

Anna Marie Diagne Institut für Linguistik-Phonetik Universität de Cologne a.diagne@uni-koeln.de	Mame Thierno Cissé Laboratoire de Phonétique Département de Linguistique Générale Faculté des Lettres et Sciences Humaines Université Cheikh Anta Diop de Dakar thiernoc@gmail.com
--	---

## COMPTE RENDU DE LA FORMATION ANIMÉE PAR LE RÉSEAU

### *LEXICOLOGIE TERMINOLOGIE ET TRADUCTION*

LORS DES JOURNÉES D'ANIMATION RÉGIONALE DE BRAZZAVILLE 3 et 5 mars 2007

#### **Formatage de corpus oraux pour l'exploitation linguistique et le partage des données**

### **ENJEUX ET OBJECTIFS DE LA FORMATION**

Le réseau *Lexicologie, terminologie, traduction* a accepté de jouer un rôle pilote pour répondre aux problèmes spécifiques des autres réseaux concernant la création de bases de données. Qu'il s'agisse de données bibliographiques, textuelles, orales, lexicographiques, terminographiques ou autres, le rôle du réseau est de proposer des démarches simples qui garantissent l'universalité et la pérennité des données, indépendamment des systèmes informatiques.

La formation à la recherche proposée pour Brazzaville s'inscrit ainsi dans la continuité des actions déjà menées par le réseau LTT pour la prise en charge de cette question avec notamment le soutien apporté à l'action de recherche en réseau « Expérimentation de normes de balisage en langues partenaires » développée en 2003 et 2005 entre la Mauritanie et le Sénégal, et la formation « Conception et utilisation d'un clavier virtuel Unicode » proposée lors des Premières Journées scientifiques communes de Ouagadougou en 2004.

Le développement de l'Internet et des logiciels libres rend de nos jours accessibles des solutions de numérisation et de transcription de corpus linguistiques dont la maîtrise est relativement simple. A cette disponibilité des outils s'est ajoutée la mise au point par la communauté scientifique de standards de formatage de données et de polices de caractères qui rendent le partage et l'exploitation des données possibles pour n'importe quel chercheur disposant d'un ordinateur à travers le monde.

L'application de ces principes en contexte africain contribuera à l'élaboration de vastes corpus et au désenclavement des chercheurs. Elle permettra également le partage des données entre chercheurs dans la même discipline ou dans des disciplines différentes au sein d'une même institution et entre institutions partenaires. L'enjeu est considérable pour la sauvegarde du patrimoine linguistique dans les pays du Sud par la constitution de grands corpus. Les outils proposés lors des journées scientifiques de Brazzaville sont des logiciels couramment utilisés et gratuits tels que *SimpleOCR* pour la numérisation de corpus écrits, *Audacity* pour la numérisation de corpus oraux et *Elan* pour la transcription et l'exploitation des données. Pour finir des stratégies de partage et de conservation des données ont été exposées.

### **CADRE GÉNÉRAL DE LA FORMATION**

#### **Salle informatique**

Le Bureau de l'AUF de Brazzaville a accueilli la formation et a mis à notre disposition une salle avec un nombre suffisant de postes informatiques et un vidéo projecteur. MM. Kabirou et Goma et tous les membres du bureau de l'AUF de Brazzaville ont aidé à un déroulement satisfaisant de la formation. Nous avons pu exploiter les deux jours avant le début de la formation pour participer à la configuration sous Windows des postes de travail et à terminer l'installation des logiciels à utiliser. Nous profitons de ce compte rendu pour leur adresser nos sincères remerciements.

## Participants

Nous avons effectué deux séances de formation avec une vingtaine de stagiaires à chaque fois. Le samedi 3 mars avec le groupe des « littéraires » et le lundi 5 mars avec le groupe des « linguistes ». Des collègues des autres réseaux et de l'Université Marien Ngouabi ont aussi suivi la formation. Tous ont été particulièrement intéressés par le protocole développé par LTT.

Lors des séances de manipulation, environ deux heures en fin de journée, nous avons remarqué combien étaient grandes l'envie d'apprendre des stagiaires et leur capacité de compréhension. Quelques uns avaient des problèmes de maîtrise de l'outil informatique ce qui a nécessité un suivi plus personnel.

En fin de stage ils ont été tous conscients de l'importance de se familiariser davantage à ce types d'outils et ils se sont rendu compte de leur facilité d'utilisation ce qui était l'objectif de la formation.

## PROGRAMME PROPOSÉ

### Matinée

- ☒ Format des enregistrements et types d'enregistreurs.
- ☒ Numérisation de matériel sonore.
- ☒ Choix d'un logiciel de transcription : le cas de *Elan*.
- ☒ Configuration de la base de données.

### Après-midi

- ☒ Conservation des données sur divers supports.
- ☒ Exportation à partir d'*Elan* vers d'autres logiciels.
- ☒ Exemple d'utilisation des techniques présentées. En lieu et place d'une présentation du projet « Conception d'un dictionnaire électronique unilingue wolof et bilingue wolof-français » en cours à Dakar comme initialement prévu, nous avons préféré laisser les stagiaires choisir une préoccupation propre pour faire une application.

## BILAN

Au bout des deux journées et après la réunion convoquée par M. Marc Cheymol pour permettre aux stagiaires d'apprécier les formations reçues, nous pouvons retenir que :

- 1- ce type de formation est nécessaire et les stagiaires sont unanimes à apprécier la formation LTT ; ils ont demandé à ce que la durée soit plus importante encore pour les manipulations notamment ; il ont été satisfaits par le contact qu'ils ont eu avec d'autres enseignants,
- 2- il est tout aussi nécessaire de faire le recensement des besoins et du profil des stagiaires dans la phase de conception des formations et d'essayer, autant que possible, de prendre en charge ces besoins dans le programme de formation,
- 3- le suivi des stagiaires est à envisager sous une forme ou une autre.

Il nous semble que les objectifs visés par ces formations ont été atteints d'autant plus que les collègues des autres réseaux qui ont suivi les formations sont aussi intéressés par le protocole développé par LTT notamment M. Foucault, Coordonnateur du Réseau « littératures d'enfance », qui a suivi entièrement les formations et qui envisage de proposer une collaboration entre son réseau et LTT du fait que dans leurs travaux ils ont beaucoup de traductions en français de documents en langues partenaires.

Les collègues de l'Université Marien Ngouabi ont montré leur intérêt pour une formation d'une durée plus importante. Anna Marie Diagne a eu à configurer quelques bases de données en marge du colloque et nous

avons configuré les machines de tous les stagiaires et collègues qui l'ont demandé.

Des collègues du Mali, de la Côte d'Ivoire, du Burkina, et du Cameroun ont manifesté leur intérêt pour la mise en place, dans leurs institutions respectives, de petites équipes de recherche avec des doctorants en lexicologie et terminologie ; ils ont posé la question de savoir si le réseau LTT pourrait, le cas échéant, les aider en leur organisant ce type de formation.