

Effet de l'anthropomorphisme des machines sur le français adressé aux robots: Étude du débit de parole et de la fluence

Natalia Kalashnikova¹ Mathilde Hutin^{1,2} Ioana Vasilescu¹
Laurence Devillers^{1,3}

(1) Université Paris-Saclay, CNRS, LISN, 91400, Orsay, France

(2) Institut Langage & Communication, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgique

(3) Sorbonne Université, Paris, France

{natalia.kalashnikova,hutin,ioana,devil}@lisn.upsaclay.fr

RÉSUMÉ

"Robot-directed speech" désigne la parole adressée à un appareil robotique, des petites enceintes domestiques aux robots humanoïdes grandeur-nature. Les études passées ont analysé les propriétés phonétiques et linguistiques de ce type de parole ou encore l'effet de l'anthropomorphisme des appareils sur la sociabilité des interactions, mais l'effet de l'anthropomorphisme sur les réalisations linguistiques n'a encore jamais été exploré. Notre étude propose de combler ce manque avec l'analyse d'un paramètre phonétique (débit de parole) et d'un paramètre linguistique (fréquence des pauses remplies) sur la parole adressée à l'enceinte *vs* au robot humanoïde *vs* à l'humain. Les données de 71 francophones natifs indiquent que les énoncés adressés aux humains sont plus longs, plus rapides et plus dysfluents que ceux adressés à l'enceinte et au robot. La parole adressée à l'enceinte et au robot est significativement différente de la parole adressée à l'humain, mais pas l'une de l'autre, indiquant l'existence d'un type particulier de la parole adressée aux machines.

ABSTRACT

The Effect of Human-Likeness in French Robot-Directed Speech : A study of Speech Rate and Fluency

Robot-directed speech refers to speech to a robotic device, ranging from small home speakers to full-size humanoid robots. Studies have investigated the phonetic and linguistic properties of this type of speech or the effect of anthropomorphism of the devices on the social aspect of interaction. However, none have investigated the effect of the device's human-likeness on linguistic realizations. This preliminary study proposes to fill this gap by investigating one phonetic parameter (speech rate) and one linguistic parameter (use of filled pauses) in speech directed at a home speaker *vs* a humanoid robot *vs* a human. The data from 71 native speakers of French indicate that human-directed speech shows longer utterances at a faster speech rate and more filled pauses than speech directed at a home speaker and a robot. Speaker- and robot-directed speech is significantly different from human-directed speech, but not from each other, indicating a unique device-directed type of speech.

MOTS-CLÉS : parole adressée aux robots, interaction humain-machine, débit de parole, pauses remplies, anthropomorphisme.

KEYWORDS: robot-directed speech, human-computer interaction, speech rate, filled pause, human-likeness.

1 Introduction

Robot-directed speech (RDS), littéralement "parole adressée au robot", désigne les productions langagières d'un humain à un outil robotique, qu'il s'agisse de petites enceintes intelligentes ou de robots humanoïdes grandeur nature. Ce type de parole peut être comparé à *computer-directed speech* ("parole adressée à l'ordinateur") en ce qu'il appartient lui aussi à la catégorie du *device-directed speech* ("parole adressée à la machine"), dans la catégorie des registres langagiers spéciaux, tels que la parole adressée aux enfants, aux locuteurs non-natifs, ou même aux animaux. L'exploration de tels styles de parole permet non seulement une meilleure compréhension des interactions humain-machine, mais aussi de l'ajustement dans le dialogue et de l'accommodation entre les interlocuteurs, ce qui à son tour peut permettre de développer des outils concrets tels que la détection de destinataire (*addressee-detection*).

Certaines études ont spécifiquement analysé les propriétés acoustiques de la parole adressée à l'ordinateur ou au robot à celle adressée aux adultes. Ces travaux de recherche ont montré que, lorsqu'ils s'adressent aux ordinateurs, les humains ont tendance à produire plus d'énoncés (Amalberti *et al.*, 1993) et à hyperarticuler les voyelles (en termes de formants et de durée) sans avoir une fréquence fondamentale (F0) plus élevée (Burnham *et al.*, 2010). En revanche, dans la parole adressée aux robots, les voyelles sont aussi hyperarticulées (Kriz *et al.*, 2010) mais le débit de parole montre peu de différence (Raveh *et al.*, 2019) et la F0 et l'intensité sont plus élevées (Kriz *et al.*, 2010; Raveh *et al.*, 2019). En ce qui concerne les particularités linguistiques de la communication avec les ordinateurs, les humains ont tendance à contrôler et simplifier leur usage du langage (moins de dysfluences, quantité limitée d'informations par proposition, etc.) : en particulier, les locuteurs utilisent moins de pauses remplies (comme *uh* ou *um*) (Amalberti *et al.*, 1993). Toutes ces études indiquent également que les différences entre la parole adressée à un appareil et la parole adressée à un humain diminuent au cours de la conversation (Amalberti *et al.*, 1993) et que les adultes montrent plus de variation intra-locuteur dans la parole adressée à un appareil que dans la parole adressée à un enfant (Kriz *et al.*, 2009; Fischer *et al.*, 2011).

Ces différences entre la parole adressée à l'ordinateur et la parole adressée au robot laissent supposer que les appareils ne sont pas tous considérés de la même façon par les humains. De plus, Gong (2008) a montré que, sur une échelle de quatre niveaux d'anthropomorphisme, plus l'appareil ressemble à un humain, plus il reçoit de réponses sociales de la part des utilisateurs lors de l'interaction. Ces résultats ont été confirmés d'un point de vue neurologique, puisque plus l'appareil est anthropomorphique, plus l'activité corticale est forte dans les régions du cerveau associées au raisonnement à propos de l'intention d'autrui (la théorie de l'esprit) (Krach *et al.*, 2008).

Dans le présent article, nous présentons les résultats préliminaires d'une étude plus large en opposant les participants dialoguant avec une enceinte Google Home vs un robot Pepper vs un autre humain. Plus précisément, nous analysons les facteurs suivants :

- les **facteurs phonétiques** : si les participants produisent (i) des énoncés (en l'occurrence des tours de parole) plus longs ; (ii) avec un débit de parole moins rapide pour la condition enceinte vs humain ;
- les **facteurs linguistiques** : si les participants produisent (iii) moins de pauses remplies pour la condition enceinte vs humain ;
- la **variation** : (iv) si ces mesures diffèrent plus intra-locuteurs dans la parole adressée à l'enceinte que dans la parole adressée à l'humain et (v) évoluent au cours de la conversation ;
- l'**effet de l'anthropomorphisme** : (vi) si la parole adressée au robot Pepper ressemble plus à

la parole adressée à l'humain ou à la parole adressée à l'enceinte (ou se positionne quelque part entre les deux).

2 Méthode

2.1 Protocole expérimental

Procédure. D'abord, les membres de notre équipe expliquent le déroulé de l'expérience aux participants, qui signent la notice de consentement et remplissent un formulaire de questions de base. Dans ce questionnaire, ils notent leur degré de volonté à adopter une série de 8 habitudes écologiques. Ensuite, ils suivent un ou une bénévole dans l'une des 3 salles (qui correspondent chacune à un agent conversationnel), où deux membres de notre équipe contrôlent la configuration de l'expérience. À la fin de l'enregistrement, les expérimentateurs remercient les participants et les invitent à retourner auprès des organisateurs où ils peuvent poser leurs questions sur l'expérience en prenant une collation. Le déroulement de l'expérience a été validé par un comité d'éthique et a respecté les mesures sanitaires contre le covid-19.

Enregistrements. La procédure d'enregistrement a été inspirée de l'étude de [Mehenni *et al.* \(2020\)](#). Lors de l'enregistrement, l'agent conversationnel (enceinte, robot Pepper ou humain) pose des questions sur les habitudes écologiques des participants. Les protocoles de l'enceinte et du robot Pepper sont réalisés dans un dispositif en Magicien d'Oz, c'est-à-dire que les réponses pré-enregistrées sont envoyées à la machine par un des expérimentateurs sans que le participant s'en rende compte. La voix de synthèse fournie par les paramètres par défaut du robot Pepper est une voix enfantine, qui a été enregistrée séparément afin de pouvoir l'utiliser aussi pour l'enceinte. La condition de la parole adressée à l'humain est réalisée par un des membres de notre équipe, qui lit à haute voix le même script qui a été enregistré pour le robot et l'enceinte. L'échange oral consiste en 4 étapes :

- S0 : L'agent instaure la conversation avec le participant en posant quelques questions anodines.
- S1 : L'agent présente les situations hypothétiques dans lesquelles les participants doivent choisir entre l'option par défaut et l'option respectueuse de l'environnement qui demande plus d'investissement (d'argent ou de temps).
- S2 : L'agent fournit une information présentant les conséquences sur l'environnement de chaque habitude et pose cette fois les questions de base du formulaire écrit.
- S3 : L'agent présente d'autres situations hypothétiques et les participants doivent à nouveau choisir entre l'option par défaut et l'option respectueuse de l'environnement.

Les données audio ont été enregistrées avec le microphone unidirectionnel (AKG45) sur Audacity à 44.1 kHz.

Participants. En avril et juin 2022, notre équipe a recruté les visiteurs, le personnel et les étudiants de Collège des Bernardins à Paris, France. Les 71 locuteurs natifs du français (46 femmes, 25 hommes) se répartissent de façon équilibrée dans les groupes d'âge (de 18 à 65+ ans). Parmi tous les participants, 21 (16 femmes, 5 hommes) ont participé à la condition avec l'agent humain, 28 (18 femmes, 10 hommes) à la condition avec l'agent robot, et 22 (12 femmes, 10 hommes) à la condition avec l'agent enceinte. Au total, 16 heures de parole ont été enregistrées.

2.2 Méthodologie

Les fichiers audio ont été segmentés manuellement et transcrits en français par deux annotateurs. Les pauses remplies ont été annotées dans la transcription. Les fichiers finaux de transcription contiennent l'horodatage des tours de parole. Tous les calculs sont faits avec Python 3 (Van Rossum & Drake, 2009).

Nous utilisons ces fichiers d'annotation pour calculer la durée totale des énoncés en secondes pour chaque tour de parole de chaque participant. Pour le calcul du débit de parole, nous utilisons l'espace entre les mots de la transcription pour segmenter le discours en unités de type (pseudo-)mots. Nous calculons ensuite la somme des unités pour chaque étape de l'expérience et divisons cette somme par la durée de l'étape. Ainsi, nous obtenons le débit de parole exprimé en nombre d'unités par seconde.

Pour les pauses remplies, nous calculons le nombre total des pauses remplies et divisons ce nombre par la durée de tours de parole pour obtenir le nombre de pauses remplies par seconde.

Pour l'analyse de la variation intra-individuelle, nous calculons l'écart-type de chaque paramètre pour chaque participant et ensuite la moyenne des écarts-types pour chaque condition.

La significativité de nos résultats est testée avec un *t-test* pour les deux échantillons indépendants appliqués avec SciPy (Virtanen *et al.*, 2020). Nous avons ensuite appliqué la correction de Bonferroni (seuil $\alpha = 0.5$) avec Statsmodels (Seabold & Perktold, 2010).

3 Résultats

3.1 Longueur de l'énoncé et débit de parole

De façon générale, les participants ont tendance à produire des énoncés plus longs en parlant avec un humain (moyenne = 372 sec.) qu'en parlant avec l'enceinte (moyenne = 191 sec., $\Delta=181$ sec., $t=2,19$, $p=0,09$) et surtout au robot humanoïde (moyenne = 147 sec., $\Delta=225$ sec., $t=2,87$, $p=0,01$). La durée des énoncés de la parole adressée au robot et à l'enceinte n'est pas significativement différente ($\Delta=44$ sec., $t=-1,08$, $p=0,85$). En regardant chaque étape séparément, comme dans la Figure 1, il apparaît

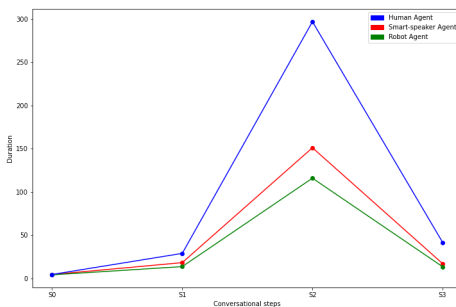


FIGURE 1 – Durée moyenne des énoncés de la parole adressée à l'humain (bleu), à l'enceinte (rouge) et au robot humanoïde (vert) à chaque étape de la conversation.

que les seules différences significatives sont observées entre la parole adressée à l'humain et celle adressée au robot dans les étapes "S2" ($\Delta=181,15$ sec., $p=0,01$), et "S3" ($\Delta=28,14$ sec., $p=0,02$).

Contrairement à la littérature comparant la parole adressée à l'appareil et l'enfant (Kriz *et al.*, 2009; Fischer *et al.*, 2011), les participants montrent plus de variation intra-locuteur dans la condition de la parole adressée à l'humain, avec un écart-type à travers des participants de 137,54 secondes, que dans la parole adressée à l'enceinte (écart-type moyen = 69,57, $\Delta=67,97$ sec., $t=2,47$, $p=0,06$). Ce paramètre est significativement différent entre la parole adressée à l'humain et au robot (écart-type moyen = 53,18 sec., $\Delta=84,36$ sec., $t=3,24$, $p=0,004$). Néanmoins, les différences entre la parole adressée à l'enceinte et au robot ne sont pas significativement différents l'un de l'autre ($\Delta=16,39$, $t=-1,3$, $p=0,2$).

En ce qui concerne le débit de parole (Figure 2), la parole adressée à l'humain est plus lente au début mais s'accélère autour de l'étape S1, au moment où la conversation commence réellement. De manière générale, les participants s'adressent à l'humain avec un débit de 2,96 unités/sec, au robot avec un débit de 2,63 unités/sec, et à l'enceinte avec un débit de 2,5 unités/sec. Comme attendu, la différence entre la parole adressée à l'humain et la parole adressée à l'enceinte est significative ($\Delta=0,46$ unités/sec, $t=3,43$, $p=0,004$). Contrairement à Raveh *et al.* (2019), la différence entre la parole adressée à l'humain et la parole adressée au robot est aussi significative ($\Delta=0,32$ unités/sec, $t=2,8$, $p=0,009$), mais pas entre la parole adressée au robot et la parole adressée à l'enceinte ($\Delta=0,14$ unités/sec, $t=1,32$, $p=0,58$). L'analyse de chaque étape montre que les deux différences significatives

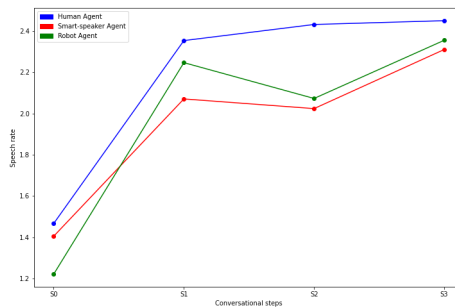


FIGURE 2 – Débit de parole moyen des énoncés de la parole adressée à l'humain (bleu), à l'enceinte (rouge) et au robot humanoïde (vert) à chaque étape de la conversation.

du débit de parole sont observées à l'étape "S2" dans les conditions entre l'humain et l'enceinte ($\Delta=0,4$ unités/sec., $p=0,002$), et entre l'humain et le robot sur la même étape ($\Delta=0,36$ token/sec., $p=0,003$). Aucune différence significative n'a été observée entre les étapes de la parole adressée au robot et la parole adressée à l'enceinte.

Contrairement à ce qui était attendu, on observe moins de variation intra-locuteur dans la parole adressée à l'enceinte, avec un écart-type moyen de 0,6 unités/sec. contre un écart-type moyen de 0,68 unités/sec. dans la parole adressée au robot ($\Delta=-0,08$ unités/sec., $t=0,92$, $p>1$.) et un écart-type moyen de 0,62 unités/sec dans la parole adressée à l'humain ($\Delta=-0,02$ unités/sec., $t=0,19$, $p>1$.) La différence entre les paroles adressées au robot et à l'humain n'est pas significative ($\Delta=0,06$ unités/sec., $t=-0,7$, $p>1$.)

Ces résultats indiquent que l'anthropomorphisme des appareils impacte de façon modérée la longueur

des énoncés et le débit de parole. Les participants ont tendance à parler plus longtemps à l'humain qu'au robot, mais pas à l'encontre, tandis qu'ils parlent plus rapidement à l'humain qu'à l'encontre, et surtout plus rapidement qu'au robot. Cependant, lorsqu'on compare les résultats pour le robot et l'encontre, aucune différence de longueur des énoncés ni de débit de la parole n'est observée.

3.2 Pauses remplies

Les pauses remplies sont des instances de dysfluences, des accidents dans la production de la parole qui sont extrêmement fréquents dans les interactions humain-humain (Shriberg, 1994). De façon générale, les participants produisent significativement moins de pauses remplies quand ils parlent à l'encontre que quand ils parlent à l'humain ($\Delta=-0,4$, $t=2,92$, $p=0,01$) et au robot ($\Delta=-0,29$, $t=2,62$, $p=0,03$). Cependant, la différence de fréquence des pauses remplies entre la parole adressée à l'humain et la parole adressée au robot est non-significative ($\Delta=0,11$, $t=0,8$, $p>1$).

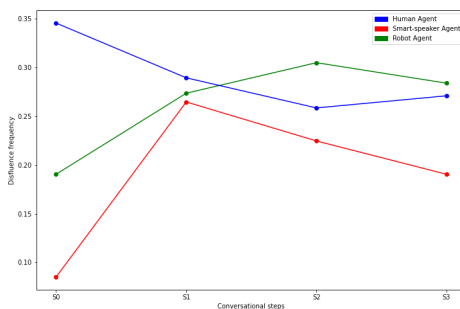


FIGURE 3 – Score moyen des disfluences des énoncés de la parole adressée à l'humain (bleu), à l'encontre (rouge) et au robot humanoïde (vert) à chaque étape de la conversation.

Cette situation évolue avec le temps (Figure 3). Au début de la conversation, surtout à l'étape "S0" (échange de banalités), les participants produisent plus des pauses remplies quand ils parlent à l'humain (0,35 pauses remplies par seconde), que quand ils parlent au robot (0,19 pauses/sec., $\Delta=0,16$ pauses/sec., $p=0,4$), ou surtout à l'encontre (0,09 pauses/sec., $\Delta=0,26$ pauses/sec., $p=0,04$). A la fin de l'interaction, par exemple à l'étape "S3", les participants produisent presque autant de pauses remplies quand ils parlent à l'humain (0,27 pauses/sec.), que quand ils parlent au robot (0,28 pauses/sec., $\Delta=0,01$ pauses/sec., $p>1$), mais plus quand ils parlent à l'encontre (0,19 pauses/sec., $\Delta=0,08$, $p=0,4$).

La variation intra-locuteur est similaire dans les trois conditions, avec un écart-type de 0,17, 0,17, ($\Delta=0$, $t=0,004$, $p=1$) et 0,12 ($\Delta=0,05$, $t=1,1$, $p=0,28$) pauses remplies par seconde dans la parole adressée respectivement à l'humain, au robot et à l'encontre.

Ces résultats en accord avec les recherches passées (Amalberti *et al.*, 1993) indiquent que les participants produisent plus de dysfluences dans la parole adressée à l'humain que dans la parole adressée à un appareil, mais cet écart diminue avec le temps. Cependant, la parole adressée au robot n'est pas significativement différente de la parole adressée ni à l'humain ni à l'encontre.

4 Conclusion et discussion

Dans cet article, nous avons analysé 16 heures de parole de 71 locuteurs natifs du français pour étudier le débit de parole et l'utilisation des pauses remplies dans la parole adressée à l'humain, à l'enceinte et au robot. Partiellement conforme aux recherches passées sur la parole adressée à l'ordinateur (Amalberti *et al.*, 1993), nous montrons que les participants parlent plus vite et produisent plus de pauses remplies en parlant avec l'humain qu'avec l'enceinte. En ce qui concerne le robot humanoïde, en se basant sur les études préconisant l'effet de l'anthropomorphisme sur les interactions parlées (Gong, 2008; Krach *et al.*, 2008), nous avons fait l'hypothèse que la parole adressée au robot partagerait plus de caractéristiques avec la parole adressée à l'humain qu'avec la parole adressée à l'enceinte. Cependant, les participants ont produit des énoncés plus longs de la façon plus rapide avec plus de pauses remplies dans la parole adressée à l'humain que dans la parole adressée au robot. De plus, la parole adressée au robot et celle adressée à l'enceinte ne sont pas différentes en longueur de l'énoncé, ni en débit de parole et en fréquence de dysfluences. Nous nous attendions également à plus de variation intra-locuteur dans la parole adressée à l'appareil qu'à l'humain : nous trouvons la tendance inverse pour la longueur des énoncés, et aucune différence pour la fréquence des pauses remplies et le débit de la parole. De façon générale, la parole adressée à l'humain est opposée à la parole adressée à l'enceinte et la parole adressée au robot.

La similarité entre les paroles adressées au robot et à l'enceinte peut être due au fait que les deux appareils communiquent avec une voix d'enfant, tandis que l'agent conversationnel humain était adulte. Il est possible que les participants se soient davantage alignés avec la parole adressée à l'enfant qu'avec la parole adressée à l'appareil. Cependant, les conditions du robot et de l'enceinte sont comparables mais ne montrent aucune différence significative, indiquant que le comportement des participants peut être considéré comme un type unique de "parole adressée à l'appareil". Enfin, le robot et l'enceinte ont utilisé la même voix alors que l'agent humain a été incarné à tour de rôle par 3 femmes et 1 homme. Il est possible que les données de la condition avec l'agent humain aient été impactées par la différence entre la parole adressée à la femme et celle adressée à l'homme.

Les futures analyses devront aussi prendre en compte l'âge des participants et leur utilisation quotidienne des appareils robotiques pour établir l'effet de la familiarité et de l'habitude. Dans les futures étapes de cette recherche, nous envisageons d'étudier les propriétés linguistiques du discours, tels que les marqueurs discursifs, la complexité de la syntaxe, etc. Nous comptons également analyser la similarité des caractéristiques phonétiques et linguistiques entre la parole des agents et la parole des sujets afin d'étudier les mécanismes de l'alignement linguistique pendant la conversation.

Remerciements

Les auteurs remercient les relecteurs pour leurs commentaires et leurs suggestions. Cet article est écrit dans le cadre de la thèse financée par Chaire AI HUMAINE (ANR-19-CHIA-0019) dirigé par Laurence Devillers.

Références

AMALBERTI R., CARBONELL N. & FALZON P. (1993). User representations of computer systems

in human-computer speech interaction. *International Journal of Man-Machine Studies*, **38**(4), 547–566. DOI : <https://doi.org/10.1006/imms.1993.1026>.

BENAMARA F., HATOUT N., MULLER P. & OZDOWSKA S., Éd.s. (2007). *Actes de TALN 2007 (Traitement automatique des langues naturelles)*, Toulouse. ATALA, IRIT.

BURNHAM D., JOEFFRY S. & RICE L. (2010). Computer-and human-directed speech before and after correction. In *Proceedings Of The 13Th Australasian International Conference On Speech Science And Technology*, volume 6, p. 13–17.

DIAS G., Éd. (2015). *Actes de TALN 2015 (Traitement automatique des langues naturelles)*, Caen. ATALA, HULTECH.

FISCHER K., FOTH K., ROHLFING K. J. & WREDE B. (2011). Mindful tutors : Linguistic choice and action demonstration in speech to infants and a simulated robot. *Interaction Studies*, **12**, 134–161. DOI : <https://doi.org/10.1075/is.12.1.06fis>.

GONG L. (2008). How social is social responses to computers ? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, **24**(4), 1494–1509. Including the Special Issue : Integration of Human Factors in Networked Computing, DOI : <https://doi.org/10.1016/j.chb.2007.05.007>.

KRACH S., HEGEL F., WREDE B., SAGERER G., BINKOFSKI F. & KIRCHER T. (2008). Can Machines Think ? Interaction and Perspective Taking with Robots Investigated via fMRI. *PLoS ONE*, **3**(7), 1494–1509. Including the Special Issue : Integration of Human Factors in Networked Computing, DOI : <https://doi.org/10.1371/journal.pone.0002597>.

KRIZ S., ANDERSON G., BUGAJSKA M. & TRAFTON J. G. (2009). Robot-directed speech as a means of exploring conceptualizations of robots. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, HRI '09, p. 271–272, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/1514095.1514171](https://doi.org/10.1145/1514095.1514171).

KRIZ S., ANDERSON G. & TRAFTON J. G. (2010). Robot-directed speech : Using language to assess first-time users' conceptualizations of a robot. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, p. 267–274. DOI : [10.1109/HRI.2010.5453187](https://doi.org/10.1109/HRI.2010.5453187).

LAIGNELET M. & RIOULT F. (2009). Repérer automatiquement les segments obsolètes à l'aide d'indices sémantiques et discursifs. In A. NAZARENKO & T. POIBEAU, Éd.s., *Actes de TALN 2009 (Traitement automatique des langues naturelles)*, Senlis : ATALA LIPN.

LANGLAIS P. & PATRY A. (2007). Enrichissement d'un lexique bilingue par analogie. In (Benamara et al., 2007), p. 101–110.

MEHENNI H. A., KOPYLYANSKAYA S., VASILESCU I. & DEVILLERS L. (2020). Nudges with conversational agents and social robots : a first experiment with children at a primary school. In *11th International Workshop on Spoken Dialog System Technology*, Madrid, Spain. HAL : [hal-03083526](https://hal.archives-ouvertes.fr/hal-03083526).

RAVEH E., STEINER I., SIEGERT I., GESSINGER I. & MÖBIUS B. (2019). Comparing phonetic changes in computer-directed and human-directed speech. In P. BIRKHOLZ & S. STONE, Éd.s., *Studentexte zur Sprachkommunikation : Elektronische Sprachsignalverarbeitung 2019*, p. 42–49 : TUDpress, Dresden.

SEABOLD S. & PERKTOLD J. (2010). statsmodels : Econometric and statistical modeling with python. In *9th Python in Science Conference*.

SERETAN V. & WEHRLI E. (2007). Collocation translation based on sentence alignment and parsing. In (Benamara et al., 2007), p. 401–410.

SHRIBERG E. (1994). *Preliminaries to a Theory of Speech Disfluencies*. Thèse de doctorat, University of California, Berkeley.

VAN ROSSUM G. & DRAKE F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA : CreateSpace.

VIRTANEN P., GOMMERS R., OLIPHANT T. E., HABERLAND M., REDDY T., COURNAPEAU D., BUROVSKI E., PETERSON P., WECKESSER W., BRIGHT J., VAN DER WALT S. J., BRETT M., WILSON J., MILLMAN K. J., MAYOROV N., NELSON A. R. J., JONES E., KERN R., LARSON E., CAREY C. J., POLAT İ., FENG Y., MOORE E. W., VANDERPLAS J., LAXALDE D., PERKTOLD J., CIMRMAN R., HENRIKSEN I., QUINTERO E. A., HARRIS C. R., ARCHIBALD A. M., RIBEIRO A. H., PEDREGOSA F., VAN MULBREGT P. & SCIPY 1.0 CONTRIBUTORS (2020). SciPy 1.0 : Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, **17**, 261–272. DOI : [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).