

Étude contrastive des emprunts autochtones —cas des emprunts à l'arabe et au japonais —

Motoyuki MIYAGAWA

1. Introduction

La question concernant les variétés et les variations des langues ne se traite pas par un seul critère, mais bien au contraire, il doit y avoir plusieurs concepts et cadres d'analyse pour étudier cette question. Pour aborder ce sujet, nous allons envisager la situation linguistique en Tunisie ainsi que celle au Japon pour qu'on puisse établir une véritable comparaison, i.e. une étude contrastive pour faire saillir les particularités de ces deux environnements linguistiques. La situation linguistique en Tunisie se distingue par un plurilinguisme (croisement et contact de différentes langues comme l'arabe, le français, l'anglais, l'allemand, l'italien, l'espagnole, le chinois, etc.) et par une diglossie du fait même qu'il existe deux codes langagiers : l'arabe littéral et dialectal. Nous recourons à un corpus d'une grande ampleur qui nous permettrait :

- d'élaborer des descriptions lexicales précises des formes et emplois autochtones,
- de mieux connaître les procédés et processus de formation lexicales et d'emplois autochtones,
- d'élaborer un dictionnaire de géosynonymes.

Nous partirons du lexique et nous focalisons sur les trois domaines suivants : les emprunts autochtones, les géosynonymes et les mots construits parce qu'ils illustrent bien les spécificités les plus saillantes du français en Tunisie. Nous nous basons sur la définition proposée par S. Mejri (2011) de ces trois domaines que nous analysons. Parmi ces trois notions, nous allons d'abord aborder la première, et nous nous limiterons à traiter les emprunts autochtones, en envisageant une comparaison avec un autre cas, celui des emprunts faits au japonais par le français, un cas qui est bien différent et qui pourra certainement servir à leur donner un premier contour. Parce que, sans faire aucune comparaison, on n'arrivera jamais à éclaircir telle ou telle particularité, ni à rester neutre et objectif pour étudier ce genre de phénomène.

2. Emprunts, importation vs substitution ?

Pour parler des emprunts, il faudrait d'abord mentionner la différence entre ces deux notions suivantes : l'importation et la substitution. Selon Haugen (1950), l'importation signifie littéralement les importations de modèles de la langue empruntée, conservant la similarité morphologique et/ou phonologique de ces modèles originels. Par contre la substitution signifie les remplacements de modèles originels par le format de la langue emprunteuse. Suivant ce critère, il a classifié les emprunts en trois catégories :

1. emprunts lexicaux (loanwords)¹

Importation morphématique, pas de substitution. Cette catégorie se divise ensuite en trois sous-catégories selon de degré de substitution phonémique, i.e. nulle, partielle ou complète.

2. emprunts-mélanges (loanblends)

Importation morphématique ainsi que substitution. Toute substitution implique un certain degré d'analyse par le locuteur du modèle qu'il imite.

3. emprunts-décalage (loanshifts)

Substitution morphématique seulement. Cette catégorie comporte ce qu'on appelle normalement les calques et les emprunts sémantiques. Le terme décalage (shift) est suggéré parce que cette sorte des emprunts apparaît seulement sous la forme du décalage fonctionnel des morphèmes natifs.

Dès le début, l'arabe ainsi que le japonais ne partagent pas le système d'écriture avec le français, l'importation des morphèmes dans un sens rigoureux est impossible. Alors aucuns emprunts doivent être classés comme les emprunts lexicaux stricto sensu, parce que dès que l'on essaie d'écrire un mot, nous serons obligés tout de suite de changer ce mot en quelque sorte que ça soit intégré à l'autre système. Cependant, si l'on considère tous les cas des remplacements comme substitution des modèles originels, on devrait désormais classer tous les exemples sous la catégorie « emprunts-décalages », et la distinction entre les cas des calques et/ou emprunts sémantiques n'est plus valable. Comme les calques et les emprunts sémantiques sont deux phénomènes linguistiques complètement distinctifs des autres cas, il faudrait les traiter autrement. Concernant ce problème, nous proposons après une classification plus détaillée.

3. Emprunts autochtones

La notion des emprunts autochtones n'est pas un concept qui s'oppose aux trois catégories mentionnées ci-dessus. C'est un concept d'une autre dimension qui est sociolinguistique. Il s'agit plutôt de la circonstance dans laquelle les emprunts se font d'une langue par une autre langue. Dans les emprunts autochtones, on peut donc prévoir aussi l'existence de toutes les trois catégories, soit emprunts, emprunts-mélanges et emprunts-décalages.

Suivant la définition proposée par Mejri (2011), nous décrivons ici les emprunts autochtones comme les emprunts qui se font à la langue (aux langues) autochtone(s) par une autre langue non maternelle (ou étrangère). C'est-à-dire, pour parler des emprunts autochtones, il faut d'abord avoir comme condition préalable un territoire où une ou plusieurs langues maternelles coexistent avec une ou plusieurs langues étrangères totalement ou

¹Nous utilisons désormais LW, LB et LS respectivement pour ces trois catégories comme abréviation.

partiellement partagées par la même communauté linguistique. Par cette définition, la situation du japonais, plus précisément, la situation autour des emprunts du français au japonais, cela ne serait pas évident du tout pour l'instant, puisque la situation linguistique du Japon est déjà connue fortement comme celle de « monolingue² », l'existence des emprunts autochtones est assez douteuse ou au moins controversée. Ce qu'il nous faut faire est donc d'ores et déjà de vérifier ce point-là. Autrement dit, ces deux situations sont considérablement écartées, et par ce fait même nous pouvons caractériser et distinguer nettement ces deux phénomènes linguistiques.

3.1. Cas à l'arabe en Tunisie

Plusieurs travaux ont été menés sur les particularités du français en Tunisie de nature contrastive. Cela peut varier depuis la phonologie, la syntaxe et le lexique, et de nombreuses recherches sont consacrées à marquer le trait de la coexistence de différents systèmes linguistiques. Chaque domaine enrichissant toutes les recherches sur les particularités de cette variété, nous nous limitons ici à traiter le lexique, car les mots sont tout d'abord un point de départ pour les autres domaines de la linguistique, tels que la phonologie la syntaxe, la morphologie, la sémantique et la pragmatique. Étant toujours servis comme les appuis dans ces autres domaines, les mots peuvent être en même temps une sorte de premiers récepteurs de la réalité concrète ou abstraite, qui nous entourent, et représentent aussi toutes les caractéristiques d'une société.

Nous avons déjà un inventaire du lexique au français en Tunisie, qui reflète clairement la situation actuelle et la particularité de cette variété. Neffati et Queffélec (2004) ont établi une bonne entrée du lexique basé sur des corpus écrits et oraux, extrait des réalisations écrites et discursives du français en Tunisie et cela consiste en 1,835 entrées³ en total. Ici, nous allons nous servir de cet inventaire comme référence non négligeable, mais pour compléter cette liste et pour une recherche plus exhaustive, nous allons aller plus loin en utilisant la méthodologie du traitement automatique des langues. Il faut noter en même temps qu'existent certaines critiques dans lesquelles on ne considère l'inventaire que comme une simple liste des mots arabes et/ou arabes dialectaux transcrits en alphabet latin, mais en profitant de ce moyen de traiter des corpus d'une grande ampleur ainsi qu'en appliquant le cadre d'analyse pour sonder que nous allons mentionner ci-dessous, de l'intégration des mots empruntés dans une autre langue, nous croyons pouvoir montrer qu'il s'agit bien des cas où l'on peut cerner certains traits d'une variété vivante de langue.

3.2. Cas au japonais

Par contre, la question concernant les emprunts du français au japonais est forcément bien limitée vu la nature de la circonstance linguistique de ce territoire. La version 2007 du dictionnaire le Petit Robert, qui est largement

²Il est à noter quand même qu'il n'est pas le cas que cette notion soit parfaitement partagée.

³Par contre, le Petit Robert (édition 2007) ne comporte que 31 lemmes dont l'étymologie arabe maghrébine est constatée. (200 comme arabe d'où 233 pour l'arabe en total, dont 2 viennent de l'arabe de l'Espagne)

répandu et dont on s'est servi comme la première référence avant d'élargir après nos références à plusieurs grands dictionnaires, comporte néanmoins 67 entrées (le nombre total des entrées est de 58,125) comme mots venant du japonais. Cependant, il n'en est pas tout de même clair si ces mots-là sont déjà des mots français bien intégrés et qu'ils peuvent être considérés comme pouvant constituer des unités lexicales comme les autres mots français. A part cette question, il nous reste toujours une question plus naïve et intuitive mais non négligeable et importante, pourquoi ce mot-ci mais pas ce mot-là, c'est-à-dire, le choix des mots à emprunter.

Nous allons donc prendre le cas des emprunts au japonais dans le français qui est utilisé dans la communauté francophone au Japon. Et comme cadre des analyses, nous prenons le degré de l'intégration (naturalisation) des mots empruntés. Selon les recherches précédentes, nous distinguons quatre degrés de l'intégration :

Cannon (1994a)

1. caractère gras, en italique ou avec des guillemets
2. le stade précoce de l'adaptation phonétique, grammaticale, syntaxique et sémantique
3. enregistrés dans tous les dictionnaires non abrégés, plus ou moins productifs, mais pas complètement absorbés dans le lexique général
4. profiter de large usage international, apparaissant dans presque tous les derniers grands dictionnaires transportant diverses significations

Kimura (2000)

1. adaptation de la prononciation et de l'orthographe (accentuation, allongement, diphtongaison, adaptation orthographique)
2. usage limité comme attributs des noms (composé substantif = mot emprunté + mot explicatif)
3. acquisition de la productivité (mot composé, suffixation, changement fonctionnel de nom)
4. glissement sémantique (transfert de sens, l'extension de sens, l'utilisation métaphorique)

Notons tout de même que ces degrés étaient fixés pour classer les emprunts d'abord au malais, après au japonais dans l'anglais, il faudrait donc refixer les degrés eux-mêmes. Cependant on devrait faire référence à ces classifications, d'autant plus que cela nous montre le processus de l'intégration des mots empruntés, et, comme Cannon (1994b) a dit, ce n'était conçu pas seulement pour le cas spécifique mais aussi pour d'autres cas. Et en vérifiant et classifiant ces processus, nous pourrions en même temps répondre à la critique que l'on avait faite en ne considérant les emprunts autochtones que comme simples transcriptions en alphabet latin des empruntés.

4. Traitement automatique

Pour récupérer les mots empruntés dont la liste préétablie n'existe pas, on ne possède pas pour l'instant une

méthode définitive qui puisse les trier facilement en une seule fois. Alors il faudrait exécuter plusieurs méthodes, voire en créer nous-mêmes une qui est assez souple et efficace, puissante et exhaustive. Pour ce faire, nous envisageons de traiter nos corpus avec des données de grands dictionnaires du français, ainsi que des langues de programmation telles que Python et/ou Perl, pour qu'on puisse établir avec les expressions régulières un programme qui peut servir à obtenir tous les exemples susceptibles d'être emprunts autochtones. Il faut traiter les mots qui sont déjà sur certains dictionnaires comme emprunts ainsi que les autres mots pas encore lemmatisés.

4.1. Corpus et lavage

Comme corpus du français au Japon, nous allons prendre tous les articles écrits en français du journal « OVNI⁴ », dont l'éditeur se situe à Paris mais distribué aussi au Japon dans la plupart des établissements francophones. Il n'y a pas d'ailleurs beaucoup de publication en français dans la communauté nipponne. (sauf en ligne, des blogs plus en plus nombreux, on traitera dans nos projets à venir) Par contre, on peut trouver pas mal de journaux, magazines, revues en français en Tunisie. Nous allons trier ces documents convenablement pour qu'on puisse faire de véritables recherches contrastives. Pour cette fois-ci, nous avons traité les articles du magazine « Réalités⁵ », hebdomadaire indépendant. Effectivement, ces deux publications ont certainement leur politique éditoriale bien différente l'une à l'autre, soit OVNI qui est destiné plutôt à une petite communauté en traitant principalement les propos de la vie quotidienne, soit Réalités qui traite surtout les problèmes politiques et qui s'est fait même l'objet de plusieurs suspensions de publication, saisies et de procès au cours des années quatre-vingts. De ce fait, même s'il n'est pas possible de généraliser tout de suite les caractéristiques de chaque cas, on pourrait illustrer globalement mais assez clairement certains aspects qui peuvent varier selon les emprunts à telle ou telle langue.

Comme lavage des corpus traités, étant donné que tous les textes sont rédigés en français, nous n'avons donné aucune différence dans les deux traitements. Comme dans les autres cas, nous avons ôté tous les signes de ponctuations ainsi que les autres éléments statistiquement non significatifs, et avons divisé tous les textes en tokens, que l'on peut manipuler automatiquement par les langues de programmation et les expressions régulières.

4.2. Hypothèses

1. étant donné que ces critères de la naturalisation ne sont pas à qualifier initialement les emprunts au japonais ni à l'arabe dans le français, il serait probable que l'on sera obligé de redéfinir les critères eux-mêmes.

2. les situations dans lesquelles sont empruntés les mots « autochtones » (soit des mots japonais soit des mots

⁴Un journal bimensuel publié par les éditions Ilyfunet depuis 1974 dont le tirage compte environ 60,000.

⁵Un magazine hebdomadaire publié par Maghreb Média depuis 1979 tirant à peu près 23,500 exemplaires.

arabes ou d'autres langues) sont tout à fait différentes par excellence. On prévoit donc une certaine différence entre les processus de la naturalisation des mots empruntés au japonais ou à l'arabe.

3. il y aurait, tout de même, entre ces deux cas, des points communs qui pourraient témoigner la vivacité de la dynamique entre les langues en contact.

5. Résultat

Le nombre total des tokens traités :

« Réalités » 121,898, « OVNI » 147,676

5.1. Corpus « Réalités »

On a obtenu quelques milliers de mots qui ne sont pas enregistrés dans les dictionnaires. Cependant, la plupart d'entre eux ne sont que des fautes d'orthographe, sinon statistiquement non significatifs. Si l'on peut quand même constater une certaine tendance propre à ce corpus :

1. chutes des accents assez fréquentes.

2. une certaine série d'hypercorrections, i.e. avec une bonne connaissance de la langue française, on a parfois tendance à corriger plus.

Ex. y' a-t-il, elle *donnaît* des ordres, *bien-sûr*, etc.

Pour les exemples que l'on a pu saisir comme statistiquement significatif, nous pouvons constater la classification suivante :

a. Création relativement libre (suivant cependant toujours la règle de la formation morphologique du français, et aussi, ça arrive parfois des combinaisons des éléments venant du latin ou du grec ancien, comme si ça arrive parfois dans les domaines scientifiques et/ou chez les savants).

Ex. *avocature* (du latin), *décodable* (décoder + suffixe able), *cybersactiviste*, *cybercensure* (élément *cyber* + activiste, censure), *régimenter* (verbalisation à partir de régiment), etc.

b. Création topologique

Ex. *menzélien*, *djerbien*, etc.

c. Formation onomastique (mots créés par des noms propres, souvent depuis des noms de personne. Cela reflète bien clairement la réalité qui est autour de la société)

Ex. *bourguibiste*, *bourguibien*, *youssefiste*, *destourien*, *nassérien*, *RCDiste*, etc.

d. Emprunts autochtones

La plupart d'entre eux sont mis entre des parenthèses ou guillemets et avec la traduction en français. Cependant on a pu constater assez d'occurrences des emprunts sans aucune explication explicite en français.

Ex. *chariaâ*, *cheikh*, *kobkab*, *aghalabite*, *fatimide*, *ziride*, *hafside*, *husseinite*, etc.

5.2. Corpus « OVNI »

a. On a trouvé quelques néologismes qui reflètent la réalité actuelle de la société.

Ex. *électrophage* (reflétant la réalité actuelle de la société japonaise, concernant la politique d'économiser l'électricité), *irradiuse* avec une explication de la formation (radieux + irradiant), et toujours reflétant la réalité, *nucléo-nimbus* (cumulo-nimbus + élément nucléo), etc.

b. Pour les exemples que l'on peut éventuellement appeler emprunts autochtones, il y en a plein de mots suivant les propos qu'on traite dans ces articles du journal, mais presque tous ces mots sont :

1. substantifs
2. avec une explication en français
3. sans aucune formation morphologique du français

c. Presque tous les exemples sont avec une explication dans les parenthèses ou les guillemets. Cependant on peut constater des exemples, bien que peu nombreux, sans explication explicite.

Ex. *conbini* (*konbini* ou *kombini*, une sorte de supérette), *otaku* (une sorte de geek), *bentô* (une sorte de casse-croûte), *chara-ben* (character(angl.) + *bentô*), etc.

d. On peut constater la création topographique comme dans le corpus « Réalités ».

Ex. *akihabarien*, etc.

e. Dérivation par des suffixes.

Ex. *otakisé*, *otakuologie* (*otaku* + suffixe), *narutophile* (*naruto*, un titre d'un manga + suffixe), etc.

5.3. Statistique

Le pourcentage des tokens considérés comme emprunts autochtones pour les deux corpus :

Table 1.

Réalité	OVNI
131 tokens ⁶ / 121,898	498 tokens / 147,676
0.1 %	0.34 %

⁶Ici, le nombre signifie la fréquence des occurrences.

Répertoire des domaines de ces emprunts :

Table 2.

Réalités		OVNI	
politique	15 ⁷	aliment	38
religion	9	culture	35
histoire	6	tradition	19
vêtement	2	grade social	18
musique	1	société	15
parfum	1	religion	9
		calendrier	7
		animal	6
		onomatopée	6
		vêtement	5
		architecture	4
		transport	4
		musique	2

Pour les deux corpus, le taux des mots empruntés n'est pas forcément important. Si l'on peut noter tout de même que celui pour le cas au japonais est plus élevé ainsi que la variété du répertoire couvre plus de domaines, cela doit s'expliquer par la différence qualitative de ces deux corpus. La direction éditoriale des Réalités est celle de local et constituée de Tunisiens, par contre la plupart des articles dans le journal OVNI sont rédigés par des Français demeurant au Japon. Partant de ce fait, on peut prévoir que la nécessité et la volonté de vouloir représenter ce qui est nouveau et inconnu pour les auteurs devront être plus évidentes pour le cas au japonais. On peut dire aussi que la variété du répertoire s'explique par le fait que le corpus « OVNI » peut être classifié comme discours culturel qui couvre plusieurs domaines tandis que le corpus « Réalités » comme discours politique. Concernant cette différence, nous envisagerons de traiter prochainement d'autres corpus que l'on peut classifier comme discours culturel en Tunisie. Nous posons comme hypothèse que même si l'on traite des corpus de discours culturel, on pourra constater certaines différences par rapport à la distance culturelle, et que cette distance, autrement dit, comment une autre culture nous semble différente, peut s'éclaircir d'une manière diachronique par le degré de la fusion et de l'intégration culturelle des sociétés en contact.

6. Conclusion

Concernant la question entre les notions « importation / substitution », nous détaillons nos analyses comme suit en prenant les exemples ci-dessous. D'un point de vue qu'il est possible de penser que cela prend une forme

⁷Les chiffres dans ce tableau représentent le nombre des types des exemples.

assez similaire du modèle originel, même si c'est indéniable que l'on remplace ses lettres, l'emprunt *bentô* peut être classifié comme une importation morphématique. La sous-catégorie de ce mot est classée alors dans « sans substitution phonémique », d'autant plus que le remplacement phonologique peut être considéré comme minimum. L'emprunt *chara-ben*, tant que ce mot est épilé tel quel, ce n'est pas une importation rigoureuse du modèle originel ni une substitution par la forme française « caractère », mais cela importe le « proto-modèle » en faisant référence au mot anglais *character*. Dans ce cas-là aussi, l'emprunt est classifié comme LW, la sous-catégorisation se fait de même. L'emprunt *depâto* se fait d'abord au mot *department store* en anglais comme LW, avec la substitution phonémique partielle. Ensuite, cet emprunt subit encore une fois une transition, toujours comme LW, mais cette fois-ci sans la substitution phonémique. Le mot *département* existe en français (couramment grand magasin, mais on dit aussi magasin par départements), mais dans ce cas-là cela ne substitue pas cet emprunt, parce qu'on se sert de l'emprunt pour signifier une autre réalité.

Contrairement aux exemples ci-dessus, l'emprunt *maman(s)-bentô* nous montre un processus différent. Le modèle natif est quasiment inexistant, cependant ce mot est produit probablement par l'analogie du mot *maman-gâteau* avec l'emprunt déjà fait *bentô*. Dans ce cas-là, il nous paraît convenable de classer ce mot comme LB.

Tous ces exemples-ci nous montrent un processus nettement distinctif par rapport aux cas des calques et/ou des emprunts sémantiques. Nous devons donc les classer autrement. Cependant, comme dans le cas de ces exemples, quand on analyse des emprunts faits aux langues dont le système d'écriture n'est pas identique, par le biais en plus d'une ou plusieurs langues, une classification plus détaillée devra dorénavant être nécessaire. Car le mot *bentô* et le mot *depâto*, classifiés tous les deux comme LW, témoignent en fait deux processus différents l'un de l'autre.

Comme le degré de l'intégration des emprunts, nous proposons ces quatre degrés suivants :

1. usage limité comme substantifs, avec explication explicite, sans aucun changement morphologique.
2. acquisition de la similarité morphémique par rapport à la formation des mots français. (forme plurielle, composition des mots composés, remplacement de phonèmes)
3. acquisition de la productivité par la dérivation, la troncation.
4. glissement sémantique.

Car nous avons constaté que, pour tous les deux corpus, quasi-totalité des emprunts ne correspondent qu'aux premiers deux degrés proposés par Cannon (1994b). Et le troisième et le quatrième degré, c'est-à-dire, si les emprunts sont déjà lemmatisés dans les dictionnaires et qu'ils ont déjà acquis l'usage international ou pas, ce ne sont pas des critères intrinsèquement évidents mais des critères d'une autre dimension. Par contre, le premier et

le deuxième degrés selon Kimura (2000), i.e., l'adaptation de la prononciation et de l'orthographe et l'emploi limité comme attribut des noms, ces deux phénomènes peuvent se produire simultanément d'une manière générale, et peuvent varier fortement selon les locuteurs. De ce fait, nous avons les traités comme une seule étape de l'intégration. Or, on pourrait prévoir qu'il y aura un écart assez important entre le troisième et le quatrième degré parce que nous n'avons pas trouvé aucun exemple pour le moment qui subit un tel glissement sémantique pour les deux corpus. Nous pensons que la globalisation, surtout celle d'informatique, dans laquelle nous pouvons facilement faire référence aux sources et acquérir des connaissances assez exactes concernant la réalité que les emprunts représentent joue un rôle négligeable dans nos jours.

Pour conclure, si l'on signale les traits qu'on a pu constater à travers nos traitements, le résultat est suivant : Pour tous les deux corpus, les emplois sans explication explicite sont fortement limités. Cependant pour le cas des emprunts faits à l'arabe, l'utilisation sans explication témoigne plus d'exemples par rapport au cas des emprunts au japonais, et ce phénomène se voit surtout dans le domaine religieux. Pour tous les deux cas, la formation depuis des noms propres est assez fréquente et fournit une productivité relativement stable. Les emprunts se font surtout pour exprimer des choses, des phénomènes qui sont en train de se produire actuellement, et qui reflètent bien la réalité des sociétés concernées. Nous envisagerons dans l'avenir d'élargir nos corpus pour saisir plus de détails et élucider les processus des emprunts, surtout d'un point de vue diachronique qui pourra certainement servir à la description exhaustive et complète.

Bibliographie

- Cannon, Garland (1994a) : « Recent Japanese Borrowings into English », *American Speech* 69, Duke University, pp. 373-397
- Cannon, G., Kaye, A.S. (1994b) : *The Arabic Contribution to the English language: an historical dictionary*, Wiesbaden: Harrassowitz Verlag.
- Haugen, Einar (1950) : « The Analysis of Linguistic Borrowing », *Launquage*26, pp.210-31.
- Mejri, S., Said, M., Sfar, I. (2009) : « Plurilinguisme et diglossie en Tunisie », *Synergies-Tunisie*, n°1, Réseau Gerflint, Paris, pp.53-74.
- Mejri, Salah (2011) : « Les spécificités du français en Tunisie : emprunts autochtones, géosynonymes et mots construits », Colloque international *Convergences, divergences et question de la norme en Afrique francophone*, Université de Cologne, 2-4 février 2011.
- Kimura, Makimi (2000) : « The Naturalization Process of Japanese Loanwords as Reflected in English Dictionaries: The Four-Stage Hypothesis and Associated Problems », *Lexicographia, Series Major* 103 (Symposium on Lexicography IX), pp.293-303.

(Doctorant à l'Université de Tsukuba)