

ILES D IMESLI

Revue annuelle, **Volume 05**, décembre 2013

ISSN : 2170-113X

ILES D IMESLI Volume 5



Dossier thématique

Linguistique du corpus



**Revue de LAELA-UMMTO
S/D du Professeure Noura Tigziri**

Directeur de publication :

Professeur Rabah Kalouche

Comité de rédaction :

Professeur Remi Jolivet

Professeur Noura Tigziri

Comité scientifique de la revue :

Ahmed Boukous (IRCAM, Maroc)

Vermondo Brugnatelli (Université de Milan, Italie)

Remi Jolivet (Université de Lausanne, Suisse)

Rabah Kahlouche (Université Mouloud Mammeri, Algérie)

Eric Keller (Université de Lausanne, Suisse)

Marteen Kossman (Université de Leyde, Leiden)

Mena Lakfioui (Université de Milan, Italie)

Mortéza Mahmoudian (Université de Lausanne, Suisse)

Philippe Martin (Université Paris7, France)

Noura Tigziri (Université Mouloud Mammeri, Algérie)

Brigitte Zelner (Université de Lausanne, Suisse)

Adresse

Laboratoire d'Aménagement et d'Enseignement de la Langue Amazighe,
Département de Langue et Culture Amazighes, Faculté des Lettres et des Langues,
Université Mouloud Mammeri de Tizi-Ouzou, 15000, Tizi-Ouzou.

Site : www.umtmo.dz, Tel / fax : 026 11 33 47. E-mail : revueiledimesli@gmail.com

ISSN : 2170-113X

Dépôt légal : 1802-2009

Imprimerie : El Amal

Sommaire

Corpus : limites, représentativité et choix Mortéza MAHMOUDIAN	9-16
L'acquisition phonologique de l'enfant Kabyle Ramdane BOUKHERROUF	17-22
Vers un modèle pour le recueil d'un corpus d'apprentissage d'une langue étrangère peu dotée de ressources Violetta CAVALLI-SFORZA, Mariam EL MEZOUAR	23-42
La théorie variationniste et son application dans l'étude du contact de langues en Kabylie : Enquête et analyse des données Kamal AKLI	43-52
Collecte et exploitation des données toponymiques berbères Abdelaziz ALLATI	55-70
Situation linguistique du massif de l'Aurès Recueil et analyse des données Abdenacer GUEDJIBA	71-88
Recueil des unités lexicales dans la région des Ait Aidel (Vallée de la Soummam) : Transcription, exploitation et finalité. Arezki AOUDIA	89-96
Corpus sur le Prélangage de l'enfant Kabyle Etude sociolinguistique Rabah TABTI	97-134
De la difficulté à recueillir un corpus de gestes dans les interactions Amar NABTI	135-151
Base de données kabyles : collectes de données et applications Noura TIGZIRI	153-167
corpus (linguistique) et cartographie : l'interprétation de certains phénomènes de variation régionale (à «Tizi-Ouzou») Seïdh CHALAH	169-194
Les humanités numériques et le projet de Bibliothèque Numérique Franco-Berbère Alain VAUCELLE, Tassadit YACINE	197-213
Proposition d'un schéma d'annotation des passages entre guillemets Mohand MAHRAZI	215-229
Annotation de corpus linguistique et TEI Remi JOLIVET	231-243

La MEI (Music Encoding Initiative) : un standard au service de la musique kabyle	245-277
Sylvaine Martin DE GUISE	
L'intérêt du corpus et une idée de sa constitution en lexicographie amazighe	281-293
Abdelaziz BERKAÏ	
Elaboration du vocabulaire fondamental du Kabyle, une nécessité pour l'enseignement.	295-303
Ouerdya KIRECHE	
Pour une indexation raisonnée des corpus littéraires kabyles	307-322
Patrice POGNAN, Mohand Akli SALHI	
Initiation d'une base de données toponymique	323-340
Mustapha TIDJET, Djamel NAHALI	
Le logiciel Nooj appliqué au kabyle	341-349
Hamid ANNOUZ, Ferroudja KACI, Kamal NAÏT-ZERRAD	

Presentation

A l'ère de la grande technologie, d'internet, la linguistique de corpus est de plus en plus confrontée à des champs connexes auxquels elle est de plus en plus appelée à participer.

Le traitement et l'informatisation des corpus oraux ou écrits supposent un certain nombre d'outils théoriques et de techniques qu'on devait maîtriser. Le premier point est la définition des métadonnées. La question des métadonnées commence à se poser sérieusement lorsque se multiplient les ressources linguistiques informatisées et potentiellement accessibles en ligne. Il s'agit de se mettre d'accord sur des descripteurs qui permettront ensuite une recherche efficace dans un catalogue qui renverra aux ressources elle mêmes. Le traitement de métadonnées fait appel à un certain nombre de projets sur les métadonnées comme par exemple le Dublin Core. Mais elle entre aussi dans l'élaboration des digital humanities, dans le fait que les Humanités ou SHS sont de plus en plus technologisées.

Les logiciels d'annotation des corpus écrits sont déjà anciens et ils ne permettent pas toujours d'aboutir à un balisage robuste.

L'annotation des corpus oraux est plus récente et pose encore de très nombreux problèmes auxquels il faut réfléchir pour y apporter des solutions.

Les techniques de conservation des corpus, leur accessibilité à travers le net, les problèmes juridiques liés à la publication de ces corpus qu'ils soient écrits ou oraux sont des questions d'actualité qui appellent une réflexion et concertation avec tous les spécialistes du domaine.

Les axes:

1. Le recueil

- enquêteur(s)
- informateur(s)
- type(s) de données à recueillir
- situation(s) de recueil
- outil(s) d'élicitation (guide d'entretien; questionnaires etc...)
- outil(s) et techniques d'enregistrement
- entretien(s) de debriefing et validation post-recueil

2. L'annotation

- transcription(s)
- annotation(s): niveaux; formats (TEL...)
- le stockage
- durabilité des supports
- gestion de la mise à jour des formats

3. L'exploitation

- mécanismes d'interrogation et d'extraction de données - accès à des outils de traitement (statistiques; cartographiques...)

4. La diffusion

- questions juridiques
- périmètre de la diffusion: documents bruts, bases de données, outils d'exploitation...
- supports (papier)