoristiques par un agent conversationnel. La section 2 présente nos récents travaux sur les systèmes de dialogue oral, puis la section 3 étudie les différentes possibilités que nous avons envisagées pour introduire de l'humour dans un agent conversationnel et enfin la section 4 présente des travaux en cours pour évaluer ces systèmes, non seulement de manière objective, mais aussi subjective. Nous concluons par une discussion dans la section 5.

## 2. ADAPTATION EN LIGNE D'UN SYSTÈME DE DIALOGUE

Cette section présente nos travaux en cours pour développer un agent conversationnel basé sur des approches stochastiques pour des dialogues guidés par un but. On rappellera déjà que classiquement un moteur de dialogue est composé de trois composants principaux : le module de compréhension de la parole (*spoken language understanding* ou SLU) qui traduit les entrées verbales de l'utilisateur en une représentation interne de l'information pour le système, le module de gestion du dialogue qui contrôle le dialogue et décide des réponses, et le module de génération en langage naturel (*natural language generation* ou NLG) qui traduit la représentation interne de la réponse du système en une réponse en langage naturel compréhensible par l'utilisateur.

Ces dernières années, les approches stochastiques ont fait une percée pour le traitement du langage naturel et sont maintenant utilisées par des systèmes état de l'art pour la SLU [6, 9] et la gestion du dialogue [7, 8, 10]. La NLG utilise

toujours principalement des modèles à base de patrons et de règles, qui produisent de bons résultats pour une tâche spécifique. Récemment des approches stochastiques ont aussi été utilisées pour la NLG dans le but de régler deux principaux inconvénients des méthodes à base de patrons : le manque de passage à l'échelle pour des usages sur de larges domaines peu contraints et les fréquentes répétitions des réponses, qui confèrent à l'agent un ton "mécanique" peu engageant.

Dans de précédents travaux, nous avons proposé un nouveau modèle stochastique pour la NLG, basé sur les réseaux neuronaux récurrents déjà utilisés dans ce contexte par Wen *et al.* [16], ainsi qu'un protocole d'adaptation en ligne de ce modèle [15]. Nous procédons en deux étapes. Tout d'abord, un corpus initial réduit, constitué de références générées par patrons, est utilisé pour entraîner un modèle de génération. Puis, le modèle est à son tour utilisé pour générer des phrases, en interaction vocale avec l'utilisateur, auquel il est demandé ponctuellement de produire un énoncé plus adapté ou simplement différent.

## 2.1 Cas statique

Afin de réduire l'effort demandé à l'utilisateur et éviter des actions inutiles, nous utilisons un algorithme de bandit contre un adversaire pour décider si le système doit ou non demander l'expertise de l'utilisateur, selon les gains et coûts estimés de cette action. Une fois que le système a généré l'énoncé, il peut choisir une action (à partir d'une distribution de probabilité) parmi un ensemble  $\mathcal{I}$  de  $M$  actions. Dans une configuration préliminaire, nous prenons  $M = 3$  et  $\mathcal{I}$  est défini comme :  $\mathcal{I} := \{\text{Skip}, \text{AskDictation}, \text{AskTranscription}\}$ . Soit  $i \in \mathcal{I}$  l'indice de l'action, nous supposons que l'effort de l'utilisateur  $\phi(i) \in \mathbb{R}^+$ , peut être associé au temps nécessaire pour effectuer l'action  $i$ . Les **efforts estimés** associés à chaque action sont :

- **Skip** : n'applique aucune mise à jour au modèle. Le coût de cette action est nul ( $\phi(\text{Skip}) = 0$ ).
- **AskDictation** : affiner le modèle en considérant un énoncé alternatif proposé par l'utilisateur et transcrit automatiquement par un système de reconnaissance vocale. Dans nos travaux préliminaires, utilisant des corpus de dialogues pré-existants (comme celui décrit dans [17]), le système de reconnaissance vocale a été simulé par l'insertion automatique d'erreurs dans la transcription manuelle d'un énoncé disponible ( $\phi(\text{AskDictation}) = 1$ ). Depuis l'approche a été validée sur des données de transcription réelles.
- **AskTranscription** : demander à l'utilisateur de transcrire la correction ou l'énoncé alternatif ( $\phi(\text{AskTranscription}) = 1+l$ , avec  $l$  la taille de l'énoncé proposé).

Nous estimons ensuite le **gain** de l'action choisie  $g(i) \in [0, 1]$  comme suit :

- **Skip** : rien n'est appris, le gain est de 0 ( $g(\text{Skip}) = 0$ ).
- **AskDictation** : le gain est calculé en considérant la marge restante du score BLEU [12] de l'énoncé généré par le système, en utilisant l'énoncé proposé par l'utilisateur comme référence BLEU<sub>gen/prop</sub>. Pour tenir compte des erreurs potentielles ajoutées par le système de reconnaissance vocal, cette mesure est pénalisée par l'estimation de taux d'erreur :

$$g(\text{AskDictation}) = (1 - \text{BLEU}_{gen/prop}) \times (1 - \text{WER}) \times (1 - \text{SER}) \quad (1)$$

Le WER (*Word Error Rate*) donne le nombre moyen d'erreurs de transcription au niveau des mots (et peut-être simulé au besoin), tandis que le SER (*Slot Error Rate* [17]) évalue la présence des slots (concepts, valeurs) de l'acte de dialogue et pénalise les énoncés générés qui ne contiennent pas les informations sémantiques requises (du fait des erreurs de transcription).

- **AskTranscription** : cette action suppose de demander à l'utilisateur de transcrire manuellement l'énoncé pour éviter les erreurs de transcription automatique. Donc, l'estimation du gain ne tient compte que du score BLEU de l'énoncé généré par le système, en utilisant la phrase proposée par l'utilisateur comme référence ( $g(\text{AskTranscription}) = 1 - \text{BLEU}_{gen/prop}$ ). Enfin, une fonction de perte  $l(i) \in [0, 1]$  est définie afin que le système maximise la mesure de gain  $g(i)$  et minimise l'effort de l'utilisateur  $\phi(i)$  :

$$l(i) = \underbrace{\alpha(1 - g(i))}_{\text{amélioration du système}} + \underbrace{(1 - \alpha) \frac{\phi(i)}{\phi_{max}}}_{\text{effort de l'utilisateur}} \quad (2)$$

Le facteur  $\alpha$  permet de pondérer l'importance du gain par rapport au coût induit pour s'adapter aux préférences de l'utilisateur.  $\phi_{max}$  est quant à lui une constante de normalisation correspondant à l'effort estimé maximal.

## 2.2 Cas du bandit contre un adversaire

Nous introduisons ensuite le formalisme présenté dans un contexte qui permette de prendre en compte l'évolution de la situation dynamiquement (par exemple la stratégie de l'utilisateur qui peut apprendre à anticiper les erreurs de transcriptions). Pour cela nous considérons le scénario suivant du bandit contre un adversaire : à chaque itération  $t$ , le système produit une phrase puis choisit une action  $i_t \in \mathcal{I}$ . Une fois que l'action  $i_t$  est effectuée, le système calcule l'estimation du gain  $g(i_t)$ , l'effort de l'utilisateur  $\phi(i_t)$  et la perte  $l(i_t)$ . Le but de l'algorithme de bandit est alors de trouver  $i_1, i_2, \dots$ , afin que pour chaque  $t$ , le système minimise la perte  $l(i_t)$ .

Nous découpons l'apprentissage en blocs de  $n$  itérations chacun. À la fin de chaque bloc, nous ajoutons les énoncés proposés par l'utilisateur à notre corpus d'apprentissage et nous mettons à jour le modèle sur ce corpus étendu. Parallèlement, nous calculons la fonction de perte pour chaque choix de bandit et nous mettons à jour sa politique. Cette mise à jour du modèle NLG et de la politique décidant de l'action au bout de plusieurs itérations, plutôt qu'à chaque énoncé collecté, permet de réduire le temps de mise à jour du modèle et d'améliorer sa robustesse.

Ce protocole d'apprentissage en ligne nous a permis d'équilibrer automatiquement l'évolution des performances du système en tenant compte des préférences de l'utilisateur (aussi bien, potentiellement, pour corriger des imperfections ressenties de l'agent que pour introduire simplement de la variabilité). Ces travaux sont cependant préliminaires et nous sommes en train d'améliorer ces résultats, notamment en affinant les estimations de gains et de coûts en utilisant les sorties d'un vrai système de reconnaissance vocale pour obtenir des erreurs de transcription plus réalistes. Le but de cette recherche est donc de permettre à l'utilisateur d'agir aussi sur les générations de l'agent. Toutefois il ne pourra agir que sur des mécanismes de génération déjà intégrés sur l'agent, pas en créer. Aussi il est important d'élargir la base des capacités

interactionnelles de l'agent en y introduisant des mécanismes apparaissant lors des interactions langagières entre humains, et notamment ceux permettant d'affirmer leur rôle social.

### 3. INTÉGRATION D'UNE DIMENSION AFFECTIVE DANS UN AGENT CONVERSATIONNEL VIRTUEL : LA PRODUCTION HUMORISTIQUE

L'introduction de mécanismes reproduisant un sens de l'humour humain au sein d'agents conversationnels est une possibilité particulièrement intéressante pour améliorer leurs compétences sociales. De récentes études ont confirmé l'intérêt de capacités humoristiques, même imparfaites pour des systèmes artificiels [5, 19]. Bien entendu, nous ne prévoyons pas de reproduire une pleine capacité d'humour, telle que présente chez l'humain (et pas chez tous d'ailleurs!), qui est difficile aussi bien à définir qu'à décrire et qui requiert souvent une grande connaissance culturelle et contextuelle. Nous prévoyons plutôt d'identifier certains mécanismes réguliers et de les automatiser pour produire des événements conversationnels inattendus, rendant d'une part l'agent plus sympathique et offrant d'autre part une stratégie de médiation souple en cas de désalignement au cours du dialogue, et dans un second temps d'optimiser l'usage de ces mécanismes selon des critères de performances objectifs et subjectifs.

De précédents travaux en linguistique [13] ont permis d'établir une taxonomie possible des mécanismes d'humour dans les interactions sociales. Elle considère différents types de productions humoristiques. Par exemple, elles peuvent apparaître en réaction à une entrée de l'utilisateur ou être générées spontanément. Dans le premier cas, on détecte une opportunité dans la conversation et on y réagit en produisant de l'humour. C'est entre autres le cas de l'humour basé sur des mots polysémiques, qu'un interlocuteur peut prétendre comprendre dans un sens inattendu selon le contexte. Dans le second cas, on détecte un besoin de facilitation dans le dialogue, par exemple en cas d'incompréhension, et on produit alors une réponse humoristique indépendante, comme raconter une blague.

Les productions humoristiques en réaction au dialogue peuvent permettre une plus grande implication de l'utilisateur, puisqu'elles n'interrompent pas le dialogue. Elles sont cependant plus compliquées à produire car elles nécessitent une bonne compréhension du contexte et du sens des mots. En revanche, les productions de type génératives peuvent être vues simplement comme une tâche de classement où on cherche à sélectionner la meilleure insertion (blague) selon le contexte parmi une liste prédéfinie.

Trois modules de générations d'humour ont pour l'instant été explorés [4] (la configuration globale du système est présenté dans la Figure 1) :

- **Citation** : ce module trouve une citation humoristique en rapport avec l'entrée utilisateur, de manière à rester dans le contexte du dialogue. Un corpus de citations collecté en ligne a été indexé et une requête est construite à partir des mots-clés disponibles dans la phrase de l'utilisateur. Un score est donné à chaque citation extraite, puis elles sont classées en fonction de leur distance à un vecteur de contexte. Un historique est gardé pour éviter les répétitions. La citation est composée avec l'action dialogique suivante sélection-

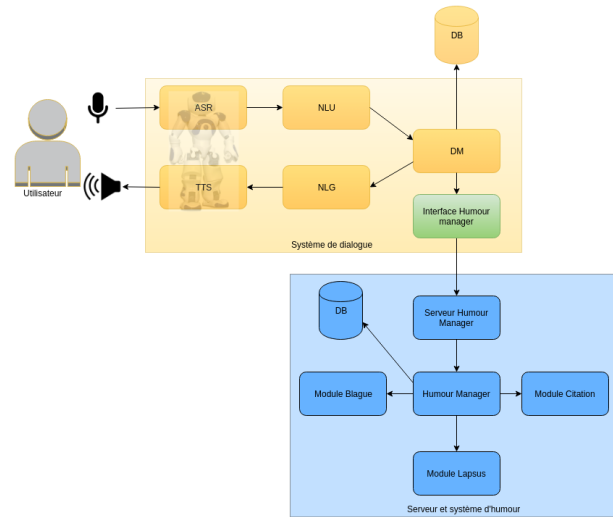


Figure 1: Configuration générale du système de dialogue et des modules de génération d'humour.

née par le gestionnaire de dialogue ;

- **Blague** : ce module a un fonctionnement proche du module de citation, le module retourne les blagues qui sont les plus proches du contexte et garde un historique pour éviter les répétitions. Les blagues sont plus longues que les citations et sont énoncées seules (elles supposent donc un retour de l'utilisateur avant un retour au flux principal du dialogue) ;
- **Lapsus** : son fonctionnement est quelque peu différent des deux autres modules. En effet le lapsus est trouvé non pas par rapport à la phrase de l'utilisateur mais par rapport à celle du système. Pour cela, on supprime les mots outils, pour ensuite y chercher un lapsus possible par rapport à la version phonétique des mots restants. Le mot sélectionné est alors remplacé par un mot d'un autre champ lexical pour avoir un effet surprenant. Une shortlist de mots-tabou a été pré-établie dans ce sens pour permettre de renforcer l'effet humoristique lorsque cela s'y prête. Le module de lapsus cherche donc à se rapprocher des qui-proquos que peuvent causer les correcteurs orthographiques sur smartphone. Par exemple : "Quelle gamme de cri (*prix*) souhaitez-vous? Modéré, élevé...?", "Quel type de poisson (*boisson*) cherchez-vous?"...

Cette première phase, introduction des trois nouveaux modules, a permis de donner une tonalité notablement différente aux interactions avec le système. Il reste qu'il est évidemment complexe de valider celle-ci par une évaluation de la qualité de l'effet produit, l'humour étant hautement subjectif l'intervalle de confiance d'une telle évaluation serait très grand. Aussi il nous semble plus prometteur de progresser sur la voie de la capacité d'adaptation de cette nouvelle compétence aux retours immédiats de l'utilisateur.

Ainsi dans une seconde phase, nous nous sommes intéressés à l'utilisation de techniques d'apprentissage par renforcement pour l'optimisation de cette nouvelle capacité. Puisque chaque utilisateur peut apprécier ou rejeter l'humour de l'agent conversationnel, nous souhaitons permettre au système d'adapter son utilisation des mécanismes humoristiques, à la fois en quantité et en qualité (quel type d'hu-

mour est apprécié, lequel ne l'est pas, à quel dose...).

L'apprentissage par renforcement, mais aussi des approches à base d'algorithme de bandit comme celle présentée dans la section 2 peuvent être utilisés pour apprendre les préférences d'un utilisateur, notamment le type d'humour à utiliser, la quantité d'humour par dialogue, tout en déterminant les moments opportuns lors des échanges pour produire de l'humour. Ces méthodes doivent permettre d'apprendre des comportements pour des profils types d'utilisateurs, mais aussi de s'adapter à un utilisateur spécifique, y compris en cours d'interaction.

Les différents types d'humour implémentés ont déjà été étudiés dans de précédents travaux [1] et automatisés avec des outils existants de traitement du langage naturel. En premier lieu construits textuellement, ils doivent encore être affinés pour le domaine oral, par exemple en modifiant la prosodie de la synthèse vocale ou en ajoutant des mouvements physiques pour accentuer l'intensité de l'intervention humoristique en cas d'incarnation physique de l'agent. L'application de techniques d'apprentissage de renforcement pour l'élaboration de stratégies différenciées d'humour est en cours de réalisation et devra permettre de passer à l'étape cruciale de l'évaluation.

#### 4. ÉVALUATION PHYSIOLOGIQUE CONTRASTIVE DES INTERACTIONS HUMAINES ET HUMAIN-MACHINE

Un des principaux problèmes avec l'ajout d'une compétence sociale dans un agent conversationnel est son évaluation. Des mesures comme le BLEU-4 [12] nous permettent de donner un score à une réponse du système par rapport à des phrases de références, tandis qu'on peut valider le système de dialogue dans sa globalité s'il permet d'atteindre l'objectif de ce dialogue de manière répétée (s'il a un objectif, par exemple une réservation d'hôtel ou la résolution du problème technique d'un utilisateur). Cependant, ce ne sont que des métriques objectives qui sont mieux adaptées à des systèmes guidés par la tâche et qui ne mesurent pas l'appréciation subjective du système par l'utilisateur. Et ce alors même que cette évaluation subjective est encore plus intéressante pour un système intégrant des productions humoristiques, puisque nous supposons que ces productions vont améliorer les interactions sociales et donc impacter positivement la perception qu'a l'utilisateur des capacités globales de l'agent.

Aussi pour mesurer cette perception, nous avons prévu d'utiliser des mesures physiologiques, la conductance cutanée (système nerveux périphérique) et l'imagerie cérébrale (système nerveux central) afin de comparer les interactions humain-humain et humain-machine durant une conversation naturelle [2]. Nous pourrions notamment comparer les réponses physiologiques face à une machine selon si les mécanismes d'humour sont utilisés par le système ou non. De telles mesures peuvent nous fournir des informations très précises pour évaluer finement nos systèmes, puisqu'elles informent directement des réactions émotionnelles de l'utilisateur [3]. De plus, en utilisant différents systèmes, capables ou non d'humour et ce à différents niveaux, les résultats devront permettre de déterminer si, et comment, les humains modifient leur perception d'une machine pendant une interaction et si, éventuellement, ils adoptent alors une posture spécifique [18].

## 5. CONCLUSION

Dans cet article, nous avons présenté différentes possibilités d'intégration de mécanismes de productions humoristiques dans un agent conversationnel virtuel oral. Cette intégration est prévue en deux étapes. D'abord, une identification et automatisation de mécanismes réguliers de productions humoristiques. Puis une optimisation de ces mécanismes en utilisant des approches d'apprentissage par renforcements. Pour évaluer la compétence sociale des agents conversationnels intégrant de l'humour, nous ne pouvons pas nous restreindre aux métriques objectives mais nous avons aussi besoin d'évaluations subjectives des ressentis de l'utilisateur. Pour effectuer ces évaluations, nous avons prévu d'utiliser des approches des sciences cognitives, et en particulier des neurosciences. Mieux associer les sciences cognitives et les agents conversationnels virtuels devra permettre à la fois d'améliorer nos agents conversationnels mais aussi par contraste d'améliorer nos connaissances à propos des interactions sociales humaines.

## Remerciements

Ce travail a été partiellement financé par le Labex BLRI (ANR-11-LABX-0036) et l'ILCB.

## 6. REFERENCES

- [1] X. Bost. De l'humour pour les systèmes d'interactions vocales (Humor in vocal interaction systems), 2013. Master Thesis.
- [2] T. Chaminade, L. Biaocchi, F. H. Wolfe, N. Nguyen, and L. Prévot. Communicative behavior and physiology in social interactions. In *Proceedings of the 1st Workshop on Modeling INTERPERSONAL Synchrony And Influence*, pages 25–30. ACM, 2015.
- [3] T. Chaminade, M. Zecca, S. J. Blakemore, A. Takanishi, C. D. Frith, S. Micera, P. Dario, G. Rizzolatti, V. Gallese, and M. A. Umiltà. Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PLoS One*, 5(7):e11577, 2010.
- [4] T. Desfrancois. Apprentissage automatique d'humour pour les systèmes de dialogues vocaux (automatic learning of humor production for the vocal dialogue systems), 2016. Master Thesis.
- [5] G. D. Duplessis, L. Béchade, M. Sehili, A. Delaborde, V. Letard, A.-L. Ligozat, P. Deléglise, Y. Estève, S. Rosset, and L. Devillers. Nao is doing humor in the chist-era joker project. In *16th Interspeech*, pages 1072–1073, 2015.
- [6] E. Ferreira, B. Jabaian, and F. Lefèvre. Zero-shot semantic parser for spoken language understanding. In *INTERSPEECH*, 2015.
- [7] E. Ferreira and F. Lefèvre. Reinforcement-learning based dialogue system for human-robot interactions with socially-inspired rewards. *Computer Speech & Language*, 34(1):256–274, 2015.
- [8] E. Ferreira, G. Milliez, F. Lefèvre, and R. Alami. Users' belief awareness in reinforcement learning-based situated human-robot dialogue management. In *IWSDS*, 2015.
- [9] E. Ferreira, A. Reiffers-Masson, B. Jabaian, and F. Lefèvre. Adversarial bandit for online interactive

- active learning of zero-shot spoken language understanding. In *ICASSP*, 2016.
- [10] G. Milliez, E. Ferreira, M. Fiore, R. Alami, and F. Lefèvre. Simulating human-robot interactions for dialogue strategy learning. In *SIMPAR*, 2014.
- [11] M. Ochs, R. Niewiadomski, C. Pelachaud, and D. Sadek. Intelligent expressions of emotions. *Affective computing and intelligent interaction*, pages 707–714, 2005.
- [12] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. Bleu : a method for automatic evaluation of machine translation. In *ACL*, 2002.
- [13] B. Priego-Valverde. *L'humour dans les interactions conversationnelles : jeux et enjeux*. PhD thesis, Aix-Marseille 1, 1999.
- [14] O. Rambow, S. Bangalore, and M. Walker. Natural language generation in dialog systems. In *HLT*, 2001.
- [15] M. Riou, B. Jabaian, S. Huet, and F. Lefèvre. Online adaptation of an attention-based neural network for natural language generation. In *INTERSPEECH*, 2017.
- [16] T.-H. Wen, M. Gašić, D. Kim, N. Mrkšić, P.-H. Su, D. Vandyke, and S. Young. Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking. In *SIGDIAL*, 2015.
- [17] T.-H. Wen, M. Gašić, N. Mrkšić, P.-H. Su, D. Vandyke, and S. Young. Semantically conditioned LSTM-based natural language generation for spoken dialogue systems. In *EMNLP*, 2015.
- [18] A. Wykowska, T. Chaminade, and G. Cheng. Embodied artificial agents for understanding human social cognition. *Phil. Trans. R. Soc. B*, 371(1693) :20150375, 2016.
- [19] D. Yang, A. Lavie, C. Dyer, and E. Hovy. Humor recognition and humor anchor extraction. In *EMNLP*, 2015.