



HAL
open science

Le conte : du terrain à Internet, pratique et déontologie

Paulette Roulon-Doko

► **To cite this version:**

Paulette Roulon-Doko. Le conte : du terrain à Internet, pratique et déontologie. Cahiers de Littérature Orale, 2010, 63-64 (Pratiques d'enquêtes), pp.295-304. halshs-00720163

HAL Id: halshs-00720163

<https://shs.hal.science/halshs-00720163>

Submitted on 23 Jul 2012

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.

Le conte : du terrain à Internet, pratique et déontologie.

Paulette ROULON-DOKO
DR au CNRS
LLACAN - UMR 8135

La collecte de contes sur le terrain par le recours à l'enregistrement de la performance sur un magnétophone analytique, ou de nos jours numérique, est pratiquée depuis près de trente ans. Les archives sonores qui témoignent d'une oralité propre à chaque culture, bien qu'étant conservées n'étaient le plus souvent pas disponibles pour un large public. Ces corpus – bandes magnétiques, minicassettes – dont l'écoute restait en générale confidentielle, limitée à ceux qui les avaient collectés peuvent maintenant, une fois convertis en format numérique – CDrom ou Internet – être écoutés par tous. Cependant ces nouveaux moyens moderne de diffusion de l'information doivent être utilisés avec un double souci, en amont, celui d'être fidèle à la performance enregistrée et respectueux des personnes qui ont fourni l'information et en aval, celui de diffuser un produit le plus complet possible restituant l'authenticité de ladite performance. Je vais donc présenter ici les étapes successives de ce cheminement qui mène de l'enregistrement sur le terrain à la diffusion sur Internet et souligner quelques uns des problèmes qui peuvent se poser. Je prendrai pour ce faire l'exemple d'un conte gbaya¹.

La première transcription

Il convient tout d'abord de rappeler que tout enregistrement fait sur le terrain doit mentionner les conditions de collecte. Il faut bien sûr préciser le lieu de la collecte, la date, le nom de la personne enregistrée, la langue parlée mais aussi spécifier les caractéristiques de la performance, en mentionnant en particulier si l'enregistrement a été réalisé en présence d'un public – veillée traditionnelle, veillée provoquée, etc. – ou sans public, le conteur étant seul. Tous ces éléments sont très importants pour cerner l'authenticité de la performance. Il faut de plus toujours transcrire le texte du conte, sur le terrain même. C'est le plus souvent en réécoutant l'enregistrement avec un ou plusieurs locuteurs de la langue que la transcription est faite. Chaque phrase est écoutée, reprise au besoin par un des locuteurs présents puis transcrite par le chercheur selon un système de notation sur lequel je ne m'attarderai pas ici. Cette première transcription est en général accompagnée d'une traduction qui aligne sous chaque terme de la langue source un terme de la langue cible, sans analyse grammaticale détaillée (*cf.* fig. 1). Le but de cette traduction étant de permettre simplement la compréhension dudit texte. Plus on connaît la langue et plus la transcription se fait aisément. Personnellement j'ai

¹ Les populations qui se reconnaissent sous le nom de Gbaya occupent un territoire situé pour les quatre cinquième à l'ouest de la République Centrafricaine et pour le dernier cinquième au centre-est du Cameroun. L'étude présente porte plus précisément sur les Gbaya 'bodoe qui font partie du groupe numériquement le plus important (160 000 locuteurs) les Gbaya kara. Ils forment un groupe homogène d'environ 5000 personnes réparties en une quarantaine de villages au sud-ouest de Bouar, en République Centrafricaine. La langue gbaya appartient, selon la classification de Greenberg, au groupe 1 de la branche orientale de la sous-famille 6 "Adamawa oriental" de la famille Niger-Congo, plus volontiers appelé Oubanguien dans les études plus récentes.

toujours commencé par transcrire sur papier, à la main, la traduction en langue cible étant facilement placée sous le texte en langue source.

Pour un chercheur, multiplier des enregistrements non identifiés et non transcrits, c'est constituer un corpus inutilisable. Il n'en va pas de même des locuteurs d'une langue qui enregistreraient eux-mêmes des contes, car ils sont à même de les comprendre et ils pourront toujours les transcrire, sans aide particulière, le moment venu. Charge à eux cependant de toujours mentionner les conditions de leur collecte.

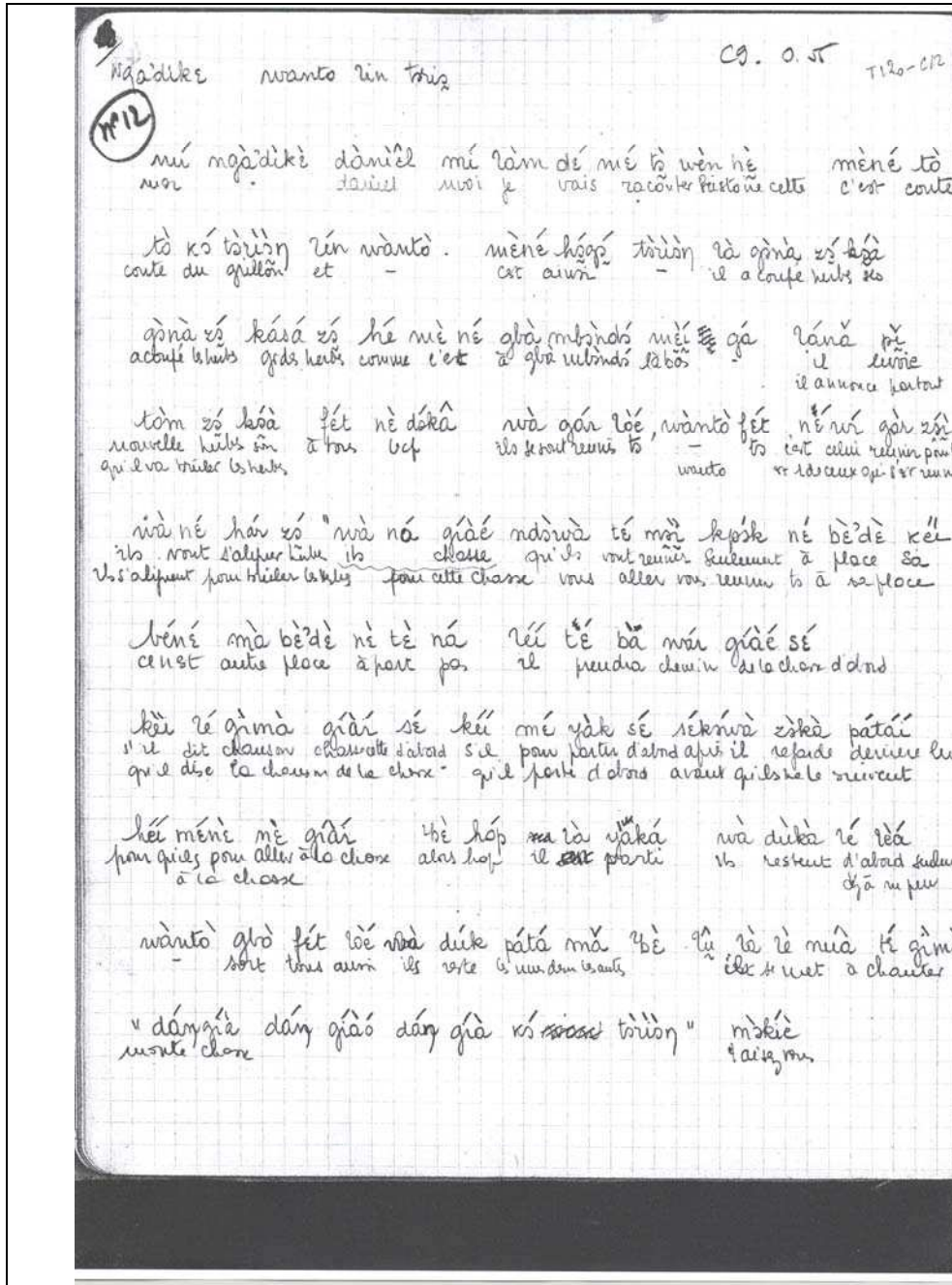


Fig. 1. Une page manuscrite d'un conte gbya.

La transcription sur ordinateur

Dans une première étape, le texte transcrit à la main va pouvoir être tapé en langue source sur un ordinateur dans un traitement de texte, au format word par exemple. Il faut pour cela pouvoir disposer d'un système de fonte (type Sil-Doulos²) et d'un clavier (type Keyman³) adéquats. Seul le texte en langue source est tapé, il occupe un espace beaucoup plus ramassé que la transcription accompagnée de sa traduction souscrite mot sous mot – 4 pages manuscrites correspondent à peu près à 1 page tapée – (cf. fig 2). Se pose alors le problème de la ponctuation qui est le plus souvent absente du premier texte, des espaces séparant tout simplement les termes ou les phrases entre eux. Il faut procéder à un découpage en phrases qui seront chacune ponctuées par l'insertion d'un point final.

mí ngàdikè danièl, mí ?àm dé-mé-t? wèn h?. m? né tò, tò kó tòrìd? ?ín wàntò. m? né h?g? tòrìd? ?à gònà z? k?à, gònà z? gásá z? hé m? né gbàm?ndó mèí gá. ?á nàá p?í tòm z? k?à fét n? d?kàà. wà gár ?d?, wàntò fét né wí gár z?. wà né hár z?. wà ná gíái ndòwà té-mòì kpók né bèdè kéí, bé né mà bèdè n? tè ná. ?éí té-bà wár gíái sé. kèì ?é gìmà gíái sé kéí mé-yàk sé sékòwà zòkà pátái héí mé-nè n? gíái. bèè hóp ?à yàká. wà dùka ?é ?èá. wàntò gbòd fét ?d?, wà dúk pátá màá. bèè ?yù ?à ?è núa té gìmà. "dáj gía, dáj gía wó, dáj gía kó tòrìd?", m?kíè.

Fig. 2. Transcription sous word de la page manuscrite de la fig. 1

Les productions orales ont de façon courante développé un système complexe de jonctifs ou coordonnants comparables à ce qui était attesté dans les premières transcription de la prose en ancien français ou en ancien anglais (Jacquesson, 2007 : 379-382). Ces éléments qui marquent les liaisons entre les phrases ou les propositions seront donc ici précédées d'un point final ou d'une virgule.

Dans une seconde étape, le texte tapé sous word va pouvoir être converti au format d'une base données, au format Shoebox⁴ par exemple. Pour ce faire, le conte au format word (.doc) sera tout simplement enregistré en format texte (.txt) qui est un format compatible avec Shoebox qui pourra permettre ensuite un découpage systématique du texte en fonction de l'analyse en amont qu'en aura fait le chercheur. Avant l'enregistrement au format texte, il convient de mentionner sur une ligne qui précède le texte, la référence qu'on souhaite donner au conte en la faisant précéder de la mention '\name'. Ainsi pour le conte ci-dessus, c'est la numérotation du conte dans le corpus de référence qui est adopté comme référence : \name T120-C12 .

En ouvrant ensuite sous shoebox le texte ainsi formaté, celui-ci est alors automatiquement divisé en lignes numérotées. La numérotation qui reprend la référence mentionnée en tête du texte est suivie du numéro de phrase et ce sont les

² Il s'agit d'un système de polices adaptées à la langue étudiée.

³ Le logiciel keyman, permet à chacun de définir un clavier qui facilite la frappe des caractères spéciaux intégrés dans la fonte choisie. Cela joue en particulier sur des éléments comme les tons, la nasalité, les glottalisées, etc.

⁴ La version la plus récente de ce logiciel est maintenant appelée toolbox.

points qui sont retenus par shoebox pour une numération systématique des phrases. 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).

<code>\ref</code>	T120-C12 001
<code>\tx</code>	mí ngàḍikè dānìèl, mí ?àm dé-mé-tỳ wèn hẹ.
<code>\ref</code>	T120-C12 002
<code>\tx</code>	mè né tò, tò kó t̀r̀ìḍḡ ?ín wànt̀.
<code>\ref</code>	T120-C12 003
<code>\tx</code>	mè né h́ḡḡ, t̀r̀ìḍḡ ?à g̀nà ź ḱà, g̀nà ź ǵsá ź hé mè né gb̀amb̀nd́ mèí ǵ.

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

L'analyse sous shoebox

Chaque phrase de texte découpée automatiquement va pouvoir être analysée et systématiquement traitée, faisant apparaître trois nouvelles lignes qui sont en colonnes, permettant un alignement mot sous mot :

un découpage morphologique ou désamalgamage⁵ (`\rm`)

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`)

⁵ Cette opération consiste à isoler les différentes parties constituant le terme de sorte que chaque élément isolé puisse ensuite être identifié comme une entrée du fichier de désamalgamage. Ce fichier en effet comprend en plus des termes retenus dans le lexique, tous les éléments grammaticaux qui n'ont pas de forme morphématique autonome comme les affixes, les infixes et les suffixes.

\ref T120-C12 003							
\tx	mè	né	hǎǎǎ,	tòrìd̩	ʔà		
\rm	m̩	né	hǎǎǎ	tòrìd̩	ʔà		
\ma	chose	être	comme ça	petit grillon	3S		
\am	N	PRED	ADV	N	PR		
\tx	g̩nà			z̩	k̩à,		
\rm	BHa-	g̩n	-H	z̩	k̩	ʔà,	
\ma	Acc-	découper	-D	herbes	de	3S	
\am	MV-	V	-FCT	N	FCT	PR	
\tx	g̩nà			z̩	g̩sá	z̩	
\rm	BHa-	g̩n	-H	z̩	g̩sá	z̩	
\ma	Acc-	découper	-D	herbes	grand	herbes	
\am	MV-	V	-FCT	N	AV	N	
\tx	hé	mè	né	gbàmbòndó	mèí	gá.	
\rm	hé	m̩	né	gbàmbòndó	mè	-í	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

Un tel travail ne peut se faire que si l'analyse grammaticale de la langue source est déjà établie, au moins dans ses grandes lignes. Chaque terme identifié peut être introduit dans un lexique dont on peut à sa guise organiser les fiches (le nombre de champs possibles n'est pas limité). Ce travail est simple pour tous les termes qui sont invariables. Pour tous les autres dont la forme peut varier selon la position dans la phrase, il convient de procéder à un désamalgamage. Ceci permet d'isoler le terme de base choisi pour être introduit comme terme de référence dans le lexique, tout en précisant le ou les éléments grammaticaux dont il est le support. La forme amalgamée est introduite dans un champ spécifique du fichier de désamalgamage auquel correspond un champ d'analyse dont chaque terme a, lui, une entrée dans ce fichier. Il est en effet nécessaire d'avoir, à cette étape du travail, choisi une convention pour la transcription des éléments grammaticaux. Ces derniers, lorsqu'ils n'ont pas de forme segmentale libre, ne peuvent figurer dans le lexique. Au fur et à mesure que l'on progresse dans l'analyse d'un texte, ce sont donc deux fichiers qui sont peu à peu constitués : un fichier de désamalgamage qui mentionne pour chaque terme lexical retenu, les variantes grammaticales possibles et leur analyse et un fichier lexical qui mentionne pour chaque terme au moins, sa catégorie grammaticale et son ou ses sens. Je prendrai deux exemples attestés dans la fig. 4 ci-dessus :

- 1) la forme verbale conjuguée du gbaya *g̩nà* où on repère l'élément 'BHa' dans la ligne de mot sous mot (\ma) qui renvoie à la modalité verbale (MV) de l'accompli et constitue une entrée dans le fichier de désamalgame.
- 2) l'anaphorique noté *-í* dans la ligne de découpage morphologique (\rm) et traduit par 'anaph' dans la ligne de mot à mot, qui constitue lui aussi une entrée dans le fichier de désamalgame.

Cette analyse dépend bien entendu du niveau d'analyse maîtrisé par le chercheur. Au minimum, on peut envisager un simple repérage des termes, faisant ressortir le terme choisi comme référence, sans analyse des éléments modificateurs ; au mieux, c'est la totalité de l'analyse grammaticale qui est alors restituée. Dans tous les cas la ligne de traduction mot à mot permet de suivre l'ordre exact des termes dans la langue source.

La ligne de traduction, tout en restant proche du texte en langue source, se plie aux contraintes de la langue cible et produit une formulation compréhensible par le lecteur de la langue cible. On peut considérer cette traduction comme une traduction littérale. On peut ensuite tendre à formuler une traduction que j'appellerai 'littéraire' qui, tout en restant fidèle, prenne plus de distance avec le texte source et intègre, en autres, des éléments d'ordre culturel qui échappent au lecteur qui ne connaît pas la culture en question. Dans cette optique, la traduction littérale de la phrase de la fig. 4 serait :

« Un jour, le grillon avait délimité dans la savane, comme qui dirait au lieu-dit Gba-Mbondo, un territoire pour sa chasse au feu. »

La mise en ligne

S'il existe une version numérique de l'enregistrement, il est alors possible d'y associer le texte traité par shoebox afin de le mettre en ligne comme cela est fait pour le programme d'archivage de corpus oraux par Christian Chanard au LLACAN⁶. Le texte ainsi mis en ligne peut être soit lu ligne par ligne, soit être écouté sans interruption étant alors la fidèle reproduction de l'enregistrement original. La transcription est, comme dans shoebox, accompagnée des trois lignes – désamalgamage, traduction mot à mot, 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 CRDO⁷ Paris (cf. Fig. 5).

S004 ▶ m̀s̀ ǹǹ g̀ f̀ h̀j̀ b̀ ǹ, s̀k̀ g̀b̀s̀ t̀á ẁnt̀ m̀ ǹ m̀ z̀ g̀ỳ k̀m̀, k̀s̀ ǹ ?áá z̀ǹ f̀ m̀,

m̀s̀	ǹǹ	g̀	f̀	h̀j̀	b̀	ǹ	s̀k̀	g̀b̀s̀	t̀á	ẁnt̀			
m̀	H B	ǹǹ	g̀	H f̀	h̀j̀	H be	ǹ	s̀k̀	g̀b̀s̀	BHa t̀	ẁnt̀		
n	Fot mN	v	n	Fot n	mn	mV v	mv	sub	Npr	mV v	Npr		
chose D	à+ V	manger	place_dégagée D	mort	ce_anaph	Inac	pouvoir_être	pas	ensuite	Cbason	Acc	dire	Ẁnt̀

m̀	ǹ	m̀	z̀	g̀ỳ	k̀m̀	k̀s̀	ǹ	?áá	z̀ǹ						
m̀	B H	ne	m̀	B H	z̀	g̀ỳ	k̀	?ám	k̀s̀	ǹ	BHa ?a	H	z̀ǹ	H	
pr	mV	mV v	pr	mV	mV v	n	fot	PP	n	sub	mV v	Fot	fot	Fot	
tu	Inj	Inj	aller	tu	Inj	Inj	déterrer	igname	de moi	igname	trépiéd	qui	Acc	se_répandre D	dans D

f̀	m̀
f̀	m̀
n	adv
champ_de_savane	là-bas

Il n'y a rien à manger à cette place mortuaire, alors Cbason dit à Wanto : Wanto, va, déterre mes ignames, les ignames trépiéd qui sont en grand nombre dans le champ là-bas,

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

⁶ llacan@vjf.cnrs.fr.

⁷ http://crdo.risc.cnrs.fr/exist/crdo/

A ce stade, il est nécessaire de réécouter le texte d'origine pour vérifier qu'il coïncide bien avec la version transcrite. Je soulignerai ici quelques cas de révisions à faire qui concernent toutes les 'corrections' volontaires ou involontaires qui ont pu se produire au moment de la première transcription. En effet, les personnes qui ont aidé à transcrire le texte sur le terrain peuvent, dans un souci de normalisation involontaire, avoir gommé des éléments que l'écoute de la bande originale restitue. Il en est ainsi des hésitations ou les reprises après erreur – se tromper dans le nom d'un personnage par exemple – spontanément supprimées et corrigées ; des interventions des auditeurs qu'on entend clairement qui ont pu également être supprimées. Dans ce dernier cas il faut trouver un procédé particulier – mise en italique, mise entre crochets, etc. – pour identifier le passage en question qu'on se doit bien sûr, s'il est audible, de restituer. Il ne me semble pas souhaitable de 'nettoyer' la bande en cherchant à effacer ces passages, car cela porte atteinte à l'authenticité de la performance. Si un passage n'est pas réellement audible, on peut toujours le signaler par une mention de 'non transcription' telle [...] par exemple. De même si, par souci d'hypercorrection, des termes prononcés dans une langue autre – emprunts au français ou à une langue véhiculaire commune dans la région par exemple – ont été remplacés par le terme correspondant de la langue cible, il convient ici de les restituer, pour suivre au plus près l'enregistrement. Enfin, dans le cas de chantefables dont le chant s'articule entre un solo dit par le conteur et le répons dit par une grande partie des auditeurs, il faut trouver un moyen de repérer les deux parties. Cela est d'autant plus important que parfois, le répons peut se poursuivre tandis que le conteur, laissant de côté le solo, continue à faire progresser les événements du conte, jusqu'à ce qu'il y mette un terme par l'emploi d'une formule consacrée. Cela est bien clair à l'écoute de la bande mais peut être source de confusion dans le rendu de la traduction. Ce ne sont là que quelques situations qui soulignent le devoir qu'a le chercheur de permettre le meilleur accès possible à l'enregistrement de la performance telle qu'elle a été produite.

Il me semble inadmissible d'envisager un nouvel enregistrement artificiel comme on pourrait le faire en faisant relire la transcription retenue par le transcripteur par un locuteur de la langue. Cela me semble fausser la véracité du témoignage et, si on a recours à un tel procédé, il doit alors être spécifié comme tel en précisant bien qu'il ne s'agit pas d'un enregistrement de la performance dans sa situation originale.

C'est là toute l'importance à accorder à la rédaction de ce qu'on appelle les métadonnées, c'est-à-dire de toutes les informations qui donnent à l'auditeur-lecteur du conte mis en ligne les éléments pour pouvoir apprécier tout le contexte de la performance initiale.

En conclusion

Les techniques modernes constituent donc une formidable avancée dans l'archivage de ces données originales. Cet archivage devient un élément dynamique qui permet aux locuteurs d'une culture de retrouver leur traditions orales en les entendant comme elles étaient dites. Elles permettent de plus à ceux qui ne sont pas locuteurs de la langue source d'écouter la performance en suivant la traduction. Enfin, pour le locuteur, comme pour le non locuteur de la langue

source, cette présentation peut donner accès, lorsque l'analyse a été poussée, à la structuration grammaticale de la langue. La mise à plat de cette analyse donne à tous ceux qui y ont accès la possibilité de se faire leur propre jugement et d'exercer leur esprit critique, car ils ont tous les atouts en main.

Références

- JACQUESSON, François, 2007, La coordination, instrument décisif de l'émergence de la prose : une discription historique, in Rousseau, A., L. Begioni, N. Quayle et D. Roulland (éds.), *La coordination*, Rennes, PUR, pp. 371-404.
- ROULON Paulette, 1977, *Wanto et l'origine des choses...*, *contes d'origine et autres contes gbaya kara*, Collection Fleuve et flamme bilingue, CILF-EDICEF, Paris, 142 p., 9 illustrations et page de couverture.
- ROULON-DOKO Paulette, 1992, Traduction d'un conte gbaya "Les deux filles et la balançoire", *L'immature*, n°4, septembre-novembre, pp. 46-47.