



Bulletin d'études orientales

Tome LIX | octobre 2010

La métrique arabe au XIII^e siècle après al-Halīl

Traitement automatique de la métrique arabe : réalisations et perspectives

Le traitement automatique de la métrique arabe : réalisations et perspectives

The computerized processing of Arabic metrics : achievements and future prospects

المعالجة الآليّة للعروض العربية: إنجازات وآفاق مستقبلية

Djamel Eddine Kouloughli



Édition électronique

URL : <http://journals.openedition.org/beo/182>

DOI : 10.4000/beo.182

ISBN : 978-2-35159-318-9

ISSN : 2077-4079

Éditeur

Presses de l'Institut français du Proche-Orient

Édition imprimée

Date de publication : 1 octobre 2010

Pagination : 17-32

ISBN : 978-2-35159-170-3

ISSN : 0253-1623

Référence électronique

Djamel Eddine Kouloughli, « Traitement automatique de la métrique arabe : réalisations et perspectives », *Bulletin d'études orientales* [En ligne], Tome LIX | octobre 2010, mis en ligne le 01 octobre 2011, consulté le 19 avril 2019. URL : <http://journals.openedition.org/beo/182> ; DOI : 10.4000/beo.182

TRAITEMENT AUTOMATIQUE DE LA MÉTRIQUE ARABE : *réalisations et perspectives*

Djamel Eddine KOULOUGHLI
(CNRS, UMR 7597)

INTRODUCTION

La réalisation d'applications de l'informatique à la langue arabe dans divers domaines (EAO, dictionnaire, TAO, analyseurs morphologiques, BDD textuelles) n'en est plus à ses débuts, même si l'on peut souvent regretter le manque d'originalité de ces applications et par suite leur adaptation parfois imparfaite aux exigences spécifiques de la langue. Mais, curieusement, il ne semble pas que le champ spécifique de la métrique arabe ait reçu jusqu'ici une attention soutenue de la part des linguistes informaticiens¹.

Les conditions théoriques semblent pourtant réunies puisque, depuis longtemps, des recherches de qualité ont été effectuées concernant la formalisation de la métrique arabe². Le sens de cette dernière remarque doit d'ailleurs être précisé : je ne veux pas laisser entendre par là que l'implémentation informatique découle, en quelque sorte automatiquement, de la disposition d'analyses descriptives et de théorisations adéquates d'un domaine linguistique donné. Je n'ai pas, en particulier, la naïveté de croire que le succès d'une implémentation informatique « prouve » la validité d'une théorisation linguistique dont elle se réclamerait.

A cet égard, je souscris entièrement à l'opinion de Bohas et Paoli (1997, p. 31-32) selon laquelle le succès d'une réalisation informatique ne saurait constituer une forme de validation d'une grammaire formelle dont elle se voudrait l'implémentation. A contre-pied de cette illusion, il faut au contraire souligner la lucidité avec laquelle ces deux auteurs observent en l'occurrence que « si l'on compare deux programmes (...) on s'aperçoit que celui qui « marche le mieux » est celui qui s'éloigne le plus de la procédure recommandée par le métricien » (BOHAS & PAOLI 1997, p. 32).

Le principal lien que je vois en réalité entre description théorisée d'un domaine linguistique et implémentation concerne la disposition d'un relevé exhaustif des données à intégrer dans le traitement automatique et éventuellement d'une structuration optimale de ces données. Pour le reste, il y a fort peu de chances que les exigences d'optimisation

1. Signalons cependant une étude récente, dont nous avons trouvé les références sur le web : EL AYECH, MAHFOUF & ZRIBI 2006, mais dont nous n'avons pas encore pu prendre connaissance.

2. On pense notamment à MALING 1973, BOHAS 1975, ou encore HARKAT 1985.

d'un traitement automatique coïncide avec celles d'une description théorique. Notons d'ailleurs que cet état de choses n'est pas foncièrement différent de celui qui existe entre différentes théories linguistiques tentant de rendre compte du même objet³ : les données seules (et possiblement leur structuration optimale) restent la référence intangible. Le type de théorisation et même son fondement épistémologique peuvent toujours être discutés.

Pour en venir à mon propre travail, il s'inscrivait, au départ, dans une perspective très modeste et ne visait pas directement le domaine de la métrique arabe. Il se rattachait plutôt à l'analyse de corpus : je me demandais, dans cette perspective, s'il était possible de concevoir une procédure algorithmique (et donc programmable sur ordinateur) de découpage syllabique d'un mot arabe quelconque. Je suis parti d'une intuition très simple, à savoir que cette opération peut être réalisée sans aucune ambiguïté par tout arabophone, notamment lorsqu'il réalise la scansion métrique d'un vers (*taqtīr*). C'est par ce biais que mes préoccupations purement « syllabiques » ont rejoint celles, plus vastes, de Georges Bohas : ce dernier, après avoir jeté les bases d'une grammaire formelle de la métrique arabe (BOHAS 1975), avait travaillé avec l'informaticien Yahya Hlal, spécialiste de TALN (traitement automatique des langues naturelles), à une première tentative de traitement automatique de la métrique arabe (BOHAS & HLAL, 1979 et 1982). Il convient de s'arrêter un instant sur cette première tentative.

1. BREF HISTORIQUE : L'IMPLÉMENTATION DE BOHAS ET HLAL (1979, 1982)

Sans trop entrer dans les détails, il faut d'abord rappeler que cette implémentation, élaborée à une époque où les micros-ordinateurs n'existaient encore qu'à l'état embryonnaire, supposait des matériels, des logiciels et des compétences informatiques très lourds qu'il n'était guère question de mettre entre toutes les mains.

Quant aux possibilités des machines de l'époque en matière de saisie des symboles alphanumériques, elles étaient très limitées : les nombres et symboles arithmétiques, l'alphabet latin majuscule et quelques symboles comme le «\$», ce qui supposait un pré-traitement important des données linguistiques à analyser, pré-traitement qui les rendait largement opaques au non-informaticien. Voici, par exemple, à quoi ressemblait la transcription » de la séquence « *alwaladāni dahabā lyawma* » dans le système élaboré par Y. Hlal :

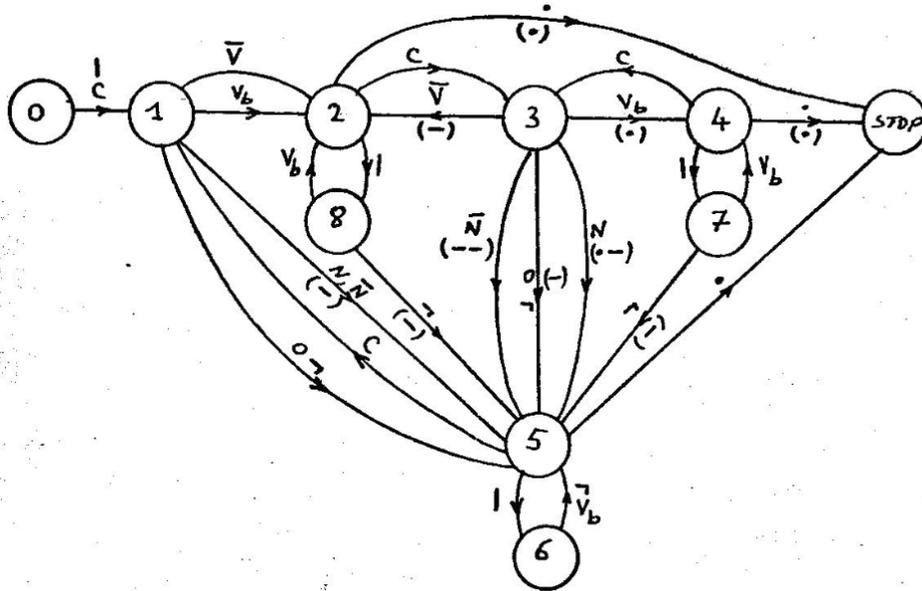
الْوَلَدَانِ ذَهَبَا الْيَوْمَ

ALOWALADANI VAGABAH ALOYAWOMA .
 01235123434751 234343475651512351 2 STOP

- . . - . . . - - .

3. On peut songer ici, tout particulièrement, à la confrontation de la théorie métrique d'al-Ḥalīl b. Aḥmad (VIII^e siècle) avec les théories concurrentes développées de nos jours pour rendre compte du même stock de données, à savoir le corpus de la poésie arabe.

Le formalisme utilisé était celui des «grammaires à nombre fini d'états», relativement complexe et lourd. Voici, à titre d'illustration, à quoi ressemblait le diagramme d'analyse syllabique de l'arabe élaboré par Y. Hlal :



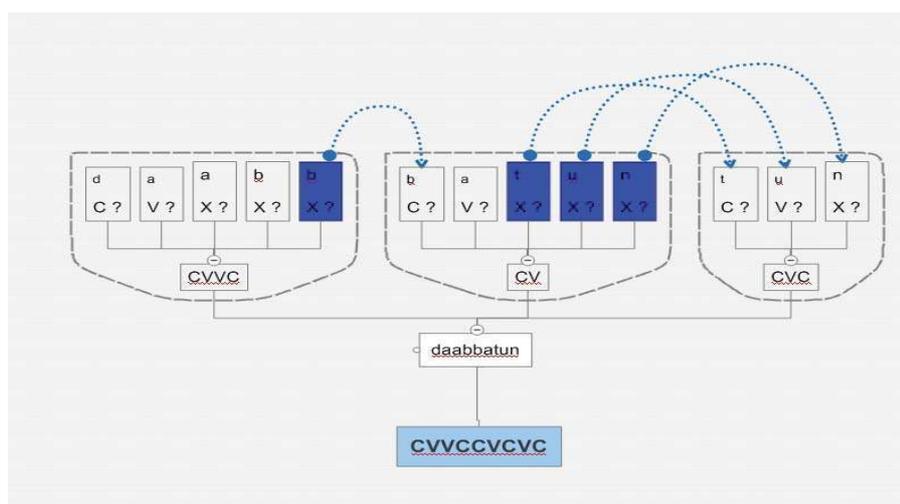
Paradoxalement, on peut ajouter que les «gros systèmes» sur lesquels étaient réalisées de telles implémentations avaient en fait des capacités de traitement assez modestes : je pense qu'il aurait été difficile à l'époque d'envisager de traiter d'un seul coup, et en quelques secondes, un *diwān* poétique entier, comme celui d'al-Aḥṭal, ce que l'on a pu faire d'emblée et sans aucune difficulté sur un micro-ordinateur actuel avec le programme *Xaliyl* que je vais présenter ci-dessous.

Mais avant de passer à cette présentation, et en conclusion de ce bref historique, il faut souligner que cette première implémentation garde un grand intérêt scientifique dans la mesure où elle a permis de démontrer, dès la fin des années soixante-dix du siècle dernier, la faisabilité d'un traitement automatique de la métrique arabe. L'importance de ce résultat sera mieux appréciée si l'on sait que, vingt ans plus tard, en 1999, Michel Bernard pouvait écrire : « Aucun système automatique n'est capable, aujourd'hui, d'accomplir une tâche aussi simple que de repérer, dans un corpus poétique, les différents mètres employés (...) » (Bernard 1999 : 82). Il est vrai que Bernard parlait du français qui, en raison notamment de sa complexité phonographématique, pose des problèmes d'analyse syllabique bien plus complexes que l'arabe.

2. LE PROGRAMME XALIYL

2.1. L'algorithme d'analyse syllabique

Ici encore, sans entrer trop avant dans le détail du fonctionnement de l'analyseur syllabique sous-jacent au programme *Xaliyl*, il nous paraît utile de donner une idée, même sommaire, de l'algorithme qu'il met en œuvre. Cet algorithme repose sur une idée fort simple : pour scanner complètement une syllabe quelconque de l'arabe, il faut disposer d'une « fenêtre de scan » comportant autant de slots qu'il y a de segments dans la syllabe la plus longue possible de la langue plus un. Le slot « surnuméraire » permet de s'assurer que le segment qui suit la syllabe à reconnaître est bien, par sa nature et sa position, étranger à la syllabe en question. Sachant que la syllabe maximale de l'arabe est la « surlongue » de structure CVVC, comme dans *daab-ba-tun*⁴ (une bête), la fenêtre de scan maximum doit comprendre, pour l'analyse syllabique de l'arabe, cinq slots. Voici un diagramme illustrant le mécanisme d'analyse syllabique du mot *daabbatun* :



Dans ce diagramme, on voit qu'après le « stockage » de cinq segments dans la première phase de l'analyse, l'algorithme reconnaît qu'il a bien affaire à une surlongue CVVC plus un segment étranger (ici de couleur foncée) qu'il renvoie à la seconde phase d'analyse. Dans cette seconde phase l'algorithme reconnaît, dans sa fenêtre de cinq segments, qu'il a affaire à une syllabe brève CV et à trois segments étrangers qu'il renvoie à la dernière phase d'analyse. Dans cette dernière phase, la fenêtre de scan se retrécit à trois slots (longueur de la chaîne qui reste à analyser) où l'algorithme reconnaît sans peine une syllabe longue CVC. L'analyse renvoie donc le résultat (correct) : CVVC-CV-CVC pour *daabbatun*.

Ce qui précède vaut pour le découpage syllabique d'un mot arabe quelconque. Mais qu'en est-il, pourrait-on se demander, de l'analyse d'un vers entier ? Pour répondre à cette question, il convient de se souvenir que, pour les métriciens arabes, un vers fonctionne,

4. En notant, à dessein, la voyelle longue « ā » comme une succession de deux brèves, ce qu'elle est du point de vue théorique. La surlongue peut être de structure CVGC (G= glide) comme dans *duwaybbatun* (petite bête).

du point de vue syllabique, un peu comme un « mot géant », commençant à la première syllabe du premier mot de l'hémistiche pour se terminer à la dernière syllabe du dernier mot de celui-ci ⁵. Cela signifie, en pratique, que tous les processus phonologiques qui affectent les mots en liaison (insertion de voyelles épenthétiques, *iltiqā' al-sākinayn*, *sandhi*, etc.) s'appliquent à l'intérieur de l'hémistiche. En d'autres termes, aucun phénomène de pause n'y est admis du point de vue métrique ⁶.

Bien entendu, l'algorithme d'analyse syllabique, dans la mesure où nous voulons qu'il parte d'une représentation graphique normale du texte poétique à analyser ⁷, et non d'une transcription phonétique qui aurait déjà noté tous les phénomènes de liaison, doit faire automatiquement tous les ajustements nécessaires pour donner du texte entré en graphie normale une représentation conforme à la forme phonétique qu'exige son traitement métrique.

D'autre part, il ne faut pas perdre de vue que la graphie arabe usuelle, bien qu'elle soit, pour l'essentiel, assez fidèle à la représentation phonétique des mots, utilise néanmoins quelques conventions purement orthographiques (c'est-à-dire sans contrepartie phonétique) comme l'*alif* du pluriel des verbes, ou des symboles spéciaux combinant valeur phonétique et valeur morphologique, comme la *tā' marbūṭa* ou l'*alif maqṣūra*. Il faut aussi penser à la *šadda* utilisée pour représenter le redoublement d'une consonne, le *sukūn*, utilisé pour représenter une absence de voyelle et enfin à la manière spécifiquement arabe de représenter les voyelles longues comme une brève suivie d'un glide homorganique.

Tous ces phénomènes, mis en lumière par l'analyse phonographématique de la langue générale ou de la langue poétique ⁸, ont dû, bien sûr, être pris en compte dans l'écriture de l'algorithme d'analyse syllabique. Par contre, suivant en cela le fonctionnement de l'orthographe arabe standard, on n'a pas jugé nécessaire de coder l'assimilation de l'article défini par les consonnes corales (*hurūf šamsiyya*), ce point n'ayant aucune incidence métrique. Il serait bien sûr aisé de remédier à ce « défaut » si cela s'avérait utile...

De tout ce qui précède, on peut déduire que l'hémistiche est l'unité effective de traitement de l'algorithme d'analyse syllabique et métrique. C'est la raison pour laquelle on demande que la frontière d'hémistiche soit explicitement déclarée (par une double barre oblique « // ») dans les textes à analyser. C'est en l'occurrence l'unique convention de « précodage » qu'exige le programme. Quant à l'extension de l'analyse au vers, à une pièce de vers, voire à plusieurs poèmes, c'est une simple question d'itération qui se réalise sans aucune intervention spécifique de l'utilisateur, le programme se poursuivant jusqu'à rencontrer la fin du fichier soumis à l'analyse.

5. A cette nuance près, et elle est de taille, que les accents de mots sont partout conservés.

6. D'ailleurs, les règles de pause en fin d'hémistiche sont spécifiques à la langue poétique : en particulier, les voyelles brèves, loin d'être effacées en cette position comme c'est le cas pour la prose, font au contraire l'objet d'une règle spécifique d'allongement.

7. C'est l'un des progrès pratiques les plus significatifs de notre programme par rapport aux implémentations antérieures. Nous précisons un peu plus loin les exigences que l'on est en droit d'avoir vis-à-vis d'un texte saisi en graphie arabe pour qu'il puisse être scanné avec succès par l'analyseur syllabique.

8. La phonographématique est la discipline qui étudie les règles et conventions qui commandent le passage de la forme phonétique d'une langue à sa forme graphique. Sur la phonographématique arabe, voir KOULOGLI 1979.

2.2. Le fonctionnement pratique du programme

Le programme *Xaliyl* est capable de réaliser automatiquement le découpage syllabique et l'analyse métrique de n'importe quel texte poétique arabe, quelle que soit sa longueur (d'un vers à plusieurs centaines). Il réalise cette tâche en ne demandant rien d'autre en entrée que le texte arabe, avec comme seule adjonction systématique une double barre oblique comme séparateur d'hémistiches (ce qui peut être fait de façon quasi-automatique une fois identifié le format du texte ⁹). En principe, le programme pourrait avoir pour input du texte en caractères arabes, mais il faudrait ajouter « à condition qu'il soit systématiquement et rigoureusement vocalisé ». En effet, le programme, s'il rencontre la moindre incohérence syllabique dans la chaîne à analyser, émet un message d'erreur et interrompt immédiatement l'analyse de la chaîne concernée. C'est la raison pour laquelle, en pratique, on utilise en input du programme une transcription « phonographématique » du texte à analyser, sachant qu'entre la transcription et le texte arabe soigneusement saisi, il existe une relation bi-univoque en sorte que l'on peut les relier automatiquement par des règles simples de transcription ¹⁰.

2.2.1. Graphie arabe ou transcription ?

Je sais que l'emploi d'une transcription au lieu de celui des caractères arabes soulève souvent, chez les Arabes, de nombreuses réticences. Il convient donc de consacrer quelques lignes à la justification de ce choix.

Il faut d'abord préciser que, notamment depuis le développement de l'informatique arabisée, il ne s'agit pas d'un choix qui serait dicté par une soi-disant inadéquation inhérente de l'écriture arabe au traitement informatique : tous ceux qui connaissent un peu les mécanismes internes de traitement de l'information linguistique par les ordinateurs savent que, pour ces machines, tous les systèmes d'écriture se ramènent finalement à des codes numériques et qu'il n'y a donc strictement aucune différence de ce point de vue entre l'écriture arabe et n'importe quelle autre écriture, notamment alphabétique.

Le problème ne réside donc pas dans les machines. Il réside dans les hommes, ou plus précisément dans ce que l'on pourrait appeler « les exigences du dialogue homme-machine ». Le fondement du problème est que, dans l'usage usuel de l'écriture arabe, les utilisateurs ont recours à toutes sortes de « simplifications », comme l'utilisation de graphèmes complexes combinant deux unités (*lām-alif*) et l'économie de certains graphèmes (notamment la *šadda* et surtout les voyelles brèves). En principe, dans la saisie soignée d'un texte, comme ce devrait être le cas pour un texte poétique que l'on veut soumettre à une analyse métrique, les éditeurs « font attention » et dotent le texte de tous les indices graphiques nécessaires à une lecture intégrale. Mais, en fait, les habitudes d'économie graphique se révèlent plus fortes que les meilleures résolutions, en sorte que pratiquement aucun texte saisi en graphie arabe n'est totalement exempt de « raccourcis » graphiques

9. Il faut noter que, comme la plupart des programmes de traitement de corpus, *Xaliyl* ne travaille que sur des données sauvegardées au format « texte ».

10. Voir en annexe la table de correspondance du système de transcription utilisé (dit « système TRS »).

qui vont s'avérer autant de causes d'erreur pour l'algorithme d'analyse syllabique. Pour illustrer la réalité de cet état de chose, nous allons présenter un exemple concret.

On trouve sur le web des textes en caractères arabes, dont des textes poétiques. La plupart, même quand ils prétendent présenter un texte « soigné », laissent voir immédiatement les très nombreuses lacunes de saisie du texte. Mais quelques uns semblent vraiment avoir été saisis avec soin et paraissent, à l'œil humain, être de parfaits candidats pour un traitement automatique rigoureux. Voici un extrait, sans modification, d'un texte de ce type, présentant en l'occurrence les premiers vers de la célèbre *Mu'allaqa* d'Imru' al-Qays :

فَقَفَا نَبْكَ مِنْ ذِكْرِي حَبِيبٍ وَمَنْزِلٍ	بَسَقَطِ اللَّوَى بَيْنَ الدُّخُولِ فَحَوْمَلٍ
فَتَوَضَّحَ فَالْمَقْرَةَ لَمْ يَعْفُ رَسْمَهَا	لَمَّا نَسَجَتْهَا مِنْ جَنُوبٍ وَشَمَالٍ
تَرَى بَعَرَ الأَرَامِ فِي عَرَصَاتِهَا	وَفَيْعَانِهَا كَأَنَّهُ حَبٌّ فَلُفْلٍ
كَأَنِّي غَدَاةَ البَيْنِ يَوْمَ تَحَمَّلُوا	لَدَى سَمَرَاتِ الحَيِّ نَاقِفٌ حَنْظَلٍ
وَفُوفًا بِهَا صَحْبِي عَلَيَّ مَطِيئِهِمْ	يَقُولُونَ لَا تَهْلِكِ أَسَى وَتَجْمَلِ

A première vue, ce texte est irréprochable. Pourtant, lorsqu'on le soumet à l'analyseur syllabique, celui-ci y décèle, rien que dans ces cinq premiers vers (sur un total de quarante-deux) une multitude de problèmes, parmi lesquels : au premier vers une *šadda* surnuméraire sur les mots *alliwā* et *aldaqhūli*¹¹ ; au deuxième vers, l'oubli de la voyelle [a] (*fatha*) devant l'*alif* d'allongement de *almiqrāti* et de *rasmuhā*¹² (mais pas pour *nasagathā*, ce qui montre que ces oublis ne font pas système); au troisième vers, oubli de la voyelle brève [a] après la *hamza* de *al'ar'āmi* et après celle de *ka'annah*¹³.

Nous en resterons là pour ne pas lasser le lecteur, mais nous préciserons que l'on pourrait continuer ainsi pour l'ensemble de cet extrait, et plus généralement pour l'ensemble du poème, pour s'apercevoir que pratiquement pas un seul vers n'est exempt de défauts plus ou moins nombreux et importants au regard d'une écriture respectant rigoureusement les exigences de la graphie arabe « stricte » et, par suite, de l'analyse syllabique systématique¹⁴.

Cette question étant, j'espère, éclaircie, je passe au point principal de mon exposé, la présentation du fonctionnement du logiciel *Xaliyl*. Pour ce faire, nous allons repartir du texte de la *Mu'allaqa* d'Imru' al-Qays, mais cette fois en transcription, et examiner ce

11. Ces types de *šadda* sur un *ḥarf šamsī* sont monnaie courante dans la graphie « soignée », mais sont en réalité strictement redondants dans cette position puisque la graphie arabe note explicitement l'article défini phonétiquement assimilé au *ḥarf šamsī*. Pour l'analyseur syllabique, toute *šadda* représentant une gémération de la consonne précédente, on se retrouverait là avec trois consonnes successives, ce qui est absolument impossible en arabe !

12. Oubli quasiment « invisible » pour le lecteur humain qui « sait » que l'*alif* d'allongement suppose une *fatha* !

13. Ici encore le lecteur humain « normal » ne détecte pas la faute.

14. Signalons en passant le cas, très fréquent, de la négation *lā*, qui théoriquement devrait être notée par la séquence *lām-fatha-alif* (d'allongement) mais qui est systématiquement graphiée *lām-alif-fatha*, au point que le système graphique arabe s'est doté du glyphe complexe *lām-alif* en principe inanalysable syllabiquement !

qu'en fait le logiciel. Voici les premiers vers du célèbre poème tels qu'ils sont soumis au programme¹⁵ :

qifaa nabki min @ikraY HabiybiN wamanzili // bisiqTi elliwaY bayna eldaxuwli faHawmali
 fatuwDiHa faelmiqraa#i lam ya&fu rasmuhaa // limaa nasajathaa min januwbiN wa\$am'ali
 taraY ba&ara el'ar'aami fiy &araSaatihaa // waqiy&aanihaa ka'an*ahu Hab*u fulfuli
 ka'an*iy gadaa#a elbayni yawma taHam*aluwe // ladaY samuraa#i elHay*i naaqifu HanZali
 wuquwfaN bihaa SaHbiy &alay*a maTiy*ahum // yaquwluwna laa tahlik 'asaYN watajam*ali
 wa'in*a \$ifaaliy &abra#uN muharaaqa#uN // fahal &inda rasmiN daarisin min mu&aw*ali
 kada'bika min 'um*i elHuwayrici qablahaa // wajaaratihaa 'um*i elrabaabi bima'sali
 'i@aa qaamataa taDaw*a&a elmisku minhumaa // nasiyma elSabaa jaa'at biray*aa elqaranfuli
 fafaaDat dumuw&u el&ayni min*iy Sabaaba#aN // &alaya elnaHri Hat*aY bal*a dam&iya miHmaliy

Le lecteur peut vérifier que le seul codage explicite, en dehors de la transcription, est l'indication de la frontière d'hémistiche. Il peut aussi constater qu'avec un tout petit peu d'habitude la lecture du système de transcription TRS, qui repose sur quelques conventions et analogies simples, n'est pas très déroutante. On observera, concernant cette transcription, qu'elle est strictement équivalente à la graphie arabe usuelle en ce sens, par exemple, qu'elle note « Y » l'*alif maqṣūra*, « * » la *šadda*, « # » la *tā' marbūṭa* ou « N » le *tanwīn*. Elle l'est également en ce qu'elle note, conformément à la graphie arabe usuelle, une *alif* purement orthographique sur le verbe à la 3^{ème} personne du masculin pluriel d'un verbe comme « taHam*aluwe » (fin du premier hémistiche du 4^{ème} vers). Notez cependant que la voyelle longue [ā] est notée « aa » et non « ae » comme le voudrait une stricte observance de l'orthographe arabe¹⁶, et que l'alif du *tanwīn naṣb* n'est pas conservé¹⁷ (par exemple dans « wuquwfaN » premier mot du cinquième vers).

Lorsque *Xaliyl* traite un fichier texte, il génère une base de données textuelle (au format Xbase) associant à chaque ligne du texte analysé un enregistrement composé de neuf champs :

VERS :	Le texte de la ligne analysée, sans modification.
H1 :	Le texte du premier hémistiche (H1).
SYLLH1 :	La séquence syllabique effective de H1.
ANASYLH1 :	La séquence syllabique « symbolique » (en C et V) de H1
METRH1 :	La scansion métrique de H1 en brèves (v, pour ∪) et longues (-, pour -).
H2 :	Le texte du second hémistiche (H2)
SYLLH2 :	La séquence syllabique effective de H2.
ANASYLH2 :	La séquence syllabique « symbolique » de H2
METRH2 :	La scansion métrique de H2.

15. Au démarrage, le programme demande le nom du texte à analyser. Il suffit de lui indiquer alors le nom du fichier texte ; tout le reste se déroule automatiquement.

16. En fait, on pourrait, sans gros problème, adopter ici aussi l'orthographe arabe. C'est uniquement pour des raisons liées à l'histoire du système de transcription TRS qu'il en va différemment.

17. C'est par principe d'économie que cette *alif* n'est pas notée, tout comme le *sukūn*, leur présence étant totalement redondante et pouvant être rétablie automatiquement lors d'un retour à la graphie arabe.

Voici par exemple une copie d'écran de l'enregistrement associé par *Xaliyl* au premier vers du texte ci-dessus :

VERS	qifaa nabki min @ikraY HabiybiN wamanzili // bisiqTi elliwaY bayna eldaxuwli faHawmali
H1	qifaa nabki min @ikraY HabiybiN wamanziliy
SYLLH1	+qi+faa+nab+ki+min+@ik+raa+Ha+biy+bin+wa+man+zi+liy
ANASYLH1	CV CVV CVC CV CVC CVC CVV CV CVC CVC CV CVC CV CVC
METR H1	v - - v - - - v - - v - v -
H2	bisiqTi elliwaY bayna eldaxuwli faHawmaliy
SYLLH2	+bi+siq+Til+li+waa+bay+nal+da+xuw+li+fa+Haw+ma+liy
ANASYLH2	CV CVC CVC CV CVV CVC CVC CV CVC CV CV CVC CV CVC
METR H2	v - - v - - - v - v v - v -

Quelques points méritent d'être signalés à ce niveau : d'abord, on peut voir que si, dans le champ « vers », le programme n'a effectué strictement aucune modification par rapport au texte qui lui a été soumis, il n'en va pas de même pour les champs H1 et H2. L'examen du texte dans ces deux champs montre que le programme y a allongé les voyelles brèves finales d'hémistiche, conformément aux exigences de la métrique arabe classique. Si on examine à présent le résultat du découpage syllabique des hémistiches, présenté dans les champs SYLLH1 et SYLLH2, on constate que le programme a systématiquement « normalisé » l'orthographe (comme le fait le *taqṭī* traditionnel), en interprétant par exemple la séquence « aY » (*faḥa + alif maqṣūra*) comme un [aa] et le « N » du *tanwīn* comme un *nūn*. On peut vérifier aussi qu'il a correctement géré les problèmes de *sandhi*, par exemple dans « siqTi elliwaY » correctement découpé en « siq+Til+li+waa ».

Le même traitement est, bien sûr, appliqué à l'ensemble des vers à analyser. Le processus complet prend moins d'une seconde pour les 82 vers de la *Mu'allaqa* sur un micro-ordinateur actuel de puissance courante. La BDD textuelle générée par *Xaliyl* est automatiquement sauvegardée sur disque pour pouvoir être ensuite diversement exploitée par l'utilisateur.

2.3. Quelques possibilités d'exploitation des résultats

Les types d'exploitation des résultats relèvent des besoins et de l'imagination de chacun. Nous voudrions, dans ce qui suit, en signaler deux qui nous paraissent particulièrement suggestifs.

Tout d'abord, il est extrêmement aisé d'aligner l'ensemble des résultats du traitement pour un champ donné, ce qui permet de faire apparaître, visuellement en quelque sorte, certaines propriétés du texte analysé. Ainsi, si l'on aligne les résultats des champs MetrH1 et MetrH2, on obtient le résultat suivant¹⁸ :

18. Que nous limitons par économie à l'analyse des neuf premiers vers. Le logiciel *Xaliyl* symbolise les syllabes brèves (U) par un v minuscule et les syllabes longues (-) par un tiret -.

METR1	METR2
V - - V - - - V - - V - V -	V - - V - - - V - V V - V -
V - V V - - - V - - V - V -	V - V V - - - V - - V - V -
V - V V - - - V - V V - V -	V - - V - V - V V - V - V -
V - - V - - - V - V V - V -	V - V V - - - V - V V - V -
V - - V - - - V - V V - V -	V - - V - - - V - V V - V -
V - V V - - - V - V V - V -	V - - V - - - V - - V - V -
V - V V - - - V - V V - V -	V - V V - - - V - V V - V -
V - - V - V - V - - V - V -	V - - V - - - V - - V - V -
V - - V - - - V - - V - V -	V - - V - - - V - V V - V -

Cetableau permet de voir immédiatement l'une des propriétés métriques fondamentales de la poésie arabe classique, propriété largement thématifiée dans Bohas (1975)¹⁹, à savoir celle du « parallélisme ». On constate en effet que des successions identiques de séquences syllabiques se reproduisent à intervalles réguliers et que, notamment, une séquence « v- » se répète systématiquement à l'initiale d'hémistiche. Dans les grammaires métriques formelles, c'est typiquement la succession réglée de telles séquences qui permet d'assigner une pièce de vers à un mètre particulier.

Mais l'examen attentif du tableau ci-dessus ne révèle pas que des répétitions régulières de séquences syllabiques. Il permet également de constater qu'en certaines positions, on trouve, selon les vers, des brèves ou des longues. C'est là une autre grande caractéristique métrique de la poésie arabe classique : à côté de positions où c'est toujours la même distribution de quantités syllabiques qui se reproduit (ce sont en quelques sortes les « constantes » d'un mètre donné), on trouve des positions où la quantité syllabique est variable. C'est par exemple le cas, dans les vers analysés ci-dessus, pour la troisième syllabe de l'hémistiche. Cette variabilité au sein du même mètre permet de parler de « variantes de réalisation » et conduit à se demander, entre autres, quelles sont les variantes possibles et éventuellement si, chez un poète donné, certaines variantes sont plus ou moins fréquentes. A ces questions, essentiellement empiriques, la simple théorie ne peut pas répondre. C'est par l'examen et le traitement métrique du plus grand nombre possible de vers du même mètre (du même poète et/ou éventuellement de poètes divers) que l'on peut apporter des éléments d'évaluation objectifs. Concrètement, l'utilisation d'un logiciel comme *Xaliyl* est ici d'un intérêt évident : il suffit en effet de l'alimenter en amont d'une quantité suffisante de données, puis de soumettre, en aval, les résultats obtenus à un traitement statistique pour commencer à avoir des éléments sérieux de réponse aux questions que l'on vient d'évoquer.

En l'occurrence, on peut appliquer aux résultats d'analyse des champs MetrH1 et MetrH2 la procédure de traitement suivante : d'abord, on regroupe l'ensemble des lignes ayant exactement la même distribution syllabique, ce qui peut se faire très simplement en triant « alphabétiquement » les champs MetrH1 et MetrH2. On peut ensuite supprimer tous les doublons (procédure classique en gestion de BDD), ce qui nous indiquera immédiatement combien de variantes de réalisation sont attestées dans la *Mu'allaqa* d'Imru' al-Qays. Les

19. Voir aussi BOHAS & PAOLI (1997, « Préliminaires »).

résultats de ces manipulations nous apprennent que dans cette *Mu'allaqa* de quatre-vingt-deux vers de mètre *ṭawīl*, on dénombre trente-six variantes de vers. La liste de ces variantes est donnée dans l'annexe 2 du présent texte.

Il est aisé d'imaginer qu'en élargissant autant qu'il sera nécessaire l'échantillon des données traitées, on arrivera à avoir une liste exhaustive de toutes les variantes de réalisation pour ce mètre et qu'en traitant statistiquement l'ensemble des données, on pourra dire quelles sont celles de ces variantes qui sont le plus usitées, chez Imru' al-Qays en particulier, ou chez un autre poète, ou encore sur l'ensemble du corpus. De telles informations peuvent s'avérer très précieuses pour l'histoire de la métrique et/ou celle de la littérature...

Concrètement, la disposition d'une liste complète de toutes les variantes possibles d'un modèle de vers permettrait aisément de doter le programme *Xaliyl* de la capacité à identifier quasi instantanément le mètre de n'importe quel vers qui lui serait soumis. D'après une rapide estimation quantitative faite par Bruno Paoli (communication personnelle), le nombre des variantes effectives des quinze modèles abstraits de mètres traditionnellement reconnus par la métrique arabe classique serait un peu supérieur à huit mille. Ce nombre peut paraître énorme, mais pour un ordinateur, reconnaître une séquence parmi huit mille (et même beaucoup plus) est une opération aisée et pratiquement instantanée.

Signalons aussi en passant que le logiciel *Xaliyl* peut aussi fonctionner comme un outil d'aide à l'édition de textes poétiques. Pour donner une idée concrète de cette possibilité, nous évoquerons deux exemples qui n'ont rien d'imaginaire puisque nous les avons rencontrés, Georges Bohas et moi-même, en préparant, il y a quelques années, une communication sur un sujet analogue à celui que nous traitons ici.

Le premier concerne le vers 6 de la *Mu'allaqa* d'Imru' al-Qays. Tel qu'il a été donné ci-dessus, il a une structure métrique irréprochable, mais c'est au prix d'une petite anomalie au niveau de la morphologie : en effet, le dernier mot du premier hémistiche, transcrit comme « muharaaqa#uN » n'est pas une forme régulière de la morphologie de la langue. Il s'agit en effet du participe passif (*ism al-maf'ūl*) du verbe *ahraqa* (variante de *harāqa* = verser) qui devrait donc avoir la forme *muharāqatun* sans voyelle brève après la deuxième consonne. C'est d'abord ainsi qu'il avait été saisi, mais lors du traitement par *Xaliyl*, il est immédiatement apparu que le découpage syllabique qui en découlait (- - v -) était totalement incompatible avec les exigences du mètre qui voulait vv - v -. Nous avons donc opté pour la « licence poétique » en *muharāqatun*, la consultation d'autres éditions du poème ayant confirmé qu'elle était celle que les métriciens avaient adoptée bien avant nous !

Le second problème est plus « sérieux », en ce sens qu'il met en évidence une faiblesse intrinsèque du logiciel *Xaliyl*. Il concerne le traitement du pronom clitique de la troisième personne du masculin singulier qui est, normalement, épélé [-hu] et devrait donc compter comme une syllabe CV. En fait, il peut compter pour une brève ou pour une longue. Plus précisément, il compte systématiquement pour une longue métrique s'il est précédé d'une brève ²⁰, mais peut valoir pour une longue ou une brève s'il est précédé d'une longue. Un exemple de cas où il doit valoir pour une longue est fourni dans le second hémistiche du vers 3 dont voici l'analyse fournie par *Xaliyl* :

20. Et, bien sûr, s'il se trouve en fin d'hémistiche, auquel cas il est systématiquement allongé, comme toute brève dans cette position. Sur ce comportement du pronom clitique de 3^{ème} personne du masculin singulier, voir BOHAS & PAOLI (1997, p. 22).

H2	waqiy&aanihaa ka'an*ahu Hab*u fulfuli
SYLLH2	+wa+qiy+&aa+ni+haa+ka+'an+na+hu+Hab+bu+ful+fu+liy
ANASYLH1	CV CVV CVV CV CVV CV CVC CV CV CVC CV CVC CV CVV
METRH1	v - - v - v - v v - v - v -

On voit que son analyse comme brève (la seule que puisse en l'état fournir le logiciel) entraîne une malformation du premier *watid* du second pied. Le problème est qu'il est difficile, sur la base d'un algorithme purement formel comme celui sur lequel est bâti *Xaliyl*, de différencier le pronom [-*hu*] de la même séquence phonétique quand elle n'a pas ce statut morphologique, comme dans le mot *waǧhu-hā* « son visage », par exemple, auquel cas elle a toujours valeur de brève. Il faut donc corriger manuellement ce type d'erreur, même si leur repérage peut être automatisé ²¹.

Signalons aussi, dans le même ordre d'idées, que le pronom clitique de première personne du singulier, normalement réalisé [-*iy*] après un nom, connaît en métrique une réalisation alternative [-*iya*] exigée par le mètre dans certains cas, comme par exemple dans le vers 9 de la *Mu'allaqa* d'Imru' al-Qays où « *dam&iy* » doit être réalisé « *dam&iya* ». Ici encore, le non respect de cette convention peut être détecté « visuellement » après l'analyse métrique, qui ferait apparaître une séquence métrique irrégulière, mais il ne peut être corrigé que manuellement, au coup par coup...

3. LES PERSPECTIVES

Nous avons évoqué ci-dessus la possibilité d'enrichir le logiciel *Xaliyl* de la capacité à reconnaître automatiquement le mètre de n'importe quel vers ou pièce qui lui serait soumis. Cette possibilité n'est guère compliquée à programmer, mais elle suppose que l'on dispose de la fameuse liste des modèles de vers et de leurs variantes. C'est aux métriciens de faire ce travail et nous leur lançons ici un appel pour qu'ils contribuent à l'avancement de ce projet.

Un autre problème, plus « pratique », doit par ailleurs être résolu. Pour le moment, comme vous avez pu le constater, *Xaliyl* ne peut être exécuté qu'à l'intérieur de l'interpréteur de commandes de son logiciel d'origine le SGBD dBASE. C'est évidemment très contraignant car il faut posséder ce logiciel et apprendre à s'en servir (ce qui est long et coûteux si c'est seulement pour faire de l'analyse de corpus poétiques en arabe !). Il est donc envisagé d'en réaliser une compilation. La compilation consiste à faire traduire *Xaliyl* en langage machine (par un « compilateur »), ce qui le rendra directement exécutable sous Windows par un simple click de souris. Nous espérons pouvoir bientôt faire état de la réalisation de cette étape du développement du logiciel.

21. Ainsi, le repérage automatique a permis de trouver trois occurrences du pronom en question qui doivent être comptées comme des syllabes longues dans la *Mu'allaqa* prise ici comme exemple. Ces trois occurrences concernent les vers 70, 73 au premier hémistiche et 74 au second, et sont particulièrement visibles car elles génèrent une séquence de quatre syllabes brèves successives, séquence totalement proscrite, comme on sait, par la métrique arabe.

ANNEXE 1

Table de conversion du système TRS

TRANSCRIPTION	CARACTERE ARABE	TRANSCRIPTION	CARACTERE ARABE
'	ء	&	ع
I	ئ	g	غ
U	ؤ	f	ف
e	ا	q	ق
Y	ى	k	ك
b	ب	l	ل
t	ت	m	م
#	ة	n	ن
c	ث	h	ه
j	ج	w	و
H	ح	y	ي
x	خ	*	ء
d	د	a	ا
@	ذ	i	ي
r	ر	u	و
z	ز	aa	أ
s	س	iy	ي
\$	ش	uw	و
S	ص	aN	أ
D	ض	iN	ء
T	ط	uN	ء
Z	ظ	aYN	ى

RÉFÉRENCES BIBLIOGRAPHIQUES

- BERNARD, Michel, 1999 : *Introduction aux études littéraires assistées par ordinateur*, Paris, PUF.
- BOHAS, Georges, 1975 : *Métrie arabe classique et moderne*, Thèse de 3ème cycle, Saint-Denis, Université de Paris-8 (n.p.).
- BOHAS, Georges & HLAL, Yahya, 1979 : « Analyse automatique de la métrique arabe », *Analyses théorie* 1, p. 94-123.
- BOHAS, Georges & HLAL, Yahya, 1982 : « Système informatique pour la métrique arabe, reconnaissance du mètre : emploi des deux hémistiches », *Bulletin d'études Orientales* 34, p. 7-31.
- BOHAS, Georges & KOULOUGHLI, Djamel E., 2001: « Towards a systematic corpus analysis of Arabic poetry », *Belgian Journal of Linguistics* 15, p. 103-112.
- BOHAS, Georges & PAOLI, Bruno, 1997 : *Aspects formels de la poésie arabe*, Toulouse, AMAM.
- HARKAT, Mustapha, 1985 : *Le modèle khalilien au centre des théories*, Doctorat d'Etat, Université de Paris 7 (N.P.).
- HLAL, Yahya, 1982 : « Découpage syllabique automatique pour la métrique arabe », *Analyses théorie* 2-3, p. 99-109.
- KOULOUGHLI, Djamel E., 1979 : « Sur la phonographématique arabe », *Analyses théorie* 1, p. 79-151.
- KOULOUGHLI, Djamel E., 2000 : « Sur l'analyse syllabique automatique de l'arabe », *Langues et Littératures du Monde Arabe* 1, p. 29-42.
- EL AYECH, Hafedh, MAHFOUF, Amine, & ZRIBI, Adnane, 2006 : « Reconnaissance de la métrique des Poèmes Arabes par Les Réseaux de neurones Artificiels », dans MERTENS, P., FAIRON, Cédric, DISTER, Anne & WATRIN, P., *Verbum ex machina*, Louvain, Université catholique, p. 462-472.
- MALING, Joan, 1977 : « The Theory of Classical Arabic Metrics », *Al-abhath* 26, p. 29-106.