

## Patrick Paroubek

Né le: 23/02/1961  
Ingénieur de Recherche CNRS (IR1)  
email:pap@limsi.fr  
URL: <http://www.limsi.fr/Individu/pap>

LIMSI-CNRS  
BP 133 - 91403 ORSAY Cedex - France  
Bâtiment 508 Université Paris XI  
fax: (33) (0)1 69 85 80 88  
tél: (33) (0)1 69 85 80 04

### Motivation pour une HDR

Ingénieur de Recherche au CNRS, affecté à la recherche dans le groupe ILES (Information Langues Écrites et Signées) du LIMSI, j'ai une activité professionnelle (publication, animation scientifique, enseignement, encadrement) qui est comparable à celle d'un chercheur en Traitement Automatique des Langues, pour ce qui concerne la recherche appliquée, c'est à dire dans mon cas l'évaluation de systèmes à composante langagière, l'analyse automatique du langage et la fouille de sentiments et d'opinions. L'obtention d'une Habilitation à Diriger les Recherches me permettrait sur le plan professionnel de pouvoir encadrer de nouveaux thésards, car j'ai, par le passé, encadré les thèses de Julien Poudade (soutenue le 21 septembre 2006) et Alexander Pak (soutenue le 13 juin 2012) ainsi que de m'impliquer plus dans la vie du campus.

### Table des matières

Titres Universitaires	1
Parcours professionnel	2
Enseignement	3
Administration	4
Activités liées à la recherche	5
Contrats et conventions de recherche	6
Encadrement	7
Synthèse des travaux	8
Publications	8-13

### Titres Universitaires

- Université Pierre & Marie Curie (Paris 6)
  - Doctorat d'Informatique au LITP (devenu depuis le LIP6), décembre 1989
  - DEA d'informatique, stage au LITP (juin 1986),
  - Maîtrise d'informatique (juin 1985) ,
  - Licence d'informatique (juin 1984) ,
  - DEUG Science et Structure de la Matière (juin 1983).

## Parcours professionnel

### • LIMSI-CNRS

*septembre 2002 - présent*

- Ingénieur de recherche au CNRS, je travaille dans le groupe ILES (Information Langues Écrites et Signées, qui s'appellait avant 2009 le groupe LIR pour Langues Information Représentations) du département Communication Homme-Machine du LIMSI (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur), où mes fonctions concernent le développement de méthodes pour l'évaluation des systèmes de traitement des informations en langue naturelle ainsi que l'analyse de sentiments et la fouille d'opinions. J'ai été responsable des activités d'évaluation dans le projet ANR PASSAGE (2006-2010) sur l'analyse syntaxique et dans le cadre du projet DOXA (CAP DIGITAL 2009-2011) sur la fouille d'opinion. Au niveau Européen, je participe aux projets Flarenet et METANET pour ce qui concerne le développement d'une infrastructure d'évaluation pour les technologies du langage. Par le passé, j'ai aussi participé à une campagne d'évaluation sur le dialogue parlé (MEDIA) et aux activités du sous-comité de normalisation ISO/TC37/SC4 sur les standards en annotation de ressources linguistiques. La première thèse que j'ai encadrée concernait l'étude des conditions pour l'émergence d'une langue dans des communautés d'entités autonomes artificielles et la seconde portait sur la fouille d'opinions et son évaluation.

### • LIMSI - CNRS

*juillet 1997 - septembre 2002*

- Ingénieur de Recherche dans le groupe Traitement du Langage Parlé, du département Communication Homme-Machine du LIMSI : réalisation d'études de l'état de l'art sur l'évaluation en ingénierie linguistique, évaluation des systèmes de dialogue oral (projets Européens DISC et DISC2), gestion scientifique du projet européen ELSE (spécification d'une infrastructure européenne pour l'évaluation en ingénierie linguistique), animation de la tâche transverse sur l'état de l'art en évaluation pour l'ingénierie linguistique pour le projet européen CLASS, participation à la conception, la gestion et l'animation de la campagne d'évaluation GRACE (campagne d'évaluation des analyseurs morpho syntaxiques du français), initiateur et responsable du projet MULTITAG (valorisation du corpus produit dans GRACE, production d'un corpus annoté d'un million de mots), participation en tant qu'expert en ingénierie linguistique au projet SILFIDE, collaboration CNRS-AUF, dont l'objectif est la mise en place de serveurs de ressources linguistiques multi-sites sur Internet, participation en tant qu'expert en ingénierie linguistique au comité scientifique TECHNOLANGUE. Représentant français au comité de standardisation l'ISO/TC159/SC5/WG3, évaluation ergonomique des systèmes de communication parlée.

### • INaLF - CNRS

*novembre 1993 - juin 1997*

- Ingénieur de Recherche à l'Institut National de la Langue Française (INaLF). Informatisation du Trésor de la Langue Française (conception et réalisation d'interface graphique), développement et évaluation d'étiqueteurs morpho-syntaxiques pour le français (co-organisation de l'action GRACE) ainsi que la construction, l'annotation, la gestion et la distribution par Internet de gros corpus textuels. Participation au projet européen PAROLE (MLAP-63-386) et LE-PAROLE (LE2-4017) en tant qu'expert linguistique, pour construire un corpus textuel multilingue annoté de grande taille.

### • Université de Rouen

*octobre 1992 - septembre 1993*

- ATER au Laboratoire d'Informatique de Rouen (LIR). Enseignement de programmation Objet et de C++ en DESS.

- **Université de Pennsylvanie**
  - *juillet 1992 - septembre 1992*
    - Chercheur invité au laboratoire LINC du Prof. A. Joshi. Préparation de la nouvelle version (1.0) de l'atelier logiciel d'analyse syntaxique XTAG.
  - **Computers Communications and Visions (c2v)**
    - *février 1992 - mai 1992*
      - Ingénieur de recherche dans une start-up informatique. Maintenance de code C pour Genese, un générateur de systèmes experts sous UNIX/SUN/X Window. Participation au lancement (analyse et gestion de projet) du projet européen de traitement automatique de la langue EUROTRA ET-10/63 sur l'application de méthodes probabilistes et de méthodes à base de corpus pour la construction de lexiques et de règles de préférences (extraction terminologique).
- **BULL SA France**
  - *octobre 1991 - décembre 1991*
    - Ingénieur logiciel. Évaluation et mise à niveau dans un environnement UNIX/C, d'un compilateur pour la notation syntaxique abstraite ASN.1 utilisée pour décrire les données dans les couches supérieures du protocole de communication OSI entre ordinateurs.
- **Université de Rouen**
  - *février 1991 à septembre 1991*
    - ATER au Laboratoire d'Informatique de Rouen (LIR), enseignement de Pascal et d'Algorithmique en DEUG.
- **Université de Pennsylvanie**
  - *février 1990 à décembre 1990*
    - Post-doctorat sous la direction du Professeur A. Joshi pour concevoir et développer le logiciel XTAG, l'interface X Window d'un atelier d'analyse syntaxique basée sur le formalisme TAG (Tree Adjoining Grammar).
- **Université Malaya, Kuala Lumpur, Malaysia**
  - *novembre 1986 à décembre 1987*
    - Maître de conférence, responsable de deux cours semestriels : gestion de projet (« Project Management and Installation ») et génie logiciel (« Software Engineering »).

## Enseignement

- **Licence**
  - *2007- présent*
    - Institut National des Langues et Civilisations Orientales (INALCO)- Cours d'introduction au traitement automatique des langues et à la linguistique de corpus (2h<sup>1/2</sup>/semaine).
- Master**
  - 2005 - 2011*
    - Master Recherche d' Informatique de l'Université PXI, cours sur l'évaluation des systèmes à composante langagière (3h).
- DESS**
  - février 2003*
    - Cours sur l'évaluation des analyseurs syntaxiques (4h) dans le DESSDEFI de l'Université Paris X.

### **Ecole d'été**

*juillet 2002*

- École d'été organisée par le réseau d'excellence européen ELSNET sur l'évaluation en ingénierie linguistique (Odense, Danemark). Cours sur l'évaluation des étiqueteurs morpho-syntaxiques et syntaxiques.

### **Journée**

*Novembre 2002*

- Séminaire à la journée TELMI en sur l'ingénierie linguistique, co-organisée par l'Université de Lille et la Chambre de Commerce de la région Nord-Pas de Calais.

### **DESS**

*2001-2003*

- Cours sur l'ingénierie linguistique (6h) dans le DESS 2IC de l'Université Technologique de Compiègne.

### **DESS**

*1992-1993*

- ATER au Laboratoire d'Informatique de Rouen (LIR). Enseignement de programmation Objet et de C++ en DESS.

### **DEUG**

*1990-1991*

- ATER au Laboratoire d'Informatique de Rouen (LIR), enseignement de Pascal et d'Algorithmique en DEUG.

### **Master**

*1986-1987*

- Université Malaya, Kuala Lumpur, Malaysia, maître de conférence, responsable de deux cours semestriels : gestion de projet (« Project Management and Installation ») et génie logiciel (« Software Engineering »).

## **Administration**

### • Laboratoire

- 2012, membre du conseil du département Communication Homme-Machine du LIMSI.
- 2005 - 2008, membre du conseil du laboratoire du LIMSI.
- 2005 - 2008, membre du conseil du département Communication Homme-Machine du LIMSI.
- 2002 - 2005, représentant des ITA au conseil scientifique du département Informatique de l'Université Paris XI

### • Comités

- Depuis juin 2008, vice-président de l'Association française pour le Traitement Automatique des Langues (ATALA), membre depuis 1995.

## Activités liées à la recherche

- Comités

- Président en 2008 et 2009 du comité d'organisation annuel du Défi Fouille de Texte (DEFT), membre depuis 2007
- 2007, 2009, 2011, comité de programme de la 3<sup>ème</sup> Language & Technology Conference, Poznan, Poland
- Depuis 1998, membre du comité scientifique de l'International Conference on Language Resources and Evaluation (LREC).
- Depuis juin 2003, membre du comité de rédaction de la revue TAL de l'Association française pour le Traitement Automatique des Langues (ATALA).
- 1999, 2005, 2006, 2007, 2008, membre du comité de programme de la conférence sur le Traitement Automatique du Langage Naturel (TALN)
- Mars 2004. Co-organisateur de la journée ATALA (Association pour le Traitement Automatique des Langues) intitulée AGENTAL (Agents et Langue), sur les relations entre systèmes Multi-Agents et Traitement Automatique des Langues.
- Mai 2004, co-organisateur de la journée ATALA (Association pour le Traitement Automatique des Langues) intitulée EVANS (Evaluation des Analyseurs syntaxiques).
- Depuis 2002, co-organise le séminaire du département Communication Homme-Machine du LIMSI.
- Co-organisateur du ACL/EACL 2001 CLASS workshop on Evaluation for Language and Dialogue Systems, les 6-7 juillet, Toulouse.
- En 2001, organisateur du Bullet Course on the paradigm of Evaluation in Speech and Language Engineering, Paris 02-03 juillet, cours international sur l'évaluation en ingénierie linguistique.
- En 2000, organisateur du Satellite Workshop: "Using Evaluation within HLT Programs: Results and Trends", in collaboration with the ELSNET Network and ELRA, the European Language Resource Association, Athènes, 30 mai 2000.
- 2003-2004. Membre du comité scientifique de l'Action Concertée Incitative « Masses de Données » (collaboration Ministère délégué à la Recherche et aux Nouvelles Technologies, INRIA, CNRS).
- Juin 2003. Membre du comité scientifique de l'atelier sur l'évaluation des analyseurs syntaxiques de la 10<sup>ème</sup> conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN'03), Batz-sur-Mer, juin 2003.
- Membre du comité scientifique de l'ISCA Tutorial and Research Workshop Multi-Modal Dialogue in Mobile Environments, June 17 - 21, 2002, Kloster Irsee, Germany.
- 2002 - 2007, membre du sous-comité ISO/TC37/SC4 sur les standards d'annotation de ressources linguistiques.
- Depuis 2001, membre du comité scientifique, de l'action transversale VENISE (réalité virtuelle) du LIMSI.
- En 2001, responsable scientifique du réseau FRANCIL, réseau en ingénierie linguistique des pays francophones (collaboration CNRS AUF).
- 2000 - 2005, membre du comité scientifique du programme TECHNOLANGUE (évaluation et ressources en ingénierie linguistique, du Ministère délégué à la Recherche et aux Nouvelles Technologies, Ministère de la Culture et de la Communication et du Ministère délégué à l'Industrie).

- 2000 - 2004, Représentant français, du groupe de travail ISO/TC159/SC5/WG3 sur l'évaluation ergonomique des systèmes de communication parlée.
- Membre du comité de programme de l'atelier, « Towards a european evaluation infrastructure for NL and speech processing », International Conference on Language Resources and Evaluation (LREC), Grenade, Spain, 2000.
- 1999-2000, réseau FRANCIL, réseau en ingénierie linguistique des pays francophones (collaboration CNRS / AUF)
- Depuis 1998, animateur de l'action transversale CORVAL (corpus et évaluation) du LIMSI, devenue l'action COPTE (CORpus Parole/Texte et Evaluation) depuis avril 2004.
- En 1998, membre du comité de programme de la conférence COLING-ACL '98, 10-14 août 1998, Montréal, Québec, Canada.
- Membre du comité de lecture des conférences: CIRSE09, ESSLI09, INTERSPEECH09, RNTI09, INTERSPEECH08, SIGDIAL09, SIGDIAL08, SIGDIAL07.

## • Contrats et conventions de recherches

### **DOXA - Projet CAP DIGITAL**

*2009-2011*

- \* Le projet vise spécifier et développer des composants, des ressources et des chaînes de traitement qui permettront détecter automatiquement les thèmes abordés ainsi que les sentiments et opinions exprimés dans de grands volumes de textes rédigés en français en anglais. Les technologies développées seront évaluées et intégrées dans une nouvelle version de la plate-forme INFOM@GIC. Dans ce projet, je suis responsable des activités d'évaluation.

### **ANR MDCA - Projet PASSAGE**

*2006-2010*

- \* Évaluation des analyseurs syntaxiques du français. Constitution d'un corpus arboré (« Treebank ») de grande taille et organisation de 2 campagnes d'évaluation. Dans ce projet, je suis responsable des activités d'évaluation.

### **TECHNOLANGUE/EVALDA**

*2003-2006*

- \* Coordonnateur scientifique. Projet EASY : évaluation des analyseurs syntaxiques.
- \* Projet MEDIA : évaluation de la compréhension dans les systèmes de dialogue oral.
- \* Projet EQueR : évaluation de systèmes de question-réponse.

### **RNTL - Projet SenseNet**

*2002-2005*

- \* Étude linguistique des usages d'Internet, c.-à-d. parcours d'exploration, profils d'internautes etc.

### **ACI TCAN MdrNT/INRIA/CNRS - Projet APCOLAN**

*2003-2004*

- \* Compréhension du langage et simulations agents.

### **INaLF-LIMSI - Projet projet ILF**

*(2002-2004*

- \* Les dictionnaires électroniques.

### **FP5 - Projet CLASS**

*2000-2002*

- \* État de l'art sur l'évaluation en ingénierie linguistique.

### **FP5 - Projet DISC2**

*2000*

\* Valorisation des résultats du projet DISC.

**CNRS - MULTITAG**

1999-2000

\* Coordonateur. Valorisation du corpus produit dans l'action GRACE.

**FP5 - ELSNET**

1998-2002

\* Réseau d'excellence européen en ingénierie linguistique.

**CNRS/AUF - Projet SILFIDE**

1998-1999

\* Serveur distribué de ressources linguistiques sur Internet.

**FP5 - Projet ELSE**

1998

\* Responsable scientifique. Proposition d'infrastructure européenne pour l'évaluation en ingénierie linguistique.

**FP5 - DISC**

1997-1999

\* État de l'art en conception et évaluation de systèmes de dialogue oral.

**CNRS - Projet GRACE**

1995-1998

\* Co-coordonateur. Campagne d'évaluation des étiqueteurs morpho-syntaxiques du français.

**FP4 - Projet LE-PAROLE**

1995-1997

\* Constitution d'un corpus multi-lingue aligné.

**FP4 - PAROLE**

1994

\* Constitution d'un corpus multilingue aligné.

• Encadrement de thèse

- 2009 - *présent*. Encadrement de la thèse de doctorat d'Alexander Pak, intitulée « *De l'évaluation des systèmes de fouille d'opinion.* ».
- 2002-2006. Co-encadrement de la thèse de doctorat de Julien Poudade (soutenue le 21/09/06), intitulée « *Emergence d'un lexique dans une population d'agents autonomes par l'action, la perception, et la contingence sensorimotrice* ».
- En 2002, contribution à l'encadrement des thèses de Gilles André et Eric Moretti.

• Encadrement de M2

- 2011 Raoum Khiari, « *Annotation pour l'analyse d'opinion.* »
- 2011 Rémi Durand, « *Définition et implémentation d'un outil de réécriture de graphes sémantiques dans un système de compréhension de textes.* »
- 2010 Marine Leglouanec, « *Annotation Annotation syntaxique des disfluences de l'oral* »
- 2009
  - \* Estelle Leton « *Annotation syntaxique d'un corpus de référence pour l'évaluation d'analyseurs syntaxiques en mode boîte noire avec des mesures quantitatives et objectives* ».
  - \* Thuy Linh Ngo « *Définition de formalisme d'annotation pour l'évaluation des analyseurs d'opinions et de sentiments sur corpus* ».
- 2005
  - \* Lionel Landwerlin « *Émergence du langage dans une population d'agents autonomes situés et découverte sensori-motrice* ».

- \* Jean-Beguec «Plateforme logicielle de pilotage et de simulation d’une expérience d’émergence du langage ».
- \* Patrice Jacq «Langage et sensations pour des agents situés ».
- Encadrement de M1
  - 2010, Erika Guilbert, « Traduire des annotations syntaxiques d’un formalisme à l’autre. »

## Synthèse des travaux

Mes travaux ont toujours portés sur le langage et les ordinateurs. Au départ purement informatiques, mes travaux concernaient les langages formels et les langages de programmation (thèse sur la sémantique des mécanismes d’exception des langages de programmation). La transition vers le traitement automatique des langues a commencé avec mon post-doc à Philadelphie dans le laboratoire d’A. K. Joshi (inventeur du formalisme TAG pour l’analyse syntaxique), où j’ai réalisé une interfaces graphique sous X11 pour le développement de grammaires TAG. Après deux années d’ATER à l’Université de Rouen, passées à enseigner les langages de programmation, cette transition s’est poursuivie avec mon recrutement au CNRS à l’INALF, où j’ai participé à l’informatisation du Trésor de la Langue Française (dictionnaire en 16 volumes) et elle s’est concrétisée avec ma mutation au LIMSI, où mes travaux portent sur l’évaluation des systèmes de traitement de l’information à composante langagière, avec la création et l’animation de communautés de chercheurs regroupées lors de campagnes d’évaluation comparative nationales ou internationales sur l’analyse automatique du langage. Les prolongements que j’envisage pour mes travaux portent d’une part sur le développement de procédure d’évaluation pour des systèmes d’analyse automatique du langage se rapprochant de plus en plus de l’interprétation humaine (complexité croissante de campagnes d’évaluation, d’abord sur l’étiquetage morpho-syntaxique, puis la syntaxe, et maintenant la fouille d’opinion), d’autre part sur une formalisation du processus d’évaluation pour l’analyse automatique du langage en vue de son automatisation complète et le développement d’une infrastructure d’évaluation pour les technologies du langage au niveau Européen et finalement sur le développement de modèles et d’algorithmes pour l’analyse de sentiments et la fouille d’opinions.

## Publications

- **Revue internationale (2)**  
*avec comité de lecture*
  - (BII) Alexander Pak, Delphine Bernhard, Patrick Paroubek, Cyril Grouin, “A Combined Approach to Emotion Detection in Suicide Notes” to appear in Biomedical Informatics Insights, 2012.
  - (NLE) Lori Lamel, Wolfgang Minker and Patrick Paroubek. ”Towards Best Practice in the Development and Evaluation of Speech Recognition Components of a Spoken Language Dialogue System”. *Revue Natural Language Engineering*, Vol. 6, part 3, pp 305-322, oct. 2000.
- **Revue francophones (6)**  
*avec comité de lecture*
  - (TAL) Alexander Pak, Patrick Paroubek, « Le microblogage pour la microanalyse des sentiments et des opinions », *Traitement Automatique des Langues*, 2010, vol. 51.3 (pp 75-100).
  - (TAL) Olivier Hamon, Patrick Paroubek, Djamel Mostefa, «SEWS : un Serveur d’Evaluation orienté Web pour la Syntaxe », in *Revue Traitement Automatique des Langues*, TAL. Vol. 49, no 2, 2008, pp 1-40



- (RNTI) Cyril Grouin, Martine Hurault-Plantet, Patrick Paroubek, et Jean-Baptiste Berthelin, «DEFT'07 : une campagne d'évaluation en fouille d'opinion », in Roche, M. et Poncelet, P. (ed). Fouille de données d'opinion, Revue des Nouvelles Technologies de l'Information, Cepaduès éditions, RNTI-E-17, 2009, pp 1-24.
- (TAL) Stéphane Chaudiron, Lynette Hirschman, Patrick Paroubek, Introduction, numéro spécial "Principes de l'évaluation en Traitement Automatique des Langues", Revue Traitement Automatique des Langues, vol. 47, no 3, à paraître en juin 2007.
- (TAL) Laurence Devillers, Hélène Maynard, and Patrick Paroubek. "Méthodologies d'évaluation des systèmes de dialogue parlé: réflexions et expériences autour de la compréhension". Revue Traitement Automatique des Langues, vol.43, no 2, pp 155-184, 2002.
- (Langues) Gilles Adda, Joseph Mariani, Patrick Paroubek, Martin Rajman, Josette Lecomte, "L'action GRACE d'évaluation de l'assignation de parties du discours pour le français", revue Langues, Vol. 2, No. 2, juin 1999, pp 119-129.

### ● Ouvrages (3)

- Mariani J., Paroubek P., Francopoulo G., Max A., Yvon F., Zweigenbaum P., The French language in the digital age / La Langue française à l'ère du numérique, p 102, European Languages Whitepaper Series, Springer, 2012.
- P. Paroubek, "Light Motif - Motif sans peine", ed. Computers Communications and Visions (C2V), 1994, pp 1-352.
- P. Paroubek, "Portabilité + Adaptabilité + Modifiabilité Réutilisabilité - Application à la compilation des mécanismes de manipulation de l'histoire d'un calcul", thèse de doctorat de l'Université Pierre et Marie Curie (Paris 6), Décembre 1989, Rapport interne IBP-LITP 9124, pp 1-166.

### ● Chapitres d'ouvrages (9)

- J-J. Mariani, G. Francopoulo, A. Max, P. Paroubek, F. Yvon, P. Zweigenbaum, French Language in the Digital Age, in print, Springer 2012
- Alexander Pak, Patrick Paroubek, « Extracting Sentiment Patterns from Syntactic Graphs », chapter to appear in « Social Media Mining and Social Network Analysis: Emerging Research », IGI Global, 2012.
- A. Pak, P. Paroubek, Text Representation using dependency tree subgraphs for sentiment analysis, 16th International Conference on Database Systems for Advanced Applications (DASFAA 2011), Hong Kong, China, April 22-25, 2011, 323-332. Paru dans Database Systems for Adanced Applications, Xu J., Yu G., Zhou S., Unland R., Lecture Notes in Computer Science.
- Paroubek P., EASY la campagne d'évaluation des analyseurs syntaxiques, dans L'évaluation en technologies de la langue : les campagnes Technolanguage, Hermès Sciences, série Cognition et traitement de l'information (IC2), Paris, pp 117-138, ISBN 13 9782746219922.
- Patrick Paroubek, "Chapter 4 - Evaluating Part Of Speech Tagging and Parsing", Evaluation of Text and Speech Systems, Edts. Laila Dybkjær, Holmer Hemsén, Wolfgang Minker, series: Text, Speech and Language Technology , Vol. 36, Kluwer Academic Publisher, ISBN-10: 1-4020-5815-2, ISBN-13: 978-1-4020-5815-8, pp 97-116, disponible à partir du 04/07/2007.
- Patrick Paroubek, "Chapitre 4 - Analyse Morpho-syntaxique et syntaxique", dans l'ouvrage "L'évaluation des systèmes de traitement de l'information", sous la direction de Stéphane Chaudiron, Collection "Information Commande Communication", aux Editions Hermes-Sciences, Lavoisier, ISBN 2-7462-0862-8, juin 2004, pp 101-125.

- J. White, L. Hirschman, J. Mariani, A. Martin, P. Paroubek, M. Rajman, B. Sundheim, "Evaluation and assessment techniques", In: E. Hovy, N. Ide, R. Frederking, J. Mariani, A. Zampolli (Eds.) "Multilingual information management: Current levels and future abilities". *Linguistica Computazionale*, 2001.
- Patrick Paroubek, Martin Rajman, "Chapitre 5: Etiquetage morpho-syntaxique", dans l'ouvrage "Ingénierie des Langues", sous la direction de Jean-Marie Pierrel, Collection "Information Commande Communication", aux Editions Hermes Science ISBN 2-7462-0113-5, octobre 2000 pp 131-148.
- G. Adda, J. Lecomte, J. Mariani, P. Paroubek, M. Rajman, « Les procédures de mesure automatique de l'action Grace pour l'évaluation des assignateurs de parties du discours pour le français, In : K. Chibout, J. Mariani, N. Masson, F. Néel (Eds.) "Ressources et Evaluation en Ingénierie des Langues". Collection Champs Linguistiques. De Boec. 2000.

### • **Conférences internationales (20)**

*avec comité de lecture*

- Mariani J., Paroubek P, Francopoulo G., Max A., Yvon F., Zweigenbaum P., The White Paper on the French Language in the Digital Age, International Symposium on Multilingualism in Cyberspace (SIMC 2012), November 21-23, Paris, France, 2012.
- Paroubek P., Tannier X., A rough set formalization of quantitative evaluation with ambiguity. Eighth International Conference on Language Resources and Evaluation (LREC 2012), May 21-27, Istanbul, Turkey, 2012.
- Pak, A., Paroubek, P., Normalization of Term Weighting Scheme for Sentiment Analysis. Fifth Language and Technology Conference, Poznan, November 25-27, 2011.
- Paroubek, P., Pak, A., Mostefa, D. (2010). Annotations for Opinion Mining Evaluation in the Industrial Context of the DOXA project. LREC 2010.
- A. Pak, P. Paroubek, Twitter as a corpus for sentiment analysis and opinion mining, Seventh International Conference on Language Resources and Evaluation (LREC 2010), Valetta, Malta, May 17-23, 2010, 6p.
- A. Vilnat, P. Paroubek, E. Villemonte De La Clergerie, G. Francopoulo, M. Guénot, Passage syntactic representation: a minimal common ground for evaluation, Seventh International Conference on Language Resources and Evaluation (LREC 2010), Valetta, Malta, May 17-23, 2010, 6p
- Eric Villemonte de la Clergerie, Olivier Hamon, Djamel Mostefa, Christelle Ayache, Patrick Paroubek and Anne Vilnat, PASSAGE: from French Parser Evaluation to Large Sized Treebank, Proceedings of the Sixth International Language Resources and Evaluation (LREC'08), 2008, European Language Resources Association (ELRA), Marrakech, Morocco, May, 28-30.
- Patrick Paroubek and Isabelle Robba and Anne Vilnat and Christelle Ayache, EASY, Evaluation of Parsers of French: what are the results?, Proceedings of the 6<sup>th</sup> International Conference on Language Resources and Evaluation (LREC)2008, Marrakech, Morocco.
- Julien Poudade, Lionel Landwerlin, Patrick Paroubek, "Cognitive situated agents learn to name actions", Proceedings of the 17th European Conference on Artificial Intelligence, ECAI'2006, Trento, Italy, August 2006, IOS Press, Frontiers in Artificial Intelligence and Applications, vol. 141, pp 51-55, ISBN 1-58603-642-4.
- Patrick Paroubek, Isabelle Robba, Anne Vilnat, Christelle Ayache, "Data, Annotations and Measures in EASY - the Evaluation Campaign for Parsers of French", In proceedings of the fifth international conference on Language Resources and Evaluation (LREC 2006), ELRA, May, 2006, Genoa, Italy, pp 315-320

- Anne Vilnat, Patrick Paroubek, Laura Monceaux, Isabelle Robba, Véronique Gendner, Gabriel Illouz, Michèle Jardino, "The Ongoing Evaluation Campaign of Syntactic Parsing of French: EASY", actes de la Fourth International Conference on Language Resources and Evaluation (LREC), Lisboa, May 2004, vol. 6, pp 2023-2026.
- Claude Barras, Gilles Adda, Martine Adda-Decker, Benoit Habert, Philippe Boula de Mareuil, Patrick Paroubek, "Automatic audio and manual transcripts alignments, time-code transfer and selection of exact transcripts", actes de la Fourth International Conference on Language Resources and Evaluation (LREC), Lisboa, May 2004, vol. 3, pp 877-880.
- Laurence Devillers, Hélène Maynard, Sophie Rosset, Patrick Paroubek, Kevin Mc Tait, Djamel Mostefa, Khalid Choukri, Laurent Charnay, Caroline Bousquet, Nadine Vigouroux, Frédéric Béchet, Laurent Romary, Jean-Yves Antoine, Jeanne Villaneau, Myriame Vergnes, Jérôme Goulian, "The French MEDIA/EVALDA projet: the evaluation of the understanding capability of Spoken Language Dialog Systems", actes de la Fourth International Conference on Language Resources and Evaluation (LREC), Lisboa, May 2004, vol. 6, pp 2131-2134.
- V. Gendner, G. Illouz, M. Jardino, L. Monceaux, P. Paroubek, I. Robba, A. Vilnat, "PEAS, the first instantiation of a comparative framework for evaluating parsers of French", actes de la 10th Conference of The European Chapter of the Association for Computational Linguistics, Budapest, Hungary, April 12-17 2003, Companion Volume, pp 95-98.
- Véronique Gendner, Gabriel Illouze, Michèle Jardino, Laura Monceaux, Patrick Paroubek, Isabelle Robba, Anne Vilnat, "A Protocol for Evaluating Analyzers of Syntax (PEAS) " in Proceedings of the Third International Conference on Language Ressources and Evaluation (LREC), Las Palmas de Gran Canaria, Spain, May 27<sup>th</sup>-June 2<sup>nd</sup> 2002, vol 2, pp 590-597
- Patrick Paroubek, "Language Resources as by-Product of Evaluation: the MULTITAG example", Second International Conference on Language Resources and Evaluation (LREC2000), Athens, Greece, May 31-June 2nd 2000, Vol. 1, pp. 151-154.
- Laila Dybkjaer, Niels Ole Bernsen, R. Carlson, Lin Chase, N. Dahlback, Klaus Failenschmid, Ulrich Heid, Paul Heisterkamp, A. Jonsson, H Kamp, Inger Karlson, Jan van Kuppevelt, Lori Lamel, Patrick Paroubek, D. Williams, "The Disc Approach to Spoken Language System Development and Evaluation", actes de la First International Conference on Language Resources and Evaluation (LREC), Granada, May 1998, vol. 1 pp 185-189.
- B. Habert, G. Adda, M. Adda-Decker, P. Boula de Mareuil, S. Ferrari, O. Ferret, G. Illouz, P. Paroubek, "The need for tokenization evaluation", actes de la First International Conference on Language Resources and Evaluation (LREC), Granada, May 1998, vol. 1 pp 427-431.
- Patrick Paroubek, Josette Lecomte, Gilles Adda, Joseph Mariani, Martin Rajman, "The GRACE French Part-Of-Speech Tagging Evaluation Task", actes de la First International Conference on Language Resources and Evaluation (LREC), Granada, May 1998, vol. 1 pp 433-441
- P. Paroubek, Y. Schabes and A. K. Joshi, "XTAG - A Graphical Workbench for Developing Tree-Adjoining Grammars", actes de la conférence 3<sup>rd</sup> Conference on Applied Natural Language Processing, Trento, April 1992, pp 216-223.

• **Conférences nationales (15)**

*avec comité de lecture*

- Alexander Pak, Patrick Paroubek, « Sentiment Polarity Classification using Dependency Tree Subgraphs Text Representation' », TALN 2011. 18<sup>ème</sup> Conférence sur le Traitement Automatique des Langues Naturelles, Montpellier, France: 2011. -6p
- Alexander Pak, Patrick Paroubek, "Constructing French Affective Lexicon using Twitter", TALN 2010. 17<sup>ème</sup> Conférence sur le Traitement Automatique des Langues Naturelles, Montréal, Canada : 2010. - 6p

- Poudade J. Landwerlin L. Paroubek P. (2006). Des agents cognitifs situés apprennent désigner des actions. Actes du 15e congrès francophone Reconnaissance des Formes et Intelligence Artificielle, 25-26 janvier 2006, Tours.
- P. Paroubek and L.-G. Pouillot and I. Robba and A. Vilnat, EASY : Campagne d'évaluation des analyseurs syntaxiques, Proceedings of the 12<sup>ème</sup> conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN), mai 2005, Dourdan, France, pp 3-12.
- Julien Poudade, Patrick Paroubek, "Apprentissage collectif et lexique", actes de la 11<sup>ème</sup> conférence annuelle sur le Traitement Automatique des Langues Naturelles, Fez, Maroc, 19-22 Avril 2004, pp 387-392.
- Anne Vilnat, Laura Monceaux, Patrick Paroubek, Isabelle Robba, Véronique Gendner, Gabriel Illouz, Michèle Jardino "Annoter en constituants pour évaluer des analyseurs syntaxiques", actes de la 11<sup>ème</sup> conférence annuelle sur le Traitement Automatique des Langues Naturelles, Fès 19-22 Avril 2004, pp 467-476.
- Martine Adda-Decker, Benoit Habert, Claude Barras, Gilles Adda, Philippe Boula de Mareüil, Patrick Paroubek, "Une étude des disfluences pour la transcription automatique de la parole spontanée et l'amélioration des modèles de langage", actes des 25<sup>ème</sup> Journées d'étude sur la Parole (JEP), Fès, 19-22 Avril 2004, pp 1-4.
- Kevin Mc Tait, Hélène Maynard, Laurence Devillers, Sophie Rosset, Patrick Paroubek, Djamel Mostefa, Khalid Choukri, Jean-Yves Antoine, Frédéric Béchet, Caroline Bousquet, Nadine Vigouroux, Olivier Bontron, Laurent Charnay, Laurent Romary, Myriame Vergnes, "Constitution d'un corpus de dialogue oral pour l'évaluation automatique de la compréhension hors et en contexte du dialogue, actes des 25<sup>ème</sup> Journées d'étude sur la Parole (JEP), Fès, 19-22 Avril 2004, pp 357-360.
- Patrick Paroubek, Martin Rajman, "MULTITAG une ressource linguistique produit du paradigme d'évaluation", actes de la 7<sup>ème</sup> conférence annuelle sur le Traitement Automatique des Langues Naturelles, Lausanne 16-18 Octobre 2000, pp. 297-306.
- Gilles Adda, Josette Lecomte, Joseph Mariani, Patrick Paroubek, Martin Rajman, "Les procédures de mesure automatique de l'action GRACE pour l'évaluation des assignateurs de Parties du Discours pour le Français", Actes des 1<sup>ères</sup> Journées Scientifiques et Techniques du Réseau Francophone de l'Ingénierie de la Langue de l'Aupelf-Uref, Avignon, Avril 1997, pp 245-257.
- Gilles Adda, Joseph Mariani, Patrick Paroubek, Martin Rajman, Josette Lecomte, "Métrique et premiers résultats de l'évaluation GRACE des étiqueteurs morpho-syntaxiques pour le français;", Actes de la conférence sur le Traitement Automatique du Langage Naturel (TALN-99), Cargèse, Corse, France, 12-17 Juillet 1999. pp 15-24
- G. Adda, P. Blache, J. Mariani, P. Paroubek, M. Rajman, "Action GRACE - Mise en place du paradigme d'évaluation. Application au domaine de l'analyse morpho-syntaxique", Actes de la conférence sur le Traitement Automatique du Langage Naturel (TALN-95), Marseille, France, Juin 1995. pp 72-79
- G. Cattaneo, P. Paroubek, "The Software support to Medium Grain Parallelism: an evaluation", Proc. of the Fourth International Symposium on Computer and Information Sciences (ISCIS IV), Cesme, Turkey, October 1989.
- V. Loia, P. Paroubek, M. Quaggetto, F.X. Testard-Vaillant, "Pascal+: an Exception Oriented Language, Proc. of the Seventh Conf. of the Chilean Computer Society (SCCC), Santiago, June 1987.
- P. Paroubek, M. Quaggetto, F.X. Testard-Vaillant, "CLIP the language and its implementation", Proc. of the Eighth Conf. of the Chilean Computer Society (SCCC), Santiago, July 1988.

• **Ateliers internationaux (13)**

*avec comité de lecture*

- Pak A., Paroubek P., Language Independent Approach to Sentiment Analysis (LIMSI Participation in ROMIP-11), Dialogue 2012, à paraître, Moscou, 2012.
- Alexander Pak, Patrick Paroubek, “Twitter for Sentiment Analysis: When Language Resources are Not Available,” DEXA, 2011, 22nd International Workshop on Database and Expert Systems Applications, Toulouse, France: pp.111-115
- Alexander Pak, Patrick Paroubek. 2010. “Twitter Based System: Using Twitter for Disambiguating Sentiment Ambiguous Adjectives”. In Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval '10). Association for Computational Linguistics, Uppsala, Sweden: 436–439
- A. Vilnat and G. Francopoulo and O. Hamon and S. Loiseau and P. Paroubek and E. Villemonte de la Clergerie, Large Scale Production of Syntactic Annotation to Move Forward, Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation (PE 2008) in conjunction with COLING, August, 2008, Manchester, pp 36-43.
- E. Villemonte de la Clergerie and Christelle Ayache and Gaël de Chalendar and Gil Francopoulo and Claire Gardent and Patrick Paroubek, Large scale production of syntactic annotations for French, 2008, Hong-Kong, January, First Workshop on Automated Syntactic Annotations for Interoperable Language Resources at IGCL'08, pp 45-52,
- Philippe Boula de Mareüil, Benoît Habert, Frédérique Bénard, Martine Adda-Decker, Claude Barras, Gilles Adda, and Patrick Paroubek. A quantitative study of disfluencies in French broadcast interviews. In Proceedings of Disfluency In Spontaneous Speech (DISS) Workshop, Aix-en-Provence, September 2005.
- Anne Vilnat, Patrick Paroubek, Laura Monceaux, Isabelle Robba, Véronique Gendner, Gabriel Illouz et Michèle Jardino, “EASY or How Difficult Can It Be to Define a Reference Treebank for French”, actes du 2nd Workshop on Treebanks and Linguistic Theories (TLT), Växjö, Sweden, November 14th-15th, 2003.
- Martine Adda-Decker, Benoit Habert, Claude Barras, Gilles Adda, Philippe Boula de Mareüil, Patrick Paroubek, “A disfluency study for cleaning spontaneous speech automatic transcripts and improving speech language models”, actes de l’atelier “Disfluency in Spontaneous Speech”, pp 67-70, 2003, Göteborg, Sweden, 5-8 Septembre.
- Laurence Devillers, Hélène Maynard, Patrick Paroubek, and Sophie Rosset, “The PEACE SLDS understanding evaluation paradigm of the French MEDIA campaign”, actes de l’atelier “Evaluation Initiatives in Natural Language Processing, are evaluation methods, metrics, and resources reusable?”, at the 10th Conference of The European Chapter of the Association for Computational Linguistics, Budapest, Hungary, April 12-17 2003, pp 11-18.
- Patrick Paroubek, “Categorical Data-Specification for Control Task Formalization and Validation in Quantitative Black Box Evaluation”, actes de l’atelier satellite à LREC2000, “Using Evaluation whitin HLT Programs: Results and Trends”, Athens, Greece, May 30th 2000 pp. 50-52.
- Joseph Mariani, Patrick Paroubek, “Human Language Technologies Evaluation in the European Framework”, actes de l’atelier DARPA Broadcast News Workshop, Whashington, February 1999, Morgan Kaufman Publishers, ISBN-1-55860-638-6, pp 237-242.
- Josette Lecomte, Gilles Adda, Joseph Mariani, Patrick Paroubek, Martin Rajman, “Progress Report on the GRACE Evaluation Program for French Part-Of-Speech Taggers, actes de l’atelier Speech And Language Technology (SALT) Club Workshop, Halifax Hall, University of Sheffield, UK, June 17-18, 1997, pp 135-142.
- P. Paroubek, M. Quaggetto, F.X. Testard-Vaillant, “Implementing Stack based Softwares”, Proc. of the Expert System Seminar, University Malaya, Malaysia, December 1987.

- **Ateliers nationaux (7)**  
avec comité de lecture

- Paroubek P., Zweigenbaum P., Forest D., Grouin C., Indexation libre et contrôlée d’articles scientifiques - Présentation et résultats du défi fouille de textes DEFT2012, Atelier Défi Fouilles de Textes (DEFT) joint à la conférence TALN, pp 3-15, 2012
- C. Grouin, D. Forest, P. Paroubek, P. Zweigenbaum, Présentation et résultats du défi fouille de texte DEFT2011. Quand un article de presse a-t-il été écrit ? À quel article scientifique correspond ce résumé ?, Défi Fouille de Textes, Montpellier, France, TALN, 2011
- C. Grouin, D. Forest, L. Da Sylva, P. Paroubek, P. Zweigenbaum, Présentation et résultats du défi fouille de texte DEFT2010 : où et quand un article de presse a-t-il été écrit ?, DEFT’10. 6ème Défi Fouille de Textes, 2010
- C. Grouin, B. Arnulphy, J-B. Berthelin, S. El Ayari, A. Garcia-Fernandez, A. Grappy, M. Hurault-Plantet, P. Paroubek, I. Robba, P. Zweigenbaum, Présentation de l’édition 2009 du défi fouille de textes (DEFT’09), DEFT’09. Défi Fouille de Textes 2009
- P. Paroubek, J-B. Berthelin, S. El Ayari, C. Grouin, T. Heitz, M. Hurault-Plantet, M. Jardino, Z. Khalis, M. Lastes, Résultats de l’édition 2007 du Défi Fouille de Textes, DEFT’07. 3ème Défi Fouille de Textes, Grenoble, France
- V. Gendner, G. Illouz, M. Jardino, P. Paroubek, L. Monceaux, I. Robba, A. Vilnat, ”Proposition de protocole d’évaluation des analyseurs syntaxiques du français: PEAS”, actes de la 10ème conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN’03), Batz-sur-Mer, 14 juin 2003, Tome 2, pp 87-94.
- Actes de l’atelier ”Evaluation Methodologies for Language and Dialog Systems”, July 06th 07th, editors David Novick and Patrick Paroubek, 39th Annual Meeting of the Association for Computational Linguistics, Toulouse, France, 2001.

#### ● Colloques & séminaires sans actes (2)

- P. Paroubek P., J. Mariani, N. Lucas, GRACE. « Présentation des résultats des essais ». Atelier de clôture de la phase d’essai de l’évaluation GRACE-I. Avignon, 11 avril 1997
- P. PAROUBEK, « Expérience de codage de textes pour Silfide », Ecole thématique ”Codage et manipulation de ressources linguistiques”. Nancy, 8-12 septembre 1997

#### ● Rapports de contrats scientifiques (20)

- Patrick Paroubek, Nadine Glas, ” Proposition de protocole d’évaluation technologique pour la chaîne complète”, DOXA livrable D-7-1-7, décembre 2011.
- Patrick Paroubek, Pierre Albert, Alexander Pak, ”Résultats finaux d’évaluation technologique pour les 3 briques thématiques, opinions/sentiments et relations thèmes-opinions/sentiments”, DOXA, livrable D-7-1-6, décembre 2011
- Alexander Pak, Pierre Albert, Patrick Paroubek, Delphine Bernhard, ”Kits d’évaluation pour les 3 briques extraction thématique, extraction d’opinions/sentiments et extraction de relations entre thèmes et opinions/sentiments”, DOXA, livrable D-7-1-5, décembre 2010
- Pierre Albert, Patrick Paroubek, Olivier Hamon, “Serveurs d’évaluation pour les 3 briques extraction thématique, extraction d’opinions/sentiments et extraction de relations entre thèmes et opinions/sentiments”, DOXA, livrable D-7-1-4, décembre 2010
- Patrick Paroubek, Pierre Albert, Delphine Bernhard, Alexander Pak, ” Prise de mesure d’évaluation technologique pour les 3 briques extraction thématique, extraction d’opinions/sentiments et extraction de relations entre thèmes et opinions/sentiments”, DOXA livrable D-7-1-3, décembre 2010
- Patrick Paroubek, Martine Hurault-Plantet, Alexander Pak, Pierre Albert, “Protocoles d’évaluation pour les briques technologiques extraction thématique, extraction d’opinions/sentiments et extraction de relations entre thèmes et opinions/sentiments”, DOXA, livrable D-7-1-1, décembre 2009

- Patrick Paroubek, “Le module d’évaluation”, PASSAGE, livrable D17, décembre 2009.
- Patrick Paroubek, “La seconde campagne d’évaluation”, PASSAGE, livrable D16, décembre 2009.
- Patrick Paroubek, “Le corpus ROVER”, PASSAGE, livrable D14, décembre 2009.
- Anne Vilnat, Olivier Hamon, Patrick Paroubek, “Rapport sur corpus de référence manuelle”, PASSAGE, livrable D8, décembre 2009.
- Patrick Paroubek, “La première campagne d’évaluation”, PASSAGE, livrable D5, décembre 2007.
- P. Paroubek, “Report on Analysis of evaluation schemes in HLT projects”, deliverable D4.1 du projet européen IST-12611, CLASS, Juillet 2002.
- P. Paroubek, “Report on questionnaire about evaluation in HLT projects”, deliverable D4.2 du projet européen IST-12611, CLASS, Juillet 2002.
- W. Turner, P. Paroubek, C. Jacquemin, R. Gherbi, M. de Saint-Leger, Bilan d’une expérience de veille scientifique : “Microbiologie-résistance aux antibiotiques”, CNRS/SDV-INIST/URI-CNRS/LIMSI. 30 octobre 2000.
- Niels Ole Bernsen, Nicoletta Calzolari, Jean-Pierre Chanod, Khalid Choukri, Laila Dybkjær, Robert Gaizauskas, Edts: Marc Blasband & Patrick Paroubek, “A Blueprint for a General Infrastructure for Natural Language Processing Systems Evaluation Using Semi-Automatic Quantitative Black Box Approach in a Multilingual Environment”, Deliverable D1.1 Projet ELSE, Evaluation in Language and Speech Engineering LE4-8340 , April 30th 1999, p. 62.
- Niels Ole Bernsen, Nicoletta Calzolari, Jean-Pierre Chanod, Khalid Choukri, Laila Dybkjær, Robert Gaizauskas, Edts: Marc Blasband & Patrick Paroubek, “Follow-up Evaluation Proposal”, Deliverable D1.2 Projet ELSE, Evaluation in Language and Speech Engineering LE4-8340 , April 30th 1999, p. 18.
- L. Chase, L. Lamel, P. Paroubek, K. Elnieus, “Working paper on speech recognition current practice”, DISC Deliverable D2.1. Mai 1998.
- L. Chase, L. Lamel, P. Paroubek, K. Elnieus, “Survey of existing platforms and methods for speech recognition”, DISC Deliverable D2.1. Mai 1998.
- L. Chase, L. Lamel, P. Paroubek, “Guidelines and testing protocols for the development of speech recognition components for SLDSs”, DISC Deliverable D2.2. October 1998.
- U. Heid, J. Van Kuppevelt, L. Chase, P. Paroubek, L. Lamel, “Issues in natural language understanding and generation in SLDSs” , DISC Deliverable D1.4. Mai 1998.