



**HAL**  
open science

## Corpus et langue arabe : un changement de paradigme

Catherine Pinon

► **To cite this version:**

Catherine Pinon. Corpus et langue arabe : un changement de paradigme. Dossiers d'HEL, 2017, Analyse et exploitation des données de corpus linguistiques, 11, pp.29-39. hal-01511222

**HAL Id: hal-01511222**

**<https://hal.science/hal-01511222>**

Submitted on 20 Apr 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **CORPUS ET LANGUE ARABE : UN CHANGEMENT DE PARADIGME**

**Catherine PINON**

**IFPO (Institut français du Proche-Orient)**

## Résumé

Nous interrogeons le rapport entre la grammaire arabe et les corpus pour expliquer la place de plus en plus importante que tiennent ces derniers en linguistique arabe. La transmission des connaissances tirées du corpus initial (Coran, poésie archaïque) sur lequel s'est fondée la tradition grammaticale arabe a mené à une longue période de stagnation des grammaires. Avec les outils numériques et le développement de vastes corpus d'arabe moderne ou contemporain, nous assistons maintenant à un bouleversement d'ordre épistémologique qui a des conséquences au niveau de la transmission des savoirs linguistiques et amène à réfléchir à la manière de traiter les données héritées de la tradition grammaticale ancestrale et de constituer une grammaire de l'arabe qui décrive davantage la réalité des usages de la langue.

## Mots clés

grammaire arabe, grammaire réaliste, arabe contemporain, épistémologie, linguistique de corpus, linguistique historique, Internet

## Abstract

We examine the relationship between Arabic grammar and the corpora to explain the increasing importance of corpora in Arabic linguistics. The transmission of knowledge, from the original corpus (Koran, archaic poetry) on which the Arabic grammatical tradition was based led to a long period of stagnation of grammar. With numerical tools and the development of vast corpora of modern or contemporary Arabic, we are now witnessing an epistemological upheaval that has consequences on the transmission of linguistic knowledge and leads us to think about how to analyse the data inherited from the ancestral grammatical tradition and how to constitute a grammar of Arabic which describes the reality of the uses of language in a more accurate way.

## Keywords

Arabic grammar, realistic grammar, contemporary Arabic, epistemology, corpus linguistics, historical linguistics, Internet

Un arabisant qui s'intéresse au développement de la grammaire arabe et qui, par curiosité, se documente sur les processus qui ont amené à l'élaboration des premières grammaires du français, peut se retrouver déconcerté en constatant que le grammairien ou le linguiste arabisant du XXI<sup>e</sup> siècle qui prend pour objet d'étude l'arabe contemporain se retrouve à peu près dans la même situation que celle des grammairiens français du XVI<sup>e</sup> siècle. Ce constat peut paraître étonnant, surtout lorsque l'on sait que la tradition grammaticale arabe remonte au VIII<sup>e</sup> siècle et qu'elle a produit un nombre considérable d'ouvrages de qualité. C'est pourtant bien le cas, comme le montrent les nombreux parallèles qu'il est possible d'établir entre les deux traditions à la lecture de

l'article de Pagani-Naudet (2010) intitulé « l'exemple dans quelques grammaires du XVI<sup>e</sup> siècle ». En effet, lorsqu'elle résume les préoccupations des grammairiens de l'époque concernant la question « comment décrire une langue vernaculaire, instable, soumise à une forte variation ? » et qu'elle estime qu'écrire une grammaire du français à l'époque est un « acte militant » ; quand elle dit de la situation épistémologique au XVI<sup>e</sup> siècle qu'elle paraît inextricable, du fait que « les premiers observateurs sont confrontés massivement à des “données récalcitrantes” », le linguiste arabisant réfléchissant à l'élaboration d'une grammaire de l'arabe standard contemporain<sup>1</sup> ne peut qu'adhérer pleinement à ces propos et les faire siens.

Vouloir élaborer une grammaire de l'arabe contemporain amène effectivement *de facto* à s'interroger sur la langue objet des descriptions. Or, l'arabe standard, utilisé sur le vaste territoire qu'est le monde arabe et même au-delà<sup>2</sup> par des locuteurs de langues maternelles différentes<sup>3</sup>, est soumis à de nombreux facteurs de variations. Plutôt qu'instable, nous préfererions dire qu'il est hautement malléable et qu'il est prompt à évoluer en fonction des besoins et des moyens toujours croissants de communication<sup>4</sup>, comme l'est n'importe quelle langue vivante. Force est pourtant de constater que, d'un point de vue épistémologique, la situation à laquelle les observateurs de l'arabe sont confrontés paraît en effet bien compliquée et parfois insurmontable, d'une part du fait de la confrontation à des données massives très disparates, mais aussi à cause du poids de la tradition grammaticale arabe qui oriente inévitablement les descriptions vers une comparaison avec « l'arabe classique ».

Comprendre pourquoi, au XXI<sup>e</sup> siècle, le linguiste arabisant désireux d'établir une grammaire de l'arabe est confronté à des préoccupations épistémologiques qui ont été celles de grammairiens d'autres langues il y a des siècles, amène à voir comment ces questionnements doivent guider les grammairiens et les linguistes dans leurs réflexions sur l'élaboration d'une grammaire réaliste de l'arabe contemporain. L'objet de cet article est donc de sonder le poids de la tradition grammaticale arabe pour interroger les rapports entre corpus et constitution des savoirs linguistiques, afin de proposer des outils méthodologiquement fiables pour la langue arabe.

---

<sup>1</sup> Par « arabe standard contemporain », nous désignons la variété d'arabe véhiculaire actuellement employée dans le monde arabe : un arabe moderne, assez homogène, où peut bien évidemment figurer du dialecte, mais où la variété vernaculaire n'est pas prédominante.

<sup>2</sup> « L'arabe » (sans plus de précision) est la langue officielle de 22 pays s'étendant sur un vaste territoire allant de la Mauritanie à l'Irak en passant par la Somalie et les États de la péninsule Arabique. Cet arabe standard est aussi utilisé par la presse et par les Arabes établis en dehors du monde arabe. C'est la langue de communication par excellence sur Internet où les frontières nationales sont parfois abolies.

<sup>3</sup> L'arabe est un cas d'école en matière de pluriglossie. On oppose habituellement l'arabe *littéral* aux dialectes, la langue maternelle de tout Arabe. Les dialectes arabes, dont la répartition n'est pas liée aux frontières politiques actuelles, sont habituellement regroupés en familles : arabe maghrébin, nilotique, syro-libanais, péninsulaire et mésopotamien. Le domaine arabe s'étend sur un vaste territoire où se parlent donc de nombreux dialectes, mais aussi d'autres langues sémitiques, africaines ou indo-européennes (entre autres le kurde, l'hébreu, les langues berbères, le peul, le wolof, le somali ainsi que le français et l'anglais). La carte linguistique du monde arabe est très complexe. Elle illustre le nombre des variétés de l'arabe et ses contacts nombreux avec d'autres langues : pour établir une grammaire, la sélection des données du corpus va être aussi cruciale que compliquée.

<sup>4</sup> Par exemple, avec l'arrivée des outils informatiques et électroniques, la question du script de la langue s'est posée. On a donc rapidement assisté au développement d'un système de transcription de l'arabe en caractères alphanumériques latins : les discours transcrits dans un tel système ne doivent pas, à notre sens, être exclus des données langagières observables.

## TRADITION GRAMMATICALE ARABE ET CORPUS

La tradition grammaticale arabe a fourni de véritables sommes sur la langue. Mais de quelle langue s'agit-il au juste ? Qu'ont réellement décrit les grammairiens, qui se sont toujours basés sur des corpus (Ditters 1990) ? Quelle a été l'influence de ces descriptions sur le développement ultérieur de la grammaire arabe ? En fait, il est possible de déceler trois grands moments dans le processus de développement de la grammaire arabe : l'établissement originel de la grammaire arabe, la transmission de ces savoirs ancestraux et la rupture induite récemment par les réflexions autour de la description de l'arabe moderne.

Les premiers grammairiens, à partir du VIII<sup>e</sup> siècle, travaillent sur un corpus constitué principalement de la poésie archaïque, du Coran et du *kalām al-'Arab*, le « parler des Arabes », étant entendu ici qu'Arabes signifie Bédouins. En effet, dans le cadre de la théorie de la corruption de la langue<sup>5</sup>, les Bédouins et les membres de la tribu du prophète ont été, pour des raisons théologiques, élus au rang de détenteurs de l'arabe le plus pur. L'objet des tous premiers grammairiens était de fournir une description complète de la *'arabiyya*, en particulier pour expliquer le système des désinences casuelles. Il ne visait certainement pas à décrire les parlers arabes de l'époque. Pour se faire, les grammairiens se sont appuyés sur un corpus composite et fragmentaire : des pièces poétiques longtemps transmises à l'oral, un texte coranique qui a mis près de 150 ans avant de se fixer par écrit et des discours prétendument recueillis directement auprès des Bédouins. Du point de vue de la linguistique historique, on considère que ces fragments de textes qui nous sont parvenus ressortissent à de l'arabe pré-classique, mais c'est bien à partir de ce corpus que les grammairiens ont peu à peu forgé l'arabe *classique* : celui de la classe des lettrés enseigné aux savants de l'époque (Larcher 2005a, 2005b, 2008).

Les grammaires anciennes ont ensuite elles-mêmes servi de corpus-exemplier pour les grammairiens postérieurs, jusqu'à l'époque moderne, selon un phénomène d'accrétion qui a contribué à construire ce que j'ai nommé une *sui-langue arabe classique*, c'est-à-dire un ensemble de descriptions linguistiques qui s'autonourrit (Pinon 2011)<sup>6</sup>. En effet, les grammairiens ont continué d'étudier le langage dans sa forme écrite à travers le corpus qui avait été formé par les pionniers, faisant de l'objet de la recherche linguistique un corpus clos (Ditters, 1990). Pour Carter (1973, p. 146), alors que

---

<sup>5</sup> Une certaine vision mythologique de l'arabe, qui prédomine encore, a l'habitude d'opposer aux arabes dialectaux l'arabe dit *classique*. Il s'agit d'une vision traditionnelle *idéologique* selon laquelle, à l'origine, il n'y aurait eu qu'un arabe dont le Coran représenterait la perfection inimitable. Cet arabe était appelé *al-'arabiyya al-fuṣḥā*, « l'arabe le plus châtié » ou tout simplement *al-'arabiyya*, « l'arabe » que l'on qualifie souvent de *classique*. Ensuite, du fait des contacts avec les populations non-arabophones, cet arabe se serait « dégradé » et aurait donné naissance aux multiples dialectes. Cette théorie dite de la « corruption de la langue » (*fasād al-luġa*), chère à la tradition arabe, ne tient pas à l'épreuve de la linguistique historique, car bien évidemment ce n'est pas l'arabe classique qui a donné naissance aux dialectes et personne n'a jamais eu comme langue maternelle ni même n'a réellement parlé l'arabe tel qu'il est décrit dans les grammaires, en particulier avec les désinences casuelles (voir notamment Larcher 2008).

<sup>6</sup> Ce qu'avant moi, pour ne citer que lui, Cantarino (1974, p. viii) nommait *ageless Arabic linguistics image*, une image de langue arabe intemporelle créée par les orientalistes à force de reprendre le matériau des grammairiens classiques. On peut en effet considérer que la langue des grammaires est une langue générée par l'amas des descriptions ; elle n'est pas réellement le reflet de ce qu'était la langue arabe dans son état classique (entendu du point de vue linguistique historique, c'est-à-dire dans son état d'évolution entre le pré-classique et le moderne).

Sibawayhi, considéré comme le père fondateur de la grammaire arabe, visait à produire une description complète de l'arabe classique, ses successeurs l'ont transformée en une grammaire normative et prescriptive qui est maintenant vue comme représentative de toute la discipline. Les premiers grammairiens sont donc devenus, malgré eux, la source d'une tradition certes riche mais figée. Quoi qu'il en soit, l'arabe tel qu'il apparaît dans les grammaires est avant tout une construction des grammairiens (nous aurons l'occasion d'illustrer ceci plus bas). Les grammaires de l'arabe classique n'ont *jamais* été élaborées à partir de l'étude des textes des auteurs de la période classique. De fait, il n'y a toujours pas de grammaire décrivant réellement la langue arabe dans son état classique, malgré les nombreuses grammaires dites « d'arabe classique » existantes. Ce que l'on possède, ce sont en fait des grammaires transmettant un savoir élaboré au Moyen-Âge sur un objet non clairement défini (cf. Illustration 1).

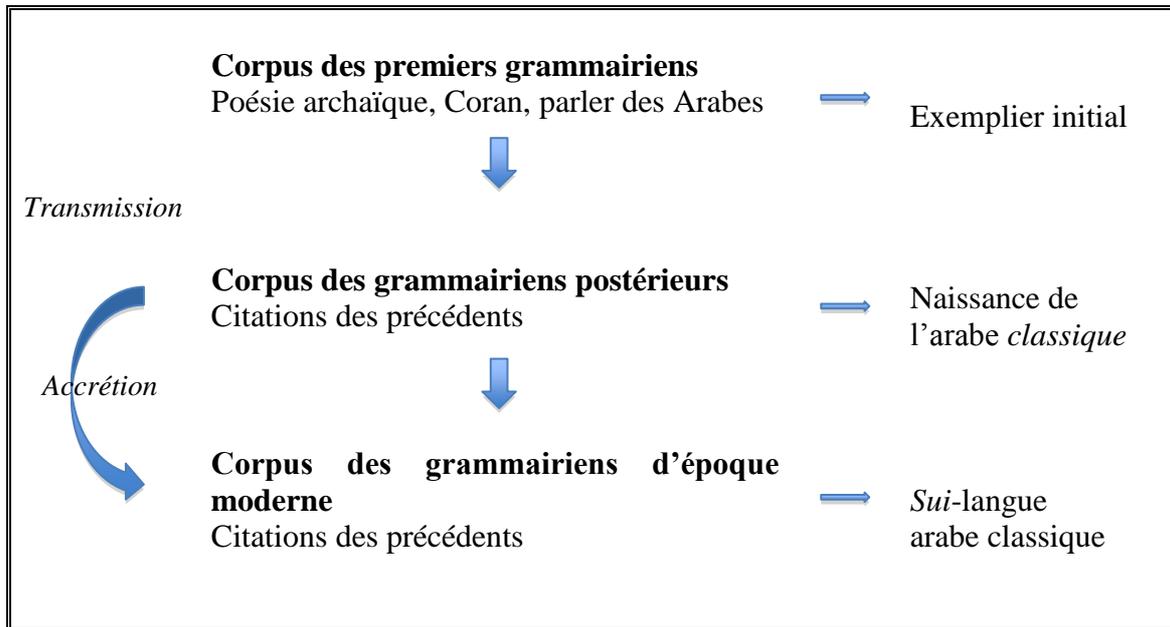


Illustration 1. Le développement de la grammaire arabe

Le processus d'accrétion a suivi son cours jusqu'à maintenant et les savoirs sur la langue arabe se sont transmis et continuent de se transmettre de grammaire en grammaire, bien que l'état de la langue ait évolué. Il existe de nombreuses grammaires de l'arabe moderne qui reprennent exactement les grammaires classiques, parfois en apportant quelques « modernismes » de surface concernant le lexique, mais sans réellement traiter de l'arabe moderne. Force est de constater que, pour ce qui est de l'élaboration des grammaires, nous, les arabisants, sommes très en retard par rapport à d'autres langues. Nous sommes encore maintenus sur le fauteuil de « l'arabe classique » par le poids de la tradition grammaticale et idéologico-mythologique. On peine grandement à s'en lever pour aller s'asseoir sur des bancs plus modernes aux fondements méthodologiques et épistémologiques réfléchis plus qu'hérités.

## ÉVOLUTION RÉCENTE DES SOURCES ET DES MÉTHODES

Néanmoins, depuis une dizaine d'années<sup>7</sup>, des chercheurs ont fait le choix de partir d'un corpus d'arabe moderne pour élaborer leur grammaire. Jusqu'à récemment, la principale source des linguistes était donc l'héritage des grammairiens arabes puis arabisants. Depuis peu, les corpus de textes prennent de plus en plus de place comme matériau indispensable à la constitution des savoirs linguistiques. L'idée est évidemment que le chercheur se confronte à des données réelles pour décrire la langue arabe, qu'il ne se limite plus à retransmettre le contenu des grammaires antérieures. D'ailleurs, en étudiant le système de la langue à partir d'un corpus de textes, les traits ou régularités que le linguiste va pouvoir relever vont souvent entrer en contradiction avec les connaissances transmises par les grammaires classiques. L'étude des corpus de textes apparaît comme complémentaire de la tradition grammaticale arabe et surtout indispensable pour acquérir de nouvelles connaissances sur la langue arabe. On peut assurément dire que c'est l'essor de la linguistique de corpus qui a amené au renouveau des grammaires arabes. Toutes ces réflexions nées autour du corpus ont abouti à une réelle rupture épistémologique, car dès lors que l'on souhaite élaborer un corpus « représentatif » de la langue arabe, on est amené à définir ce que l'on entend par langue arabe et ce que l'on peut inclure comme sources textuelles.

### DES EXEMPLES CONCRETS

Suite à la rupture épistémologique amorcée par la linguistique de corpus dans les études arabes, on en vient à se demander s'il faut déconstruire les grammaires. En effet, la difficulté pour le linguiste à s'affranchir des grammaires précédentes génère différents problèmes d'inadéquation descriptive<sup>8</sup>, comme le maintien de traits inusités, la grande difficulté à intégrer de nouvelles parties de grammaire du fait qu'elles ne figuraient pas dans les ouvrages antérieurs ou encore l'incapacité à s'affranchir d'une typologie qui ne convient plus à la description de l'état moderne de la langue (pour des exemples, voir Pinon 2012).

Deux exemples concrets vont nous être utiles pour illustrer les problématiques liées à ces questions de grammaire, visant à montrer comment le Coran continue d'influencer les descriptions et comment la peur des dialectes amène à sélectionner des traits au détriment d'autres traits tout autant, voire plus usités.

---

<sup>7</sup> Quelques auteurs, comme Cantarino, font figure de pionniers. Dans l'introduction à son ouvrage en trois volumes qui représente l'une des premières références suivant une méthodologie de linguistique de corpus proche de celle que nous connaissons actuellement, Cantarino (1974, p. viii), explique comment la focalisation sur le corpus « primitif » a amené les orientalistes à exagérer le fait que la langue littéraire est restée essentiellement inchangée depuis des siècles. La tendance prononcée à ne pas tenir compte de la personnalité de l'arabe littéraire moderne en tant que phénomène linguistique vivant et à écarter des ouvrages de référence l'usage contemporain (considéré, en cas d'évolution notable, comme une corruption des formes classiques), explique que l'arabe soit enseignée comme une langue médiévale morte. Mais le constat est bien antérieur : il y a plus d'un siècle déjà, Périer (1911) rapportait que le père Louis Cheikho estimait qu'une « grammaire faite d'après les auteurs [...] nous débarrasserait d'une foule d'inexactitudes que les grammairiens colportent depuis des siècles... les grammairiens poussent parfois la subtilité jusqu'au ridicule, et les meilleurs auteurs leur donnent en plus d'un endroit un démenti formel ».

<sup>8</sup> Pour une analyse des interactions entre norme linguistique et norme culturelle, du passage de la norme au mythe dans la civilisation arabo-islamique et de leurs conséquences dans l'élaboration des dictionnaires et des grammaires, cf. Reig 1993.

On constate que certains faits ne figurent pas dans les grammaires d'arabe moderne parce qu'ils n'ont jamais été intégrés dans les grammaires antérieures. C'est le cas d'une question aussi importante en langue que l'expression des modalités. En arabe, les modalités aléthiques classiques du nécessaire et du possible sont prises en charge, entre autres, par le verbe outil *kāna*. *Kāna* (forme accomplie du verbe) est utilisé dans le Coran dans de nombreuses phrases du type « Dieu *est* omniscient et sage » (*kāna -Llāh 'ālim<sup>an</sup> ḥakīm<sup>an</sup>*). Mais en dehors de cet emploi, la nécessité n'est pas marquée par *kāna* qui sert principalement d'exposant temporel du passé ou tout simplement d'outil syntaxique. Pourtant, c'est la seule mention des modalités que l'on trouve dans les grammaires et cette présence est due au simple fait que ce type d'emploi figure dans le Coran<sup>9</sup>. En réalité, l'emploi modal du verbe *kāna* le plus courant sert à marquer le potentiel, sous la forme inaccomplie *yakūn* du verbe. Or, si l'emploi de la forme accomplie de ce dernier pour marquer le nécessaire figure dans les grammaires parce qu'il est très présent dans le Coran, les nombreux emplois de la forme inaccomplie de ce verbe (*yakūn*) pour marquer le possible en sont totalement absents<sup>10</sup>. Nous pouvons suggérer que cette absence est causée, en partie du moins, par le recours à ce même emploi dans les dialectes arabes, selon un processus de rejet voulant que les traits présents dans les dialectes amènent à la sélection, par le grammairien ou le locuteur, d'un trait différent dans la variété standard de la langue. Quoi qu'il en soit, la seule mention des modalités dans les grammaires est due à une phraséologie coranique (donc uniquement attestée dans ce texte), alors que les nombreux emplois du verbe *kāna* pour marquer le potentiel, courants depuis l'époque classique, ne figurent pas dans les grammaires. L'exemple le plus parlant est tiré des grammaires arabes elles-mêmes, où la phrase suivante revient souvent : « le verbe *kāna* peut être bivalent ou peut être monovalent » (*kāna takūn nāqisa wa-takūn tāmma*). Dans cette description de la valence du verbe *kāna*, on utilise le verbe *kāna* lui-même à l'inaccompli (*yakūn*) pour marquer le fait qu'il peut admettre un ou deux arguments (cf. Illustration 2)<sup>11</sup>.

L'expression des modalités du nécessaire et du possible		
En langue		Dans les grammaires
<ul style="list-style-type: none"> <li>arabe coranique <i>kāna</i> = nécessité (emploi rare)</li> </ul>	<i>Poids du Coran</i> 	<i>kāna</i> = nécessité
<ul style="list-style-type: none"> <li>arabe classique / moderne <i>yakūnu</i> = possible (emploi courant)</li> </ul>		

<sup>9</sup> Précisément, les grammairiens arabes anciens n'en parlent pas en termes de modalité, mais d'omnitemporalité, ce que le linguiste peut réinterpréter en termes de nécessité. Par suite, il réinterprétera aussi la forme inaccomplie *yakūnu* en termes de sporadicité (régularité) ou possibilité.

<sup>10</sup> Dans la tradition grammaticale arabe, la transmission de l'exemple a souvent conditionné l'observation elle-même de la langue, en occultant tout un ensemble d'emplois. Par la suite, comme pour les grammaires du français, « la description de la langue s'opère en quelque sorte en vase clos » (Pagani-Naudet 2010, p. 233).

<sup>11</sup> Pinon 2012, 2013.

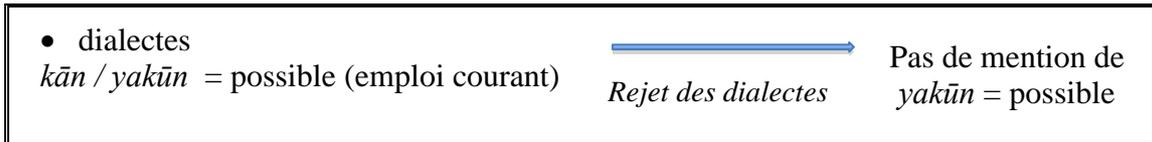


Illustration 2. Influence du corpus coranique dans les grammaires au détriment de la réalité des usages en langue

Pour illustrer cette peur des dialectes et montrer comment les choix opérés par les grammairiens ont influencé le développement de la langue même, on peut citer le cas de la négation du passé. À l'époque classique, on utilisait régulièrement la particule *mā* suivi d'un verbe à l'accompli. Mais comme c'est aussi la forme de négation employée dans les dialectes, on a petit à petit privilégié l'emploi d'une autre particule (*lam*) suivie d'un verbe à l'inaccompli. La négation classique en *mā* + accompli, correspondant à la négation dialectale, a peu à peu disparu des écrits (cf. Illustration 3).

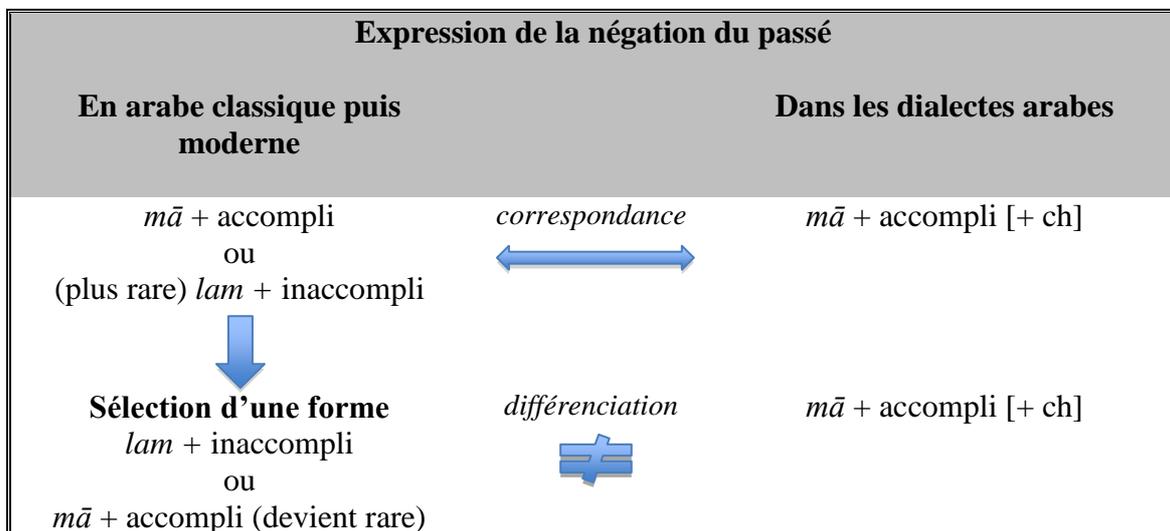


Illustration 3. Sélection d'un trait par réaction à son emploi dans les dialectes

Quand on voit comment les choix opérés par les grammairiens peuvent parfois influencer l'évolution de la langue, on comprend qu'il soit d'autant plus crucial de réfléchir à ces questions fondamentales de choix des sources textuelles pour le linguiste.

Un autre exemple d'inadéquation descriptive réside dans la présence de certains traits inusités ou ayant évolué dans les grammaires, héritage des grammaires antérieures dont le maintien n'apparaît pas justifié à l'épreuve des textes. C'est le cas de valeurs attribuées à certaines particules, valeurs qu'elles possédaient dans l'état classique, voire pré-classique de la langue mais dont sont dénués les emplois modernes. Par exemple, pourquoi continuer de présenter *idā* comme une particule temporelle (« quand ») alors qu'elle n'est plus que conditionnelle en arabe contemporain (« si ») ? Pourquoi refuser d'admettre une catégorie de particules introductrices de questions indirectes en continuant de dire qu'il s'agit de particules du conditionnel ? Pourquoi refuser de parler d'adverbes ou d'adverbiaux en arabe, sous prétexte que cette catégorie n'existe pas chez

les grammairiens qui ont, depuis l'origine de la grammaire arabe, regroupé tous les mots en trois parties du discours (nom, verbe, particule) ?

Une question aussi fondamentale que l'ordre des mots peut aussi constituer un bon exemple des effets pervers de cette transmission des savoirs linguistique par accréation. En effet, on constate que les grammaires actuelles sont beaucoup plus restrictives que les grammaires classiques : de génération en génération, on est passé de grammaires de type encyclopédique à des grammaires scolaires, restreignant les descriptions à des cas généraux ne se référant plus à une quelconque réalité. Depuis des siècles, il semble que l'on assiste à une sorte de défoliation des savoirs linguistiques, comme si ces derniers passaient dans un entonnoir de transmission du savoir réducteur. De l'objectif descriptif et explicatif des premiers grammairiens qui avaient la volonté de décrire toutes les possibilités du système au besoin de restreindre les descriptions à un système facile à enseigner, beaucoup de pages se sont envolées et l'on est passé à des grammaires de type prescriptif de plus en plus détachées de la langue. Pour beaucoup d'usagers, ce processus a amené à confondre les grammaires avec la langue elle-même. Dans les grammaires arabes classiques, tout est possible en matière d'ordre des mots ; dans les grammaires actuelles, tout ou presque est figé de manière prescriptive (l'ordre canonique est : verbe, sujet, complément). Or, l'ordre des mots en arabe contemporain est beaucoup plus varié et aucune grammaire ne propose de liste des ordres des mots attestés ou d'explications linguistiques et pragmatiques de ces différentes possibilités d'ordonner les syntagmes dans la phrase.

#### QUESTIONS DE TYPOLOGIE

Enfin, les questions liées à l'activité taxinomique me semblent fondamentales. L'organisation des grammaires actuelles reste encore largement tributaire des classifications élaborées par les grammairiens classiques. Prenons l'exemple du passif en arabe. Il en existe deux sortes : un passif vocalique apophonique et un passif syntaxique périphrastique. Dans les grammaires classiques, seul le passif vocalique consistant à modifier le schéma vocalique du verbe pour le mettre au passif est mentionné (ainsi, *fu'ila* / *yuf'alu* est la forme passive de *fa'ala* / *yaf'alu*). Il existe pourtant, depuis toujours, un passif périphrastique faisant appel à une tournure du type « verbe auxiliaire (*tamma*) + groupe nominal », ce qui donnerait en français « le trésor a été volé » et « le vol du trésor a eu lieu ». Ce passif périphrastique, qui était absent des grammaires anciennes bien qu'il ait été présent dans l'état classique de la langue, a été intégré dans certaines grammaires récentes. Mais on le trouve alors mentionné dans un chapitre spécialement dédié au verbe auxiliaire *tamma* employé dans cette tournure. Pourquoi ne pas imaginer un chapitre dédié au passif en arabe, où les deux manières d'exprimer les actions passives seraient réunies ? En effet, l'éclatement du passif dans deux chapitres totalement séparés d'une grammaire, outre qu'il pose pédagogiquement et pratiquement problème, laisse à penser que le passif périphrastique est un trait nouveau, récemment apparu et marginal, alors qu'on en trouve des attestations anciennes. Ceci illustre la grande difficulté que nous avons à nous affranchir des catégories élaborées par les grammairiens classiques<sup>12</sup>. Pourtant, il semble évident que la rupture dans la transmission de l'héritage grammatical doit amener à la recherche de nouvelles formes adaptées aux descriptions, donc à de nouvelles typologies potentiellement innovantes par rapport à la classification traditionnelle. Au-delà de la

<sup>12</sup> Nous pourrions prendre d'autres exemples, comme la tripartition classique des mots en verbes, noms et particules évoquée plus haut, qui amène à des regroupements linguistiquement insensés.

simple question du classement des phénomènes linguistiques, intégrer les différents modes d'expression d'un point particulier dans un même chapitre conduirait aussi à réfléchir aux différences sémantiques des uns et des autres et, partant, à enfin intégrer dans les grammaires de l'arabe des explications stylistiques et pragmatiques.

#### VERS UNE MÉTHODOLOGIE RECONSIDÉRÉE

Les corpus ont amené à réfléchir à la langue arabe et aux descriptions qu'on peut en faire. Mais c'est là une première étape qu'il faudrait dépasser en développant la réactivité des linguistes arabisants face aux phénomènes linguistiques de l'arabe contemporain, en encourageant leur capacité à renouveler la grammaire, c'est-à-dire à accepter de ne pas transmettre un fait de langue désuet ou non attesté, et à prendre l'initiative d'intégrer de nouvelles catégories et de nouveaux phénomènes<sup>13</sup>. Interroger les rapports entre corpus et constitution des savoirs linguistiques amène naturellement à réfléchir à des outils méthodologiquement fiables et adaptés à la langue arabe contemporaine.

#### POUR ALLER PLUS LOIN : UN (RE)TOUR VERS LA LINGUISTIQUE HISTORIQUE ?

Il convient maintenant d'aller encore un peu plus loin pour s'interroger sur le changement de paradigme qui s'opère dès lors que l'on a accepté de dépasser les considérations socio-mystico-religieuses qui pèsent sur la langue arabe. On ne considère alors plus l'arabe comme une langue figée dans le temps, mais bien comme une langue vivante ayant évolué d'un état pré-classique à un état classique, puis moderne, dont le contemporain n'est qu'une photographie en synchronie. On peut donc s'autoriser des études de linguistique historique. Pour le linguiste qui souhaite étudier l'évolution d'un trait propre à l'arabe, la situation est encore délicate parce que si ce trait apparaît dans les grammaires classiques, cela ne signifie pas forcément qu'il était présent en arabe classique du fait que ces grammaires ne se sont pas fondées sur des textes représentatifs de l'arabe classique. D'autre part, qu'un fait soit absent des grammaires ne signifie certainement pas qu'il n'existait pas : il peut tout simplement avoir été omis des descriptions précédentes. Nous en avons vu des exemples avec les modalités et le passif périphrastique. Il semble que tout le travail soit à (re)faire maintenant que l'on a admis les bases méthodologiques de la linguistique de corpus. D'une grammaire réaliste des différents états de la langue à une grammaire historique, le pas sera alors naturel car finalement, pour aboutir à une vision complète de l'évolution de l'arabe, il faudrait établir des corpus synchroniques contemporains et philologiques des différentes époques pour comparer et mesurer ensuite les évolutions. Nous pourrions alors imaginer des outils jusque là inexistantes pour la langue arabe, comme un dictionnaire historique et étymologique. Réfléchir à la place et au statut du corpus en linguistique arabe nous a permis d'expliquer comment cette lacune a pu perdurer jusqu'à ce jour. Les réflexions liées au choix du corpus servent aussi de base méthodologique ; elles contribuent à changer radicalement de paradigme et à faire évoluer positivement les connaissances sur la langue arabe.

#### CONCLUSION

On a évoqué les carcans traditionnels, mythologiques, religieux et idéologiques qui font que les études de linguistique arabe sont considérablement en retard par rapport à

---

<sup>13</sup> « Nouveau » compris à la fois dans le sens de « nouvellement introduit dans une grammaire » et dans celui d'une « évolution originale de la langue ».

d'autres langues. C'est le cas de l'élaboration des grammaires en particulier. On a expliqué comment la linguistique de corpus avait opéré un changement de paradigme dans les études linguistiques arabes, en remplaçant la langue au cœur des préoccupations, et non pas les savoirs accumulés sur une *sui*-langue. Travailler sur corpus, c'est d'abord constituer un corpus. Et constituer un corpus, c'est faire des choix cruciaux : chaque choix reflètera le point de vue du chercheur sur la langue arabe et orientera les descriptions<sup>14</sup>. Réfléchir à la manière de construire un corpus tendant à un maximum de représentativité amène à réfléchir à la langue elle-même, dans son infinie richesse et dans toute sa complexité. On est obligé de se confronter aux questions de pluriglossie, de diatopie, de genre, pour voir comment on peut intégrer ces dimensions dans la recherche. La méthodologie du corpus amène *de facto* à s'interroger sur la réalité de la langue, puis sur la finalité d'un travail de description technique de la langue. On brise le vase clos d'une transmission théorique de savoirs accumulés sur la langue, au profit d'un cercle récursif qui part du corpus pour y revenir : le corpus est nécessaire, mais comment le construire, et pour quoi faire ?

Devant le rôle « régénérateur » du corpus en grammaire et linguistique arabe et, plus généralement, devant l'importance des questions épistémologiques dans l'évolution d'un domaine de recherche, on ne peut que déplorer le peu de cas fait de la méthodologie en général, dans les études universitaires, quand on voit que le renouveau de la grammaire arabe est dû à des questions purement méthodologiques et épistémologiques. D'autre part, pour que la linguistique de corpus puisse s'épanouir, il ne faut pas omettre le partage des informations, des corpus, des données : là encore, il reste un grand pas à faire au sein de la communauté des linguistes arabisants.

#### BIBLIOGRAPHIE

- BOHAS, Georges, 1981. « Quelques aspects de l'argumentation et de l'explication chez les grammairiens arabes », *Arabica* XXVIII, 2-3, Leiden, Brill, 204-221.
- CANTARINO, Vicente, 1974. *Syntax of Modern Arabic Prose. Vol. I: The simple sentence*. Bloomington / London, Indiana University Press (Asian Studies Research Institute, Oriental series 4).
- CANTARINO, Vicente, 1975. *Syntax of Modern Arabic Prose. Vol. II: The expanded sentence*. Bloomington / London, Indiana University Press (Asian Studies Research Institute, Oriental series 4).
- CANTARINO, Vicente, 1975. *Syntax of Modern Prose. Vol. III: The compound sentence*. Bloomington / London, Indiana University Press (Asian Studies Research Institute, Oriental series 4).
- DITTERS, Everhard, 1990. "Arabic Corpus linguistics in past and present", in Versteegh, Kees & Carter, Michael G. (ed.), *The History of Arabic Grammar II*, University of Nijmegen, Amsterdam/Philadelphia, Benjamins, 129-141.
- DELESALLE, Simone & MAZIÈRE, Francine, 2002. « La liste dans le développement des grammaires », *Histoire Épistémologie Langage*, 24/1, 65-92.
- LARCHER, Pierre, 2005a. « D'Ibn Fâris à al-Farrâ' ou un retour aux sources sur la *lughâ al-fushâ* », *Asiatische Studien/Etudes asiatiques*, LIX, 3/2005, 797-814.

<sup>14</sup> Reig 1993, notamment sur le « je » du dictionnaire.

- LARCHER, Pierre, 2005b. « Arabe préislamique, arabe coranique, arabe classique : un continuum ? », in Ohlig, Karl-Heinz & Puin, Gerd-Rüdiger (ed.), *Die dunklen Anfänge. Neue Forschungen zur Entstehung und frühen Geschichte des Islam*, Berlin, Verlag Hans Schiler, 248-265.
- LARCHER, Pierre, 2008, « Al-lugha al-fushâ : archéologie d'un concept "idéolinguistique" », *Revue des mondes musulmans et de la Méditerranée* 124, 263-278.
- PAGANI-NAUDET, Cendrine, 2010. « Le statut de l'exemple dans quelques grammaires du XVI<sup>e</sup> siècle », in Cappeau, Paul, Chuquet, Hélène & Valetopoulos, Freiderikos (éd.), *L'exemple et le corpus, quel statut ?*, Rennes, Presses Universitaires de Rennes (Travaux linguistiques du Cerlico 23), 219-234.
- PÉRIER, Auguste, 1911. *Nouvelle grammaire arabe*, Paris, Ernest Leroux.
- PINON, Catherine, 2011. « La grammaire arabe : entre théories linguistiques et applications didactiques », *Synergie-Monde arabe* 7, *Essais de linguistique arabe*, 75-86.
- PINON, Catherine, 2012. « Les enjeux épistémologiques et didactiques d'une grammaire arabe fondée sur corpus », in Arnavielle, Teddy (éd.), *Voyages grammairiens*, Paris, L'Harmattan, 83-101.
- PINON, Catherine, 2013. « Quel corpus peut aider à fonder la grammaire d'une langue pluriglossique ? Exemple de l'arabe contemporain », *Cahiers de Praxématique* 54-55, *Corpus, données, modèles*, Montpellier, Presses Universitaires de Montpellier, 39-58.
- REIG, Daniel, 1993. « Dictionnaire et idéologie dans la culture arabe », *Studia Islamica* 78, 63-97.
- SARFATI, Georges-Elia, 1995. *Dire, agir, définir : dictionnaires et langue ordinaire*, Paris, L'Harmattan.