

Vers un corpus électronique du *Kitāb* de Sībawayhi

Djamel Eddine Kouloughli,
Georges Bohas¹

0. INTRODUCTION

Le *Kitāb* de Sībawayhi (m. vers 796) est, comme on sait, le plus ancien et sans doute le plus important traité de grammaire arabe médiévale qui nous soit parvenu. Considéré comme la source directe ou indirecte de l'essentiel de la tradition grammaticale arabe, cet ouvrage a fait l'objet d'énormément de « lectures » et de conjectures théoriques².

La plupart resteront cependant difficiles à tester empiriquement tant que l'on ne pourra pas véritablement « fouiller » le texte à partir d'index et des concordances systématiques. Partant de cette conviction, nous avons, à partir de l'édition du *Kitāb* faite par Hārūn³, fait saisir le texte sur support électronique, dans l'état dans lequel il a été, jusqu'à présent, édité, c'est-à-dire sous forme essentiellement non vocalisée. Nous avons pu réaliser, à partir du texte brut saisi électroniquement, un corpus étiqueté (utilisant notamment comme balises les noms de chapitres des deux grandes éditions en usage, celle de Derenbourg et celle de Hārūn). Ce travail, en dépit de ses limitations, a déjà permis, en l'état, de procéder à un certain nombre d'exploitations intéressantes⁴.

Cependant, il est évident que seule une édition électronique intégralement vocalisée du texte, édition qui n'a encore jamais été réalisée, permettrait de générer un lexique précis et des concordances non ambiguës, et de fournir ainsi un véritable outil de recherche sur cet ouvrage fondamental. Diverses considérations pratiques et théoriques justifient cette conviction. Citons-en les deux plus importantes :

1. D. E. Kouloughli : CNRS ; G. Bohas : ENS-Lyon.

2. Cf. Carter, 2004, Baalbaki, 2008 et Marogy, 2010, pour ne citer que quelques exemples récents.

3. Rappelons qu'il existe trois éditions du *Kitāb* : la première, en deux volumes, publiée à l'Imprimerie nationale (Paris) par les soins de l'arabisant Hartwig Derenbourg en 1881, la seconde, également en deux volumes, publiée aux Éditions Būlāq (Le Caire) en 1898-1899 et la troisième, en cinq volumes, publiée aux Éditions *Al-Ķāngī* (Le Caire) entre 1966 et 1977 par 'Abd al-Salām Hārūn.

4. Cf. par exemple le n° 5 de la revue *Langues et Littératures du Monde Arabe* (LLMA) consacré au préambule du *Kitāb* de Sībawayhi, et le n° 8 de la même revue où, à l'occasion d'un compte-rendu d'ouvrage, on présente certains résultats de l'exploitation du corpus, déjà intéressants malgré son état encore embryonnaire.

– d’une part il y a le fait que, malgré son statut d’ouvrage fondateur de la grammaire arabe, il n’existe à ce jour aucune édition intégralement vocalisée de cet ouvrage. Ceci n’est pourtant pas un luxe compte tenu de la difficulté du texte et de son importance ;

– d’autre part, la disposition d’un corpus électronique intégralement vocalisé de cet ouvrage devrait permettre d’en produire de vrais index, fiables et complets, et des concordances qui ouvriraient enfin la voie à une exploration systématique et exhaustive.

En ce qui concerne ce dernier point, on sait que l’arabisant Gérard Troupeau a publié en 1976⁵ un « Lexique-Index du *Kitāb* de *Sībawayhi* ». Mais, outre le fait que l’index de Troupeau ne prend pas en compte les « mots outils », ce qui empêche de l’utiliser pour des recherches les concernant, la simple disposition d’une version électronique du *Kitāb*, même partiellement vocalisée, nous a permis de découvrir que, malheureusement, l’index de Troupeau pouvait être grossièrement erroné. C’est incontestablement le cas concernant un fait nullement négligeable, à savoir le nombre de fois que le nom d’*Al-Ḳalīl Ibn ’Aḥmad* est cité dans l’ouvrage de *Sībawayhi*.

Troupeau affirme en effet que *Al-Ḳalīl*, le principal maître de *Sībawayhi*, est cité 608 fois dans le *Kitāb*, chiffre repris, sans contrôle, dans de nombreux travaux, notamment Carter (2004) et Baalbaki (2008). Or, vérification faite, cette évaluation se révèle totalement inexacte. En effet, l’interrogation du corpus électronique permet de vérifier qu’*Al-Ḳalīl* n’est cité nommément qu’un peu plus de 400 fois dans le *Kitāb*⁶.

Cette erreur, compte tenu de sa nature, nous paraît particulièrement grave. Or il est hélas très probable que ce n’est pas la seule et que, lorsqu’un travail systématique aura été fait sur l’indexation du *Kitāb*, on pourra mesurer l’écart immense qui sépare le travail fait à la main par un seul homme et le travail fait par l’ordinateur sur un corpus soigneusement établi et révisé par plusieurs.

1. LES ÉTAPES DU PROJET

Pour pouvoir mener à bien le projet d’une vocalisation complète du *Kitāb*, il fallait pouvoir s’assurer la collaboration directe d’un *Ustād* dépositaire autorisé d’une des traditions de lecture orale intégrale de l’ouvrage, concrètement d’un *Ustād* diplômé de l’Université d’*Al-’Azhar* au Caire. Nous avons trouvé « la perle rare », en la personne de l’*Ustād* *’Aḥmad al-Qādirī*, titulaire de la *’Ālamiyyat Al-’Azhar* et éditeur professionnel de textes classiques, attaché à l’Institut français du Proche-Orient (Damas).

5. Troupeau, 1976.

6. En outre, une vérification systématique des références de Troupeau (qui, heureusement, donne les pages et les lignes de l’édition Derenbourg sur laquelle il a travaillé) a permis de constater que, comme le corpus électronique le laissait craindre, ses décomptes étaient factuellement inexacts.

Le projet a donc démarré en 2009, et a impliqué, au fur et à mesure de sa mise en œuvre, la participation de quatre personnes : Djamel Kouloughli (chercheur CNRS, Paris) et Georges Bohas (professeur à l'ENS Lyon), pour la phase de conception et de démarrage, Aḥmad al-Qādirī (IFPO, Damas), pour la vocalisation du texte et la relecture des textes, et Ahyaf Sinno (professeur à l'Université Saint-Joseph, Beyrouth) pour la relecture globale et la mise en forme du texte.

Nous sommes donc partis d'une saisie électronique du *Kitāb* à partir de l'édition *Hārūn*, la plus répandue aujourd'hui dans le Monde rabe. Elle n'est que très partiellement vocalisée, et pas toujours de façon sûre. On s'est autorisé, bien sûr, à recourir aux deux éditions antérieures, particulièrement celle de Derenbourg, quand cela nous a paru utile. Voici quelques copies d'écran illustrant les étapes du processus de vocalisation du texte. Le texte de départ doté d'un premier balisage en chapitres ressemble à ce qui suit :

(هذا باب ما يختار فيه الرفع إذا ذكرت المصدر الذي يكون علاجاً) | وذلك إذا
كان الآخر هو الأول | وذلك نحو قولك له صوتٌ صوتٌ حسنٌ لأنك إنما أردت
الوصف كأنك قلت له صوتٌ حسنٌ وإنما ذكرت الصوتَ توكيداً ولم تُرد أن تحمله
على الفعل لما كان صفةً وكان الآخر هو الأول كما قلت ما أنت إلا قائمٌ وقاعدٌ
حملت الآخر على أنت لما كان الآخر هو الأول | ومثل ذلك له صوتٌ أيما صوتٍ
وله صوتٌ مثل صوتِ الحمارِ لأنَّ أياً والمثل صفةٌ أبداً | وإذا قلت أيما صوتٍ
فكأنك قلت له صوتٌ حسنٌ جداً وهذا صوتٌ شبيهٌ بذلك | فأَيٌّ ومثلُهما الأولُ |
فالرفعُ في هذا أحسنٌ لأنك ذكرت اسماً يحسن أن يكون هذا الكلامُ منه يحملُ عليه
كقولك هذا رجلٌ مثلكَ وهذا رجلٌ حسنٌ وهذا رجلٌ أيما رجلٍ | وأما له صوتٌ
صوتٌ حمارٍ فقد علمت أن صوتَ حمارٍ ليس الصوتُ الأولُ وإنما جاز لك رفعه
على سعةِ الكلامِ كما جاز لك أن تقول ما أنت إلا سَيْرٌ

Le texte initial

Nous avons alors demandé à l'Ustād Al-Qādirī d'établir, à partir du tapuscrit, une version du texte intégralement vocalisée à la main. Voici la même page vocalisée (et éventuellement corrigée) par l'Ustād :

(هذا باب ما يختار فيه الرفع إذا ذكرت المصدر الذي يكون علاجاً) | وذلك إذا
كان الآخر هو الأول | وذلك نحو قولك له صوتٌ صوتٌ حسنٌ لأنك إنما أردت
الوصف كأنك قلت له صوتٌ حسنٌ وإنما ذكرت الصوتَ توكيداً ولم ترد أن تحمله
على الفعل لما كان صفةً وكان الآخر هو الأول كما قلت ما أنت إلا قائمٌ وقاعدٌ
حملت الآخر على أنت لما كان الآخر هو الأول | ومثل ذلك له صوتٌ أيما صوتٍ
وله صوتٌ مثل صوتِ الحمارِ لأن أياً والمثل صفةٌ أبداً | وإذا قلت أيما صوتٍ
فكأنك قلت له صوتٌ حسنٌ جداً وهذا صوتٌ شبيهٌ بذلك | فأىٌ ومثلُهما الأولُ |
فالرفعُ في هذا أحسنٌ لأنك ذكرت اسماً يحسنُ أن يكون هذا الكلامُ منه يحملُ عليه
كقولك هذا رجلٌ مثلكَ وهذا رجلٌ حسنٌ وهذا رجلٌ أيما رجلٍ | وأما له صوتٌ
صوتٌ حمارٍ فقد علمت أن صوتَ حمارٍ ليس الصوتُ الأولُ وإنما جاز لك رفعه
على سعةِ الكلامِ كما جاز لك أن تقول ما أنت إلا سَيْرٌ

بالصوت

Le texte vocalisé manuellement

Nous avons ensuite réalisé une transcription entièrement automatisée du même texte et y avons inséré les voyelles à partir du texte vocalisé par l'Ustād Al-Qādirī. Nous avons utilisé pour la transcription le système TRS, mis au point par l'un d'entre nous, et utilisé depuis quelques années dans une large variété de travaux de recherche sur l'arabe⁷. Ce système qui n'utilise que les caractères ASCII de base et est donc transportable sur n'importe quelle plateforme, a été établi à partir d'une analyse phonographématique de la graphie arabe. Il peut donc être converti de façon entièrement automatique en écriture arabe. Le travail sur texte transcrit permet de contrôler de façon automatique la correction de la saisie en soumettant le résultat à diverses procédures de vérification et d'édition automatique : vérification de la cohérence syllabique du texte, vérification des hapax, voire vocalisation automatique de toutes les séquences (mots ou segments répétés) susceptibles d'une seule vocalisation. Voici en TRS le contenu de la page précédente :

7. Pour plus de détails sur le système TRS, cf. Kouloughli, 2008 et/ou 2010. Ce système est particulièrement utile pour l'analyse métrique automatique de vers arabes, domaine où l'expérience a révélé que la plupart des textes vocalisés en écriture arabe comportent des oublis et des conventions graphiques qui génèrent des erreurs lors de l'analyse syllabique automatique des vers.

(ha@ae baebu mae yuxtaeru fiyhi elraf&u 'i@ae @akarta elmaSdara el*a@iy yakuwnu &ilaejaN) | wa@alika 'i@ae kaena el'aaxiru huwa el'aw*ala | wa@alika naHwu qawlika lahu SawtuN SawtuN HasanuN li'an*aka 'in*amae 'aradta elwaSfa ka'an*aka qulta lahu SawtuN HasanuN wa'in*amae @akarta elSawta tawkiydaN walam turid 'an taHmilahu &alaY elfi&li lam*ae kaena Sifa#aN wakaena el'aaxaru huwa el'aw*ala kamae qulta mae 'anta 'il*ae qaelimuN waqae&iduN Hamalta el'aaxira &alaY 'anta lam*ae kaena el'aaxiru huwa el'aw*ala.

| wamiclu @alika lahu SawtuN 'ay*umae SawtiN walahu SawtuN miclu Sawti elHimaeri li'an*a 'ay*aN waelmicla Sifa#uN 'abadaN | wa'i@ae qulta 'ay*umae SawtiN faka'an*aka qulta lahu SawtuN HasanuN jid*aN waha@ae SawtuN \$abiyhuN bi@alika | fa'ay*un wamicluN humae el'aw*alu.

| faelraf&u fiy ha@ae 'aHsanu li'an*aka @akarta esmaN yaHsunu 'an yakuwna ha@ae elkalaemu minhu yuHmalu &alayhi kaqawlika ha@ae rajuluN micluka waha@ae rajuluN HasanuN waha@ae rajuluN 'ay*umae rajuliN.

| wa'am*ae lahu SawtuN Sawtu HimaeriN faqad &alimta 'an*a Sawta HimaeriN laysa elSawta el'aw*ala wa'in*amae jaeza laka raf&uhu &alaY sa&a#i elkalaemi kamae jaeza laka 'an taquwla mae 'anta 'il*ae sayruN.

Le texte vocalisé en transcription

Ces vérifications faites, nous retranscrivons automatiquement le texte en écriture arabe pour une dernière relecture, a priori plus facile dans l'écriture originale. Voici à présent le résultat final du processus :

(هذا باب ما يُختارُ فيه الرقعُ إذا ذكرتِ المصدرَ الذي يكونُ عِلْجًا)

| وذلك إذا كان الآخرُ هو الأولُ | وذلك نحو قولك له صوتٌ صوتٌ حسنٌ لأنك إنما أردتِ الموصفَ كأنك قلتَ له صوتٌ حسنٌ وإنما ذكرتِ الصوتَ توكيدًا ولم تُرد أن تحمله على الفعلِ لما كان صيغةً وكان الآخرُ هو الأولُ كما قلتَ ما قلتَ إلا قائمٌ وقاعدٌ حملتِ الآخرَ على 'تنتَ لما كان الآخرُ هو الأولُ | ومثلُ ذلكَ له صوتٌ أيما صوتٌ وله صوتٌ مثلُ صوتِ الجمارِ لأنَّ أيًا والمثلُ صيغةٌ أبداً | وإذا قلتَ أيما صوتٌ فكأنك قلتَ له صوتٌ حسنٌ جدًا وهذا صوتٌ شبيهةٌ بذلكَ أيٌّ ومثلُهما الأولُ فالرقعُ في هذا 'حسنٌ لأنك ذكرتِ اسمًا يحسنُ 'ن يكونُ هذا الكلامُ منه يُحملُ عليه كقولك هذا رجلٌ حسنٌ وهذا رجلٌ 'يما رجلٌ وأما له صوتٌ صوتٌ جمارٍ فقد علمتَ أن صوتَ جمارٍ ليس بالصوتِ الأولِ وإنما جازَ لك رقعهُ على سعةِ الكلامِ كما جازَ لك أن تقولَ ما أنتَ إذا سترَ

Le texte vocalisé en caractères arabes

À ce stade, le texte peut être géré par un logiciel de traitement de corpus et l'on peut commencer à l'exploiter de diverses manières. Voici, à titre d'exemple, le début de la concordance du mot « الخليل », qui nous a permis d'établir le nombre réel d'occurrences de ce nom dans le *Kitāb*⁸ :

8. Noter que la couleur bleue du mot pivot de la concordance est rajoutée par le logiciel d'interrogation du corpus, mais peut être supprimée à l'édition.

لَكَ مَا أَحْسَنَ عَبْدَ اللَّهِ . زَعَمَ الْخَلِيلُ أَنَّه بِمَثَرَلَةِ قَوْلِكَ شَيْءٌ أَحْسَنَ
 وَفِي مَعْنَى النَّصَبِ . وَهَذَا قَوْلُ الْخَلِيلِ رَحِمَهُ اللَّهُ . وَإِذَا قُلْتَ
 زَيْدًا لَقَيْتَ أَخَاهُ . وَهُوَ قَوْلُ الْخَلِيلِ . وَمِثْلُ ذَلِكَ مَا أَذْرِي أَزِيدًا
 الْأَوَّلُ مَعْنَى الْأَمَّاكِينِ . وَزَعَمَ الْخَلِيلُ رَحِمَهُ اللَّهُ أَنَّهُمْ يَقُولُونَ مُطِرْنَا
 الْحِجْحَ إِِنْ شَاءَ اللَّهُ . . . وَقَالَ الْخَلِيلُ هُوَ كَأَنَّ أُخْبِكَ عَلَى الْاسْتِخْفَافِ
 أَحَ وَالْحَسَنَ وَحَوَّ الْأَخَ . وَكَانَ الْخَلِيلُ يَرَاهُ . وَإِنْ شِئْتَ قُلْتَ هَذَا ضَرْبُ
 خَرِّ فَقَالَ اتَّقِ الْمِرَاءَ . وَقَالَ الْخَلِيلُ لَوْ أَنَّ رَجُلًا قَالَ إِيَّاكَ تَفْسِكَ
 . . . وَحَدَّثَنِي مَنْ لَأ أَنَّهُمْ عَنِ الْخَلِيلِ أَنَّهُ سَمِعَ أَعْرَابِيًّا يَقُولُ إِذَا
 أَمَرَ وَتَدَخَّلَهُ فِي آخَرَ . وَقَالَ الْخَلِيلُ كَأَنَّهُ عَلَى ذَلِكَ الْمَعْنَى كَأَنَّكَ
 ذَكَرُ . وَمِثْلُ ذَلِكَ فِيمَا زَعَمَ الْخَلِيلُ : // إِذَا تَعَنَّيَ الْحَمَامُ الْوَرَقُ
 رَبَّتْ عَنْهَا أُمُّ عَمَّارٍ // قَالَ الْخَلِيلُ رَحِمَهُ اللَّهُ لَمَّا قَالَ هَبَّجَنِي
 مَّارٍ . وَمِثْلُ ذَلِكَ أَيْضًا قَوْلُ الْخَلِيلِ رَحِمَهُ اللَّهُ وَهُوَ قَوْلُ أَبِي عَمْرٍو
 أَيْ مِنَ الْفَلْظِ بِالْفِعْلِ . وَزَعَمَ الْخَلِيلُ رَحِمَهُ اللَّهُ أَنَّهُ سَمِعَ بَعْضَ الْعَرَبِ

Début de la concordance de « الخليل »

Ajoutons que la version préparée pour l'exploitation électronique du texte comporte des « balises », c'est-à-dire des marques spéciales associées au début et à la fin de segments du texte, donc l'identification automatique présente potentiellement un intérêt pour l'exploration du texte. En l'état actuel du corpus, il y a des balises pour les vers témoins (šawāhid), constituées par deux barres obliques, des balises pour les versets coraniques et, bien sûr, des balises pour les chapitres ('abwāb) de l'ouvrage. On peut ainsi repérer systématiquement tous les segments de texte balisés et éventuellement les extraire pour en faire une étude spécifique. Voici par exemple ce que donne l'extraction automatique des vers témoins du Kitāb :

//قواطع مكة من ورق الحبي //
 //كنزاح ريش خمامة تجديفة % ومسحلت بالفقتن عصفت الإتمد //
 //دار لسعدى إذ من هواكا //
 //مطرت بمصلي في بتملات % دوامي الأيد يخطن السرخا //
 //فلسنت بآبيه ولا استطيعه % ولك استغني إن كان ماؤك ذا فضل //
 //فإن ياك غما أو سميئا فإني % ساجحل عتبه لنفسه مغمما //
 //وأخو العوان متى يشأ يصرمته % وتعدن أعداء نعتد ودام //
 //تشي يداها الحصى في كل هاجرة % نفي الدناير ثقاد الصباريف //
 //مهلا أعادل فة جرئت من خلفي % آبي أجود لأقوام وإن صنوا //
 //له رجل كالمه صنوت حاد % إذا طلب الوسيقة أو زمير //
 //وأيقن أن المثل إن تلقين به % يكن لمسيل القمل بعهده آبر //
 //أو مغبر الطهر يني عن وليه % ما حج ربه في الدنيا ولا احتمرا //
 //وما له من مجد تليد وما له % من الريح حظ لا الجيوب ولا الصنا //
 //نباة في دار صديق فة أقام بها % حيا نعلنا وما نعلنه //
 //صدهت فأطولت الصندود وقلنا % وصال على طول الصندود يدونه //
 //ولا تطلق الفسشاء من كان منهم % إذا حنسوا منا ولا من سواننا //
 //وما قصدت من أهلها لسوانكا //

Les premiers vers témoins du Kitāb

9. Noter que le surlignage jaune et le soulignement des exemples sont ajoutés par le logiciel d'interrogation du corpus, mais peuvent être supprimés à l'édition.

2. ÉTAT ACTUEL DU PROJET ET PERSPECTIVES

En cette fin d'année 2011, la partie du *Kitāb* consacrée à la syntaxe (*naḥw*), partie qui correspond au tome I des éditions Derenbourg et Būlāq, a été intégralement traitée selon la procédure qui vient d'être présentée. Il reste à en faire une relecture globale portant notamment l'attention sur la ponctuation et le découpage en paragraphes, éléments importants pour la lisibilité du texte.

Une fois complètement achevée la mise au point du texte, pourra commencer son exploitation électronique, et notamment la production d'un index intégral des mots, et celle d'index spécifiques pour les vers témoins, les versets coraniques, voire les exemples en prose.

La seconde partie du *Kitāb*, consacrée aux processus morphophonologiques (*ṣarf*), a été entièrement vocalisée manuellement. Il reste à saisir cette vocalisation sur support électronique puis à la vérifier et enfin à en mettre au point la structuration textuelle.

Une édition papier du texte du *Kitāb* intégralement vocalisé est envisagée, compte tenu de la place éminente que cet ouvrage occupe dans la culture arabe en général et dans l'histoire de la grammaire arabe en particulier.

On envisage également la mise au point d'une version électronique du texte pouvant être affichée sous forme de pages Web et qui pourrait être consultable et interrogeable en ligne à l'instar des textes grammaticaux du *Corpus Grammaticorum Latinorum*. Comme celui-ci, cette publication électronique pourrait se faire sur le site du Laboratoire d'Histoire des Théories Linguistiques (UMR 7597 du CNRS).

Références

- BAALBAKI R., 2008m *The Legacy of the Kitāb*, Leyde, Brill.
- CARTER M., 2004, *Sībawayhi*, Londres, Tauris.
- KOULOUGHLI D.-E., 2008, « Initiation pratique à la constitution et à l'exploitation de corpus électroniques en langue arabe (III^e partie) », *LLMA* n° 7, p. 75-93.
- 2009, Compte rendu de *The Legacy of the Kitāb: Sībawayhi's Analytical Methods within the Context of the Arabic Grammatical Tradition* de Ramzi Baalbaki, *LLMA* n° 8, p. 135-147.
- 2010, « Traitement automatique de la métrique arabe : réalisations et perspectives », *Bulletin d'études orientales*, vol. LIX, p. 17-31.
- CARTER M., BOHAS G., DIAB-DURANTON S., GUILLAUME J.-P., KOULOUGHLI D.-E., 2004, « Autour du préambule du *Kitāb* de Sībawayhi », *LLMA* n° 4.
- MAROGY A., 2010, *Kitāb Sībawayhi : Syntax and Pragmatics*, Leyde, Brill.
- TROUPEAU G., 1976, *Lexique-Index du Kitāb de Sībawayhi*, Paris, Klincksieck.