



République du Sénégal
Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation
UNIVERSITÉ CHEIKH ANTA DIOP de DAKAR



Centre de Linguistique Appliquée de Dakar

FICHE TECHNIQUE

De la linguistique de terrain à l'élaboration de corpus contrastifs



Conçue par :

Dr Noël Bernard BIAGUI

Chargé de recherche en linguistique descriptive et didactique des langues

Dr Augustin NDIONE

Chargé de recherche en linguistique descriptive et didactique des langues

JANVIER 2021

Résumé. Les méthodologies de constitution de corpus linguistiques ont été amplement étudiées, mais sont moins abondantes quand il s'agit de corpus oraux ; ces méthodologies sont encore plus rares en ce qui concerne l'inter langue orale. Partant du constat que l'intérêt des linguistes pour la langue orale a systématiquement été en retard par rapport à celui porté à la langue écrite, nous nous intéressons dans cette fiche aux techniques de collectes de l'oralité et de ses concrétisations à travers l'outil informatique.

Mots-clés : Techniques de collecte, terrain, informatique, description.

Plan de la fiche technique

PLAN DE LA FICHE TECHNIQUE	1
INTRODUCTION	1
1. LES CONDITIONS DE LA RECOLTE	1
2. DES PETITS « TRUCS » POUR LE TERRAIN	2
3. CONDUITE DES ENQUETES DE TERRAIN	3
4. LA PLACE DE L'INFORMATIQUE	3
5. LA PRE-ENQUETE.....	8
6. LA PLACE DE L'ENQUETEUR	8
7. L'INFORMATEUR	9
CONCLUSION	9
BIBLIOGRAPHIE.....	9

Introduction

Les langues sur lesquelles les linguistes de terrain travaillent sont souvent des langues « rares », « en danger » ou « minoritaires ». En fait, il s'agit de langues souvent sans tradition d'écriture, parlées par peu de locuteurs, dans des endroits où ce sont d'autres langues qui sont utilisées par les administrations, les écoles, les médias, etc. Le problème de la transcription est alors ici central, contrairement aux corpus sur les langues qui possèdent écriture et orthographe. Les corpus sont pour la plupart composés d'enregistrements de parole : récits, dialogues, chants, cérémonies, questionnaires, élicitations, etc. Ces enregistrements sont par ailleurs transcrits, traduits et analysés au minimum jusqu'à un niveau morpho-phonologique. La quantité de données que l'on peut posséder sur une langue particulière est aussi une caractéristique de la linguistique de terrain. En général, les corpus sur une langue ne sont constitués que de quelques heures d'enregistrements. Souvent, seule une partie des enregistrements a été transcrite, ou bien l'analyse de certaines parties n'a pas été poussée au même niveau que celle des autres. Par exemple, il est fréquent qu'un corpus soit composé de textes inégalement décrits, certains ne comportant qu'une transcription « au kilomètre », d'autres étant traduits en différentes langues et glosés morphème par morphème. Quelquefois, on possède un lexique ou un dictionnaire de la langue contenant entre trois et quatre mille mots. Enfin, avec un peu de chance, il est possible que d'autres linguistes possèdent des ressources sur cette même langue, ou qu'il existe des descriptions anciennes de celle-ci dans de vieux ouvrages.

1. Les conditions de la récolte

Nous entendons par « enquêtes de terrain linguistiques » les procédures de constitution de données linguistiques empiriques réalisées dans des rencontres in situ (hors laboratoires) avec des informateurs. Les terrains d'enquêtes sont parfois difficiles d'accès (plusieurs journées de marche) et sous-équipés (en électricité par exemple). Les linguistes doivent alors se munir en batteries ou piles afin d'être le plus

autonome possible une fois sur place. Le choix du matériel à emporter se fait alors en fonction du poids, de l'encombrement et de la consommation d'énergie et parfois au détriment de la qualité. Actuellement, de nombreux linguistes choisissent des enregistreurs numériques qui sont souvent de bonne qualité que les matériels qu'ils utilisaient par le passé (enregistreurs analogiques), mais qui sont beaucoup plus encombrants, plus fragiles et plus lourds. Les enregistrements étant faits « sur le terrain », en milieu naturel (par opposition aux travaux faits en laboratoire), ils sont donc bruités par des éléments extérieurs de l'environnement. Cet aspect des conditions d'enregistrement rapproche ainsi les données de ce que l'on pourrait appeler la parole spontanée par opposition au côté artificiel des données que l'on obtient dans des conditions de laboratoire.

2. Des petits « trucs » pour le terrain

L'enquête de terrain est certainement un des maillons les plus importants dans la chaîne conduisant à l'analyse des faits linguistiques en Afrique, « terre de traditions orales » des « peuples sans écriture », dont la connaissance scientifique des langues, des relations entre langues et à l'intérieur des langues doit être approfondie. Plusieurs disciplines des sciences humaines et sociales dont la linguistique ont tâté le terrain et continuent à le labourer pour en tirer la substance, c'est-à-dire **décrire** et **expliquer** les faits. Naturellement les chercheurs dans ces disciplines ont, de par leurs pratiques et leurs objectifs, adopté plusieurs méthodes d'approche.

Loin d'embrasser l'ensemble des problèmes liés à l'enquête de terrain linguistique, ce cours se focalisera sur trois (3) points que sont :

- (a)- la place de l'informatique ;
- (b)- la nécessité de la pré-enquête et ;
- (c)- le choix des informateurs, qui méritent une attention particulière car ils influent sur les résultats (ou du moins sur leur analyse) de la recherche de terrain.

Ces trois (3) points cités ci-dessus seront examinés à la lumière de mon expérience de terrain à travers des enquêtes menées entre 2009 et 2010 en Casamance.

Il est cependant important de souligner, d'ores et déjà, que les écoles américaines, anglaises, allemandes, françaises, etc. ont chacune essayé de synthétiser cette expérience de terrain afin de guider les étudiants et les chercheurs nouvellement engagés sur les traces de leurs prédécesseurs.

Plusieurs documents ont été produits concernant la méthode d'enquête linguistique sur le terrain. On retiendra celle, en plusieurs volumes, de Luc Bouquiaux et Jacqueline M. C. Thomas : (1976). *Enquête et description des langues à tradition orale*. (1971 et 1976) qui fait autorité pour l'école française et qui a marqué des générations de chercheurs qui ont décrit les langues de l'Afrique noire francophone. Quel que soit leur domaine en linguistique :

- phonétique et phonologie,
- morphosyntaxe,
- lexico-sémantique,
- linguistique contrastive,
- linguistique comparée,
- sociolinguistique, etc.

Il est difficile de faire une totale abstraction de ces ouvrages classiques quand on débute sur le terrain et qu'on veut recueillir des documents oraux, les transcrire, les interpréter, les conserver ou les diffuser. A

ce jour, et à notre connaissance, il n'existe pas de documents de synthèse d'envergure, produits par les linguistes africains pour rendre compte de toute l'expérience qu'ils ont accumulée au cours de nombreuses recherches ayant abouti à la description systématique de leurs langues. La méthode pour aborder le terrain est donc un des points principaux qui impose à tout chercheur travaillant sur les langues à tradition orale un premier choix. Elle peut conditionner, dans une certaine mesure, les résultats et surtout l'énergie à dépenser pour y parvenir. Si, pour les descriptions synchroniques de la phonologie ou de la morphosyntaxe d'une langue ou d'un dialecte, il est tout à fait loisible d'opter pour une enquête intensive, en description lexico-sémantique, la méthode extensive est quasiment indispensable. En sociolinguistique par contre, on peut faire le choix entre l'une ou l'autre des deux méthodes ou encore les combiner. Si le choix de la méthode est important, l'élaboration du questionnaire aussi est déterminante car c'est de ce protocole que dépendent les résultats qu'on cherche. Il est tout aussi utile de s'appuyer sur de très bons questionnaires et de disposer d'informateurs qui acceptent le jeu des questions-réponses sans essayer de tricher.

3. Conduite des enquêtes de terrain

Plusieurs paramètres doivent être pris en compte pour la conduite des enquêtes de terrain. Il s'agit des conditions :

- sociologiques :
 - milieu rural ou urbain,
 - religion,
 - sexe de l'enquêté,
 - position sociale de l'enquêté,
 - durée de séjour en dehors de sa communauté,
 - le nombre de langue qu'il parle, etc.
- des principes d'enquête (directe ou indirecte),
- des aspects matériels,
- des informateurs occasionnels ou de référence, etc.

4. La place de l'informatique

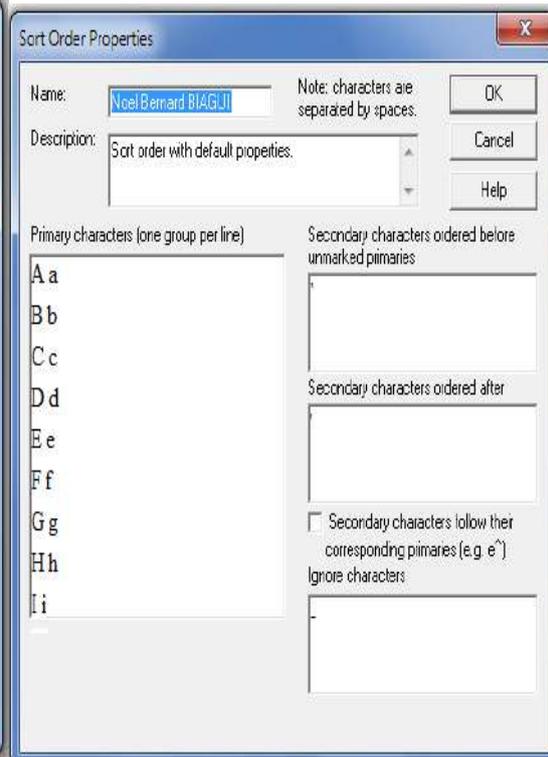
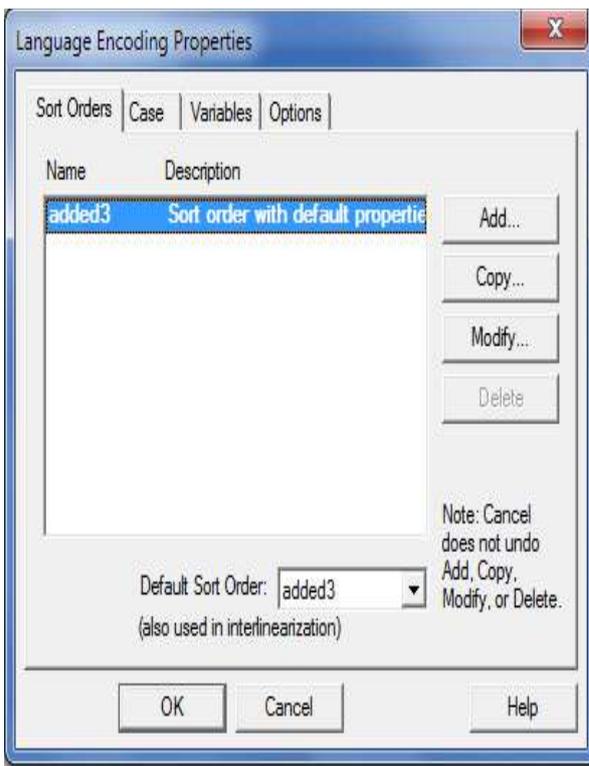
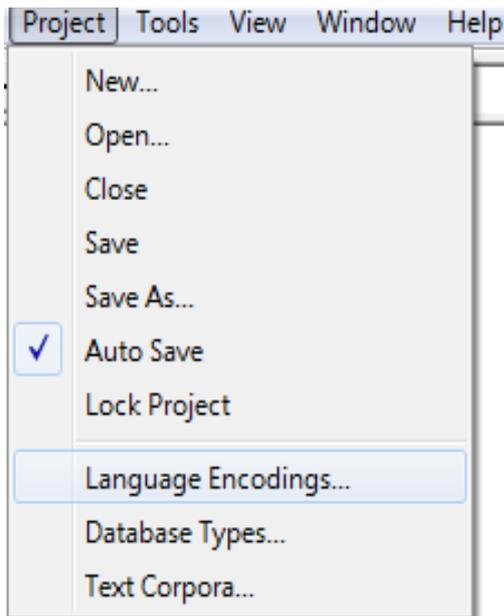
Il y a une vingtaine d'années encore le linguiste ne jurait que par la gomme, le crayon, les fiches cartonnées et le magnétophone qui étaient ses outils indispensables de travail dans la conduite des enquêtes de terrain. Aujourd'hui l'informatique est un outil technologique dont il est difficile de se passer. Il s'est imposé dans tous les secteurs de la vie et la linguistique n'y a pas échappé. Il n'est point important d'évoquer les raisons, il faut seulement faire le constat de son importance et le recours à l'outil pour le traitement des données, leur vérification et souvent même leur analyse. L'outil informatique a cependant ses contraintes. La fiche, le protocole d'enquête ou encore le questionnaire d'enquête (il s'agit de la même chose selon le domaine de la linguistique) une fois élaborée n'a plus d'ajouts ou de suppressions. Les différents champs ou rubriques doivent être scrupuleusement respectés pour le traitement des données. Cette stricte organisation impose des contraintes aux chercheurs qui doivent avoir le questionnaire le mieux adapté aux besoins de leurs recherches, d'où la nécessité de pré-enquêtes pour mettre à l'épreuve le questionnaire avant de le soumettre aux informateurs pour la phase définitive de l'enquête. A cause de l'informatique, la fiche est à la fois fiche d'enquête et de traitement des données. Il existe sur le marché

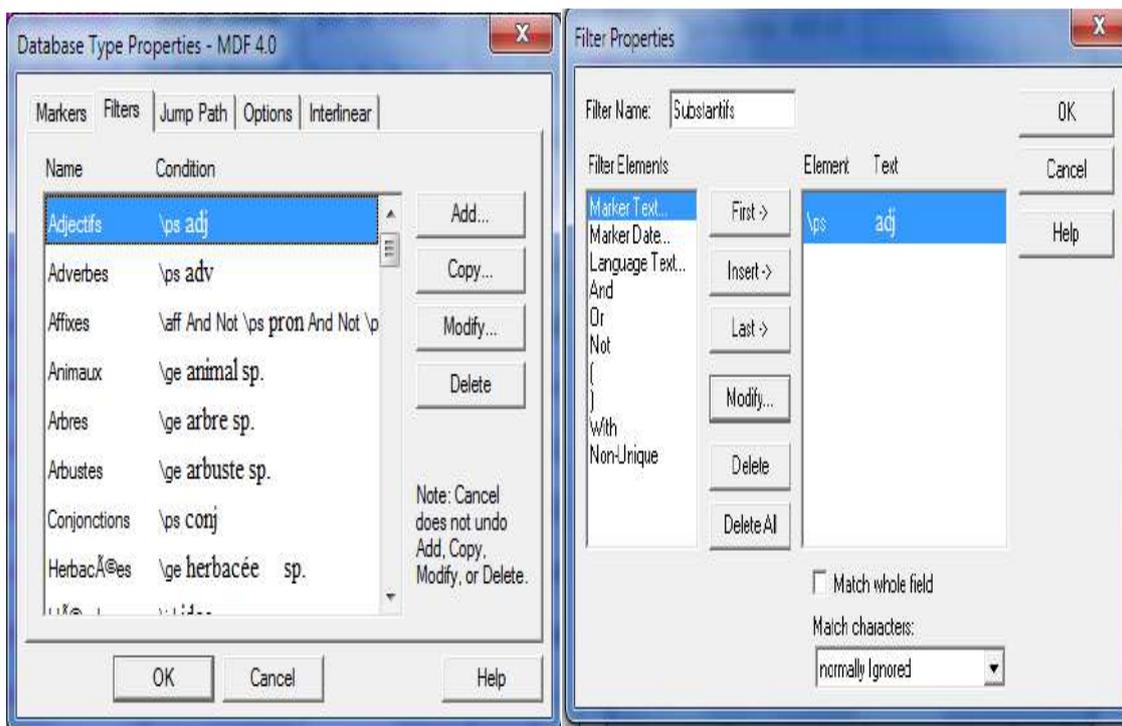
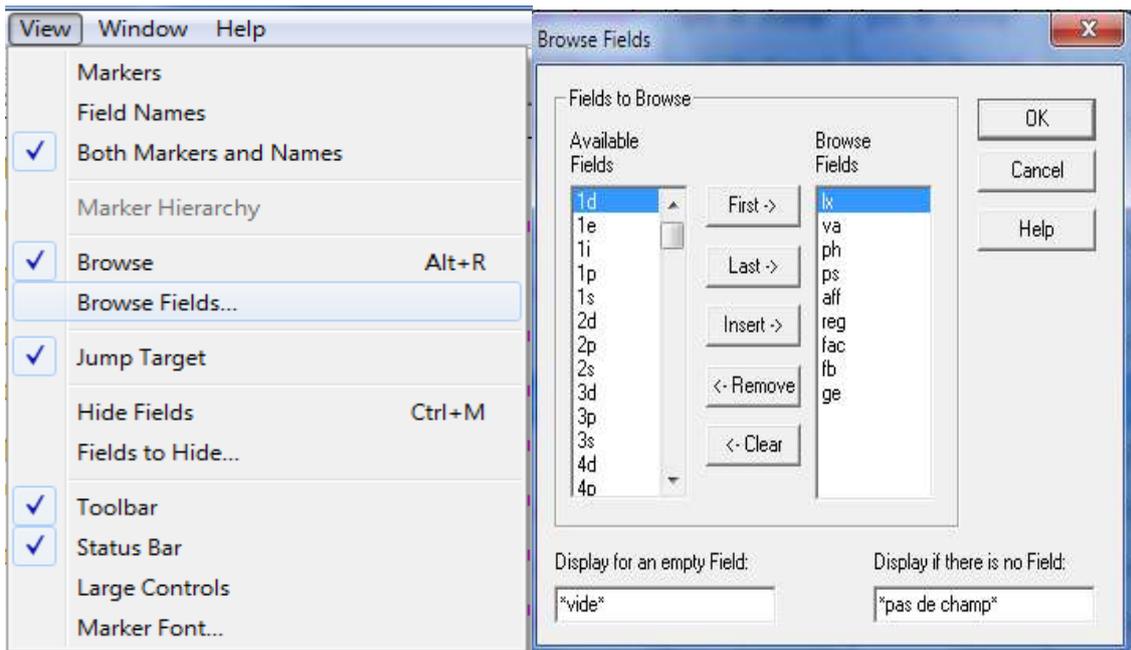
des logiciels courants de traitement de texte dont les plus connus sont Word et Excel ou peu connus comme la famille de logiciels de la Société Internationale de Linguistique (SIL) qui comportent des systèmes de gestion de bases de données fonctionnant par champs et enregistrements (**Toolbox**). Un champ correspond à une rubrique de la fiche. Il y a autant de champs dans un enregistrement que de rubriques sur la fiche cartonnée. Il s'établit une bijection entre l'ensemble des champs du logiciel et l'ensemble des rubriques sur le papier. Les champs, qu'ils soient facultatifs ou obligatoires, doivent toujours figurer dans l'enregistrement selon un ordre rigoureux. L'enregistrement est donc l'équivalent de la fiche cartonnée jadis utilisée par le linguiste et la base de données ne représente rien d'autre que le fichier que le linguiste compile durant des années de terrain. Ces types de bases de données, conçus avec des logiciels de traitement de textes, peuvent aisément être reproduits en texte avec les polices, les corps et les styles de caractères souhaités pour la mise en page définitive. Tous ces logiciels permettent une meilleure exploitation des données et donnent même accès aux diagrammes et aux courbes qui facilitent l'analyse.

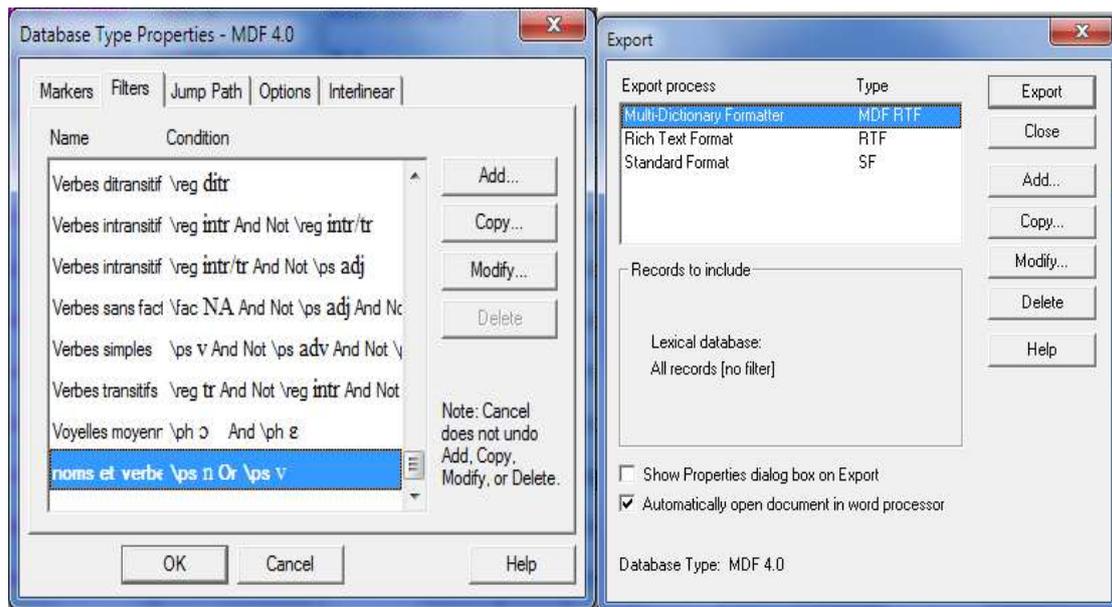


Les logiciels de la SIL comme **Toolbox** que nous maîtrisons sont spécialisés pour les linguistes, quel que soit leur champ d'action. Ils sont des aides précieuses à l'analyse phonologique, morphosyntaxique ou lexicale. On peut aussi utiliser des logiciels spécialisés de gestion de base de données comme **Flex** le plus récent logiciel de la SIL.

Ces types de logiciels **Toolbox** ou **Flex** permettent d'avoir des champs nettement séparés et de reproduire la fiche cartonnée telle quelle à l'ordinateur. Ils conviennent très bien au domaine de la lexicographie, particulièrement pour l'élaboration de dictionnaires électroniques.







abakati

anu

A - a

abakati [abaŋkati] *n* arbre sp (avocatier).
abertu [aŋbertu] *adj.* 1) ouvert i **ka boŋ disá**
porta di kintal abertu na fuska-fuska 2)
abertuda [aberŋtuda] *n* ouverture.
a-bo [aŋbo] *pron* toi.
abortá [aborŋta] *v* avorter.
abortu [aŋbortu] *n* avortement.
a-bos [abos] *pron* vous.
abril [aŋbril] *n* avril.
acá [aŋca] *inj.* 1) allez!, ouste! **acá forsa bay!**
acá sey li!
acukut [acuŋkut] *n* confectionné avec les
nervures de la palme du rônier, ce panier a une
forme d'outre et il est utilisé pour mettre du
poisson au moment de la pêche.
adorá [adoŋra] *v* adorer qqch/qqn. **gintiw ta**
adorá baloba l'animiste vénère le fétiche
adorantá [adoraŋta] *v fac.* faire faire adorer
qqm d'autre.
afeteré [afeteŋre] *n* habit sp.
aforsa [aŋforsa] *adv.* de force. i **mesté wo i**
negá i na bay aforsa
áfrika [aŋfrika] *n* Afrique.
afrikanu [afriŋkanu] *n* africain.
afumban [afumŋban] *n* herbacée sp (sote de
combres avec des piquants). *Ref:* A1
agmentá [agmenta] *v* maudire quelqu'un.
ago [aŋgo] *Variant: agora.* *adv.* et alors!
agora [aŋgora] *Variant: ago.* *adv.* alors.
agustu [aŋgustu] *n* août.
aju [aŋa] *n* ail.
ajuda [aŋuda] *n* aide.
akanjora [akaŋjora] *n* poisson sp. *Ref:*
A1-5
akara [akaŋra] *n* beignet de haricots.
alaman [alaŋman] *n* amende.
alegriya [aleŋgrija] *n* allégresse.
alfayatu [alfajatu] *n* tailleur.
algeŋ [alŋgeŋ] *pron indef.* 1) quelqu'un.
2) individu.
algodon [algoŋdon] *Variant: logodon.* *n*
arbuste sp (cotonnier fait 50 à 60 cm de haut).
ali [aŋli] *adv.* voici.
alimentá [alimta] *n* 1) aliment. 2) nourri

alkol [alŋkol] *n* alcool chirurgical.
alma [aŋalma] *n* âme.
almadija [almaŋdi] *n* outil de pêche constitué
de plusieurs hameçons disposés de manière
horizontale sur une file.
almentu [almentu] *n* profil.
almentuda [almentuda] *n* corpulence.
almosu [alŋmosu] *n* petit déjeuner. *Syn: matá*
bicu.
altar [alŋtar] *n* autel.
altu [aŋaltu] *adj.* 1) haut, géant. 2)
altuda [alŋtuda] *n* hauteur.
altura [alŋtura] *n* alentours.
altusé [altuŋse] *v* arriver en retard au lieu du
rendez-vous. **pabiya di bo N ta altusé pa bay**
féra tudu diya à cause de toi je pars tard au
marché tous les jours
amañá [amaŋna] *adv.* demain, après demain.
amaŋen [amaŋen] *n* oiseau sp.
ambisoŋ [ambiŋsoŋ] *n* ambition.
amen [aŋmen] *adv.* amen.
a-mi [aŋmi] *Variant: mi.* *pron.* moi.
amigu [aŋmigu] *n* ami.
amigundadi [amiguŋdadi] *n* camaraderie.
amontondadi [amontondadi] *n* paresse.
amontoy [amonŋtoŋ] *adj.* 1) paresseux. 2)
amonga [amonga] *adj.* 1) lépreux.
amor [aŋmor] *n* amour.
anagó [anaŋgo] *n* vêtement masculin.
andriyagu [andriŋjagu] *n* arbuste sp (peut
atteindre 3 à 5m de haut).
anel [aŋnel] *Variant: nel.* *n* bague.
a-nos [aŋnos] *sub.* nous.
ansa [aŋsa] *n* aile.
ansalma [aŋsalma] *n* esprits des ancêtres.
ansol [aŋsol] *n* hameçon.
antú [antu] *adv.* avant.
antá [aŋta] *inj.* ça alors.
antigadu [antiŋgadu] *adv.* 1) autrefois. 2) jadis.
3) anciennement.
antigu [antiŋgu] *adj.* 1) ancien.
antu di [antu di] *Variant: antu ku.* *conj sub*
temp. avant.
antu lu [antu lu] *conj sub temp.* avant que

5. La pré-enquête

Le questionnaire se situe au début et à la fin de tout travail d'enquête linguistique. C'est la raison pour laquelle il doit être aussi fiable que possible. Ceci conduit obligatoirement à faire des enquêtes préalables pour le mettre à l'épreuve en y apportant les modifications indispensables à son maniement au cours de l'enquête définitive. Soit dit en passant, il n'y a jamais d'enquête définitive tant que des éléments de dernière heure peuvent survenir à travers des informations complémentaires sinon capitales. La pré-enquête suppose le dépouillement des documents écrits. Cette phase est appelée enquête savante chez certains. Il s'agit des premières recherches qu'on effectue sur les travaux déjà réalisés sur la langue ou les langues sur lesquelles portera l'enquête. C'est l'occasion de dépouiller dictionnaires, lexiques, thèses, mémoires et tous les autres documents écrits disponibles.

Cette première phase permet de faire un début de collecte des *items lexicaux*, des expressions, et autres dictons, etc. de se faire une idée des caractéristiques morphosyntaxiques et des unités grammaticales, d'apprécier le sens des éléments, les nuances sémantiques, les synonymes, les antonymes, etc. Elle montre toute son efficacité pour l'assemblage des premiers éléments sur lesquels doit porter l'enquête, pour la confirmation ou pour l'infirmité des hypothèses ou des théories préalablement avancées. C'est un travail de laboratoire qui donne l'occasion à plusieurs chercheurs de travailler ensemble ou le cas échéant au principal acteur de prendre les avis des autres. Ce travail a donc pour mérite de procéder aux tout premiers ajustements qui mettent le chercheur sur la piste de l'item le plus évident mais ô combien fondamental à la réalisation de la fiche finale qui nécessitera des sorties sur le terrain. Les premières sorties sont ciblées sur un petit nombre d'informateurs pour permettre la modification du questionnaire initial pour qu'il réponde aux objectifs du travail à accomplir. Les points d'enquête sont choisis parce qu'ils offrent la possibilité de questionner des locuteurs qui apporteront les éléments indispensables à l'amélioration de la qualité du questionnaire. Au bout de plusieurs tentatives, le chercheur disposera d'un questionnaire satisfaisant qui permettra de mener une enquête à grande échelle.

6. La place de l'enquêteur

Si un questionnaire adéquat est primordial, la responsabilité de l'enquêteur (l'étudiant) est non moins négligeable. Une bonne partie du travail de collecte peut reposer sur lui à partir du moment où il est bien formé. C'est pourquoi il serait bon de l'associer au travail de confection du questionnaire en lui demandant assez souvent son avis. Il faut l'amener à se poser des questions sur la manière d'aborder les informateurs, de façon à ce que son interrogatoire (discussion avec l'informateur) obtienne des réponses objectives (c'est-à-dire sans calcul). La préparation de l'enquêteur doit prendre donc en compte tous les paramètres possibles pour affronter le terrain. Si par le passé on devait se contenter d'enquêteurs de bas niveau scolaire, aujourd'hui les étudiants de linguistique présentent le meilleur profil pour ce genre de travail. Ils ont non seulement la pratique de la langue de recherche (français), mais ils disposent également d'une formation théorique de base, qui sont des atouts non négligeables. Ils sont donc qualifiés pour être de vrais assistants au chercheur à qui ils peuvent beaucoup apporter. Au-delà d'une bonne pratique de la langue de recherche et d'une formation théorique de base qu'ils reçoivent, leur motivation est très grande, ce qui permet une meilleure qualité dans le travail.

Avant, les enquêteurs avaient pour seule motivation la gratification qu'ils reçoivent des chercheurs. Prendre des étudiants de linguistique a non seulement le mérite de les former davantage à leur métier mais aussi d'avoir plus d'assurance quant à la qualité du travail d'enquête. De plus, ils peuvent participer

aisément à la confection du questionnaire ainsi qu'au dépouillement des enquêtes une fois réalisées. Il est de ce fait plus avantageux d'avoir dans son effectif des étudiants avertis que d'avoir à faire à un personnel sans formation de base dans le domaine de la recherche, personnel qui très souvent ne comprend pas les tenants et les aboutissants du travail qu'il doit accomplir.

7. L'informateur

Il est au centre du travail car c'est à lui qu'on demande un service. Sa mise en confiance est très importante. Il peut connaître ou ignorer l'information recherchée. Il peut être tenté de parler pour plaire à l'enquêteur ou pour le mettre sur une fausse piste. Il peut attendre en retour de l'information, une rétribution. Il peut tirer une certaine fierté du contact qu'il a avec l'enquêteur. C'est donc un personnage assez complexe qui doit vivre une certaine complicité avec l'enquêteur - d'où la nécessité de ne pas le brusquer, de prendre le temps qu'il faut pour lui faire comprendre le jeu dans lequel il entre et dans lequel il n'a pas le droit de tricher. Dans ce cas, la liste (le questionnaire) préparée par le linguiste doit lui être soumise pour traduction. Au cas où il ne saurait lire, selon Bouquiaux et Thomas (1976, II : 272) « lui faire retenir de mémoire une dizaine de termes à enregistrer (moins ou plus selon ses capacités) et procéder à l'enregistrement par segments » serait la solution. Plusieurs fois, il est arrivé que des informateurs donnent des réponses pour le simple plaisir de l'enquêteur. Celui-ci, grâce à ses compétences, détecte « le réflexe trompeur » de l'informateur et le ramène dans les règles du jeu en le sensibilisant sur l'importance de son rôle. On peut se trouver face à des informateurs, dans le cadre d'enquête extensive, qui se refusent de parler à l'enquêteur. On met en avant les conditions sociologiques du milieu, qui veulent qu'on passe par le chef de village pour toute information à donner. Le chef du village lui-même, ou très souvent un de ses proches, se fait le devoir de désigner l'informateur. Une fois encore l'intervention de l'enquêteur est déterminante. Il est donc difficile de dissocier le rôle de l'informateur de la place que doit occuper l'enquêteur dans les enquêtes linguistiques, mais un questionnaire bien élaboré entre les mains d'un enquêteur bien formé et conscient de sa tâche doit pouvoir tirer le maximum d'une enquête en linguistique à travers des informateurs.

Conclusion

L'enquête de terrain est très complexe et il n'est pas évident de faire ressortir dans cette ébauche tous les aspects que certains ont par ailleurs décrit dans plusieurs volumes. Notre souci a été de montrer que l'enquête est avant tout un travail de longue haleine et que le linguiste sur le terrain doit faire preuve de tact et de simplicité. Son travail ne consiste pas seulement à obtenir des résultats scientifiques, mais doit veiller aussi à ne pas blesser son informateur et à en tirer le maximum de données sur la langue. Il doit aussi tisser des rapports humains et scientifiques avec les autres. En dépit de l'apport inestimable de l'informatique, pour le linguiste étranger, intégrer la culture par le biais de la langue est certainement un atout non négligeable pour l'analyse des faits linguistiques.

Bibliographie

- BOUQUIAUX, Luc & THOMAS, J. M. C. (dirs) (1976). Enquête et description des langues à tradition orale. Volume II. Paris : SELAF, 2ème éd. revue et augmentée.
- BAUDE, O. (2006), Corpus oraux, Guides des bonnes pratiques, 2006, CNRS-Éditions et Presses Universitaires d'Orléans.

- BAUDE, O. (2004) « Les corpus oraux entre science et patrimoine. L'expérience de l'observatoire des pratiques linguistiques », Actes du Colloque international du GRESEC « La publicisation de la science » (Grenoble) : 7-11.
- MONDADA, L. (1998) « Technologies et interactions sur le terrain du linguiste. Le travail du chercheur sur le terrain. Questionner les pratiques, les méthodes, les techniques de l'enquête ». Actes du Colloque de Lausanne 13-14.12.1998, Cahiers de l'ILSL 10 : 39-68.
- MONDADA, L. (2006) « Video recording as the reflexive preservation-configuration of phenomenal features for analysis », Knoblauch, H., Raab, J., H.-G. Soeffner, Schnettler, B. eds.
- MONDADA, L. (à paraître) « La demande d'autorisation comme moment structurant pour l'enregistrement et l'analyse des pratiques bilingues », Tranel, Université de Neuchâtel.