

Sur la conception de la recherche étymologique ou L'étymologie en réseau

dernière mise à jour 10 /10/2008

Ce texte est le développement de la communication faite par Roland Laffitte à l'Université de Tlemcen, le 20 mai 2008.

Introduction

Pour ce qui concerne les travaux en langue française, l'étymologie des mots d'origine arabe et orientale dans les langues européennes est depuis très longtemps davantage un exercice auquel se livrent des linguistes ou des historiens comme travail annexe à leurs recherches, qu'une entreprise systématique, comme cela peut être le cas en langue espagnole, italienne ou allemande. Le fait que l'étymologie des mots d'origine arabe et orientale ne se présente en langue française pas comme une discipline à part entière possédant ses propres règles peut laisser à place à des improvisations et à des fantaisies parfois déroutantes.

C'est la raison pour laquelle nous appellerons dans un premier temps à dépasser l'étymologie populaire. Nous essaierons, dans un second temps, d'envisager un éventail de possibilités étymologiques et inviterons, dans un troisième temps, à replacer les étymons étudiés dans leur parcours linguistique d'ensemble. Ces conseils que nous donnons ici comme premier bilan de six années de travail collectif et de publication du *Bulletin de la SELEFA* ne sont qu'une première approche de la conception de la recherche étymologique à laquelle nous essayons de faire vivre.

1. Dépasser l'étymologie populaire

* Paronymies et étymologie * Le syndrome de Cheikh Zoubir

Le premier écueil consiste à prendre pour argent comptant l'étymologie populaire, c'est-à-dire les explications qui, en s'appuyant sur certaines ressemblances formelles ou paronymies sémantiques ou phonétiques, rattachent un mot inconnu à un autre connu avec lequel il n'a pas de racine commune.

Ainsi le français *choucroute* semble provenir de *chou* et de *croûte* alors qu'il s'agit du terme alsacien *surkrut*, qui correspond à l'allemand *sauerkraut*, littéralement « herbe amère ». Ne serait-ce que pour des raisons mnémotechniques, le locuteur français qui entend le mot *surkrut* le rattache de façon inconsciente à des termes familiers de même allure phonétique. Il va de soi que cela n'a rien à voir avec une explication étymologique, mais ce phénomène présente un intérêt, car il permet d'expliquer l'orthographe du terme. En nous renseignant sur les conditions de la réception du mot, ce qui est particulièrement vrai pour l'emprunt, l'examen de l'étymologie populaire est partie intégrante de la recherche étymologique.

Prenons un exemple en langue arabe, celui du terme *الزبانا / الزباني al-Zubānā* qui le nom d'un des 28 *منزل القمر manāzil al-qamar*. Il faut savoir que les étoiles repères de cette mansion lunaire sont placées depuis Babylone sur la figure *الميزان d'al-Mizan*, « la Balance », laquelle trouve place entre les *قرون العقرب Qurūn al-ʿaqrab*, soit « les Pincettes du Scorpion ». Partant de là, les linguistes arabes classiques ont cru pouvoir faire dériver le terme *الزباني al-Zubānā* de la racine ZBN avec le sens de « pousser ». En effet comme l'explique ʿAbd al-Raḥmān al-Ṣūfī dans son *Kitāb ṣuwar al-kawākib al-tābita*, « *الزبانا* vient de *الزبن*, chacune d'elles étant poussée l'une de l'autre sans se rencontrer ». En fait, il ne s'agit pas d'un mot arabe, mais d'un emprunt à l'akkadien *Zibānītu* qui est signifié « balance » et est effectivement utilisé pour désigner la constellation de la *Balance*.

Un des paroxysmes de l'étymologie populaire est atteint dans notre domaine lorsqu'il s'agit de trouver une explication fantaisiste par la langue arabe à des termes européens. On pourrait nommer ce travers le « syndrome de Cheikh Zoubir », en référence à cette rumeur ridicule selon laquelle Shakespeare serait d'origine arabe... Les ravages de cette tendance sont grands. L'auteur d'un livre sur *L'origine arabe de la langue française* qui a un certain succès de librairie, a beau prétendre avoir « évité l'exagération de certains mots français à une origine arabe », le premier mot français qu'il aborde est *abbé* qu'il fait venir de l'arabe ʿ*abbād*, « adorateur », ce qui ne correspond à aucune réalité ni linguistique ni historique. Il s'agit de l'araméen *abā*, qui correspond à l'arabe *abu*, « père », par l'intermédiaire du grec ecclésiastique *abba*. Si le terme arabe ʿ*abbād* présente une ressemblance formelle avec l'espagnol *abad* vient du passage par le latin dont la forme est *abbatem* à l'accusatif. Cela prouve que le titre de docteur ès lettres de l'Université n'est nullement une garantie contre le ridicule pour ce linguiste dont le nom sera tu pour lui éviter la honte publique dans son pays.

2. Envisager l'éventail des possibilités étymologiques



La recherche étymologique ne mène pas toujours à trouver un lien direct et simple entre un emprunt ou un calque et son étymon dans la langue d'origine. Il faut au contraire analyser le legs lexical comme un processus complexe, parfois contradictoire. Nous pouvons nous en rendre compte en examinant une gamme des rapports entre un terme considéré dans la langue d'arrivée et son étymon dans la langue de départ, sans pour autant prétendre à une typologie exhaustive :

* **L'étymon pur :** Certes, il existe une proportion non négligeable de cas où nous trouvons une correspondance immédiate entre un terme de la langue d'arrivée et son étymon dans une langue étrangère. C'est souvent le cas de termes scientifiques ou techniques comme l'arabe الإنبيق *al-inbīq*, qui a donné *alambic*.

Mais même dans ce domaine, le rapport entre l'étymon et le terme légué, qu'il s'agisse d'un emprunt ou d'un calque, est plus complexe que l'on pourrait s'y attendre au premier abord. Le terme legs lexical peut en effet cacher toute un éventail de modifications sémantiques possibles. À côté de la correspondance étroite, nous avons en effet :

1. La spécialisation : c'est le cas de l'arabe الجبر *al-jabr*, déjà évoqué *supra*, le sens mathématique du terme étant dérivé de la notion initiale de « contrainte », toujours en usage dans l'arabe commun.

2. L'extension sémantique, comme c'est le cas de لاجورد *lājuward* / لازورد *lāzuward*, « lapis lazuli », c'est-à-dire la pierre de couleur bleue dont le nom s'étend, par synecdoque, à la couleur elle-même, soit *azur*.

* **Le croisement :** il s'agit de l'action de deux mots qui agissent l'un sur l'autre par contamination.

1^{er} cas : croisement de 2 termes d'amont : l'arabe خروب *ḥarrūb* semble être le croisement de deux termes. Nous avons bien l'araméen חרובה *ḥarūbā*, ce qui n'explique pas mais il faut expliquer la gémination du /r-/, le passage du phonème /h-/ en araméen au /ḥ-/ en arabe, ainsi que le doublet خرنوب *ḥarnūb* (voir *Bull. SELEFA* n° 9, 21-28).

2^{ème} cas : croisement d'un emprunt avec un terme de la langue d'arrivée.

* transformation de l'emprunt par assimilation phonétique à un mot de la langue d'arrivée. C'est là que se manifeste l'étymologie populaire : l'expression « s'en fichier comme de l'an quarante » dérive, par étymologie populaire de « s'en fichier comme de l'alcoran ».

Mais il peut aussi y avoir l'influence de l'étymologie savante : ainsi, quand Johann Bayer demande, vers 1600, à un ami linguiste d'où peut venir le nom *Sartai* pour l'étoile *Gamma* du *Bélier*, cet ami, féru d'hébreu et qui pense que l'arabe vient de cette langue, lui répond que cela peut être l'hébreu משארתיים *m^ošār^ot^hīm*, « servants », terme utilisé dans le lexique grammatical, en suite de quoi l'astronome renomma cette étoile *Mesarthim*. Il fallut attendre quelques temps pour que les textes arabes permettent d'établir qu'il s'agissait de الشرتانن *al-Šaratān*, littéralement « les Deux Marques », c'est-à-dire celles qui parquent l'équinoxe de printemps et le premier منزل القمر *manzil al-qamar*, d'où le nom donné conséquemment à l'étoile *Alpha* du *Bélier* qui est aujourd'hui *Shératan*.

* suffixation : أخل النهر *aḥir al-nahar* devient le latin médiéval *Achernar* et le français *achernar* par influence du suffixe *-ard*, du francisque hart, « dur », cf. le nom *Bernard*.

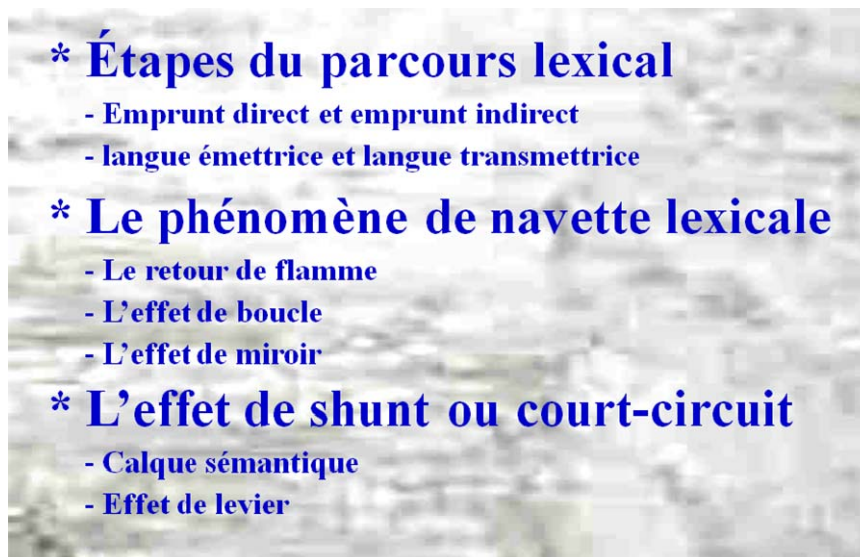
* transformation de la transcription par assimilation à un schème graphique de la langue d'arrivée :

* **L'interférence** : c'est le phénomène qui, en matière d'emprunt, se caractérise par l'utilisation, dans la langue d'arrivée, d'un élément lexical ou une forme propre à la langue de départ.

1. Une forme d'interférence que nous avons déjà rencontré est le **calque sémantique** (voir *supra*) Un exemple est fourni par le français populaire contemporain quand *calculer* prend l'acception « considérer, remarquer », qui est une interférence évidente de l'arabe حسب *ḥasaba*, où ce sens est dérivé de « calculer ».

2. Une autre forme d'interférence est le résultat de ce qu'on pourrait appeler l'**effet de levier**. Dans ce cas, un mot existant est propulsé sous l'effet de son correspondant dans une langue étrangère.

3. Replacer l'étymon dans son parcours linguistique



* Étapes élémentaires du parcours lexical

1. **Emprunt direct et emprunt indirect.** On peut parler d'emprunt au sens strict lorsque une langue A intègre ou finit par intégrer un terme qui existait précédemment dans une autre langue. Au sens strict, le français *algarade*, est un emprunt à l'espagnol *algarada*, lui-même un emprunt à l'arabe الغارة *al-gāra*, « raid, incursion en territoire ennemi ». Parallèlement, le moyen français *arsenal* doit être considéré comme un emprunt au vénitien *arsenal* qui est lui-même un emprunt à l'arabe دارا لصناعة *dār al-ṣināʿa*, alors que la première forme attestée en français avait été l'ancien français *tarsenal*, directement emprunté au même étymon arabe dans le royaume de Jérusalem. Il faudrait donc distinguer ici « emprunt direct », comme dans le cas de *tarsenal* et « emprunt indirect » dans le cas d'*algarade* ou de *arsenal*.

Il faut toutefois faire remarquer que l'emprunt indirect peut suivre lui-même un long parcours entre la langue arabe et la langue d'arrivée, comme c'est le cas d'*amiral* que le français doit probablement au latin *admiratus* au byzantin ἀμυρᾶς, lui-même emprunté à l'arabe أمير *amir*. Le parcours sera encore allongé d'une étape si l'on poursuit jusqu'à l'anglais *admiral*.

Quoiqu'il en soit, on considère tous ces termes comme des « arabismes », quel que soit le parcours qui est celui de ces termes à l'intérieur des langues européennes.

Reste à savoir, pour rester dans les emprunts à la langue arabe, si l'étymon arabe considéré se rattache à une racine authentiquement arabe ou est lui-même un terme emprunté.

2. **Langue émettrice ou langue transmettrice.** Dans de nombreux cas, la langue qui a fourni l'étymon est bien l'émettrice de l'étymon. Te est le cas de nombreux termes que nous avons jusqu'ici mentionnés dans cet exposé : أخل النهر *aḥir al-nahar*, أمير *amir*, الثانية الجبر *al-tāniya, al-jabr*, الغارة *al-gāra*, دارا لصناعة *dār al-ṣināʿa*. Nous avons néanmoins rencontré le cas de الزباني / الزبانا *al-Zubānā* qui a donné, pour des noms d'étoiles toujours utilisés, *Zuben* et *Acubens*, qui est lui-même l'akkadien *Zibānītu*, « la Balance », par l'intermédiaire d'un dialecte araméen, sachant que nous avons en mandéen la forme *zibānītā* pour « la balance ». En réalité, bien qu'il a existé historiquement de fortes polémiques entre linguistes pour

savoir l'origine de nombreux termes, notamment pour ce qui concerne les mots étrangers dans le *Qur'ān*, il existe en langue arabe de nombreux termes reconnus comme des emprunts. Il suffit pour cela de se référer à l'ouvrage d'Al-Jawāliqī écrit au début du XII^e s. / h. V^e s., le fameux *Mu'arrab*.

Ce linguiste met avant tout en lumière les mots persans comme c'est le cas de nombreux mots passés dans les langues européennes. Citons par exemple لشکر *laškar*, qui, d'un côté par l'urdu et le portugais puis, dans une deuxième vague par l'arabe العسكر *al-^caskar*, a donné plus récemment le français *lascar*, ou encore le palhavi *šalwār* – passé tel quel dans le persan شلوار – qui explique l'arabe سروال *sarwāl*, qui a donné le français *sarouel*, en vogue aujourd'hui chez grands couturiers.

Mais bien d'autres langues ont enrichi le lexique arabe. Nous avons d'un côté les langues asiatiques, notamment le sanscrit, les langues dravidiennes, le malais et le chinois. Nous avons également le substrat sémitique ancien et toutes les langues de la famille sémitique dont le corpus est largement sous-évalué : notamment l'akkadien, le sudarabique et l'éthiopien, l'hébreu ancien et surtout l'araméen, notamment dans sa variété syriaque, en particulier du fait de son rôle d'intermédiaire dans les traductions du grec à l'arabe et dans sa contribution à la terminologie scientifique et intellectuelle de l'arabe classique. Et ceci sans parler des apports modernes des langues européennes, le français, l'anglais ou l'italien...

2. Le phénomène de navette lexicale

Nous pouvons parler de navette lorsqu'un terme emprunté par une langue donnée revient à sa langue d'origine par une voie ou par une autre.

* Il peut s'agir d'un simple **retour de flamme** : ainsi le terme *français* qui se dit en dialectal maghrébin فرانساوي *frānsāwī* croisé avec le terme originel entre dans le français colonial sous la forme de *frankaoui* pour désigner les Français de métropole, et se retrouve aujourd'hui, à partir des banlieues, dans la langue des jeunes pour désigner un « Français du type beauf, franchouillard ». De même, l'arabe algérien كَرُو *gārrū*, « cigarette », qui vient de l'espagnol *cigaro* par l'intermédiaire des colons d'origine hispanique, se retrouve aujourd'hui, par l'intermédiaire de l'immigration maghrébine en France, dans la langue des jeunes sous la forme *garo*.

* Le phénomène de navette peut être bien plus complexe et traduire un **effet de boucle** complexe entre les langues de deux rives de la Méditerranée.

Ce terme a réellement une histoire infinie, reflet de celle du fruit qu'il désigne. Originnaire de Chine, l'*abricot* a commencé sa carrière méditerranéenne en Syrie. Les Grecs l'avaient appelé `armeniakon « [fruit de l']Arménie », parce que cette région était sa provenance immédiate. Pour les Latins, *prunus armeniaca* « prune d'Arménie » ou *persica praecocia* (cf. Corominas, *DCELC*, t. I, p. 84) se nommait aussi *praecoquum* « [fruit] précoce », terme passé dans le grec tardif sous la forme πραικόκιον. C'est bien ce dernier qui fut adopté, via le syriaque بارقوق *barqūqyā* (Payne Smith, col. 620), par les Arabes qui ont cultivé le fruit sous le terme بارقوق *barqūq* lequel s'adapta dans la péninsule Ibérique où le mot *albaricoque* est attesté en 1330 puis, plus au nord et notamment en France, au XVI^e siècle. (*Bull. SELEFA* n° 1, 13-15).

On pourrait donner maints autres exemples : le mot persan شاه *šāh*, « roi », pièce maîtresse du jeu de شترنگ *šatrang*, qui a donné par le canal de l'arabe et du latin médiéval, le vieux français *eschec*, « échec », puis l'anglais *chek* « contrôle », d'où le nom de

l'instrument monétaire appelé en anglais *cheque* et en français *chèque*, se retrouve dans l'arabe شيك *šīk* / شك *šik* et le persan چك *ček* pour dire « chèque ».

* Une variété de mot-navette est ce qu'on pourrait appeler le **mot-miroir**. Il existe, dans la langue des jeunes, des emprunts à la langue arabe très particuliers qui ne sont pas le produit d'un emprunt étranger pour décrire une réalité nouvelle en français mais qui correspondent à l'utilisation de mots arabes par des locuteurs bilingues pour donner du piquant à des expressions idiomatiques du français populaire. Ces termes furent utilisés, au moins au départ, par des jeunes arabophones, non tant dans un but cryptique que pour une sorte de clin d'œil identitaire avec une forte charge ludique. Certains de ces mots sont restés confinés dans les cercles arabophones mais d'autres ont échappé à ce domaine pour circuler dans les canaux plus larges de la langue des jeunes. On peut ainsi donner, à titre d'exemple, *zebda*, dans le sens de *beur* utilisé par effet de calembour par homophonie avec *beurre* et qui constitue probablement le prototype de ce phénomène linguistique (voir *Bull. SELEFA* n° 7, 19-21). On peut aussi relever *haraguer*, dans le sens de « accabler de reproches » (voir *Bull. SELEFA* n° 4, 20), ou, plus rarement *fakht* et *gombbron* pour « pigeon, dupe » (voir *Bull. SELEFA* n° 7, 19-21).

3. L'effet de shunt ou court-circuit : lorsque le parcours étymologique d'un terme est détourné par une explication étymologique apparente exemple de *girofle*.

L'effet de shunt ou court-circuit étymologique est particulièrement frappant dans le domaine des emprunts ou des calques des langues romanes à l'arabe. Il traduit dans le domaine de l'histoire des mots la tentative faite à la Renaissance de renier l'héritage culturel arabe.

Un exemple déjà donné concerne le terme *minute* donné par les dictionnaires comme le latin [*minutum*] *secundum*, ce qui conduit à effacer le fait que cette expression n'existe pas dans les textes latins classiques mais est une manière de rendre l'arabe [الدقيق] الثانية [*al-daqa'iq*] *al-tāniya*, qui est lui-même le calque du grec δεύτερον [*λεπτον*]. Le court-circuit en respecte donc pas l'histoire du mot.

Autre exemple, le terme *métaphysique*, formé sur le grec φυσικά *physiká*, la « nature » et son étude, bref, la « physique » affublé du suffixe μετά *metá*, pour dire « après », comme si le mot avait été pris au grec. En réalité », l'histoire est bien plus complexe (voir *Bull. SELEFA* n° 12, 1-6) : le mot n'existe pas en grec chez Aristote mais apparaît dans le catalogue d'un bibliothécaire de l'époque hellénistique pour classer sous une même rubrique tous les ouvrages ne faisant pas partie chez le maître de la grec φυσικά *physiká*. Pour concevoir toute cette partie de la philosophie comme un concept, il faut attendre Al-Farabī qui écrit précisément une "رسالة في ما بعد الطبيعة" *Risāla fī mā ba'd al-ṭabī'a*, littéralement « Épitre sur "l'Après-physique" ». Les traducteurs latins qui trouvèrent ce terme chez ابن سينا *Ibn Sīnā* / Avicenne et chez ابن رشد *Ibn Rušd* / Averroès qui le met en valeur dans ses trois ouvrages consacrés à la métaphysique d'Aristote, *Jawāmi' kitāb* [...], *Talḥīṣ* [...] et *Šarḥ mā* ..., et n'ayant pas accès aux textes grecs donnant le mot μεταφυσικά forgèrent un mot latin d'étymologie grecque *metaphysica* comme calque de l'arabe *mā ba'd al-ṭabī'a*.

Pour ce qui est des emprunts, nous pouvons donner l'italien *cùpola*, qui a donné le français *coupole*. Les romanisants le font tout naturellement dériver du bas latin *cupula*, « tonnelet, petite cuve », diminutif de *cupa*, « cuve ». Mais il y a de fortes présomptions pour que le sicilien *cùbbula* et l'italien *cùpola* ne soient dérivés de l'arabe قبة *qubba*, « dôme, coupole », avec attraction paronymique du latin *cupula*.

Conclusion

Tous ces éléments montrent qu'il est indispensable d'aborder la recherche étymologique non pas comme la quête de liens simples, directs et immédiats, mais comme l'appréhension d'un ensemble de flux d'influences complexes, parfois contradictoires et croisés, ce qui fournit l'idée d'un réseau. C'est bien ce qui est rendu par l'image qui figure sur la page d'accueil du site de la SELEFA :



Nous avons ici les bras de l'Amazone, multiples, convergents et divergents, dont certains peuvent même présenter un courant qui s'inverse...