

---

# Introduction au numéro spécial sur le traitement automatique du langage parlé

**Laurent Besacier\*** — **Wolfgang Minker\*\***

\* *Univ. Grenoble Alpes, LIG, Grenoble, France*  
*laurent.besacier@imag.fr*

\*\* *Univ. of Ulm, Institute of Information Technology, Ulm, Germany*  
*wolfgang.minker@uni-ulm.de*

---

*RÉSUMÉ. Cette introduction présente l'appel à communications de ce numéro spécial, le processus de relecture ainsi qu'un bref résumé des contributions acceptées.*

*ABSTRACT. This introduction presents the call for papers for this special issue, the review process and a brief summary of the contributions.*

*MOTS-CLÉS : traitement automatique du langage parlé, reconnaissance automatique de la parole, traduction de la parole, corpus oraux, dialectes oraux, détection d'émotions, intelligibilité de la parole, corpus de dialogue, détection de coréférences, apprentissage automatique.*

*KEYWORDS: spoken language processing, speech recognition, speech translation, spoken language corpora, spoken dialects, emotion detection, speech intelligibility, dialogue corpus, coreference resolution, machine learning.*

---

## 1. Introduction

La communication orale reste le moyen le plus naturel pour dialoguer et interagir avec la machine ou avec une autre personne. Le traitement automatique du langage parlé (TALP) et le dialogue trouvent désormais de nombreuses applications directes dans des domaines divers tels que la recherche d'information, l'interaction en langue naturelle avec des dispositifs mobiles, la robotique sociale, les technologies d'assistance à la personne, l'apprentissage des langues, etc. Cependant, le TALP pose des problèmes spécifiques liés à la nature même du matériau traité. En effet, on est amené à traiter des énoncés de parole plus ou moins spontanée et contenant de nombreux traits paralinguistiques. Par exemple, la présence de disfluences orales (répétitions, reprises, incises...) réduit la régularité syntaxique des énoncés ; les énoncés oraux sont également riches d'informations liées aux affects, etc. Par ailleurs, l'étape de transcription automatique, souvent nécessaire avant l'application de traitements de plus haut niveau (compréhension, traduction, analyse, etc.), rend des sorties bruitées (contenant des erreurs) qui nécessitent des analyses robustes et un couplage étroit entre étapes de traitement.

Pour ce numéro spécial, nous avons invité des contributions portant sur tous les aspects (théoriques, méthodologiques et pratiques) relatifs au traitement automatique du langage parlé et à la communication orale. En particulier, les thèmes listés dans l'appel à communication étaient :

- reconnaissance automatique de la parole ;
- compréhension automatique de la parole ;
- traduction de la parole ;
- synthèse de la parole ;
- dialogue oral homme-machine ;
- analyse robuste de la langue parlée ;
- analyse des affects sociaux ou des émotions dans des énoncés oraux ;
- fouille de documents à composantes orales ;
- applications à composantes orales ;
- outils d'aide à l'apprentissage d'une langue seconde ;
- aspects multilingues du traitement automatique du langage parlé ;
- évaluation de systèmes de traitement du langage parlé ;
- corpus et ressources pour l'oral ;
- analyse du discours oral ;
- dialogue adaptatif au contexte et au profil de l'utilisateur ;
- analyse des traits paralinguistiques dans des énoncés oraux.

Pour ce numéro spécial, nous avons reçu neuf soumissions complètes. Chaque article a été relu par trois relecteurs : deux relecteurs issus du comité scientifique *ad hoc*, constitué de spécialistes du traitement de la parole et du langage parlé ainsi

qu'un relecteur issu du comité permanent du journal TAL. À l'issue d'un processus de sélection en deux phases, nous avons conservé six articles originaux.

On y retrouve des travaux sur la traduction de parole (Segal *et al.*, 2015) et la transcription automatique de la parole (Illina *et al.*, 2015). La collection de ressources pour le traitement automatique du langage parlé fait aussi l'objet de deux contributions : pour le traitement de l'arabe dialectal (Boujelbane *et al.*, 2015) et pour la résolution de coréférences à l'oral (Désoyer *et al.*, 2015). Enfin, la parole est abordée comme véhicule d'informations paralinguistiques dans (Devillers *et al.*, 2015) tandis que Fontan *et al.* (2015) s'intéressent à la simulation de mesures perceptives humaines à l'aide d'outils de reconnaissance automatique de la parole. Une présentation succincte de chacun de ces articles est proposée dans le paragraphe suivant.

## 2. Présentation des contributions

Un des défis de la traduction automatique de la parole (TrAP) est notamment le manque de corpus oraux parallèles. Comme l'indique Segal *et al.* (2015), cette rareté des données spécifiques implique d'utiliser, pour l'apprentissage des modèles de traduction, des données parallèles textuelles qui ressemblent peu aux réelles sorties des systèmes de reconnaissance automatique de la parole (RAP). Les auteurs proposent de contourner ce problème dans le cadre particulier de la traduction de conférences (*TED Talks*). Ils quantifient l'impact des erreurs de transcription, de segmentation et de ponctuation sur les performances des systèmes de TAP (écart de plus de dix points BLEU sur la qualité de la traduction automatique par rapport à une traduction de l'écrit) et pointent les voies d'amélioration les plus prometteuses. Cet article constitue également une très bonne introduction en français au domaine de la TrAP.

Concernant la reconnaissance automatique de la parole, le problème des mots hors vocabulaire – et notamment des noms propres – est abordé dans (Illina *et al.*, 2015) pour la transcription d'émissions de radio en français. L'idée force repose sur l'hypothèse que certains noms propres apparaissent dans des documents relatifs à la même période temporelle et dans des contextes lexicaux similaires. À partir d'un corpus initial de grande taille (données textuelles d'agences de presse), un sous-corpus composé des documents qui sont contemporains de l'émission à transcrire, est obtenu pour augmenter le lexique du système de RAP. Plusieurs méthodes de sélection sont proposées et évaluées pour augmenter de façon dynamique le vocabulaire en utilisant des informations lexicales et temporelles. Les résultats expérimentaux montrent une réduction significative du taux d'erreur de noms propres pour une tâche de transcription d'émissions de radio.

Le cas particulier des dialectes, présents seulement sous une forme orale, est un autre défi pour le traitement automatique des langues. Ainsi, Boujelbane *et al.* (2015) proposent une première contribution originale au traitement automatique de l'oral parlé dans les médias tunisiens. Les auteurs collectent des transcriptions d'émissions tunisiennes contenant de nombreux cas d'alternance codique entre arabe standard et

dialectal. Ils proposent ensuite de projeter des ressources, initialement disponibles pour l'arabe standard, vers le dialecte ciblé. Un corpus de textes standard de grande taille est « transformé » en arabe dialectal afin d'entraîner un modèle de langue spécifique. L'évaluation – en terme de couverture lexicale et de perplexité – montre le potentiel de l'approche pour construire rapidement un système de transcription avec peu de ressources dialectales.

Le manque de ressources orales pour adapter des systèmes de TAL à des phénomènes propres à l'oral est également abordé dans (Désoyer *et al.*, 2015). Cet article présente un système original de résolution des coréférences en français reposant sur des techniques d'apprentissage automatique à partir de données exclusivement orales (corpus ANCOR : anaphore et coréférence dans les corpus oraux). Les performances du système de résolution de coréférences sont proches de celles de l'état de l'art des systèmes conçus pour l'écrit. La méthodologie et les ressources présentées par Désoyer *et al.* (2015) constituent par ailleurs un très bon point de départ pour de futures études des chaînes de coréférences en français parlé.

Enfin, les deux derniers articles sont spécifiques au signal de parole qui véhicule, entre autres, des informations paralinguistiques. Devillers *et al.* (2015) aborde la détection des émotions dans un contexte d'interaction homme-robot en utilisant des descripteurs multiples issus du signal vocal. Les descripteurs non verbaux extraits sont présentés comme complémentaires au canal verbal pour la reconnaissance de quatre états émotionnels : la joie, la tristesse, la colère et un état neutre. De son côté, Fontan *et al.* (2015) proposent d'utiliser des outils du traitement automatique de la parole pour une application de réglage de prothèses auditives. Un système de RAP est utilisé comme mesure perceptive. Pour cela, une étude des corrélations entre performances humaines et performances machines est réalisée à partir d'un corpus de parole dégradée. Les résultats montrent que la prédiction, par la machine, de l'intelligibilité de la parole est envisageable à terme.

#### Remerciements

Nous remercions Béatrice Daille et Pascale Sébillot pour leur aide précieuse et leurs conseils tout au long de la réalisation de ce numéro. Nous remercions également le comité scientifique de ce numéro spécial (par ordre alphabétique) : Gilles Adda, Jean-Yves Antoine, Véronique Aubergé, Jérôme Bellegarda, Hélène Bonneau-Maynard, Christophe Cerisara, Jan Cernocky, Géraldine Damnati, Laurence Devillers, Thierry Dutoit, Yannick Esteve, Maxine Eskenazi, Benoît Favre, Isabelle Ferrane, Guillaume Gravier, Denis Jovet, Juliette Kahn, Benjamin Lecouteux, Fabrice Lefevre, Georges Linares, Sylvain Meignier, Olivier Pietquin, Andrei Popescu-Belis, Sophie Rosset.

### 3. Bibliographie

- Boujelbane R., Ellouze M., Béchet F., Belguith L., « De l'arabe standard vers l'arabe dialectal : projection de corpus et ressources linguistiques en vue du traitement automatique de l'oral dans les médias tunisiens », *Traitement automatique des langues (TAL)*, 2015.
- Désoyer A., Landragin F., Tellier I., Lefeuvre A., Antoine J.-Y., « Les coréférences à l'oral : une expérience d'apprentissage automatique sur le corpus ANCOR », *Traitement automatique des langues (TAL)*, 2015.
- Devillers L., Tahon M., Sehili M., Delaborde A., « Détection des états affectifs lors d'interactions parlées : robustesse des indices non verbaux », *Traitement automatique des langues (TAL)*, 2015.
- Fontan L., Magnen C., Tardieu J., Ferrané I., Pinquier J., Farinas J., Gaillard P., Aumont X., « Comparaison de mesures perceptives et automatiques de l'intelligibilité : application à de la parole simulant la presbyacousie », *Traitement automatique des langues (TAL)*, 2015.
- Illina I., Fohr D., Linarès G., « Ajout de nouveaux noms propres au vocabulaire d'un système de transcription en utilisant un corpus diachronique », *Traitement automatique des langues (TAL)*, 2015.
- Segal N., Bonneau-Maynard H., Yvon F., « Traduire la parole : le cas des *TED Talks* », *Traitement automatique des langues (TAL)*, 2015.