

# Un cadre expérimental pour les Sciences de la Parole

Gilles ADDA

LIMSI/CNRS Rue John von Neumann Université Paris-Sud 91403 ORSAY

[gilles.adda@limsi.fr](mailto:gilles.adda@limsi.fr)

## RÉSUMÉ

---

Cet article est une prise de position pour la mise en place d'un cadre théorique et pratique permettant de faire émerger une science empirique de la parole. Cette science doit se fonder sur l'apport de toutes les sciences, du traitement automatique ou de la linguistique, dont l'objet d'étude est la parole. Au cœur de ce rapprochement se trouve l'idée que les systèmes automatiques peuvent être utilisés comme des *instruments* afin d'explorer les très grandes quantités de données à notre disposition et d'en tirer des connaissances nouvelles qui, en retour, permettront d'améliorer les modélisations utilisées en traitement automatique. Quelques points cruciaux sont abordés ici, comme la définition de l'observable, l'étude du résiduel en tant que diagnostic de l'écart entre la modélisation et la réalité, et la mise en place de centres instrumentaux permettant la mutualisation du développement et de la maintenance de ces instruments complexes que sont les systèmes de traitement automatique de la parole.

## ABSTRACT

---

### **An experimental framework for speech sciences**

This article is a position paper in favor of the establishment of a theoretical and practical framework to bring out an empirical science of speech, based on the contribution of all the sciences whose object of study is the speech. Central to this re-convergence is the idea that automatic systems can be used as *instruments* to explore large amounts of data at our disposal and to derive new linguistic knowledge which, in turn, will allow to improve the models used in the automatic systems. Some crucial points are discussed, such as the definition of the observable, the study of the residual as a diagnostic of the gap between modeling and reality, and the development of instrumental centers for the sharing of development and maintenance of these complex instruments which are automatic speech processing systems.

---

**MOTS-CLÉS :** analyse d'erreurs ; structuration de la recherche en parole.

**KEYWORDS:** Epistemologic study ; error analysis ; structuration of speech sciences.

---

## 1 Introduction

Cet article est une prise de position pour la mise en place d'un cadre théorique et pratique permettant de faire émerger une science empirique de la parole. Le but de cette prise de position est d'apporter mon soutien à un mouvement que l'on voit apparaître dans les différentes communautés des sciences du langage, et en particulier dans un certain nombre de disciplines dont l'objet est l'étude de la parole. Ce mouvement tend à considérer que nous sommes arrivés à une maturité des systèmes de reconnaissance automatique (au sens large) qui peut nous permettre de passer un cap scientifique, et de rapprocher la communauté du traitement automatique et

les communautés des sciences humaines afin qu'elles s'enrichissent mutuellement, voire qu'elles collaborent véritablement autour des mêmes objets et des mêmes instruments, dans un cadre expérimental commun (Adda-Decker, 2006). Mais la constatation d'un rapprochement, ne doit pas nous faire sous-estimer tout le travail autant théorique que pratique à mettre en œuvre, si nous voulons que ce rapprochement devienne une réalité scientifique. Parmi les questions théoriques qu'il nous faut aborder, nous pouvons citer : Quel est le statut de la connaissance que nous produisons, comment la qualifier par rapport à d'autres sciences ? Est-il possible d'autonomiser les sciences de la parole en une véritable science, en essayant de trouver à la fois quel est son observable et le moyen d'améliorer la manière de l'observer, et d'en tirer des connaissances généralisables ? Et parmi les questions pratiques : La structure actuelle de la recherche est-elle adéquate pour accueillir et faire prospérer une science empirique réunissant les deux communautés ? Les modes d'évaluation et de production scientifiques sont-ils adaptés ?

Il serait illusoire ici de vouloir toutes les aborder. Le but de cet article étant finalement de susciter le débat voire l'intérêt autour de ce sujet, pas de l'épuiser<sup>1</sup>, je ne survolerai, après une courte mise en perspective, que certaines d'entre elles, c'est-à-dire une définition de l'observable, et un mode opératoire afin d'en tirer des connaissances en particulier à travers l'étude des erreurs, et une proposition concrète de création de centres expérimentaux permettant de mettre pratiquement en œuvre ce cadre expérimental.

## 2 Historique

Au tournant des années 80, le traitement de la parole a été confronté à de profonds changements, où se sont mis conjointement en place 2 faits majeurs :

- le développement de l'approche statistique, fondée sur la modélisation statistique de la parole, issue de la théorie de l'information (Jelinek, 1976) ;
- l'introduction du principe de l'évaluation comparative des systèmes (Pallet, 1985).

L'introduction de ces deux principes structure la recherche en parole depuis 40 ans, au point où, plus qu'une approche dominante, ils constituent un couple de pratiques quasi-hégémonique dans la production scientifique actuelle en traitement automatique de la parole. Ils ont permis au traitement automatique de la parole de mettre en défaut les critiques fondamentales quant à son caractère scientifique qui sont apparues à la fin des années 60 (Pierce, 1969).

Même si l'application de ces principes semble aujourd'hui avoir été un long fleuve tranquille, de nombreuses critiques sont apparues. Même en désaccord avec elles, elles nous permettent de nous interroger sur la réalité de la mise en place du cadre actuel, et peut nous permettre d'entrevoir quelques pistes d'évolution. Parmi ces critiques, nous pouvons citer celle formulée par Stephen E. Levinson (Levinson, 1994), qui remet en cause le caractère réellement scientifique des avancées obtenues dans le cadre de l'évaluation comparative, car elles ne reposeraient pas sur une théorisation. Autre critique, celle de Roger K. Moore (Moore, 2007), qui par extrapolation des courbes des tailles des corpus en fonction des performances des systèmes, conclut à la nécessité d'augmenter de manière démesurée les corpus afin d'amener les systèmes à des performances comparables à celles de l'être humain et aboutit à la conclusion qu'il faut changer de paradigme ; il suggère l'introduction d'une approche complètement différente, la « Cognitive Informatics », qui est fondée sur une étude transdisciplinaire de la compétence humaine. Cette approche est très intéressante (voir section 3), mais elle n'est pas du tout incompatible avec une poursuite des

---

1. On pourra trouver un survol un peu moins rapide de ces questions dans (Adda, 2011)

recherches sur l'amélioration des systèmes, et en particulier l'amélioration de l'utilisation des données (voir section 4).

Ces critiques mettent en exergue plusieurs points que l'on peut penser perfectibles dans l'état actuel des sciences de la parole : mieux définir, en tant que science expérimentale, le cadre scientifique de nos travaux en traitement automatique ; mieux définir et mieux utiliser les corpus qui sont le cœur des modélisations statistiques et de l'évaluation comparative, par exemple en permettant une plus grande interaction avec les autres sciences de la parole.

### 3 Linguistique et traitement automatique de la parole

Un rapprochement des communautés des sciences humaines et sociales ayant pour objet la parole au sens large et la communauté du traitement automatique de la parole a déjà eu lieu sur plusieurs points ces dernières années. Parmi ces domaines intéressants, nous pouvons citer la collaboration entre modèles de perception humaine et modèles statistiques utilisés pour la reconnaissance automatique. M.A. Huckvale a très tôt introduit la position alors iconoclaste que les systèmes de reconnaissance pouvait offrir une vue nouvelle pour les sciences cognitives, et en particulier pour la reconnaissance de mots par les humains (Huckvale, 1997, 1998). Dans (Moore et Cutler, 2001), un parallèle très complet est fait entre les buts des études de la reconnaissance par des humains et des machines. Plus récemment, dans (Scharenborg *et al.*, 2005; Scharenborg, 2005) est fait le parallèle entre reconnaissance humaine (Human Speech recognition, HSR) et reconnaissance automatique (Automatic Speech Recognition, ASR) et le développement d'un modèle computationnel de la reconnaissance de mot par un humain.

Mais le point crucial que je veux mettre en avant ici est la **linguistique à l'instrument**. Il s'agit ici de l'une des évolutions récentes qui est fondamentale dans le développement d'un cadre expérimental : considérer les outils de traitement automatique comme des *instruments* permettant d'accéder aux très grandes quantités de données (de toutes sortes, et en particulier de parole), afin de pouvoir en extraire une connaissance phonétique, sociolinguistique, lexicale, syntaxique, ... Cette évolution fait écho au « Portrait de linguiste(s) à l'instrument », dans lequel Benoît Habert (Habert, 2005) fait état d'une évolution de la linguistique, qui jusqu'alors était dominée par l'approche générativiste, et déniait l'intérêt du concept d'instrument. B. Habert introduit deux idées clés : les instruments permettent de voir de nouveaux phénomènes, ce sont des outils de perception ; il est nécessaire d'adapter les données aux instruments : un instrument est un capteur imparfait, qui sert à prélever une information, et il est donc nécessaire de « voir » avec cet instrument des données qui peuvent être visibles, et pour lesquelles la précision de l'instrument permet d'extraire une information pertinente. Plus récemment Mark Liberman (Liberman, 2010) fait le parallèle entre la situation actuelle, et l'état de la science en 1610 « Hypothesis : 2010 is like 1610 (...) We've invented the linguistic telescope and microscope (...) We can observe linguistic patterns ». Il montre que les sciences du langage peuvent être un paradigme des « e-Sciences », fondée sur l'utilisation intensive des ordinateurs, utilisant des bases de données de très grande dimension, dans un environnement hautement distribué.

Les systèmes automatiques de traitement de la parole (reconnaissance du locuteur, segmentation audio, transcription orthographique et phonémique) peuvent être considérés comme des instruments capables d'explorer des quantités de données jusqu'alors inatteignables, et de faire émerger des problématiques ou de valider des hypothèses linguistiques. Cette utilisation nécessite cependant un rapprochement des différentes communautés, car les systèmes automatiques,

comme l'a souligné B. Habert, sont des *instruments* et non des *outils*<sup>2</sup> : ils nécessitent une mise au point, des développements spécifiques, un « savoir-faire », qui n'est pas réductible à la mise à disposition des communautés linguistiques de « boîtes à outils ».

## 4 Définir un observable

Le rapprochement des différentes sciences de la parole peut se faire autour d'une famille d'instruments, les systèmes automatiques de traitement de la parole, sur (par exemple) les sujets scientifiques décrits dans la section précédente. Cependant, pour que les différentes approches puissent s'enrichir mutuellement, il convient qu'elles s'accordent sur un *style de raisonnement scientifique*, en reprenant la terminologie introduite par Ian Hacking (Hacking, 1992), style qui doit prendre en compte le style spécifique des sciences utilisant la statistique. Selon Hacking, chaque style définit ses critères de vérité, et l'objet de sa recherche. Quel peut être ici cet objet ? En prenant en compte les résultats obtenus par les deux principes fondateurs énoncés en section 2, et ce que l'on observe en « Linguistique de corpus », cet objet semble être le corpus. Mais ce concept de corpus est flou, voire contradictoire selon ses acceptations. Pratiquement, ce que l'on utilise comme propriété de cet objet, ce n'est pas d'être un recueil de faits isolés de langue (par exemple une phrase ou un phonème particulier énoncé par un locuteur particulier dans un contexte défini) mais qu'il permette des mesures statistiques significatives et *intéressantes*. Ces mesures sont pertinentes par rapport à un certain nombre de paramètres implicites ou explicites que l'on prête aux données langagières contenues dans le corpus, et que sont supposés traiter les systèmes automatiques (parole lue ou spontanée, spécifique à une tâche donnée, bande large ou étroite, ...). Je fais l'hypothèse ici que l'objet commun est un *espace langagier*, dont le corpus serait un échantillon supposé pertinent, délimité par des valeurs d'un jeu de paramètres, certains explicites (choisis explicitement lors de la constitution du corpus) ou implicites (présents implicitement par le choix spécifique du corpus), où les phénomènes ont des caractéristiques stables et intéressantes et où les mesures sur cet objet sont les sorties<sup>3</sup> de systèmes automatiques sur le corpus ; j'appellerai par la suite cet objet *espace langagier isoparamétrique* (ELI). L'intérêt de ces définitions de l'objet commun et de la manière dont on l'observe, est qu'elles permettent de dire comment on peut améliorer la qualité de l'observation : en expérimentant sur les contenu ou les frontières de cet objet. Citons comme exemple l'évolution des performances lors de l'accroissement du corpus d'apprentissage au sein d'un ELI : une stagnation ou même une augmentation du taux d'erreur doit nous orienter vers la recherche de nouveaux paramètres à expliciter. Autre moyen d'améliorer l'observation, l'étude des erreurs faites sur un ELI, qui est abordée dans la section 5 : celle-ci permet de mettre à jour de nouvelles dimensions intrinsèques du corpus, ou d'orienter les recherches afin d'améliorer les systèmes.

## 5 Analyse du résiduel

Dans le cas de la transcription de la parole, on mesure des taux d'erreurs qui vont représenter l'écart entre la modélisation obtenue sur le corpus d'apprentissage, et les performances sur

---

2. Je reprends dans la définition introduite par B. Habert, le fait que l'outil est générique, polyvalent, alors que l'instrument est spécifique à un type de données.

3. Les sorties (mots, phonèmes ou autres) seront utilisées comme *faits* par les différentes sciences de la parole.

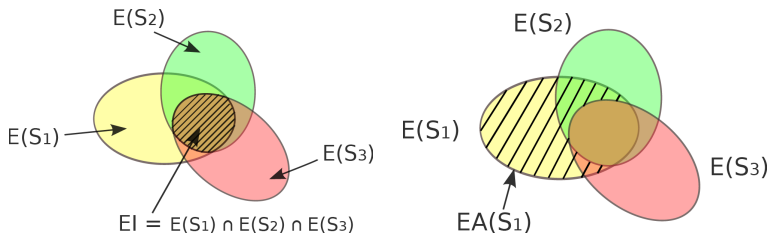


FIGURE 1 – (gauche) Visualisation de classe des erreurs irréductibles, égale à l'intersection stricte des erreurs des 3 systèmes S1, S2 et S3. (droite) Visualisation de la classe des erreurs atteignables du système S1, comme la classe des erreurs de S1 qui ne sont pas des erreurs pour les systèmes S2 et S3.

le corpus de développement et d'évaluation (Adda, 2011; Abney, 2011). Cet écart entre la modélisation et la réalité est communément nommé *résiduel*. Ce résiduel, ou « erreurs », a été assez peu étudié en parole. Or dans un cadre expérimental, l'étude des erreurs est un domaine crucial. Cette étude est multiforme ; elle couvre par exemple l'étude comparée des erreurs des systèmes et des humains (Vasilescu *et al.*, 2009), l'étude diagnostique des erreurs (Goldwater *et al.*, 2010). Il faut souligner ici que l'étude des erreurs est intéressante quand il y a des erreurs en nombre assez faible pour que l'on puisse les classer aisément : pour des tâches où le taux d'erreur est trop important, les différents types d'erreurs interfèrent profondément, rendant toute tâche d'analyse très difficile. Aussi, l'étude des erreurs est intéressante en particulier sur des tâches/corpus que l'on a jugées « résolues », c'est-à-dire que les efforts en terme de développement simple (en particulier par augmentation de la taille du corpus d'apprentissage), n'apporteront qu'un gain faible, alors que les résultats sont assez « satisfaisants ». C'est en particulier pour ces tâches que les erreurs résiduelles, qui semblent rétives aux modélisations classiques, sont intéressantes à étudier. D'un autre côté, nous avons souligné que les systèmes de traitement de la parole pouvaient (devaient) être utilisés comme instruments pour explorer les corpus, afin de tester certaines hypothèses, de découvrir certains faits linguistiques, phonétiques etc ; dans ce paradigme, l'étude des erreurs nous permet de mesurer la précision des systèmes comme instruments, selon les corpus.

Examiner les erreurs est également utile pour découvrir les principales directions à explorer afin de déterminer de nouvelles techniques et à terme améliorer les performances des systèmes. Il est intéressant de se placer dans l'optique de la *résolution* des erreurs. Si nous voulons résoudre des problèmes, il s'agit de les identifier en séparant le problème global en sous-problèmes atteignables ; à ce titre une typologie des erreurs en fonction de leur solution est utile, car nous ne devons pas tenter de résoudre avec les outils méthodologiques existants, des problèmes qui ne peuvent pas être résolus, ou qui ne sont pas des problèmes, ou encore qui sont d'une importance mineure. Afin de pouvoir explorer les limites des modélisations actuelles sur une tâche, je préconise l'usage du *Rover Oracle* (RO), c'est-à-dire le choix oracle de la bonne solution si elle existe dans le graphe des différentes possibilités fournies par les différents systèmes suivant la méthode ROVER (Fiscus, 1997). Le Rover Oracle nous offre ainsi une approximation de la limite supérieure de ce que peut atteindre un système de l'état de l'art, qui utiliserait de manière optimale les différentes modélisations présentes dans les différents systèmes, et ayant également le réglage optimal de ses différents paramètres. En plus d'une indication sur la limite inférieure

du taux d'erreurs sur un corpus donné, le Rover Oracle nous permet un accès à différentes classes d'erreurs intéressantes. La classe d'erreurs la plus intéressante pour le chercheur, et qui permet d'approximer la précision des systèmes de traitement en tant qu'instrument, est la classe des **erreurs irréductibles** (EI) (voir figure 1, gauche), définie par  $EI = ERR(RO)$  où  $ERR(RO)$  est l'ensemble des erreurs produites par le Rover Oracle. On obtient ainsi la classe des solutions correctes qu'aucun système n'avait envisagées. Autre classe d'erreurs intéressantes, la classe des **erreurs atteignables** d'un système  $S$  ( $EA(S)$ ) (voir figure 1, droite). Sa définition est  $EA(S) = ERR(S) - ERR(RO)$ . De manière explicite,  $EA(S)$  représente la classe des erreurs du système  $S$  qui ont pu être corrigées en utilisant le Rover Oracle. Pour un système  $S$ , elle représente l'ensemble des erreurs pour lesquelles au moins un autre système a pu trouver la bonne solution.

Dans la mesure où l'étude des erreurs est intéressante là où il y en a peu, cela oblige, si l'on veut pouvoir trouver des corrélations utiles, à avoir de grands corpus d'évaluation. 10 heures de parole, avec un taux d'erreurs de 5% produisent environ 5000 erreurs, ce qui est bien peu pour extraire des classes. On peut donc extrapoler que ces analyses seront réellement fructueuses sur des corpus de l'ordre de 100 heures, ce qui est parfaitement atteignable avec les corpus actuels.

## 6 Centres Instrumentaux pour les sciences de la parole

Dans le mode de fonctionnement actuel de l'évaluation comparative, certaines équipes/laboratoires obtiennent par leur participation aux évaluations majeures une *certification* qui leur accorde une crédibilité. Cette crédibilité leur apportera tout à la fois une plus grande facilité à faire publier leurs résultats et une plus grande confiance de la communauté dans la validité de ces résultats. Il y a dans ce mode de fonctionnement, une importante déperdition de temps et de travail, car il n'accorde pas assez de place pour les équipes qui n'ont pas les moyens humains ou matériels de participer à des évaluations importantes, mais impose également aux équipes certifiées de consacrer beaucoup de temps sur ces évaluations et donc potentiellement moins sur l'innovation. En caricaturant la situation actuelle, nous avons le choix entre une uniformité fiable ou une innovation peu fiable, puisque si une innovation est introduite par un équipe qui n'a pas été certifiée ou par un système qui n'a pas été calibré par une évaluation, il n'y aura pas de confiance dans le résultat, au moins tant que celui-ci n'aura pas été reproduit par un couple système/équipe certifié. Par ailleurs, par rapport au modèle américain qui peut se résumer à développer quelques centres d'excellence, financés par les projets gouvernementaux, autour de centres d'évaluation et de mises à disposition de corpus (NIST<sup>4</sup> et LDC<sup>5</sup>), eux-mêmes financés directement ou indirectement par des projets gouvernementaux, on ne peut que constater que, malgré l'existence de quelques centres d'excellence, il n'existe pas de modèle efficace européen.

Si les systèmes de traitement du langage peuvent être considérés comme des instruments, il doivent être considérés comme des instruments complexes, difficile à maîtriser et coûteux à développer et à maintenir, et donc pour lesquels il est intéressant de développer une structure de mutualisation, comme par exemple, toute proportion gardée par rapport aux investissements en jeu, pour les anneaux d'accélération utilisés en physique des particules, au CERN<sup>6</sup>. L'existence de plusieurs instruments différents est intéressante, pour maintenir une certaine compétitivité,

---

4. [www.itl.nist.gov/](http://www.itl.nist.gov/)

5. [www ldc.upenn.edu/](http://www ldc.upenn.edu/)

6. Organisation européenne pour la recherche nucléaire, [public.web.cern.ch/public/](http://public.web.cern.ch/public/)

mais également par souci de complémentarité (voir section 5). On peut donc penser à la mise en place de cette mutualisation au sein de plusieurs laboratoires qui seraient alors des *centres instrumentaux* mutualisés. Le statut de centre instrumental amènerait des moyens (financiers et humains) supplémentaires ; les centres instrumentaux seraient impliqués dans les projets expérimentaux d'autres laboratoires qui n'auraient pas le statut de centre instrumental ; en échange, le centre accueillerait des équipes et des chercheurs choisis sur des projets nécessitant l'emploi d'un système état de l'art et/ou de corpus avec un financement spécifique, un partage des résultats (et des publications), et une pérennisation des acquis à un même endroit. L'ensemble des laboratoires utilisateurs et le centre instrumental formeraient un réseau, qui permettraient un développement des recherches et des instruments.

Dans le schéma de recherche que je propose, l'ensemble des expériences pourraient être menées avec des systèmes de l'état de l'art et sur des bases de données pertinentes. Les moyens de mise au point des instruments seraient concentrés sur quelques centres, et les autres moyens étant dévolus à des expériences précises. Ce schéma conserve la fiabilité directement issue de l'évaluation comparative, mais permet une plus grande innovation.

## 7 Conclusion

La mise en place de l'évaluation comparative et ses bénéfices visibles actuels n'est qu'une étape dans la mise en place d'une structure scientifique à même de permettre un développement à long terme de notre domaine. Ce développement passe en particulier selon moi par la pérennisation d'une vraie science empirique, qui elle-même nécessite en particulier :

- Une plus grande formalisation des buts, et des moyens à mettre en œuvre pour atteindre ces buts. Cela signifie définir clairement l'observable de cette science empirique, le moyen de l'atteindre, de l'observer, puis comment apprendre à partir de celui-ci. Je suggère d'utiliser comme observable une extension de la notion de corpus, l'espace langagier isoparamétrique, qui est défini par ses paramètres intrinsèques et extrinsèques.
- L'étude des erreurs des systèmes comme indicateur de l'écart entre la modélisation et la réalité telle qu'observée par l'évaluation sur un corpus.
- Une meilleure structuration de la production scientifique par une mutualisation des systèmes de traitement automatique comme instruments, systèmes complexes et coûteux à mettre en œuvre et à maintenir au meilleur niveau.

A l'heure actuelle le domaine du traitement de la parole bénéficie, relativement et avec de fortes disparités, de moyens qui font envie à bien d'autres domaines scientifiques. Mais les risques inhérents aux grands cycles technologiques et économiques sont grands. En formalisant plus notre discipline de manière à faire émerger une science empirique stable, et en utilisant mieux le potentiel actuel en terme de moyens humains et matériel, la connaissance que nous pourrions générer serait plus importante, et les sciences de la parole pourraient peser ainsi d'un poids plus important dans les choix scientifiques.

## Remerciements

Ce travail a été réalisé en partie dans le cadre du projet ANR EDyLex (ANR-09-CORD-008).

## Références

- ABNEY, S. (2011). Data-intensive experimental linguistics. *Linguistic Issues in Language Technology*, 6(2).
- ADDA, G. (2011). Approches empiriques et modélisation statistique de la parole. Habilitation à Diriger les Recherches de l'Université Paris-Sud (spécialité : informatique).
- ADDA-DECKER, M. (2006). De la reconnaissance automatique de la parole à l'analyse linguistique de corpus oraux. In *XXVIes Journées d'Étude sur la Parole, JEP*, Dinard.
- FISCUS, J. G. (1997). A post-processing system to yield reduced word error rates : Recognizer output voting error reduction (ROVER). In *Automatic Speech Recognition and Understanding, 1997. Proceedings., 1997 IEEE Workshop on*, pages 347–354.
- GOLDWATER, S., JURAFSKY, D. et MANNING, C. D. (2010). Which words are hard to recognize ? prosodic, lexical, and disfluency factors that increase speech recognition error rates. *Speech Communication*, 52(3):181–200.
- HABERT, B. (2005). Portrait de linguiste(s) à l'instrument. *Texte !*, 10(4).
- HACKING, I. (1992). Statistical language, statistical truth, and statistical reason : The self-authentication of a style of scientific reasoning. *Social Dimensions of Science*, pages 130–157.
- HUCKVALE, M. (1997). 10 things engineers have discovered about speech recognition. In *NATO ASI Workshop on Speech Pattern Processing*.
- HUCKVALE, M. (1998). Opportunities for re-convergence of engineering and cognitive science accounts of spoken word recognition. In *Proceedings of the Institute of Acoustics Conference on Speech and Hearing*, pages 9–20.
- JELINEK, F. (1976). Continuous speech recognition by statistical methods. *Proceedings of the IEEE*, 64:532–556.
- LEVINSON, S. E. (1994). *Speech recognition technology : a critique*, pages 159–164. National Academy Press, Washington, DC, USA.
- LIBERMAN, M. (2010). The future of computational linguistics : or, what would antonio zampolli do ? In *Antonio Zampolli Prize speech, presented at LREC2010*, Valletta, Malta.
- MOORE, R. K. (2007). Spoken language processing : Piecing together the puzzle. *Speech Communication*, 49:418–435.
- MOORE, R. K. et CUTLER, A. (2001). Constraints on theories of human vs. machine recognition of speech. In *Proceedings SPRAAC workshop on Human Speech Recognition as pattern Classification*. Max-Planck-Institute for Psycholinguistics.
- PALLET, D. S. (1985). Performance assessment of automatic speech recognizers. *J. J. Res. Natl. Inst. Stand. Technol.*, 90:371–387.
- PIERCE, J. R. (1969). Whither Speech Recognition ? *Acoustical Society of America Journal*, 46:1049.
- SCHARENBERG, O. (2005). Parallels between HSR and ASR : how ASR can contribute to HSR. In *Proceeding of Interspeech*, pages 1237–1240.
- SCHARENBERG, O., NORRIS, D., ten BOSCH, L. et McQUEEN, J. M. (2005). How should a speech recognizer work ? *Cognitive Science*, pages 867–918.
- VASILESCU, I., ADDA-DECKER, M., LAMEL, L. et HALLÉ, P. (September, 2009). A Perceptual Investigation of Speech Transcription Errors Involving Frequent Near-Homophones in French and American English . In *Interspeech'09*, pages 144–147, Brighton, UK.