
Préface

La phonétique et la phonologie à la rencontre du traitement automatique des langues

Il existe entre les recherches sur la forme sonore du langage et le traitement automatique des langues (TAL) des connexions bien établies depuis longtemps. La construction des systèmes de synthèse et de reconnaissance automatiques de la parole en a été l'un des moteurs. Au cours de ces dernières années, ces relations se sont nettement renforcées sous l'influence d'un certain nombre de facteurs. Le premier point de convergence tient aux techniques de recueil et d'exploitation partagée des ressources considérables dont nous disposons à présent dans le domaine de l'oral. Ces ressources jouent à présent un rôle de premier plan pour les phonologues et les phonéticiens qui, de plus en plus nombreux, entreprennent de soumettre leurs hypothèses théoriques à l'épreuve des grands corpus oraux. Le second point de convergence est lié à l'intérêt croissant des spécialistes du TAL pour la langue parlée. Les ressources concernant l'oral ont d'abord consisté en enregistrements audio de la parole en mode de lecture. Aujourd'hui, elles consistent massivement en données bi ou plurimodales de parole spontanée en situation d'interaction, ce qui soulève des problèmes de nature théorique et méthodologique redoutables que les phonologues/phonéticiens et les spécialistes du TAL ont entrepris d'aborder de front.

Dans les recherches sur l'oral, on voit ainsi se généraliser l'utilisation d'un large ensemble d'instruments et de méthodes permettant de procéder de manière automatique au traitement et à l'analyse des données recueillies : convertisseurs graphèmes-phonèmes, alignement automatique du texte sur le signal de parole, segmentation automatique du signal en unités de tailles différentes (des événements acoustiques aux tours de parole), procédures d'annotation morphosyntaxique, etc. Les études sur corpus de grande ou très grande échelle menées en phonologie et en phonétique, font ainsi de plus en plus souvent appel à des outils initialement développés en TAL, des dictionnaires électroniques aux systèmes complets de reconnaissance automatique de la parole. Les spécialistes du TAL et les phonologues/phonéticiens ont également contribué de manière conjointe au développement des systèmes d'annotation multiniveaux (des événements articulatoires/acoustiques au niveau pragmatique en passant par la prosodie et la syntaxe) de la parole.

Dans ce contexte très favorable à l'établissement de passerelles interdisciplinaires autour de l'oral, les connaissances et les ressources accumulées

par les phonologues et les phonéticiens sont aujourd'hui exploitées par les spécialistes du TAL, qu'il s'agisse pour eux de construire des bases de données lexicales à partir de corpus oraux, de développer des systèmes de RAP capables de traiter les variations régionales dans la forme sonore d'une langue, ou encore de concevoir des systèmes de synthèse de type visage parlant dans la communication homme-machine.

Dans cette configuration interdisciplinaire nouvelle, s'est imposée à nous la nécessité de procéder à un état des lieux des recherches actuelles à l'interface entre phonétique, phonologie et traitement automatique de la parole. Les sept contributions rassemblées dans ce volume thématique de *Traitement automatique des langues* constituent l'aboutissement de ce projet, et montrent sous quelle forme cette interface s'est aujourd'hui mise en place. Elles sont pour une part issues d'une journée d'étude de l'ATALA que nous avons organisée en 2006. Toutes les contributions retenues, et les nouvelles suscitées, ont été reprises pour les besoins du présent numéro thématique de TAL.

L'article de Adda-Decker, Gendrot, et Nguyen, « *Contributions du traitement automatique de la parole à l'étude des voyelles orales du français* », présente d'abord un aperçu des perspectives nouvelles qui s'offrent aujourd'hui aux phonologues et aux phonéticiens avec le développement des grandes bases de données de parole, et celui des outils qu'il est maintenant possible d'employer pour exploiter ces bases de données de manière automatique. Parmi les outils présentés, l'accent est placé sur les dictionnaires de prononciation associés aux systèmes de RAP, et sur les aligneurs automatiques. Le domaine d'application de ces outils est illustré au travers d'analyses à grande échelle portant sur la durée et la forme spectrale des voyelles du français, et sur les variations présentées par ces paramètres en fonction du style de parole et de l'accent régional du locuteur.

La contribution de Bazillon, Jousse, Béchet, Estève, Linarès, et Luzzati, « *La parole spontanée : transcription et traitement* », est centrée sur la parole spontanée et les problèmes hautement complexes que celle-ci peut soulever pour les systèmes de reconnaissance automatique de la parole. Les auteurs commencent par caractériser la parole spontanée, dans ce qu'elle a de spécifique par rapport à la parole préparée. Ils proposent huit critères pertinents à différents niveaux d'analyse (phonologique, morphosyntaxique, lexical, prosodique). Ils procèdent ensuite à un inventaire des logiciels d'aide à la transcription, des corpus disponibles, et des conventions d'annotation utilisables dans les études à grande échelle sur la parole spontanée. Différentes expériences sont présentées dont l'objectif est de développer un système offrant la possibilité de détecter de manière automatique des zones de parole spontanée au sein d'une base de données de parole. Les auteurs entreprennent ensuite de montrer de quelle manière il est possible d'améliorer un système de RAP afin de le rendre capable de transcrire automatiquement de la parole spontanée, en présentant une analyse détaillée des principales erreurs commises par les systèmes actuels dans ce domaine.

L'article de Beller, Veaux, Degottex, Obin, Lanchantin, et Rodet, « *IrcamCorpusTools : Plate-forme pour les corpus de parole* », vise à présenter un système développé par les auteurs pour la création, la gestion et l'exploitation des corpus de parole. Ce système est basé sur l'environnement de programmation Matlab/Octave, et il se présente comme une plate-forme de travail permettant de couvrir de manière unifiée l'ensemble des étapes relatives à la mise en place et à l'utilisation d'un corpus de parole, depuis l'annotation multiniveaux des signaux enregistrés, jusqu'aux analyses dont les données peuvent en être l'objet, et jusqu'aux applications possibles dans le domaine de la synthèse vocale. Des illustrations sont fournies qui portent sur l'analyse de la qualité vocale et différents phénomènes prosodiques.

Bertrand, Blache, Espesser, Ferré, Meunier, Priego-Valverde, et Rauzy (« *Le CID – Corpus of Interactional Data – Annotation et exploitation multimodale de parole conversationnelle* ») centrent leur contribution sur un corpus audio-vidéo constitué dans le but d'étudier les relations qui s'établissent entre différents niveaux d'analyse (du niveau phonétique au niveau discursif/interactionnel) dans le français parlé en situation d'interaction conversationnelle. Ce corpus, le CID, rassemble huit dialogues filmés d'une heure chacun. Après avoir exposé le protocole adopté pour la mise en place et l'enregistrement de ces dialogues, les auteurs fournissent des éléments détaillés sur la manière dont les données audiovisuelles ont été annotées à différents niveaux (orthographique, phonétique, prosodique, morphosyntaxique, syntaxique, discursif et mimo-gestuel). Des premiers résultats de référence sont présentés, portant sur la précision de l'alignement automatique phonèmes – signal de parole et sur les fréquences d'occurrences des phonèmes dans la parole conversationnelle. Les durées des voyelles et les facteurs présidant à la production des signaux de backchannel sont également étudiés.

Boula de Mareüil, Vieru-Dimulescu, Woehrling, et Adda-Decker, « *Accents étrangers et régionaux en français – Caractérisation et identification* », présentent les résultats d'une étude destinée à caractériser la façon dont les accents étrangers et les accents régionaux sont identifiés par des auditeurs humains, d'une part, et par un système de classification automatique, d'autre part. Le premier volet de cette investigation a porté sur l'identification de six accents étrangers en lecture et dans des entretiens semi-directifs. Les auteurs montrent que ces accents sont identifiables par des auditeurs francophones dans des proportions bien supérieures au hasard. Des analyses phonétiques et prosodiques, ainsi que des analyses automatiques menées au moyen d'un aligneur équipé d'un dictionnaire de prononciation étendu, font apparaître que les différences qui se présentent entre les accents vont dans le sens des tendances déjà établies dans la littérature. Un classificateur automatique basé sur un ensemble de traits hiérarchisés en fonction de leur pouvoir discriminant entre accents, aboutit lui aussi à de bons taux d'identification. Dans un second volet, cette investigation est étendue à l'identification des accents régionaux (six au nord de la Loire, cinq dans le sud de la France, un en Suisse romande). Les résultats montrent qu'une grande partition est établie par les auditeurs entre français du nord,

français du midi, et français suisse romand, sur la base d'un ensemble de caractéristiques phonétiques que les auteurs s'attachent à identifier.

Bürki, Gendrot, Gravier, Linarès, et Fougeron, « *Alignement automatique et analyse phonétique : comparaison de différents systèmes pour l'analyse du schwa* », ont entrepris de déterminer la fiabilité de trois systèmes d'alignement automatique, en centrant cette évaluation sur le schwa en français. Leurs analyses ont porté sur les schwas alternants dans de la parole radiophonique produite par près de 600 locuteurs. Les trois systèmes évalués (LIA, IRISA monophones et IRISA triphones) présentent des différences liées au dictionnaire de prononciations et aux modèles acoustiques utilisés. Leurs performances ont été évaluées par comparaison avec les alignements réalisés manuellement par deux transcripateurs. Les résultats montrent que ces performances sont influencées à la fois par les caractéristiques structurelles du système utilisé et par la nature de la tâche (détection ou segmentation du schwa). Dans leur discussion, les auteurs soulignent qu'il est important de prendre en considération l'ensemble de ces éléments dans le choix d'un aligneur automatique pour les besoins d'une analyse phonétique.

Vasilescu, Adda-Decker, et Remoto, « *Caractéristiques acoustiques et prosodiques des hésitations vocaliques dans trois langues* », se sont intéressées à la manière dont les voyelles d'hésitation sont phonétiquement réalisées en anglais américain, en français et en espagnol. Leurs analyses ont porté sur des enregistrements relatifs à des émissions radiophoniques et à des débats parlementaires. Elles ont permis aux auteurs de comparer la forme sonore des voyelles d'hésitation dans les trois langues. Les auteurs ont également entrepris d'examiner la relation entre la forme sonore des voyelles d'hésitation et celles des voyelles dites intralexicales à l'intérieur du système vocalique de chaque langue. Les résultats montrent que les voyelles d'hésitation ont des timbres qui varient d'une langue à l'autre. Ces timbres sont proches de timbres infralexicaux sans nécessairement se confondre avec eux.

Pour clore ce préambule, nous souhaitons exprimer tous nos remerciements à la revue TAL et à son comité de rédaction pour nous avoir offert de coordonner la réalisation de ce numéro thématique. Nous remercions également les membres du comité de lecture mis en place pour les besoins de ce numéro. Le temps et l'attention qu'ils ont bien voulu consacrer à l'évaluation des différentes contributions nous ont été très précieux.

Bernard Laks
Laboratoire MoDyCo
Université de Paris Ouest Nanterre et
CNRS
200 avenue de la République
92001 Nanterre
bernard.laks@u-paris10.fr

Noël Nguyen
Laboratoire Parole et Langage
Université Aix-Marseille et CNRS
5 avenue Pasteur
13100 Aix-en-Provence
noel.nguyen@lpl-aix.fr

Comité de lecture

Martine ADDA-DECKER, LIMSI, CNRS, Université Paris VI, Université Paris XI, Orsay, France

Roxane BERTRAND, LPL, CNRS, Aix-Marseille Université, Aix-en-Provence, France

Philippe BLACHE, LPL, CNRS, Aix-Marseille Université, Aix-en-Provence, France

Cédric GENDROT, LPP, CNRS, Université de Paris III, France

John GOLDSMITH, Université de Chicago, USA

Guillaume GRAVIER, Irisa, CNRS/INRIA et Université de Rennes I, France

Jonathan HARRINGTON, IPS, Université de Munich, Allemagne

Lori LAMEL, LIMSI, CNRS, Université Paris VI, Université Paris XI, Orsay, France

François PELLEGRINO, DDL, CNRS, Université de Lyon II, France

François POIRE, Université de Western Ontario, Canada

Yvan ROSE, Memorial University of Newfoundland, Canada

Tobias SCHEER, BCL, CNRS, Université de Nice, France

Atanas TCHOBANOV, MoDyCo, CNRS, Université de Paris X, France

Jacqueline VAISSIERE, LPP, CNRS, Université de Paris III, France

Nathalie VALLEE, DPC-GIPSA, CNRS, INPG, UJF, Université Grenoble III, France

Qu'ils soient tous chaleureusement remerciés !