

NÉOLOGISMES ET TERMES D'EMPRUNT DANS LE FRANÇAIS DES AFFAIRES

Une étude de la terminologie en usage dans les offres d'emploi de la presse française

Jean SOUBRIER

INSA de Lyon, France

Depuis 1972, date à laquelle fut créée au ministère de l'Économie et des Finances la toute première commission de terminologie, pas moins de six arrêtés¹ relatifs à la terminologie économique et financière ont été adoptés par les gouvernements successifs. Ces arrêtés, qui rendent obligatoires ou fortement conseillés plus de 250 termes dans les documents officiels ont pour but avoué de faire disparaître de la langue des affaires bon nombre de termes empruntés à l'anglo-américain. Largement relayés par la presse et par divers organismes publics ou privés, les travaux des terminologues ont bien sûr contribué à l'enrichissement du français avec un certain succès. Dans quelle entreprise n'utilise-t-on pas quotidiennement aujourd'hui des mots comme *crédit-bail*, *logiciel*, *départ-usine* et *savoir-faire* au lieu de *leasing*, *software*, *ex-works* et *know-how*, qui étaient les seuls termes en usage voici vingt ans ? Toutefois ces réussites terminologiques ne doivent pas occulter la réalité des faits. Les termes de la langue des affaires empruntés à l'anglo-américain demeurent toujours très prisés dans les milieux professionnels. Il n'est pour s'en convaincre que de lire les titres des ouvrages exposés au rayon «entreprise», d'une librairie ou, plus simplement encore, de parcourir les pages économiques et commerciales de la presse française. Pourtant, en l'absence de données chiffrées, notre appréciation du phénomène ne peut être que subjective et c'est pourquoi nous avons décidé d'entreprendre une étude sur la terminologie en usage dans les offres d'emploi publiées dans la presse pour cadres. Ce choix s'est imposé à nous pour différentes raisons alliant des exigences d'ordre pratique à des considérations plus théoriques.

Rédigées à l'intention du candidat idéal, les offres d'emploi concentrent en quelques mots l'image que l'entreprise veut donner d'elle-même. Elles constituent à ce titre un terrain d'observation privilégié. Par ailleurs il faut souligner que le support que nous avons retenu (presse généraliste) publie des annonces couvrant de très nombreux secteurs du monde professionnel et cela nous a permis d'avoir une vision assez large du phénomène. Enfin, et ce n'est pas la moindre de nos préoccupations, les offres d'emploi, publiées selon un rythme régulier et facilement repérables au coeur d'un magazine simplifient grandement la constitution d'un corpus.

¹ Arrêtés du : 29 novembre 1973, 18 février 1987, 6 janvier 1989, 11 janvier 1990, 30 septembre 1991, 11 février 1993.

Cette étude, qui ne porte que sur un domaine très circonscrit de la langue des affaires, s'est fixé un triple objectif :

- Repérer et quantifier les emprunts à l'anglo-américain dans les offres d'emploi.
- Vérifier l'hypothèse selon laquelle les néologismes « officiels » ne connaissent qu'une très faible diffusion auprès des utilisateurs professionnels.
- Analyser, dans une perspective linguistique mais également culturelle, les principaux facteurs qui freinent la diffusion des néologismes dans le monde des affaires et qui contribuent de ce fait à l'utilisation plus ou moins généralisée d'une terminologie d'emprunt.

PRÉSENTATION DE L'ÉTUDE

Afin de constituer notre corpus, nous avons choisi le supplément « Réussir » publié chaque semaine par le magazine *L'Express*. Ce supplément détachable, qui contient en moyenne plus d'une centaine d'offres d'emploi, présente en outre l'avantage d'appartenir à un très vaste service d'offres d'emploi accessible sur Internet auquel nous avons eu recours ultérieurement. Nous avons procédé au dépouillement de 12 numéros successifs (n° 2389 au n° 2400 entre le 17 avril et le 3 juillet 1997) et étudié 1473 offres d'emploi rédigées en français (les quelques offres d'emploi rédigées en anglais n'ont bien sûr pas été prises en compte).

MÉTHODOLOGIE

La première question qui s'est posée à nous fut celle du repérage des emprunts à l'anglais dans notre corpus. Sans rentrer dans le détail de la problématique du repérage qui reste fondamentalement liée à toute activité terminologique² nous résumerons ici les choix que nous avons faits :

« Quand on parle d'emprunt linguistique, c'est d'abord aux mots que l'on pense »³ et bien naturellement nous avons relevé tous les termes dont la graphie trahissait à l'évidence l'importation lexicématique, c'est-à-dire « l'emprunt de signe complet à la fois comme expression et comme contenu »⁴ par exemple : *leader*, *start up*, *sourcing*, *challenge*, etc.

Nous avons aussi relevé quelques graphies hybrides manifestement situées entre anglais et français comme *marchandising*, *mixt*, *télémarketing*, *entreprenarial* ou encore nées d'une inversion de deux termes comme *mix marketing*.

« C'est avoir une idée bien superficielle de l'anglicisme que de ne le considérer que comme un mot emprunté. D'une part nous n'empruntons pas que des mots, et d'autre part il ne s'agit pas toujours du signe complet mais de son expression ou de son contenu

² CHUKWU, Uzoma (1993): *Le repérage des termes dans un corpus bilingue anglais/français*. Thèse de Doctorat dirigée par Ph. Thoiron, Faculté des langues, Université Lumière Lyon 2. 545p.

³ DEROY, Louis (1956): *L'emprunt linguistique*, Bruxelles, Sté. d'édition Les belles lettres, p.67.

⁴ HÖFLER, Manfred (1982) : *Dictionnaire des anglicismes*, Paris, Larousse, p.VI.

seulement»⁵. Cette remarque de Josette Rey-Debove nous a incité à relever, en marge de notre liste principale, une autre série d'emprunts traditionnellement décrits comme des anglicismes clandestins ou anglicismes sémantiques. Nous avons ainsi réuni des mots (cf. Annexe B) ayant déjà fait l'objet d'un débat ancien (ex : *contrôle*, *opportunité*, *approche*, *attractif*) et répertoriés dans de nombreux ouvrages⁶, mais aussi des mots d'introduction beaucoup plus récente qui témoignent de l'évolution de notre langue (*developpeur* = responsable de l'accroissement du volume d'affaires, *junior* = débutant, *senior* = confirmé, *être en charge* = être responsable de, etc.). Toutefois, nous avons pris le parti de ne pas comptabiliser ces emprunts sémantiques dans notre liste principale dans la mesure où il n'y a pas véritablement d'importation lexicématique mais seulement importation de structure sémantique.

Nous n'avons pas non plus retenu dans notre liste principale les emprunts à l'anglais qui ont été «officiellement» francisés après consultation des diverses commissions de terminologie. Ces emprunts officiels sont des termes dont la forme initiale est parfaitement compatible avec notre système morpho-phonologique et qui, de ce fait, ne posent pas de réel problème d'assimilation (ex. : *management*, *maintenance* etc.).

Nous n'avons pas davantage retenu les dérivations d'emprunts totalement francisées dans leur forme (ex. : «vous *managerez* un centre de profit...» ou «vos qualités *managériales*...») même si celles-ci sont parfaitement contestables comme, par exemple, le terme *profitabilité*.

RÉSULTATS

Au terme de notre phase de repérage nous avons pu établir une liste de 142 emprunts lexicaux à l'anglais (mots simples ou mots composés) relevés dans 1473 offres d'emploi. Cette liste que nous présentons sous forme de tableau en Annexe A, appelle un certain nombre de commentaires.

L'élément le plus important de cette liste figure dans les chiffres placés en regard de chaque entrée. Ces chiffres indiquent les occurrences enregistrées pour chacun des emprunts. Ces occurrences varient de 542 à 1. Mais le fait le plus surprenant est que deux mots, *marketing* et *leader*, totalisent à eux seuls près de 65 % des occurrences relevées.

Inversement, on constate que plus de la moitié des emprunts relevés (83 sur 142) n'ont qu'une seule occurrence dans notre corpus.

Par ailleurs nous avons cherché à quantifier la fréquence moyenne d'emploi des termes d'emprunts dans l'ensemble de notre corpus.

⁵ REY-DEBOVE, J. et G. GAGNON (1980) : *Le dictionnaire des anglicismes*, Paris, Le Robert, p. IX.

⁶ REY-DEBOVE, J. et G. GAGNON (1980) : *Le dictionnaire des anglicismes*, Paris, Le Robert. WALTER, Henriette et Gérard WALTER (1991) : *Dictionnaire des mots d'origine étrangère*, Paris, Larousse.

WALTER, Henriette (1997) : *L'aventure des mots venus d'ailleurs*, Paris, Robert Laffont.

N'ayant pas eu la possibilité matérielle de procéder à un traitement informatique complet (utilisation d'un scanner, transcription en fichier texte et comptage automatique des mots), nous avons dû procéder à une estimation moyenne à partir d'un comptage manuel portant sur 147 annonces sélectionnées de manière aléatoire, ce qui correspond environ à 10 % de l'ensemble des offres dépouillées. Ce comptage, qui n'a porté que sur le seul texte des annonces et pas sur les raisons sociales et les adresses des annonceurs, nous a permis d'évaluer la taille de notre corpus à 207 000 mots. Rapproché du nombre total d'occurrences (1619) cumulées (c-à-d. calculé en prenant en compte tous les mots à l'intérieur d'une entrée composée) nous obtenons un chiffre de 1 emprunt tous les 128 mots ce qui, traduit en pourcentage, nous donne une proportion moyenne d'emprunts égale à 0,78 % sur l'ensemble du corpus.

Ce chiffre, qu'il convient bien sûr de considérer comme une simple estimation, est néanmoins loin de correspondre au déferlement d'américanismes que l'on nous annonce régulièrement depuis le célèbre pamphlet d'Etienneble⁷. Il semble même à peine plus élevé que la proportion des anglicismes relevés il y a vingt ans dans le journal *Le Monde* et que l'on estimait à 0,6 % (soit un emprunt tous les 166 mots)⁸.

Si l'on considère de surcroît que la langue des affaires compte parmi les langues de spécialité dont la terminologie a été le plus influencée par les emprunts à l'américain nous pourrions même trouver dans ces chiffres un motif de satisfaction quant à la capacité du français à ne pas dépasser un certain «seuil de tolérance» dans le volume de ses importations.

Plus encore, le fait que seuls 11 termes de notre liste enregistrent des occurrences supérieures à 10 nous invite à relativiser davantage la portée du phénomène ainsi que le faisait déjà Jacques Cellard en écrivant : «Dans l'ensemble, il en est de l'investissement de notre langue par l'anglais ce qu'il en est, révérence parler, des figurants du Châtelet : en en faisant passer et repasser une douzaine sur scène, on donne à bon compte l'illusion d'une armée. Peu ou prou on en revient toujours aux mêmes...»⁹.

Étant ainsi parvenu à une estimation chiffrée de l'anglicisation de la langue des offres d'emploi, nous avons ensuite cherché à vérifier l'hypothèse selon laquelle les néologismes proposés par les commissions de terminologie n'étaient que très rarement employés par les professionnels du commerce et de l'industrie. Naturellement nous avons rencontré certains de ces néologismes au cours de la première partie de notre travail mais dans des proportions si réduites qu'il nous apparaissait alors dérisoire d'en faire état. Nous avons donc choisi d'adopter une démarche inverse et d'utiliser à des fins terminologiques les ressources d'Internet.

Ainsi nous avons eu recours au serveur «Cadres on line» (<http://www.cadresonline.com>) qui regroupe les offres d'emploi publiées dans 17 titres de

⁷ ETIEMBLE, René (1964) : *Parlez-vous franglais ?*, Paris, Gallimard.

⁸ KLEIN, Virginia (1977) : *A study of Franglais in twelve issues of Le Monde*, (Mémoire de maîtrise dirigé par Guy J. Forgeue), Université de Paris III.

⁹ CELLARD, Jacques (1979) : *La vie du langage*, coll. «L'ordre des mots», Paris, Le Robert, p.9.

la presse française¹⁰ (dont en autres *L'Express*) et dans deux titres de la presse britannique¹¹.

Mis à jour deux fois par mois, ce serveur présente l'avantage de posséder un moteur de recherche qui permet de sélectionner une offre selon plusieurs critères : fonction (*direction, production, commercial...*) secteur (*BTP, santé, tourisme, mécanique, énergie, chimie...*) lieux (*région parisienne, province...*). Mais surtout il existe aussi une possibilité de recherche par mots-clés et c'est à partir de celle-ci que nous avons poursuivi notre étude.

Nous avons retenu dans un premier temps tous les emprunts de notre liste totalisant un nombre d'occurrences supérieur à 5 et susceptibles par ailleurs de constituer un mot-clé afin de déterminer ceux qui étaient reconnus par le serveur. C'est ainsi que nous avons pu lancer une recherche à partir des termes suivants :

<i>marketing</i> (129)	<i>high tech</i> (60)
<i>leader</i> (401)	<i>merchandise</i> (1)
<i>challenge</i> (65)	<i>prospect</i> (10)
<i>manager</i> (71)	<i>reporting</i> (16)
<i>merchandising</i> (12)	<i>holding</i> (15)
<i>packaging</i> (8)	<i>mix</i> (4)
<i>business to business</i> (12)	<i>process</i> (18)
<i>leadership</i> (26)	

Les chiffres qui figurent entre parenthèses sont les nombres d'offres correspondant à la demande et affichés par le serveur en fonction du mot-clé utilisé. Les chiffres donnés ici sont ceux obtenus lors d'une recherche sur «Cadres on line» entre le 13 juin et le 3 juillet 1997 à partir de 2043 offres d'emploi. Nous ne les communiquons qu'à titre indicatif et il ne peuvent pas être rapprochés du nombre d'occurrences signalé dans notre liste d'emprunts lexicaux (Annexe A) car chaque mot, même répété plusieurs fois dans une annonce, n'est compté qu'une seule fois par le système de recherche.

En revanche nous avons veillé à ne pas prendre en compte les offres rédigées en anglais et signalées par le serveur à l'aide d'un petit drapeau britannique.

Nous avons ensuite effectué une recherche en utilisant cette fois comme mots-clés, non plus les termes anglo-américains cités plus haut mais au contraire les néologismes dont l'usage est fortement encouragé par les pouvoirs publics.

Comme nous le pressentions, les résultats ont été extrêmement décevants.

En ce qui concerne les néologismes suivants :

mercatique (= *marketing*)
chalenge (= *challenge*)

¹⁰ *L'Express, Le Monde, L'usine nouvelle, 01 Informatique, Electronique International, 01 Réseaux, Internet professionnel, Décision Micro et Réseaux, Caractère, France agricole, Agro Distribution, RIA, Néo-Restoration, LSA, L'Echo, Le moniteur, L'Argus.*

¹¹ *Property week, Building.*

manager (= *manager*)
marchandisage (= *merchandising*)
marchandiseur (= *merchandiser*)
mercatique après-vente (= *reporting*)
tenante (= *holding*)
marchéage (= *mix*)

le serveur nous a répondu à chaque tentative de recherche :

«Nous sommes désolés. Aucune offre ne correspond à votre demande».

(Notons au passage qu'une simple modification orthographique du terme emprunté visant à le mettre en conformité avec notre système morpho-phonologique (*manager* = *manager*, *challenge* = *chalenge*) suffit à faire obstacle à son intégration.)

Même réponse négative à partir de la périphrase descriptive :

négociation d'entreprise à entreprise (*business to business*)

En revanche, lorsque nous avons tenté une recherche portant sur un essai de traduction du terme d'emprunt, nous avons enregistré quelques réponses positives en particulier pour :

conditionnement (9) qui obtient un résultat légèrement supérieur à *packaging*

de même que *procédé* (7) qui pourrait se substituer à *process*.

Mais comme nous le verrons plus tard la traduction d'un emprunt par un terme univoque n'est pas toujours possible. Ainsi une recherche menée sur les mots *meneur* et *autorité* en traduction de *leader* et *leadership* renvoie aux qualités requises chez un candidat et ne rend pas compte de la notion de position dominante occupée par une entreprise sur un secteur donné.

Conforté, au terme de cette expérience télématique, dans notre opinion concernant la très faible diffusion des néologismes dans le monde professionnel, nous avons essayé d'analyser les différents facteurs pouvant expliquer une telle indifférence à l'égard de la terminologie officielle.

UNE INFORMATION SUFFISANTE

D'emblée nous avons éliminé l'argument selon lequel ce phénomène relève d'un simple manque d'information.

En effet les travaux des commissions ministérielles de terminologie ont toujours fait l'objet d'une très large diffusion auprès des professionnels et du grand public, soit par le biais d'organes officiels (Les Notes Bleues de Bercy), soit par la presse grand public (*Le Monde*, *Libération*...), soit encore par des organismes privés tels que l'APFA (Association pour Promouvoir le Français des Affaires), association à but non lucratif qui organise chaque année sous le patronage du Commissariat général de la langue française «La journée du français des affaires» et qui publie sous forme de dépliant «Les 700 mots d'aujourd'hui pour les affaires».

TRADUCTIONS TARDIVES

En fait, la première raison qui vient naturellement à l'esprit pour expliquer les réticences des professionnels à utiliser les néologismes officiels se situe dans le décalage que l'on enregistre entre le moment où un terme anglo-saxon est introduit dans notre langue et le moment où un équivalent français est enfin proposé.

L'exemple le plus connu est bien sûr celui de *marketing/mercatique*.

Introduit en France dans les années 1950, le concept nouveau de *marketing* est resté sans traduction jusqu'en 1973, date à laquelle l'emploi de *mercatique* fut d'abord conseillé, puis rendu obligatoire en 1987. *Mercatique* et ses dérivés *mercaticien(-ne)* se proposent de traduire la multitude d'appellations apparues anarchiquement dans les entreprises (*marketing man, marketing manager, marketing management...*) mais, ainsi que nous avons pu le constater dans notre étude, ils demeurent complètement inconnus des professionnels.

Ce décalage est encore accentué par l'évolution très rapide des sciences et des techniques. Les terminologues ont de plus en plus de mal à coller à l'événement linguistique ainsi que le rappelle Claude Hagège : «...le rythme des apports étrangers, accrédités par les médias, est aujourd'hui trop fébrile pour laisser place aux créations lexicales qui arrivent longtemps après la chose.»¹²

Sans doute est-ce ce retard qui permet d'expliquer d'autres échecs comme *marchéage* (*marketing mix*), *marchandisage* (*merchandising*), *minimarge* (*discount*), *minimargeur* (*discounter*), *stylique* (*design*)...

Comment fait-on aujourd'hui pour diffuser les équivalents de *trade marketing* (mercatique de la distribution) et de *reporting* (mercatique après-vente) alors que *mercatique* est lui-même ignoré ?

Le *management* décrit dès 1918 par Henri Fayol¹³, a été officiellement francisé en 1973. Mais combien de temps faudra-t-il encore pour diffuser les équivalents de ses dérivés les plus récents : *disease management* (prise en charge globale de la maladie), *data management* (gestion des données), *facilities management* (gestion des infrastructures) ?

TRADUCTIONS MULTIPLES

À la décharge des utilisateurs d'une terminologie d'emprunt il faut rappeler les hésitations qui caractérisent parfois les directives officielles. Ainsi devant l'échec manifeste d'une première traduction, les commissions de terminologie, après plusieurs années de réflexion, proposent un autre équivalent qui se trouve, pendant quelque temps, en concurrence directe avec le premier terme proposé. On imagine l'embarras du locuteur francophone qui, contraint de faire un choix délicat, optera finalement pour l'expression anglo-saxonne.

¹² HAGÈGE, Claude (1987) : *Le français et les siècles*, Paris, Éditions Odile Jacob, p. 150.

¹³ FAYOL, Henri (1918) : *Administration industrielle et générale*, Paris, Dunod.

Rappelons certaines de ces traductions multiples que l'on a pu rencontrer au cours de ces vingt dernières années :

consumerism fut traduit par *consommérisme* (traduction proposée en 1979 par l'Académie des Sciences Commerciales avec la caution du Conseil International de la Langue Française) alors qu'une circulaire ministérielle du 3 janvier 1994 recommandait l'usage des termes *consommatisisme* et *consommaction*;

ou bien

sponsor, traduit par *commanditaire* dans l'arrêté du 17 mars 1982 relatif au vocabulaire du tourisme et par *parrain* et/ou *parraineur* par l'arrêté du 6 janvier 1989 relatif à la terminologie économique et financière;

ou encore

design, qui fut pendant longtemps traduit par *esthétique industrielle* jusqu'à ce que soit inventé le terme *stylique* en 1983.

EMPRUNTS POLYSÉMIQUES

Il arrive parfois qu'un emprunt s'établisse durablement dans sa langue d'accueil lorsqu'il possède un champ sémantique plus large que celui couvert par ses différentes traductions. Ainsi le mot *challenge* (que l'on tente d'adapter en français sous la forme *challenge* en raison d'une étymologie ancienne et de la règle de doublement des consonnes)¹⁴ ne peut se satisfaire de la traduction *défi*, car ce terme contient également la notion de *remise en question*, d'*entreprise périlleuse* ou encore de *pari*.

Les annonces que nous avons dépouillées en témoignent :

«Homme de challenge vous aimez relever les défis ?» (*L'Express* n° 2390 p. 8)

«Si vous êtes un homme ou une femme de défi et de challenge, rejoignez-nous» (*L'Express* n° 2394 p. 58)

Nous retrouvons le même phénomène avec les mots *leader* et *leadership* que nous avons évoqués précédemment. Empruntés au XIX^e siècle au vocabulaire anglo-saxon de la politique, ces deux termes couramment utilisés aujourd'hui dans les offres d'emploi doivent leur succès à leur multiplicité de sens.

leader : chef, meneur d'homme, chef de file, numéro un, premier, principal..., ex :

«Vous êtes avant tout un leader, capable d'organiser et fédérer votre équipe» (*L'Express* n° 2391 p. 23)

«Nous sommes leader européen de la distribution par catalogue des matériels électroniques...» (*L'Express* n° 2391 p. 28)

¹⁴ VOIROL, Michel (1989) : *Anglicismes et anglomanie*, Paris, Éd. Centre de formation et de perfectionnement des journalistes, p. 19.

«Vous prenez en charge le produit leader du groupe.» (*L'Express* n° 2391 p. 35)

leadership : autorité, sens du commandement, place de premier, position dominante..., ex :

«Excellent communicateur, faisant preuve de leadership, vos qualités d'organisateur et de coordinateur ne sont plus à démontrer.» (*L'Express* n° 2392 p. 20)

«Nous vous offrons :...Un leadership incontesté, des marques de forte notoriété...» (*L'Express* n° 2390 p. 27)

C'est encore la polysémie qui peut justifier l'emprunt du terme *business* tel qu'il apparaît dans les offres d'emploi :

«Vous gérez votre business et pilotez ce centre de profit avec une réelle autonomie» (*L'Express* n° 2396 p. 38)

Traduit en français *business* oblige à faire un choix entre *les affaires*, *l'affaire*, *l'entreprise*, *le commerce*, *la clientèle* ou *le secteur d'activité*...

EMPRUNTS PLUS COURTS

On ne peut parler de l'anglicisation du français sans évoquer le fait que les emprunts à l'anglais sont souvent beaucoup plus courts, et donc moins coûteux à produire que leurs équivalents français. Ce détail n'a pourtant pas échappé aux terminologues qui se sont efforcés de créer des termes au moins aussi courts que les emprunts qu'ils souhaitaient remplacer. *Mercatique*, *informatique*, *ingénierie*, *stylique*, *tenante*, en sont quelques exemples. Malheureusement la structure de notre langue nous force à des créations plus longues.

Quelles chances, de ce point vue, avons-nous de voir le mot *reporting* être remplacé par *mercatique après-vente* (domaine mercatique), *déclaration des ordres* (domaine bancaire) ou encore *reddition des comptes* (domaine comptable) ? Il en est de même pour *démarchage téléphonique* (phoning), *publipostage* (mailing), *réunion préparatoire* (briefing) ou la *mercatique de la distribution* (trade marketing) parfois nommée *distrimercatique*¹⁵.

EMPRUNTS CONNOTATIFS

Les arguments que nous venons d'évoquer permettent de justifier la nécessité matérielle d'utiliser dans une langue professionnelle une terminologie empruntée à l'anglo-américain. Ces arguments, qui ne prennent en compte que la fonction dénotative des emprunts lexicaux, peuvent être acceptés dans la perspective d'une langue technique, fonctionnelle et univoque.

Cependant, qui pourrait croire que la langue des offres d'emploi est technique au point d'être réductible à sa seule fonction référentielle ? Outre le fait qu'il n'existe pas, dans les langues naturelles, de terme purement dénotatif, il est facile de démontrer que

¹⁵ MOULINIER, René (1997) : *Les 500 mots-clés de la vente*, Paris, Dunod, p. 351.

chaque terme d'emprunt est toujours plus ou moins utilisé en fonction de sa valeur connotative

Prenons le terme *manager*. Ce terme est en anglais relativement neutre dans la mesure où il peut s'appliquer indifféremment à un directeur de société ou à un gérant de station-service. Or, si cet emprunt prolifère dans les offres d'emploi, c'est essentiellement en raison de ses connotations. Il désigne ainsi une personne occupant un poste de responsabilités qui adopte nécessairement des méthodes de gestion modernes importées des États-Unis. Naturellement la modernité, l'efficacité voire la jeunesse ne sont connotées que dans la mesure où le terme est perçu comme américain. Le travestir dans la graphie francisée *manageur* lui ôte toute sa puissance évocatrice et le rabaisse au rang de nos *responsables*, *directeurs* et autres *chefs de service*. Les entreprises ne s'y trompent pas et usent de cet artifice pour le recrutement de leurs cadres :

«Chef des ventes.....grâce à vos qualités de manager...» (*L'Express* n° 2394 p. 37)

«Chefs de secteur (managers commerciaux)» (*L'Express* n° 2394 p. 37)

«Technical Sales Manager (responsable de zone)» (*L'Express* n° 2394 p. 40)

ou encore cette définition qui laisserait rêveur un lecteur anglophone :

«Meneur d'hommes, Entrepreneur, en un mot Manager» (*L'Express* n° 2394 p. 32)

Dans bon nombre de cas, il est évident que la valeur connotative de l'emprunt constitue sa seule justification. Nous le constatons lorsque l'emprunt n'est pas introduit pour désigner une réalité anglo-saxonne encore mal connue ou mal acclimatée en France, mais au contraire pour exprimer un concept parfaitement français et pour lequel notre langue possède sa propre désignation.

Que dire en effet des termes *packaging* (conditionnement), *sales* (ventes), *discount* (rabais), *consumer* (consommateur), *retail* (détail) ou encore *process* (procédé) ?

Notons à ce propos que lorsque l'emprunt n'est pas motivé par un réel besoin dénotatif il est soumis aux caprices de la mode et sa durée de vie en français peut être très variable. Ainsi les termes *know-how*, *team*, *winner*, *deal*, très courants dans les années 1980 ont aujourd'hui presque tous disparu.

Même si la majorité des emprunts ont été à l'origine introduits pour palier une lacune linguistique réelle, la plupart se maintiennent dans notre langue par la seule force de leurs connotations. Et c'est en fait un véritable discours qui se met en place sous l'alibi d'une langue technique. Un discours très idéologique qui n'est autre que celui du culte de l'Amérique comme modèle de réussite économique et commerciale. Dénoncée dans les années 1960, «l'américanolâtrie» a été relayée depuis cette époque par les cadres français de l'industrie et du commerce qui adoptent avec enthousiasme un vocabulaire d'outre-Atlantique, dans le seul but de faire rejaillir sur notre réalité économique hexagonale une partie du prestige du modèle américain.

Les terminologues ont beau s'acharner, ni la mercatique, ni le marchandisage n'évoqueront le rêve américain et les entreprises continueront de recruter des *sales managers*...

Naturellement, la maîtrise d'un vocabulaire pseudo-technique joue un rôle social important dans la mesure où elle renforce l'image d'un groupe d'initiés.

Comme le soulignait Claude Hagège : «Les anglicismes techniques tiennent le profane à distance. Ils flattent celui qu'ils distinguent»¹⁶. Mais cette recherche de la distinction à tout prix marque aussi les limites du discours technique. Perçu par le grand public comme une garantie de savoir, l'usage injustifié d'une terminologie d'emprunt n'est pas sans évoquer le latin de cuisine des médecins de Molière et témoigne souvent d'une indigence de la pensée. En effet, si l'emprunt permet de faire l'économie d'une traduction difficile, il dispense également son utilisateur d'une indispensable réflexion à propos de l'objet dénoté. «Le connoté prend alors le pas sur le dénoté et le terme d'emprunt perd toute sa justification du point de vue d'un discours qui se prétend, avant tout, technique et monoréférentiel»¹⁷.

CONCLUSION

Cette étude sur la terminologie en usage dans les offres d'emploi nous a tout d'abord permis de relativiser l'importance de l'emprunt lexical dans ce domaine très circonscrit de la langue des affaires. Loin du raz de marée qu'une lecture un peu rapide de la presse pour cadres pourrait laisser envisager, le phénomène se limite à la généralisation de deux termes (*marketing* et *leader*) qui selon toute vraisemblance s'imposeront définitivement dans notre langue avec leur graphie originale, comme se sont autrefois imposés *week-end*, *gadget* ou *hold-up*. L'emploi des autres termes dépend plus de la nouveauté qu'ils représentent et des connotations qui leur sont attachées, que d'une véritable nécessité dénomminative.

En revanche il faut constater une réelle résistance de la part des milieux professionnels à utiliser la terminologie officielle mise en place par les pouvoirs publics. Certes une langue ne se gouverne pas par décrets et son usage réel s'établit au niveau de toute une communauté dont les réactions peuvent être orientées mais jamais imposées.

On pourrait toutefois déplorer la méfiance collective de la communauté francophone de France à l'égard de toute activité néologique. Les vains efforts des futures ministres du nouveau gouvernement de Lionel Jospin pour imposer l'expression «Madame la Ministre» en sont un très bel exemple.

L'héritage de trois cents ans de purisme langagier, et ce n'est pas le moindre des paradoxes, a hérissé notre langue de tant d'interdits que les francophones ainsi «privés de leur dynamisme inventif»¹⁸ préférèrent emprunter un terme étranger qu'utiliser une expression nouvelle qu'ils jugent, a priori, suspecte.

¹⁶ HAGÈGE, Claude : *op.cit.* p. 115.

¹⁷ SOUBRIER, Jean (1985) : *Le français économique et commercial : Ambiguïté d'une langue parallèle*, Thèse de Doctorat, Université Lyon 2, p. 174.

¹⁸ HAGÈGE, Claude : *op.cit.* p. 172.

Annexe A1

Emprunts lexicaux - Classement par occurrences

marketing	552	géomarketing	2
leader	425	hardware	2
challenge	96	key account	2
manager	61	Key Account Manager	2
merchandising	26	marketing mix	2
packaging	23	marketing on line	2
business to business	15	mediaplanning	2
leadership	15	outdoor	2
high tech	12	project director	2
merchandise	11	screening	2
prospect	11	show room	2
reporting	9	success story	2
holding	7	web	2
mix	7	back office	1
process	6	biomedical operation manager	1
B to B	5	brief	1
design	5	business analyst	1
helpdesk	5	business plan	1
sportswear	5	business school	1
British	4	business unit	1
business	4	business unit manager	1
corporate	4	cash and carry	1
data management	4	challenger	1
engineering	4	co-leader	1
laser	4	coaching	1
mix marketing	4	commando	1
start up	4	consulting	1
télémarketing	4	consumer marketing manager	1
trade marketing	4	credit manager	1
broadcast	3	data communication	1
CD-Rom	3	data manager	1
internet	3	digital mapping	1
mailing	3	direct marketing	1
medico-marketing	3	drug master files	1
mixt	3	drug safety officer	1
monitoring	3	entertainment	1
planning	3	entrepreneurial	1
pressing	3	Europe's first marketing agency	1
pricing	3	European Key Accounts	1
scanner	3	facilities management	1
business development	2	fax	1
cocktail	2	fluent	1
consumer	2	fluid handling components	1
data	2	full service	1
discount	2	fun	1
disease management	2	get yourself connected	1

hard	1
healthcare	1
incentive	1
interviewer	1
intranet	1
job	1
kit	1
labelling	1
laser disc	1
leasing	1
marchandising	1
marketing B to B	1
marketing promotion	1
meeting	1
micro-marketing	1
mini disc	1
new biz	1
office product	1
on line	1
on time delivery	1
organizer	1
outplacement	1
outsourcing	1
package carrier	1
package express	1
pool	1
product line management	1
provider	1
purchasing manager	1
push technology	1
recording	1
remodeling	1
retail	1
retail marketing manager	1
royalties	1
sales and marketing	1
sales engineer	1
sales manager	1
sales operation	1
search	1
senior management program	1
soft drink	1
sourcing	1
stock option	1
stress	1
team leader	1
technical sales manager	1
test	1
top	1
total look	1

Annexe A2**Emprunts lexicaux - Classement
alphabétique**

B to B	5	full service	1
back office	1	fun	1
biomedical operation manager	1	géomarketing	2
brief	1	get yourself connected	1
British	4	hard	1
broadcast	3	hardware	2
business	4	healthcare	1
business analyst	1	helpdesk	5
business development	2	high tech	12
business plan	1	holding	7
business school	1	incentive	1
business to business	15	internet	3
business unit	1	interviewer	1
business unit manager	1	intranet	1
cash and carry	1	job	1
CD-Rom	3	key account	2
challenge	96	Key Account Manager	2
challenger	1	kit	1
co-leader	1	labelling	1
coaching	1	laser	4
cocktail	2	laser disc	1
commando	1	leader	25
consulting	1	leadership	15
consumer	2	leasing	1
consumer marketing manager	1	mailing	3
corporate	4	manager	61
credit manager	1	marchandising	1
data	2	marketing	552
data communication	1	marketing B to B	1
data management	4	marketing mix	2
data manager	1	marketing on line	2
design	5	marketing promotion	1
digital mapping	1	mediaplanning	2
direct marketing	1	medico-marketing	3
discount	2	meeting	1
disease management	2	merchandiser	11
drug master files	1	merchandising	26
drug safety officer	1	micro-marketing	1
engineering	4	mini disc	1
entertainment	1	mix	7
entreprenarial	1	mix marketing	4
Europe's first marketing agency	1	mixt	3
European Key Accounts	1	monitoring	3
facilities management	1	new biz	1
fax	1	office product	1
fluent	1	on line	1
fluid handling components	1	on time delivery	1

organizer	1	sales and marketing	1
outdoor	2	sales engineer	1
outplacement	1	sales manager	1
outsourcing	1	sales operation	1
package carrier	1	scanner	3
package express	1	screening	2
packaging	23	search	1
planning	3	senior management program	1
pool	1	show room	2
pressing	3	soft drink	1
pricing	3	sourcing	1
process	6	sportswear	5
product line management	1	start up	4
project director	2	stock option	1
prospect	11	stress	1
provider	1	success story	2
purchasing manager	1	team leader	1
push technology	1	technical sales manager	1
recording	1	télémarketing	4
remodeling	1	test	1
reporting	9	top	1
retail	1	total look	1
retail marketing manager	1	trade marketing	4
royalties	1	web	2

Annexe B

Emprunts sémantiques - Classement par occurrences

approche	manière d'aborder (un problème)	4
attractif	attirant	31
auditeur	vérificateur,	
	commissaires aux comptes	3
contrôle	maîtrise, direction	4
contrôler	diriger	2
développer	accroître	12
développeur	chargé d'accroître	
	(un volume d'affaires)	10
en charge de	responsable de	9
junior	débutant	31
major (adj.)	principal	1
opportunité	occasion	80
rapportant (part. pré)	(to report to) : rendre compte à	3
reportant (part.pré)	(to report to) : dépendre	
	hiérarchiquement	2
senior (adj.)	confirmé	51

Emprunts sémantiques - Classement par occurrences

opportunité	occasion	80
senior (adj.)	confirmé	
attractif	attirant	31
junior	débutant	
Développer	accroître	
développeur	chargé d'accroître	
	(un volume d'affaires)	10
en charge de	responsable de	9
approche	manière d'aborder (un problème)	4
contrôle	maîtrise, direction	4
auditeur	vérificateur, commissaires	
	aux comptes	3
rapportant (part. pré)	(to report to) : rendre compte à	3
contrôler	diriger	
reportant (part.pré)	(to report to) : dépendre	
	hiérarchiquement	2
major (adj.)	principal	

ENTRE STOCKAGE ET RÉEMPLOI EN TERMINOLOGIE : UN PROCESSUS D'ANAMNÈSE ?

Michèle A. LORGNET

Université de Bologne-Forlì, Italie

On présente ici une hypothèse de recherche qui a pour origine une réflexion sur l'utilisation de la terminologie dans ses deux aspects et fonctions : le stockage des connaissances et leur réemploi successif en traduction vers la langue étrangère. La recherche emprunte une méthode empirique analysant le double apprentissage des moyens terminologiques et terminographiques et de leur utilisation en traduction professionnelle pour de jeunes apprenants-traducteurs.

Les apprenants concernés, les étudiants en traduction et interprétation de l'Université de Bologne-Forlì (SSLiMIT), ont été observés pendant la période de rédaction de leur mémoire de maîtrise. Pour vérifier, ou anticiper, durant cette période, la capacité professionnelle de ces futurs traducteurs *in situ*, (Gouadec 1994 : 163-182), c'est-à-dire dans des entreprises et des organismes de la Romagne, nous avons constitué un groupe de travail de chercheurs et d'enseignants pour assister ces «impétrants» en traductologie, et ce pour toutes les langues enseignées à l'école, c'est-à-dire l'anglais, le français, l'allemand, l'espagnol et le russe. Une période de stage et de collecte de matériel à l'intérieur des entreprises a donc été doublée et suivie d'une période d'informatisation assistée.

Dans les pays de langue minoritaire comme l'Italie, la plupart des travaux de traductions techniques requis par les petites et moyennes entreprises sont des traductions vers la langue étrangère. Il s'agit dans la grande majorité des cas de textes pour la présentation ou l'entretien de matériel, textes allant du simple opuscule au livret technique élaboré, du simple feuillet illustratif accompagnant la machine ou le produit, au volume présenté en librairie, par exemple, en ce qui concerne un des organismes concernés. L'hypothèse de recherche collective a donc envisagé plusieurs stades dans l'élaboration pédagogique nécessaire à une telle assistance. Pour pouvoir développer «l'autonomie de production de traduction» des étudiants, il a d'abord fallu leur demander d'informatiser un matériel linguistique à l'état brut, concernant une terminologie extrêmement variée, difficile à classer, en micro-glossaires bi- ou trilingues. Pour tester leur capacité de réemploi du matériel terminologique informatisé par leurs soins lors de traductions vers la langue étrangère à partir de l'italien, il a été ensuite nécessaire de contrôler les micro-corpus en corpus parallèles (Lorget 1997 : 7-9).

La période de stages financés par la municipalité, la Chambre de Commerce et les banques de Forlì, avait garanti le côté officiel de cette première entrée dans le monde du

travail. Les étudiants effectuaient ainsi la période de collecte de matériel à l'intérieur des entreprises, travaillant généralement par groupe de deux, pour un maximum de trois cents heures de recherche. Ces étudiants étaient tous inscrits à la dernière année de leurs cursus universitaires. En même temps, ils constituaient les fiches terminologiques bi- ou trilingues des micro-glossaires dans une série de domaines qui étaient, par conséquent, en même temps extrêmement restreints au niveau de la spécialisation et extrêmement larges et flous au niveau de la nature des textes traduits par la suite.

Les domaines concernés étaient les suivants : artisanat, agriculture intégrée, industrie mécanique — élévateurs hydrauliques, chaudières à gaz, agrafeuse, machines agricoles pour jardinage —, biologie, préservation de l'environnement d'un parc national, le Parc National du «Casentino» ayant, à l'occasion, demandé la traduction d'un important volume concernant l'histoire, la géologie, la faune et la flore de ce milieu naturel. La constitution des fiches de données était toujours soumise à un triple contrôle : contrôle linguistique des enseignants, confrontation avec les documents sur la normalisation des termes à la Chambre de commerce de Forlì, consultation des banques de données existant sur le marché et consultation technique des ingénieurs et spécialistes des entreprises. L'assistance technique du travail informatique a été effectuée par Franco Bertaccini.

La fiabilité du travail terminologique accompli était vérifiée en retour dans des traductions de documents professionnels vers la langue étrangère avec confrontation directe du travail par les partenaires étrangers des entreprises.

À la suite de résultats assez fluctuants dans la recherche, puis ensuite dans l'utilisation des termes, leur classification et le réemploi successifs de notions techniques ou technologiques dans une langue étrangère (Assal 1994 : 461-462), il a fallu essayer d'identifier le pourquoi de ces nombreuses difficultés d'apprentissage.

Tout d'abord, l'abstraction notionnelle nécessaire à l'opération de stockage de la terminologie semblait poser des problèmes dans l'opération successive du réemploi en langue étrangère, surtout lorsque le réemploi advenait en un second temps assez éloigné. «Les résidus sémantiques appartenant à un lexème restent collés à lui quand il devient terme» (Assal, 1994 : 462). Le problème ardu de la sélection des synonymes ou quasi synonymes (Kocourek 1982 : 164-167; Lethuillier 1989 : 443-450, Gross et Clas 1997 : 149-150) d'un domaine spécifique ne semblait devoir être résolu que par l'appel à un parlant de langue maternelle (dans notre cas toujours un enseignant de notre groupe de travail).

La prise de conscience de la globalité du contexte terminographique (De Bessé 1991 : 112-113) apparaissait en réalité plus facile à circonscrire lors du stockage que lors du réemploi. La cohérence notionnelle, la consistance des groupements préférentiels (Thoiront et Béjoint 1989 : 666-667), si précieuses dans la langue maternelle, ne semblait souvent à nos traducteurs en herbe bien plus une difficulté seconde qu'un atout lors de leur travail de mémorisation à long terme. Les imbrications des phénomènes sémantiques avec les connexions syntaxiques (Mel'cuk 1978 : 288-289), particulièrement utiles lors du stockage notionnel, nécessitant toujours une activité contrôlée très forte dans la langue non-maternelle (Gaonac'h 1996 : 27-28), introduisaient quelquefois même un blocage du travail du traducteur vers la langue non maternelle. Le repérage des unités terminologiques dans le contexte de la traduction spécialisée (Mareschal 1989 : 377-378), — effectivement

plus aisé à réaliser sur la base de données textuelles — et la lecture paradigmatique en profondeur du texte à traduire, dans le secteur spécialisé et technique comme dans celui du texte littéraire (Lorgnet 1995), ne garantissaient pas toutefois la même habileté lors de la production en traduction. Après une période d'observation qui n'a somme toute été faite que lors de cette première année de recherche, il nous a semblé que l'opération raisonnée du stockage terme à terme est effectivement partiellement insuffisante pour l'appropriation plus ou moins automatique des termes dans une langue étrangère, appropriation nécessaire lors des traductions rapides effectuées en entreprises. D'autre part, la vitesse de réemploi des syntagmes terminologiques est étroitement liée à la faculté associative des apprenants, même lorsque l'analyse des corpus en parallèle facilite ce réemploi.

Il s'agissait donc de relever, à l'intérieur de ces groupes figés, quels étaient les groupes de mots les mieux stockés, c'est-à-dire qui ne nécessitent plus la consultation des fiches terminologiques en mémoire, pour comprendre si la constitution de réseaux notionnels (Blampain 1992) ou de représentation sémantique est ou non soumise à une mémorisation a-notionnelle fondée sur d'autres facteurs de mémorisation.

La distinction classique entre la capacité de mémorisation à court terme (valeur extensive des termes) et la mémoire à long terme (valeur intensive du terme) semble remise en question lors du réemploi, surtout si l'opération de réemploi intervient sous contrainte temporelle forte. L'instance de contrôle nécessaire à la reverbération en traduction vers la langue étrangère apparaît effectivement n'intervenir qu'en complément des processus d'automatisation (Gaonac'h 1996 : 34). Le travail sur la nature des synonymes en stade de stockage avait utilisé le schéma du CILF pour rendre optimale la compréhension notionnelle des termes. Pour une telle opération, le repérage en langue maternelle du sème culminant sémantiquement est bien sûr très utile (Cigada 1988). Lors du réemploi, on repère malheureusement une sorte de brouillage de la mémoire notionnelle distincte et, lors de la traduction d'un texte d'une certaine longueur, la consultation répétée des fiches en mémoire allonge considérablement le travail.

Exemples

Terme source

terme cible

a) représentation complète en langue source et en langue cible

acclimatamento della pianta

adattamento

compatibilità

acclimation

adaptation

compatibilité

b) termes équivalents présentant plusieurs synonymes

b1) plusieurs synonymes en langue cible

stolone

selvatico

surgeon

drageon

balais de sorcière

verticillo

verticille

patte d'oie

argilla
mama

argile
marne
glaise
molasse

b2) les termes de la langue source n'ont pas d'équivalent; les synonymes cibles approximatifs ne sont pas équivalents

trattrice agricola

.....

.....
tracteur
véhicule tous terrains

b3) les termes de la langue source n'ont pas d'équivalence synonymique; la langue cible a des synonymes

impagliata

.....

.....
plat à légumes ?

b4) le terme source est acceptable en terme cible

Tatura trellis

Tatura trellis

c1) les termes cibles présentent une traduction oblique

radicelle

(ébauche racinaire)
broussin

c2) le terme cible traduit un synonyme de la langue source

non crescita del meristema

morìa del meristema

ramo spoglio

avortement du méristème
.....
chicot
branche défoliée

Le problème des équivalences sémantiques et de la traduction (Lyadri 1997 : 143) touche donc de très près le rapport avec certains phénomènes linguistiques importants tels que la métonymie (Horacek 1996 : 113) ou si l'on veut la métaphore au sens large et aristotélicien du terme (Lorgnet 1995).

La recherche en cours a donc tenté de repérer quels étaient, dans un domaine sectoriel donné, les syntagmes figés dont la reconnaissance initiale au cours du stockage produit, à long terme, une meilleure production en traduction.

Si l'on part du principe, cher à Halliday, que l'importance du niveau lexicographique l'emporte sur celle du niveau syntactico-sémantique, on peut alors analyser l'efficacité de certains critères comme celui de la boucle phonologique (Baddeley, Papagno, Vallai 1988 : 586-595) et les réseaux associatifs rhétoriques du verbal et du non-verbal.

Avant toute chose, le «style» d'un texte technique ou technologique particulier, c'est-à-dire le surplus de son sens prescriptif et normatif dans une direction-cibliste unilatérale (Dimarco, Mah 1994 : 21-41) se doit d'être perçu et analysé par le traducteur. L'ouverture des choix lexicaux paradigmatiques en langue cible ne semble toutefois perceptible et productive que si la mémorisation syntagmatique s'est, en un certain sens, vidée du sens prescriptif et/ou notionnel. Sans en revenir à un behaviourisme polémique,

il est certain, de par cette première phase de notre recherche, que la mémorisation en série parallèle telle que la représente la mémoire terminologique n'est pas suffisante pour la production en langue étrangère d'un texte cohérent et cohésif. Les «magasins de la mémoire» platonicienne ont d'autres clefs, et il n'est pas si sûr que la mémorisation d'une langue étrangère passe par les mêmes voies que celle de la langue maternelle. Il faudrait à ce propos revoir les assertions et les polémiques sur le bilinguisme précoce.

Dans une série de termes synonymes, la mémorisation la meilleure semble presque toujours, de par notre première expérience, concerner le premier terme de la liste qui suit. Quand la mémorisation manque complètement, la recherche notionnelle des synonymes en langue source, ou bien l'analyse de l'apex sémantique, régénère le terme cible oblique, c'est-à-dire le terme en forme de syntagme ou altéré par une transformation sémantique morphématique ou métaphorique.

Fil d'Archal

fil de laiton

cruchon

pichet

broc

dormance

latence

habiller les racines

pralinage

préparation des racines

bouquet de mai

rosette de bourgeons (de feuilles)

sevrage

taille au collet

balais de sorcière

surgeon

CONCLUSION

Une analyse prolongée et systématique de ces intuitions pourra sans doute être mieux élaborée par la suite, en multipliant les sujets d'observation. Il reste toutefois que l'effort mnémotechnique des apprenants en terminologie pour la traduction vers la langue étrangère fait appel à des mécanismes échappant à un notionalisme pur et dur, et qu'une meilleure prise de conscience de libres associations synonymiques en langue maternelle grâce à des exercices préparatoires en classe de traduction aide les futurs traducteurs à développer une fonction parallèle d'anamnèse également dans la langue étrangère.

RÉFÉRENCES

- ASSAL, A. (1994) : «La notion de notion en terminologie», *META*, XXXIX, 3, pp. 460-464.
- BADDELEY, A.D., PAPAGNO, C., et G. VALLAR (1988) : «When long-term learning depends on short-term storage», *Journal of Memory and Language*, 27, pp. 586-595.
- BLAMPAIN, D. (1992) : «Traduction et écosystèmes terminologiques», *Équivalences*.
- De BESSE, B. (1991) : «Le contexte terminographique», *META*, XXX (1), pp. 111-120.
- CIGADA, S. (1988) : «I meccanismi del senso : Il culminatore semantico», E. Rigotti et C. Cipolli (dir), *Ricerche di Semantica testuale*, Brescia, La Scuola, pp. 25-70.
- DIMARCO, Ch. et K. MAH (1994) : «A Model of Comparative Stylistics for Machine Translation», *Machine Translation*, pp. 21-60.
- GAONAC'H, D. (1996) : «Processus cognitifs de base dans l'acquisition des langues», *Bulletin suisse de linguistique appliquée*, 63, pp. 25-36.
- GOUADEC, D. (1994) : «Terminologie, phraséologie et informatique dans la formation des traducteurs professionnels», M. Lorgnet (dir), *Atti della Fiera Internazionale della Traduzione, Forum di Forlì, 3-6 dicembre 1992*, «Biblioteca della Scuola Superiore di Lingue Moderne per Interpreti e Traduttori», 4, Bologna, CLUEB, pp. 163-183.
- GROSS, G. et A. CLAS (1997) : «Synonymie, polysémie et classes d'objets», *META*, XLII (2), pp. 147-154.
- HORACEK, H. (1996) : «On expressing Metonymic Relations in Multiple Languages», *Machine Translation*, 11, pp. 109-158.
- KOCOUREK, R. (1982) : *La langue française de la technique et de la science*, Wiesbaden, Branstetter Verlag.
- LETHUILLIER, J. (1989) : «La synonymie en langue de spécialité», *META*, XXXIV (3), pp. 443-449.
- LORGNET, M.A. (1995) : *Pour une traduction holistique. Recueil d'exemples pour l'analyse et la traduction*, Bologna, Clueb, pp. 135.
- LORGNET, M.A. (à paraître) : «Coherence into Non-mother Tongue Terminology», *Proceedings of the Conference "Translation into Non-Mother Tongues"*, Ljubljana, 29-30 May 1997.
- LYADRI, R. (1997) : «Problématique des équivalences sémantiques et de la traduction dans des dictionnaires arabe-français», *META*, XLII (1), pp. 142-146.
- MARESCHAL, G. (1989) : «Repérage d'unités terminologiques dans le contexte de l'enseignement de la traduction spécialisée», *META*, XXXIV (3), pp. 377-380.
- MEL'CUK, I. (1978) : «Théorie de langage, théorie de traduction», *META*, XXIII (4), pp. 271-302.
- THOIRON Ph. et H. BÉJOINT (1989) : «Pour un index évolutif et cumulatif de cooccurrents en langue techno-scientifique sectorielle», *META*, XXXIV (4), pp. 661-671.

INSTITUTIONS LEXICALES ET DICTIONNAIRES TRADITIONNELS MALGACHES*

Jean-Yves MORIN

Université de Montréal, Canada

1. INTRODUCTION

Les *dictionnaires traditionnels*, malgré leurs défauts, ont l'immense avantage d'être aisément manipulables par les humains (et, jusqu'à un certain point, par les machines). Dans les bases lexicales à couverture large, on aimerait pouvoir intégrer ces divers types d'information et générer, de façon automatique ou semi-automatique, des représentations riches et souples. Les premiers travaux portant sur l'exploitation informatique des dictionnaires traditionnels (sous format électronique)¹ avaient des buts à la fois trop ambitieux et trop limités. D'une part, ils voulaient dériver automatiquement une représentation sémantique formelle à partir des définitions contenues dans ces dictionnaires en sous-estimant grandement la complexité et le manque de fiabilité d'une telle transformation. D'autre part, la sémantique lexicale sous-jacente était extrêmement étroite (pour ne pas dire simpliste) et l'on n'utilisait généralement qu'une seule source de données². Ce que nous proposons ici, c'est de reconsidérer le problème sous un autre angle, en tirant le meilleur profit possible des dictionnaires traditionnels. Il ne s'agit pas de proposer un nouveau formalisme de représentation de l'information lexicale, il en existe déjà plusieurs³, ni un ensemble d'algorithmes de restructuration de lexiques⁴, mais plutôt un ensemble de méthodes d'analyse fonctionnelle des données lexicales, les *institutions lexicales*, qui permettent un typage rapide de propriétés en termes de contenus et qui

* Une version préliminaire de ce travail a été présentée aux V^{es} Journées scientifiques du réseau LTT *La mémoire des mots* à Tunis, en septembre 1997. Ce travail a bénéficié de l'aide du réseau LTT à l'équipe «TAO et théories linguistiques» (Anne Abeillé, Philippe Blache, Jean-Yves Morin et Éric Wehrli) ainsi que d'une subvention de recherche du fonds FCAR du Gouvernement du Québec. Pour des raisons éditoriales, le contenu de cette présentation a été scindé en trois parties autonomes : «Lexiques-grammaires et institutions lexicales», «Institutions lexicales et DEC» et le présent article.

¹ Cf. par exemple Boguraev & Briscoe (1988).

² Généralement le dictionnaire dont on se trouvait à disposer sous format électronique.

³ Par exemple DATR (Evans & Gazdar 1996), le lexique génératif (Pustejovsky 1995), INTEX (Silberztein 1993).

⁴ Bien que nous utilisions effectivement de tels algorithmes, à titre expérimental.

facilitent l'identification des *conditions de collage* entre institutions lexicales distinctes en fonction des *domaines* qu'elles recouvrent.

Les dictionnaires traditionnels du malgache, comme ceux des langues sémitiques, sont classés par racines, elles-mêmes classées dans l'ordre alphabétique. Si un tel classement peut poser des obstacles à l'utilisateur non malgachophone, qui a souvent du mal à déterminer sous quelle racine chercher une forme donnée⁵, il est particulièrement utile pour la morphologie computationnelle, dans la mesure où le réseau dérivationnel d'une forme y est présenté de façon synoptique. Au niveau de la création de bases lexicales, cette discipline est également très pratique, puisqu'elle permet de créer de façon systématique, en même temps que la racine, un ensemble de dérivés potentiels qui constitueront autant de fragments d'entrées lexicales à compléter, reliés à la racine par des hyperliens.

Les dictionnaires traditionnels ne comportent évidemment pas de représentation sémantique explicite. La composante sémantique y est représentée par des traductions ou des définitions. Il en est de même pour la composante syntaxique, qui se réduit à l'attribution d'une partie du discours à chaque forme. Cependant, la richesse de l'information morphologique et colocationnelle consignée dans ces dictionnaires permet de construire le squelette d'une base de données lexicales susceptible d'être enrichie. En principe, la construction d'une telle base de données constitue un travail long et délicat (comme tout travail lexicographique) qui demanderait non seulement une connaissance parfaite de la langue, mais en outre l'accès à des corpus étendus. Nous nous contentons d'adopter ici une approche à la fois *opportuniste*, qui tente d'exploiter au maximum les informations rassemblées par les lexicographes traditionnels et de les rendre accessibles par des chemins multiples (qui, eux, ne sont pas traditionnels) et *paresseuse*, qui permet de retarder les tâches difficiles (par exemple, l'analyse sémantique fine). La compilation de l'ensemble des formes (racines ou dérivés) en un module (arbre, trie, automate, etc.) permettant d'indexer l'information lexicale sur toutes les formes (et dans toutes les directions) ne pose aucun problème de principe. Une partie importante de l'information lexicale peut donc être directement accessible à un utilisateur quelconque (qui peut aussi bien être un programme qu'un humain) depuis les formes dérivées (et fléchies) vers les formes de base, mais également par plusieurs autres chemins.

À titre d'exemple, considérons l'entrée *varotra* dans trois dictionnaires traditionnels (deux dictionnaires bilingues malgache-français et un dictionnaire monolingue malgache) : Abinal & Malzac (1888, A&M), Rajemisa-Raolison (1985, RR) et Rajaonarimanana (1995, NR). Voyons comment on pourrait exploiter ces informations en déterminant les *conditions de collage* de ces trois entrées à l'intérieur d'un squelette commun.

⁵ Le *Rakibolana malagasy-italiana/Dizionario italiano malgascio* du père Profita (Profita 1969) est organisé de façon alphabétique tant dans sa partie malgache-italienne que dans la partie italienne-malgache et ne contient pratiquement pas de renvois. Il est intéressant de noter qu'à notre connaissance, un des seuls autres dictionnaires où les formes malgaches soient classées par ordre alphabétique (et donc plus accessibles à qui ne connaît pas bien la morphologie malgache) est le *Rakibolana malagasy* (Rajemisa-Raolison 1985, RR), un dictionnaire monolingue malgache, qui s'adresse donc précisément aux malgachophones.

	VAROTRA , <i>s.</i> Vente, commerce, négoce, trafic, marchandises. (Au passif prend. <i>Amidy</i> . Voy. VIDY).
	Mi —, <i>va.</i> Vendre, faire le commerce, trafiquer. Mivaròta , <i>imp.</i> Fi —, <i>s.</i> Ce qu'on peut vendre, manière de vendre. Mpi —, <i>s.</i> Celui qui vend, le marchand, le commerçant.
	Fivaròtana , <i>s.</i> L'action de vendre, de faire le commerce, de trafiquer, le prix, le lieu, la cause. Ivaròtana , <i>rel.</i> Ariary telopolo no nivarotako ny omby mifahiko. J'ai vendu trente piastres le bœuf que j'avais engraisé. Ivaròty , <i>imp.</i> Prend Mampi—, Mifampi—.
	Maha —, <i>Qui peut ou sait vendre.</i>
	Varobàrotra , <i>dupl. dim. de Varotra.</i>
1	Mamola-bàrotra , <i>Chercher à faire baisser, à réduire le prix.</i>
2	Manongom-bàrotra , <i>Voy. SONGONA.</i>
3	Tombom-bàrotra , <i>s.</i> Profit, bénéfice.
4	Varotra ambongàdiny , <i>s.</i> Vente en gros, par pièce.
5	Varotra an-trano tòkana , <i>s.</i> Monopole.
6	Varotra an-tsinjàrany , <i>s.</i> Vente au détail.
7	Varo-bàbo , <i>s.</i> Très bon marché, à vil prix. ⁶
8	Varo-balòngana , <i>Voy. VALONGANA.</i>
9	Varo-baventy , <i>s.</i> Commerce en grand; <i>fig.</i> tout vol où on expose sa vie.
10	Varo-bazàha , <i>s.</i> Vente à prix fixe.
11	Varo-bòba , <i>s.</i> Comme VARO-BABO.
12	Varo-dròà , <i>s.</i> Action de vendre encore ce qu'on a vendu, de promettre à l'un ce qu'on a promis à l'autre.
13	Varo-jabo an-tsena, ka samy mandoka ny azy ho mäsaka , <i>prov.</i> Chacun vante sa marchandise, comme les marchands de Jabo au marché.
14	Varo-mahay mody et Varo-mièra , <i>s.</i> Vente d'un article qui peut être rendu si l'acquéreur n'en est pas content; action de prendre à crédit avec obligation de rendre l'objet à défaut du prix.
15	Varo-mahèry , <i>s.</i> Vente forcée.
16	Varo-màika , <i>s.</i> Vente faite à vil prix pour cause d'urgence.
17	Varo-mamindro , <i>s.</i> Achat dont on fait traîner la discussion dans l'espérance de trouver mieux.
18	Varo-mandèha , <i>s.</i> Colportage.
19	Varo-màty , <i>s.</i> Vente définitive..
20	Varo-miàndry , <i>s.</i> Objets qu'on garde longtemps, qui ne se vendent que par circonstance, dont on demande un bon prix.
21	Varotry ny mananotèna , <i>lit.</i> Commerce des veuves; <i>fig.</i> vente faite avec hésitation, dans la crainte de se tromper.
22	Varotr'omby anaty ambiàty , <i>lit.</i> Vente de bœufs cachés dans les ambiaty; <i>fig.</i> vente malhonnête.
23	Varo-tsy azon-dàhy , <i>s.</i> Marché sur lequel on ne s'entend pas, malentendu dans une vente ou un achat.
24	Varo-tsy mifòdy , <i>s. lit.</i> Vente sur laquelle on ne revient pas; <i>fig.</i> la mort.

Figure 1 : Varotra (Abinal & Malzac 1888 : 815)

⁶ La forme *varo-babo*, qui apparaît dans Abinal et Malzac, n'est pas reconnue par nos informateurs, seule la forme *varo-boba* existe pour eux. C'est également la seule forme répertoriée dans les dictionnaires plus récents. Ou bien la forme *varo-babo* s'est perdue au profit de son synonyme et quasi-homonyme (équivalent contrepétique) *varo-boba*, ou bien il s'agit d'une erreur dans A&M. Cependant *varo-babo* y apparaît aussi bien sous l'entrée *babo* que sous celle de *varotra*, si erreur il y a, elle n'est donc pas accidentelle.

	Varotra a. : Fanakalozan-javatra amin'ny vola, hitadiavantombony : <i>Lasa nanao varotra any an-tsiraka izy.</i>
	♦ Entana amidy : <i>Aza hitsahinao ity varotr'olona.</i>
	• <i>t.i.f.</i> : Voavarotra / mivarotra, fivarotra, mpivarotra / ivarotana, fivarotana / mampivarotra, mifampivarotra, mahavarotra / varobarotra.
	• F.-p. :
1	<i>Ady varotra</i> : fifamaliana ataon'ny mpivarotra sy ny mpividy momba ny vidin'entana ka ny mpividy mangataka mba hahenan'ny mpivarotra io vidiny io, ny mpivarotra kosa tsy mety mampihena avy hatrany.
2 (< A&M 1)	<i>/ Mamola-barotra</i> : mampitamby ny mpivarotra ny mpividy mba hampihenany ny vidin'ny entana iadiam-barotra, na mampitamby ny mpividy ny mpivarotra mba hanekenany ny vidiny araka ny tiany hivarotana azy.
3 (< A&M 2)	<i>/ Manongom-barotra</i> : mahita olon-kafa mitady hividy ilay zavatra efa nokasaina hovidina ka manondrotra indray ny vidiny efa nomena.
4	<i>/ Mivaro-tena</i> : mahazo vola amin'ny fijanganjangana.
5	<i>/ Varotra alika</i> : varotra ifanerena.
6 (< A&M 4)	<i>/ Varotra ambongadiny</i> : varotra be no atao fa tsy mitsipotipotika.
7	<i>/ Varotra an'elakela-trano, varotra an-kobohobo</i> : varotra atao any antakontakona tsy sahy mazava noho ny mpivarotra ta hanondrotra ny vidin'entana tsy ara-drariny.
8 (< A&M 5)	<i>/ Varotra an-trano tokana</i> : varotra entana tsy mety hita any an-kafa fa ao amin'io toerana io ihany.
9 (< A&M 6)	<i>/ Varotra antsingarany</i> : fivarotana zavatra tsipotehina madinidinika.
10	<i>/ Varotra olona</i> : fivarotana olona atao mpanompo.
11 (< A&M 22)	<i>/ Varotra omby anaty ambiaty</i> : varo-javatra tsy eo ambany maso.
12 (< A&M 21)	<i>/ Varotry ny mananotena</i> : varotry ny mpivarotra misalasala noho izy sady ta hahazo vola no toa tsy mahafoy entana.
13 (< A&M 10)	<i>/ Varo-bazaha</i> : varotra tsy misy ady varotra fa efa raikitra hatrany tsy azo ovana ny vidin'entana.
14 (< A&M 11)	<i>/ Varo-boba</i> : varotra mora dia mora
15	<i>/ Varo-dratsy</i> : fivarotana zavatra nangalarina.
16 (< A&M 12)	<i>/ Varodroa</i> : enti-milaza olona roa samy mivarotra zavatra iray (oh. trano iray izay karakaraina mpanera roa samy hafa).
17 (< A&M 14)	<i>/ Varotra mahay mody</i> : varotra azo averina na afody.
18 (< A&M 16)	<i>/ Varo-maika na Varotry ny maika hiala, Varotry ny maika hody</i> : varo-mora.
19	<i>/ Varo-maizina</i> : mitovy amin'ny varotra an'elakela-trano.
20 (< A&M 18)	<i>/ Varo-mandeha</i> : fivarotana entana mifindrafindra amin'ny toerana maro samihafa fa tsy amin'ny toerana anankiray raikitra.
21 (< A&M 19)	<i>/ Varo-maty</i> : varotra entana tsy azo averina intsony na manao ahoana na manao ahoana. ♦ h.o. : Fahafatesana.
22 (< A&M 20)	<i>/ Varo-miandry</i> : varotra entana izay avela hijanona eo raha mbola tsy lafo.
23	<i>/ Varo-miera na Varo-mifody</i> : varotra entana azon'ny mpividy averina raha tsy mahafa-po azy.
24	<i>/ Varotra misy fehi-vavany</i> : varotra misy fepetra izay azo ravana.
25	<i>/ Varotra tsy miverimbodiondry</i> : varotra entana nanomezana vola mialoha ho fandraiketam-barotra, ka raha tahiny tsy raikitra ny varotra any aoriana, dia lany maina iny vola iny fa tsy averin'ilay olona tompon'entana intsony.

26 (< A&M 24) / *Varo-tsy mifody* : varotra entana tsy azo averina intsony. ♦ h.o. : Fahafatesana.

Figure 2 : Varotra (Rajemisa-Raolison 1985 : 1017)

Voici une traduction approximative de cet article (avec les indices pour faciliter les références)⁷.

«**Varotra** a. : Commerce, vente. Échange d'objet contre de l'argent, dans le but est d'en retirer un profit : *Lasa nanao varotra any an-tsiraka izy. Il est parti faire du commerce dans les régions côtières.* ♦ Marchandise à vendre : *Aza hitsahinao ity varotr'olona. Ne piétine pas ces marchandises.*

•,• Mots de même racine : **Voavarotra** / **mivarotra**, **fivarotra**, **mpivarotra** / **ivarotana** **fivarotana** / **mampivarotra**, **mifampivarotra**, **mahavarotra** / **varobarotra**.

• Expressions : (1) *Ady varotra* : Marchandage. Débat entre vendeur et acheteur sur le prix de la marchandise où l'acheteur demande au vendeur de réduire son prix et le vendeur essaie de baisser petit à petit le prix initial. (2) / *Mamola-barotra* : Négociation sur le prix. L'acheteur menace le vendeur de partir s'il ne veut pas diminuer le prix de sa marchandise ou c'est le vendeur qui fait croire à l'autre que le prix qu'il propose est une bonne aubaine et qu'il doit l'accepter. (3) / *Manongom-barotra* : Enchérissement. Le vendeur augmente le prix sur lequel on s'est mis d'accord parce qu'il a trouvé un autre client qui pourrait acheter la marchandise à un prix plus intéressant pour lui. (4) / *Mivaro-tena* : Se prostituer. Gagner de l'argent avec son activité sexuelle. (5) / *Varotra alika* : litt. "vente de chien" Vente forcée. (6) / *Varotra ambongadiny* : Commerce de gros. Commerce où l'on vend les choses par gros lots, mais en petites quantités. (7) / *Varotra an'elakela-trano, varotra ankobohobo* : litt. "vente dans les petites ruelles", "vente derrière les maisons". Marché noir. Vente illicite pratiquée dans un endroit caché par un vendeur voulant vendre ses marchandises à un prix déraisonnable. (8) / *Varotra an-trano tokana* : Monopole. Commerce où l'on vend des choses qu'on ne peut pas trouver ailleurs. (9) / *Varotra antsingarany* : Commerce de détail. Vente de choses en petites quantités. (10) / *Varotra olona* : Traite d'esclaves. Vente de personnes destinées à l'esclavage. (11) / *Varotra omby anaty ambiaty* : litt. "vente de boeufs dans les ambiaty" Vente d'une chose qui n'est même pas visible. (12) / *Varotry ny mananotena* : Vente où c'est le vendeur qui hésite à savoir s'il peut empocher l'argent tout en gardant la marchandise. (13) / *Varo-bazaha* : litt. "vente à l'étrangère" Vente à prix fixe. Vente sans marchandage. Le prix est fixe et inchangeable. (14) / *Varo-boba* : Liquidation. Vente où les prix sont très bas. (15) / *Varo-dratsy* : litt. "vente méchante". Vente d'objets volés. (16) / *Varodroua* : Vente où deux personnes essaient de vendre la même chose. (Par exemple, deux agents immobiliers qui essaient de vendre une même maison.) (17) / *Varotra mahay mody* : litt. "vente connaissant le retour". Vente où l'on peut échanger ou retourner et se faire rembourser les marchandises. (18) / *Varo-maika na Varotry ny maika hiala, Varotry ny maika hody* : litt. "vente rapide" ou "vente des gens pressés (de rentrer)" Liquidation, vente à bon marché. (19) / *Varo-maizina* : Marché noir. Équivalant à l'expression "varotra an'elakela-trano". (20) / *Varo-mandeha* : Commerce ambulancier, colportage. Commerce de marchandises qui se fait en se déplaçant d'un endroit à l'autre, et non dans un seul et même endroit. (21) / *Varo-maty* : Vente définitive. Vente de marchandises qui ne peuvent être ni échangées, ni remboursées, et cela peu importe la raison. ♦ Figuré : la mort. (22) / *Varo-miandry* : litt. "vente dormante" Vente où les marchandises doivent être exposées tant qu'elles ne sont pas

⁷ Cette traduction est présentée suivant la typographie originale de RR, avec l'entrée lexicale en blocs compacts.

vendues. (23) / *Varo-miera* na *Varo-mifody* : Vente à l'essai. Commerce dont la politique consiste à rembourser les marchandises non satisfaisantes (24) / *Varotra misy fehi-vavany* : litt. "vente à bouche cousue" Vente conditionnée par des règles modifiables. (25) / *Varotra tsy miverimbodiondry* : Vente sans retour des arrhes (litt. "vente sans retour du derrière de mouton"). Transaction où l'acheteur doit donner une certaine somme d'avance pour garantir qu'il s'engage à acheter la marchandise mise de côté. Dans le cas où l'acheteur ne voudrait plus prendre la marchandise, son argent ne lui serait pas remis. (26) / *Varo-tsy mifody* : Vente définitive. Vente où les marchandises ne peuvent être retournées. ♦
Figuré : la mort.»

	vàrotra	n.	Vente, commerce, marchandises.
1 (< A&M 18, RR 20)	<i>Varo-mandeha,</i>		colportage.
2	<i>Tranom-barotra.</i>		maison, société de commerce.
3 (< RR 1)	<i>Ady varotra,</i>		marchandage, négociation.
4	<i>Fifanakalozam-barotra,</i>		échanges commerciaux.
5	<i>Miady varotra,</i>		marchander, négociier.
6 (< A&M 3)	<i>Tombom-barotra,</i>		bénéfice.
7 (< A&M 24, RR 26)	<i>Varotsimifody,</i>		litt. "vente sur laquelle on ne revient pas", l'au-delà, l'autre monde.
8 (< A&M 11, RR 14)	<i>Varo-boba.</i>		vente à vil prix, solde.
9	<i>Varotra ivelany.</i>		commerce extérieur.
10 (< A&M 19, RR 21)	<i>Varo-maty,</i>		vente définitive.
11 (< A&M 16, RR 18)	<i>Varo-maika,</i>		vente d'urgence.
12 (< RR 19)	<i>Varo-maizina,</i>		marché noir.
13 (< A&M 4, RR 6)	<i>Varotra ambongadiny,</i>		vente en gros.
14 (< A&M 6, RR 9)	<i>Varotra antsinarany,</i>		vente au détail.
15	<i>Vodiondrim-barotra,</i>		arrhes.
16	<i>Varotra iombonana,</i>		communauté économique.
17	<i>Varotra fampisehoana,</i>		exposition-vente.
18	<i>Lalam-barotra.</i>		débouché.
	♦ <i>Mi-</i> ,	v. a.	Vendre, trafiquer.
	♦ <i>Mpivarotra.</i>		commerçant, marchand.
19	<i>Mpivaro-kena,</i>		boucher.
20	<i>Mpivaro-tena,</i>		prostituée.
21	<i>Mpivarotra amoron-dalana,</i>		marchand du bord des rues.
22	<i>Mpivaro-mandeha,</i>		marchand ambulant.
23	<i>Mpivarotra omby,</i>		marchand de bestiaux.
	♦ <i>Fivarotana,</i>		commerce.
24	<i>Fivarotam-boky,</i>		librairie.
25	<i>Fivarotam-panafody,</i>		pharmacie.
26	<i>Fivarotan-tena,</i>		prostitution.
	♦ Dupli.,		Varobarotra

Figure 3 : Varotra (Rajaonarimanana 1995 : 325)⁸

⁸ Dans la typographie originale de NR, les sous-entrées ne sont pas séparées par des retours à la ligne et toute l'entrée lexicale apparaît en un seul bloc, formaté en drapeau sous la vedette.

On note, outre la richesse et la précision de l'information lexicale présentée, le haut degré d'organisation de la présentation elle-même⁹.

Particularités des différents dictionnaires

Abinal & Malzac (1888) (A&M)

Chez A&M, les différentes sous-entrées sont distinguées systématiquement par l'utilisation de caractères gras pour la vedette de chaque sous-entrée (qui apparaît, formatée en drapeau, à la suite de la vedette), tandis que les différents types de sous-entrées (formes fléchies, formes dérivées, composés, expressions) ne sont pas distinguées typographiquement entre elles¹⁰. Les exemples, par contre, sont distingués typographiquement : ils sont en italiques et regroupés dans le même paragraphe que la sous-entrée qu'ils illustrent. On notera aussi, chez A&M, que pour les formes préfixées en *Ci-* (où $C \in \{m, f, mp\}$), le préfixe de verbalisation *mi-* (réalisant la fonction lexicale P_0 dans le vocabulaire des fonctions lexicales (FL) de la théorie sens-texte (TST), cf. Mel'cuk, Clas et Polguère (1995)) est présenté en premier, suivi des préfixes de nominalisation d'action *fi-* (réalisant la FL S_0 de TST) et de nominalisation d'acteur *mpi-* (réalisant la FL S_1 de TST), alors que pour les formes relatives (circonfixées), c'est la forme nominalisée *fivarotana* (circonfixée en *fi--ana* et réalisant une fonction lexicale complexe $S_{0_{rel}}$) qui apparaît en premier et la forme verbalisée *ivarotana* (circonfixée en *i--ana* et réalisant une fonction lexicale complexe $P_{0_{rel}}$) qui apparaît en second. La forme *mpivarotana*, théoriquement possible (et qui pourrait signifier quelque chose comme «la personne chez qui, pour qui ou à cause de qui l'on vend») n'est jamais lexicalisée. Cet ordre de

⁹ Nous avons quelque peu modifié la typographie de façon à mieux séparer les différentes sous-entrées et nous avons numéroté les composés, phrasèmes et exemples pour faciliter les références croisées d'un dictionnaire à l'autre. Nous avons également modernisé l'orthographe d'Abinal et Malzac (1888). Tous les accents aigus ont ainsi été remplacés par des accents graves, conformément à l'usage contemporain. Cet accent note l'accent d'intensité, particulièrement marqué en malgache. Dans l'orthographe usuelle, on ne note l'accent que s'il ne suit pas les lois régulières d'accentuation. Ces lois sont fort simples : (i) pour les monosyllabes, accent sur cette syllabe; (ii) pour les formes polysyllabiques, accent sur la pénultième en général et sur l'antépénultième pour certains mots dont la dernière syllabe est *kV*, *trV* ou *nV* (< *k*, *t*, *r*, *n* fermant cette syllabe en proto-malgache). Nous avons tout de même conservé ici les accents qui ne seraient pas notés dans l'orthographe usuelle. Dans A&M, l'accent est marqué systématiquement sur le dernier mot de la vedette, et uniquement sur celui-ci, aussi bien pour l'entrée principale que pour les dérivés ou les exemples. Même dans le cas où deux exemples sont coordonnés métalinguistiquement, seul le dernier mot du deuxième exemple porte l'accent : dans «Varo-mahay mody et Varo-mièra», *miera* est accentué (alors qu'il ne le serait pas en orthographe usuelle), tandis que *mody* ne l'est pas (*et* est ici une conjonction métalinguistique). Dans NR, seule la vedette principale est accentuée. Dans RR, seuls sont accentués les mots qui portent l'accent en orthographe usuelle.

¹⁰ En fait, on remarque que, dans la typographie d'A&M, les sous-entrées perçues comme plus régulières (formes fléchies ou dérivées) ne sont pas séparées par un retour à la ligne et les formes dérivées dans ces sous-entrées ne sont pas en caractères gras, mais en petites majuscules, comme les renvois. Ainsi, au début de l'article *varotra*, les formes *Fi—* et *MPI—*, qui sont des préfixes complexes de nominalisation régulièrement associés au préfixe complexe de verbalisation *MI—* (comme les préfixes complexes *fan(a)-* et *mpan(a)-* sont associés à *man(a)-*) sont groupées avec cette dernière dans un même paragraphe.

présentation (verbalisation avant nominalisation pour l'actif et nominalisation avant verbalisation pour le relatif ou circonstanciel) est assez systématique chez A&M et semble représenter une perception intuitive de l'importance ou de la fréquence des formes qui est conforme à nos observations.

Rajemisa-Raolison (1985) (RR)

Chez RR, les formes dérivées sont regroupées dans une rubrique spéciale de l'entrée. Elles apparaissent après le symbole «•,•» et l'abréviation *t.i.f.* (*teny iray fototra*, «mots de même racine»). Elles sont en caractères italiques gras et groupées en séries apparentées. Les séries sont séparées par des diagonales («/»). Les termes de ces séries ne sont pas explicités mais constituent autant de renvois aux entrées correspondantes¹¹. Les composés et les phrasèmes apparaissent en italiques, après le symbole «•» et l'abréviation *F.-p.* (*fombam-piteny*, «expressions» — littéralement «façons de dire») séparés par une diagonale «/». Les différents sens sont séparés par le symbole «♦». Les sens figurés sont indiqués par le symbole «♦» suivi de l'abréviation *h.o.* (*hevitra an'ohatra*, «sens figuré»)¹².

Notons enfin que RR est un dictionnaire illustré et qu'une entrée lexicale peut également faire référence à une ou plusieurs illustrations ou encore à un tableau¹³. Le lien entre entrée ou illustration est indiqué soit par une flèche pointant de l'entrée vers l'illustration, soit par une reprise, à l'intérieur de l'illustration, de la vedette de l'entrée. Nous n'avons pas tenu compte de cet aspect ici, puisqu'il s'agit de connaissances encyclopédiques plutôt que linguistiques, mais il est certain que dans une représentation multimédia, de telles représentations graphiques ont une grande importance. Une chose est certaine, comme l'institution RR est la seule à contenir de telles références, il n'y aurait aucun problème à coller celles-ci (ou, en fait, toute autre illustration pertinente) sous l'entrée lexicale appropriée¹⁴.

¹¹ Puisque, comme on l'a déjà mentionné, ce dictionnaire est le seul où les dérivés constituent des entrées autonomes apparaissant telles quelles à leur place alphabétique dans le dictionnaire.

¹² Les synonymes (il n'y en a pas dans cette entrée) apparaissent après le symbole «¶» et l'abréviation *M.h.* (*mitovy hevitra*, «sens équivalent»). Les proverbes (extrêmement nombreux et courants en malgache) apparaissent après le symbole «★» et l'abréviation *Ohab.* (*ohabolana*, «proverbe»).

¹³ Ainsi, pour *ohabolana* ('proverbe'), on a un tableau de proverbes classés par thèmes (RR, pp. 775-777), pour *varimbazaha* (litt. 'riz d'Européen', 'blé'), on a une photo de blé sur pied (RR, p. 1016).

¹⁴ Le lien entre information linguistique et information graphique est cependant loin d'être trivial dans RR, même pour un agent humain. Par exemple, à la même page (RR, p. 1016) on retrouve une photo de lémurien (un maki, *lemur catta*, *lemuroïde*) sans aucune entrée qui pointe vers elle, ni identification sur la photo. On doit donc parcourir l'ensemble des entrées de la page pour savoir laquelle est pertinente. Il s'agit de l'entrée «**Varika** a. : Karazana gidro kely, ny rambony misoratra mitapatapaka.» ('Espèce de petit lémurien dont la queue est striée.') Cette définition, qui correspond à la traduction qu'A&M donnent de *varika* («Petit Lémurien : le macoco à queue annelée, *Lemur catta*, L.») semblerait plutôt appropriée pour *maky* que pour *varika* (*lemur fulvus*). RR a peut-être repris A&M sans vérification. Il faut cependant noter que l'usage courant (surtout celui des citadins) ne respecte pas nécessairement celui des spécialistes.

Rajaonarimanana (1995) (NR)

Chez NR, l'organisation est plus simple que chez A&M ou RR. L'information lexicale est relativement moins riche : il y a beaucoup moins d'exemples, les dérivés ne sont pas mentionnés systématiquement et la forme de l'impératif (ici avec la consonne thématique *-t-*) doit être déduite de la forme du dérivé *fivarotana*. Par contre, on retrouve plus de termes associés aux réalités contemporaines (cf. les items 4, 9, 16, 17, 18) ainsi que des composés ou des phrasèmes eux-mêmes construits sur les dérivés (cf. les items 19 à 26). C'est cette récursivité qui rend les entrées, en apparence assez simples, difficiles à analyser aussi bien pour l'humain que pour un programme. En ce qui concerne l'organisation typographique, les composés et phrasèmes apparaissent en italiques minces, tandis que les dérivés apparaissent en italiques grasses, précédés du symbole «♦». Les dérivés de dérivés apparaissent en principe précédés du symbole «•» (mais cet usage n'est pas systématique).

Recoupements des différents dictionnaires

On peut déjà constater à partir de cet exemple que la tradition lexicographique malgache ne le cède en rien à la tradition lexicographique française qui consiste à copi(er) systématiquement ses prédécesseurs et ce, non seulement pour la nomenclature (où les données linguistiques elles-mêmes font qu'il est impossible d'échapper aux répétitions), mais pour les définitions et les exemples. Ainsi les items 2, 3, 6, 8, 9, 11 à 22 et 26 de RR reprennent essentiellement A&M, tandis que l'item 6, les items 3 et 12, les items 1, 7, 8, 10, 11 13 et 14 de NR reprennent respectivement A&M, RR et à la fois A&M et RR.

Sous des apparences superficiellement assez distinctes et des organisations diverses (syntaxiquement et typographiquement), on sent bien que ces dictionnaires disent à peu près la même chose sur un certain nombre de points et qu'ils se complètent sur d'autres. Ces différentes entrées contiennent essentiellement un noyau d'informations communes et des satellites spécifiques. C'est cette intuition que nous essayons de mettre en oeuvre. Ceci suggère la possibilité de poursuivre la tradition et de reprendre ces différentes données, en les fusionnant dans une base lexicale ouverte.

Syntaxe des entrées lexicales dans les dictionnaires traditionnels

Examinons de plus près la forme de ces entrées. Nous utiliserons des expressions régulières de façon à pouvoir en caractériser succinctement la forme générale¹⁵.

¹⁵ Ces expressions régulières ne caractérisent pas directement les entrées lexicales, mais en constituent une description abstraite et schématique en ce sens que les symboles qui y apparaissent sont aussi bien des catégories abstraites (par exemple *L*, *R*, *C*, *T*, etc.) que des symboles spécifiques (par exemple ♦, •, ♦•, *t.i.f.*, •, etc.). De plus, nous ne tenons pas compte de certains détails de typographie, de ponctuation, etc. Les opérateurs propres employés dans ces expressions seront :

- l'opérateur unaire postfixé dit «étoile de Kleene» («*») qui signifie que le terme précédent peut-être répété un nombre indéfini de fois (y compris aucune);
- l'opérateur unaire postfixé («+») qui signifie que le terme précédent peut-être répété un nombre indéfini de fois (mais au moins une : $x^+ = xx^*$);

Abinal & Malzac (1888) (A&M)

L'entrée de *varotra*, de même que presque toutes les entrées dans A&M¹⁶ peut être décrite abstraitement au moyen de l'expression régulière suivante :

L (**R**) (**C** **T*** (**E** **T**)* (**F** **C**)* **N**)*

(D (**C**) **T*** (**E** **T**)* (**F** **C**)* **N**)*

(P (**C**) **T*** (**E** **T**)* (**N**))*

où

- L** est une vedette,
- R** est une racine¹⁷,
- C** une catégorie,
- T** une traduction,
- E** un exemple d'emploi,
- F** une forme fléchie (généralement l'impératif¹⁸),
- N** une annotation ou un renvoi,
- D** un affixe de dérivation ou une forme dérivée,
- P** une expression (ou un proverbe).

— l'opérateur n-aire de disjonction («{ $-_1 \mid -_2 \dots$ }») qui indique un choix exclusif entre les termes $-_1, -_2, \dots$ qu'il renferme;

— l'opérateur implicite de concaténation.

Rappelons que les parenthèses sont des opérateurs impropres qui ont deux rôles :

- 1°) indiquer l'optionnalité d'un terme et
- 2°) regrouper des termes en un seul.

¹⁶ Les entrées qui ne peuvent pas être décrites par ces expressions régulières sont des entrées exceptionnelles. Ou bien on y trouve une combinaison booléenne de termes, ou bien des termes apparaissant dans une position exceptionnelle. Par exemple «*HALAVÏRINA, HALAVÏTINA*, p. de Lavitra.» (p. 209, conjonction de vedettes), «*LÈMBA* (trivial), *adj.* Mort.» (p. 399, annotation immédiatement après la vedette) ou encore «*KÒTAKA* et *KOTAKÒTAKA* (*dupl.* plus usité), s. Tumulte, turbulence, révolution. Comme *HOTAKOTAKA*.» (p. 361, conjonction de vedettes et annotation immédiatement après l'entrée). Tenir compte de ces cas ne pose aucun problème de principe, mais aurait compliqué indûment l'expression.

¹⁷ Cette racine apparaît (entre parenthèses) quand la vedette n'est pas elle-même une forme racine, mais une forme dérivée idiosyncratique (par exemple «*KITRANOTRÀNO* (*trano*), s. Petites maisonnettes que font les enfants.» (p. 340). La dérivation en *ki-duplicatif* (ou *tsi-duplicatif*) pour indiquer un jeu impliquant le sens de la racine (ici *trano* : 'maison') est idiosyncratique) ou opaque (par exemple «*MÀINA* (*haina*), *adj.* Sec. desséché, tari; *fig.* désappointé, déchu, surpris, qui n'a rien à dire. [...]» (p. 419), «*MÀINTY*, (*inty*) *adj.* Noir. [...]» (ibidem), «*TSATSÏAKA* (*tsiaka*), s. Action de déchirer avec bruit, de faire du bruit en se déchirant, en se fendant. [...]» (p. 756). Le dérivé duplicatif attendu de *tsiaka* serait *tsiatsiaka*, qui existe effectivement et apparaît sous l'entrée de *tsiaka*).

¹⁸ L'impératif est une forme importante, puisqu'il est relativement imprédictible et que l'allomorphe du radical (incluant souvent une consonne finale) sur lequel il est formé sert de base dans les dérivations suffixales (et circonfixiales).

On remarque que cette description est divisée en trois zones :

(a) l'entrée principale :
 $L (R) (C T^* (E T)^* (F C)^*) N^*$,

(b) les dérivés :
 $(D (C) T^* (E T)^* (F C)^* N^*)^*$

et

(c) les expressions (ou proverbes) associés :
 $(P (C) T^* (E T)^* (N))^*$.

Ainsi sont bien distinguées par leur position respective les informations globales sur l'item lexical (sa catégorie, ses traductions, ses exemples d'emploi avec leurs traductions, ses formes flexionnelles, ses liens sémantiques à d'autres items, etc.), qui apparaissent dans la première zone et les informations morphologiques ou phraséologiques sur ses dérivés ou ses expressions associées qui apparaissent respectivement dans sa deuxième et sa troisième zone¹⁹.

Rajemisa-Raolison (1985) (RR)

Les entrées de RR peuvent être décrites par l'expression régulière suivante :

$L (d) (R) (C) (S (: E^+)) (\diamond S (: E^+))^*$
 $(\bullet, \bullet \text{ t.i.f. } D^+ (/ D^*)^*)$
 $(\bullet \text{ F.p. } : P (: S (: E^+) \diamond H.o. S) (/ P (: S (: E^+) \diamond H.o. S)^*)^*)$
 $(\star \text{ Ohab. } (O : S)^+)$
 $(\P M.h. : M^+)$

où

L	est une vedette,
d	est une indication dialectale ou étymologique ²⁰ ,
R	est une racine,
C	une catégorie,
S	une définition,
E	un exemple d'emploi,
D	une forme dérivée,
P	une expression,
M	une forme synonyme,
♦	le délimiteur des significations différentes,

¹⁹ Ces informations, de même que les liens facultatifs vers la racine ou vers des synonymes globaux de la première zone, constitueraient autant d'hyperliens vers d'autres formes dans une représentation en hypertexte.

²⁰ Par exemple : Marotoa (*fr. marteau*) a. : Tantanam-by kely. ('petit marteau de fer').

- ,• *t.i.f.* le délimiteur de la liste des dérivés (*teny iray fototra*, «mots de même racine»),
- *F.p.* le délimiteur de la liste des expressions (*fombam-piteny*, «manières de parler»),
- ★ *Ohab.* le délimiteur de la liste des proverbes (*ohabolana*)

et

- ¶ *M.h.* le délimiteur de la liste des synonymes²¹ (*mitovy hevitra*, «même sens»).

Chacun des termes complexes de l'expression régulière correspond donc à une zone définie :

- (a) l'entrée principale :

$$L(d)(R)(C)(S(:E^+))(\diamond S(:E^+))^*$$
- (b) les dérivés :

$$(\bullet, \bullet \text{ t.i.f. } D^+ (/ D^*)^*),$$
- (c) les expressions associées :

$$(\bullet F.p. : P(:S(:E^+) \diamond H\alpha S)(/ P(:S(:E^+) \diamond H\alpha S)^*)^*),$$
- (d) les proverbes associés :

$$(\star Ohab. (O : S)^+),$$
- (e) les synonymes (et antonymes) :

$$(\parallel M.h. : M^+)$$

On remarque que les zones sont plus nombreuses et plus faciles à identifier, du fait qu'elles sont systématiquement indiquées au niveau typographique.

²¹ Les antonymes (*mifanohi-kevitra*), peu nombreux, sont aussi notés sous cette rubrique. On remarque qu'il n'y a pas de lien entre cette liste et la liste des définitions dans l'entrée principale. Généralement, cette liste fait référence au sens premier du terme (i.e. le premier élément S de l'entrée principale). Comme les homonymes font l'objet d'entrées distinctes, ceci ne pose pas de problème (du moins pas à un humain).

Rajaonarimanana (1995) (NR)

Quant aux entrées dans NR, elles peuvent être décrites par l'expression régulière suivante :

$$L \quad (R) (C) (T) ((n)(C) (n)(T) (F (C) (T)) \{ (E T) | N \}^* \\ (\diamond \{ D | Dupli.(D) \} (C) (T) ((n)(C) (n)(T) (F (C) (T)) \\ \{ (E T) | N \}^* | (\bullet D (C) (T) ((n)(C) (n)(T) (F (C) (T)) \{ (E T) | N \}^*)^*)^*$$

où

L	est une vedette,
R	est une racine,
C	une catégorie,
T	une traduction,
n	un numéro de sous-entrée,
F	une forme fléchie (généralement l'impératif),
E	un exemple d'emploi ou une expression associée,
D	une forme dérivée,
Dupli.	l'indication d'une forme redoublée,
N	une annotation indiquant un synonyme, un antonyme, une forme dialectale, une variante, un renvoi, etc.,
♦	le délimiteur des dérivés,
•	le délimiteur des sous-dérivés.

Cette expression est plus complexe que les précédentes du fait que l'entrée n'est pas organisée linéairement en zones distinctes, mais hiérarchiquement en sous-entrées²². Cette récursivité (que des dictionnaires rigoureux comme le DEC (Mel'cuk et al. 1984, 1988, 1992; Mel'cuk, Clas et Polguère 1995) excluent par principe) pose des problèmes de lecture et d'interprétation assez difficiles. De plus, les exemples d'emploi ne sont pas distingués typographiquement des expressions.

²² Il serait donc plus simple de représenter les entrées de NR au moyen d'un réseau de transitions récursif ou d'une grammaire hors-contexte, par exemple :

ENTRÉE → Lex (Rac) Cat Trad (FORME) (ILLUSTRATION) (DÉRIVÉS) (SOUS-ENTRÉES)

FORME → Lex (Cat) (Trad)

ILLUSTRATION → (Ex Trad) (Note) (ILLUSTRATION)

DÉRIVÉS → DÉRIVÉ (SOUS-DÉRIVÉS) (DÉRIVÉS)

DÉRIVÉ → ♦ SOUS-ENTRÉE

DÉRIVÉ → ♦ Duplicatif (SOUS-ENTRÉE)

SOUS-DÉRIVÉS → • DÉRIVÉ (SOUS-DÉRIVÉS)

SOUS-ENTRÉES → SOUS-ENTRÉE (SOUS-ENTRÉES)

SOUS-ENTRÉE → (n) Lex (Cat) (Trad) (FORME) (ILLUSTRATION)

où Lex est une forme quelconque, Rac est une racine, Trad est une traduction, Ex est un exemple, Note est une annotation quelconque, Duplicatif est une indication de duplicatif, n est un nombre.

Fusion et filtrage

Si l'on disposait de ces dictionnaires sous format électronique (ce qui n'est pas le cas, notons-le), on pourrait, étant donné une telle caractérisation, les fusionner en une base lexicale complexe gardant (ou non) la trace de la source de ses données²³.

On se contentera d'utiliser ici la technique classique et heuristiquement productive du *fantasme*. C'est-à-dire que l'on présupposera tout simplement que l'on dispose de ces dictionnaires sous format électronique. En pratique, le problème de saisie se combine à celui du *filtrage*. En fait, grâce aux outils que nous avons développés, les deux opérations peuvent être combinées jusqu'à un certain point. Plutôt que de saisir globalement chacun des dictionnaires, pour les filtrer ensuite, on en a saisi des fragments suffisants pour construire semi-automatiquement un squelette de base de données lexicales contenant plus de 10 000 entrées de base (racines ou dérivés). Ce squelette a lui-même été retravaillé de diverses manières et a donné lieu à plusieurs expérimentations. Il est actuellement enrichi semi-automatiquement de façon systématique à partir de données contenues dans les dictionnaires traditionnels (cf. outre A&M, NR et RR, Malzac 1908; Profita 1969; Rakotonaivo 1996), et dans d'autres institutions lexicales, comme les lexiques-grammaires (Rabenilaina 1987, 1996). Chaque entrée lexicale de chacun des dictionnaires peut être analysée selon la grammaire propre au dictionnaire, puis standardisée selon un schéma commun. Les différents schémas ainsi obtenus peuvent alors être fusionnés en un schéma unique²⁴. C'est précisément à la caractérisation de tels schémas que serviront les institutions lexicales.

Nous avons développé, à titre expérimental, quelques maquettes de programmes d'analyse et de fusion pour les entrées lexicales dans ces différents formats. On trouvera, en annexe, le résultat de l'application d'un de ces programmes aux trois entrées de *varotra*.

²³ La saisie de ces textes pose toutes sortes de problèmes techniques. Dans le cas d'A&M, par exemple, l'impression la plus récente (1993) reproduit photostatiquement l'édition de 1888. Une saisie par lecteur optique est donc pratiquement exclue dans ce cas, du fait de la faible qualité graphique de l'original. Par ailleurs, il y a les délicats problèmes de droits d'auteur posés par les dictionnaires récents. Certains considèrent que même leur nomenclature devrait être protégée, ce qui est soit vide de sens (ajouter ou enlever un seul terme ferait que l'on aurait une autre nomenclature, ou alors il faudrait établir des mesures de distance entre nomenclatures, en précisant des seuils au-delà desquels des nomenclatures seraient distinctes), soit ridicule (tout nouveau traitement d'un terme déjà couvert par un dictionnaire serait exclu). Comme la nomenclature, les traductions et même les exemples sont souvent eux-mêmes empruntés à A&M, ces problèmes sont loin d'être évidents. Le fait de noter la source des données permettrait probablement de contourner certains problèmes de droits d'auteur, à condition de ne pas en faire un usage commercial, bien sûr.

²⁴ On pourrait penser que les techniques classiques d'alignement (Gale et Church 1993; Blank 1995) seraient utiles pour appairer les entrées correspondantes des différents dictionnaires. Tel n'est pas le cas. En effet, l'équivalence entre dictionnaires n'est pas du tout du même type que l'équivalence entre un texte et sa traduction. Un dictionnaire n'est pas la «traduction» d'un autre, même si les emprunts, voire les traductions (cf. RR, qui utilise parfois comme définition d'un terme malgache, la retraduction en malgache de l'équivalent français du terme donné par A&M) sont fréquents. Il n'y a pas de correspondance statistiquement significative entre dictionnaires. L'appariement, s'il est possible, doit se faire de façon symbolique.

3. INSTITUTIONS LEXICALES

Une *institution* est l'ensemble des prédicats implicites ou explicites (mais le plus souvent implicites) qui caractérisent un cadre théorique ou descriptif²⁵. Appliquée au domaine lexical, cette notion suggère l'idée suivante. Si l'on sait de quoi une institution lexicale (i.e. un type de lexique, de dictionnaire, etc.) parle et si l'on sait traduire d'une institution à l'autre, on peut fusionner des informations provenant de plusieurs institutions lexicales²⁶. Autrement dit, si la sémantique est claire, la syntaxe (comme l'intendance) devrait pouvoir suivre. Ainsi, on sent bien que, sous les divergences de forme, les trois dictionnaires traditionnels que nous avons mentionnés parlent bien (du moins partiellement) des mêmes choses. Si l'on arrive à déterminer de façon plus précise quel est le domaine couvert par chacun de ces types de description, on pourra déterminer les conditions permettant de les fusionner.

Il y a deux bonnes nouvelles ici. La première, qui peut sembler paradoxale, c'est que si deux institutions ne parlent pas de la même chose (i.e. si leurs domaines sont entièrement distincts, s'il n'y a pas d'interface entre elles), on peut toujours les combiner. Par exemple, si chacun des dictionnaires traditionnels contenait des informations entièrement distinctes, on pourrait tous les combiner facilement. Évidemment, pour qu'une combinaison soit intéressante, il faut qu'il y ait une interface²⁷. Cependant cette interface n'a pas à être *totale*. Les parties d'une institution dont le domaine n'a pas d'équivalent dans une autre institution sont combinables à cette dernière sans problème. La deuxième, c'est que, lorsque deux institutions parlent effectivement de la même chose, on peut soit fusionner les informations, lorsqu'on sait qu'elles sont compatibles et qu'on sait les fusionner, soit tout simplement les cumuler, en indiquant la source²⁸ quand on ne sait pas si elles sont compatibles ou comment les fusionner.

Ainsi, si l'on sait

- (a) que l'institution lexicale A&M attribue à une entrée dont la VEDETTE est **VÀROTRA** la CATÉGORIE *s.* et la TRADUCTION «Vente, commerce, négoce, trafic, marchandises.»,

²⁵ Cf. Blache & Morin (1997).

²⁶ L'indéfini (*on, l'on*) dans ces énoncés est voulu. Les agents peuvent être humains, machines, ou les deux. En fait, ce que la philosophie des institutions lexicales favorise, ce sont justement des traitements interactifs, coopératifs, partiellement automatisés. Ceci permet une très grande souplesse (et une indépendance totale) par rapport aux grandes entreprises de normalisation (HTML, SGML, XML, etc.).

²⁷ On peut imaginer des cas de figure où un tel assemblage hétéroclite aurait de l'intérêt, une encyclopédie, par exemple, où les articles de littérature comparée et ceux de géologie auraient peu de chance de porter sur des domaines en intersection (encore que les formes — strates, couches, etc. — puissent être en intersection). Mais une véritable encyclopédie ne saurait se limiter à ces deux domaines et il est probable, que, de proche en proche, les domaines soient finalement connexes.

²⁸ Cette indication sert, en quelque sorte, de ceinture de protection contre l'incohérence. Associer à une entité E deux propriétés P et P' qui sont incompatibles produit une entité incohérente. Mais associer à E deux propriétés X(P) et Y(P') peut permettre d'éviter cela jusqu'à ce qu'un agent plus compétent parvienne à les concilier.

- (b) que l'institution lexicale NR attribue à une entrée dont la VEDETTE est **vàrotra** la CATÉGORIE n. et la TRADUCTION «Vente, commerce, marchandises.»,
- (c) que l'institution lexicale RR attribue à une entrée dont la VEDETTE est **Varotra** la CATÉGORIE a. et les DÉFINITIONS «Fanakalozan-javatra amin'ny vola, hitadiavan-tombony» et «Entana amidy»

et que l'on sait

- (d) identifier les vedettes **VÀROTRA** de A&M, **vàrotra** de NR et **Varotra** de RR et les fusionner en une vedette 'varotra',
- (e) identifier les catégories s. de A&M, n. de NR et a. de RR et les fusionner en une catégorie 'N',

mais que l'on ne sait pas (pour le moment²⁹)

- (f) fusionner «Vente, commerce, négoce, trafic, marchandises.» de A&M et «Vente, commerce, marchandises.» de NR (ni même si elles sont compatibles)

ni

- (g) si les TRADUCTIONS de A&M et NR sont compatibles avec les DÉFINITIONS de RR,

on pourra déjà créer, dans une institution lexicale A&MNRRR, une entrée lexicale partielle comme la suivante³⁰ :

²⁹ Cette qualification est importante. Les institutions lexicales sont des entités souples et dynamiques. Si, par la suite, un agent (humain ou non) qui sait fusionner ces définitions intervient, il pourra les modifier. Par exemple, un lexicographe ou un programme maximaliste, qui ferait l'union des termes de deux ensembles de traductions pourrait fusionner ces deux ensembles en un seul, équivalent au premier (puisque le deuxième est un sous-ensemble du premier). Un agent qui pourrait faire correspondre une définition en langue naturelle et une définition de type DEC pourrait produire la définition et les fonctions lexicales suivantes. La définition est donnée en pseudo-malgache métalinguistique, suivie d'un équivalent en pseudo-français métalinguistique, les indices indiquent des étiquettes sémantiques qui constituent autant de contraintes de type (cf. Miličević 1997).

Définition

fiasa an'ny fiaraha-monina : ~ Y amin' Z amin' W X
 entana olona vola olona
 (action sociale : X ~ Y à Z pour W
 personne marchandise personne argent)
 Fonctions lexicales : S₀(varotra) = varotra, S₂(varotra) = varotra

³⁰ La forme spécifique de la représentation n'a pas d'importance à ce niveau. En fait, les objets créés dans nos maquettes sont des objets prolog, que l'on pourrait aisément traduire en objets HTML, SGML ou XML au besoin. On se rend compte qu'une bonne partie du travail créateur de traitement de l'information (qu'elle soit lexicale ou non) consiste à manipuler et à expérimenter avec des représentations diverses. Le fait d'utiliser des représentations prolog pour les institutions lexicales facilite ce travail. Les institutions lexicales sont des entités plus abstraites et plus souples (et surtout moins lourdes) que les DTD de SGML, par exemple, même s'il y a évidemment des liens entre les deux. Ce ne sont pas des normes, mais des caractérisations abstraites des domaines sémantiques recouverts par des objets descriptifs. Elles

VEDETTE	varotra	
CATÉGORIE	N	
TRADUCTIONS	"Vente, commerce, négoce, trafic, marchandises."	(A&M)
	"Vente, commerce, marchandises.",	(NR)
DÉFINITIONS	"Fanakalozan-javatra amin'ny vola, hitadiavan-tombony"	(RR)
	"Entana amidy"	(RR)

Figure 4 : Entrée lexicale partiellement fusionnée pour *varotra*

Quels sont donc les domaines recouverts par ces trois dictionnaires ? Quelle est la structure conceptuelle des entrées lexicales qu'on y retrouve ? Une entrée lexicale complète³¹ se compose obligatoirement d'une VEDETTE, d'une CATÉGORIE et d'une liste de TRADUCTIONS ou de DÉFINITIONS³². Toute entrée lexicale se construira donc autour d'un squelette [VEDETTE,CATÉGORIE,TRADUCTIONS ou DÉFINITIONS]. Les éléments de ce squelette forment respectivement le noyau des trois zones [PHONOLOGIQUE/GRAPHIQUE,SYNTAXIQUE,SÉMANTIQUE].

Évidemment, un tel squelette est très incomplet, particulièrement aux niveaux SYNTAXIQUE et SÉMANTIQUE. Tout ce qu'il contient au niveau SYNTAXIQUE, c'est une *partie du discours*. Bien qu'une partie du discours constitue un type syntaxique pauvre, ce caractère peu contraignant lui permet de se combiner librement avec de l'information syntaxique plus riche (liste et formes d'arguments) provenant d'une autre source³³. Quant à la SÉMANTIQUE, une traduction ou une définition n'est évidemment pas une représentation satisfaisante. Mais, suivant en cela notre tendance au retardement des tâches difficiles (équivalent en quelque sorte à de l'évaluation paresseuse), on considère que les traductions ou les définitions ne sont que des *noms* que nous attribuons à la représentation sémantique

sont orientées vers des traitements spécifiques interactifs où chacun des agents (humain ou machine) est chargé de tâches pour lesquelles il est mieux adapté (les tâches pouvant être redistribuées dynamiquement) et non vers des traitements de masse entièrement automatisés.

³¹ Une entrée lexicale incomplète est une entrée où l'une ou l'autre de ces rubriques manque. Dans le cas des entrées lexicales principales, la VEDETTE est toujours présente, mais pour les sous-entrées (qui constituent les DÉRIVÉS), elle peut être absente ou suggérée par des affixes (*mi-*, *fi-*, *mpi-*, etc.) ou des opérations (*Duplicatif*). Les autres cas d'entrées lexicales incomplètes sont des cas où la CATÉGORIE et/ou les TRADUCTIONS/DÉFINITIONS sont absentes. Dans les entrées principales, elles sont remplacées par un simple RENVOI (un type particulier d'ANNOTATION) à une entrée lexicale complète. Dans les sous-entrées, elles sont implicites (et généralement reconstituables à partir des affixes ou des opérations).

³² TRADUCTIONS et DÉFINITIONS sont des composantes conceptuellement équivalentes, puisqu'on peut voir une définition comme une traduction dans un métalangage de définition qui peut utiliser soit le vocabulaire et la syntaxe de la langue décrite elle-même, soit celui d'une autre langue. En fait, elles correspondent l'une ou l'autre à la zone sémantique de l'entrée.

³³ Par exemple, si l'on sait qu'un lexique-grammaire (Rabenilaina 1987, 1996) contient de l'information sur la distribution syntaxique des arguments d'un prédicat et que les dictionnaires traditionnels ne contiennent aucune information sur ce domaine, on peut tout simplement coller l'information distributionnelle provenant du lexique-grammaire et l'information provenant des dictionnaires traditionnels. C'est ce que nous examinons dans un autre travail (Morin 1997).

effective. Que celle-ci existe ou pas n'invalide pas l'usage de ces chaînes comme des noms, ni l'usage de ceux-ci pour distinguer des acceptions.

Une entrée lexicale ne contenant que ce squelette est une *entrée lexicale simple*. Une entrée lexicale peut aussi être composite et comporter un ensemble de rubriques facultatives. Elle peut contenir une liste de DÉRIVÉS, une liste d'EXPRESSIONS, une liste d'EXEMPLES et/ou une liste d'ANNOTATIONS. On ajoutera donc implicitement au squelette ci-haut de telles rubriques de façon à obtenir pour chaque entrée lexicale, une structure virtuelle comme la suivante :

VEDETTE	<i>vedette</i>
CATÉGORIE	<i>catégorie</i>
TRADUCTIONS/DÉFINITIONS	<i>liste de traductions/définitions</i>
DÉRIVÉS	<i>liste de dérivés</i>
EXPRESSIONS	<i>liste d'expressions</i>
EXEMPLES	<i>liste d'exemples</i>
ANNOTATIONS	<i>liste d'annotations</i>

Figure 5 : Squelette d'une entrée lexicale fusionnée

EXPRESSIONS et EXEMPLES

Les entrées des dictionnaires traditionnels peuvent contenir des listes d'EXPRESSIONS associées et des listes d'EXEMPLES d'emploi. Les EXPRESSIONS et EXEMPLES sont distingués typographiquement des DÉRIVÉS dans chacun des trois dictionnaires. Chaque élément de ces EXPRESSIONS ou EXEMPLES est formé d'une VEDETTE, elle même associée à une liste de TRADUCTIONS/DÉFINITIONS et/ou à une liste d'ANNOTATIONS³⁴. Les EXPRESSIONS sont distinguées typographiquement des EXEMPLES

³⁴ Pour les EXPRESSIONS, on a généralement aussi une CATÉGORIE chez A&M. On notera aussi que les PROVERBES sont traités comme des EXPRESSIONS par A&M, alors que RR les isole en fin d'entrée (avant la liste des synonymes) et leur donne un statut particulier. Comme nous ne nous intéressons ici qu'à la dimension strictement linguistique, nous n'avons pas créé de rubrique particulière pour les proverbes qui, du fait de leur structure particulière, seront classés avec les exemples d'emploi.

aussi bien dans A&M³⁵ que dans RR³⁶, mais pas dans NR³⁷. Les EXPRESSIONS peuvent être formées d'un seul mot-forme, ou d'une séquence de mots-formes et être plus ou moins compositionnelles.

Expressions compositionnelles, semi-compositionnelles et non compositionnelles

À propos de la distinction entre *expressions compositionnelles* (dont le sens {est représentable, peut être reconstruit} à partir de celui des composantes et de règles générales, *expressions non compositionnelles* (dont le sens \b\bc\{(\a\ac(n'est pas représentable, ne peut pas être reconstruit)) à partir de celui des composantes et de règles générales) et *expressions semi-compositionnelles* (dont le sens \b\bc\{(\a\ac(n'est représentable, ne peut être reconstruit)) que partiellement à partir de celui des composantes et de règles générales), il y a plusieurs choses à dire.

Premièrement, elle est souvent assez difficile à tracer, même dans des cas simples. Par exemple, si *mpivaro-boky* (*mpi-varotra-boky*, 'AGENT-commerce-livre' = 'libraire') est compositionnel, qu'en est-il de *mpivarotena* (*mpi-varotra-tena*, 'AGENT-commerce-{personne, corps, substance, éléments d'une chose, PRONOM RÉFLÉCHI}' = 'prostitué(e)') ? On voudrait certainement dire que ce dernier est «plus» compositionnel que *mitena* (*mi-tena*, 'VERBEPRÉSENT-{personne, corps, substance, éléments d'une chose, PRONOM RÉFLÉCHI}' = 'porter tous les jours les mêmes habits'), qui ne serait pourtant pas nécessairement totalement non compositionnel. Il y aurait donc des degrés de compositionnalité à établir.

Deuxièmement, si l'on savait construire, à partir des définitions présentes dans les institutions sources, une définition dans un langage standard, les (sous-)entrées correspondantes ne contiendraient que de l'information *héritée* pour les expressions *compositionnelles*, que de l'information *spécifiée* pour les expressions *non compositionnelles* et un mélange des deux pour les expressions *semi-compositionnelles*. Dès lors, les degrés de compositionnalité correspondraient au rapport entre (nombre de)

³⁵ Où la **vedette** de chaque expression est en caractères gras et débute un nouveau paragraphe, alors que les exemples sont en minces et viennent immédiatement à la suite de la forme qu'ils illustrent.

La forme illustrée par l'exemple y est marquée en italiques. Ainsi dans l'entrée de *varotra* ci-haut, la forme **ivarotana** est illustrée par un exemple :

«**Ivarõtana**, *rel.* Ariary telopolo no *nivarotako* ny omby mifahiko. J'ai vendu trente piastres le bœuf que j'avais engraisé.»

où *nivarotako* (= n_{PASSÉ}+ivarotana+ahop_{PRONOM}, 1 P.S.) est la forme du passé (préfixe *n-*), avec AGENT enclitique de première personne du singulier d'*ivarotana*.

³⁶ Où les expressions sont regroupées, après le délimiteur «• F.p.» et sont notées, avec leur *vedette* en italiques, immédiatement après le séparateur «/», et avant une définition et/ou des annotations en caractères minces, alors que les *exemples* sont entièrement en italiques.

³⁷ Où *expressions* et *exemples* d'emploi sont tous en italiques et non démarqués les uns des autres, avec les traductions ou annotations pertinentes en caractères minces. En fait, comme les EXEMPLES sont très peu nombreux dans ce dernier dictionnaire, on peut se contenter d'une procédure très simple pour les isoler, quitte à reclassifier certaines formes par la suite. Dans un premier temps, la procédure que nous utilisons, qui est efficace quel que soit l'agent qui l'effectue (humain ou non), consiste à considérer comme un EXEMPLE uniquement les entrées dont la VEDETTE a la forme extérieure d'une phrase, c'est-à-dire se termine par une ponctuation forte, ou contient une ponctuation quelconque.

composantes héritées et (nombre de) composantes spécifiées. Cependant, cette opération est, en général, très difficile même pour les agents humains (et *a fortiori* pour les machines).

Troisièmement, dans une base lexicale à accès multiples, cette distinction perd beaucoup de son importance. En effet, contrairement à la compositionnalité syntaxique, où une expression syntaxique compositionnelle est aussi bien analysable que synthétisable et n'a pas à être apprise individuellement, dans la compositionnalité lexicale en général, une expression lexicale compositionnelle analysable ou synthétisable de façon univoque doit être apprise. Ainsi, même si *mpivarokena* (*mpi-varotra-hena*, 'AGENT-commerce-{viande, chair}') = 'boucher') est une expression compositionnelle, le locuteur malgache doit apprendre qu'elle a bien ce sens et que c'est celle-là et pas une autre qui a ce sens particulier³⁸. Donc, compositionnel ou non, tout sens lexical doit être appris et stocké quelque part. La seule différence éventuelle, quand on sera en mesure de représenter systématiquement le sens lexical, sera que le sens des expressions compositionnelles sera représentable de façon beaucoup plus compacte que le sens des expressions non compositionnelles, sera plus facile à apprendre, à reconstruire en cas de perte, etc.

DÉRIVÉS

Une entrée lexicale peut aussi contenir une liste de DÉRIVÉS. On a vu que si celle-ci n'est qu'une simple liste structurée de formes, qui constituent autant de renvois, chez RR, elle peut être une liste de sous-entrées chez A&M. Elle peut même être récursive chez NR et contenir des dérivés de dérivés. Dans une base lexicale électronique (comme dans un dictionnaire formalisé du type DEC), cette récursivité sera éliminée, puisqu'il n'y aura pas de sous-entrées. On ne conservera que des *liens*, comme chez RR. Cependant, lors de la création de la base, il est plus pratique de suivre pas à pas les entrées lexicales des institutions sources et de créer des entrées complexes, qui pourront être «aplaties» par la suite. Un DÉRIVÉ est donc tout simplement une (sous-)entrée³⁹ apte à se détacher de son entrée d'origine en acquérant des liens bidirectionnels⁴⁰.

Les DÉRIVÉS se divisent en dérivés par *affixation* et dérivés par *réduplication*. Les deux types sont bien distingués aussi bien chez A&M, que chez NR⁴¹. Ce qui varie de

³⁸ De même, en français, si *fruitier* 'marchand de fruits' est compositionnel (par opposition à *boucher*, qui est non compositionnel), il faut apprendre que cette expression existe effectivement avec ce sens, alors que *légumier*, qui serait tout aussi compositionnel, n'a le sens de 'marchand de légumes' qu'en Belgique, à notre connaissance.

³⁹ Chez A&M, les sous-entrées des dérivés ne contiennent généralement pas de composés, ni de phrasèmes associés. Mais ils peuvent contenir des exemples. Chez Rajaonarimanana, les dérivés forment des sous-entrées de plein droit, comme on peut le voir ci-dessus pour *mpivarotra* et *fivarotana*.

⁴⁰ Du dérivé à la forme source (généralement une racine, mais éventuellement un dérivé) et de la forme source au dérivé.

⁴¹ Chez RR, comme la liste des dérivés est une simple liste de formes complètes, on doit disposer d'un analyseur morphologique pour identifier le type morphologique de chacune des formes et sa décomposition. Comme la dérivation est assez complexe, mais régulière, en malgache, nous avons pu construire un tel analyseur que nous comptons appliquer aux listes de formes dérivées de RR (saisies manuellement). En fait donc, il devrait être possible de reconstruire la forme complète d'un dérivé à partir de sa source et de son type, tout comme de reconstruire son type à partir de sa forme complète et de celle de sa source. Quant à la sémantique

dictionnaire à dictionnaire et même d'entrée à entrée dans un même dictionnaire, c'est la façon dont les formes sont présentées. Pour les formes affixées, on donne parfois simplement l'afixe, parfois une forme complète (cf. **Mi—**, **Maha—** vs **Fivaròtana**, **Ivaròtana** dans l'entrée *varotra* d'A&M ci-haut ou **Mi-** vs **Mpivarotra** et **Fivarotana** dans l'entrée *varotra* de NR⁴²). Pour les formes redoublées, elles sont toujours associées à une annotation (qui peut varier de forme) dans A&M. Dans NR, elles apparaissent parfois à la suite d'une annotation (en principe, l'abréviation «Dupli.»), parfois, elles restent implicites, seule l'annotation étant donnée.

On pourra donc ajouter au squelette d'entrée lexicale ci-haut une zone de DÉRIVÉS. Chaque DÉRIVÉ aura un TYPE MORPHOLOGIQUE (*préfixe*, *suffixe*, *circonfixe* ou *infixe* pour les dérivés affixaux et *duplicatif* pour les formes redoublées), en plus des autres rubriques communes à toutes les entrées : une VEDETTE (la forme dérivée complète), une CATÉGORIE⁴³, une liste (éventuellement vide) de TRADUCTIONS/DÉFINITIONS. Pour les dérivés affixaux, la forme de l'afixe pourra apparaître comme valeur de l'attribut correspondant au TYPE (*préfixe*, *suffixe*, *circonfixe* ou *infixe*).

Si donc on sait traduire les dérivés de chacune des institutions lexicales A&M, RR et NR, on peut les fusionner dans une institution A&MRRNR⁴⁴. Quand on détachera les dérivés de l'entrée principale où ils ont été créés, on les remplacera par un ensemble de liens correspondant aux différentes fonctions morphologiques possibles (i.e. les différents types de dérivés), mais également aux fonctions sémantiques ($\approx$ les *fonctions lexicales* de la lexicographie explicative et combinatoire, Mel'cuk, Clas et Polguère (1995)) que ces différents types de dérivés instantancient. Chaque dérivé sera donc accessible, depuis une racine par plusieurs chemins. Inversement, dans l'entrée du dérivé, les fonctions inverses (morphologiques ou sémantiques) retourneront l'entrée lexicale de la racine. Par exemple, appliquées à l'entrée *varotra*, la fonction morphologique *mp-* (ou *préfixe_{mp-}*), de même que la fonction sémantique *nominalisation-S_I* retourneront toutes deux le dérivé *mpivarotra* avec son entrée lexicale. Et la fonction *mp-⁻¹* (ou *préfixe_{mp-}⁻¹*) de même que la fonction *S_I⁻¹* appliquées à l'entrée *mpivarotra* retourneront l'entrée *varotra*. Ces fonctions inverses, particulièrement utiles pour l'apprenant de la langue, aussi bien que pour les programmes d'analyse, sont créées automatiquement. Étant donnée une racine *r*, on peut systématiquement créer un ensemble de *fonctions morphologiques* et de *fonctions sémantiques* régulièrement associées.

associée, elle devrait également être restructurable, du moins pour les affixes correspondant à des *fonctions lexicales* (au sens de la théorie sens-texte) précises, ce qui est le cas de la plupart des affixes en malgache (cf. infra).

⁴² Pour un même affixe, *mi-*, par exemple, on trouve dans RR 89 % d'entrées (957) avec l'afixe comme vedette du dérivé et 11 % (117) avec la forme complète. Comme cet affixe à finale vocalique n'entraîne pas d'allomorphie (sauf devant une base à «i» initial où les deux «i» se fusionnent généralement, mais pas toujours : *mi* + *iray* = *miray*, mais *mi* + *isa* = *miisa*), les formes complètes sont tout à fait prédictibles.

⁴³ Dans le cas où une institution source ne mentionne pas la catégorie du dérivé, celle-ci est toujours facile à reconstruire à partir du TYPE du DÉRIVÉ et de la CATÉGORIE de l'entrée contenant ce DÉRIVÉ. Pour les dérivés affixaux, chaque affixe définit la CATÉGORIE de son résultat. Pour les formes redoublées, le résultat est de la même CATÉGORIE que la forme de départ.

⁴⁴ Voir la zone des DÉRIVÉS de l'annexe qui donne une représentation externe du résultat de la fusion des trois entrées de *varotra*.

FONCTION MORPHOLOGIQUE	TYPE	FONCTION SÉMANTIQUE
<i>préfixe</i> _{an-}	RADICAL → V _{an}	<i>actif</i>
<i>préfixe</i> _{i-}	RADICAL → V _i	<i>moyen</i>
<i>préfixe</i> _{a-}	RADICAL → V _a	<i>statif</i>
<i>suffixe</i> _{-ina}	RADICAL _{thème} → V _{ina}	<i>passif</i>
<i>suffixe</i> _{-ana}	RADICAL _{thème} → V _{ana}	<i>passif</i>
<i>circonfixe</i> _{an--ana}	RADICAL _{thème} → V _{an-ana}	<i>circonstanciel</i>
<i>circonfixe</i> _{i--ana}	RADICAL _{thème} → V _{i-ana}	<i>circonstanciel</i>
<i>préfixe</i> _{tafa-}	RADICAL → X _{tafa}	<i>participe passif perfectif</i>
<i>préfixe</i> _{voa-}	RADICAL → X _{voa}	<i>participe passif perfectif</i>
<i>préfixe</i> _{f-}	V _{an/i} → N _f	<i>S0</i>
<i>préfixe</i> _{h-}	V _a → N _h	<i>S0</i>
<i>préfixe</i> _{mp-}	V* _{ina/ana} → N _{mp}	<i>S1</i>
<i>préfixe</i> _{ah-}	V _a → V _{ah}	<i>abilitatif</i>
<i>préfixe</i> _{amp-}	V* _{ina/ana} → V _{amp}	<i>causatif</i>
<i>préfixe</i> _{ank-}	V _{i/a} → V _{ank}	<i>transformatif</i>
<i>préfixe</i> _{if-}	V* _i → V _{if}	<i>réciproque</i>
<i>préfixe</i> _{ih-}	V _a → V _{ih}	<i>progressif</i>
<i>préfixe</i> _{m-}	V* _{ina/ana} → V _m	<i>présent</i>
<i>préfixe</i> _{n-}	V → V _n	<i>passé</i>
<i>préfixe</i> _{t-}	Prép → Prép	<i>passé</i>
<i>préfixe</i> _{h-}	V → V _h	<i>futur</i>
<i>redoublement</i>	RADICAL → RADICAL _{dupl}	<i>diminutif, fréquentatif, intensif, etc.</i>
<i>préfixe</i> _{ki-}	RADICAL _{dupl} → X _{ki}	<i>diminutif, jeu</i>
<i>préfixe</i> _{tsi-}	RADICAL _{dupl} → X _{tsi}	<i>diminutif, jeu</i>
<i>préfixe</i> _{tsi-}	NUMÉRAL → Adv _{numéral}	<i>'n par n'</i>
<i>préfixe</i> _{in-}	NUMÉRAL → Adv _{numéral}	<i>'n fois'</i>

préfixe <i>an-</i>	$N \rightarrow Adv_{loc}$	locatif

Figure 6 : Fonctions morphologiques typées et fonctions sémantiques en malgache

Ce tableau résume l'essentiel de la morphologie dérivationnelle (et une partie de la morphologie flexionnelle⁴⁵) malgache. La première colonne décrit la fonction morphologique (le type d'affixe ou de processus dont il s'agit), la deuxième colonne identifie le type d'input et d'output de la fonction morphologique⁴⁶ et la troisième colonne donne une brève identification la fonction sémantique associée⁴⁷.

Lors de la création d'une base lexicale, il suffit donc que l'agent lexicographe (humain ou non) applique cette table et vérifie systématiquement lesquelles de ces relations sont effectivement réalisées, quelles sont leurs propriétés particulières et, éventuellement, quelles autres relations plus idiosyncratiques existent pour la racine en question⁴⁸.

⁴⁵ La morphologie flexionnelle est très limitée en malgache. Il n'y a pas de marque de personne, ni de nombre sur le verbe. Les modes se réduisent à deux, l'infinitif et l'impératif. Il n'y a pas de genre ni de nombre sur le nom ou l'adjectif et donc pas d'accord. Par contre les prépositions peuvent être tensées et il existe un système complexe de prépositions et d'adverbes locatifs.

⁴⁶ *RADICAL* et *RADICAL*_{thème} font référence aux formes de base. Le *RADICAL*_{thème} est une forme, contenant généralement une consonne thématique qui n'est pas totalement prévisible et qui apparaît devant suffixe (i.e., à l'impératif, au passif suffixé et au circonstanciel). *RADICAL*_{dupl} est la catégorie des radicaux redoublés. Pour le *RADICAL*_{varotra}, le *RADICAL*_{thème} est *varot-* et le *RADICAL*_{dupl} est *varobarotra*.

⁴⁷ Nous n'indiquons pas ici les types sémantiques de l'input et de l'output des fonctions sémantiques, puisque ce travail est en cours, mais le nom de la fonction sémantique constitue une indication grossière du type de son output. On peut rapprocher ces fonctions sémantiques des fonctions lexicales de la lexicologie explicative et combinatoire. Cependant, pour la lexicologie malgache, il nous semble plus intéressant de construire d'abord des fonctions morphologiques et de les associer à des fonctions sémantiques, plutôt que l'inverse, puisqu'une bonne partie de la morphologie dérivationnelle est très régulière (et peut donc être générée automatiquement et vérifiée de façon efficace). Les types morphologiques sont des catégories indicées. Comme il s'agit d'une représentation par contraintes, on peut poser des contraintes négatives sur ces indices. Par exemple, la spécification *V*ina/ana* sur les préfixes *m-* et *mp-* indique qu'ils peuvent s'appliquer à n'importe quelle forme verbale sauf celles à suffixe (ou circonfixe contenant) *-ina* ou *-ana*. Notons aussi que l'on peut considérer les circonfixes malgaches comme une combinaison d'un préfixe et d'un suffixe. Dans ces conditions, on aurait les spécifications suivantes pour le suffixe *-ana* :

suffixe <i>-ana</i>	<i>RADICAL</i> _{thème} $\rightarrow V_{ana}$	passif
suffixe <i>-ana</i>	$V_{i/a} \rightarrow V_{i/a-ana}$	circonstanciel

⁴⁸ Il existe aussi quelques infixes fossiles à productivité très limitée, mais qui seront tout de même inclus dans nos bases lexicales.

Nous avons effectivement créé un environnement expérimental permettant de faciliter ce travail. Pour une forme donnée, les fonctions de la table sont automatiquement appliquées. L'agent lexicographe n'a plus qu'à compléter l'information (essentiellement sémantique) laissée implicite et à filtrer les données incorrectes. Les entrées correspondantes sont automatiquement créées, avec l'information partielle régulière qu'elles contiennent et associées à la racine par des hyperliens étiquetés (par les fonctions morphologiques et sémantiques) et bidirectionnels (les liens inverses étant évidemment étiquetés par les fonctions inverses). Les entrées résiduelles (i.e. ne correspondant pas à des formes effectives) peuvent être filtrées (automatiquement ou non)⁴⁹.

Si la base est créée exclusivement à partir d'institutions lexicales (i.e., sans tenir compte des connaissances linguistiques de l'agent et sans recours à des corpus externes), on peut tout de même obtenir des résultats intéressants. Ainsi, à partir de l'analyse des entrées de *varotra* dans les trois institutions lexicales A&M, RR et NR, on peut obtenir l'entrée complexe dont une représentation externe est donnée en annexe.

ANNOTATIONS

Les ANNOTATIONS sont très nombreuses⁵⁰ et peuvent être de plusieurs types dans les dictionnaires traditionnels malgaches.

D'abord, un type particulier : le renvoi éventuel à une RACINE. La plupart des entrées lexicales incomplètes et certaines entrées lexicales complètes contiennent une telle rubrique.

Par exemple :

« MARÀRY , Voy. RARY.»	(A&M, p. 434. Il s'agit du premier homonyme <i>rary</i> , 'mal, douleur', <i>marary</i> , 'malade, souffrant')
« Gija a. : Jer. Geja.»	(RR, p. 392. <i>Geja</i> 'serrer, étreindre, retenir')
« lahàsa V. asa.»	(NR, p. 180. Il y a quatre entrées homonymes <i>asa</i> . Il s'agit ici de <i>asa2</i> , 'travail', dont <i>lahasa</i> ,

<i>infixe-in-</i>	RADICAL → V _{in}	<i>passif</i>
<i>infixe-om-</i>	RADICAL → V _{om}	?

Le premier permet de relier *vinidy* ('acheté') à *vidy* ('achat, prix') et le second *homehy* ('riant') à *hehy* ('rire (nom)'). Il y a également d'autres préfixes fossiles identifiables formellement mais non sémantiquement : *bo-*, *fo-*, *kan-*, *ko-*, *mo-*, *sa-*, *san-*, *tan-*, *to-*, *to-*, *tsin-*, *vo-*, *za-*. D'autres infixes servent à former des dérivés sur des catégories mineures (*-re-*, pluriel, *-za-*, lieu caché, *-n-*, locatif). Les fonctions morphologiques correspondantes existent, mais sans fonctions sémantiques associées. Elles peuvent être activées au besoin. Ainsi *bohesika* ('action d'entrer en masse') peut être traité comme un RADICAL simple ou comme un RADICAL dérivé de *sesika* ('bourse, bourrage').

⁴⁹ Les fonctions correspondantes retourneraient, par exemple, la liste vide.

⁵⁰ Ainsi, NR contient plus de 2 500 annotations pour un peu plus de 4 500 entrées.

qui n'apparaît pas dans l'entrée en question, est plus ou moins synonyme — plutôt spécialisé comme 'travail des champs' —).

Les entrées lexicales complètes qui contiennent un renvoi RACINE sont celles qui ne sont pas elles-mêmes des racines et dont la relation avec la racine n'est pas graphiquement transparente.

Par exemple :

«**KAVÌA** (*havia*), *adj.* Gaucher. **Kaviavia**, *adj.* Maladroit, lourd. **Mi—**, *vn*, Être maladroit. *Mikaviavia* izaitsizy izy, raha mamankona. Il est très maladroit pour raboter. **FI—**, *s.* La maladresse. **MPI—**, *s.* Un maladroit. **Fikaviaviàzana**, *s.* La maladresse. **Ikaviaviàazana**, *rel.*»

(A&M, p. 319. *Ka-* n'est pas un préfixe dérivationnel régulier.)

«**Maneza** (*heza*) *mt. p.* (*manezà*) : Manisy heza.»

(RR, p. 617. *Manisy heza* 'appliquer le heza', le heza est (selon A&M, p. 243) une «Eau extraite des cendres de certaines herbes et servant de mordant dans la teinture de la soie.» *heza* est la racine, *mt. p.*, la catégorie *matoantenin'ny mpanao*, 'litt. verbe d'acteur', *manezà*, la forme de l'impératif)

«**nahàndro** (*handro*) *n.* Mets, nourriture.»

(NR, p. 213. *Na-* n'est pas un préfixe dérivationnel régulier. Littéralement, il s'agit de la forme flexionnelle du passé de *mahandro*, 'cuisiner'.)

Dans une base lexicale, les liens sont établis systématiquement pour toutes les formes vers leurs racines, que les relations soient transparentes ou pas. On ne conservera donc pas ces renvois spécifiques.

Les autres renvois dans les dictionnaires traditionnels concernent des synonymes, des antonymes, des variantes, des abréviations, des surnoms, des formes supplétives (par exemple *amidy* comme forme passive de *varotra* dans A&M), le renvoi à des illustrations (dans RR), etc. Enfin, un certain nombre d'annotations constituent des marques d'usage. Dans nos bases lexicales, nous conservons ces annotations classées par type, en attendant de les distribuer et de les compléter de façon plus systématique.

4. CONCLUSION

Dans cet article, nous avons tenté de montrer comment on pouvait combiner de l'information lexicale provenant d'institutions lexicales distinctes dans une base lexicale dynamique. Nous avons examiné la structure des entrées lexicales dans trois institutions lexicales : Abinal & Malzac (1888), Rajemisa-Raolison (1985) et Rajaonarimanana (1995). Les dictionnaires traditionnels malgaches sont particulièrement intéressants à cet égard, puisqu'ils sont très riches et généralement bien structurés. De plus, comme la langue malgache possède une morphologie dérivationnelle complexe, mais très régulière, il est possible d'utiliser ces régularités pour filtrer et compléter les dictionnaires traditionnels. Le but à long terme de cette recherche est de favoriser l'émergence d'une véritable lexicographie expérimentale qui utilise au maximum les données des institutions lexicales traditionnelles.

RÉFÉRENCES

- ABINAL & MALZAC, SJ. (1888) : *Dictionnaire malgache-français*, Fianarantsoa, Ambozontany.
- BLACHE, P. & J.-Y. MORIN (1997) : «TAO et théories linguistiques : institutions grammaticales», communication aux V^{es} Journées scientifiques du réseau LTT, *La mémoire des mots*, Tunis, septembre 1997.
- BLANK, I. (1995) : «Sentence alignment : Methods and implementations», *T.A.L.*, 36, 1-2, pp. 81-99.
- CHOMSKY, N. (1995) : «Bare phrase structure», dans Webelbuth, G. (1995 réd.), *Government and Binding Theory and the Minimalist Program*, Oxford, Blackwell, pp. 383-439.
- DOMENICHINI-RAMIARAMANANA, B. (1983) : *Du ohabolana au hainteny, langue, littérature et politique à Madagascar*, Paris, Karthala.
- EVANS, R. & G. GAZDAR (1996) : «DATR : A language for lexical knowledge representation», *Computational Linguistics*, Vol. 22, 2, pp. 167-216.
- GALE, W. & K. CHURCH (1993) : «A program for aligning sentences in bilingual corpora», *Computational Linguistics*, 19, 1, pp. 121-142.
- MEL'CUK, I. et al. (1984, 1988, 1992) : *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques*, 3 volumes parus, Montréal, Presses de l'Université de Montréal.
- MEL'CUK, I., CLAS, A. et A. POLGUÈRE (1995) : *Introduction à la lexicologie explicative et combinatoire*, Montréal, AUPELF-UREF, Gembloux, Duculot.
- MILIČEVIĆ, J. (1997) : *Étiquettes sémantiques dans un dictionnaire formalisé du type Dictionnaire Explicatif et Combinatoire*, Mémoire de maîtrise, Département de linguistique et de traduction, Université de Montréal.
- MORIN, Jean-Yves (1997) : «Lexiques-grammaires et institutions lexicales», communication aux V^{es} Journées scientifiques du réseau LTT, *La mémoire des mots*, Tunis, septembre 1997.
- PROFITA, P. (1969) : *Rakibolana malagasy-italianina/Dizionario italiano malgascio*, Antananarivo, P. Pietro Profita.
- PUSTEJOVSKY, J. (1995) : *The Generative Lexicon*, Cambridge, Mass., MIT Press.
- RABENILAINA, R.-B. (1987) : *Lexique-grammaire du malgache. Constructions transitives et intransitives*, Antananarivo, FOFIPA.
- RABENILAINA, R.-B. (1996) : *Le verbe malgache. Constructions transitives et intransitives*, Montréal, GRESLET et AUPELF-UREF.
- RAJAONARIMANANA, Narivelo (1995) : *Dictionnaire du malgache contemporain*, Paris, Karthala.
- RAKOTONAIVO, F. (1996) : *Rakibolana frantsay-malagasy*, Fianarantsoa, Ambozontany.
- SILBERZTEIN, M. (1993) : *Dictionnaires électroniques et analyse automatique de textes*, Paris, Masson.

ANNEXE : Une représentation externe du résultat de la fusion des trois entrées de *varotra*⁵¹

VEDETTE **varotra**

CAT n

TRADUCTIONS

Vente, commerce, marchandises NR
Vente, commerce, négoce, trafic, marchandises AM

DÉRIVÉS

DÉRIVÉ

TYPE MORPHO Duplicatif
 VEDETTE **varobarotra**
 CAT n
 TRADUCTIONS [dupl. dim. de
varotra] AM
 AM RR NR

DÉRIVÉ

TYPE MORPHO [Circonfixe **fi-**
-ana]
 VEDETTE **fivarotana**
 CAT n
 TRADUCTIONS
L'action de vendre, de faire le commerce, de trafiquer, le prix, le lieu, la cause AM
 commerce NR
 EXPRESSIONS
 [MALGACHE **Fivarotam-boky**

⁵¹ Pour des raisons de concision, nous n'avons pas inclus dans cette représentation les définitions en malgache de RR, ni les champs ne contenant aucune valeur spécifiée positivement, ni non plus les fonctions sémantiques. Les abréviations AM NR et RR indiquent l'origine d'une information. Leur portée, implicite ici, est explicitement spécifiée dans les représentations internes. Cet exemple de représentation externe réduite est générée automatiquement à partir de la représentation interne. On pourrait aussi générer une représentation HTML, par exemple. La représentation interne est, une représentation prolog et contient beaucoup plus d'informations.

FRANÇAIS *librairie*
 MALGACHE **Fivarotam-panafody**
 FRANÇAIS *pharmacie*
 MALGACHE **Fivarotan-tena**
 FRANÇAIS *prostitution*
 NR
 AM RR NR

DÉRIVÉ

TYPE MORPHO [Circonfixe
i--ana]
 VEDETTE **ivarotana**
 CAT v_circ
 EXEMPLES
 MALGACHE **Ariary telopolo no nivarotako ny omby mifahiko.**
 FRANÇAIS *J'ai vendu trente piastres le boeuf que j'avais engraisé.* AM
 NOTES [IMPÉRATIF
ivaròty] AM
 AM RR

DÉRIVÉ

TYPE MORPHO [Préfixe **fi-**]
 VEDETTE **fivarotra**
 CAT n
 TRADUCTIONS
Ce qu'on peut vendre, manière de vendre AM
 AM RR

DÉRIVÉ

TYPE MORPHO [Préfixe
maha-]
 VEDETTE **mahavarotra**
 CAT v_abil
 TRADUCTIONS *Qui peut ou sait vendre* AM
 AM RR

DÉRIVÉ

TYPE MORPHO [Préfixe
mampi-]
 VEDETTE **mampivarotra**
 CAT v_caus
 AM RR

DÉRIVÉ

TYPE MORPHO [Préfixe
mi-]

VEDETTE **mivarotra**
 CAT v_actif
 TRADUCTIONS
Vendre, faire le commerce,
trafiquer AM
Vendre, trafiquer NR
 EXPRESSIONS
 MALGACHE **Mivaro-**
tena RR
 NOTES [IMPÉRATIF
mivarôta] AM
 AM RR NR

DÉRIVÉ
 TYPE MORPHO [Préfixe
mifampi-]
 VEDETTE **mifampivarotra**
 CAT v_récip
 AM RR

DÉRIVÉ
 TYPE MORPHO [Préfixe
mpi-]
 VEDETTE **mpivarotra**
 CAT n
 TRADUCTIONS *Celui qui vend, le*
marchand, le commerçant AM
commerçant, marchand NR
 EXPRESSIONS
 MALGACHE **Mpivaro-**
kena
 FRANÇAIS *boucher*
 MALGACHE **Mpivaro-**
tena
 FRANÇAIS *prostituée*
 MALGACHE **Mpivaro-**
mandeha
 FRANÇAIS *marchand*
ambulant
 MALGACHE
Mpivarotra amoron-
dalana
 FRANÇAIS *marchand du bord des*
rues
 MALGACHE **Mpivarotra**
omby
 FRANÇAIS *marchand de*
bestiaux NR
 AM RR NR

DÉRIVÉ
 TYPE MORPHO [Préfixe **voa-**]
 VEDETTE **voavarotra**
 CAT v_passif
 RR

EXPRESSIONS
 MALGACHE **Fifanakalozam-**
barotra
 FRANÇAIS *échanges commerciaux*
 NR
 MALGACHE **Fivarotam-boky**
 FRANÇAIS *librairie* NR
 MALGACHE **Fivarotam-panafody**
 FRANÇAIS *pharmacie* NR
 MALGACHE **Fivarotan-tena**
 FRANÇAIS *prostitution* NR
 MALGACHE **Lalam-barotra**
 FRANÇAIS *débouché* NR
 MALGACHE **Mamola-barotra**
 FRANÇAIS *Chercher à faire baisser, à*
réduire le prix AM
 MALGACHE **Mamola-barotra**
 RR
 MALGACHE **Manongom-barotra**
 [CF Songona] AM
 AM RR
 MALGACHE **Tombom-barotra**
 FRANÇAIS *Profit, bénéfice* AM
 FRANÇAIS *bénéfice* NR
 MALGACHE **Tranom-barotra**
 FRANÇAIS *maison, société de*
commerce NR
 MALGACHE **Varo-babo**
 FRANÇAIS *Très bon marché, à vil*
prix AM
 MALGACHE **Varo-balongana**
 [CF Valongana] AM
 MALGACHE **Varo-baventy**
 FRANÇAIS *Commerce en grand*
 [FIGURÉ *tout vol où on expose sa vie*]
 AM
 MALGACHE **Varo-bazaha**
 FRANÇAIS *Vente à prix fixe* AM
 AM RR
 MALGACHE **Varo-boba**
 [CF Varo-babo] AM
 FRANÇAIS *vente à vil prix, solde*
 NR
 AM RR NR
 MALGACHE **Varo-dratsy** RR
 MALGACHE **Varo-droa**
 FRANÇAIS *Action de vendre encore*
ce qu'on a vendu, de promettre à l'un ce qu'on
a promis à l'autre AM
 MALGACHE **Varodroa** RR
 MALGACHE **Varo-mahery**
 FRANÇAIS *Vente forcée* AM
 MALGACHE **Varo-maika**

FRANÇAIS *Vente faite à vil prix pour cause d'urgence* AM
 FRANÇAIS *vente d'urgence* NR
 AM RR NR
 MALGACHE **Varo-maizina**
 FRANÇAIS *marché noir* NR
 RR NR
 MALGACHE **Varo-mamindro**
 FRANÇAIS *Achat dont on fait traîner la discussion dans l'espérance de trouver mieux* AM
 MALGACHE **Varo-mandeha**
 FRANÇAIS *Colportage* AM
 NR
 AM RR NR
 MALGACHE **Varo-maty**
 FRANÇAIS *Vente définitive* AM NR
 AM RR NR
 MALGACHE **Varo-miandry**
 FRANÇAIS *Objets qu'on garde longtemps, qui ne se vendent que par circonstance, dont on demande un bon prix* AM
 AM RR
 MALGACHE **Varo-miera**
 FRANÇAIS *Vente d'un article qui peut être rendu si l'acquéreur n'en est pas content; action de prendre à crédit avec obligation de rendre l'objet à défaut du prix* AM
 AM RR
 MALGACHE **Varo-mifody** RR
 MALGACHE **Vodiondrim-barotra**
 FRANÇAIS *arrhes* NR
 MALGACHE **Ady varotra**
 FRANÇAIS *marchandage, négociation* NR
 RR NR
 MALGACHE **Miady varotra**
 FRANÇAIS *marchander, négocier* NR
 MALGACHE **Varo-mahay mody**
 FRANÇAIS *Vente d'un article qui peut être rendu si l'acquéreur n'en est pas content; action de prendre à crédit avec obligation de rendre l'objet à défaut du prix* AM
 MALGACHE **Varotra mahay mody** RR
 MALGACHE **Varo-tsy azon-dahy**
 FRANÇAIS *Marché sur lequel on ne s'entend pas, malentendu dans une vente ou un achat* AM
 MALGACHE **Varo-tsy mifody**
 [LITTÉRALEMENT *Vente sur laquelle on ne revient pas*]
 [FIGURÉ *la mort*] AM

AM RR
 MALGACHE **Varotsimifody**
 FRANÇAIS *l'au-delà, l'autre monde*
 [LITTÉRALEMENT *vente sur laquelle on ne revient pas*] NR
 MALGACHE
Varotr'omby anaty ambiaty
 [LITTÉRALEMENT *Vente de boeufs cachés dans les ambiaty*]
 [FIGURÉ *vente malhonnête*] AM
 MALGACHE
Varotra omby anaty ambiaty
 RR
 MALGACHE **Varotra alika** RR
 MALGACHE **Varotra**
ambongadiny
 FRANÇAIS *Vente en gros, par pièce* AM
 FRANÇAIS *vente en gros* NR
 AM RR NR
 MALGACHE **Varotra an'elakela-trano**
 RR
 MALGACHE **Varotra an-trano**
tokana
 FRANÇAIS *Monopole* AM
 AM RR
 MALGACHE **Varotra an-tsinjarany** AM
 MALGACHE **Varotra antsinjarany** RR NR
 FRANÇAIS *Vente au détail*
 AM RR NR
 MALGACHE **Varotra**
fampisehoana
 FRANÇAIS *exposition-vente* NR
 MALGACHE **Varotra iombonana**
 FRANÇAIS *communauté économique* NR
 MALGACHE **Varotra ivelany**
 FRANÇAIS *commerce extérieur* NR
 MALGACHE **Varotra misy fehi-vavany** RR
 MALGACHE **Varotra olona** RR
 MALGACHE
Varotra tsy miverimbodiondry RR
 MALGACHE **Varotry ny maika hiala** RR
 MALGACHE **Varotry ny maika hody** RR
 MALGACHE **Varotry ny mananotena**

[LITTÉRALEMENT Commerce des
veuves]

[FIGURÉ vente faite avec hésitation,
dans la crainte de se tromper] AM

AM RR

MALGACHE Varotra an-
kobohobo RR

MALGACHE

Varo-jabo an-tsena, ka samy
mandoka ny azy ho masaka.

FRANÇAIS (prov.) Chacun vante sa
marchandise, comme les marchands de Jabo
au marché. AM

EXEMPLES

MALGACHE

Aza hitsahinao ity
varotr'olona. R R

MALGACHE

Lasa nanao varotra any an-
tsiraka izy.

RR

NOTES

[PASSIF Amidy] [CF Vidy]

AM

ÉTUDE POUR UNE EXTRACTION AUTOMATIQUE DE NÉOLOGISMES

Yvette Yannick MATHIEU

CNRS, Laboratoire de Linguistique Informatique, Université Paris 13, France

Le but de cette étude est le repérage automatique de néologismes de forme dans des textes écrits. Nous avons comparé, par ordinateur, le vocabulaire d'une année d'un journal quotidien (*Le Monde* 1993) avec la nomenclature de l'année 1992 du dictionnaire électronique *DELA*¹ qui contient les formes fléchies des mots simples. Ce dictionnaire a été constitué au Laboratoire d'Automatique Documentaire et Linguistique (LADL) par Blandine Courtois (1990). Les résultats de cette comparaison ont été analysés afin de définir des algorithmes de traitements informatiques pour discriminer les erreurs et reconnaître les néologismes potentiels.

Cet article complète celui de Mathieu, Gross et Fouqueré (1998) dans lequel nous avons étudié en détail les mots commençant par les lettres **a**, **b** et **c**. L'étude que nous présentons ici est maintenant exhaustive puisque nos résultats concernent le corpus entier; précisons également qu'un dictionnaire de référence différent a été utilisé².

1. RÉSULTATS

Le corpus du journal *Le Monde* de l'année 1993 contient 23 millions de chaînes de caractères³ dont 201 775 différentes. Chaque chaîne a été recherchée dans le dictionnaire, celles qui en sont absentes sont des noms propres, des mots avec tiret ou des mots simples.

Il est naturel de ne trouver aucun mot avec tiret dans le *DELA*, car ce dictionnaire ne contient que des mots simples.

Nous avons dénombré 123 044 noms propres. Nous appelons «nom propre» tout mot qui commence par une lettre majuscule, qui n'est pas un sigle et qui n'est pas le début d'une phrase, c'est-à-dire qui n'a pas été reconnu dans sa graphie commençant par une

¹ Dictionnaire Electronique du LADL des Formes fléchies.

² La problématique du choix du dictionnaire de référence ne sera pas abordée ici.

³ Nous appelons chaîne de caractères toute séquence de lettres contiguës contenant éventuellement un ou plusieurs traits d'union mais ni blanc ni signe de ponctuation. Les chaînes en majuscules et minuscules sont distinguées.

minuscule. Ces mots représentent la grande majorité des mots inconnus dans les corpus de journaux. Max Silberstein (1995) a fait une étude comparative des corpus du *Monde* 1992 et 1993 et a dénombré 89 % de mots en majuscules parmi les mots absents du *DELA*F.

Il y a 15 230 mots simples du *Monde* 1993 qui ne sont pas reconnus par le *DELA*F. Parmi ces mots, il y a des néologismes possibles et des erreurs de diverses natures.

1.1 Les néologismes possibles

Nous avons fait une étude précise de tout le corpus des mots simples qui peuvent être des néologismes. Nous pouvons distinguer :

- des noms et adjectifs dérivés de nom de ville, de région ou pays : *ajacciens*, *catalanisation*, *corrèzien*, *corsitude*, *marocanisation*, *paimpolaise*;

- des noms et adjectifs dérivés de nom de personnalité artistique ou politique. Ces mots sont pour la plupart liés à l'actualité et leur emploi est probablement éphémère : *balladurien(ienne)*, *balladurisation*, *balladurisé*, *balladurisme*, *chiraquien(ienne)*, *chiraquie*, *chiraquisme*, etc. Mais il y a également des dérivés de personnages célèbres plus anciens : des compositeurs (*berlioziste*, *brahmsien*), des peintres (*caravagesque*, *chagallien*, *raphaëlesque*), des écrivains (*brechtien*, *brechtisme*, *conradienne*, *courtelinesque*, *dickensien*, *éluardien*, *perecquien*), des cinéastes (*fellinien*, *rohmérien*), des musiciens (*coltranien*). Ces mots dérivés, qui ne vont probablement pas disparaître de la langue, pourraient soit être intégrés au dictionnaire de référence, soit constituer un dictionnaire à part entière.

- des noms de vins : *bourgueil*, *margaux*, *yquem*;

- des abréviations ou diminutifs, utilisés dans le langage courant : *ados* (adolescents), *allocs* (allocations), *clodo* (clochard), *cata* (catastrophe), *séropo* (séropositif). Ces mots ont été étudiés par Groud et Serna (1996);

- des mots préfixés et des formes en *o-* qui sont des réductions d'adjectif ou de nom coordonné (Dugas, Mathieu et Courtois 1995). Certains préfixes sont très fréquents : *anti* (*antibourgeois*, *anticapitaliste*, *antihéros*), *co* (*coanimateur*, *codécouvreur*, *coprésence*), *dé* (*décohabitations*, *déprivatisées*, *déréférencier*);

Le tableau suivant indique le nombre d'occurrences des préfixes les plus courants.

Préfixes	Nombre
<i>anti</i>	275
<i>auto</i>	112
<i>co</i>	158
<i>dé</i>	114
<i>euro</i>	42
<i>hyper</i>	35
<i>inter</i>	83
<i>micro</i>	38
<i>multi</i>	69
<i>néo</i>	40
<i>pré</i>	83
<i>pro</i>	34
<i>re</i>	89
<i>ré</i>	57

Tableau 1

- des mots étrangers, essentiellement anglais, italiens, allemands et latins : *addict*, *network*, *nocciola*, *overseas*. Certains sont utilisés couramment dans la langue comme *bodybuilding*, *brunch*, *bulldog*, *fitness* (Girardin 1996) et produisent parfois des dérivés francisés : *bodybuildé*, *folkeux*;

- des mots nouveaux (c'est-à-dire qui ne sont pas dans le *DELAF*) qui ne font pas partie des catégories déjà mentionnées comme : *accessoirisé*, *accréditation*, *anabolisant*, *animalerie*, *audimat*, *automaticiens*, *avionique*, *crédibiliser*, *contractualiser*, *photocopillage*, *scénarisation*. Certains mots sont construits à partir d'initiales. Ainsi *bédéphile* vient de B.D. (Bandes Dessinées), *pédégères* vient de P.D.G. (Président Directeur Général). Des adjectifs sont créés à partir de ceux existants en ajoutant un suffixe pour amplifier ou modifier le trait : *brillantissime*, *élégantissime*, *kitchissime*, *lentissime*, *nullissime*, *jargonnesque*, *spiralesque*, *vampiresque*. Une intensification est également donnée par une orthographe différente : *hénaurme* pour *énorme*. Des adjectifs sont créés en ajoutant le suffixe *phobe* : *homophobe*, *motophobe*, *publiphobe*. Il existe une tendance à la féminisation de noms ou d'adjectifs : *baroudeuses*, *forçate*, *écrivaine*, *loubarde*.

Cette typologie est résumée dans le tableau 2.

Types	Nombre	% des néologismes potentiels
Néologismes potentiels	7170	
<i>N</i> et <i>Adj</i> dérivés de noms géographiques	855	11,9
<i>N</i> et <i>Adj</i> dérivés de noms de personne	370	5
Abréviations et diminutifs	170	2,37
<i>N</i> , <i>V</i> et <i>Adj</i> préfixés	2030	28,3
Autres mots simples	2280	31,8

Tableau 2

Parmi les mots simples inconnus du *DELA*F, il y a de nombreuses erreurs. Nous en avons fait également une typologie, afin de définir des algorithmes pour les éliminer automatiquement par ordinateur.

1.2 Les erreurs

Les erreurs sont principalement typographiques : mots collés comme : *leministre*, *lacrération*, *étédédécidées*, *catastrophiquedes*, *françaiset* (1 340 en tout), mots coupés : *onneur*, *sec*, *uions*, et des mots mal orthographiés, la frontière entre ces deux types d'erreur n'étant pas clairement délimitée. (Voir Mathieu et *alii.* (1997) pour une étude détaillée.)

2. STRATÉGIE POUR RECONNAÎTRE LES «CANDIDATS-NÉOLOGISMES»

L'étude précédente permet de définir les différentes étapes d'un traitement informatique. Pour établir si un mot inconnu du dictionnaire de référence peut être un néologisme, nous proposons :

- 1) d'isoler les mots simples;
- 2) d'éliminer les erreurs : les lettres isolées, les mots qui ont 3 consonnes identiques qui se suivent, les mots collés, les mots qui ont des erreurs d'accent (cette dernière étape en utilisant un correcteur orthographique);
- 3) d'éliminer les mots qui présentent une malformation régulière comme *arborigène* et les mots tronqués, par consultation de dictionnaires préalablement établis.

Une analyse du corpus restant doit permettre de reconnaître les mots qui peuvent être des néologismes. Un certain nombre d'opérations peuvent être exécutées, comme :

- extraire les mots préfixés, avec une méthode que nous avons décrite dans Mathieu et *alii.* (1997);
- extraire les mots contenant certains suffixes comme *esque* ou *issime*;
- repérer les mots issus de noms géographiques ou de noms propres. Ce repérage pourrait se faire grâce à une étude sur les processus de dérivation de ces noms. Denis Maurel (1994) a commencé ce travail pour les noms de géographie;
- reconnaître certaines catégories de mots, par consultation de dictionnaires précédemment établis : les mots étrangers, les noms de vins, etc.

La démarche que nous venons de décrire est rendue possible, d'une part, par la disponibilité de grands corpus de textes et, d'autre part, par l'existence ou le développement d'importantes ressources linguistiques électroniques comme les dictionnaires de formes fléchies, les dictionnaires géographiques, etc.

CONCLUSION

Cette étude s'inscrit dans un courant de recherche qui privilégie l'étude de phénomènes linguistiques à travers de grands corpus. Des travaux se sont plus particulièrement intéressés à la détection de néologismes dans la presse (voir, par exemple, Cabré et de Yzaguirre (1995) pour le catalan et le castillan). L'objectif principal de notre travail est d'alléger considérablement la tâche du linguiste, en lui proposant des procédures semi-automatiques pour constituer un corpus de néologismes éventuels. Mais il reste néanmoins un travail important à faire : établir si un néologisme potentiel est ou non un vrai néologisme.

RÉFÉRENCES

- CABRÉ, Maria Teresa et Luis de YZAGUIRRE (1995) : «Stratégie pour la détection semi-automatique des néologismes de presse», *Traduction Terminologie Rédaction*, vol. VIII, n° 2, Canada, Université Concordia, pp. 89-100.
- COURTOIS, Blandine (1990) : «Un système de dictionnaires électroniques pour les mots simples du français», *Langue française*, 87, Paris, Larousse.
- DUGAS, André, Yvette Yannick MATHIEU et Blandine COURTOIS (1995) : «Formes simples en o-», *Linguisticae Investigationes*, vol. XIX, n° 2, Amsterdam / Philadelphia, John Benjamins, pp. 429-441.
- GIRARDIN, Chantalle (1996) : «Trader : aux frontières du néologisme», *Terminologies nouvelles*, Rint (Réseau international de néologie et terminologie), n° 14, Ministère de la Culture et des Affaires sociales de Belgique, Bruxelles, pp. 31-34.
- GROUD, Claudette et Nicole SERNA (1996) : *De abdom à zoo, Regards sur la troncation en français contemporain*, CNRS-INaLF, Paris, Didier Érudition, 161 pages.
- MATHIEU, Yvette Yannick, Gaston GROSS et Christophe FOUQUERÉ (1998) : «Vers une extraction automatique des néologismes», *Cahiers de lexicologie*, Paris, Didier Érudition.
- MAUREL, Denis (1994) : «Le traitement informatique de la dérivation des noms de ville», *T.A.L.*, vol. 35, n° 2, Paris, ATALA, pp. 111-127.
- SILBERZTEIN, Max (1995) : «Dictionnaires électroniques et comptage de mots», *JADT*, Rome, pp. 93-101.

REPÈRES MÉMORIELS DU LEXIQUE DANS LES LANGUES BANTU. ILLUSTRATION À PARTIR DES STRUCTURES MORPHO-SÉMANTIQUES DES SUBSTANTIFS KIRUÚNDI

Jean B. NTAIRUTIMANA

GRESLET, Université de Montréal, Canada

INTRODUCTION

Le groupe bantu représente un très large ensemble de langues parlées en Afrique centrale, orientale et australe. Globalement, la zone bantu s'étend au sud d'une ligne reliant le sud du Nigeria au sud de la Somalie (cf. carte en annexe).

Les langues bantu sont des langues remarquablement homogènes quasiment à tous les niveaux linguistiques, tellement homogènes que toutes les études historico-comparatives leur attribuent aisément et unanimement une maternité linguistique commune, la proto-langue ayant été reconstituée à la fin du XIX^e siècle par Meinhof (1899) qui l'a nommée Ur-Bantu¹.

Les langues appartenant à la famille bantu possèdent une structure morpho-lexicale particulière par sa grande régularité. Cette régularité facilite une grande prédictibilité de la valeur sémantique de chaque mot, ainsi qu'une grande souplesse dans la créativité lexicoterminologique; ce qui, comme nous le montrerons, amplifie considérablement la mémoire linguistique de chaque mot.

Le kiruúndi appartient au domaine linguistique bantu². Sa structure linguistique correspond au modèle typique des structures linguistiques bantu, c'est pourquoi nous partirons du lexique kiruúndi, particulièrement du substantif, pour illustrer nos propos sur les repères mémoriels du lexique bantu.

Nous présenterons, dans cette communication, les principaux éléments morphologiques du lexique kiruúndi où s'ancre préférentiellement la mémoire linguistique, ainsi que l'interactivité de ces repères d'ancrage au niveau intra et extra-lexical. Nous en

¹ L'Ur-Bantu est également appelé Proto-Bantu ou bantu commun.

² Le kiruúndi fait partie du grand groupe I.A.5., selon la classification de Greenberg (1966); il porte le numéro d'identification D62 dans la classification de Guthrie (1971), et le numéro J62 selon Meeussen (1967).

montrons, par la suite, les corrélations avec les structures morphologiques types des autres langues bantu.

Nous proposerons également une façon d'organiser et de gérer les données relatives à ces structures morphologiques et, enfin, nous présenterons les applications possibles de cet aspect de la mémoire des mots dans divers domaines linguistiques et paralinguistiques.

1. SUBSTANTIF KIRUÚNDI

1.1 Structure morphologique

Le nom³ kiruúndi est constitué de 6 types de morphèmes, à savoir :

1. l'augment
2. le préfixe
3. l'infixe
4. le radical
5. le suffixe
6. la finale

N.B. : L'ensemble formé par le radical, ses suffixes éventuels et la finale, est appelé thème.

Le schéma ci-dessous montre comment les morphèmes — encadrés — se combinent entre eux pour constituer un nom, à travers une série d'étapes strictement ordonnées, partant du point de départ 0 → vers le point d'arrivée → 0.

La flèche qui contourne un morphème, de gauche à droite, signifie que le morphème en question peut être présent ou non dans la structure substantivale. C'est le cas de tous les morphèmes nominaux, à l'exception du préfixe et du radical qui ne peuvent jamais manquer dans un nom.

La flèche qui contourne un morphème, de droite à gauche, signifie que l'on peut retrouver plus d'un morphème de cette catégorie au sein d'un même substantif, situation très fréquente pour les infixes et les suffixes.

La flèche qui contourne, de droite à gauche, une série de morphèmes, veut dire que la série peut être redupliquée à l'intérieur d'un même nom, comme on le verra en 1.2.2.

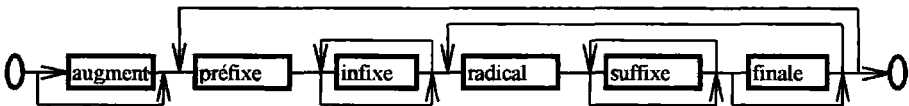


Figure 1 : Schéma de la structure morphologique du substantif kiruúndi

³ Ici, nous ne faisons aucune distinction entre les notions de «nom» et de «substantif».

Le mécanisme de construction substantivale schématisé ci-dessus est très productif en kiruúndi. Théoriquement, il est même capable de générer une infinité de noms différents, qu'aucun ouvrage ne pourrait inventorier au complet.

Le tableau suivant présente, à titre d'exemples, quelques-uns des mots construits sur le modèle du schéma précédent, tandis que les découpages morphologiques illustrent la variété et la souplesse de leurs combinaisons.

SUBSTANTIFS	STRUCTURES MORPHOLOGIQUES					
	Augment	Préfixe	Infixe	Radical	Suffixe	Finale
1. <i>Abatáhahára</i> ≡ 'qui sont toujours présents'	a-	-ba-	-ta- -ha-	-har-		-a
2. <i>Agakomére</i> ≡ 'petite blessure'	a-	-ka-		-kom-	-ir-	-e
3. <i>Agatéerantímba</i> ≡ 'source d'inquiétudes'	a- i- ⁴	-ka- -n-		-ta- -tiimb-	-ir-	-a -a
4. <i>Imigoondooro</i> ≡ 'pousses [végétales]'	i-	-mi-		-goond-	-uur-	-a
5. <i>Inkórabára</i> ≡ 'qui est très dangereux'	i- i-	-n- -i-		-kor- -bar-		-a -a
6. <i>Intóboro</i> ≡ 'trou'	i-	-n-		-tob-	-ur-	-o
7. <i>Ishúurwé</i> ≡ 'fleur'	i-	-i-		-shuur-	-u-	-e
8. <i>Ubugúfi</i> ≡ 'état de ce qui est court'	u-	-bu-		-gufi		
9. <i>Udukomére</i> ≡ 'petites blessures'	u-	-tu-		-kom-	-ir-	-e
10. <i>Ugutóoratoora</i> ≡ 'ramassage répété'	u-	-ku-		-ta- -ta-	-uur- -uur-	-a -a
11. <i>Umukóbekóbe</i> ≡ 'figus'	u-	-mu-		-kobe- -kobe		
12. <i>Umutázitéeraruungu</i> ≡ 'qui ne les [vaches] ennuie pas'	u- i-	-mu- -i-	-ta- -zi-	-ta- -ruung-	-ir-	-a -u
13. <i>Umutúuhúrana</i> ≡ 'qui n'induit pas en erreur'	u-	-mu-	-ta-	-uh-	-ur-an-	-a
14. <i>Umuuntu</i> ≡ 'personne [humaine]'	u-	-mu-		-ntu		
15. <i>Umwímenyerezo</i> ≡ 'exercice'	u-	-mu-	-ii-	-meny-	-ir- -ir- -i-	-o

⁴ La lettre-morphème barrée subit un amuïssement dans la structure de surface.

16. <i>Urukurúkurú</i> ≡ 'rumeur'	u-	-ru-		-kur- -kur-		-u -u
17. <i>Urushíngé</i> ≡ 'aiguille'	u-	-ru-		-shiing-		-e

Tableau 1 : Structure morphologique de quelques substantifs kiruúndi

On ne peut dénombrer tous les radicaux du kiruúndi, et il s'en crée, de temps à autre, de nouveaux pour combler les besoins de communication; ce qui fait que le nombre de radicaux est théoriquement illimité. Contrairement au radical, les autres morphèmes lexicaux sont en nombre limité. Ils constituent des listes fermées, faciles à inventorier.

Ainsi donc, la création lexicale se fait autour du même noyau radical, auquel sont adjoints les différents morphèmes illustrés par la figure ci-dessus. Tous ces morphèmes interagissent entre eux et ont chacun un rôle particulier à jouer dans la structure du mot. Chaque morphème possède une valeur linguistique propre, non autonome et incomplète, en interaction avec les valeurs linguistiques des autres morphèmes; tandis que la valeur linguistique de la totalité du mot créé dépend des valeurs combinées des morphèmes qui le constituent.

Ce mode de création est extrêmement productif et régulier, de façon que tout locuteur soit apte à créer de nouvelles unités lexicales. Il est aussi normalement capable de deviner le sens de tout mot, même nouveau, si la structure linguistique du mot est correcte, et si la mémoire du locuteur en reconnaît le sens du radical de base.

1.2 Repères mémoriels

Nous considérons que la valeur linguistique du mot correspond à sa charge mémorielle, laquelle permet que le mot soit inscrit dans la mémoire individuelle du locuteur ainsi que dans la mémoire collective de la société où ce mot s'emploie.

Bien entendu, à chaque niveau linguistique correspond une mémoire linguistique propre. Chaque langue est dotée en effet de ses propres mémoires phonologique, syllabique, morphématique, lexicale, propositionnelle, sémantique, etc. qui, respectivement, régissent la structuration, la génération, le stockage et la combinaison des phonèmes, des syllabes, des morphèmes, des mots, des propositions, de la charge sémantique des éléments linguistiques, etc.

Par ailleurs, les éléments de chaque niveau mémoriel restent en constante interaction entre eux et servent de base à l'élaboration du niveau mémoriel directement supérieur; les autres niveaux linguistiques servant de repères mémoriels supplémentaires dans le discours.

Or, comme les mots sont constitués de morphèmes, il devient utile de comprendre que la mémoire linguistique d'un mot se manifeste au niveau de chaque morphème lexical. Celui-ci fonctionne alors comme principal repère mémoriel pour guider la mémoire dans les processus mentaux de synthèse (construction) et d'analyse (compréhension) des unités lexicales.

Nous survolerons, ci-dessous, et à titre simplement illustratif, certains lieux d'interaction entre les principaux repères mémoriels lexicaux — à savoir les morphèmes — et les niveaux phonologique, lexical et sémantique, dans la structuration des mots.

1.2.1 Niveau phonologique

STRUCTURE PHONOLOGIQUE DE L'AUGMENT

L'augment, appelé encore prépréfixe par Lemaréchal (1989 : 41), est la voyelle initiale du substantif kiruúndi, toujours suivie du préfixe de classe. Elle ne peut être que l'une des 3 voyelles du triangle vocalique du kiruúndi⁵ — en l'occurrence [i], [a] ou [u] — et doit toujours être identique à la voyelle du préfixe.

(1)

- | | |
|-----|---|
| [i] | <i>imigoondooro</i> (<i>i-mi-goond-uur-o</i>) |
| [a] | <i>agatéerantĩmba</i> (<i>a-ka-ta-ir-a-i-n-tiimb-a</i>) |
| [u] | <i>umuuntu</i> (<i>u-mu-ntu</i>) |

Ainsi, la structure de l'augment prend pour repère principal la structure vocalique du préfixe de classe.

L'augment est la voyelle caractéristique de la catégorie substantivale où on la retrouve exclusivement⁶.

C'est également une voyelle qui subit de fréquents amuïssesments, notamment quand le substantif est à la forme appellative, lorsqu'il est précédé d'un référent, d'une négation, etc., et surtout lors des créations lexicales par combinaison de plusieurs mots (cf. tableau 1, colonne 2).

LOI DE DAHL

Il a été observé que, dans le discours, les consonnes sourdes des affixes précédant le radical deviennent sonores lorsque la consonne initiale du radical est sourde. Ce phénomène de dissimilation consonantique, particulièrement fréquent dans les langues bantu de la région orientale de l'Afrique, a été baptisé «loi de Dahl» par les bantuistes. En kiruúndi, il concerne les consonnes *k* et *t* — les seules consonnes non voisées des préfixes de classes — qui se transforment respectivement en *g* et *d*.

(2)

- | | |
|-----------------------|-----------------------------------|
| <i>Udukomére</i> | < <i>u-tu-kom-ir-e</i> |
| <i>Agatéerantĩmba</i> | < <i>a-ka-ta-ir-a-i-n-tiimb-a</i> |

⁵ Le kiruúndi a 5 voyelles, ainsi disposées dans le triangle vocalique :

[i]		[u]
	[e]	[o]
	[a]	

⁶ C'est ce qui a fait croire, à tort, à certains linguistes, que l'augment correspondait à l'article en kiruúndi.

Les exemples ci-dessus montrent ainsi comment le radical sert de repère pour la structuration phonologique des morphèmes qui le précèdent, justifiant ainsi les rapports d'interaction énoncés ci-haut.

HARMONIE VOCALIQUE

Bien souvent, des suffixes changent de voyelles selon le type de radical qui les précède. C'est le cas du suffixe *-ir* qui se métamorphose en *-er* sous l'influence du *o* du radical, dans les noms suivants :

- (3)
- | | | |
|-------------------|------------------|-------------------|
| <i>Agatóorero</i> | < a-ka-toor-ir-o | < a-ka-ta-ur-ir-o |
| <i>Udukomére</i> | < u-tu-kom-ir-e | |

Ces changements sont dus à une assimilation vocalique du suffixe, conditionnée par la voyelle pénultième du radical, dans une optique de rapprochement des timbres des voyelles concernées. Le radical devient ainsi un repère pour la structure vocalique du suffixe, comme il l'était pour la structure consonantique du préfixe selon la loi de Dahl.

1.2.2 Niveau lexical

SUBSTANTIVATION PAR PRÉFIXATION ET PRÉPRÉFIXATION

Toute unité lexicale, de quelque nature qu'elle soit, est susceptible d'être substantivée par adjonction d'un préfixe de classe et d'un augment :

- (4)
- | | |
|------------------------|--|
| <i>Ugutóoratoora</i> : | Le verbe <i>gutóoratoora</i> (ku-ta-ur-a-ta-ur-a) se nominalise en troquant le préfixe de l'infinitif <i>gu-</i> contre le préfixe <i>-ku-</i> de classe 15, naturellement précédé du prépréfixe correspondant <i>u-</i> . |
| <i>Ubugúfi</i> : | L'adjectif <i>-gufi</i> se transforme en substantif par préfixation en <i>-bu-</i> (classe 14) et par prépréfixation en <i>u-</i> . |

Ce mode de création lexicale permet de comprendre la portée de la charge mémorielle particulière du préfixe de classe et du prépréfixe correspondant, dans la structure lexicale.

REDOUBLEMENT THÉMATIQUE

Il existe un autre mode de création lexicale consistant à redoubler le thème, en vue d'ajouter un sens diminutif ou répétitif à la charge sémantique du radical :

- (5)
- | | |
|------------------------|---|
| <i>urukurúkuru</i> : | répétition du thème <i>-kurú</i> pour désigner une <i>rumeur</i> (<i>urukurúkuru</i>) relativement moins importante qu'une vraie <i>nouvelle</i> (<i>inkurú</i>). |
| <i>ugutóoratoora</i> : | redoublement du thème <i>-toora</i> (<i>ramasser</i>), pour signifier l'action de ramasser à plusieurs reprises. |

Ce mécanisme d'enrichissement lexical recourt donc aux mêmes repères mémoriels que la simple création lexicale, mais exploite doublement les repères entrant dans la constitution du thème nominal, en l'occurrence le radical, ses suffixes et la finale.

AGGLUTINATION MULTILEXICALE

En plus d'exploiter doublement certains morphèmes lexicaux pour créer de nouvelles unités lexicales, le kiruúndi, comme d'autres langues bantu, procède à la combinaison de plusieurs mots de radicaux et de sens différents pour créer de nouvelles unités lexicales à charge mémorielle plus dense :

(6)

Agatéerantíimba : agglutination de deux unités lexicales issues de deux thèmes, *-teera* et *-tiimba*

Umutázitéeraruungu : agglutination de deux unités lexicales issues de deux thèmes, *-teera* et *-ruungu*

Les mots formés par ce mode d'agglutination possèdent une charge mémorielle plus importante parce qu'issue de morphèmes de plus d'une unité lexicale; ce qui équivaut à des repères mémoriels plus nombreux et plus variés.

ACCORDS DE CLASSES

Au-delà des repères morphologiques au sein d'une même unité lexicale, le substantif kiruúndi peut exercer une certaine influence sur les mots en contact avec lui dans un énoncé. Par son influence sur les autres unités lexicales, le substantif transfère une partie de sa charge mémorielle en prenant toujours pour repère le préfixe de classe. Effectivement, dans n'importe quelle proposition kiruúndi, c'est la classe du préfixe nominal qui commande tous les accords et concordances des mots à morphologie variable, tels que le verbe, l'adjectif, le pronom, etc. Voici des exemples d'accord dans la phrase suivante :

(7)

abaantu bagúfi bagira udutáambwé tugúfi

/ a-**ba**-ntu / **ba**-gufi / **ba**-gir-a / u-**tu**-tamb-u-e / **tu**-gufi /

/ humains / courts / font / petits pas / courts /

'les gens de petite taille font de petits pas'

Cet énoncé illustre comment, d'une part, le préfixe **-ba-** (classe 2) de *abaantu* commande l'accord de l'adjectif *bagúfi* et celui du verbe *bagira*; tandis que d'autre part, le préfixe **-tu-** (classe 13) de *udutáambwé* commande l'accord de l'adjectif *tugúfi*, issu, tout comme *bagúfi*, du même adjectif *-gufi*, soumis à l'effet de deux classes différentes.

1.2.3 niveau sémantique

MÉMOIRE SÉMANTIQUE DE L'AUGMENT

À regarder de près la structure sémantique du substantif kiruúndi, il semble que l'augment ne représente pas d'intérêt particulier au niveau de la mémoire sémantique du mot.

Le peu d'intérêt sémantique que représente l'augment fait que son fréquent amuïssement ne nuise en rien à la valeur sémantique des unités lexicales concernées.

MÉMOIRE SÉMANTIQUE DES PRÉFIXES

Dans les langues bantu, le nombre de préfixes nominaux (classes) varie entre 5 et 20; le kiruúndi en compte 16, comme l'illustrent les exemples ci-dessous.

(8)

1 -mu-	ex. (u-mu-ntu) umuuntu = un humain
2 -ba-	ex. (a-ba-ntu) abaantu = des humains
3 -mu-	ex. (u-mu-vumu) umuvumú = ficus
4 -mi-	ex. (i-mi-vumu) imivumú = ficus
5 i- (-ri-)	ex. (i-i-vi) iví = genou
6 -ma-	ex. (a-ma-vi) amaví = genoux
7 -ki-	ex. (i-ki-ntu) ikiintu = chose
8 -bi-	ex. (i-bi-ntu) ibiintu = choses
9 -n-	ex. (i-n-ka) inká = vache
10 -n-	ex. (i-n-ka) inká = vaches
11 -ru-	ex. (u-ru-ntu) uruuntu = grosse chose
12 -ka-	ex. (a-ka-ntu) akaantu = petite chose
13 -tu-	ex. (u-tu-ntu) utuuntu = petites choses
14 -bu-	ex. (u-bu-ntu) ubuuntu = personnalité
15 -ku-	ex. (u-ku-ntu) ukuuntu = manière
16 -ha-	ex. (a-ha-ntu) ahaantu = lieu

Ce dénombrement des préfixes de classes ne tient pas compte de l'opposition singulier/pluriel, car dans les habitudes linguistiques bantuistes, on considère qu'un mot au singulier appartient à une classe différente de celle du même mot au pluriel.

Par ailleurs, on remarque que des mots appartenant aux mêmes catégories sémantiques⁷ ont tendance à se regrouper au sein des mêmes classes.

Très souvent, en effet, les classes :

1/2	sont réservées aux humains;
3/4	regroupent beaucoup de noms de végétaux;
5/6	sont utilisées pour les parties du corps;
6	à elle seule désigne les liquides et les masses;
7/8	- se retrouvent dans beaucoup de noms d'animaux, - et donnent parfois un sens augmentatif;
9/10	regroupent également beaucoup de noms d'animaux;
11	est augmentative;
12/13	donnent aux substantifs une valeur diminutive, l'augmentatif et le diminutif pouvant, bien sûr, s'interpréter avec une nuance méliorative ou péjorative;
14	désigne des réalités abstraites ou le diminutif;
15	est consacrée - aux infinitifs substantivés,

⁷ Dans ce contexte, la notion de «catégorie sémantique» serait à rapprocher de celle de «classe d'objets» (cf. Gross et Clas 1997).

16 - aux différentes parties du corps (comme 5/6);
regroupe les diverses modalités de lieux.

Le sémantisme ci-dessus n'est cependant pas à prendre au sens absolu. En effet, il n'est pas rare de rencontrer ça et là, d'une part, des mots appartenant aux mêmes catégories sémantiques alors qu'ils portent des préfixes de classes différentes; tout comme, d'autre part, des mots de même préfixe de classes peuvent appartenir à des catégories sémantiques bien distinctes. Cependant, le caractère sémantique flou dont il est ici question ne concerne que certains préfixes kiruúndi; car il semblerait que les dix-neuf préfixes de classes du proto-bantu (Guthrie 1971 et Meeussen 1967) aient eu des valeurs sémantiques nettes, une netteté qui se retrouve en totalité ou en partie dans l'une ou l'autre langue bantu (Spitulnik 1987).

MÉMOIRE SÉMANTIQUE DES RADICAUX

Étant donné la multiplicité des valeurs sémantiques qui peuvent être générées par un même morphème radical, on peut soutenir, sans risque de se tromper, que la valeur sémantique du morphème radical est rarement monovalente. Bien souvent, que ce soit dans les langues bantu ou dans bien d'autres langues, le radical est sémantiquement polyvalent, ambigu, parfois même flou.

En effet, lorsqu'on observe en (8) les significations si différentes des unités lexicales formées à partir du radical -ntu, il devient extrêmement difficile de donner le sens exact du dit radical.

Par ailleurs, il n'est pas rare que deux mots de radicaux identiques expriment des idées contraires, tout simplement à cause de la valeur sémantique de leurs affixes (préfixes, infixes ou suffixes).

Tenant compte des faits évoqués ci-dessus, on peut donc affirmer que la mémoire sémantique d'un mot se modèle et se clarifie au fur et à mesure que les autres morphèmes lexicaux s'agglutinent au morphème radical. Seul le résultat final de cette agglutination permet de clarifier le sens du mot, qui reste, malgré tout, axé sur le sens du radical.

Malgré ses contours sémantiques flous, le morphème radical demeure le repère-clé dans la construction du sens des mots, principalement grâce à la force d'attraction que le radical exerce sur les autres morphèmes venant clarifier son sémantisme, mais également par son adaptabilité, lui permettant de se combiner avec divers affixes, en s'appropriant, en quelque sorte, leur valeur sémantique.

MÉMOIRE SÉMANTIQUE DES AUTRES MORPHÈMES

Contrairement à la charge sémantique de l'augment, du préfixe et du radical, celle des autres morphèmes lexicaux est assez nette, ce qui leur permet d'influencer de manière déterminante la mémoire sémantique des mots qu'ils constituent.

Voici, par exemple, la signification nettement exprimée par certains morphèmes du tableau 1 :

(9)

- les infixes :

-ha- :	lieu :	<i>abatáhahára (a-ba-ta-ha-har-a)</i>
-ii- :	réfléchi :	<i>umwímenyerezo (u-mu-ii-meny-ir-ir-i-o)</i>
-ta- :	négation :	<i>abatáhahára (a-ba-ta-ha-har-a)</i>

- suffixes :

-an- :	ensemble :	<i>umutúuhúrana (u-mu-ta-uh-ur-an-a)</i>
-ir- :	applicatif :	<i>agatéerantúmba (a-ka-ta-ir-a-n-tiimb-a)</i>
-uur- :	reversif :	<i>imigondooro (i-mi-goond-uur-o)</i>

- finales :

-a- :	résultat :	<i>Umutúuhúrana (u-mu-ta-uh-ur-an-a)</i>
-e- :	état :	<i>urushúngé (u-ru-shiing-e)</i>
-o- :	action/résultat :	<i>umwímenyerezo (u-mu-ii-meny-ir-ir-i-o)</i>

Tous ces morphèmes servent de repères à la mémoire sémantique du mot et ont chacun une valeur sémantique propre, mais pas totale, car non autonome. Toutes ces valeurs sémantiques se prennent mutuellement pour référence et se repèrent principalement à la valeur sémantique du radical qu'elles viennent compléter.

2. AU-DELÀ DES SUBSTANTIFS

Les morphèmes lexicaux décrits dans les paragraphes précédents ne sont pas la propriété exclusive de la catégorie nominale. En effet, grâce aux nombreux mécanismes de création lexicale, et plus particulièrement la dérivation, plusieurs unités lexicales issues de catégories différentes partagent les mêmes morphèmes, et par conséquent, les mêmes repères mémoriels. Elles partageront, entre autres, le morphème radical qui, comme nous l'avons déjà signalé, sert à construire une quantité foisonnante de mots appartenant à des catégories lexicales variées.

En outre, même si les mots entrant dans la construction d'un énoncé possèdent, chacun, une mémoire particulière avec des repères propres, ils doivent tous partager des repères mémoriels communs, afin d'assurer une cohérence grammaticale de l'énoncé. C'est pour rester conforme à cette cohérence que, en kiruúndi, comme évoqué en 1.2.2, les mots prédicatifs s'accordent en classe et en nombre avec le substantif argument de chaque énoncé minimal. Bien entendu, pour ce genre d'opérations le repère mémoriel le plus influent reste le préfixe de classe qui, dans les langues bantu, commande tous les accords. C'est ainsi que, pour sa charge grammaticale, le préfixe de classe est considéré comme repère mémoriel par excellence.

Les limites du présent travail ne nous permettent pas de rentrer dans les détails des fonctionnements morphologiques des différentes unités lexicales, car ils ressemblent, à quelques différences près, à ceux du substantif. Du reste, ce dernier n'a pas, lui non plus, été analysé à fond, à cause de ces contraintes liées à la nature même de ce travail. À la section précédente, nous avons voulu tout simplement montrer que les structures morphologiques et les repères mémoriels qu'ils représentent, au sein d'unités lexicales comme le verbe, l'adjectif, le pronom, etc. sont similaires à ceux du substantif. De plus, comme vu précédemment, toutes ces unités lexicales sont en constante interaction avec le substantif qui en commande les accords.

3. AU-DELÀ DU KIRUÚNDI

Tout comme nous n'avons pas voulu limiter nos analyses au seul substantif kiruúndi, nous voudrions, de la même manière, étendre leur portée, ne fût-ce que sommairement, à l'ensemble des langues bantu. En effet, la structure morphologique du nom kiruúndi semble correspondre au modèle morphologique type des noms en langues bantu; car, comme nous l'avons déjà signalé ça et là, lors de nos commentaires, l'essentiel des caractéristiques morphologiques du kiruúndi se retrouvent dans toutes les langues bantu.

Effectivement, mis à part l'augment absent du substantif dans l'une ou l'autre langue bantu, toutes partageant l'essentiel des morphèmes décrits en 1.

Du reste, si Bleek (1862) a donné le nom *bantu* à la famille linguistique dont il est ici question, c'est parce que les études historico-comparativistes ont montré que toutes les langues appartenant à cette famille utilisent le mot *muntu* (ou des formes phonétiques apparentées) pour nommer le concept de 'personne humaine', tandis que le nom *bantu* (ou des formes phonétiquement apparentées) désigne le même concept au pluriel; d'où les préfixes de classes 1 et 2 (-*mu*- et -*ba*-) réservés aux humains.

Les langues bantu disposent donc d'un héritage linguistique commun, aisément détectable à travers leurs comportements linguistiques, et plus particulièrement au niveau des structures morphologiques. Ce sont ces comportements linguistiques qui ont permis de reconstituer l'ancêtre commun de ces langues avec une relative facilité, et d'en comprendre les mécanismes de différenciation.

Par des distances linguistiques — liées, entre autres, aux mouvements des populations et à l'effet du temps — les membres de cette famille linguistique ont progressivement pris des identités linguistiques différentes, en modifiant petit à petit les structures phonologiques ainsi que les valeurs linguistiques de certains morphèmes.

Malgré leurs différences, les langues bantu ont cependant gardé un grand stock d'éléments linguistiques communs et des comportements linguistiques communs, notamment :

- la création lexicale régulière et prolifique, par agglutination autour du radical, la plupart des morphèmes radicaux étant communs à plusieurs langues;
- l'accord grammatical strict de la classe, le repère mémoriel classificatoire étant le préfixe nominal;
- ...

L'on comprend donc que les langues bantu avaient, au départ, les mêmes repères mémoriels, dont les plus importants étaient le radical et le préfixe de classe. Ces morphèmes sont toujours considérés comme les deux éléments lexicaux fondamentaux, car ils sont indispensables dans la construction de n'importe quelle unité lexicale en langue bantu.

C'est pour leur nature lexicale essentielle que nous avons pris le préfixe et le radical comme modèles illustratifs des repères mémoriels, le premier comme repère mémoriel par excellence pour sa charge grammaticale, et le second comme repère mémoriel d'importance capitale au niveau sémantique.

4. AU-DELÀ DE LA MÉMOIRE LINGUISTIQUE

Nous avons essayé de montrer, ci-dessus, comment les langues bantu sont organisées en de riches et réguliers réseaux de morphèmes lexicaux, qui constituent des repères incontournables de la mémoire linguistique des énoncés et de leurs constituants.

Après avoir observé des réseaux si exubérants et en même temps si bien organisés, nous croyons que les éléments de ces structures morphologiques gagneraient à être répertoriés et classés dans des dictionnaires, tout comme on le fait à grande échelle pour les unités lexicales, et dans une moindre mesure pour les locutions.

Bien entendu, pour être optimalement exploitables, ces structures seraient répertoriées dans des dictionnaires électroniques, afin de permettre un accès rapide à l'information et un inventaire le plus exhaustif de données, sans beaucoup de contraintes spatiales.

De plus, pareil dictionnaire permettrait que les repères morphématiques dépassent le niveau de l'habituelle mémoire linguistique — individuelle et collective — exclusivement réservée à ceux qui maîtrisent la langue, pour atteindre la mémoire numérique, directement accessible par la machine, et par conséquent, théoriquement exploitable par tous, y compris les non-locuteurs de la langue concernée.

Pour une meilleure efficacité, le dictionnaire dont il est question ici serait structuré comme une base de données contenant les listes des différents morphèmes⁸ avec leurs valeurs linguistiques respectives, comme la position du morphème dans le mot, ses cooccurents, sa valeur sémantique, etc. Toutes ces informations seraient progressivement rentrées dans la base de données au fur et à mesure de la décomposition morphologique de mots, tirés de textes expressément choisis à cet effet.

Nous avons vu, par ailleurs, comment les morphèmes du kiruúndi sont organisés en réseaux de repères mémoriels. Ces réseaux pourraient aisément servir de guides pour créer des liens dans la base de données. Celle-ci serait, par conséquent, une base de données relationnelle, susceptible de permettre une reconstitution automatique des liens intra-lexicaux unissant les différents morphèmes d'un mot, ainsi que les liens extra-lexicaux qui commandent les accords de classes dans un énoncé.

Une base de données pareille prendrait sûrement beaucoup de temps et d'efforts avant d'être exploitable. En retour, eu égard à la représentativité, à la richesse et à la variété de ses enregistrements, la base de données serait d'un intérêt linguistique certain, primordialement en linguistique théorique, à cause des nombreuses analyses statistiques qu'elle permettrait de réaliser, comme par exemple :

- établir la fréquence d'utilisation de tel ou tel morphème lexical;

⁸ Comme les radicaux forment des listes ouvertes, la base de données enregistrerait les morphèmes radicaux au fur et à mesure qu'ils seraient identifiés.

- résoudre la question non encore bien tranchée (Hyman 1993; Meeussen 1967 etc.) au sujet de la compatibilité et de l'ordre des affixes⁹.

Bien entendu, pour une plus grande efficacité, la base de données serait couplée à une interface d'analyse et de synthèse des unités lexicales, qui comprendrait, entre autres :

- un programme intégrant les règles phonologiques régissant les combinaisons des morphèmes, ainsi que les règles d'accords et de concordances;
- un programme de découpage morphologique, capable de décomposer chaque mot en ses morphèmes. Ce programme fonctionnerait comme les programmes de parsing qui décomposent les propositions en vue d'en dégager les unités lexicales constitutives;
- d'autres informations linguistiques pertinentes, au sujet des mécanismes de découpage ou de combinaisons morphologiques;
- ...

Évidemment, cette interface analytico-synthétique aurait comme dictionnaire-repère la base de données, avec ses unités lexicales et morphologiques. Ceci signifie que l'interface se servirait des mots et des morphèmes de la base de données comme modèles d'identification d'autres mots dans les textes à analyser ou pour vérifier l'exactitude des mots et morphèmes reconstitués.

Ainsi, du point de vue analytique, et à l'aide de l'interface ci-dessus, cette base de données permettrait d'analyser des textes, pour y détecter les unités lexicales et en extraire les morphèmes lexicaux. Ces derniers pourraient être réintroduits dans la base de données et servir ainsi à l'enrichir.

Par ailleurs, comme la base de données contiendrait les valeurs sémantiques de ses morphèmes, l'étape précédente d'extraction des morphèmes pourrait être complétée par leur interprétation sémantique, et contribuer ainsi à la compréhension de mots analysés, éventuellement à leur traduction, peut-être même à la traduction automatique de certains textes.

Sur le plan de la synthèse automatique, et par le truchement de l'interface décrite précédemment, la base de données serait également la source tout indiquée pour la génération automatiquement des mots, comme ceux naturellement produits par les locuteurs de langues bantu, par combinaison des différentes valeurs sémantiques des morphèmes. Ces mécanismes de création synthétique des mots seraient de précieux adjuvants accélérant la création de nouveaux termes, principalement pour désigner les concepts techniques nouveaux¹⁰. Cependant, personne ne devrait en exagérer l'utilisation, sans risque de retirer aux mots leur essence naturelle. La génération de mots de synthèse ne serait qu'une solution de dernier recours, utilisable seulement si la mémoire linguistique naturelle est à cours

⁹ On sait, jusqu'à présent, que plusieurs infixes et suffixes peuvent se retrouver dans une même unité lexicale, mais ils ne se combinent pas de manière anarchique, tout comme ils n'acceptent pas n'importe quel morphème. À l'heure actuelle, il ne semble pas y avoir, en ce qui concerne le kiruúndi en tout cas, d'étude pointue sur ces règles — que tout locuteur suit sans s'en rendre compte — régissant l'ordre et la combinabilité de ces affixes.

¹⁰ Nous avons déjà expérimenté pareilles procédures en générant automatiquement un vocabulaire kiruúndi de termes de cuisine ainsi que leurs traductions françaises.

d'inspiration; et le produit de cette génération devraient impérativement être soumis à la vérification et à l'approbation d'un nombre significatif de locuteurs usagers avant d'être soumis au grand public.

CONCLUSION

Nous sommes parti d'une analyse du substantif kiruúndi, pour illustrer la régularité des structures morphologiques de l'ensemble du lexique kiruúndi. Nous avons montré comment ces structures fournissent à la langue une mémoire lexicale très prolifique et en même temps rigoureusement ordonnée par les morphèmes lexicaux. C'est pourquoi les morphèmes ont été décrits comme des repères indispensables dans l'élaboration de la mémoire lexicale, une mémoire agissante à tous les niveaux linguistiques.

Comme la langue kiruúndi a beaucoup d'affinités avec les autres langues bantu, il n'a pas été difficile de justifier que la mémoire lexicale du kiruúndi est similaire à celles des autres langues-sœurs.

Nous avons ensuite parlé des possibilités de classer les éléments de cette mémoire linguistique, à la fois si abondante et si régulière, en stockant dans une base de données les différents morphèmes, ainsi que leurs valeurs linguistiques respectives.

Une base de données relationnelle comme celle décrite ci-dessus est susceptible de nombreuses applications, en linguistique comme dans d'autres disciplines connexes. Nous avons sommairement évoqué certaines des applications que cette base de données pourrait permettre de réaliser, que ce soit en linguistique théorique ou en linguistique appliquée à l'analyse ou à la synthèse automatique des unités lexicales.

Dans les conditions optimales, notre modèle de base de données pourrait devenir contrastif et contenir des informations linguistiques en provenance de plusieurs langues bantu, afin d'offrir une vue plus large de leurs repères mémoriels lexicaux. Ainsi, la valeur linguistique de chaque morphème dans une langue serait mise en parallèle avec la valeur linguistique du même morphème dans les autres langues bantu. Conséquemment, les valeurs linguistiques identiques dans différentes langues seraient également mises en relation avec les morphèmes correspondants.

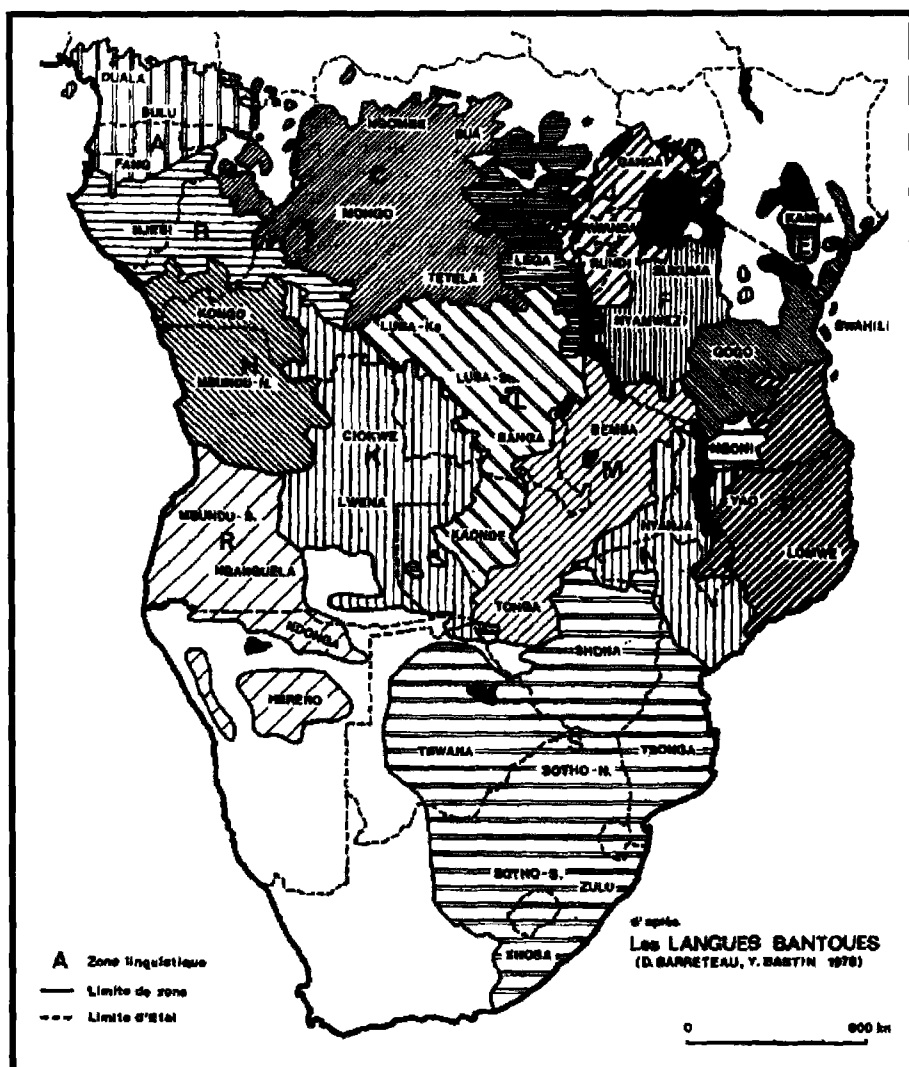
Pour donner une vue d'ensemble des repères mémoriels lexicaux en langues bantu, l'aspect contrastif porterait sur le maximum possible de langues, et idéalement sur toutes. Cependant, ce projet paraît trop gros pour pouvoir être entrepris dans l'immédiat.

Malgré tout, si le projet de base de données se montre concluant dans l'une ou l'autre langue, il sera alors possible de l'étendre, petit à petit, sur d'autres langues¹¹ et d'entreprendre une mise en relation de leurs bases de données, créant ainsi une méga base de données multilingues. Seul l'avenir nous dira si le rêve que nous caressons depuis longtemps, de créer une base de données contrastives sur l'ensemble des données morpho-lexicales des langues bantu, serait réalisable ou non.

¹¹ Toutes les langues devront être, bien entendu, préalablement soumises au même modèle de description linguistique.

RÉFÉRENCES

- BASTIN, Y. (1985) : *Les relations sémantiques dans les langues bantoues*, Bruxelles, ARSOM, 86 p.
- BENDOR-SAMUEL, J. & R. L. HARTELL (1984) : *The Niger-Congo Languages : a classification and description of Africa's largest language family*, New York, University Press of America, 505 p.
- BLEEK, W. H. I. (1862) : *A comparative grammar of South African Languages*, part 2, Cape Town, Juta.
- FAÏK-NZUJI, C. M. (1992) : *Éléments de phonologie et de morphologie des langues bantu*, Louvain-la-Neuve, Peeters, 163 p.
- FODOR, I. (1982) : *A fallacy of contemporary linguistics. J. H. Greenberg's classification of the african languages and his «comparative method»*, Hamburg, Buske.
- GREENBERG, J. H. (1970) : *The languages of Africa*, 3rded., Bloomington, Indiana University, 180 p.
- GROSS, G. et A. CLAS (1997) : «Synonymie, polysémie et classes d'objets», *Meta*, vol 42, n° 1, (mars 1997), Montréal, PUM, pp. 147-154.
- GUTHRIE, M. (1971) : *Comparative bantu*, Farnborough, Gregg International Publishers, vol. 1-143 p., vol. 2-180 p., vol. 3-326 p., vol. 4-248 p.
- HYMAN, Larry M. (1993) : «Conceptual issues in the comparative study of the bantu verb stem», Mufwene S. S. & Moshi L., *Topics in African Linguistics*, Philadelphia, John Benjamins Publishing Company, pp. 3-34.
- LEMARÉCHAL, Alain (1989) : *Les parties du discours. Sémantique et syntaxe*, Paris, PUF, 272 p.
- MANN, M. & D. DALBY (1987) : *A thesaurus of african languages : a classified and annotated inventory of the spoken languages of Africa with an appendix on their written representation*, London, Hans Zell Publishers, 325 p.
- MEEUSSEN, A. E. (1959) : *Essai de grammaire rundi*, Tervuren, AMRAC, 236 p.
- MEEUSSEN, A. E. (1967) : *Bantu grammatical reconstructions*, Tervuren, AMRAC, 236 p.
- MEINHOF, C. (1899) : *Grundriss einer Lautlehre der Bantusprachen nebst einer Anleitung zur Aufnahme von Bantusprachen*, Berlin, Dietrich Reimer.
- MONINO, Y. (1995) : *Le proto-gbaya : Essai de linguistique comparative sur vingt-et-une langues d'Afrique centrale*, Paris, Peeters, 725 p.
- MUFWENE S. S. & L. MOSHI (1993) : *Topics in African Linguistics*, Philadelphia, John Benjamins Publishing Company, 307 p.
- NTAKIRUTIMANA, J. B. (1993) : *La créativité lexicale en kiruúndi, Étude lexico-sémantique de la néologie dans les manuels de lecture à l'école primaire*, mémoire de Maîtrise, Montréal, Université de Montréal, 119 p.
- NTAKIRUTIMANA, J. B. (1993) : *Vers une traduction automatique kiruúndi-français : Approche présémantique des substantifs de la terminologie culinaire*, Séminaire de Traduction automatique, hiver 1993, travail de fin de session, Montréal, Université de Montréal, (inédit).
- RODEGEM, F. M. (1970) : *Dictionnaire rundi-français*, Tervuren, AMRAC, 644 p.
- SPITULNIK, D. A. (1987) : *Semantic Superstructuring and infrastructuring : Nominal classes struggle in Chibemba*, Bloomington, Indiana University Linguistic Club, 146 p.



Carte des langues bantoues

LES CANADIANISMES EN FRANÇAIS ET EN ANGLAIS¹

Roda P. ROBERTS et Chantale GRENON-NYENHUIS

Université d'Ottawa, Canada

INTRODUCTION

Comme le signalent Jean Dubois et Claude Dubois (1971 : 8), les dictionnaires sont «des objets culturels, intégrés comme tels à cette culture : ils témoignent d'une civilisation». Cet avis, partagé par beaucoup, est fondé sur le double constat que le dictionnaire a pour vocation première d'informer sur la langue, et que la langue et la culture sont étroitement liées (Rialland-Addach 1995 : 91). En effet, le dictionnaire s'articule sur deux plans en termes hjelmsléviens : celui de l'expression, «une langue et son lexique, des usages et leurs vocabulaires, certains discours et leurs occurrences», et celui du contenu, «un univers exprimé, des visions culturelles du monde, des articulations conceptuelles correspondant à des classes extensionnelles» (Rey et Delasalle 1979 : 5).

Mais si le dictionnaire unilingue est bien reconnu comme «*le livre privilégié de référence* à la connaissance et au savoir linguistique et culturel» d'une communauté (Dubois et Dubois 1971 : 8), le dictionnaire bilingue l'est beaucoup moins. Selon Dubois et Dubois (1971 : 34-35), le dictionnaire bilingue repose sur l'identité supposée des concepts et des manières de percevoir le monde dans deux langues; ainsi, «les cultures transmises peuvent être ramenées à un certain nombre d'invariants communs : les visions du monde sont identiques, imposées par les principes logiques qui sont à la base du fonctionnement d'une pensée universelle». En d'autres termes, selon eux, le dictionnaire bilingue présente la culture partagée plutôt que les particularités culturelles de chaque langue, car il cherche à donner des paires de synonymes. Cependant, on commence peu à peu à reconnaître qu'on ne peut s'attendre à trouver des «équivalents complets» dans le dictionnaire bilingue, car «*the conceptual world evolves differently in different languages as a result of, for example, historical, geographical, social, cultural, and economic differences between the countries where the different languages are used*» (Svensén 1993 : 140). On commence à accepter l'idée que le dictionnaire bilingue peut inclure des équivalents partiels ou même aucun équivalent, pour autant que le lexicographe identifie et indique les différences entre les deux langues (Svensén 1993 : 142). Cette nouvelle façon de voir le rôle du dictionnaire bilingue en fait, lui aussi, un ouvrage culturel, qui reflète non pas la culture d'une communauté linguistique, ni même la culture partagée, mais deux cultures différentes.

¹ Cette recherche a été subventionnée par le Conseil de recherches en sciences humaines du Canada.

Un tel ouvrage est d'autant plus nécessaire quand deux langues et donc deux cultures coexistent, comme c'est le cas au Canada avec l'anglais et le français. Même si ces deux langues sont utilisées sur le même territoire, la culture véhiculée par chacune n'est pas la même. En outre, ce qui complique d'autant plus la production d'un dictionnaire canadien bilingue, c'est que l'anglais et le français du Canada diffèrent sur certains points du français de France et de l'anglais britannique et américain, car ces deux langues ont connu une évolution différente dans ce pays et reflètent des réalités différentes. En d'autres mots, l'anglais et le français canadiens ont des particularités linguistiques et culturelles non seulement l'un par rapport à l'autre, mais aussi par rapport à l'anglais et au français tels qu'ils sont utilisés ailleurs. Ces particularités linguistiques canadiennes, souvent appelées «canadianismes», se manifestent dans la prononciation, la syntaxe, la morphologie, mais surtout le vocabulaire. Ce sont les canadianismes lexicaux qui seront examinés ici dans l'optique de la production d'un nouveau dictionnaire bilingue, le *Dictionnaire bilingue canadien* (DBC), qui, ses rédacteurs l'espèrent tout au moins, sera l'ouvrage culturel par excellence de ce pays bilingue².

1. L'ORIGINE DES CANADIANISMES

On peut attribuer l'origine des canadianismes en anglais ainsi qu'en français aux circonstances dans lesquelles ces langues se sont implantées au Canada. Le fait que l'historique du français au Canada diffère de celui de l'anglais au Canada explique en grande partie les divergences entre les canadianismes dans ces deux langues.

1.1 Historique du français canadien

Le Canada est découvert en 1534 par Jacques Cartier, qui prend possession du territoire au nom du roi de France. Le français est donc la première langue européenne qui tente de conquérir ce vaste territoire. Mais les ambitions françaises se limitant aux richesses qui peuvent être acquises dans ce pays, la Nouvelle-France n'est pas immédiatement colonisée. Ce n'est qu'après la fondation de la ville de Québec par Samuel de Champlain en 1608 et la fondation subséquente de Ville-Marie et de Trois-Rivières que la population francophone commence à s'accroître, quoique à un rythme relativement lent. Nous parlons de la population francophone de l'époque, même s'il est difficile de déterminer si le français s'est développé avant ou après l'arrivée des immigrants venus de France. Selon l'hypothèse traditionnelle, les patois, originaires surtout de l'Île-de-France, de l'Ouest et du Nord-Ouest, se sont rencontrés en arrivant en Nouvelle-France et ont constitué la base de ce qui est devenu le français canadien. Cependant, d'après une nouvelle théorie (Mougeon et Beniak 1994 : 1-55), il est probable que les immigrants, qui devaient passer un certain temps dans la même ville avant de s'embarquer pour l'Amérique, aient appris le français comme langue commune avant d'entreprendre le voyage.

Quoi qu'il en soit, le français au Canada est influencé dès ses débuts par les langues autochtones, dont il emprunte certains vocables pour désigner de nouvelles réalités auxquelles sont exposés les Européens lors de leur arrivée sur le continent nord-américain

² Il s'agit d'un projet interuniversitaire, mené en collaboration par l'Université d'Ottawa, l'Université de Montréal et l'Université Laval, qui vise la production en 2003 d'un dictionnaire anglais-français, français-anglais reflétant les usages canadiens.

(ex. **babiche**, **boucane**, et **toboggan**, issus de langues amérindiennes, et **mukluk** et **umiak**, relevés des dialectes inuit).

Mais l'impact critique sur le français canadien vient de l'anglais, qui commence à s'établir en Amérique du Nord à partir de 1713, quand les Britanniques s'installent sur la côte est de ce qui constitue maintenant les États-Unis, dans le but de coloniser ce nouveau territoire. En peu de temps, la population de langue anglaise devient supérieure à celle de langue française. Les Britanniques commencent à revendiquer une partie des terres canadiennes et, en 1759, s'emparent de la ville de Québec. Enfin, en 1763, le Traité de Paris oblige la France à céder la Nouvelle-France aux Anglais et le français du Canada est isolé du français de France. C'est à cette époque que l'anglais est introduit dans la colonie et qu'une certaine assimilation des francophones et de leur culture commence.

Si l'Acte de Québec, signé en 1774, permet aux francophones de revenir au système de droit civil français et d'occuper des postes politiques tout en continuant de pratiquer la religion catholique, le rapport de Lord Durham, rédigé à la suite des Rébellions des Patriotes en 1837-1838, préconise l'assimilation des francophones (Mallory 1984 : 340) et recommande l'union du Bas-Canada (majoritairement francophone) et du Haut-Canada (majoritairement anglophone). En 1840, l'Acte d'Union entre en vigueur et la politique d'assimilation de la population francophone se poursuit.

Mais l'Église catholique, qui exerce une grande influence sur les francophones du Bas-Canada, s'assure de garder séparées la population francophone et la population anglophone. Ainsi, le français survit, du moins au Bas-Canada, même si les liens avec le français de France sont rompus et si le français occupe une position d'infériorité par rapport à l'anglais. Mais le français évolue différemment sur les deux continents; en Amérique du Nord, son évolution est influencée par l'anglais. Il faut attendre la Révolution tranquille des années 60 pour qu'un effort concerté soit fait afin de libérer le français de l'anglais et pour améliorer la qualité du français canadien. Grâce aux travaux terminologiques de l'Office de la langue française, créé en 1961, et aux lois linguistiques approuvées par le Gouvernement du Québec, le français reprend sa vigueur et devient un outil de travail quotidien.

Ce bref résumé historique explique d'abord pourquoi il y a des canadianismes en français et donne ensuite une indication des types de canadianismes qui s'y retrouvent : emprunts aux langues autochtones, mots et sens tombés en désuétude en France, anglicismes, et terminologie moderne particulière.

1.2 Historique de l'anglais canadien

Comme nous l'avons signalé précédemment, les Britanniques s'installent en Amérique du Nord après les Français. Comme ces derniers, ils empruntent des mots autochtones pour décrire les réalités nord-américaines; ces emprunts s'effectuent généralement directement aux langues autochtones, mais parfois aussi à la suite des contacts avec la population francophone qui avait déjà fait des emprunts à ces langues.

Devenant rapidement supérieurs en nombre, les Anglais réclament le territoire français au Canada et l'acquièrent après le Traité de Paris en 1763. À partir de cette date, l'anglais établit sa dominance au Canada.

Mais de quel anglais s'agit-il ? Les anglophones qui viennent s'installer au Canada ne sont pas uniquement originaires de Grande-Bretagne. En fait, beaucoup d'Américains immigrent vers le nord, surtout après la Déclaration d'indépendance des États-Unis en 1776, quand un bon nombre de Tories et de Loyalistes sont obligés de s'exiler à cause de leurs convictions politiques. Ainsi se développe petit à petit un anglais canadien qui n'est ni britannique ni américain, mais qui possède certaines caractéristiques des deux langues, car les anglophones du Canada, contrairement aux francophones, ont toujours gardé contact avec la Grande-Bretagne et les États-Unis. En outre, bien entendu, l'anglais canadien a acquis un certain nombre de mots pour désigner des réalités typiquement canadiennes. Mais la grande ressemblance de l'anglais canadien avec l'anglais américain, d'une part, et l'anglais britannique, de l'autre, a longtemps entravé la reconnaissance des particularités de cette variété de langue.

Ce bref historique de l'anglais canadien laisse supposer que la plupart des canadianismes dans cette langue seraient des mots autochtones et des mots désignant les réalités propres au Canada.

2. IDENTIFICATION DES CANADIANISMES

Ce sont ces canadianismes en anglais et en français qui reflètent la culture typiquement canadienne. Il est donc de rigueur de les inclure dans un dictionnaire canadien bilingue. Mais comment les identifier sans effectuer une comparaison systématique, qui serait longue et pénible, de tout le vocabulaire de l'anglais canadien avec les vocabulaires américain et britannique et une comparaison du vocabulaire du français canadien avec celui du français hexagonal ? Le DBC a utilisé quatre types de sources différentes pour y arriver.

Le point de départ était la documentation lexicographique. D'abord, nous avons examiné les dictionnaires canadiens anglais et français pour constituer des listes préliminaires. Nous nous sommes servis principalement du *Gage Canadian Dictionary* (GAGE) (1983), du *Funk and Wagnalls Canadian College Dictionary* (FUN) (1989) et du *Penguin Canadian Dictionary* (PEN) (1990) pour l'anglais, ainsi que du *Dictionnaire du français Plus* (PLUS) (1988) et du *Dictionnaire québécois d'aujourd'hui* (RQ) (1992) pour le français. En dépouillant les dictionnaires canadiens anglais, nous avons repéré tous les mots et les emplois marqués «Canadian». Et quand ces ouvrages n'étaient pas d'accord sur la marque géographique (c'est-à-dire si le mot ou l'emploi n'était pas marqué dans un des dictionnaires ou portait une autre marque géographique comme «North American»), nous avons fait des vérifications supplémentaires dans un dictionnaire américain, le *Random House Webster's College Dictionary* (RHWEB) (1991), ainsi que dans un dictionnaire britannique, le *Collins English Dictionary* (COLL) (1986). Le processus d'identification des canadianismes français dans les dictionnaires n'était pas identique, pour la simple raison que les dictionnaires canadiens mentionnés ci-dessus ont comme politique de ne pas marquer ce qui est canadien, car ils prennent comme point de référence le français du Canada. Ainsi, il a fallu laisser le repérage des canadianismes français à nos spécialistes du français canadien, qui les identifiaient d'abord en jumelant l'examen des deux dictionnaires canadiens et leurs propres connaissances linguistiques, puis en cherchant les éléments retenus dans les dictionnaires hexagonaux (le *Petit Robert*, le *Lexis* et le *Petit Larousse*) : l'absence de ces éléments dans ceux-ci ou leur marquage comme mot «canadien» permettait de confirmer leur statut de canadianisme. Les listes préliminaires des canadianismes anglais et français, dressées comme nous l'avons indiqué ci-dessus étaient ensuite élargies en

consultant d'autres ouvrages lexicographiques plus restreints et plus spécifiques, comme le *Concise Dictionary of Canadianisms* (CDC) (1973), *Richesses et particularités de la langue écrite au Québec* (RP) (1980), et le *Dictionnaire du français canadien* (DCF) (1990), ainsi que le fichier lexical du Trésor de la langue française au Québec (TLFQ)³.

Des consultants et des informateurs constituaient notre deuxième source d'information. Nous avons déjà signalé le rôle essentiel que jouaient nos spécialistes du français canadien dans l'établissement de la liste initiale. Ceux-ci, à leur tour, ont consulté au besoin les lexicographes du TLFQ. Pour l'anglais, nous avons tenté de faire une consultation plus large en faisant appel aux associations d'études canadiennes, sans grand succès, cependant.

Comme troisième source d'information sur les canadianismes, nous avons utilisé des articles traitant de l'anglais et du français canadiens. Beaucoup d'entre eux sont des articles de recherche (ex. R.J. Gregg 1989 : 151-187; C. Poirier 1980 : 43-89). D'autres sont des articles journalistiques. L'examen de ces articles nous a permis de nous assurer que nous avions relevé tous les canadianismes importants.

Enfin, la quatrième source d'information, peut-être la plus innovatrice, était, et continue d'être, des corpus de textes. Le DBC se sert principalement de deux bases de données textuelles : TEXTUM et TRANSBASE. TEXTUM, qui a été établie en fonction des besoins du DBC, contient des textes surtout journalistiques, mais aussi littéraires et scientifiques; tandis que la plupart des textes sont canadiens d'origine, cette base de données inclut un certain nombre de textes du français hexagonal et de l'anglais américain, qui nous permettent de faire des comparaisons afin de mieux identifier les usages canadiens⁴. TRANSBASE, base de données constituée par le Centre d'innovation en technologies de l'information, contient un corpus de traduction qui comporte des textes de départ et des traductions couvrant une période de huit ans du journal des débats du Parlement canadien, communément appelé le *Hansard*.

En utilisant ces quatre sources, nous avons établi deux listes de canadianismes (l'une pour l'anglais, l'autre pour le français) à inclure dans le DBC. Il faut noter qu'il s'agit de listes ouvertes : d'abord parce que, étant donné la durée du projet (une dizaine d'années), de nouveaux canadianismes risquent d'apparaître, surtout en français où la terminologie technique est encore en train de s'établir; puis, parce que nous arrivons parfois à repérer des sens canadiens «oubliés» lors de l'étude d'une concordance pour un mot-vedette particulier. Ce n'est, en effet, qu'en examinant TEXTUM pendant la préparation de l'article pour le nom *aîné* que nous avons identifié l'un de ses sens, celui d'une personne d'une tribu autochtone, qui est généralement âgée et qui a beaucoup d'influence sur la vie de la tribu. C'est un des sous-corpus canadiens, constitué d'articles de la *Presse canadienne française*, qui a révélé ce sens qui ne se trouve dans aucun dictionnaire, même canadien. Voici quelques exemples d'*aîné* dans ce sens, tirés du sous-corpus :

³ Nous tenons à exprimer notre reconnaissance envers l'équipe du TLFQ pour toute l'aide qu'elle nous a fournie.

⁴ Bien que TEXTUM ne contienne pas de textes britanniques, nous nous servons des *MicroConcord Corpus Collections* sur CD-ROM, qui contiennent d'une part des textes journalistiques, d'autre part des textes «académiques» dans une variété de domaines.

Presse canadienne française

>> aine

1 : 2726 matches

>> "aine "

2 : 1098 matches

>> 2 near "cri "

3 : 3 matches

>> pr

179365963,... demandé à un aîné cri, à l'occasion du procès qui a précédé le..

176165820,... demandé à un aîné cri, à l'occasion du procès qui a précédé le..

164006111,... sagesse de l'aîné Cri que vous citez, en disant modestement com..

>> pr.200 shift.-100

164006011,...n souci d'équité et de justice. En terminant, permettez-nous de vous assurer que nous respectons la sagesse de l'aîné Cri que vous citez, en disant modestement comme lui que nous ne possédons pas la..

179365863,...u Grand conseil des Cris du Québec et vice-président de l'Administration régionale crie. Quand on a demandé à un aîné cri, à l'occasion du procès qui a précédé le jugement Malouf dans l'affaire de la..

>> 2 near inuit

10 : one match

>> pr

293531366,...lo Pudlat, un aîné respecté parmi les artistes inuit du Canada j..

>> pr.200 shift.-100

293531266,...anisée dans l'Arctique canadien. L'exposition Pudlo. Une célébration est consacrée à l'oeuvre de Pudlo Pudlat, un aîné respecté parmi les artistes inuit du Canada jusqu'à sa mort récente. L'ouverture..

Pour nous assurer qu'il s'agissait bien d'un canadianisme, nous avons fait la même recherche dans le sous-corpus hexagonal, *Le Monde*. Comme le montre l'historique de nos recherches, présenté ci-dessous, ce sens ne semble pas exister en France.

Le Monde

>> aine

1 : 315 matches

>> "aine "

2 : 146 matches

>> 2 near cri

3 : one match (not relevant)

>> pr

46430207,...Paris, Dorbon aîné. DOC:AVEC UN DESSIN TIRE D'UNE BANDE DESSIN..

3. SÉLECTION DES CANADIANISMES POUR LA NOMENCLATURE DU DBC

En utilisant les procédures décrites ci-dessus, les chercheurs du DBC ont identifié un grand nombre de canadianismes en français et en anglais à inclure dans le dictionnaire. Mais nos listes ne sont pas exhaustives car, pour avoir droit de cité dans notre dictionnaire, les canadianismes doivent répondre à certains critères, qui ne sont pas exactement les mêmes pour l'anglais et le français.

Pour l'anglais, nous avons retenu presque tous les canadianismes qui se trouvaient dans deux sources lexicographiques (autres que les dictionnaires historiques) ou dans les corpus, car il ne s'agit pas de plusieurs milliers de mots ou d'emplois : nous n'avons éliminé que ceux dont le statut de canadianisme n'est pas confirmé, ceux qui sont utilisés seulement dans une région particulière (ex. le mot terre-neuvien **airsome**, adjectif, dans le sens de froid ou frais) ainsi qu'un certain nombre de canadianismes qui seraient aujourd'hui considérés comme étant archaïques (ex. **dime party**, dans le sens d'une soirée à laquelle on pouvait assister en offrant un don de dix sous) ou qui reflètent une réalité qui ne serait plus comprise aujourd'hui (ex. **fur board**, dans le sens de «thin board inserted in a cased skin to hold its shape until dry» - CDC). En appliquant ces quelques critères, nous nous sommes retrouvés avec moins de trois cents canadianismes lexicaux.

Pour le français, les critères de sélection ont été à la fois plus sévères et plus souples. Les canadianismes qui ont été exclus systématiquement, comme pour l'anglais, sont ceux qui sont régionaux, archaïques, et «historiques». Mais on a remplacé le critère de présence des canadianismes dans deux sources lexicographiques (non historiques) ou dans les corpus par celui de la fréquence relative de leur usage, selon le jugement de nos spécialistes du français canadien. Ce changement s'explique d'abord par le fait que les dictionnaires vraiment canadiens sont peu nombreux en français et que ni le *PLUS* ni le *RQ* ne sont considérés un reflet fidèle du français canadien, puis par le fait que beaucoup de canadianismes français, étant de niveau familier, ne se retrouvent pas dans les textes écrits qui constituent la plus grande partie de nos corpus. L'application de ces critères a mené à la sélection de quelques 3 000 canadianismes.

4. PROPORTION DE CANADIANISMES EN ANGLAIS ET EN FRANÇAIS

De prime abord, le nombre de canadianismes sélectionnés pour le DBC en anglais et en français peut sembler déséquilibré. En effet, le DBC comptera dix fois plus de canadianismes français que de canadianismes anglais. Mais un point important qui ressort de notre identification des canadianismes est qu'il existe beaucoup plus de canadianismes en français qu'en anglais.

Pour donner une idée plus claire de la proportion des canadianismes dans chaque langue (par rapport aux mots et aux emplois partagés avec d'autres variétés géographiques de chaque langue), examinons la nomenclature du *GAGE* et du *RQ*. Ces deux dictionnaires canadiens unilingues, dont nous nous sommes servis, entre autres, pour établir la nomenclature du DBC, incluent un plus grand nombre de canadianismes que les autres dictionnaires généraux canadiens. Même s'ils ne sont pas de taille comparable, ils se prêtent bien à une comparaison de la proportion des canadianismes dans chaque langue.

Ainsi, sur environ 70 000 entrées, le *GAGE* compte environ 800 canadianismes, si l'on inclut dans le total non seulement les mots ou les emplois qui contiennent la marque géographique *Cdn*, mais aussi ceux dont la particularité canadienne ne se manifeste que dans la définition; c'est donc dire que les canadianismes constituent moins de 1,5 % du total. Par contre, sur environ 40 000 entrées, le *RQ* compte à peu près 4 000 canadianismes; en d'autres termes, les canadianismes constituent 10 % du total. Ces statistiques très rudimentaires ont besoin d'être nuancées, car, d'une part, nous avons fait un calcul approximatif du nombre d'entrées et, d'autre part, les canadianismes ne sont pas tous présentés sous forme d'entrées séparées, parce qu'ils comprennent les emplois ainsi que les mots. Par ailleurs, bien que les deux dictionnaires se vantent de présenter des canadianismes, leurs préfaces sont muettes au sujet des critères utilisés pour les sélectionner. Mais, malgré ces réserves, on peut conclure que la proportion des canadianismes est beaucoup plus élevée en français qu'en anglais.

Le grand nombre de canadianismes français peut s'expliquer par les deux siècles d'isolement qu'a connu le Canada français. Cette période d'isolement a forcé le français canadien à se développer indépendamment du français hexagonal. Pendant ces deux cents ans, les francophones du Canada ont dû adapter leur langue aux réalités qui les entouraient. C'est dans le but de désigner certaines de ces réalités qu'ils ont donné des sens nouveaux à des mots qui étaient connus en France et qu'ils ont emprunté d'autres mots aux langues avec lesquelles ils étaient en contact. Cet isolement explique aussi pourquoi certains vocables ou sens qui sont sortis de l'usage ailleurs dans la francophonie sont encore employés régulièrement au Canada, s'ajoutant ainsi au stock de canadianismes.

D'autre part, le nombre apparemment moins élevé de particularités lexicales en anglais canadien pourrait être expliqué, nous croyons, par le fait que l'anglais canadien n'a jamais été isolé des autres sources d'anglais. Il a toujours joui de l'influence de l'anglais britannique et de l'anglais américain, même si l'importance de ces influences varie d'une époque à l'autre.

Ainsi, ce sont des facteurs historiques, politiques et même économiques qui semblent être responsables de l'important écart qui existe entre le nombre de canadianismes français et de canadianismes anglais.

5. TYPES D'UNITÉS LEXICALES QUE REPRÉSENTENT LES CANADIANISMES EN FRANÇAIS ET EN ANGLAIS

Il existe aussi des divergences entre les types d'unités lexicales que représentent les canadianismes en français et en anglais. En préparant la nomenclature du DBC, nous avons eu l'impression que la plupart des canadianismes en anglais sont des mots simples ou des mots composés, tandis que les canadianismes en français comprennent aussi des unités plus complexes comme les collocations et les expressions figées⁵. Pour confirmer cette impression, nous avons examiné à fond les canadianismes classés sous les lettres *D*, *F*, et

⁵ Dans cette partie de l'étude, nous ne faisons pas de distinctions entre les canadianismes de forme et les canadianismes sémantiques. Ainsi, si un sens canadien s'ajoute à un mot simple français commun au Canada français et à la France, on considère ici le mot simple comme canadianisme. La même chose s'applique aux mots composés, aux collocations et aux expressions figées.

S de la nomenclatures des canadianismes élaborée pour le DBC, les classant dans l'une ou l'autre de ces catégories⁶.

Des 65 canadianismes anglais à l'étude, 29 (soit 44,6 %) sont des mots simples (ex. **decker**, **dude**, **sourdough**); 35 (soit 53,8 %) sont des mots composés (ex. **deputy minister**, **diamond willow**, **fire hall**, **snow goose** et **sugaring off**); nous n'avons trouvé aucune expression figée, bien que nous ayons identifié une collocation (soit 1,5 %) (**to make a décharge**).

Par ailleurs, des 896 canadianismes français étudiés, 436 (soit 48,6 %) sont des mots simples (ex. **dépisteuse/dépisteuse**, **désencrage**, **fichument**, **flasher**, **francophonisation**, **sacoché**, **salaison** et **siffleux**); 199 (soit 22,2 %) sont des mots composés (ex. **chat sauvage**, **Département de santé communautaire**, **directeur-gérant** (dans le domaine du sport), **disco-mobile**, **fonds de pension**, **sac d'école** et **sous-verre**); il y a 51 collocations (soit 5,6 %) (ex. **dépouiller l'arbre de Noël**, **description de poste**, **se faire frapper**, **filer mal** et **saison régulière**); enfin, nous y avons repéré 210 expressions figées (soit 23,44 %) (ex. **le diable est aux vaches**, **être à la fine pointe de**, **mouiller à siaux** et **sacrer son camp**).

Il apparaît donc clairement que la typologie des canadianismes anglais et des canadianismes français est très différente : deux catégories — mots simples et mots composés — accaparent les canadianismes anglais, avec les mots composés comptant pour plus de la moitié, tandis qu'en français, les quatre types d'unités sont représentés, avec les mots simples en dominance sur les autres.

Nous croyons que ces différences quant aux types d'unités lexicales entre l'anglais canadien et le français canadien s'expliquent également par l'évolution de ces deux langues au Canada. Le contact continu avec la Grande-Bretagne et les États-Unis qui a caractérisé le développement du Canada anglais fait que l'anglais canadien a senti surtout le besoin d'unités lexicales lui permettant de désigner des réalités concrètes caractéristiques du pays : par conséquent, il a eu recours à des mots simples et des mots composés. Le français canadien, d'autre part, isolé du français hexagonal par l'histoire, a acquis des particularités non seulement dans sa façon de désigner les réalités typiquement canadiennes, mais aussi dans la manière de s'exprimer à des niveaux plus abstraits, ce qui expliquerait la répartition des canadianismes français sur les quatre types d'unités.

6. TYPES DE CONCEPTS DÉSIGNÉS PAR LES CANADIANISMES ANGLAIS ET FRANÇAIS

Pour confirmer notre impression que les canadianismes anglais représentent surtout des réalités concrètes, tandis que les canadianismes français désignent aussi bien des réalités abstraites que concrètes, nous avons examiné les 50 premiers mots simples et composés

⁶ Nous tenons à préciser que, dans quelques cas, notre classification des unités lexicales étudiées pourrait porter à discussion. En effet, il arrive que la distinction entre deux types d'unités complexes ne soit pas évidente. Ainsi, nous avons parfois considéré qu'une unité lexicale était une collocation alors que d'autres auraient tout aussi bien pu juger qu'il s'agissait d'une expression figée par exemple. Mais nous ne croyons pas que ces quelques cas affectent sérieusement les résultats obtenus.

canadiens (ou avec un sens canadien) qui commence avec *D* dans le *GAGE* et le *RQ* (voir annexe)⁷.

Nous avons d'abord analysé ces mots selon leur degré d'abstraction. Nous les avons divisés en deux catégories : mots désignant des concepts concrets (objets, personnes, lieux, actions et qualificatifs concrets), et ceux désignant des concepts abstraits (idéas, émotions, états, activités intellectuelles, morales, ou émotives, qualificatifs abstraits, etc.). Selon cette catégorisation assez rudimentaire, 47 canadianismes anglais sur 50 désignent clairement des concepts concrets (ex. **deputy minister**, **dogsled**, **dew-worm**). Les trois qui restent semblent plus abstraits par rapport aux autres : **darb** («any thing or person thought to be especially large, good, etc.» - *GAGE*); **disallow** («of the Federal Government, nullify an act of a provincial legislature» - *GAGE*); et **disallowance** (1. «the power of the Federal Government to annul provincial legislation»; 2. «an act of the Federal Government exercising this power» - *GAGE*). Mais il ne s'agit pas d'un degré très élevé d'abstraction. D'autre part, sur les 50 canadianismes français, 9 sont carrément abstraits : **damné** («exprimant la désapprobation» - *RQ*); **débile** («Iron. extraordinaire, exceptionnel, fantastique» - *RQ*); **débouler** («(personnes) faire irruption» - *RQ*); **déculotter** («causer la ruine de qqn, lui enlever tout ce qu'il possède» - *RQ*); **se déculotter** («ne pas compter à la dépense, ne pas agir de façon mesquine» - *RQ*); **définitivement** («certainement, sans aucun doute» - *RQ*); **dégêner** («faire perdre sa gêne, sa timidité à qqn» - *RQ*); **se dégêner** (perdre sa gêne, sa timidité⁸); **se dégot(t)er** (se trouver, se dénicher⁹). Ainsi, les canadianismes français examinés comptent au moins trois fois plus de mots abstraits que les canadianismes anglais examinés.

Deuxièmement, nous avons réanalysé les mêmes vocables pour voir combien d'entre eux représentent des réalités typiquement canadiennes. Trente canadianismes anglais sur 50 désignent des réalités évidemment canadiennes : **dag**, **Dall(s) sheep**, **dan**, **Persian darnel**, **debt**, **debtor**, **Dene (n.)**, **Dene (adj.)**, **Dene Nation**, **deputy**, **deputy minister**, **deputy returning officer**, **detachment**, **development road**, **dew-worm**, **diamond willow**, **dickey (dickie, dicky)**, **Digby chicken**, **disallow**, **disallowance**, **Dogrib (n.)**, **Dogrib (adj.)**, **Dominion (adjl.)**, **Dominion Day**, **don**, **doré**, **dory**, **draegerman**, **drum dance**, **dugout**. À cette liste, nous aurions pu ajouter quelques autres vocables comme **dalles**, **dogsled**, **dog-sledge**, **dog-train**, et **Doukhobo(u)r**, qui représentent des réalités principalement canadiennes. Ainsi, la proportion des canadianismes anglais dénotant des réalités canadiennes est élevée. Cependant, parmi les 50 canadianismes français, il n'y en a que deux qui désignent une réalité purement canadienne — **D.E.C.** et **délégation** — même s'il y en a d'autres, comme les trois termes reliés au hockey, qui pourraient être considérés comme représentant des concepts principalement canadiens. Cette analyse montre donc clairement que les canadianismes en anglais servent principalement à présenter les réalités canadiennes, tandis que ceux en français désignent très souvent des concepts qui ne sont

⁷ Il est intéressant de noter qu'en nous limitant à 50 premiers mots qui commencent avec un *D*, nous épuisons les canadianismes anglais qui commence par cette lettre, mais il nous reste au moins trois fois plus de canadianismes en français.

⁸ C'est nous qui avons formulé cette définition, car le *RQ* n'en donne pas une pour le verbe pronominal.

⁹ C'est nous qui avons formulé cette définition, car le *RQ* n'en donne pas une pour le verbe pronominal.

pas propres au Canada. En d'autres termes, les canadianismes anglais donnent une idée de la vie et du paysage au Canada, tandis que les canadianismes français, tout en remplissant aussi ce rôle, présentent souvent des réalités que le Canada partage avec d'autres pays, mais d'un point de vue canadien.

CONCLUSION

Notre analyse des canadianismes anglais et français présente un portrait du Canada anglophone et francophone. Le portrait du Canada anglais qui en ressort est celui d'un pays avec certaines particularités géographiques, politiques, historiques qui ont besoin d'être nommées, mais d'un pays avec un nombre limité de ces particularités, car la culture canadienne n'est pas tellement différente de la culture américaine d'une part, et de la culture britannique de l'autre. Le fait qu'il y a peu de canadianismes anglais récents est peut-être une indication du fait qu'il y a relativement peu d'innovations modernes typiquement canadiennes. Les canadianismes français nomment aussi beaucoup de particularités canadiennes. Mais, parce que le Canada français a longtemps été isolé de la France, ces vocables présentent aussi d'autres réalités, abstraites aussi bien que concrètes. En outre, la présence des canadianismes modernes en français est une indication du fait que les Canadiens français veulent participer à la vie nord-américaine, mais dans leur propre langue. Ainsi, le français canadien donne l'impression d'un peuple plus dynamique que ne le fait l'anglais canadien.

ANNEXE

Canadianismes anglais dans le *GAGE*

- | | | |
|----|--------------------|--|
| 1. | dag, n. | - a heavy, flat, double-edged triangular blade, used by the Indians as a weapon and tool |
| 2. | dalles, n.pl. | - a narrow stretch of river between high rock walls, characterized by whirlpools, rapids, and treacherous currents |
| 3. | Dall('s) sheep, n. | - a white North American sheep... of the western mountains found from northern British Columbia to the Arctic Ocean. |
| 4. | dan, n. | - in the North : 1. a sealskin used as a container for oil; 2. a buoy made of inflated sealskin or sheepskin sewn airtight, used as a mark in deep-sea fishing |
| 5. | darb, n. | - any thing or person thought to be especially large, good, etc. (<i>slang</i>) |
| 6. | Persian darnel | - a common and troublesome Canadian weed introduced from Asia |
| 7. | dasher, n. | - the fence surrounding the ice of a hockey rink; the boards |
| 8. | deadhead, n. | - a log or fallen tree partly or entirely submerged in a lake, etc., usually with one end embedded in the bottom |
| 9. | debt, n. | - 1. a credit given at a trading post to hunters and trappers in the form of supplies to be paid for out of |

- the next season's catch; 2. the amount of credit given;
3. in the North, credit taken at any store
10. debtor, n. - a person owing or taking debt at a trading store
11. décharge, n. - a shallow stretch in a water course, where it is necessary to unload a canoe, etc. in order to make way by tracking or paddling
12. decker, n. (Baseball) a catcher's mitt
13. decoy, v. - deke
14. deer, n. - caribou
15. deke, v. - (Hockey) a fake shot or movement intended to draw a defending player out of position (*slang*)
16. Dene n.pl. - the Athapascan Indian peoples of the Northwest Territories
17. Dene, adj. - of, having to do with, or designating the Dene
18. Dene Nation - the official organization representing the Athapascan peoples of the Northwest Territories
19. deputy, n. - a representative to or in certain assemblies. In Quebec, the members of the National Assembly are often called deputies
20. deputy minister - in Canada, a senior civil servant who acts as assistant to a cabinet minister
21. deputy returning officer - in Canada, an official appointed by the returning officer of a constituency to look after the procedure of voting at a particular polling station
22. detachment, n. - the smallest unit in the organization of the Royal Canadian Mounted Police or other police force. Some rural detachments of the RCMP have only one or two officers.
23. development road - in the North, a road or one of a system of access roads intended to help the exploitation of natural resources
24. dew-worm, n. - any large earthworm that comes to the surface at night when there is dew on the grass. Dew-worms are often used as fishing bait.
25. diamond willow willow wood having a diamond-patterned grain, resulting from an abnormal growth of the stems that may occur in any species of willow. It is prized for making lamps, walking sticks, ornaments, etc.
26. dickey (dickie, dicky) n. - a pull-over garment for the upper body, having a hood made of duffel or skins; parka; atigi
27. Digby chicken - a small smoke-cured herring
28. dime, n. - a coin of Canada or the United States, equal to one tenth of a dollar; a ten-cent coin
29. disallow, v. of the Federal Government, nullify an act of a provincial legislature
30. disallowance, n. - 1. the power of the Federal Government to annual provincial legislation; 2. an act of the Federal Government exercising this power
31. dogan, n. - a Roman Catholic, especially one of Irish background

32. Dogrib, n. - 1. a member of an Amerindian people who live in the Northwest territories... 2. the Athapascan language of the Dogrib
33. Dogrib, adj. - of or having to do with the Dogrib or their language
34. dogsled, n. - a sled that is pulled by dogs
35. dog-sledge, n. - dogsled
36. dog-train, n. - a sled pulled by a team of dogs
37. dome fastener - a metal or plastic fastener consisting of two parts, one with a small, rounded projection in the centre that snaps into a sock in the centre of the other
38. Dominion, adjl. - 1. in Canada, under the control or authority of the federal government; 2. in Canada, relating to the country as a whole; national in scope
39. Dominion Day - a national holiday commemorating the establishment of the Dominion of Canada in 1867; Canada Day
40. don, n. - in some Canadian universities and colleges, an official in charge of a student residence
41. doré, n. - walleye
42. dory, n. - 1. John Dory, an edible sea fish; 2 walleye; doré
43. Doukhobo(u)r, n. - a member of a 200-year old Christian sect originally from Russia that traditionally believes that every person knows what is right and must be guided by this knowledge rather than by an outside authority. Several thousand Doukhobors left Russia in 1898 and settled in western Canada
44. draegerman, n. - especially in the Maritimes, a coal miner trained in underground rescue work and the use of special oxygen equipment effective in gas-filled mines
45. dram, n. - (*Historical*) (Logging) a section of a timber raft, made up of several cribs lashed together
46. driveway, n. - a road, especially one that is lined with trees and lawns
47. drum dance - 1. an Inuit dance consisting of expressive body movements and gestures, performed to the accompaniment of drums and with the dancer or dancers singing as they dance; 2. any of various Indian dances accompanied by drums
48. dude, n. - in the Western parts of Canada and the United States, a city-bred person, especially one who spends a holiday on a ranch
49. dugout, n. - especially on the Prairies, a large excavation used to hold water collected there from the spring thaw and from rainfall
50. duplex, n. - 1. a building consisting of two dwellings under one rook, either side by side or one above the other; 2. one of the dwellings of such a building

Canadianismes français dans le RQ

- | | | |
|-----|----------------------------------|---|
| 51. | dactylo, n.f. | - machine à écrire (<i>fam</i>) |
| 52. | dalle, n.f. | - gouttière (<i>fam</i>) |
| 53. | dalot, n.m. | - 1. bordure latérale creuse de l'allée de quilles; 2. partie profonde de la gouttière - tuyau de descente de la gouttière |
| 54. | damné(e), interj. | - exprimant la désapprobation (<i>fam</i>) |
| 55. | danse carrée, n.f. | - quadrille |
| 56. | dard, n.f. | - fléchette |
| 57. | dardage, n.m. | - (Hockey) action de darder; son résultat |
| 58. | darder, v.tr. | - (Hockey) donner un coup à un joueur avec la pointe de la lame du hockey (<i>fam</i>) |
| 59. | se darder, v.pron. réfl. | - (vers, sur) se précipiter, se jeter, s'élancer sur, vers |
| 60. | carré aux dattes, n.m. | - pâtisserie faite de dattes cuites disposées entre deux couches d'une préparation à base d'avoine ou de gruau |
| 61. | débarbouillette, n.f. | - petit carré de tissu éponge avec lequel on fait sa toilette, on se débarbouille |
| 62. | débarquer, v.intr. | - descendre de (<i>fam</i>) |
| 63. | débarrer, v.tr. | - ouvrir ce qui est fermé à l'aide d'un mécanisme (verrou, taquet, etc.), spécialt ouvrir avec une clé |
| 64. | débattre, v.intr. | - palpiter, battre très fort d'une manière déréglée |
| 65. | débenture, n.f. | - obligation non garantie |
| 66. | débile, adj. | - Iron. extraordinaire, exceptionnel, fantastique |
| 67. | déboîter, v.tr. | - détruire, réduire en morceaux (<i>fam</i>) |
| 68. | débossier, v.tr. | - débosser (<i>fam</i>) |
| 69. | débossage, n.m. | - débosselage (<i>fam</i>) |
| 70. | débouche, n.m. | - Abrév. de débouche-bouteilles |
| 71. | débouche-bouteilles, n.m. invar. | - ustensile servant à ouvrir les bouteilles capsulées |
| 72. | débouler, v.intr. | - (personnes) faire irruption |
| 73. | débrocher, v.tr. | - enlever les agrafes qui attachent des feuilles (<i>fam</i>) |
| 74. | débrocheuse, n.f. | - dégraffeuse (<i>fam</i>) |
| 75. | D.E.C., n.m. invar. | - Abrév. de <i>diplôme d'études collégiales</i> |
| 76. | décharge, n.f. | - cours d'eau dans lequel un lac déverse son trop-plein d'eau, spécialt l'endroit d'un lac d'où part ce cours d'eau |
| 77. | déclin, n.m. | - matériau de revêtement extérieur dont les éléments (planches, tôles, etc.) sont disposés horizontalement de manière à se chevaucher partiellement les uns les autres; ce revêtement |
| 78. | décrassage, n.m. | - [(Tech) opération qui consiste à débarrasser la grille d'un foyer des matières non combustibles] ¹⁰ le décrassage d'un poêle |
| 79. | décret, n.m. | - ensemble de règles définissant les normes de travail d'un groupe de salariés |
| 80. | décrochage (scolaire), n.m. | - abandon des études avant la fin de la période d'obligation scolaire |

¹⁰ Cette définition est tiré du PLUS, car le RQ ne donne qu'un exemple d'usage.

81. décrocher (de l'école) v.intr. -quitter l'école
82. décrocheur, -euse, n. - élève qui abandonne ses études
83. décrochir, v.tr. - redonner une forme droite à (*fam*)
84. déculotter, v.tr. - causer la ruine de qqn, lui enlever tout ce qu'il possède (*fam*)
85. se déculotter, v. prom. réfl. - ne pas compter à la dépense, ne pas agir de façon mesquine
86. déductible, n.m. - part d'un dommage qu'un assuré doit assumer (*fam*) (*Anglic.*)
87. défensif, -ive, adj. - (Hockey) Joueur défensif, avant qui se préoccupe davantage de la défense que de l'attaque; (Sports) Un match défensif, où la défense domine
88. défensive, n.f. - (Sports) la défense
89. définitivement, adv. - certainement, sans aucun doute (*fam*) (*Anglic.*)
90. défoliateur, -trice, adj. - qui cause la chute des feuilles d'un arbre
91. dégêner, v.tr. - faire perdre sa gêne, sa timidité à qqn
92. se dégêner, v. prom. réfl. - [perdre sa gêne, sa timidité]¹¹
93. se dégot(t)er, v. pron. - [se trouver, se dénicher]¹²
94. dégouttière, n.f. - 1. (surtout au plur.) liquide qui dégoutte, spécialt l'eau; 2. eau de pluie qui tombe d'un toit, d'une gouttière; glaçon, glace fondante qui dégoutte; 3. eau qui s'infiltré dans une construction et qui dégoutte
95. dégraffeuse, n.f. - petit instrument de bureau servant à enlever les agrafes
96. dégriffer, v.tr. - enlever les griffes à un animal, spécialt au chat
97. déjeuner, n.m. - repas du matin
98. déjouer, v.tr. - (Sports) manoeuvrer la rondelle, le ballon... de manière à empêcher un adversaire de s'emparer
99. délégation, n.f. - représentation permanente d'une province auprès d'un état étranger; fonction de délégué - ensemble du personnel assurant cette mission; résidence du délégué, locaux et services.
100. deltaplanneur, -euse, n. - adepte du deltaplane

¹¹ C'est nous qui avons formulé cette définition, car le RQ n'en donne pas une pour le verbe pronominal.

¹² C'est nous qui avons formulé cette définition, car le RQ n'en donne pas une pour le verbe pronominal.

RÉFÉRENCES

1. Ouvrages théoriques

- DUBOIS, J. et C. DUBOIS (1971) : *Introduction à la lexicographie : le dictionnaire*, coll. «Langue et langage», Paris, Larousse, 217 p.
- GREGG, R. J. (1989) : «La lexicographie de l'anglais canadien», *Revue québécoise de linguistique*, 18-1, pp. 151-187.
- MALLORY, J. R. (1984) : *The Structure of Canadian Government*, édition revue, Toronto, Gage Educational Publishing Company, 472 p.
- POIRIER, C. (1980) : «Le lexique québécois : son évolution et ses composantes», *Stanford French Review*, IV-1/2, pp. 43-89.
- REY, A. et S. DELASALLE (1979) : «Problèmes et conflits lexicographiques», *Langue française*, 43, pp. 4-26.
- RIALLAND-ADDACH, V. (1995) : «Culture partagée et dictionnaire monolingue de français langue étrangère», *Études de linguistique appliquée*, 97, pp. 91-103.
- SVENSÉN, B. (1993) : *Practical Lexicography : Principles and Methods of Dictionary-Making*, Oxford/New York, Oxford University Press, 285 p.

2. Ouvrages lexicographiques

- Collins English Dictionary* (1986) : Glasgow, Collins.
- Concise Dictionary of Canadianisms* (1973) : Toronto, Gage.
- Dictionnaire du français canadien* (1990) : Toronto, Stoddart.
- Dictionnaire du français Plus* (1988) : Montréal, CEC.
- Dictionnaire québécois d'aujourd'hui* (1992) : Saint-Laurent, Québec, Dicorobert.
- Funk and Wagnalls Canadian College Dictionary* (1989) : Toronto, Fitzhenry and Whiteside.
- Gage Canadian Dictionary* (1983) : Toronto, Gage.
- Lexis. Dictionnaire de la langue française* (1987) : Paris, Larousse.
- Penguin Canadian Dictionary* (1990) : Markham / Mississauga, Penguin / Copp Clark Pitman.
- Petit Larousse illustré* (1991) : Paris, Larousse.
- Petit Robert* (1991) : Paris, Robert.
- Random House Webster's College Dictionary* (1991) : New York, Random House.
- Richesses et particularités de la langue écrite au Canada* (1980) : Montréal, Université de Montréal.

PROBLÈMES DE L'ÉTABLISSEMENT DE LA TERMINOLOGIE JURIDIQUE DANS UNE LANGUE SANS OFFICIALITÉ HISTORIQUE. LE CAS DU BASQUE

Antton ELOSEGI ALDASORO

Université du Pays Basque, Donostia, Espagne

LA RELATION DE LA LANGUE BASQUE AVEC LE DROIT

La relation de la langue basque avec le droit a des aspects que l'on serait tenté de qualifier de paradoxaux —si cela n'étaient pas le cas de tant de langues, petites et grandes. Les pays parlant basque ont développé, tout au long de leur histoire, un bon nombre d'institutions juridiques propres, tant de droit public (les *Biltzar* du Labourd, les *Juntas Generales* de la Biscaye ou le Royaume de Navarre), que de droit privé. On trouve même, de nos jours, des institutions spécifiquement basques de droit privé actives dans le système juridique espagnol, et dont le nom basque n'a jamais été traduit (comme le testament dit de '*hilburuko*'). Pourtant, le droit basque n'a jamais été écrit en basque par des Basques mais plutôt en latin, romance navarrais, occitan provençal, occitan gascon, espagnol ou français.

À côté de ce paradoxe historique, nous nous trouvons actuellement face à un autre paradoxe : celui d'être en train de créer pour une vieille langue un monde conceptuel et terminologique, celui du droit, dont la caractéristique la plus manifeste est censée être sa traditionnalité, le fait d'être intimement liée à la mémoire historique de la langue qui a véhiculé ce droit. Mais nous savons bien que notre cas n'a rien d'extraordinaire comme le prouvent tant de pays qui ont été gouvernés dans des langues autres que la leur propre¹.

1. QUELQUES DONNÉES SOCIOLINGUISTIQUES ACTUELLES

La langue basque est parlée dans sept provinces historiques, 3 françaises et quatre espagnoles, par une partie variable de la population. Quelque 700 000 personnes en tout sur une population totale de quelque 2 860 000 habitants, dont 2 640 000 en Espagne et 220 000 en France. Il reste un nombre résiduel de bascophones monolingues (30 000), les autres ayant une bonne ou une très bonne compétence en espagnol ou en français. C'est-à-dire que les bascophones actuels sont une minorité bilingue de 23 % sur le territoire historique basque.

¹ Voir le livre de J.L. Mark Cooray (1985) qui décrit les problèmes auxquels le Sri Lanka indépendant a dû faire face pour arriver à faire fonctionner son droit dans les langues autochtones récemment officialisées.

Le Pays basque espagnol étant une région fortement industrialisée, la population bascophone est, dans sa grande majorité, urbaine et industrielle, mais les noyaux les plus compacts à dominance bascophone sont constitués de petites villes et de villages mixtes ruraux-industriels, hors des grandes villes comme Bilbao, Pampelune, Saint Sébastien. Du côté français, le monde rural est plus fortement représenté.

Un aperçu historique nous offre une image plus nuancée de la réalité; la situation ayant fortement changé depuis que dans la partie espagnole la langue basque a été reconnue co-officielle. D'après les données officielles, corroborées par une enquête sociolinguistique en 1996², on compte quelque 100 000 nouveaux bascophones depuis les 15 dernières années, dans une population à démographie stabilisée. Ce renouveau quantitatif de la langue basque est doublé d'une forte extension des usages sociaux dont cette langue avait été exclue historiquement. C'est justement un champ d'usage nouvellement ouvert, celui du droit, qui fait l'objet de cette communication.

Ces données montrent l'inversion du processus historique de perte continue tant du nombre des locuteurs que des relations sociales où la langue était employée. Processus accéléré avec l'extension de l'alphabétisation et l'importance croissante de la langue écrite qui condamnent sans merci les langues à usage uniquement oral.

Les facteurs de récupération linguistique sont bien établis. L'école vient en tout premier lieu : aujourd'hui tous les enfants d'une bonne partie du pays reçoivent leur éducation totalement ou partiellement en langue basque, ou bien apprennent le basque. Les progrès dus à l'enseignement primaire et secondaire s'ajoutent aux effets d'une chaîne de télévision émettant en basque toute la journée ainsi qu'aux moyens de communication en général, avec un quotidien et plusieurs hebdomadaires monolingues basques. Le troisième pilier de la récupération est l'emploi de la langue dans l'administration et à l'université; nous en parlerons plus tard.

Il reste tout de même de grands «trous noirs» où le basque n'a guère de présence, notamment dans la vie économique, les entreprises et les services privés et aussi, ce qui nous concerne plus directement, dans l'administration de la justice.

2. LA LANGUE BASQUE. DEUX MOTS SUR SON HISTOIRE ET SUR SON LEXIQUE

Du point de vue typologique le basque ou *euskara* est la seule langue non indoeuropéenne de l'ouest de l'Europe. Elle n'a pas de liens génétiques connus³ même si l'on a essayé de lui trouver des cousins chez nos voisins les Berbères ou chez les Kartveliens de Georgie ou bien, plus loin, chez les Aïnos du Japon.

Du point de vue historique, bien que de très sérieux indices lui donnent une ancienneté de quelques milliers d'années sur son territoire, son existence dans le Sud-Ouest

2 II^e Enquête sociolinguistique, 1996, Département de la culture du Gouvernement basque, Direction générale de politique linguistique du Gouvernement de la Navarre, et l'Institut culturel basque pour la partie française du territoire basque.

3 Trop de recherches et trop de pseudo-science ont été consacrées au «dévoilement des mystères» de l'origine du basque (voir J.L.Trask (1997) dans son dernier livre).

de la France et au Sud des Pyrénées depuis quelque dix-huit siècles est indéniable, puisque constatée épigraphiquement (Michelena 1984; Trask 1997). Ce territoire, jadis très vaste, est allé en se rétrécissant depuis la Rome Impériale, (Caro Baroja 1945) avec quelques rares épisodes d'expansion, sous la pression du latin et de ses langues filles. Mais il est toujours resté un noyau dur, qui comprend grosso modo le Pays basque français actuel (une moitié du département des Pyrénées atlantiques), plus, en Espagne, une bonne partie de la Navarre et de la région autonome qui porte le nom de Pais Vasco ou Euskadi.

Au cours de l'histoire, le Pays basque n'a jamais été totalement monolingue, puisque la langue d'écriture a toujours été autre — ce qui suppose une élite connaissant le latin et/ou d'autres langues romanes. De plus, entourée de langues plus «prestigieuses et cultivées», le basque a adopté un bon nombre d'éléments des langues voisines, modifiant progressivement, comme c'est le cas pour toute langue vivante, sa structure phonologique, sa syntaxe et surtout son lexique.

Le basque actuel garde dans son lexique de très importants champs sémantiques où la présence du basque primitif est manifeste. Mais, comme il fallait s'y attendre, les emprunts lexicaux au latin classique ou tardif, à l'occitan gascon et au castillan sont très nombreux et ce dans tous les champs «ajoutés» de la vie, depuis très longtemps. Ces derniers siècles, l'influence du français a été aussi considérable et, depuis quelques années, l'anglais commence aussi à laisser sa trace sur le lexique basque, à travers le filtre du français ou de l'espagnol.

Ce changement-enrichissement lexical prend une telle ampleur que la majorité des mots employés aujourd'hui dans le parler populaire, même rural, est étymologiquement non basque; leur adaptation phonologique et l'ancienneté de leur usage fait que les locuteurs n'ont souvent pas conscience de l'origine romane de ces mots de tous les jours⁴.

3. DU MANQUE D'OFFICIALITÉ À LA RECONNAISSANCE OFFICIELLE

3.1 L'histoire politique

La langue basque, paraît-il, n'a jamais été un véhicule écrit de la loi ni des relations juridiques privées, bien que différents territoires linguistiquement basques ont été politiquement souverains. La plus connue de ces entités politiques est le Royaume de Navarre, indépendant jusqu'au XVI^e siècle, à composant ethno-linguistique majoritaire basque, mais qui n'a jamais employé cette langue «officiellement».

L'histoire linguistique officielle des autres territoires, jouissant eux aussi d'un haut degré d'autonomie politique est semblable : leurs élites ont toujours préféré l'une ou l'autre langue néolatine.

4 Le lexique du basque actuel est un lexique «en voie de développement» assez avancée. La quantité et la variété des publications de différentes disciplines parues chaque année (1 305 en 1995, dont 467 classées 3-, Sciences sociales), et les douzaines de dictionnaires terminologiques publiés ainsi que des recherches comme la Collecte systématique du basque actuel (EEBS), corpus de 3 000 000 d'entrées lexicographiques en font la preuve.

L'histoire moderne du Pays basque est celle de la perte progressive de l'entité politique et de la fusion graduelle dans les grands États-nations, la France et l'Espagne, qui l'englobent. La perte de la personnalité politique et, donc, de l'aspiration à l'officialité apparaît comme définitive du côté français avec la Révolution⁵. Du côté espagnol, des guerres et des épisodes de violence, aux XIX^e et XX^e siècles, ont sûrement quelque chose à voir avec la résistance des Basques à la dilution de leur culture dans des ensembles uniformisateurs. Mise à part une courte expérience de 9 mois en 1936-1937, pendant la guerre suivant le soulèvement fasciste de Franco, où la langue basque fut déclarée officielle, jamais jusqu'à présent, c'est-à-dire jusqu'à la Constitution espagnole de 1978, la langue basque n'a eu de reconnaissance officielle.

On trouve, tout de même, des documents basques à valeur juridique depuis le XVI^e siècle : des témoignages transcrits littéralement, des traductions d'actes notariés, des registres communaux, des lettres entre des communes basques des deux côtés de la frontière, de très intéressants documents juridico-politiques des premières années révolutionnaires en France et, à partir du XIX^e siècle, au Pays basque d'Espagne, quelques ordonnances et des avis municipaux, ainsi que quelques contrats privés et d'autres documents commencent à être traduits en basque.

3.2 La situation actuelle dans l'administration et dans la vie juridique

En Pays basque de France le français reste seule langue officielle et valable; du côté espagnol, la Navarre, offre à l'*euskara* une demi-officialité dans une partie de son territoire. Les données qui suivent ne concernent que l'entité politique et administrative qui porte le nom de *Comunidad Autónoma del País Vasco* ou *Euskadi*.

Dans cette partie, la plus peuplée du pays (74 % de la population basque et 79 % des bascophones y habitent) et où les partis nationalistes sont au pouvoir régional depuis 1980, il s'est développé une politique d'application de l'officialité.

Ainsi, toutes les normes juridiques émanant des pouvoirs régionaux et locaux sont publiées en *euskara* en même temps qu'en espagnol. Chacun a le droit de s'adresser aux autorités dans la langue de son choix, et celles-ci sont (théoriquement) tenues de répondre dans la même langue. C'est surtout aux administrations locales des territoires à grande proportion de bascophones que la politique de basquisation de l'administration a été mise en vigueur⁶, faisant de l'*euskara* leur langue de travail. Mais pour la plupart des organes administratifs, le basque est une langue vers laquelle sont traduits des documents finaux dont l'instruction a été faite en espagnol.

5 La Révolution et la Grande Guerre ont marqué très négativement l'avenir de la langue basque en France, mais dans des différents épisodes de la Révolution l'*euskara* a eu plus de «reconnaissance» que sous n'importe quel autre régime. Voir Rica Esnaola (1975) et *1789 et les Basques* (1991).

6 Aujourd'hui, 28 communes entre 200 et 20 000 habitants se sont syndiquées dans le UEMA ou Mancomunidad des communes bascophones avec le but d'employer uniquement l'*euskara* comme langue d'administration. Seules les communications dirigées vers des non bascophones sont traduites en espagnol.

La politique de *capacitacion* linguistique (langue, traduction, rédaction) des fonctionnaires publics menée par l'IVAP (Institut basque de l'administration publique) et l'exigence de la connaissance de cette langue pour certains postes ont permis au bilinguisme de l'administration de faire des progrès.

La situation dans la vie juridique privée n'est pas si avancée. C'est surtout des entreprises de services publics, des caisses d'épargne et des compagnies d'assurances qui offrent leurs formulaires en format bilingue. Bien qu'officiellement valables, peu de contrats ou de testaments se font en *euskara*.

À l'université, les études de droit en basque se sont peu à peu développées depuis une dizaine d'années; aujourd'hui les deux facultés de Droit de la Communauté Autonome offrent la possibilité de faire une Licence en Droit en langue basque. Plus de 15 % des étudiants de la plus importante d'entre elles, l'Université du Pays basque, choisissent la «ligne basque» où la loi — presque toujours écrite en espagnol — est enseignée en basque. Les études juridiques de troisième cycle viennent juste de commencer : une thèse doctorale juridique a déjà été écrite en basque.

4. PROBLÈMES DE L'ÉTABLISSEMENT DU LANGAGE JURIDIQUE EN BASQUE

L'arrivée de l'officialité a supposé le passage d'une situation de vide presque total de langage juridique à une nouvelle situation où l'on produit chaque jour de nombreux textes juridiques, c'est-à-dire, des textes en langage juridique. Ce changement a demandé la résolution de différentes sortes de problèmes, concernant soit le langage du droit lui-même, soit la situation de la langue basque quant à son statut et quant à son corpus.

4.1 Problèmes concernant le langage du droit

De nombreux auteurs ont décrit (Cornu 1990) les traits caractéristiques du langage du droit et de sa terminologie. Je ne retiendrai que ceux qui supposent une difficulté ajoutée à l'établissement de nouveaux langages de droit.

- Le droit, comme système de régulation de la société, est tout d'abord un fait social et non pas une science, même s'il existe une science du droit. La terminologie du droit, préexistante à la science, n'a pas été créée ni structurée par des scientifiques.
- La terminologie du droit est dans sa grande majorité le résultat d'une spécialisation du lexique commun. La plupart des termes simples du droit sont des mots connus des simples usagers de la langue, mais le droit leur donne un sens spécialisé, quelquefois différent ou opposé à celui du langage commun.
- Le droit existe partout mais ses concepts ne sont pas universels. Beaucoup de termes juridiques designent des notions spécifiques des droits nationaux et sont intraduisibles *stricto sensu*, comme l'anglais *equity* ou les mots basques *hilburuko* et *elkarporoso*.

- Le droit est un fait social à grande continuité historique. Les changements politiques et même les révolutions sociales n'effacent jamais la continuité du vocabulaire traditionnel du droit.⁷
- Le droit est une réalité historique étroitement liée à la langue «nationale» de chaque pays. Il existe aussi de grands espaces de «culture juridique» supra-nationaux, mais la réalité «état-nationale» l'emporte toujours.
- Le droit, la loi écrite, a façonné, d'une certaine façon, les grandes langues porteuses du droit, les textes légaux ayant souvent été le modèle de prestige pour la langue écrite comme les *Leges XII Tabularum* ou les *Serments de Strasbourg*.
- La pratique juridique est souvent un exercice linguistique de commentaire de texte et d'interprétation, très difficile à faire dans une langue autre que celle du texte de la loi.

4.2 Problèmes de statut

- Une situation de minorisation linguistique : la population bascophone est minoritaire, bilingue et habituée à employer le basque uniquement comme langue B, réservant à l'espagnol ou au français les usages à prestige social.
- De différents statuts légaux : le langage du droit n'est pas nécessaire du côté français; du côté espagnol, il a deux statuts différents : celui de la Navarre et celui de la Communauté autonome du Pays basque. La coordination d'ensemble reste, donc, impossible.
- Une officialité de deuxième classe. Quelques autorités en Pays basque, et en tout premier lieu les autorités judiciaires, ne reconnaissent pas dans la pratique le droit à la communication bilatérale en basque, notamment aux procès judiciaires. Une voie essentielle pour la normalisation du basque juridique reste ainsi fortement bloquée.
- Un usage symbolique de la langue : le basque est souvent employé non pas pour des besoins de communication mais comme symbole de la différence culturelle. Une partie importante des textes juridiques publiés en basque a rempli cette fonction; peu importe si l'on manquait de terminologie adéquate, ou si les textes traduits allaient être incompréhensibles pour les bascophones.
- Un mauvais usage des recours : cet usage symbolique, quoique nécessaire, et le manque de planification adéquate, ont dévié des ressources humaines et financières qui auraient pu être mieux employées du point de vue de la normalisation linguistique.

4.3 Problèmes de corpus

- Une langue récemment unifiée : en 1978 quand la production de textes juridiques commence, la langue basque est loin de la standardisation. La *koiné* basque ou *euskara batua* crée par *Euskaltzaindia*, l'Académie de la langue basque, pour l'unification du basque écrit sur les dialectes, comptait à peine 10 ans.

⁷ Voir par exemple l'étude de Dubois (1962).

- Manque de dictionnaire du basque unifié, ce qui laisse aux rédacteurs juridiques — des traducteurs — le choix de beaucoup de mots du langage courant et, bien sûr, des termes juridiques. Par exemple, les traducteurs ont eu à utiliser intensivement un dictionnaire datant de 1905⁸, à grand intérêt philologique, mais qui avait écarté les mots jugés étrangers, bien qu'étant d'usage généralisé. Même si la situation a totalement changé ces dernières années⁹, les origines du basque juridique ont été marquées par le manque de base lexicographique suffisamment ample et acceptée.
- Manque de tradition juridique : les traducteurs ont eu à travailler sans modèle de référence propre, sur un langage fortement marqué par la tradition, tant pour le vocabulaire que pour la phraséologie, la syntaxe ou les différents registres linguistiques.

4.4 Problèmes concernant le corpus et le statut

- L'unité de la langue face au double système juridique : le basque unifié étant conçu pour servir de véhicule unitaire aux Basques de France ou d'Espagne, le langage créé pour le droit devait répondre en même temps aux besoins des deux systèmes juridiques. Problème sans solution dans beaucoup de cas, car les institutions juridiques ne sont pas identiques des deux côtés de la frontière. Par exemple, le basque (d'origine romane) *prokuradore* signifie *procureur* du côté français et *avoué* du côté espagnol.
- Un langage produit par des traducteurs : tandis que d'autres variétés ou technolèctes basques se sont développés, à partir surtout de l'école et des manuels scolaires, le langage du droit a été l'oeuvre des traducteurs du Parlement et des administrations publiques.
- Le basque est la langue du bout de la chaîne : le processus de production d'un texte se déroule entièrement en espagnol jusqu'à la phase finale où le traducteur intervient pour que le texte soit publié ou arrive chez son destinataire dans les deux langues officielles.

5. PRODUCTION DE TEXTES JURIDIQUES ET DE DICTIONNAIRES

5.1 Production de textes juridiques

Comme nous l'avons écrit, le langage juridique basque fut d'abord un langage créé par des traducteurs de l'administration publique. Depuis, la production de textes s'est diversifiée : la rédaction directement en langue basque se développe et ces textes appartiennent à des secteurs de plus en plus variés.

8 Le dictionnaire trilingue *Diccionario Vasco-Español-Francés* de R.M. Azkue est resté la référence principale jusqu'aux années 1980.

9 La lexicographie basque a fait des grands pas en avant. En plus de l'EEBS et des dictionnaires terminologiques cités, on dispose maintenant du volumineux dictionnaire général basque (OEH) (Michelena 1987) avec un corpus de 5 800 000 entrées lexicales qui reflète le basque des tous les textes traditionnels, partiellement publié, et l'on trouve sur le marché des dictionnaires de la langue actuelle comme le *Euskal Hiztegia* (Sarasola 1996) ainsi que des dictionnaires encyclopédiques en plusieurs volumes, en plus de dictionnaires bilingues (espagnol, français, anglais). *Euskaltzaindia*, l'Académie de la langue basque, prépare son dictionnaire du basque unifié qui supposera un pas très ferme dans la normalisation de l'*euskara*.

Ainsi, dès la fin des années 1980, l'*euskara* commence à être une langue d'enseignement du droit : des professeurs et des étudiants sont les nouveaux acteurs de la production de textes, ici aussi traduits dans la plupart des cas. Un certain nombre de manuels universitaires de droit les plus connus ont été traduits de l'espagnol, déjà quelques-uns ont été directement écrits en basque, et la production d'articles scientifiques sur des sujets juridiques commence à être normale dans des revues bilingues¹⁰. Il existe même une revue juridique rédigée exclusivement en langue basque¹¹.

5.2 Terminographie et normalisation du langage juridique

Au début, c'est les traducteurs eux-mêmes qui ont eu à jouer le rôle de terminologue et de terminographe, en assumant la responsabilité des choix terminologiques et en faisant des recopiations de termes. Vu la diversité des choix, des coordinations se sont créées assez tôt, mais bientôt on a pu constater que faute d'une autorité unique, les modèles linguistiques devenaient de plus en plus divergents.

Deux types d'institutions ont eu dès le début une incidence importante sur l'établissement du basque juridique : d'une part l'IVAP, déjà mentionné, et d'autre part, l'Institut basque de services universitaires (UZEI), institut privé à but non lucratif dédié à la confection de dictionnaires terminologiques; bien plus tard une troisième institution, l'université, s'y est ajoutée.

C'est UZEI qui en 1985 a publié le seul dictionnaire général de droit existant aujourd'hui. Ce dictionnaire de droit est l'oeuvre la plus marquante de la terminologie juridique basque, bien que contesté par un courant¹². Ce dictionnaire et la presque soixantaine d'autres dictionnaires spécialisés édités par UZEI représentent le fondement de la base de données terminologiques *Euskalterm* accessible «en ligne» et mise à la disposition de toutes les administrations publiques.

L'Institut basque d'administration publique à travers ses sections de traduction, terminologie et standardisation du langage administratif a surveillé la traduction des textes légaux et a surtout produit des manuels pour la rédaction et la traduction de textes administratifs ainsi que des recueils terminologiques partiels.

Quant à l'université, même si des universitaires ont toujours été consultés pour l'élaboration des terminologies, elle est arrivée depuis peu de temps à la production de recueils terminologiques en collaboration avec l'UZEI ou l'IVAP.

10 En tout cas, le droit est une discipline en retard sur beaucoup d'autres à l'université. La comparaison du nombre des thèses doctorales juridiques en basque (une seule) avec les plus de 70 thèses de tout genre lues en basque à la seule Université du Pays basque semble le démontrer.

11 La revue juridique *ELERIA* éditée par la Société des Études basques *Eusko Ikaskuntza* ne publie que des articles originellement écrits en basque sur des sujets de droit et de linguistique juridique.

12 Ce dictionnaire, *Zuzenbide Hiztegia* (UZEI : 1985), en deux volumes, dont l'un est monolingue basque et encyclopédique et l'autre un vocabulaire trilingue basque-espagnol-français, avec plus de 10 000 entrées à partir du basque, a été considéré représentatif du courant B que nous décrivons plus tard. Les critiques adressées à son égard lui reprochent surtout d'avoir voulu établir une terminologie avant même qu'il existe un dictionnaire général de langue commune moderne.

6. DIVERGENCES ET ESSAIS DE SOLUTIONS

Faute d'une tradition de référence et d'autorité terminologique capable d'imposer un modèle, les traducteurs qui ont créé le langage juridique ont essayé de fournir une réponse aux problèmes décrits, mais leurs directions ont été assez divergentes. Puisqu'il n'y a pas de véritable vie juridique en basque qui mette à l'épreuve la validité des textes produits et les choix terminologiques, les divergences loin de se diluer sont devenus des courants. Les différends existants à propos du modèle de modernisation de la langue en général, assez idéologisés d'ailleurs, se sont manifestés très explicitement à propos du langage juridique et administratif.

Ces diverses tendances ont produit non seulement des modèles textuels différents, mais aussi des normes et des terminologies assez divergentes. Il y a des cas où différentes traductions d'un même texte légal n'ont que 20 % de coïncidences lexicales, et peut-être encore moins pour les autres aspects textuels.

Même au risque d'en faire une caricature, nous allons essayer de signaler très schématiquement les traits distinctifs des deux principaux courants, que nous appellerons courant A et courant B (Berasategi 1989 : 29-45).

Courant A :

- Il veut être plus près de la langue orale traditionnelle, même aux dépens de la précision technique.
- Il donne plus d'autonomie au texte basque par rapport à l'espagnol : ses traductions s'éloignent davantage de l'équivalence formelle.
- Il est plus puriste tant du point de vue morphosyntaxique que lexical : il s'attache davantage aux structures considérées «plus pures» ou «moins latines», et priorise toujours les mots appartenant au fonds lexical basque par rapport aux emprunts.
- Ses textes sont souvent difficiles à comprendre, mais ils remplissent «mieux» la fonction symbolique décrite, puisqu'ils sont censés être plus «purement basques».
- Il est souvent en relation avec les organes supérieurs de l'administration autonome basque.

Courant B :

- Il cherche à répondre rapidement aux efforts de modernisation et d'adaptation de la langue à de nouveaux usages.
- Il cherche un plus grand rapprochement avec le parler des moyens urbains, sans trop se soucier de l'emploi des emprunts lexicaux.
- Il rejette l'emploi purement symbolique de la langue.

Il est préoccupé par la précision technique, essayant de trouver des équivalents lexicaux pour tous les mots espagnols, sans trop distinguer les termes techniques et les unités figées par la tradition.

Cette tendance est implantée surtout dans les administrations municipales et dans la terminographie spécialisée.

Il faut quand même dire que dès le début des années 1990 les vents poussent à la convergence. Même si des points de discussion restent encore sans accord, tant la coopération pratique que les débats qui se déroulent mènent à des solutions communes. Il est des thèmes où l'accord paraît bien établi, comme par exemple, la critique à l'emploi purement symbolique de la langue, ou la nécessité de s'adresser aux bascophones actuels dans un langage actuel, et non pas dans une langue idéale qu'il faudrait leur apprendre. Depuis quelques années, des commissions mixtes, représentant tous les courants et une grande variété de secteurs (académiciens, philologues, terminologues, représentants des diverses administrations, etc.), ont travaillé à l'élaboration de recueils terminologiques, ou de formulaires administratifs et judiciaires, arrivant à des solutions de consensus approuvées par tous.

Mais si l'accord entre les tendances est tout à fait nécessaire pour trancher des différends ayant à voir avec les idéologies à propos de la langue, il est clair qu'un approfondissement des problèmes plus spécifiquement linguistiques aiderait beaucoup à trouver des solutions.

Il faut tenir compte de la grande dépendance du langage de spécialité juridique à l'égard du langage commun. Plus que dans n'importe quel autre technolècte, la plupart des termes juridiques sont à l'origine des mots de la langue commune. Les travaux lexicographiques en cours sur le basque unifié fixeront, sans doute, une base beaucoup plus sûre pour la terminographie juridique, mais ils devraient tenir compte du «côté terminologique» de beaucoup de mots.

Quant à elle, la terminographie juridique doit passer — elle est déjà en train de passer — à une nouvelle phase. Les dictionnaires juridiques et administratifs existants sont pour la plupart normatifs plutôt que descriptifs; ce sont des recueils de propositions plutôt que des descriptions de la réalité. C'est bien compréhensible puisque le corpus existant au moment de leur confection était plutôt maigre. Mais nous avons actuellement une masse assez considérable de textes juridiques qui n'a jamais été dépouillée.

Des travaux de dépouillement sur corpus nous permettraient de connaître le degré d'acceptation des diverses propositions terminologiques faites par les dictionnaires existants et de commencer à étudier le poids des emprunts, des calques et des autres formations néologiques dans l'établissement de la nouvelle terminologie juridique. Actuellement, des travaux terminographiques partiels ont déjà emprunté cette voie de dépouillement de corpus.

À part cet aspect quantitatif, nous avons aujourd'hui une partie de la terminologie juridico-administrative qui est effectivement employée comme langue de communication et de travail. Des études sociolinguistiques à ce sujet nous permettraient de mieux connaître

le degré d'usage, d'acceptation, et donc, de compréhensibilité des termes proposés par les différents dictionnaires.

7. CONCLUSION

Il est clair que seules les langues qui ont joui historiquement d'une reconnaissance de la part des autorités politiques ont pu développer sans entraves les différents langages de spécialité. Il est aussi évident que le technolècte du droit est la variété linguistique la plus dépendante de la politique : pourquoi écrirait-on des textes juridiques s'ils n'ont pas de valeur officielle ?

Ainsi, la langue basque a perdu la mémoire de son droit, et seulement quand elle a conquis son officialité le langage et la terminologie du droit ont commencé à être nécessaires. Mais l'arrivée de l'officialité ayant été vécu comme une victoire et une libération, les textes juridiques ont eu souvent à remplir une fonction symbolique plutôt que pratique. La traduction est devenue la source la plus importante de production de textes juridiques, au détriment de la création propre, et donc, de l'autonomie de la langue. Pendant un certain temps, des réponses improvisées à des besoins considérés comme pressants ont configuré des modèles de langage et des terminologies assez divergents.

Mais actuellement, à l'administration et à l'université surtout, le technolècte juridique est en train de passer l'épreuve de la pratique. La fonction symbolique prendra l'espace nécessaire mais réduit qui lui correspond. D'autre part, la lexicographie basque a fait d'énormes progrès ces dernières années, ce qui permettra de trancher beaucoup de discussions sur le lexique standard commun employé dans les textes juridiques. Il reste que les travaux terminologiques actuels, basés dans une meilleure connaissance du corpus existant, permettent de fixer une base plus ferme pour ce langage juridique que nous voulons authentiquement basque et pleinement communicatif et technique.

RÉFÉRENCES EN LANGUE NON BASQUE

- 1789 et les Basques* (1991) : Bordeaux, Presses Universitaires de Bordeaux, pp. 171-194.
- AGIRRE GANDARIAS, S. (1992) : «La oficialidad del euskera en documentos postmedievales», *ASJU*, XXVI,I, San Sebastian, Seminario de Filología Vasca Julio de Urquijo, pp. 259-279.
- AGIRRE GARAI, J. (1988) : «El lenguaje administrativo en euskera : situación y perspectivas», *Revista de llengua y dret*, 11, pp. 91-98.
- AZKARATE, M. (1988) : «Neologisms and Lexicography in the Basque Language», *Budalex Proceedings*, Budapest, Akadémiai Kiadó.
- BERASATEGI, J.I. (1989) : «La traducción jurídico-administrativa en el País Vasco», *Revista de Llengua i Dret*, 13, pp. 29-45.
- CARO BAROJA, J. (1945) : *Materiales para una historia de la lengua vasca en relación con la latina*, San Sebastian, Universidad de Salamanca.

COORAY, L.J. Mark (1985) : *Changing the Language of the Law. The Sri Lanka experience*, Québec, Les Presses de l'Université Laval.

CORNU, G. (1990) : *Linguistique juridique*, Paris, Montchrestien.

MICHELENA, L. (1984) : *Sobre la historia de la lengua vasca*, San Sebastian, Seminario de Filología Vasca Julio de Urquijo.

RICA ESNAOLA, M. (1975) : «Traduction en basque de termes politiques sous la révolution», *ASJU*, IX, pp. 3-172.

TRASK, L. R. (1997) : *The History of Basque*, London & New York, Routledge.

URRUTIA, A. (1987) : «Derecho, planificación lingüística y euskara», *HAEA/RVAP*, 17, pp. 119-123.

DICTIONNAIRES

AGIRRE, J. et A. DIAZ DE LEZANA, (1991) : *Vocabulaire des traités européens/ Europako Ituneeen Hiztegia*, Bilbao, IVAP.

AZKARATE, M., X. KINTANA et X. MENDIGUREN BEREZIARTU (dir) (1996) : *Elhuyar Hiztegia*, (dic. Basque-espagnol / Espagnol-Basque), Usurbil, Elhuyar.

MICHELENA, L. (1987) : *Diccionario General Vasco/ Orotariko Euskal Hiztegia*, Bilbao, Euskaltzaindia, Desclée, Mensajero.

SARASOLA, I. (1996) : *Euskal Hiztegia*, Donostia, Kutxa Fundazioa.

XARRITON, P. et X. KINTANA (1997) : *Dictionnaire/Hiztegia. Euskara-Frantsesa. Français-Basque*, Donostia, Elkar.

UZEI (1983) : *Administrazioa, Zirkulazioa Hiztegia*, [Dictionnaire d'administration et de circulation], Donostia, Elkar.

UZEI (1985) : *Zuzenbidea Hiztegia*, [Dictionnaire de droit], Donostia, Elkar.

UZEI (1994) : *Hirigintza Hiztegia*, [Dictionnaire d'urbanisme], Donostia, Elkar.