Traitement Automatique des Langues Africaines 2016
Préface

Mathieu Mangeot[1]
(1) LIG, campus, 38041 Grenoble cedex 9, France
mathieu.mangeot@imag.fr

---

1 Motivations et objectifs

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 les actes : http://talaf.imag.fr/2012/). Le second a eu lieu lors de la conférence TALN 2014 le premier juillet 2014 à Marseille (voir les actes : 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, 65 Rue des Grands Moulins, 75013 Paris. Pour vous inscrire, rendez-vous sur le site de la conférence JEP-TALN-RECITAL 2016.

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 ;

- 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

2 Présentation des articles

L'atelier a reçu 12 soumission. 8 articles ont été rédigés en français et 4 en anglais. Pour mémoire, l'édition 2012 de TALAf avait reçu 12 soumissions et l'édition 2014 en avait reçu 13. Parmi ces articles, 10 ont été acceptés et 2 rejetés. 8 articles portent sur le traitement de l'écrit et 2 articles sur le traitement de l'oral. Tous les articles ont été relus par au moins deux relecteurs.

La diversité linguistique est présente puisque huit langues figurent dans les articles acceptés : amazighe, bambara, comorien, igbo, maninka, peul, swahili, wolof. Nous remarquons l'arrivée de langues d'Afrique de l'Est comme le swahili et le comorien. La langue la plus traitée reste le wolof avec 4 articles.

3 Programme de l'atelier

09h30-10h00	**Valentin Vydrin, Andrij Rovenchak & Kirill Maslinsky** *Maninka Reference Corpus: A Presentation*.
10h00-10h30	**Ikechukwu Onyenwe, Mark Hepple & Uchechukwu Chinedu** *Improving Accuracy of Igbo Corpus Annotation Using Morphological Reconstruction and Transformation-Based Learning*.
10h30-11h00	Pause café
11h00-11h30	**Moneim Abdourahamane, Christian Boitet, Valérie Bellynck, Lingxiao Wang & Hervé Blanchon** *Construction d'un corpus parallèle français-comorien en utilisant de la TA français-swahili*.
11h30-12h00	**David Blachon, Elodie Gauthier, Laurent Besacier, Guy-Noël Kouarata, Martine Adda-Decker & Annie Rialland** *Collecte de parole pour l'étude des langues peu dotées ou en danger avec l'application mobile Lig-Aikuma*.
12h00-14h00	Pause repas
14h00-14h30	**Michael Melese Woldeyohannis, Laurent Besacier & Meshesha Million** *Amharic Speech Recognition for Speech Translation*.
14h30-15h00	**El Hadji Malick Fall, El Hadji Mamadou Nguer, Sokhna Bao Diop, Mouhamadou Khoulé, Mathieu Mangeot & Mame Thierno Cissé** *Digraphie des langues ouest africaines : Latin2Ajami : un algorithme de translittération automatique*.
15h00-15h30	**Fatimazahra Nejme, Siham Boulaknadel & Driss Aboutajdine** *Développement de ressources pour la langue amazighe : Le Lexique Morphologique El-AmaLex*.
15h30-16h00	**Alla Lo, Elhadji Mamadou Nguer, Abdoulaye Youssoupha Ndiaye, Cheikh Bamba Dione, Mathieu Mangeot, Mouhamadou Khoule, Sokhna Bao Diop & Mame Thierno Cisse** *Correction orthographique pour la langue wolof : état de l'art et perspectives*.
16h00-16h30	Pause café
16h30-17h00	**Mouhamadou Khoule, Mathieu Mangeot, El Hadji Mamadou Nguer & Mame Thierno Cisse** *iBaatukaay : un projet de base lexicale multilingue contributive sur le web à structure pivot pour les langues africaines notamment sénégalaises*.
17h00-17h30	**Chérif Mbodj & Chantal Enguehard** *Production et mise en ligne d'un dictionnaire électronique du wolof*.

4 Comité scientifique

- 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)

5 Conclusion

Cette troisième édition confirme l'intérêt d'un atelier francophone sur le traitement automatique des langues africaines. Le TAL en Afrique a pris son envol. Cet atelier et de la liste de discussion par courriel talaf_AROBASE_imag.fr permettent de construire et de structurer la communauté qui se met en place progressivement. Les savoirs et savoirs-faire doivent également être capitalisés pour resservir pour d'autres langues et d'autres contextes.

Le prochain atelier TALAf est prévu pour 2018. Il est prévu de l'organiser conjointement avec la conférence TALN. À moyen terme, nous envisageons d'organiser cet atelier dans un pays africain moteur du domaine.