

**TALAf 2016 : Traitement automatique des langues africaines (écrit et parole)**  
**Atelier JEP-TALN-RECITAL 2016 - Paris le 4 juillet 2016**

Version HTML : <http://talaf.imag.fr/>

===== VERSION FRANÇAISE == ENGLISH VERSION BELOW =====

## **PRÉSENTATION**

Les ateliers TALAf ont lieu tous les deux ans. Le premier atelier a eu lieu lors de la conférence JEP-TALN-RÉCITAL 2012 le 8 juin 2012 à Grenoble (voir <http://talaf.imag.fr/2012/Actes>). Le second a eu lieu lors de la conférence TALN 2014 le premier juillet 2014 à Marseille (voir <http://talaf.imag.fr/2014/Actes>).

La troisième édition de TALAf aura lieu lors de la conférence JEP-TALN-RECITAL, le 4 juillet 2016 dans les locaux de l'INALCO, à Paris.

Le traitement automatique des langues est en plein essor en Afrique. En effet, dans de nombreux pays, nous assistons à une reconnaissance officielle des langues nationales :

- Au Niger, les arrêtés définissant les alphabets haoussa, kanouri, tamajaq et zarma ont été publiés en 1999. Depuis, l'Assemblée Nationale a mis en place des traductions simultanées des débats en trois langues : français, haoussa et zarma;
- Au Maroc, l'Institut Royal de la Culture Amazighe (IRCAM) qui œuvre pour la promotion de la culture amazighe et du développement de la langue berbère a été fondé par décret royal en 2001 ;
- Au Sénégal, la reconnaissance des langues nationales est mentionnée dès l'article premier de sa Constitution du 7 janvier 2001 : « *La langue officielle de la République du Sénégal est le Français. Les langues nationales sont le Diola, le Malinké, le Pular, le Sérère, le Soninké, le Wolof et toute autre langue nationale qui sera codifiée* ». Le ministère de l'Enseignement technique, de la Formation professionnelle, de l'Alphabétisation et des Langues nationales (METFPALN) en est chargé. Depuis le 9 décembre 2014, les propos des parlementaires sénégalais sont traduits en simultané grâce à un système d'interprétation dans les six langues nationales (peul, sérère, wolof, diola, mandingue et soninké) en plus du français, permettant à la majorité des députés de s'exprimer dans leur langue maternelle.

Par ailleurs, un certain nombre de collègues / universitaires africains formés dans les pays du Nord reviennent dans leur pays avec la volonté de continuer leur travail sur les langues locales. Il y a également des diaspora disposant de moyens technologiques leur permettant de contribuer directement en ligne et de manière bénévole.

À cela s'ajoute le développement de programmes d'enseignement bilingues (langue officielle / langue nationale) dans les écoles primaires de nombreux pays. La langue officielle restant pour la plupart celle de l'ancien colonisateur (français, anglais, portugais...).

D'autre part, le téléphone mobile se répand à grande vitesse : avec 650 millions d'unités, l'Afrique a dépassé les États-Unis et l'Europe. Dans de nombreuses régions, il est plus facile d'installer une antenne de téléphonie mobile que d'installer des lignes fixes. De ce fait, les personnes qui s'équipent pour la première fois d'un téléphone le font avec un terminal mobile. Des applications se développent comme le transfert d'argent ou la diffusion de bulletins météo.

Le financement de projets de recherche portant sur ces langues peut être obtenu depuis de nombreuses années auprès de l'Organisation Internationale de la Francophonie avec ses appels à projets du fonds francophone des inforoutes (voir par exemple les projets [DiLAF](#) ou [flore](#)) ou de l'Agence Universitaire de la Francophonie. La France, également, finance des projets sur ces langues à travers l'Agence Nationale pour la Recherche (voir par exemple le projet [ALFFA](#)).

Les conditions sont donc réunies pour l'essor du traitement automatique des langues en Afrique, pour l'écrit comme pour la parole.

Dans ce contexte, les rôles de l'atelier TALAf sont les suivants :

- mettre en relation les chercheurs du domaine grâce aux rencontres lors de l'atelier mais aussi avec la liste de diffusion [talaf@imag.fr](mailto:talaf@imag.fr) ;
- mutualiser les savoirs en utilisant des outils en source ouverte, des standards (ISO, Unicode), et en publiant les ressources produites sous licence ouverte (Creative Commons), afin d'éviter, entre autres, la perte d'informations lorsqu'un projet s'arrête et ne peut être repris immédiatement faute de moyens ;
- développer un ensemble de bonnes pratiques fondées sur l'expérience des chercheurs du domaine. Il s'agit de mettre au point des méthodologies simples et économes en coût d'achat de logiciels pour l'élaboration de ressources, d'échanger sur les techniques permettant de se passer de certaines ressources inexistantes et enfin d'éviter des pertes de temps et d'énergie.

Les ateliers TALAf sont soutenus par l'association Lexicologie Terminologie Traduction : <http://www.ltt.auf.org/index.php>

Dans ce cadre, nous invitons tous les chercheurs en traitement automatique des langues travaillant sur des langues africaines, y compris les langues créoles d'Afrique, que ce soit à l'écrit ou à l'oral, à soumettre une communication pour cet atelier.

## **TYPES DE COMMUNICATION**

Les publications devront comprendre entre 6 et 12 pages. Les auteurs sont invités à soumettre des articles présentant des travaux de recherche originaux notamment sur les thèmes proposés ci-dessous.

Les auteurs francophones sont invités à rédiger en français avec un résumé dans la langue de leur choix. Les auteurs non francophones peuvent rédiger en anglais avec un résumé en français et un autre dans la langue de leur choix.

## **THÈMES**

L'atelier est ouvert à la présentation de travaux de recherche portant sur les thèmes suivants :

Ressources :

- constitution de corpus écrits (monolingues, bilingues alignés ou comparables)
- constitution de corpus oraux (incluant la transcription)
- élaboration de lexiques et dictionnaires (monolingues, bilingues)
- enrichissement de ressources
- évaluation de la qualité des ressources

Outils :

- analyseurs morphologiques, correcteurs orthographiques

- analyseurs syntaxiques, correcteurs grammaticaux
- systèmes de traduction automatique (statistique ou symbolique)
- reconnaissance de la parole
- synthèse vocale
- translittération

## **CRITÈRES DE SÉLECTION**

Les soumissions seront examinées par au moins deux spécialistes du domaine.

Pour les travaux de recherches, seront considérées en particulier :

- l'adéquation aux thèmes de l'atelier,
- l'importance et l'originalité de la contribution,
- la correction du contenu scientifique et technique,
- l'organisation et la clarté de la présentation.

## **MODALITÉS DE SOUMISSION**

Les formats de soumission sont disponibles pour OpenOffice, Word et Latex sur le site de TALN :

<https://jep-taln2016.limsi.fr/styles/jeptaln2016-v2.zip>

Les propositions de communications doivent être envoyées sous forme pdf à l'adresse suivante :

<http://talaf.imag.fr/2016/soumission.php>

## **COMITÉ DE PROGRAMME**

Martine Adda-Decker (CNRS-LPP & LIMSI, Paris, France)

Laurent Besacier (LIG, Grenoble, France)

Sokhna Bao Diop (Université Gaston Berger, St Louis du Sénégal, Sénégal)

Philippe Bretier (Voxygen, Pleumeur-Bodou, France)

Khalid Choukri (ELDA, Paris, France)

Mame Thierno Cissé (ARCIV, Université Cheikh Anta Diop, Dakar, Sénégal)

Chantal Enguehard (LINA, Nantes, France)

Núria Gala (LIF, Marseille, France)

Modi Issouf (Ministère de l'Éducation, Niamey, Niger)

Fary Silate Ka (IFAN, Université Cheikh Anta Diop, Dakar, Sénégal)

Mathieu Mangeot (LIG, Grenoble, France)

Chérif Mbodj, (Centre de Linguistique Appliquée de Dakar, Sénégal)

Kamal Naït-Zerrad (INALCO, Paris, France)

El Hadj Mamadou Nguer (Université Gaston Berger, St Louis du Sénégal, Sénégal)

Donald Osborn (Bisharat, ltd.)

Francois Pellegrino, (DDL, Lyon, France)

Olivier Rosec (Voxygen, Pleumeur-Bodou, France)

Fatiha Sadat (UQAM, Montréal, Canada)

Aliou Ngoné Seck (FLSH, Université Cheikh Anta Diop, Dakar, Sénégal)

Emmanuel Schang (Université d'Orléans, Orléans, France)

Gilles Sérasset (LIG, Grenoble, France)

Max Silberztein (ELLIADD, Université de Franche-Comté, Besançon, France)

Sylvie Voisin (DDL, Lyon, France)

Valentin Vydrin (LLACAN-INALCO, Paris, France)

## CALENDRIER

- Date limite de soumission : 8 mai 2016
- Notification aux auteurs : 23 mai 2016
- Date limite de soumission des versions définitives : 3 juin 2016
- Atelier : 4 juillet 2016

## CONTACT

Pour toute question, contacter [Mathieu.Mangeot@imag.fr](mailto:Mathieu.Mangeot@imag.fr)

===== ENGLISH VERSION =====

## PRESENTATION

TALAf workshops take place every two years. The first workshop was held during the JEP-TALN-RÉCITAL 2012 conference on June 8, 2012 in Grenoble (see proceedings: <http://aclweb.org/anthology/W/W12/#1300>). The second one took place during the TALN 2014 conference on July 1, 2014 in Marseille (see proceedings: <http://www.taln2014.org/site/actes-en-ligne/actes-en-ligne-ateliers/>).

The third edition of TALAf will be held during the JEP-TALN-RECITAL conference, on July 4, 2016 at INALCO in Paris.

Natural language processing is booming in Africa. Indeed, in many countries, there is an ongoing official recognition of national languages, for instance:

- In Niger, laws defining alphabets for Hausa, Kanuri, and Tamajaq Zarma were published in 1999. Since then, the National Assembly has set up simultaneous translation of the debates in three languages: French, Hausa and Zarma;
- In Morocco, the Royal Institute of Amazigh Culture (IRCAM), which works for the promotion of Amazigh culture and development of the Berber language was founded by royal decree in 2001;
- In Senegal, the recognition of national languages of the recognition was mentioned in the first article of the Constitution of 7 January 2001: "The official language of the Republic of Senegal is French. The national languages are Diola, Malinke, Pular, Serer, Soninke, Wolof and other national language to be codified." The Department of Technical Education, Vocational Training, Literacy and National Languages (METFPALN) is responsible for this. Since December 9, 2014, the Senegalese parliamentarians debates are translated simultaneously through an interpretation system in six national languages (Fulani, Serere, Wolof, Jola, Mandinka and Soninke) in addition to French, allowing the majority of members to speak in their mother tongue.

Moreover, a number of colleagues / African scholars trained in the North return to their country with the will to continue their work in local languages. There are also some Diasporas that have technological material allowing them to contribute directly online and on a voluntary basis.

Added to this, the development of bilingual education programs (official / national language) in primary schools in many countries is growing. The official language remaining mostly that of the

former colonial country (French, English, Portuguese ...).

On the other hand, mobile phones are spreading fast: with 650 million units, Africa has surpassed the United States and Europe. In many areas, it is easier to install a mobile antenna than fixed lines. Therefore, the people who use a telephone for the first time do it with a mobile terminal. Applications are developed such as money transfer or dissemination of weather reports.

The funding of research projects on these languages can now be obtained from the "Organisation Internationale de la Francophonie" with their calls for projects of the "fonds francophone des inforoutes" (see eg [DiLAF](#) or [flore](#) projects) or the "Agence Universitaire de la Francophonie". France also supports projects on these languages through the National Agency for Research (see eg [ALFFA](#) project).

So the conditions are gathered for the development of natural language processing in Africa, both written and spoken.

In this context, the roles of TALAf workshop are:

- bring together researchers in the field through meetings at the workshop but also with the [talaf@imag.fr](mailto:talaf@imag.fr) mailing list;
- pooling knowledge using open source tools, standards (ISO, Unicode), and publishing the resources produced with an open license (Creative Commons) to avoid including the loss of information when a project stops and can not be resumed immediately for lack of resources;
- develop a set of best practices based on the experience of researchers ; set up simple efficient methodologies based on free or very cheap software for the development of resources, exchange about techniques that can avoid the use of non-existent resources and finally avoid loss of time and energy.

TALAf workshops are supported by the non-profit organisation "Lexicologie Terminologie Traduction": <http://www.ltt.auf.org/index.php>

We invite all researchers in natural language processing working on African languages, including Creole languages of Africa, whether written or oral, to submit a paper to this workshop.

## **TYPES OF COMMUNICATION**

Publications should contain between 6 and 12 pages. Authors are invited to submit papers presenting original research particularly on the themes listed below.

French speaking authors are invited to write in French with a summary in the language of their choice. Non-French speaking writers can write in English with a summary in French and another in the language of their choice.

## **TOPICS**

The workshop is open to research works on the following topics:

Resources:

- written corpora (monolingual, bilingual aligned or comparable)
- speech corpora (including transcription)
- lexicons, dictionaries and databases (monolingual, bilingual, multilingual)

- resources enrichment
- resources quality evaluation

Tools:

- morphological analyzers, spell-checkers
- syntactic analyzers, grammar checkers
- machine translation systems (empirical or rule-based)
- speech recognition
- text-to-speech synthesis
- transliteration

## **SELECTION CRITERIA**

Submissions will be reviewed by at least two specialists of the domain.

The following points will be taken into account:

- adequacy to the workshop topics
- importance and originality of the contribution,
- scientific and technical content precision,
- organization and clarity of the presentation.

## **SUBMISSIONS**

The submission formats will be available for OpenOffice, Word and Latex and accessible from:

<https://jep-taln2016.limsi.fr/styles/jeptaln2016-v2.zip>

The communication proposals must be sent in PDF format to the following address:

<http://talaf.imag.fr/2016/soumission.php>

## **PROGRAMME COMMITTEE**

Martine Adda-Decker (CNRS-LPP & LIMSI, Paris, France)

Laurent Besacier (LIG, Grenoble, France)

Sokhna Bao Diop (Université Gaston Berger, St Louis du Sénégal, Sénégal)

Philippe Bretier (Voxygen, Pleumeur-Bodou, France)

Khalid Choukri (ELDA, Paris, France)

Mame Thierno Cissé (ARCIV, Université Cheikh Anta Diop, Dakar, Sénégal)

Chantal Enguehard (LINA, Nantes, France)

Núria Gala (LIF, Marseille, France)

Modi Issouf (Ministère de l'Éducation, Niamey, Niger)

Fary Silate Ka (IFAN, Université Cheikh Anta Diop, Dakar, Sénégal)

Mathieu Mangeot (LIG, Grenoble, France)

Chérif Mbodj, (Centre de Linguistique Appliquée de Dakar, Sénégal)

Kamal Naït-Zerrad (INALCO, Paris, France)

El Hadj Mamadou Nguer (Université Gaston Berger, St Louis du Sénégal, Sénégal)

Donald Osborn (Bisharat, ltd.)

Francois Pellegrino, (DDL, Lyon, France)

Olivier Rosec (Voxygen, Pleumeur-Bodou, France)

Fatiha Sadat (UQAM, Montréal, Canada)

Aliou Ngoné Seck (FLSH, Université Cheikh Anta Diop, Dakar, Sénégal)

Emmanuel Schang (Université d'Orléans, Orléans, France)

Gilles Sérasset (LIG, Grenoble, France)

Max Silberztein (ELLIADD, Université de Franche-Comté, Besançon, France)

Sylvie Voisin (DDL, Lyon, France)

Valentin Vydrin (LLACAN-INALCO, Paris, France)

### **IMPORTANT DATES**

- Submission deadline: 8 May 2016
- Notification of acceptance: 23 May 2016
- Final Submission Deadline: 3 June 2016
- Workshop: 4 July 2016

### **CONTACT**

For any enquiries, please contact [Mathieu.Mangeot@imag.fr](mailto:Mathieu.Mangeot@imag.fr)