



HAL
open science

Corpus oral de bedja (couchitique)

Martine Vanhove

► **To cite this version:**

Martine Vanhove. Corpus oral de bedja (couchitique). Journée d'étude CORLI: Traitements et standardisation des corpus multimodaux et web 2.0., May 2018, Paris, France. hal-01873875

HAL Id: hal-01873875

<https://hal.science/hal-01873875v1>

Submitted on 13 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Corpus oral de bedja (couchitique)

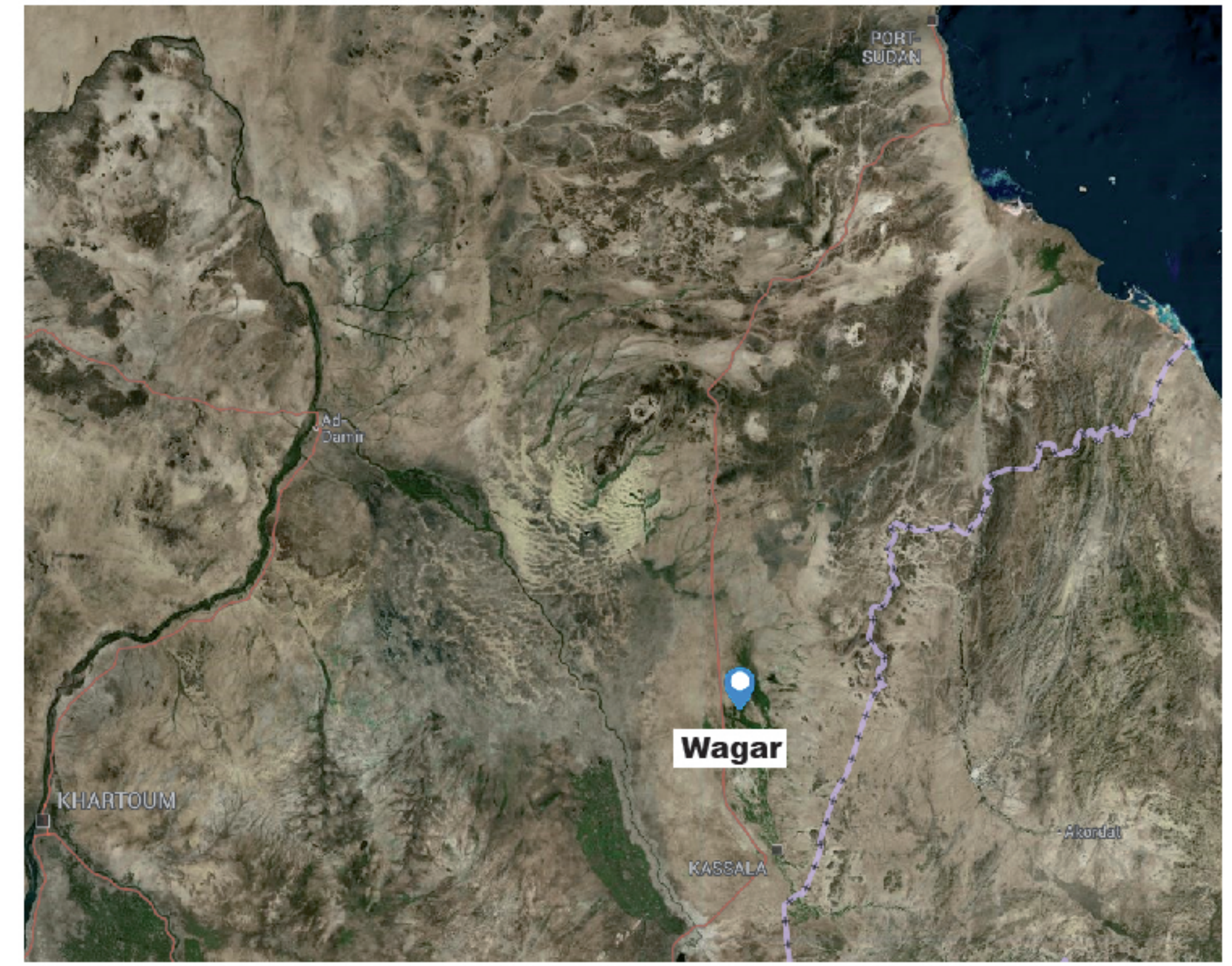
Présentation langue et corpus

Bedja :

Famille couchitique, phylum Afroasiatique, langue non écrite, parlée au Soudan. Trois variétés dialectales (nord, centre, sud)



Textes recueillis sur le terrain entre 2001 et 2006 à Port-Soudan, Erkowit, Sinkat, Wagar

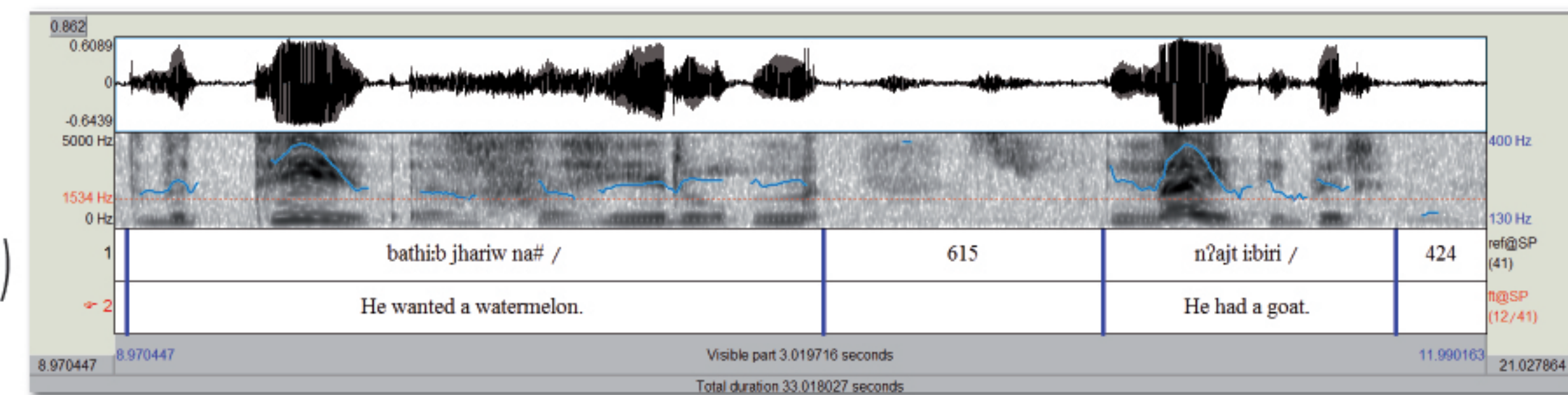


Genres : contes, poèmes, devinettes, anecdotes humoristiques, récits historiques, descriptions d'activités et jeux traditionnels, conversations.

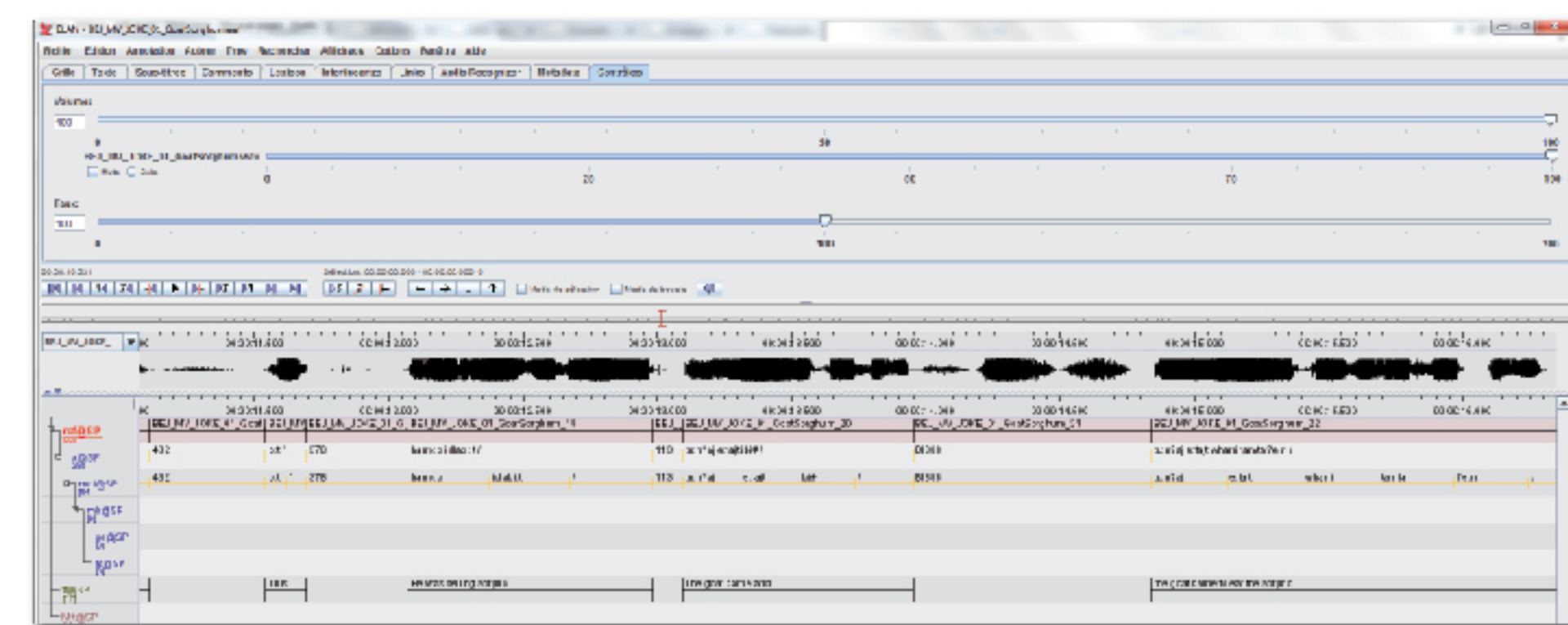
Locuteurs : tous âges de 7 à 77 ans, hommes et femmes

Méthode

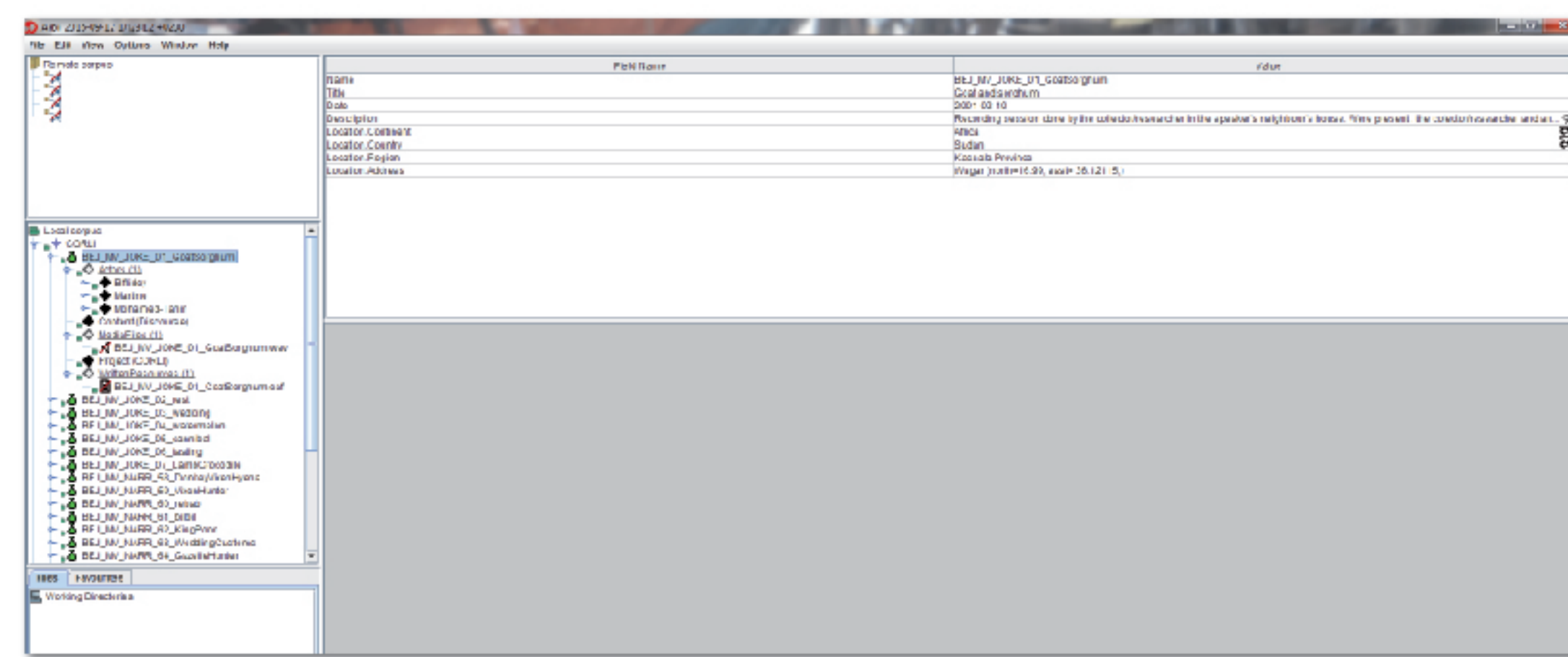
- Découpage en unités prosodiques** sous PRAAT (2 vacataires - Lisa Grillo et Anne Armand) ; vérification par chercheuse
- Transfert des transcriptions** depuis WORD dans PRAAT (vacataires)
- Transfert des traductions** depuis WORD (chercheuse)



- Transfert de PRAAT vers ELAN CorpA** (chercheuse)



- Metadonnées** dans ARBIL (chercheuse)



Résultats

- 7h30 de corpus traité** par vacataires (sur 12 prévues car financement moindre que demandé + temps de traitement sousestimé)
- Transfert sur la plateforme COCOON** en cours (effectué par ingénieur LLACAN, Christian Chanard, et Michel Jacobson (BNF))
- 41mn de corpus** recueilli en 2001 **déjà en ligne** sur COCOON

<http://cocoon.huma-num.fr/exist/crdo/search2.xql?language=http://lexvo.org/id/iso639-3/bej>

