Langage, Langues et Cultures d'Afrique Noire

Llacan

CE QUE LA TECHNOLOGIE A CHANGÉ DANS L'ÉLABORATION D'UNE GRAMMAIRE DE RÉFÉRENCE POUR LES LANGUES AFRICAINES ORALES

Paulette Roulon-Doko
roulon@vjf.cnrs.fr

PLAN GÉNÉRAL

- Ecouter et transcrire l'oral
- Qu'apporte l'enregistrement numérique au chercheur?
- La constitution d'un corpus de référence
- Associer enregistrement numérique et transcription
- Quelques exemples tirés de la grammaire du gbaya

Le support d'enregistrement a totalement changé

Avant le développement des techniques modernes d'enregistrements numériques dont nous disposons actuellement, l'enregistrement depuis les années 1960 se faisait avec un magnétophone et des bandes (UHER, NAGRA). Ces bandes étant jugées fragiles, il était conseillé d'éviter de les écouter souvent, ce qui conduisait le linguiste à en faire une transcription qui devenait ensuite la base de son travail. Les bandes étaient rangées comme archives. En fin de compte bien qu'à la base il s'agisse d'une production orale, le travail du linguiste se faisait essentiellement sur la transcription effectuée. Actuellement de tels enregistrements peuvent être numérisés en wav. et se retrouvent sur un support numérique qui, lui, peut être entendu aussi souvent que nécessaire.

Un support fiable reproductible et réécoutable autant que nécessaire

Des cahiers à une base de données

Autrefois, les cahiers et les fiches, aujourd'hui des logiciels de base de données comme Toolbox permettent d'introduire toutes ces données dans l'ordinateur, modifiant le travail de compilation. Il est possible de créer autant de fichiers de données qu'on le souhaite, chaque fichier engrangeant des informations spécifiques à un domaine, pas nécessairement utiles pour un autre. La copie d'une fiche d'un fichier à l'autre, en cas de besoins, ne pose aucun problème.

De plus un logiciel comme Toolbox permet l'interlinéarisation des textes transcrits, ce qui impose une rigueur dans l'analyse en amont, et indique très vite si une analyse ne convient pas.

L'ensemble de ces textes, constituent un corpus qui peut être interrogé sans avoir besoin d'ouvrir chaque fichier. On peut ainsi chercher une forme donnée dans un tel corpus et avoir une idée de sa fréquence en discours.

Compilation et statistique de fréquence facilitées.

Langage, Langues et Cultures d'Afrique Noire

Transcription manuscrite

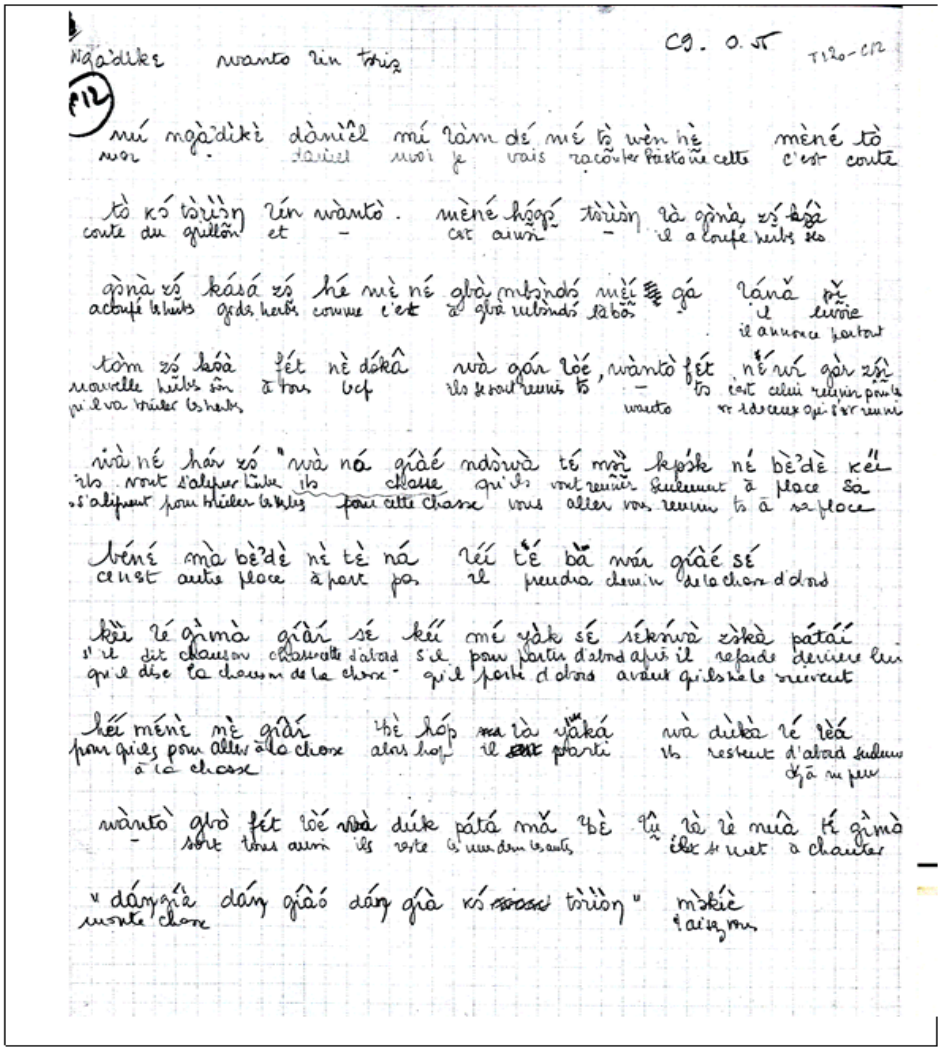

Fig. 1. Une page manuscrite d'un conte gbaya.

INDEXATION DES PHRASES DU CORPUS

Le texte présente alors deux lignes, la ligne de référence (\ref) et la ligne de texte (\tx) – ce conte comporte 68 phrases numérotées – (*cf.* fig. 3).

\ref	T120-C12 001
\tx	**mí ngàdîkɛ̀ dànìɛ̂l, mí ʔàm dɛ́-mɛ́-tɔ̰ wèn hɛ̰.**
\ref	T120-C12 002
\tx	**mɛ̀ né tò, tò kɔ́ tɔ̀rìɔ̀ŋ ʔín wàntò.**
\ref	T120-C12 003
\tx	**mɛ̀ né hɔ́gɔ́, tɔ̀rìɔ̀ŋ ʔà̰ gɔ̀nà zɔ̰́ kɔ́à̰, gɔ̀nà zɔ̰́ gásá zɔ̰́ hé mɛ̀ né gbàmbɔ̀ndɔ́ mɛ̀í gá.**

Fig. 3. Les 3 premières phrases du conte numérotées sous Shoebox.

un mot sous mot (\ma)
la mention des catégories grammaticales (\am)
et une ligne indépendante des précédentes, la traduction dans la langue cible
traduction (\tr)

\ref T120-C12 003

\tx	**mɛ̀**	**né**	**hɔ́gɔ́,**		**tɔ̀rìɔ̀ŋ**	**ʔà**						
\rm	mɔ̀	né	hɔ́gɔ́		tɔ̀rìɔ̀ŋ	ʔà						
\ma	chose	être	comme ça		petit grillon	3S						
\am	N	PRED	ADV		N	PR						
\tx	**gɔ̀nà**						**zɔ́**		**kɔ́à,**			
\rm	BHa-	gɔn		-H			zɔ́		kɔ́	ʔà,		
\ma	Acc-	découper		-D			herbes		de	3S		
\am	MV-	V		-FCT			N		FCT	PR		
\tx	**gɔ̀nà**				**zɔ́**	**gásá**		**zɔ́**				
\rm	BHa-	gɔn	-H		zɔ́	gásá		zɔ́				
\ma	Acc-	découper	-D		herbes	grand		herbes				
\am	MV-	V	-FCT		N	AV		N				
\tx	**hé**	**mɛ̀**	**né**	**gbàmbɔ̀ndɔ́**	**mɛ̀í**		**gá.**					
\rm	hé	mɔ̀	né	gbàmbɔ̀ndɔ́	mɛ̀	-ìí	gá.					
\ma	comme	chose	être	herbes sp	là-bas	-anaph	comme					
\am	SUB	N	PRED	NPR	ADV	-MOD	SUB					

\tr C'est ainsi que le Grillon il a délimité son territoire de chasse, il a délimité un territoire, un grand territoire, comme qui dirait Gbambondo là-bas.

Fig. 4. Désamalgamage de la phrase 3 sous Shoebox

Langage, Langues et Cultures d'Afrique Noire
Llacan

Qu'apporte l'enregistrement numérique?

Des logiciels spécialisés associant enregistrement et transcription

La possibilité permanente d'aller écouter le produit original enregistré permet de mieux prendre en compte **la spécificité de cette production orale**. En effet, lorsqu'on transcrit, le plus souvent avec l'aide d'un [ou des] locuteur[s], celui-ci peut facilement reformater une phrase, un syntagme, sans même parler des 'disfluencies' qui sont le plus souvent supprimées. L'écoute à l'oreille de l'enregistrement, même s'il peut être répété à l'infini, laisse ce problème entier.

Des logiciels comme **ELAN** ou **PRAAT** [ce dernier plus adaptée pour une transcription phonétique très fine] permettent de **visualiser le flux oral**, d'en identifier les différents segments et donc de ne rien laisser de côté. Cela permet aussi de resituer une forme grammaticale dans son co-texte, et de voir si celui-ci peut être systématisé.

Le but du travail sur corpus n’est pas une analyse fine de la prononciation, c’est une **méthode pour étudier la sémantique et la syntaxe** d’une langue donnée.

Image, son, transcription et analyse associés

L'utilisation conjointe d'ELAN et de Toolbox permet d'importer l'analyse de la transcription dans ELAN et de l'associer à la bande sonore. Plus récemment le LLACAN a développé des outils pour pouvoir faire cette interlinéarisation directement sous ELAN, ce qui produit le même résultat que l'association précédemment mentionnée, avec un plus : des recherches multiples qu'on ne peut pas faire sous Toolbox.

J'ajouterai qu'ELAN, développé à la base pour étudier le langage des sourds, permet aussi d'intégrer la vidéo. Il devient alors possible d'avoir sur un même support, vidéo, audio et analyse grammaticale. Cela ouvre la voix à une prise en compte beaucoup plus fine de la complexité du corpus étudié.

Tous ces moyens ont de plus l'avantage de pouvoir **mettre le corpus à la disposition de tous**, et d'être accessible en particulier aux locuteurs de ces langues. Cela ouvre donc une nouvelle ère, qui permettra tant aux chercheurs qu'aux locuteurs de pouvoir 'juger sur pièces' les analyses proposées.

catégorie grammaticale – et suivie par une ligne libre de traduction. Je donne ci-après le fac-similé d'une ligne d'un conte gbaya T25-C11, consultable en ligne sur le site du CRDO[7] (Centre de ressources pour la description de l'oral) Paris (*cf.* Fig. 5).

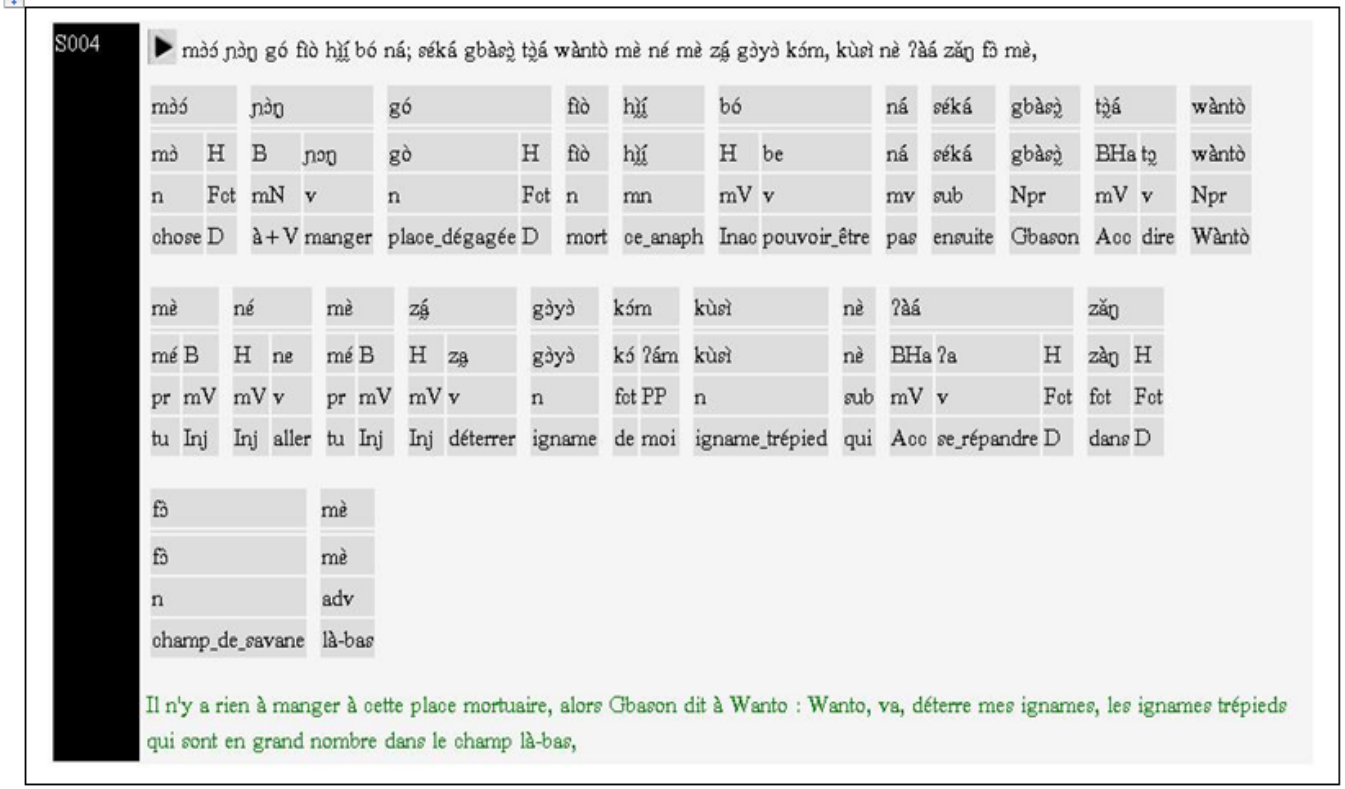

Fig. 5. Une phrase en ligne du conte T25-C11 sur Internet

L'EXEMPLE DE LA GRAMMAIRE GBAYA

Le gbaya du nord (gya, Niger-Congo, Adamawa-Ubangui) qui est parlé en Afrique Centrale (2/3 RCA, 1/3 Cameroun) par environ 500 000 locuteurs.
Le miratif,
Des termes qui semblent synonymes : bú ~ búkꭥ « dix »,
kóré ~ kórá : Le cas de « sec », formé sur le verbe *kɔr* « sécher », qui présente soit un a en application de la règle de formation *kɔ́rá*, soit un e en remplacement du a final *kóré* que j'ai longtemps considéré comme des variantes libres, doit en fin de compte être analysé (cf VI.1.1.3.1.b) comme deux éléments distincts. L'adjectif verbal *kɔ́rá* « sec » et le nom déverbatif *kóré* « aliment séché ».
Les déictiques : la distinction entre adverbe et démonstratif
Les modalités d'énoncés, des éléments qui ne peuvent être compris qu'en situation et dans un corpus suffisant

EXEMPLE 1.

libres. L'exemple suivant m'incite à reconsidérer cette analyse.

1. **ɓásí zɔ̀m *kóré sàdî ʔèá dún bɔ̀tɔ̀ híp***
   IMP.porter AUG.D ~ viande seulement I.ACC.remplir panier_*sp.* à_ras_bord

   Apporte beaucoup de morceaux de viande séchée et bourre bien le panier.

En effet, si l'augmentatif portait sur la valeur de l'adjectif « sec », la détermination par l'augmentatif signifierai « très sec », or ici, la signification produite est « beaucoup de viande séchée ». Le comportement de *kóré* n'est donc pas identique à celui de *kórá* qui, déterminé par *zɔ̀m*, signifie « très sec ». On doit donc distinguer entre :

2. **zɔ̀m kórá nù**
   AUG.D ~ terre

   Un sol très sec

3. **zɔ̀m kóré sàdî**
   AUG.D ~ animal

   Beaucoup de viande séchée

Parlant du buffle qu'ils ont tué, le frère et la sœur,

4. **wà dɛ́ kóréà ʔá yík wèè ʔèá**
   3P INAC.faire aliment_séché.DEF I.ACC.mettre surface feu seulement

   Ils en font de la nourriture séchée en la faisant boucaner sur le feu. (T9-C7 027)

L'inventaire de ces 'modalités d'énoncés' en gbaya ɓodoe :

wó	vraiment, à vrai dire, en vérité	wá	hélas
yè	finalement, enfin, en fin de compte	gɔ̀	assurément
yè	en effet	ʔɔ̀óyè	bien entendu
ʔɛ́	donc		

Tableau 1. Les modalités d'énonciation

Les exemples de wó dans un corpus enregistré de plus de 2 heures.

1. **ʔám sàà mɛ́ wó**
je ACC+appeler+D toi ME

Oui *je t'appelle !*

2. **ʔà tɛ́-gbìní wó**
elle V.INAC+casser+ANAPH ME

Elle va ***sûrement*** *casser.* [le locuteur vient de dire que cela va casser et l'énonciateur reprend avec cet énoncé]

3. **ʔèì nɛ́ gíà sɔ́sɔ̀ɔ̀ wó**
SUB+on INAC+aller chasse aujourd'hui ME

[L'annonce d'un chasse vient d'être faite] *Alors on va* ***donc*** *à la chasse aujourd'hui.*

4. **mɛ́ kɔ́ sɔ́m wó**
tu INAC+aimer REVOLU+moi ME

Ah si tu m'aimais ***vraiment****.* [elle veut aguicher le garçon]

5. **ʔá-nɛ̀ dìlà nɛ́ wó**
voilà que lion INAC+aller ME

[Le lion a entrepris de faire le travail pour son beau-père]. *Et le lion y va* ***donc****.*

6. **ʔà dí ná wó**
il INAC+être beau pas ME

Il n'est ***vraiment*** *pas beau.*

EXEMPLE 2 (FIN)

wó	vraiment	assertion validée par l'énonciateur	E
yè	finalement	conclusion logique de la situation pour le locuteur ou l'acteur	L
yɛ̀	en effet	la situation est un fait avéré qui vient d'être présenté	E
ʔɛ́	donc	a) conséquence perçue par le locuteur ou l'acteur b) surprise du locuteur ou l'acteur	L L
wá	hélas	plainte, supplique du locuteur ou l'acteur	L
gɔ̀	assurément	l'énonciateur se porte garant de la vérité de la situation	E
ʔɔ̀ɔ́yɛ̀	bien entendu	situation conforme à l'ordre des choses	E

e = énonciateur ; l = locuteur ou acteur

Tableau 2. Valeur sémantique des modalités d'énonciation

CONCLUSION

L'intérêt d'un **travail sur corpus** est de pouvoir systématiquement contrôler la fréquence d'emploi d'une forme identifiée. Cette information me semble fondamentale pour définir la place de ladite forme dans le système. Certaines langues par exemple font un grand usage de la topicalisation tandis que d'autres utilisent plus la focalisation. Sans étude de corpus, seuls les procédés utilisés peuvent être présentés et non la mention de la place que chacun y occupe.

L'**élicitation** qui incite à recourir à la compétence/performance du locuteur peut très vite créer des énoncés dont l'emploi n'est pas véritablement attesté. Ce type de travail conduit souvent aux questionnaires, utiles sans doute pour multiplier les paradigmes morphologiques par exemple, mais beaucoup moins fiables lorsqu'il s'agit d'autres domaines de la linguistique (problème de calque de la langue cible entre autres).

La mise en ligne accessible à tous, y compris les locuteurs de la langue, permet (i) un archivage, (ii) une diffusion et (iii) une analyse qui peut être contrôlée voire discutée par tous, chercheurs ou locuteurs.