

## **WORKSHOP TALN 2003**

### **TRAITEMENT AUTOMATIQUE DES LANGUES MINORITAIRES ET DES PETITES LANGUES**

VVF - Batz-sur-Mer (44)  
du mercredi 11 au samedi 14 juin 2003

organisée par l'IRIN, Université de Nantes  
avec la collaboration des laboratoires ADICORE et VALORIA,  
de l'Université de Bretagne Sud et de l'IRISA, INRIA Rennes.

[www.sciences.univ-nantes.fr/irin/taln2003/](http://www.sciences.univ-nantes.fr/irin/taln2003/)

#### **ARRIERE-PLAN**

Au cours des dernières années, les langues minoritaires ont attiré une attention considérable. Des projets qui visent à la revitalisation, la standardisation et à la normalisation linguistique ont été lancés pour favoriser l'usage de ces langues et pour contribuer à leur survie. Les parleurs de langues minoritaires ont pris conscience du fait que leurs langues appartiennent à l'acquis culturel du monde, et ils sont de plus en plus inclinés à utiliser leurs langues maternelles à une échelle plus large. L'augmentation de pages internet en petites langues en est la preuve.

#### **DESCRIPTION DU PROBLÈME**

Cet atelier approchera le problème des langues minoritaires du point de vue informatique. L'atelier cible essentiellement sur les langues minoritaires avec peu de ressources informatiques (occitan, hakka, la langue corse, nahuatl, ..., mais aussi des langues de geste). Les langues minoritaires avec des ressources informatiques importantes, par exemple le catalan, ne sont pas exclues a priori, dans la mesure où elles servent d'exemple pour une langue minoritaire qui a réussi. Également sont acceptées des communications sur les langues majoritaires avec peu de ressources informatiques au cas où elles traitent des problèmes informatiques partagés par les langues minoritaires.

Le but de l'atelier est d'obtenir un panorama des activités et des accomplissements dans le domaine du traitement automatique de langues minoritaires pour favoriser la recherche dans ce domaine et de mettre plus prestige à telles entreprises.

Le traitement automatique de langues minoritaires doit surmonter un certain nombre de difficultés qui résultent de leur statut spécial.

\* Car ces langues ont peu de parleurs, il y a peu de linguistes indigènes et même peu de linguistes informaticiens. Les approches basées sur les règles peuvent donc difficilement être appliquées aux problèmes comme l'étiquetage, analyse syntaxique, etc.

\* Le soutien financier maigre pour ces langues semble également exclure des approches basées sur les règles à cause de la quantité de travail humain que ces approches exigent normalement. Ce problème pourrait être surmonté si des cadres informatiques développés pour d'autres langues peuvent être adoptés.

\* Les approches basées sur corpus sont seulement applicables si des corpus adéquats sont disponibles. Cependant, la création d'un corpus coûte temps et argent et exige des conceptions linguistiques valides, surtout si des corpus d'usage général doivent être créés.

\* Les approches basées sur des exemples semblent plus prometteuses dans cette lumière si ce n'est pas un corpus d'usage général, mais des listes d'exemples spécifiques qui sont requis. Il semble également plus facile de mettre en application une compilation d'exemples spéciaux que d'écrire des règles formelles. Pourtant, nous ignorons si une telle approche aux langues minoritaires serait praticable.

\* Des techniques informatiques simples peuvent être développées ou sont déjà appliquées. Ces techniques profitent d'une spécificité d'une langue ou d'une famille de langues. Ceci cependant peut entraver le transfert de l'approche d'une langue à l'autre. À titre d'exemple, quelques techniques pourraient fonctionner avec des langues analytiques mais pas avec des langues agglutinatives. Les différents systèmes d'écriture pourraient également empêcher qu'une approche simple soit appliquée à un autre langage. On souhaite que l'atelier stimule la recherche dans ce domaine. Nous invitons des soumissions de communication qui traitent les suivants sujets sans y être limités.

## **SUJETS D'INTÉRÊT**

\* la relation entre le TALN et le support de langues minoritaires,

\* le développement d'applications spécifiques pour le traitement automatique des langues minoritaires, p.e. pour l'étiquetage, l'analyse morphologique, l'analyse syntaxique, la recherche documentaire, la traduction automatique

\* le développement des corpus et des dictionnaires électroniques pour les langues minoritaires,

\* la présentation de techniques simples au TALN qui pourraient être appliquées aux langues minoritaires,

\* les études générales qui décrivent la situation actuelle du TALN pour les langues minoritaires d'un pays, d'une région ou d'un type de langage,

\* l'analyse comparative de différentes approches au TALN de différents langues minoritaires et types de langues divers,

\* ressources libres pour le TALN, leurs domaines d'application et leur limitations,

\* les besoins d'applications du TALN pour des groupes particuliers de langues minoritaires.

## COMITÉ DE PROGRAMME

Shin-Hsi Chen	National Taiwan University hh_chen@csie.ntu.edu.tw
Vitelio Herrera	Union Latine, Direction Terminologia et Industries de la Langue, Paris v.herrera@unilat.org
Leonid Iomdin	Academia Nauk Moscow, Laboratory of Computer Linguistics iomdin@iitp.ru
Harold Somers	Centre for Computational Linguistics, UMIST Harold.Somers@umist.ac.uk
Oliver Streiter	EURAC, European Academy, Language & Law ostreiter@eurac.edu
Mathias Stuflesser	SPELL, Servisc de Planificazion y Elaborazion dl Lingaz Ladin spell-mathias@ladinia.net
Leonhard Voltmer	EURAC, European Academy, Minorities lvoltmer@eurac.edu
Wolfgang Wölck	University at Buffalo, SUNNY wwolck@acsu.buffalo.edu

## CALENDRIER

Date limite de soumission des articles : 19.03.2003  
Notification d'acceptation aux auteurs : 31.03.2003  
Réception des papiers finaux (prêts à cliquer) : 28.04.2003

## SOUSSION DE COMMUNICATION

Les articles soumis seront rédigés en français (ou en anglais par les non-francophones). Ils ne devront pas dépasser 10 pages en Times 12, espacement simple, figures, exemples et références compris. Les soumissions devront parvenir au comité d'organisation, sous la forme d'un document attaché (fichiers rtf, ps ou pdf au format A4 et non Lettre US), par courrier électronique, adressées à [ostreiter@eurac.edu](mailto:ostreiter@eurac.edu). Le message contiendra le titre de la soumission et les noms des auteurs.

Les formats à utiliser pour la rédaction sont disponibles sur le site de la conférence : [www.sciences.univ-nantes.fr/irin/taln2003/](http://www.sciences.univ-nantes.fr/irin/taln2003/) (rubrique TALN)

## CONTACT

Pour toute question relative à l'organisation de l'atelier et la soumission de communications veuillez contacter :

Oliver Streiter  
European Academy  
Language and law  
mail: [ostreiter@eurac.edu](mailto:ostreiter@eurac.edu)  
tel: +39 0471 055 115  
fax: +39 0471 055 199