

VOCABULAIRE DU ROMAN KABYLE (1981-1995) UNE ÉTUDE QUANTITATIVE ¹

par

Sinikka Loikkanen

Une étude quantitative sur le vocabulaire du roman kabyle peut, à première vue, sembler étrange car l'existence du roman kabyle paraît être encore modeste pour le grand public. Pourtant, il y en a déjà plusieurs romans publiés dans cette langue.

L'histoire du roman écrit kabyle n'a pas encore atteint ses vingt ans. Le premier roman fut publié en 1981 et le plus récent, inclus dans cette étude, en 1995. Cela fait en moyenne un roman tous les deux ans et demi. Ce n'est pas beaucoup pour une société qui a une tradition littéraire écrite vieille de près d'un siècle (*cf.* Chaker 1989 : 19-23 ; 1992). La naissance de ce genre littéraire peut être comprise comme un aspect de la revendication culturelle et identitaire qui a bousculé le monde berbère à partir des années quatre-vingt. Cette réalité elle-même se reflète dans la thématique de ces romans.

Le point de départ de cette étude était totalement pratique. Voulant apprendre la langue kabyle j'ai voulu savoir quels sont les mots et les verbes les plus fréquemment utilisés en cette langue. Pour réaliser cette étude, on a considéré le vocabulaire du roman kabyle comme une unité. On a voulu savoir de quoi ce vocabulaire est composé. Ainsi, on n'a pas tenu compte de la variation sociale ou locale des mots, ni non plus de leur origine, ni de leur emploi dans la phrase, ou du style. Une étude quantitative de ce type est, à ma connaissance, la première étude sur le vocabulaire romanesque kabyle.

L'écriture et la notation utilisées dans les romans étudiés sont très variables et causent certaines difficultés. La manière d'écrire diffère non seulement entre les romans mais aussi au sein d'un même roman. Les écrivains ont utilisé tantôt la notation phonologique, tantôt la notation phonétique avec ou sans assimilation phonétique ou tiret entre le nom ou le verbe et leurs affixes. Ainsi, on peut trouver dans les textes, par exemple, pour l'expression « à côté de » les

1. Cette étude est la synthèse d'un mémoire de DEA (*Vocabulaire du roman kabyle*) soutenu à l'INALCO en 1998 sous la direction de S. Chaker.

notations *t tama*, *ttama*, *t-tama*, pour les mots « vêtements » *icettiden* ou *icettiden*, « avec lui » *yides* ou *yid-es* et pour la « malédiction » *dazwessu* ou *daawassu*. On écrit aussi *yalwa* pour *yal wa* « tous ceux » ou *sekra* pour *s kra* « avec un peu, par tous ceux ». Il y a donc des variations à plusieurs niveaux.

Une telle variation de la transcription du texte indique qu'il existe encore une grande hésitation parmi les auteurs quant à la manière d'écrire les mots ou les expressions kabyles. Dans ces conditions, on peut conclure que le passage de l'oralité à l'écriture n'est pas encore complètement achevé. Sans exagérer, on peut dire que chaque écrivain applique ses règles d'écriture et que celles-ci sont variables même pour un même auteur. Cela rappelle le cas de l'ancien français où les textes du XVII^e siècle, surtout avant les 1660, ont de telles variations orthographiques qu'« il serait assez vain de faire entrer dans une statistique des variations qui n'ont souvent d'autre cause que la fantaisie ou l'ignorance d'un imprimeur » ce qui « crée évidemment des difficultés supplémentaires » (Muller 1979 : 125n). Pour éviter des difficultés dues à l'instabilité de la notation, on a dû reconstruire un certain métaniveau. Afin de le former, on a recherché les équivalences des mots, puis, par la lemmatisation on a regroupé les mots sous les vocables correspondants, ce qui permet de passer au vocabulaire du texte. La notation employée dans cette lemmatisation se base fondamentalement, avec quelques exceptions, sur celle utilisée dans le *Dictionnaire kabyle-français* de J.-M. Dallet (1982).

I. CORPUS

Le corpus comporte les six premiers romans écrits en kabyle qui sont :

- Rachid Aliche, *Asfel*. Mussidan : Fédérop, 1981. 139 p.
- Saïd Sadi, *Askuti*. Paris : Imedyazen, 1983 / Alger : Asalu, 1991. 184 p.
- Rachid Aliche, *Faffa. Igujen irgazen ur ttrun*. Mussidan : Fédérop, 1986. 142 p.
- Aëmer Mezdad, *Iđ d wass*. Alger : Asalu, Azar, 1990. 182 p.
- Eëmer Whemza, *Si tedyant yer tayed....* Lyon : Editions Berbères, 1994. 199 p.
- Salem Zenia, *Tafrara. Aurore*. Paris : L'Harmattan, AWAL, 1995. 181 p.

Bien que ces romans ne soient pas quantitativement comparables entre eux, *i.e.* ils diffèrent par le nombre de pages et de mots, on les a traités et considérés de la même façon. On a étudié tout le texte narratif de chaque roman. On a exclu uniquement les numéros et les rubriques éventuels des chapitres, les numéros

des pages, les notes en bas de page qui expliquent les mots, les lexiques, les préfaces ainsi que les phrases écrites entièrement en langue étrangère.

On a considéré ici comme une langue étrangère ce qui n'est pas écrit en kabyle, *i.e.* dans ce contexte, ce qui est écrit en arabe ou en français. Mais, dans quelle mesure peut-on compter ces dernières comme des langues étrangères dans le contexte maghrébin de nos jours? Est-ce que l'arabe est une langue étrangère? Est-ce que le français est une langue plus étrangère? Les sociétés ainsi que les cultures maghrébines sont profondément imprégnées de ces deux dimensions, ce qui a des conséquences aussi sur la langue parlée localement. Il existe des emprunts de part et d'autre, et il n'est guère judicieux de les séparer et supprimer les éléments dits étrangers. La délimitation entre les éléments « purs » et « étrangers » est souvent comme un trait tracé dans l'eau. L'élaboration des néologismes est bien compréhensible pour les mots et les expressions qui font défaut, mais la traduction et la suppression systématiques des anciens emprunts n'est pas toujours justifiable. Alors, si un kabyle dit : « *hemdullilah!* » ou dit qu'il va au *ssuq*, est-ce qu'il faut considérer qu'il parle l'arabe et pas le kabyle! Maintenant, on peut trouver écrit dans le corpus étudiés des néologismes *illu* pour le dieu et *ulzuz* pour le marché.

L'étude se base sur la forme. Si on relève dans le texte par exemple le mot *garçon*, on le classe comme mot étranger. Mais si, par contre, on trouve *agarsun*, on le considère comme un mot kabylisé et il sera classé avec les autres mots normaux kabyles. La délimitation entre les mots français et les mots empruntés au français kabylisés est assez facile. Mais avec l'arabe ceci n'est pas toujours aussi clair. Selon une étude, 38 % du lexique kabyle est emprunté à l'arabe et dans certains domaines (vie religieuse et politique) le vocabulaire est presque totalement arabisé (*cf.* Chaker 1984 : 116-229). En plus, ces emprunts sont bien intégrés dans la structure de la langue kabyle. Mais, étant donné que le point de départ du dépouillement est la forme, on a dû dans certains cas classer les mots venant de l'arabe parmi les mots étrangers. Il s'agit des cas où l'écrivain lui-même les a considérés comme tels en utilisant l'arabe comme traduction et quand il les a placés entre crochets ou parenthèses, comme par exemple dans les phrases suivantes dans le roman *Si tedyant yer tayed...* de Emeṛ Wḥemza (1994).

Awal amezwaru n Naṣeṛ : *Lžžayer galzeum anney galzeum awen* [tkeffi-na wa tkeffi-kum]! (p. 11).

Adrar nni ymuger, mazal-t ar ass-a yeṭtef isem-is : Adrar [*Ġebel*] Tariq, zelqen-t kan çitūḥ (p. 54).

Ma tenniḍ-as dayen : S yisem n Yillu aṛeḥḥam yesserḥamen [*Bismi Llahi rrahmani rrahim*] ad k-in yerr : (p. 65)...

Ibn Xeldun yenna yakan di tmiḍi tis mraw kuḥ (14) : Ma teṭuzeṛreb texṛeb (ida eurribet xurribet) (p. 104).

Yuzzel wawal-is, ur yesqewqiw, ur yesleqliđ, ur yeqqar: Yenna-d (qal) (p. 114).

On a donc ignoré dans les calculs les mots *tkeffi-na wa tkeffi-kum*, *ğebel*, *bismi llahi rrahmani rrahim*, *ida zurribet xurribet* et *qal* étant donné que ce sont des traductions.

II. DE L'ÉTUDE STATISTIQUE LEXICALE

Dans la recherche statistique lexicale, on compte le nombre de mots qu'un texte ou un corpus contient. Celle-ci inclut deux phases : d'abord, on segmente le texte, c'est-à-dire, on opère un découpage du texte et on compte le nombre de mots – d'occurrences – qu'il contient. Puis, on opère un classement des éléments ainsi obtenus et on calcule le nombre et la fréquence de mots différents – des vocables – en reconstituant ainsi le vocabulaire de ce texte étudié. (Pour la méthodologie et les notions, cf. par exemple Guiraud 1954; Muller 1977; 1979: 125-143, 197-206; 1985: 125-136.)

Un mot est défini généralement par une unité graphique séparée des autres par des blancs ou par des signes de ponctuation. Ceci est un idéal. Mais souvent l'unité graphique contient plusieurs morphèmes (par exemple en français *aux*, *lesquelles*; en kabyle *mulac* «si non», *yides* «avec lui») ou un mot peut être constitué par plusieurs unités graphiques (par exemple *pomme de terre*, *chemin de fer*; en kabyle *adrar ufud* «le tibia»). Pour compter les occurrences et puis, pour regrouper les vocables, il faut élaborer les critères avec lesquels on traite le texte et on reconstitue le lexique. Le kabyle semble différer du français en ce que dans le premier il n'existe pas – si on suit la notation proposée dans le Dallet (1982) – beaucoup de mots composés écrits ensemble. Selon le Dallet, nous écrivons *m^a ulac* et *yid-es* au lieu de les écrire en un seul mot. En considérant l'apostrophe et le tiret comme des signes de ponctuation, il est facile de définir les règles de séparation.

III. DÉCOUPAGE DU TEXTE

Le découpage du texte en mots à base des unités graphiques s'est avéré insuffisant à cause de la notation utilisée dans les romans kabyles étudiés. Parfois, on a écrit plusieurs éléments différents ensemble, comme par exemple les affixes personnels avec les prépositions. Pour les traiter et les compter isolément, on a dû les séparer l'un de l'autre. Parfois, on a dû réunir et/ou

composer autrement les éléments distincts ou ceux séparés l'un de l'autre d'une façon différente de celle utilisée dans les manuels du kabyle, comme par exemple dans les cas suivants où les unités sont mal identifiées :

yell-is → *yell-i | s*
gmat-sen → *gma | t-sen*
gara-neγ → *gar | a-neγ*
aqla-k → *aql | a-k*

Pour compter les mots, on a extrait les occurrences des différents romans d'après les mêmes principes suivant la grammaire kabyle pour avoir un recensement aussi uniforme que possible. On a retenu les critères suivants de découpage du texte :

Les noms

Les noms, les substantifs sont tous traités isolément. On leur a enlevé tous les éléments périphériques éventuels, comme les particules prédicatives, les prépositions, les affixes personnels ainsi que les suffixes démonstratifs. Ainsi, on a obtenu au minimum deux occurrences, par exemple :

les particules prédicatives : *t-tamγart* « c'est une vieille » → *t | tamγart*
 les prépositions : *t-temγart* « avec une vieille » → *t | temγart*
 les affixes personnels : *babas* « son père » → *baba | s*
 les suffixes démonstratifs : *argaz-agi* « cet homme-ci » → *argaz | agi*.

On a ainsi traité aussi l'expression *iman-iw*, *iman-is* « moi-même », « lui-même », bien qu'*iman* ne soit utilisé qu'avec des affixes personnels.

Les noms composés qui ont un signifié unique, mais qui sont écrits séparément, comme *adrar ufud* « (le) tibia », sont décomposés. Ici, ils forment autant d'occurrences que de mots dont ils sont composés – *adrar ufud* constitue donc deux occurrences qui sont séparées et traitées isolément. Mais, les noms composés commençant par *bu*, qui sont écrits ensemble avec ou sans tiret, comme *tabuzeggaγt* « rougeole » ou *bu-yefrax* « maladie du muguet », ne sont pas séparés.

On n'a pas segmenté les noms dérivés, comme les noms d'agents, les noms d'instruments ou les noms verbaux.

Les verbes

Tous les éléments, comme les particules d'orientations, celles de l'aoriste ou

de l'intensif ainsi que les affixes personnels, attachés aux verbes avec ou sans tirets, sont ainsi segmentés, par exemple :

kecmen-d «ils sont entrés» constitue deux occurrences, ainsi que
annawed «nous arriverons», deux occurrences, et
 (a) *k-d-iniγ* «je te dirai», trois occurrences.

Les prépositions

Il existe une quantité de prépositions qu'on utilise avec des affixes personnels en les écrivant en un seul mot avec ceux-ci avec ou sans tiret. Il y a onze prépositions de ce type: *ddaw* «sous», *deffir* «derrière», *deg* «dans», *fell* «sur», *gar* «entre», *nnig* «au-dessus de», *γer* «vers», *γur* «chez», *yid* «avec», *yis* «avec» et *zdat* «devant». On les a toutes séparées de leur affixes et comptées séparément, comme par exemple :

γures / *γur-es* «chez lui» → *γur* | *es*
yidi «avec moi» → *yid* | *i*
zdatzen «devant eux» → *zdat* | *sen*.

Les affixes personnels

Dans tous les cas où l'affixe personnel est attaché avec ou sans tiret avec n'importe quel autre élément – nom, verbe ou préposition – on l'a séparé de celui-ci, par exemple :

nom	:	<i>axxam-is</i> «sa maison»	→	<i>axxam</i> <i>is</i>
		<i>yemmas</i> «sa mère»	→	<i>yemma</i> <i>s</i>
verbe	:	<i>inna-yas</i> «il lui a dit»	→	<i>inna</i> <i>yas</i>
préposition	:	<i>yides</i> «avec lui»	→	<i>yid</i> <i>es</i>
		<i>ack-it</i> «qu'il est bon»	→	<i>ack</i> <i>it</i>
		<i>aqliyi</i> «me voici»	→	<i>aql</i> <i>iyi</i> .

Les mots grammaticaux

Ceux qui se sont réduits et soudés à l'élément suivant par une apostrophe, comme par exemple *ma* et *mi* dans *m'ulac*, *m'ur* et *m'ara*, sont tous séparés.

Les numéraux

Ils suivent aussi la règle générale ; les numéros écrits en lettres ont été séparés des autres éléments. Par exemple *xemsin alef* forment deux occurrences. Quant aux ordinaux, que l'on écrit tantôt en un mot, tantôt à part avec un élément *wis / tis* + un nombre cardinal, ils ont tous été séparés. Mais quand il s'agit d'un nom propre ou de celui d'un souverain suivi d'un numéro, par exemple *Massinissa II*, on les compte ensemble comme une seule unité.

En plus des cas cités ci-dessus, il existe de nombreux mots écrits ensemble mais qu'on ne sépare pas ici. Cela concerne les mots qui ont un signifié figé et bien défini dans le langage. Cette règle concerne entre autres certains mots interrogatifs, conjonctions et expressions figées, comme par exemple : *acuγef?* « pourquoi? », *acimi?* « pourquoi? », *asmi* « lorsque », *awufan* « puisse », *iwumi?* « à qui? », *ulamek* « il n'a pas de moyen », *wissen?* « qui sait? ».

IV. REGROUPEMENT DES VOCABLES

Après que tous les mots aient été dépouillés et comptés, on les a regroupés sous les lemmes. Par la lemmatisation, qu'on peut définir comme une chasse aux classes d'équivalence des mots, on peut passer des mots aux vocables correspondants, *i.e.* aux formes lexicales des mots. La lemmatisation se base sur les critères de regroupement suivants :

– Les mots sont tous regroupés selon leur sens et selon leur classe grammaticale. Dans cette phase, on a réuni sous le lemme correspondant toutes les notations possibles qu'un mot a pu avoir dans les textes étudiés, par exemple le lemme *tawwurt*, nom féminin « (la) porte », englobe les formes variantes comme *tawwurt* et *tabburt* avec toutes leurs formes déclinées, le lemme *s*, préposition « avec, au moyen de » englobe les formes *s*, *es*, *yis*, *yiss*, et le lemme *ad*, particule de l'aoriste, englobe les formes variantes, comme *a*, *an*, *at*, *aṭ*.

– Les homographes sont regroupés sous les lemmes convenables à chacun par le sens et par la classification grammaticale. Par exemple le mot comme *yelli* est reparti entre trois lemmes : *yelli*, nom féminin « (ma) fille », *ili*, verbe « être » et *lli*, verbe « ouvrir ».

– Dans quelques rares cas, un lemme peut englober plus qu'un mot. Ce problème ne concerne que quelques homographes qui ont la même classe grammaticale et totalement la même conjugaison bien qu'ils aient des sens différents, comme par exemple le verbe *ssiwel* « 1) appeler, dire, 2) faire bouillir » et la particule *ara* 1) particule de l'aoriste, 2) particule complétive de négation.

– La seule exception est le lemme *d* qui contient toutes les attestations de *d*, comme la particule présentative et la préposition ainsi que les particules d’orientation et disjonctives. On a retenu cette solution parce que la distinction entre les diverses unités n’est pas toujours très claire.

V. RÉSULTATS

Le corpus des six romans traités contient dans sa totalité 234 238 occurrences (mots) parmi lesquelles on a identifié 8 088 vocables (lemmes), ou mots différents. Les mots et les lemmes repartissent entre les romans étudiés de la manière suivante :

	<i>Asfel</i>	<i>Askuti</i>	<i>Faffa</i>	<i>Id d wass</i>	<i>Si tedyant yer tayed</i>	<i>Tafrara</i>
Lemmes :	2 057	2 400	3 015	3 011	3 440	3 718
Mots :	14 578	31 508	23 768	47 906	62 763	53 715

La répartition des lemmes distribués en fonction de la fréquence nous révèle le degré de compréhension, c’est-à-dire le nombre relatif cumulé de lemmes nécessaires à la compréhension du texte :

Degré de compréhension (%)	Nombre cumulé de lemmes
11,66	2
20,66	6
30,30	14
40,42	30
50,05	58
60,21	112
70,02	220
80,15	500
90,15	1 331
100	8 088

Les mots dont la fréquence (F) est la plus haute sont des particules, des prépositions et des affixes. Les premiers deux verbes ne se manifestent qu’aux

14^e (*ili*) et 15^e (*ini*) rangs, quant aux deux premiers substantifs, ils ne se retrouvent qu'aux rangs 29 (*ass*) et 42 (*iman*) :

Rang	Lemme/Classe	F
1	d <i>part</i>	18 670
2	s <i>aff</i>	8 635
3	ad <i>part</i>	6 505
4	deg <i>prép</i>	5 661
5	n <i>prép</i>	4 746
6	t <i>aff</i>	4 174
7	i <i>pron</i>	3 478
8	wer <i>part</i>	3 289
9	s <i>prép</i>	3 011
10	sen <i>aff</i>	2 835
11	i <i>prép</i>	2 801
12	ara <i>part</i>	2 623
13	γer <i>prép</i>	2 463
14	ili <i>v</i>	2 078
15	ini <i>v</i>	1 886
	:	:
29	ass <i>n m</i>	1 194
42	iman <i>n m</i>	772

La distribution des 8 088 lemmes selon la classe grammaticale (voir ci-dessous) nous montre que les fréquences des substantifs et des verbes sont pratiquement égales au niveau du texte ($\simeq 20\%$), c'est-à-dire qu'on les a utilisés autant de fois dans le corpus, mais au niveau des lemmes les verbes ne représentent plus que 30,23 % alors que les substantifs plus de la moitié, 51,59 %. Dans le tableau ci-dessous, on donne d'abord les fréquences dans les textes, et puis, au niveau des lemmes. En plus des fréquences absolues, c'est-à-dire les occurrences relevées dans le corpus, on donne les fréquences relatives en pourcentage (%):

Distribution des classes grammaticales

Classe	Fréquence dans le texte		Fréquence par lemmes	
	F	%	F	%
Adjectifs	4 651	1,99	478	5,91
Adverbes	15 099	6,45	123	1,52
Affixes et pronoms	38 220	16,32	73	0,90
Conjonctions	9 881	4,22	46	0,57
Démonstratifs	3 926	1,68	16	0,20
Expressions et interjections	1 349	0,58	120	1,48
Mots étrangers	153	0,07	68	0,84
Noms propres	4 580	1,96	465	5,75
Numéraux	2 243	0,96	41	0,51
Particules	32 527	13,89	7	0,09
Prépositions	24 980	10,66	26	0,32
Substantifs	47 759	20,39	4 173	51,59
Verbes	48 863	20,86	2 445	30,23
Mots non-classés ²	7	0,00	7	0,09
Σ =	234 238	100	8 088	100

2. Les mots non-classés sont des mots qu'on n'a pas pu identifier à cause de la notation ou bien parce qu'ils sont des créations d'auteur.

Cette étude quantitative nous dévoile pour la première fois statistiquement quels sont les adjectifs, les substantifs et les verbes les plus fréquemment utilisés. Mais, il faut se rappeler que la thématique des romans étudiés est limitée. Les quinze en tête sont regroupés dans le tableau ci-dessous :

	Adjectifs		Substantifs		Verbes	
1	amezyan	« petit »	ass	« jour »	ili	« être »
2	amezwaru	« premier »	iman	« personne »	ini	« dire »
3	ažžayri	« algérien »	amdan	« homme »	uɣal	« venir »
4	amaziy	« berbère »	abrid	« chemin »	rnu	« ajouter »
5	aneggaru	« dernier »	tamurt	« pays »	awi	« apporter »
6	ameqqran	« grand »	awal	« mot »	ffeɣ	« sortir »
7	aqbayli	« kabyle »	axxam	« maison »	af	« trouver »
8	aɣrab	« arabe »	taddart	« village »	err	« rendre »
9	afransis	« français »	tiɛ	« œil »	qqim	« rester »
10	ir	« mauvais »	afus	« main »	wali	« regarder »
11	amcum	« méchant »	mmi	« fils »	eğğ	« laisser »
12	aberkan	« noir »	argaz	« homme »	kcem	« entrer »
13	amellal	« blanc »	tikkelt	« fois »	ddu	« aller »
14	amenzu	« premier »	ul	« cœur »	zer	« voir »
15	amecɛuɣ	« petit »	tameɛttut	« femme »	bɣu	« vouloir »

VI. CONCLUSION

Cette étude porte sur le vocabulaire du roman kabyle. Le corpus est constitué de six romans écrits et publiés en langue kabyle entre 1981 et 1995. Elle propose les premières données statistiques concernant le vocabulaire du roman kabyle.

La langue et l'écriture dans les romans sont très variables. Cette étude a révélé l'instabilité dans l'écriture du kabyle, notamment les difficultés concernant l'emphatisation, l'assimilation, l'enchaînement des mots, le choix des lettres et l'utilisation du tiret. Cette variabilité dans l'écriture, qui est encore sans convention normative et sans standardisation, rend très difficile la recherche automatique sur ordinateur. Le traitement automatique des textes demande une notation univoque ; on ne peut pas écrire une règle spéciale pour chaque cas.

La langue utilisée vient des quatre coins de la Kabylie ainsi que du monde autour de celle-ci. Dans cette langue écrite, il y a un trait frappant par rapport à la langue parlée, c'est l'utilisation intensive des néologismes berbères. Leur sens n'est pas toujours expliqué dans le texte. D'après les premières estimations, on a relevé un peu plus de cinq cent néologismes, ce qui fait environ 6 à

7 pour cent des lemmes du corpus. Ces chiffres demandent encore une étude plus approfondie.

Nous avons obtenu quatre listes des mots et cinq tableaux statistiques. Les six romans se composent de 234 238 mots qui se regroupent en 8 088 lemmes différents. On a pu établir que 58 mots différents suffisent pour comprendre 50 % du texte du corpus, et que 1 331 mots permettent d'en comprendre 90 %. Pour comprendre le reste, il nous faut encore 6 757 mots. Sur la base de ces résultats nous pouvons, pour la première fois, former un vocabulaire de base fondé statistiquement.

SINIKKA LOIKKANEN

RÉFÉRENCES

- CHAKER S., *Textes en linguistique berbère*. Introduction au domaine berbère. Paris, CNRS, 1984.
- 1989: *Berbères aujourd'hui*. Paris: Éditions L'Harmattan.
 - 1992: La naissance d'une littérature écrite. Le cas berbère (Kabylie). *Bulletin des Études africaines* 17-18: 7-21.
- DALLET J.-M., *Dictionnaire kabyle français*. Parler des At Mangellat, Algérie. Paris: Selif, 1982.
- GUIRAUD P., *Les caractères statistiques du vocabulaire*. Essai de méthodologie. Paris: P.U.F., 1954.
- MULLER Ch., *Principes et méthodes de statistique lexicale*. Paris: Hachette, 1977.
- 1979: *Langue française et linguistique quantitative*. Recueil d'articles. Genève: Éditions Slatkine.
 - 1985: *Langue française, linguistique quantitative, informatique*. Recueil d'articles, 1980-1984. Genève-Paris: Slatkine-Champion.