
LES 23ÈMES JOURNÉES D'ÉTUDE SUR LA PAROLE

AUSOIS – FRANCE, 19-23 JUIN 2000

Gérard BAILLY et Pierre BADIN

Institut de la Communication Parlée (ICP)
46, av. Félix Viallet – 38031 Grenoble Cedex
Mel : {bailly,badin}@icp.inpg.fr

1. Préambule

Les Journées d'Étude sur la Parole (JEP) sont un congrès international qui a pour vocation de réunir tous les deux ans l'ensemble des chercheurs francophones travaillant sur la parole, tant sur les aspects fondamentaux de la communication parlée que sur les différents problèmes de son traitement automatique. La thématique des JEP couvre donc autant les technologies vocales (synthèse, codage, reconnaissance de la parole, de la langue et du locuteur...), l'ingénierie linguistique (traitement automatique de la langue, systèmes de dialogue, bases de données textuelles et parlées...) que les sciences qui les alimentent (traitement du signal et des images, phonétique, phonologie, linguistique, psychologie, motricité, perception...).

2. Un peu d'histoire...

Les premières JEP ont été organisées à Grenoble en 1970 sous l'égide du Groupe Communication Parlée du Groupement des Acousticiens de Langue Française (GALF). Suivent alors Aix-en-Provence (71), Lannion (72), Bruxelles (73), Orsay (74), Toulouse (75), Nancy (76), Aix-en-Provence (77), Lannion (78), Grenoble (79), Strasbourg (80), Montréal (81), Bruxelles (84), Paris (85), Aix-en-Provence (86). En 1987, le GALF devient Société Française d'Acoustique (SFA) et c'est sous son égide que les JEP sont organisées à Hammamet (87), Nancy (88), Montréal (90) puis Bruxelles (92). C'est en 1994 qu'apparaît le patronage conjoint de la SFA et de l'European Speech Communication Association (ESCA) pour l'organisation des JEP à Lannion. Suivent alors au rythme maintenant régulier d'une réunion tous les deux ans, Avignon (96), Martigny (98) et Aussois à présent.

Grâce notamment à ces rendez-vous réguliers, cette communauté francophone est restée fortement soudée et structurée. Cette cohésion et ce dynamisme doivent continuer à s'affirmer au sein d'une communauté européenne qui structure largement la recherche en parole au niveau international (par le biais notamment de congrès internationaux comme EuroSpeech et de la revue *Speech Communication*). Le GFCP est ainsi devenu l'un des quelques Special Interest Group (SIG) de l'ISCA, donnant ainsi à la communauté francophone une place privilégiée dans cette association internationale en croissance régulière.

3. Le mode de fonctionnement des JEP

Au fil des ans, les JEP ont convergé vers un mode de fonctionnement spécifique et original tant au niveau de la sélection des papiers que de l'organisation des sessions de la conférence. Ainsi les papiers soumis

doivent être complets (4 pages) et sont lus par au moins deux experts. L'avis de ces relecteurs est discuté dans une session plénière du comité scientifique qui dure environ deux jours et ressemble parfois aux débats d'un jury de cour d'appel : il arrive qu'un papier soit en fait relu par 5 ou 6 personnes du comité ou experts contactés par fax et sommés de donner un avis circonstancié dans la journée. Peu d'articles sont rejetés lors de cette phase mais l'acceptation définitive de plus de la moitié des papiers est souvent conditionnée à des modifications impératives : s'instaure ainsi parfois un dialogue enrichissant entre auteur(s) et relecteurs, ces derniers n'hésitant pas parfois à lever l'anonymat de leur jugement. Cette "communion" entre l'auteur et son public continue jusque dans la conférence elle-même puisqu'aucune session en parallèle n'est organisée : tous les participants assistent à toutes les sessions (sauf les inévitables escapades buissonnières !).

4. Les JEP à Aussois

4.1. Les innovations des JEP de cette fin de millénaire...

Tout d'abord sur le mode d'hébergement. Pour la première fois, les participants ont communiqué, mangé et dormi en un même lieu : le centre Paul Langevin du CAES du CNRS. La salle de conférence est à quelques longueurs de couloir des chambres de deux (ou individuelles pour les plus fortunés) et de la salle à manger : cette unicité de lieu constitue une occasion unique pour de nombreux chercheurs de présenter leurs travaux, dans la langue dans laquelle ils ont été conçus et développés, puis de pouvoir poursuivre tard dans la nuit les discussions engagées lors des sessions orales ou affichées.

Ensuite sur la disponibilité des actes : ces discussions gagnent en efficacité si les interventions (et les textes des conférences invitées) sont disponibles avant la conférence. La conception du serveur *web* (voir en titre) a permis d'offrir ce service plusieurs semaines avant l'ouverture des journées. Il a permis en outre à certains auteurs d'améliorer encore leurs textes par les commentaires reçus et aux organisateurs de produire un *cédérom* qui fonctionne et des actes presque sans faute en l'espace de quelques semaines.

Enfin sur la longue marche... plus d'une centaine de scientifiques a été lâchée à 2600 mètres d'altitude le mercredi 21 juin à 15 heures sur les chemins d'alpage, après une raclette copieusement arrosée en plein soleil. Miraculeusement, seule une légère entorse a été déplorée.

4.2. Les chiffres...

Parmi les 112 papiers reçus, plus de la moitié a été acceptée sous la condition de modifications impératives. Seuls 9 papiers — dont 2 étaient clairement hors sujet — ont été rejetés. Par ailleurs, 5 papiers invités ont été ajoutés aux actes. Parmi les 130 participants, qui ont — à plus de 90 % — réussi à rester pendant la totalité des 5 jours avec nous, nous trouvons aux côtés de chercheurs français, suisses, algériens, marocains, belges et canadiens, des auteurs affiliés à des laboratoires en Allemagne, Corée du sud, Etats-Unis, Hollande, Italie, Japon, Pays-Bas, Tchéquie et Suède. Le fait que des auteurs non-francophones aient choisi le français pour présenter leurs travaux montre une fois de plus que les JEP sont un lieu d'échange riche et convivial.

4.3. Les JEP de la cognition ?

Chaque conférence est teintée par le laboratoire organisateur... L'ICP situe l'étude de la parole au centre d'un triptyque signaux-langage-cognition. La parole peut être en effet considérée comme un système de communication qui émerge de contraintes imposées par le système biophysique permettant de véhiculer du sens en produisant et percevant des gestes audibles, visibles, voire touchables. L'étude de la parole s'abreuve ainsi constamment de techniques instrumentales et de simulation, de méthodologies d'expérimentation, de connaissances sur le vivant et de cadres théoriques. Grâce à nos collègues en psychologie expérimentale et cognitive dont la participation fut massive aux JEP de Bruxelles et de Martigny, nous avons pu apprécier les dimensions cognitives désormais indispensables et inséparables de notre champ de recherches. Les nouvelles techniques d'imagerie cérébrale permettent maintenant de confirmer les schémas fonctionnels suggérés par la psychologie et de faire des hypothèses sur les structures corticales dans les tâches de production ou de perception de parole, montrant ainsi les recrutements étonnants des aires motrices en perception et vice versa.

C'est pourquoi la cognition était l'un des thèmes privilégiés de l'appel à communications. La conférence invitée de J.F. Demonet (*Imageries fonctionnelles cérébrales : vers une physiologie de la cognition humaine*) — remplacé à brûle-pourpoint et avec brio par C. Abry — répondait à cette initiative. Initiative encouragée par l'apparition pour la première fois aux JEP d'une session orale intitulée *Parole et cognition*. L'introduction de cette nouvelle discipline dans nos jeux est une manifestation de la volonté constante d'ouverture scientifique qui caractérise les JEP depuis leur création.

Cet éclairage sur l'ancrage physiologique de la cognition humaine a été complété par la conférence invitée de A. Di Cristo (*Interpréter la prosodie*) qui a insisté sur la prise en compte du contexte situationnel de l'interaction verbale dans l'étude des signaux prosodiques : le *comment dire* de l'acte de langage, œuvrant en coopération avec le contenu linguistique du discours, ne prend sens qu'une fois interprété au sein des représentations mentales de l'auditeur ou construites lors de l'interaction.

Un autre séminaire d'ouverture a été donné par A. De Cheveigné (Analyse de scène auditive et parole) faisant ainsi écho à celui de J.-L. Gauvain (Systèmes de reconnaissance à grands vocabulaires : progrès et défis) : les tâches de transcription de dialogues de plus en plus réalistes auxquelles les systèmes de reconnaissance automatique de parole sont confrontés permettent d'estimer à la fois les progrès considérables accomplis grâce aux modèles statistiques et la distance qui reste à parcourir pour égaler notre capacité à analyser une scène audiovisuelle.

Si l'analyse d'une scène audiovisuelle complexe est un défi majeur des technologies d'indexation automatique des multiples documents que notre "société de l'information" produit et stocke chaque jour, la synthèse de scènes audiovisuelles est un marché en pleine expansion. L'usage de réalités virtuelles est généralisé dans l'industrie du jeu, et l'émergence de la norme MPEG4 suppose que nos télécommunications seront elles aussi virtuelles dans quelques dizaines d'années. Déjà en 1997, C.-E. Liedtke et son équipe de l'Université de Hanovre montraient, dans son intervention au Symposium International sur la Télévision à Montreux : "How far are we from the virtual actor and the virtual scene?" que les progrès en simulation numérique, modélisation biomécanique et synthèse graphique permettent de piloter des clones vidéo-réalistes de sujets humains à partir de mouvements estimés à partir de caméras multiples. Lors des JEP, le premier prix *Christian Benoît* remis à Tony Ezzat du Medialab (MIT) a permis de montrer que nous n'étions pas loin du locuteur virtuel ...

Ce numéro spécial offre un éventail d'articles qui, s'ils sont loin de couvrir entièrement la diversité des recherches sur la communication parlée, illustrent cependant des aspects aussi différents que la modélisation du langage pour l'amélioration de la reconnaissance automatique de parole (Zitouni), la caractérisation de la prosodie dans le cas de la parole émotionnelle (Mozziconacci et Hermes) ou dans le cas spécifique d'un patient aphasique produisant un jargon phonémique (Louis, Di Cristo, Habib et Hirst), les différences de nature de l'accès lexical aux noms de personnes et d'objets (Evrard), ou encore la nature de la conscience phonologique chez les enfants dyslexiques phonologiques (Sabater, Daffaure, De Martino et Rey).

Nous espérons que cette brève introduction vous donnera envie de lire les articles de cette édition spéciale de *In Cognito* qui rassemble des versions étendues d'une sélection de papiers présentés à Aussois. Ne manquent plus que le son et l'image ! et la lecture de ces articles vous auront sensibilisé, intéressé ou même gagné à cette aventure passionnante qu'est la compréhension des mécanismes de production et de perception de ce formidable outil de communication qu'est la parole. Et pour paraphraser notre histrion des JEP2000, Joël Martin ("Faits cosmiques et faibles masses") : *Que votre verbe soit en joie !*