

JEP-TALN-RECITAL 2012

JEP : Journées d'Études sur la Parole
TALN : Traitement Automatique des Langues Naturelles
RECITAL : Rencontre des Étudiants Chercheurs en Informatique
pour le Traitement Automatique des Langues

Actes de la conférence conjointe JEP-TALN-RECITAL 2012

Volume 4 : conférences invitées

Éditeurs

Laurent Besacier
Hervé Blanchon
Marie-Paule Jacques
Nathalie Vallée
Gilles Sérasset

4 – 8 Juin 2012
Grenoble, France

© 2012 Association Francophone pour la Communication Parlée (AFCP) et
Association pour le Traitement Automatique des Langues (ATALA)

Des versions imprimées de ces actes peuvent être achetées auprès de :

GETALP-LIG
Laurent Besacier
BP 53
38041 Grenoble Cedex 9
France
Laurent.Besacier@imag.fr

Préface

Pour la quatrième fois, après Nancy en 2002, Fès en 2004, et Avignon en 2008, l'AFCP (Association Francophone pour la Communication Parlée) et l'ATALA (Association pour le Traitement Automatique des Langues) organisent conjointement leurs principales conférences afin de réunir en un seul lieu les deux communautés du traitement de la parole et de la langue écrite pour favoriser les interactions entre nos deux communautés.

Plus précisément, la conférence JEP-TALN-RECITAL'2012 réunit cette année la vingt-neuvième édition des Journées d'Étude sur la Parole (JEP'2012), la dix-neuvième édition de la conférence sur le Traitement Automatique des Langues Naturelles (TALN'2012) et la quinzième édition des Rencontres des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL'2012).

Nous avons souhaité organiser cet événement sur le campus universitaire de l'université de Grenoble, au plus proche des trois laboratoires co-organisateurs (LIG, LIDILEM, GIPSA-Lab). L'université Stendhal-Grenoble 3 (consacrée aux disciplines des humanités) nous accueille dans ses locaux à cette occasion.

Par ailleurs, JEP-TALN-RECITAL'2012 accueille quatre ateliers ; la septième édition du « Défi Fouille de Texte » (DEFT), la seconde édition du « Défi Geste Langue et Signe » (DEGELS), ainsi que deux nouveaux auxquels nous souhaitons longue vie : « Interactions Langagières pour personnes Agées Dans les habitats Intelligents » (ILADI) et « Traitement Automatique des Langues Africaines – écrit et parole » (TALAF). Quatre conférenciers renommés ont accepté notre invitation pour des sessions plénières communes. Nous espérons que leur hauteur de vue et leur ouverture d'esprit permettront des discussions intéressantes et ouvriront des perspectives prometteuses.

Quelques informations sur les processus de sélection pour cette édition sont présentées ci-dessous. Nous remercions tous les relecteurs et membres des différents comités de programme pour leur travail ainsi que nos sociétés savantes : l'AFCP et l'ATALA (avec son comité permanent qui assure la continuité de la forme et du fond entre les diverses éditions).

Nous avons reçu 62 propositions d'articles longs pour TALN, parmi lesquels 24 ont été sélectionnés au moyen d'un processus de relecture consciencieux, soit un taux de sélection de 39 %. 61 articles courts ont été soumis parmi lesquels 29 ont été sélectionnés au moyen d'un processus de relecture identique à celui des articles longs, soit un taux de sélection de 48 %. Comme lors de l'édition précédente de TALN, les articles courts seront présentés sous forme de sessions orales brèves (2 minutes par publication) et de poster. 10 démonstrations seront également présentées au cours d'une session dédiée.

Concernant les JEP, 145 propositions ont été reçues. À l'issue de la réunion du comité de programme qui s'est tenue à Grenoble les 15 et 16 mars, 108 articles ont été sélectionnés (74%). 28 articles seront présentés en session orale et 80 lors de sessions poster.

La désaffection grandissante des soumissions à RECITAL nous a conduit à proposer plusieurs innovations afin de remobiliser nos jeunes chercheurs. Tout d'abord, l'appel à communication a été étendu pour permettre la soumission de travaux préliminaires, projets de thèse, travaux des premiers mois de recherche (états de l'art, premières pistes...). Ensuite le processus de relecture a été modifié pour offrir à nos jeunes des relectures pédagogiques (encouragements, pistes) et permettre des échanges directs avec les relecteurs (relectures non-anonymes). Ces changements ont été accueillis très favorablement puisque nous avons reçu 42 propositions de communications parmi lesquelles 11 feront l'objet de présentations orales (27%) et 17 de présentations sous forme de poster (40%). Nous sommes également revenus à des sessions RECITAL spécifiques qui ne sont pas en parallèle avec des sessions TALN.

En ce qui concerne les actes, nous avons fourni de nouveaux styles optimisés pour une lecture à l'écran. Bien que les habitudes des auteurs aient été changées à cette occasion, nous espérons que les lecteurs nous feront des retours d'usage positifs. Un meilleur référencement des travaux présentés a aussi été l'une de nos préoccupations; aussi avons-nous choisi de les faire référencer par l'ACL (*Association for Computational Linguistics*) dans l'*ACL Anthology*¹ pour une meilleure visibilité.

Nous vous souhaitons, chers lecteurs, un parcours passionnant et passionné au fil des nombreuses pages de ces actes et, pourquoi pas, des découvertes inattendues grâce au hasard et à votre sagacité; découvertes qui seront les graines de nouvelles idées pour faire progresser nos champs de recherche.

Laurent Besacier, Président JEP

Hervé Blanchon & Georges Antoniadis, Présidents TALN

Didier Schwab & Jorge Mauricio Molina Mejia, Présidents RECITAL

1. <http://www.aclweb.org/anthology/>

Le mot de la présidente de l'Association pour le Traitement Automatique des Langues

L'Association pour le Traitement Automatique des Langues (ATALA²) soutient depuis 1959 les travaux de recherche fondamentale et appliquée en linguistique informatique.

En complément des travaux sur les modèles informatiques de la langue, il est primordial pour l'ATALA de renforcer ses liens avec des domaines connexes tels que le traitement de la parole ou la représentation des connaissances.

Ceci est d'autant plus important à un moment où, avec l'avènement des technologies de l'Internet et de l'information, les données écrites et parlées, qu'il était jusqu'alors très difficile de recueillir sont devenues, en un laps de temps très court, pléthores et très faciles d'accès. En quelques années seulement, nous sommes passé du rêve, avoir accès à plus de données, au cauchemar, avoir trop de données. L'Internet et l'utilisation généralisée des bases de données sont aujourd'hui la cause principale de la croissance exponentielle et continue des données en ligne.

De nos jours, grâce aux logiciels embarqués la plupart des types de dispositifs électroniques que nous utilisons quotidiennement sont en mesure de fournir des données pérennes. En effet, alors qu'auparavant la plupart des données disparaissaient après avoir été utilisées dans un but précis, les données sont maintenant stockées, fusionnées, distribuées et même revendues pour être analysées et interprétées dans le meilleur des cas, à des fins d'innovation ou d'avancée scientifique.

Dans un contexte en constante mutation, l'organisation conjointe entre l'AFCP et l'ATALA des journées TALN permet aux deux communautés d'échanger leurs méthodes d'analyse et de compréhension de ces données textuelles ou parlées afin de faire progresser la recherche en proposant de nouvelles méthodes et de nouveaux algorithmes sur lesquels s'appuyer pour développer de nouvelles technologies et services dans le domaine de l'analyse intelligente des données.

Frédérique Segond
Présidente de l'ATALA

2. <http://www.atala.org/>

Le mot de la présidente de l'Association Francophone de la Communication Parlée

Chers collègues,

Après les éditions de 1970 (1^{ères} JEP), 1979 (10^{èmes} JEP), et avec en 2000 un détour à Aussois (23^{èmes} JEP), les Journées d'Etude sur la Parole sont de retour à Grenoble !

L'AFCP (Association Francophone de la Communication Parlée³) se réjouit de s'associer de nouveau à l'ATALA (Association pour le Traitement Automatique des Langues) pour l'organisation de cet événement commun que sont les JEP-TALN-RECITAL. Rappelons que depuis 2002, les communautés du traitement de la langue, orale comme écrite, se retrouvent périodiquement en un même lieu afin favoriser les échanges et stimuler l'émergence de projets de recherche commun. Les éditions passées, à Nancy en 2002, à Fès en 2004, à Avignon en 2008, ont été un réel succès et nous gageons que cette édition JEP-TALN-RECITAL'2012 sera de nouveau un moment fort de rencontres et d'échanges fructueux entre les différents acteurs de nos communautés.

Pour ce qui concerne cette 29^{ème} édition des Journées d'Etude sur la Parole, 145 communications ont été soumises, ce qui est très satisfaisant (136 soumissions en 2010 à Mons, 130 en 2008 à Avignon). L'origine variée des soumissions (majoritairement de France, mais aussi de Belgique, de Suisse, du Canada, des Etats-Unis, de Tunisie, du Maroc, ...) souligne une fois encore le caractère international de ces journées francophones, qui est une priorité de l'AFCP. Sur ces 145 soumissions, 108 ont été retenues, ce qui donne un taux d'acceptation de 74% qui est similaire à celui de l'édition précédente. La couverture thématique des papiers retenus est vaste et reflète le dynamisme et la diversité des recherches sur la parole dans la communauté francophone.

Pour rappel, les communications aux JEP sont sélectionnées sur la base d'un article complet. Chaque soumission est évaluée par deux relecteurs. Le comité de programme, constitué des membres du CA de l'AFCP et de membres du comité d'organisation, se réunit pendant deux jours pour examiner les soumissions et leurs évaluations, certaines sont relues par un 3^{ème} lecteur, et la sélection finale est effectuée. Les communications sélectionnées sont alors groupées par thèmes afin de définir les sessions thématiques de la conférence, et pour chaque session, des communications orales sont choisies. Les autres communications, qui seront présentées sous forme de posters, ne sont pas regroupées thématiquement de façon à avoir des sessions poster couvrant un large spectre d'intérêts. Il est donc à noter qu'aux JEP la sélection entre communication orale et affichée s'effectue principalement sur la base d'un choix thématique pour les sessions orales et ne renvoie donc pas à un critère de qualité.

3. L'Association Francophone de la Communication Parlée (AFCP) est une structure d'animation et de réflexion de la communauté francophone travaillant sur la parole. <http://www.afcp-parole.org/>

Pour ces JEP, outre les traditionnelles bourses proposées aux étudiants et jeunes chercheurs, nous renouvelons notre action d'invitation de jeunes chercheurs appartenant à des laboratoires situés hors de France. Cinq jeunes chercheurs venant de Tunisie et d'Algérie ont été ainsi sélectionnés sur dossier et nous auront le plaisir de les accueillir à ces rencontres. Nous aurons également l'honneur de remettre lors de ces journées les prix de thèse édition 2010 et 2011, à Gwénolé Lecorvé et Juliette Kahn, respectivement.

Pour finir, l'AFCP est ravie de voir cette 29^{ème} édition des Journées d'Etude sur la Parole se tenir à Grenoble. Grenoble est depuis longtemps un haut lieu de la recherche sur la parole et a toujours eu un rôle important dans la structuration et l'animation de notre communauté parole, tant au niveau national, qu'au niveau international. Après des restructurations difficiles du pôle parole grenoblois, nous ne pouvons que nous réjouir que l'ensemble des laboratoires grenoblois, sous l'impulsion du LIG, ait entrepris l'aventure commune qu'est l'organisation de cet événement important pour la communauté francophone. Au nom de l'AFCP, je tiens donc à remercier sincèrement tous les organisateurs de ces Journées, le LIG, le LIDILEM et le GIPSA-Lab et en particulier Laurent Besacier, pour son dynamisme et son investissement dans cette entreprise.

Au nom du comité de programme, je remercie aussi vivement les 114 relecteurs pour leur temps et leur travail fait dans un esprit constructif.

Enfin, je tiens à remercier tous les auteurs, conférenciers, et participants qui sont le moteur de notre communauté scientifique si sympathique.

Je vous souhaite à tous des journées et des rencontres enrichissantes et stimulantes.

Cécile Fougeron
Présidente de l'AFCP
Présidente du Comité de Programme des XXIX^{èmes} JEP

Comité d'organisation de JEP-TALN-RECITAL'2012 :

Georges ANTONIADIS (LIDILEM, Université Grenoble 3)
Véronique AUBERGÉ (Gipsa-Lab, CNRS)
Valérie BELYNCK (LIG-GETALP, Grenoble INP)
Laurent BESACIER (LIG-GETALP, Université Grenoble 1)
Hervé BLANCHON (LIG-GETALP, Université Grenoble 2)
Francis BRUNET-MANQUAT (LIG-GETALP, Université Grenoble 2)
Emmanuelle ESPERANÇA-RODIER (LIG-GETALP, Université Grenoble 1)
Jérôme GOULIAN (LIG-GETALP, Université Grenoble 2)
Marie-Paule JACQUES (LIDILEM, Université Grenoble 3)
Olivier KRAIF (LIDILEM, Université Grenoble 3)
Alexandre LABADIÉ (LIG-GETALP, CNRS)
Thomas LEBARBÉ (LIDILEM, Université Grenoble 3)
Benjamin LECOUEUX (LIG-GETALP, Université Grenoble 2)
Mathieu MANGEOT (LIG-GETALP, Université De Savoie)
Jorge Mauricio MOLINA MEJIA (LIDILEM, Université Grenoble 3)
Claude PONTON (LIDILEM, Université Grenoble 3)
François PORTEY (LIG-GETALP, Grenoble INP)
Solange ROSSATO (LIG-GETALP, Université Grenoble 3)
Isabelle ROUSSET (LIDILEM, Université Grenoble 3)
Didier SCHWAB (LIG-GETALP, Université Grenoble 2)
Frédérique SEGOND (Pôle Innovation Viseo)
Gilles SÉRASSET (LIG-GETALP, Université Grenoble 1)
Agnès TUTIN (LIDILEM, Université Grenoble 3)
Michel VACHER (LIG-GETALP, CNRS)
Nathalie VALLÉE (Gipsa-Lab, CNRS)
Virginie ZAMPA (LIDILEM, Université Grenoble 3)

Comité de programme de JEP'2012 :

Présidents :

Laurent BESACIER (LIG-GETALP, Université Grenoble 1, France)
Cécile FOUGERON (LPP Paris)
Guillaume GRAVIER, IRISA et CNRS-INRIA Rennes)

Membres :

Gilles ADDA (LIMS1, Paris)
Melissa BARKAT-DEFRADAS (PRAXILING, Montpellier)
Loïc BARRAULT (LIUM, Le Mans)
Philippe BOULA DE MAREUIL (LIMS1, Paris)
Véronique BOULENGER (DDL Lyon)
Elisabeth DELAIS-ROUSSARIE (Lab. Linguistique Formelle, Paris)
Véronique DELVAUX (Univ. Mons, Belgique)

Didier DEMOLIN (Gipsa-Lab, Grenoble)
Laurence DEVILLERS (LIMSI, Paris)
Isabelle FERRANE (IRIT, Toulouse)
Emmanuel FERRAGNE (CLILAC-ARP, Paris)
Corinne FREDOUILLE (LIA, Avignon)
Bernard HARMEGNIES (Univ. Mons, Belgique)
Fabrice HIRSCH (PRAXILING, Montpellier)
Thomas HUEBER (Gipsa-Lab, Grenoble)
Irina ILLINA (LORIA, Nancy)
David LANGLOIS (LORIA, Nancy)
Georges LINARES (LIA, Avignon)
Hélène LOEVENBRUCK (Gipsa-Lab, Grenoble)
Egidio MARSICO (DDL, Lyon)
Sylvain MEIGNIER (LIUM, Le Mans)
Christine MEUNIER (LPL, Aix en Provence)
Yohann MEYNADIER (LPL, Aix en Provence)
François PELLEGRINO (DDL, Lyon)
Pascal PERRIER (Gipsa-Lab, Grenoble)
François PORTET (LIG-GETALP, Grenoble)
Solange ROSSATO (LIG-GETALP, Grenoble)
Sophie ROSSET (LIMSI, Paris)
Marc SATO (Gipsa-Lab, Grenoble)
Christophe SAVARIAUX (Gipsa-Lab, Grenoble)
Christine SÉNAC (IRIT, Toulouse)
Rudolph SOCK (IPS, Strasbourg)
Annemie VAN HIRTUM (Gipsa-Lab, Grenoble)
Béatrice VAXELAIRE (IPS, Strasbourg)
Chakir ZEROUAL (LPP Paris et Univ. Sidi Mohamed Ben-abdellah, Fes, Maroc)

Relecteurs additionnels :

Martine ADDA-DECKER, LPP et LIMSI Paris)
Régine ANDRE-OBRECHT (IRIT, Toulouse)
Angélique AMELOT (LPP, Paris)
Corine ASTESANO (Univ. Toulouse 2 et LPL, Aix en Provence)
Véronique AUBERGÉ (LIG et GIPSA-Lab, Grenoble)
Nicolas AUDIBERT (LPP, Paris)
Gérard BAILLY (Gipsa-Lab, Grenoble)
Claude BARRAS (LIMSI, Paris)
Denis BEAUTEMPS (Gipsa-Lab, Grenoble)
Nathalie BEDOIN (DDL, Lyon)
Roxane BERTRAND (LPL, Aix en Provence)
Benjamin BIGOT (LIA, Avignon)
Frédéric BIMBOT (IRISA et CNRS-INRIA Rennes)
Anne BONNEAU (LORIA, Nancy)
Hélène BONNEAU-MAYNARD (LIMSI, Paris)
Hervé BREDIN (LIMSI, Paris)

Nathalie CAMELIN (LIUM, Le Mans)
Christian CAVE (LPL, Aix en Provence)
Claire PILLOT-LOISEAU (LPP, Paris)
Lise CREVIER-BUCHMAN (LPP, Paris)
Mariapaola D'IMPERIO (LPL, Aix en Provence)
Paul DELÉGLISE (LIUM, Le Mans)
Christian DICANIO (UC Berkeley, États-Unis)
Cong-Thanh DO (LIMSI, Paris)
Christelle DODANE (PRAXILING, Montpellier)
Driss MATROUF (LIA, Avignon)
Sophie DUFOUR (LPL, Aix en Provence)
Elie EL-KHOURY (LIUM, Le Mans)
Robert ESPESSER (LPL, Aix en Provence)
Yannick ESTÈVE (LIUM, Le Mans)
Martine FARACO (LPL, Aix en Provence)
Jérôme FARINAS (IRIT, Toulouse)
Dominique FOHR (LORIA, Nancy)
Teddy FURON (IRISA et CNRS-INRIA Rennes)
Maeva GARNIER (Gipsa-Lab, Grenoble)
Cedric GENDROT (LPP, Paris)
Alain GHIO (LPL, Aix en Provence)
Antoine GIOVANNI (CHU Marseille et LPL Aix en Provence)
Laurent GIRIN (Gipsa-Lab, Grenoble)
Pierre HALLE (LPP, Paris)
Sophie HERMENT (LPL, Aix en Provence)
Daniel HIRST (LPL, Aix en Provence)
Kathy HUET (Univ. Mons, Belgique)
Stephane HUET (LIA, Avignon)
Denis JOUVET (LORIA, Nancy)
Juliette KAHN (LNE Paris)
Sophie KERN (DDL, Lyon)
Hélène LACHAMBRE (IRIT, Toulouse)
Muriel LALAIN (LPL, Aix en Provence)
Antoine LAURENT (LIUM, Le Mans)
Gwénoélé LECORVE (IDIAP Martigny (Suisse))
Thierry LEGOU (LPL, Aix en Provence)
Christophe LÉVY (LIA, Avignon)
Alain MARCHAL (LPL, Aix en Provence)
Odile MELLA (LORIA, Nancy)
Ilya OPARIN (LIMSI, Paris)
Caterina PETRONE (LPL, Aix en Provence)
Myriam PICCALUGA (LPL, Aix en Provence)
Julien PINQUIER (IRIT, Toulouse)
Serge PINTO (LPL, Aix en Provence)
Agnès PIQUARD-KIPFFER (LORIA, Nancy)

Michel PITERMANN (LPL, Aix en Provence)
Rachid RIDOUANE (LPP Paris)
Albert RILLIARD (LIMSI, Paris)
Mickael ROUVIER (LIUM, Le Mans)
Jérémi SAUVAGE (PRAXILING, Montpellier)
Jean-Luc SCHWARTZ (Gipsa-Lab, Grenoble)
Grégory SENAY (LIA, Avignon)
Willy SERNICLAES (ULB Bruxelles, Belgique)
Marion TELLIER (LPL, Aix en Provence)
Michel VACHER (LIG Grenoble)
Nathalie VALLÉE (Gipsa-Lab, Grenoble)
Anne VILAIN (Gipsa-Lab, Grenoble)
Coriandre VILAIN (Gipsa-Lab, Grenoble)
Emmanuel VINCENT (IRISA et CNRS-INRIA Rennes)
Pauline WELBY (LPL, Aix en Provence)

Comité de programme de TALN'2012 :

Présidents :

Georges ANTONIADIS (LIDILEM, Université Grenoble 3, France)
Hervé BLANCHON (LIG-GETALP, Université Grenoble 2, France)

Membres :

Nicholas ASHER (IRIT, CNRS et Université Toulouse 3)
Frédéric BÉCHET (LIF, Aix Marseille Université)
Yves BESTGEN (Université Catholique de Louvain, Louvain-la-Neuve, Belgique)
Philippe BLACHE (LPL, CNRS et Université de Provence)
Christian BOITET (LIG-GETALP, Université Grenoble 1)
Malek BOUALEM (France Telecom Orange Labs, Lannion)
Narjès BOUFADEN (KeaText, Montréal, Canada)
Yllias CHALI (University of Lethbridge, Lethbridge, Canada)
Laurence DANLOS (ALPAGE, Université Paris 7)
Piet DESMET (ITEC, K.U.Leuven et K.U.Leuven KULAK, Belgique)
Mark DRAS (Macquarie University, Sydney, Australie)
Denys DUCHIER (LIFO, Université d'Orléans)
Marc DYMETMAN (XRCE, Grenoble)
Dominique ESTIVAL (University of Western Sydney, Sydney, Australie)
Cédrick FAIRON (Université Catholique de Louvain, Louvain-la-Neuve, Belgique)
Olivier FERRET (CEA LIST, Palaiseau)
Michel GAGNON (École Polytechnique de Montréal, Montréal, Canada)
Claire GARDENT (LORIA, Villers lès Nancy)
Nabil HATOUT (CLLE-ERSS, CNRS et Université Toulouse II)
Sylvain KAHANE (MODYCO-ALPAGE, Université Paris 10)
Laura KALLMEYER (Heinrich-Heine-Universität, Düsseldorf, Allemagne)
Mathieu LAFOURCADE (LIRMM, Université Montpellier 2)
Philippe LANGLAIS (DIRO, Université Montréal, Canada)
Guy LAPALME (RALI, Université Montréal, Canada)

Yves LEPAGE (IPS, Université Waseda, Japon)
Emmanuel MORIN (LINA, Université Nantes)
Adeline NAZARENKO (LIPN, Université Paris 13)
Luka NERIMA (LATL, Université Genève, Suisse)
Alain POLGUÈRE (Université de Lorraine et ATILF CNRS)
Laurent PRÉVOT (LPL, CNRS et Université de Provence)
Violaine PRINCE (LIRMM, Université Montpellier 2)
Jean-Philippe PROST (LIRMM, Université Montpellier 2)
Christian RETORÉ (LaBRI et INRIA, Université Bordeaux 1)
Sophie ROSSET (LIMSI, CNRS)
Didier SCHWAB (LIG-GETALP, Université Grenoble 2)
Holger SCHWENK (LIUM, Université du Maine, Le Mans)
Pascale SÉBILLOT (IRISA, INSA de Rennes)
Gilles SÉRASSET (LIG-GETALP, Université Grenoble 1)
Agnès TUTIN (LIDILEM, Université Grenoble 3)
Anne VILNAT (LIMSI, CNRS et Université Paris Sud)
François YVON (LIMSI, CNRS et Université Paris Sud)
Virginie ZAMPA (LIDILEM, Université Grenoble 3)
Pierre ZWEIGENBAUM (LIMSI, CNRS et INALCO)

Comité Scientifique de TALN'2012 :

Présidents :

Georges ANTONIADIS (LIDILEM, Université Grenoble 3, France)
Hervé BLANCHON (LIG-GETALP, Université Grenoble 2, France)

Membres :

Les membres du comité de programme aidés de . . .

Ramzi ABBES (Techlimes, Lyon)
Stergos AFANTENOS (IRIT, Université de Toulouse)
Salah AIT-MOKHTAR (XRCE, Grenoble)
Maxime AMBLARD (LORIA, Université de Lorraine)
Jean-Yves ANTOINE (LI, Université de Tours et Lab-STICC, CNRS)
Delphine BATTISTELLI (STIH, Université Paris 4)
Denis BECHET (LINA, Université de Nantes)
Patrice BELLOT (LSIS, Université Aix-Marseille)
Delphine BERNHARD (LiPa, Université de Strasbourg)
Romaric BESANÇON (CEA-LIST, Saclay Nano-Innov)
Brigitte BIGI (LPL, Aix en Provence)
Julien BOURDAILLET (Xerox, États-Unis)
Caroline BRUN (XRCE, Grenoble)
Francis BRUNET-MANQUAT (LIG-GETALP, Université Grenoble 2)
Marie CANDITO (Alpage, Université Paris Diderot)
Thierry CHANIER (LRL, Clermont Université)
Vincent CLAVEAU (IRISA-CNRS, Rennes)
Nathalie COLINEAU (CSIRO ICT Centre, Marsfield, Australie)
Benoît CRABBÉ (Alpage, Paris 7)

Béatrice DAILLE (LINA, Université de Nantes)
Pascal DENIS (Alpage)
Iris ESHKOL-TARAVELLA (LLL, Université d'Orléans)
Cécile FABRE (CLLE-ERSS, Université Toulouse 2)
Benoit FAVRE (LIF, Université Aix-Marseille)
Dominic FOREST (Université de Montréal, Canada)
Karen FORT (INIST et LIPN, Paris 13)
George FOSTER (CNRC, Gatineau, Canada)
Nuria GALA (LIF, Université Aix-Marseille)
Bruno GAUME (CLLE-ERSS, Université Toulouse 2)
Éric GAUSSIER (LIG-GETALP, Université Grenoble 1)
Kim GERDES (LPP, Université Paris 3)
Jérôme GOULIAN (LIG-GETALP, Université Grenoble 2)
Benoît HABERT (ICAR, ENS Lyon)
Najeh HAJLAOUI (Institut de recherche Idiap, Martigny, Suisse)
Thierry HAMON (LimetBio, Université Paris 13)
Marie-Paule JACQUES (LIDILEM, Université Grenoble 1)
Guillaume JACQUET (XRCE, Grenoble)
Christine JACQUIN (LINA, Université de Nantes)
Adel JEBALI (Université Concordia, Montréal, Canada)
Leïla KOSSEIM (Université Concordia, Montréal, Canada)
Olivier KRAIF (LIDILEM, Université Grenoble 3)
Éric LAPORTE (LIGM, Université Paris-Est Marne-la-Vallée)
Dominique LAURENT (Synapse, Toulouse)
Thomas LEBARBÉ (LIDILEM, Université Grenoble 3)
Anne-Laure LIGOZAT (LIMSI, ENSIE)
Cédric LOPEZ (LIRMM, Université Montpellier 2)
Mathieu MANGEOT (LIG-GETALP, Université de Savoie)
Denis MAUREL (LI, Université de Tours)
Aurélien MAX (LIMSI, Université Paris-Sud)
Jasmina MILIĆEVIĆ (OLST, Dalhousie University, Canada)
Laura MONCEAUX (LINA, Université de Nantes)
Richard MOOT (LaBRI et SIGNES, Bordeaux)
Erwan MOREAU (Trinity College Dublin, Irlande)
Fabienne MOREAU (IRISA, Université Rennes 2)
Véronique MORICEAU (LIMSI, Université Paris-Sud)
Philippe MULLER (IRIT, Université de Toulouse)
Alexis NASR (LIF, Université Aix-Marseille)
Aurélié NÉVÉOL (NCBI, National Library of Medicine, États-Unis)
Jian-Yun NIE (RALI, Université de Montréal, Canada)
Cécile PARIS (CSIRO ICT Centre, Marsfield, Australie)
Yannick PARMENTIER (LIFO, Université d'Orléans)
Guy PERRIER (LORIA, Université de Lorraine)
Sylvain POGODALLA (LORIA, Vandoeuvre-lès-Nancy)
Thierry POIBEAU (LaTTiCe, Montrouge)
Claude PONTON (LIDILEM, Université Grenoble 3)

Andrei POPESCU-BELIS (Institut de recherche Idiap, Martigny, Suisse)
Carlos RAMISCH (LIG-GETALP, Grenoble)
Mathieu ROCHE (LIRMM, Université Montpellier 2)
Antoine ROZENKNOP (LIPN, Université Paris 13)
Benoît SAGOT (Alpage, INRIA Roquencourt)
Djamé SEDDAH (Alpage, Université Paris 4)
Kamel SMAÏLI (LORIA, Université de Lorraine)
Xavier TANNIER (LIMSI, Université Paris-Sud)
Isabelle TELLIER (LaTTiCe, Université Paris 3)
Juan-Manuel TORRES-MORENO (LIA, Université d'Avignon et des Pays de Vaucluse)
François TROUILLEUX (LRL, Université Clermont-Ferrand 2)
Lonneke VAN DER PLAS (IMS, Université de Stuttgart, Allemagne)
Fabienne VENANT (LORIA, Université Nancy 2)
Jacques VERGNE (GREYC, Université de Caen)
Éric VILLEMONTÉ DE LA CLERGERIE (Alpage, INRIA Roquencourt)
Eric WEHRLI (LATL, Université de Genève, Suisse)
Guillaume WISNIEWSKI (LIMSI, Université Paris-Sud)
Imed ZITOUNI (IBM T.J. Watson Research Center, Yorktown Heights, États-Unis)
Michael ZOCK (LIF, Marseille)
Amal ZOUAQ (Royal Military College of Canada et Athabasca University, Canada)
Mounir ZRIGUI (UTIC, Faculté des Sciences de Monastir, Tunisie)
Sandrine ZUFFEREY (ILC, Université Catholique de Louvain-la-Neuve, Belgique)

Comité de programme de RECITAL'2012 :

Présidents :

Jorge Mauricio MOLINA MEJIA (LIDILEM, Université Stendhal – Grenoble 3)
Didier SCHWAB (GETALP-LIG, Université Pierre Mendès France – Grenoble 2)

Membres :

Vanessa ANDRÉANI (Société CFH et laboratoire ERSS, Université Toulouse 2 – Le Mirail)
Nicolas AUDIBERT (Laboratoire de Phonétique et Phonologie-CNRS, Université Sorbonne-Nouvelle)
Frédéric BÉCHET (Laboratoire d'Informatique Fondamentale de Marseille, Université d'Aix-Marseille)
Patrice BELLOT (LSIS, Université d'Aix-Marseille)
Valérie BELYNCK (GETALP-LIG, Grenoble INP)
Farah BENAMARA (IRIT, Université Toulouse 3)
Christian BOITET (GETALP-LIG, Université Joseph Fourier – Grenoble 1)
Leila BOUTORA (LPL, Université d'Aix-Marseille, Marseille)
Francis BRUNET-MANQUAT (GETALP-LIG, Université Pierre Mendès France – Grenoble 2)
François-Régis CHAUMARTIN (Société Proxem, Laboratoire Alpage, UMR INRIA, Université Paris 7)
Gaël DE CHALENDAR (CEA LIST, Palaiseau)
Achille FALAISE (GETALP-LIG, Société Floralis, Université Joseph Fourier-Grenoble 1)
Olivier FERRET (CEA LIST, Palaiseau)
Nuria GALA (LIF, Université d'Aix-Marseille)
Jérôme GOULIAN (GETALP-LIG, Université Pierre Mendès France – Grenoble 2)
Thierry HAMON (LIM&BIO, Université Paris 13)
Nicolas HERNANDEZ (LINA, CNRS 6241, Nantes)

Bernard JACQUEMIN (CREM, Université de Haute Alsace, Mulhouse)
Olivier KRAIF (LIDILEM, Université Stendhal – Grenoble 3)
Alexandre LABADIÉ (GETALP-LIG, Grenoble)
Mathieu LAFOURCADE (LIRMM, Université de Montpellier 2)
Guy LAPALME (RALI, Université de Montréal, Canada)
François LAREAU (CLT, Macquarie University, Australie)
Thomas LEBARBÉ (LIDILEM, Université Stendhal – Grenoble 3)
Benjamin LECOUTEUX (LIG-GETALP, Université Pierre Mendès France – Grenoble 2)
Yves LEPAGE (Université Waseda, Japon)
Mathieu LOISEAU (LIDILEM, Université Stendhal – Grenoble 3)
Cédric LOPEZ (LIRMM, Université Montpellier 2)
Denis MAUREL (Université François Rabelais Tours)
Aurélien MAX (LIMSI-CNRS & Université Paris-Sud)
Jean-Luc MINEL (MoDyCO, UMR 7114, Université Paris-Ouest Nanterre La Défense – CNRS)
Emmanuel MORIN (LINA, CNRS 6241, Nantes)
Yayoi NAKAMURA-DELLOYE (LCAO, Université Paris VII)
Claude PONTON (LIDILEM, Université Stendhal-Grenoble 3)
François PORTET (GETALP-LIG, Grenoble INP)
Laurent PREVOT (LPL, Université d'Aix-Marseille, Marseille)
Violaine PRINCE (LIRMM, Université Montpellier 2)
Jean-Philippe PROST (LIRMM, Université Montpellier 2)
Bali RANAIVO-MALANÇON (Universiti Sarawak Malaysia, Malaisie)
Christian RETORÉ (LaBRI, Université Bordeaux 1)
Mathieu ROCHE (LIRMM, Université Montpellier 2)
Solange ROSSATO (GETALP-LIG, Université Stendhal – Grenoble 3)
Azim ROUSSANALY (LORIA, Université de Lorraine)
Isabelle ROUSSET (LIDILEM, Université Stendhal – Grenoble 3)
Fatiha SADAT (Université du Québec à Montréal, Canada)
Tristan VANRULLEN (TVSI, Marseille)
Eric WEHRLI (LATL, Université de Genève, Suisse)
Virginie ZAMPA (LIDILEM, Université Stendhal – Grenoble 3)
Haifa ZARGAYOUNA (LIPN, Université Paris 13)
Michael ZOCK (CNRS-LIF, Marseille)
Mounir ZRIGUI (Faculté des Sciences, Université de Monastir, Tunisie)
Pierre ZWEIGENBAUM (LIMSI-CNRS, Orsay)

Conférenciers invités :

Ian Maddieson (Université de Californie, Berkeley, États-Unis)
Jacqueline Léon (Laboratoire d'histoire des théories linguistiques, CNRS, Paris)
Yoshinori Sagisaka (Université de Waseda, Japon)
Hans Uszkoreit (DFKI, Sarrebruck, Allemagne)

Sponsors :



Table des matières

<i>Quality Translation for a Multilingual Continent - Priorities and Chances for European MT Research</i> Hans Uszkoreit	1
<i>Tensions entre théorie et pratique dans les systèmes de TAL. Étude historique et épistémologique</i> Jacqueline Léon	3
<i>Trans-disciplinary spoken language processing studies for scientific understanding of second language learner's characteristics</i> Yoshinori Sagisaka	5
<i>The Origin and Distribution of Complexity of Phonological Structure</i> Ian Maddieson	7

Quality Translation for a Multilingual Continent Priorities and Chances for European MT Research

Hans Uszkoreit^{1, 2}

(1) German Research Center for Artificial Intelligence (DFKI)

(2) Saarland University

Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany

uszkoreit@dfki.de

ABSTRACT

Recent progress in translation technology has caused a real boost for research and technology deployment. At the same time, other areas of language technology also experience scientific advances and economic success stories.

However, research in machine translation is still less affected by new developments in core areas of language processing than could be expected. One reason for the low level of interaction is certainly that the predominant research paradigm in MT has not started yet to systematically concentrate on high quality translation. Most of the research and nearly all of the application efforts have focused on solutions for informational inbound translation (assimilation MT).

This focus has on the one hand enabled translation of information that normally is not translated at all. In this way MT has changed work and life of many people without ever infringing on the existing translation markets.

In my talk I will present a new research approach dedicated to the analytical investigation of existing quality barriers. Such a systematic thrust can serve as the basis of scientifically guided combinations of technologies including hybrid approaches to transfer and the integration of advanced methods for syntactic and semantic processing into the translation process. Together with improved techniques for quality estimation, the expected results will drive translation technology into the direction badly needed by the multilingual European society.

KEYWORDS : machine translation, hybrid approaches, syntactic processing, semantic processing, quality estimation.

Tensions entre théorie et pratique dans les systèmes de TAL. Étude historique et épistémologique

Jacqueline Léon

CNRS, Université Paris Diderot, Sorbonne Paris Cité,
Histoire des Théories Linguistiques

jacqueline.leon@linguist.jussieu.fr

RÉSUMÉ

Le TAL est tenu pour une activité pratique hétérogène susceptible d'associer dans un même système des méthodes et des théories linguistiques ou philosophiques parfois contradictoires.

Dans mon exposé, je vais tâcher d'expliciter cette tension en m'appuyant sur une étude de tout premiers systèmes de TAL, modèles de traduction automatique et d'extraction d'information, mis au point dans les années 1950. On examinera les questions suivantes :

- Quelles sont les implications pour un système de TAL quand il a été conçu comme « technologie opérationnelle », en-dehors de tout cadre théorique linguistique, et qu'il a été doté de moyens importants devant être expertisé et évalué régulièrement ? Cela veut-il dire qu'il a été conçu en dehors de tout lien avec des disciplines scientifiques ?
- Quelles sont les implications pour un système de TAL quand la mesure de ses limites en tant qu'application pratique est prise en considération ?
- Comment faire co-exister des théories contradictoires, comme ça a été le cas, par exemple, pour la construction de langues intermédiaires à base sémantique pour la traduction automatique, associant théories du sens a priori et théories du sens en contexte ?

Trans-disciplinary spoken language processing studies for scientific understanding of second language learner's characteristics

Yoshinori Sagisaka

Global Information and Telecommunication Institute

Waseda University, Tokyo, Japan

sagisaka@waseda.jp

ABSTRACT

Language and speech proficiency is considered to be an important factor to identify human beings. Though traditional studies on language and speech can reveal many aspects of their characteristics, we have not yet had a complete view of human's language and speech ability. I believe that trans-disciplinary studies will enable us to have its scientific modeling.

In this talk, I would like to introduce our research efforts on segmental duration control as an example of research towards computational human modeling. The computational modeling of segmental duration that we have been studying around three decades not only contributes to prosody control in speech synthesis technology but also gives an integrated view of individual timing characteristics studied in phonetic science. Together with duration control modeling, a series of perceptual studies on duration modifications needed for model evaluation have suggested us a unified view of scientific understanding on rhythm and timing. Through the introduction of our current efforts on the objective evaluation of 2nd language (L2) proficiency in speech timing control, we will see that these models and findings are useful for L2 learning and acquisition.

To conclude my talk, I finally introduce research consortium called AESOP (Asian English Speech cOrpus Project) where researchers in different fields (speech science, informatics, phonetics, psychology and language education) have started to work together by collecting commonly sharable L2 language and speech data.

The Origin and Distribution of Complexity of Phonological Structure

Ian MADDIESON^{1, 2}

(1) University of California, Berkeley

(2) University of New Mexico, Albuquerque

ianm@berkeley.edu

ABSTRACT

Simple measures of phonological complexity have been shown to frequently correlate with each other such that greater elaboration of, say, the size of a consonant inventory or of a vowel inventory co-occurs with greater numbers of tone contrasts when a large ensemble of languages is studied. Further, these measures correlate positively with elaboration of the syllable canon. Most linguists would probably approve the suggestion by Comrie (1992), that it is easier to imagine that “earlier stages of human language [differed] qualitatively from those spoken today in being less complex”.

This talk will consider to what extent it is possible to usefully speculate about the origins of phonological complexity. One tool for such speculation is the examination of the patterns of distribution of measures of complexity around the world. If language originated with humans in Africa does a hypothesis of original simplicity suggest African languages should be the simplest? In fact, African languages have a higher mean basic vowel inventory than languages elsewhere, and the highest proportion of tone languages. These factors enable a correlation to be found between on the one hand a measure of complexity weighing these factors equally with consonant inventory size and on the other hand the distance from a presumed origin in Africa (Atkinson 2011), with complexity declining with distance. Atkinson’s suggested analogy to a genetic ‘founder effect’ is unconvincing for multiple reasons. Deep phylogenetic relationships between languages mirroring the population genetic patterns have not (yet) proved recoverable and the logic of phoneme-loss-by-distance is unpersuasive.

Various alternative hypotheses linking phonological patterns with external factors have been proposed. Those that will be explored include the effects of population size, social structure, language contact situation and environmental conditions. These ideas will be evaluated in light of data in a database assembling information on over 700 languages, as well as results from smaller projects addressing syllable structure and sonority patterns in selected language samples.

Index

Léon, Jacqueline, [3](#)

Maddieson, Ian, [7](#)

Sagisaka, Yoshinori, [5](#)

Uszkoreit, Hans, [1](#)