

DÉTECTION AUTOMATIQUE DE RÉMINISCENCES POTENTIELLES DE XÉNOPHON CHEZ ARRIEN

1. Introduction

Dans le domaine des études d'intertextualité, les hellénistes et latinistes s'intéressent aux citations depuis des siècles. En fait, les Anciens étaient déjà conscients de l'importance de l'imitation et de l'émulation littéraires dans leur propre processus d'écriture, et cela dans les deux civilisations et régimes linguistiques de l'Antiquité classique. Des générations entières de philologues se sont alors ingéniées à relever et à répertorier les citations d'auteurs grecs et latins chez des auteurs, grecs et latins également, d'époques postérieures¹. Un développement assez récent dans les recherches intertextuelles en philologie classique se focalise davantage sur les modalités et les mécanismes de la citation². C'est une piste de recherche qui continuera sans doute à porter ses fruits dans les années à venir.

En revanche, les citations non signalées comme telles et les réminiscences plus subtiles sont bien plus difficiles à déceler. Certes, la critique littéraire les mentionne de façon sélective et assez souvent avec une marge de subjectivité, mais un répertoire complet pour un corpus donné – dans la mesure de ce qui est humainement possible – fait cruellement défaut.

1. Citons un exemple qui dépasse les limites linguistiques : G. N. KNAUER, *Die Aeneis und Homer. Studien zur poetischen Technik Vergils mit Listen der Homerzitate in der Aeneis*, Göttingen, Vandenhoeck & Ruprecht, 1964.

2. Voir deux ouvrages collectifs émanant de cette « école » : C. DARBO-PESCHANSKI (dir.), *La Citation dans l'Antiquité : actes du colloque du PARSIA, Lyon, ENS LSH, 6-8 novembre 2002*, Grenoble, Jérôme Millon, 2004 et Chr. NICOLAS (dir.), « *Hôs ephat'* », « *dixerit quispiam* », « *comme disait l'autre* » : *mécanismes de la citation et de la mention dans les langues de l'Antiquité*, Grenoble, Université Stendhal-Grenoble III, 2006.

2. Naissance et développement du projet³

C'est face à ce constat qu'il nous a paru important de mettre en place un projet de recherche visant, dans un groupe d'auteurs choisis, à déceler, à répertorier, à classer et à commenter le nombre le plus élevé qui soit de réminiscences, allusions, échos, réutilisations, influences et autres types de réécriture non avouée⁴. Ces procédés peuvent porter sur un ou plusieurs mots ou encore sur la structure d'ensemble ou d'une grande partie d'une œuvre. On peut penser également aux motifs et thèmes⁵.

Cela dit, pour éviter que la quantité de données ne soit trop importante et difficilement exploitable, on se limite aux allusions sans mention du rapport intertextuel, en excluant les références nominatives ou les citations « avouées » ou signalées. Les recherches sont par ailleurs centrées sur une étude de cas – Xénophon chez Arrien – au sein d'un corpus bipartite constitué, d'un côté, des historiens⁶ de langue grecque de l'époque impériale⁷ chez lesquels on étudiera les processus de réutilisation et, de l'autre, d'un ensemble de textes antérieurs issus prioritairement mais pas exclusivement de l'historiographie des époques classique et hellénistique dont se sont beaucoup inspirés les auteurs du premier groupe.

Devant l'ampleur du corpus et la complexité du sujet, il fut donc convenu de nous attacher à l'étude des œuvres d'Arrien de Nicomédie, auteur si caractéristique de l'historiographie impériale et si manifestement tributaire des sources antérieures, en comparaison avec celles de Xénophon qu'il admirait. Cet ensemble de textes assurément porteurs de résultats nous a permis de tester et de préciser notre méthode de recherche qui s'articule en deux volets. D'une part, il s'est agi de parcourir l'ensemble de la littérature critique, afin de réunir les réutilisations déjà pointées par les philologues et de les enrichir d'un commentaire pointu. D'autre part, un logiciel spécifique a été développé à l'Université de Liège qui permet de confronter deux textes

3. Le projet est soutenu par un crédit de démarrage accordé en 2010 par les Fonds Spéciaux pour la Recherche de l'Université de Liège.

4. Les diverses formes d'intertextualité peuvent effectivement revêtir des statuts différents, et la distinction entre réminiscence, allusion, écho, réutilisation, réécriture, influence, référence, citation, mention, fragment, renvoi, etc. est souvent difficile à faire.

5. Voir J. ALAUX, « Usages tragiques d'Homère », dans Chr. NICOLAS (dir.), *op. cit.* (n. 2), p. 33-49.

6. En excluant dès le départ les géographes et biographes qui s'éloignent du genre historique bien qu'ils en possèdent certaines caractéristiques, mais en y incluant les histoires ecclésiastiques (Eusèbe, Socrate, mais aussi les *Actes des Apôtres* ...).

7. Les historiens de l'époque impériale pouvaient puiser dans un fonds de littérature beaucoup plus important que l'historiographie de l'époque classique ou hellénistique (qui, elle, est d'ailleurs souvent trop fragmentaire pour faire l'objet de recherches automatisées).

grecs pour mettre en évidence les passages similaires, potentielles réminiscences d'un auteur antérieur dans l'œuvre d'un second, postérieur.

Pour ce faire, un descriptif technique fut d'abord rédigé avec le concours de Pascale Renders (ULg-ATILF)⁸ afin qu'un informaticien puisse mettre au point un logiciel répondant aux exigences du projet. C'est ensuite sous la direction du Professeur Pierre Wolper (Institut Montefiore - Département d'électricité, électronique et informatique) et notamment avec la collaboration du professeur Gérard Purnelle (Directeur du CIPL - Centre Informatique de Philosophie et Lettres⁹ et Directeur Adjoint du LASLA - Laboratoire d'Analyse Statistique des Langues Anciennes¹⁰) que Stephan Guralnik, alors étudiant en dernière année de master en informatique, a élaboré le programme baptisé *ReminiFinder* – sur lequel nous reviendrons en détail ci-dessous – dans le cadre de son Travail de Fin d'Études.

Une fois ce TFE achevé, l'attaché scientifique Marc Vandersmissen entreprit parallèlement de lemmatiser les œuvres d'Arrien et de Xénophon et d'apporter une série de suggestions d'améliorations au logiciel. Il rassembla également, sans atteindre l'exhaustivité, le plus grand nombre de réminiscences potentielles déjà identifiées comme telles dans la littérature scientifique¹¹ : l'objectif ainsi poursuivi était de constituer une base de données des réminiscences déjà repérées par les philologues, afin de l'enrichir ensuite des résultats obtenus par *ReminiFinder* (voir section 4). Le répertoire ainsi constitué contribue à améliorer notre compréhension du système de reprise textuelle dans l'historiographie impériale. Nous sommes également en mesure de comparer les données recueillies manuellement avec les résultats du logiciel pour évaluer ses performances.

8. Voir <http://www.atilf.fr/>.

9. Voir <http://www.cipl.ulg.ac.be/>.

10. Voir <http://www.cipl.ulg.ac.be/Lasla/>.

11. Citons, à titre d'exemples, les ouvrages modernes qui ont révélé le plus grand nombre de réminiscences de Xénophon chez Arrien : C. RENZ, *Arrianus quatenus Xenophontis imitator sit*, Rostochii, C. Boldt, 1879 ; H. R. GRUNDMANN, *Quid in elocutione Arriani Herodoto debeatur*, Berlin, Calvary, 1884 ; Arrianus, *Quae exstant omnia*, I : *Alexandri Anabasin continens*, éd. A. G. ROOS, Leipzig, Teubner, 1967 ; Ph. A. STADTER, *Arrian of Nicomedia*, Chapel Hill, University of North Carolina Press, 1980 ; H. TONNET, *Recherches sur Arrien, sa personnalité et ses écrits atticistes*, Amsterdam, Hakkert, 1988, 2 vols ; A. B. BOSWORTH, *Commentary on Arrian's History of Alexander*, Oxford, Clarendon Press, 1980-1995, 2 vols.

3. Le logiciel *ReminiFinder*

A. Description générale

ReminiFinder est un logiciel – développé dans un environnement Java – qui compare un texte dont on veut étudier les éventuelles réminiscences avec une ou plusieurs autres œuvres d’un même auteur ou non (Annexe 1). Ces textes doivent avoir été préalablement lemmatisés : durant ce processus, les mots déjà connus dans les bases de données du LASLA sont attachés au(x) lemme(s) et catégorie(s) grammaticale(s) correspondants tandis que les formes non reconnues sont extraites du fichier en vue d’être liées à leur analyse manuellement¹² par le chercheur. Ces nouvelles données sont ensuite intégrées aux précédentes pour constituer un fichier cohérent et exploitable par le logiciel. L’ensemble du corpus Arrien - Xénophon a été traité avec ce procédé.

Sur la base de ces fichiers, le programme repère entre les deux entités comparées des tuples communs¹³, c’est-à-dire des ensembles de n formes de la première entité qui peuvent être rapprochés d’un autre ensemble de n formes de la seconde entité parce qu’ils sont attachés précisément aux mêmes analyses (lemme et catégorie grammaticale), en se limitant aux formes cooccurrentes dans un contexte dont l’étendue est préalablement délimitée (Annexe 2). Un tuple commun de niveau deux ($n = 2$) signale donc l’existence d’un couple de lemmes communs entre les deux entités – un tuple de niveau trois, celle d’un triplet, etc. – et le niveau d’un tuple définit donc également le niveau de la réminiscence potentielle ainsi mise à jour. Les résultats peuvent ensuite être classés selon leur pertinence¹⁴ (Annexe 3) et les textes grecs des deux auteurs sont alors présentés, par niveau de réminiscence¹⁵ (Annexe 4).

Il s’est néanmoins avéré qu’avec ces seuls critères, un nombre impressionnant et difficilement exploitable de résultats était généré par la machine. Plusieurs mécanismes ont alors été mis en place pour tenter de réduire la quantité de résultats non significatifs. Le premier intervient dans le processus même de recherche des tuples puisque le logiciel ne conserve que

12. Nous remercions tout particulièrement notre collègue, Paul Pietquin (Ulg - Université de Namur), qui nous a fourni un logiciel permettant de visualiser très rapidement tous les lemmes possibles d’une forme.

13. La distance maximale entre deux mots peut être adaptée par l’utilisateur selon sa recherche.

14. Quatre critères définissent cette pertinence : la fréquence des lemmes, la distance entre les lemmes, la différence des distances entre les lemmes, l’ordre similaire des lemmes.

15. Il est évidemment possible d’exporter les résultats dans un document Word selon certains critères variables mais il ne nous a pas paru pertinent d’en donner ici le détail.

les lemmes en dessous d'une certaine fréquence dans le texte (variable fixée par défaut à 2 % – Annexe 2). Ensuite, il est prévu dans l'interface de pouvoir sélectionner les catégories grammaticales intéressantes pour l'utilisateur. Dans notre projet, nous travaillons essentiellement avec quatre de ces catégories qui nous semblent les plus porteuses de sens : les substantifs, les adjectifs, les verbes et les adverbes. Toutefois, certains mots « outils » parasitaient encore les résultats et il fut décidé de créer une liste de lemmes exclus lors du parcours des fichiers par le programme (Annexe 5). Citons entre autres les verbes εἶμι ou λέγω, ou encore la négation, qui ne sont plus pris en compte lors de la recherche de tuples.

Cependant, malgré ces dispositions, le nombre de résultats aux niveaux inférieurs des réminiscences reste encore trop important, ce qui rend jusqu'à présent leur utilisation difficile. Par exemple, lorsque l'*Anabase* d'Arrien est comparée avec celle de Xénophon, 15 réminiscences potentielles de niveau cinq sont détectées tandis que 52242 le sont au niveau deux (pour les *Cynégétiques*, niveau cinq : 4 résultats ; niveau deux : 657). Avec de tels effectifs, il devient humainement impossible de parcourir attentivement tous les résultats, et l'opération devient inutile a fortiori si ceux-ci ne sont en définitive pas significatifs. On constate, en effet, que d'un niveau à l'autre, le logiciel répète de nombreux résultats, ce qui empêche l'utilisateur d'avoir aisément accès aux données réellement exploitables. Pour une réminiscence de niveau n , il semblerait que le programme présente, au niveau $n - 1$, cette même réminiscence multipliée par le nombre de combinaisons des différents couples qui composent son tuple.

Depuis le colloque romain, Nabil Cherchem, étudiant en informatique de l'Université de Namur sous la direction du professeur Paul Pietquin, a introduit un algorithme correcteur facultatif qui résout cette multiplication invasive des données. Malgré cette importante amélioration, on constate encore un nombre élevé de résultats. Par ailleurs, N. Cherchem a corrigé d'autres défauts ponctuels (comme certains problèmes d'affichage du grec dans l'interface) et a travaillé à enrichir le logiciel de nouvelles fonctionnalités. Ainsi, il est désormais possible d'exporter les résultats sous plusieurs formats de fichier (Word, HTML et PDF) et la liste des lemmes à exclure de la recherche (comme évoqué ci-dessus) peut être aisément changée par l'utilisateur, qui peut ainsi définir avec souplesse ses propres critères d'investigation. Quoiqu'il en soit, il est d'ores et déjà clair que le programme continuera à faire l'objet de plusieurs mises à jour au fil des prochains essais, afin d'en affiner toujours davantage la précision.

B. *Apports et limites*

Bien que le programme soit encore en pleine phase de maturation et d'évolution, en fonction des observations faites par ses utilisateurs, nous pouvons dès à présent tenter une esquisse de ses apports majeurs mais aussi de ses principales limites. Tout d'abord, une fois que les textes sont lemmatisés, *ReminiFinder* permet de traiter des écrits grecs de grande taille et donc, par exemple, de s'intéresser à l'ensemble du corpus Arrien - Xénophon. Il apporte également à la recherche un critère automatique et systématique que l'esprit humain atteint difficilement, limitant ainsi les erreurs et les oublis. Néanmoins, une fois confronté à cet ensemble de résultats et après les avoir comparés avec ceux de la recherche livresque, le philologue constate rapidement que les deux approches sont indissociables pour obtenir l'étude la plus complète possible. En effet, deux dimensions inéluctables limitent les performances du logiciel. Il s'agit de sa nature et des choix posés pour réduire la quantité de données.

En premier lieu, quatre caractéristiques inhérentes au programme altèrent son efficacité. Ainsi, *ReminiFinder* fonctionne sur la base de fichiers de textes lemmatisés. Or la lemmatisation complète d'œuvres comme celles d'Arrien ou de Xénophon demande un investissement en temps considérable. C'est pourquoi nous avons choisi la procédure en deux temps (lemmatisation automatique des formes connues et lemmatisation manuelle des formes non connues) décrite précédemment. La première phase de lemmatisation, bien que plus rapide, génère une série d'analyses multiples¹⁶ qui augmentent le « bruit » déjà important des résultats non significatifs. Néanmoins, un rapide test¹⁷ a permis de démontrer que le gain obtenu par une lemmatisation complète ne serait pas suffisamment important par rapport au temps investi pour réaliser cette lemmatisation. Cette méthode indiquée pour notre type de recherches sur un corpus étendu représente donc elle-même une limite du programme.

Ensuite, le logiciel travaille sur des formes attachées à des *lemmes*. Dès lors, si Arrien reprend une expression chez Xénophon mais en la modifiant de sorte que les formes conservent leur sens initial mais ne dépendent plus du même lemme (exemple 1), la formule ne peut pas être reconnue par le logiciel.

16. Une même forme peut avoir un même lemme mais dépendre de catégories grammaticales différentes (ἐπί : préposition ou adverbe) ; elle peut aussi être attachée à deux lemmes différents (κινήσεις : κίνησις ou κινέω).

17. Nous avons lié arbitrairement une et une seule analyse (lemme et catégorie grammaticale) à chaque forme du texte pour comparer les résultats de ces fichiers désambiguïsés avec ceux obtenus par la méthode décrite, qui laisse subsister des analyses multiples. La différence n'est que de 3 à 6 %.

La troisième caractéristique du programme est qu'il fonctionne sur la base de *couples* de formes aux lemmes identiques. Il est donc incapable de retrouver chez Arrien des mots isolés mais propres au vocabulaire de la langue de Xénophon et, par conséquent, révélateurs d'une reprise textuelle¹⁸ (exemple 2). Par ailleurs, même si le programme était en mesure de retrouver des termes isolés, il faudrait qu'il soit couplé avec une base de données des textes grecs munis d'un indice fréquentiel pour chaque lemme, afin de ne conserver que les lemmes spécifiques à Xénophon et ainsi éliminer un nombre considérable de résultats inutiles. Cette dernière réflexion nous amène également à souligner l'intérêt d'utiliser le TLG en parallèle avec *ReminiFinder* pour les autres données fournies. Même si une formulation apparaît dans les deux textes, il est probable qu'elle soit aussi utilisée par d'autres auteurs. Une vérification dans le reste de la littérature conservée – le TLG est ici le meilleur outil – est dès lors indispensable (exemple 3).

Même si elle peut paraître évidente, il importe de mentionner enfin une quatrième caractéristique, qui concerne une limite plus large du logiciel. Comme *ReminiFinder* repère des tuples, il ne peut dégager des réutilisations plus subtiles, comme les motifs ou les thèmes qui ne sont pas nécessairement exprimés avec les mêmes termes d'un auteur à l'autre. Notre étude se concentre donc essentiellement sur les reprises textuelles de Xénophon chez Arrien, qui sont les seules à pouvoir être repérées par le logiciel. À celles-ci s'ajoutent les réminiscences décelées par le dépouillement des études antérieures et par nos lectures de Xénophon et d'Arrien.

Par ailleurs, les choix destinés à restreindre le nombre de fausses réminiscences potentielles laissent de côté également une série de réelles reprises. En effet, en éliminant certaines catégories grammaticales a priori non porteuses de sens, de même que certains mots « outils », l'utilisateur passe à côté de certains termes significatifs en raison du contexte ou de leur position particulière dans les deux textes (exemple 4).

En conclusion, il est indispensable de bien comprendre les atouts et les faiblesses du logiciel pour pouvoir adapter la méthode de recherche en vue d'une utilisation optimale. Il semble aussi nécessaire de l'employer comme un outil supplémentaire (en parallèle avec les études antérieures, les éditions et commentaires, les recherches livresques, le TLG...) pour ne pas négliger certaines données très riches dans l'étude des réutilisations littéraires.

18. L'étude d'H. TONNET, *op. cit.* (n. 11), fait le point sur ce type de reprise et nous a fourni le matériel à soumettre à une analyse philologique.

C. Exemples¹⁹

C.1. ἡγεμών / ἡγεμονία

Ce premier exemple, probable reprise de Xénophon (*Hell.*, I, 4, 20) chez Arrien (*Anab.*, I, 1, 2)²⁰, ne pourrait pas être mis au jour par *ReminiFinder*. L'auteur de Nicoméde emploie dans son texte la forme τὴν ἡγεμονίαν, liée à son lemme ἡγεμονία, tandis que son modèle avait utilisé le substantif ἡγεμών, du lemme ἡγεμών. Le programme est ici incapable d'associer ces deux lemmes différents, bien que proches de sens, et l'utilisateur manque une information potentiellement importante. En l'occurrence, seule une consultation préalable de la bibliographie disponible pouvait nous permettre de compléter au mieux notre étude.

C.2. ὑπερύψηλος

Ce terme pointé par H. Tonnet – parmi de nombreux autres termes²¹ – comme propre au vocabulaire xénophontéen (*Anab.*, III, 5, 7) mais présent également chez Arrien (*Anab.*, I, 5, 12 ; 13, 5 ; 27, 5 ; III, 24, 5 ; IV, 21, 3 ; 27, 6) pourrait constituer une trace de l'influence du premier sur le second. Or le logiciel, basé sur la recherche de paires de lemmes communs, est incapable de trouver de telles occurrences. Pourtant, des mots isolés caractéristiques du lexique de Xénophon participent à part entière à la coloration xénophontéenne qu'Arrien cherche à donner à son œuvre. Il convient donc de les prendre en compte dans notre recherche.

C.3. ἐν ἀριστερᾷ / δεξιᾷ ἔχων

Également considérée par H. Tonnet²² comme une réutilisation, cette expression rencontrée chez Arrien (*Anab.*, I, 1, 5 ; III, 1, 3 ; 7, 3 ; 7, 7 ; VII, 17, 5 ; 22, 1) et chez Xénophon (*Anab.* I, 5, 1 ; II, 4, 28 ; IV, 3, 16 ; *Hell.*, VI, 5, 27) apparaît en fait aussi chez d'autres historiens ou biographes de langue grecque comme Hérodote, Thucydide, Strabon, Diodore de Sicile ou encore Plutarque. Le logiciel seul ne nous permet évidemment pas d'avoir accès à ces données complémentaires, indispensables toutefois pour le commentaire de cette formule. Seuls le recours au TLG et l'analyse philo-

19. Cette série d'exemples provient de la liste des réminiscences potentielles fondée sur le dépouillement de la littérature spécialisée sur Arrien. Elle tend à démontrer que certains cas de figure ne sauraient être pris en compte par le logiciel et qu'il est nécessaire de compléter la recherche avec d'autres outils informatiques ou livresques.

20. Hypothèse émise par A. B. BOSWORTH, *op. cit.* (n. 11), vol. 1, p. 49.

21. H. TONNET, *op. cit.* (n. 11), p. 261-266.

22. H. TONNET, *op. cit.* (n. 11), p. 249.

logique de tous les passages ainsi mis au jour nous permettent de savoir s'il s'agit bien d'une tournure propre aux deux auteurs étudiés – et donc plus probablement un cas de reprise – ou si c'est en fait une expression plus largement connue des auteurs grecs.

C.4. λέγεται δὴ ... μὲν ... δέ

Le dernier exemple développé ici présente une réminiscence très probable et très intéressante de Xénophon chez Arrien²³. Or le programme ne l'aurait pas détectée, par suite des mesures destinées à réduire le « bruit » dans les résultats obtenus. Nous avons, en effet, retiré la catégorie grammaticale des particules du moteur de recherche et nous avons exclu le verbe λέγω des lemmes du texte. Cependant, ces deux extraits illustrent un cas de reprise textuelle composée uniquement de ce verbe et de particules :

Λέγεται δὴ Φίλιππος μὲν τελευτῆσαι ἐπὶ ἄρχοντος Πυθοδήλου Ἀθήνησι· παραλαμβάντα δὲ τὴν βασιλείαν Ἀλέξανδρον, παῖδα ὄντα Φιλίππου, ἐς Πελοπόννησον παρελθεῖν· εἶναι δὲ τότε ἀμφὶ τὰ εἴκοσιν ἔτη Ἀλέξανδρον.
(Arr., *Anab.*, I, 1, 1)

On place la mort de Philippe sous l'archontat de Pythodélos, à Athènes. Après avoir hérité de la royauté puisqu'il était le fils de Philippe, Alexandre arriva dans le Péloponnèse. Il avait environ vingt ans²⁴.

Πατὴρ δὲ μὲν δὴ ὁ Κῦρος λέγεται γενέσθαι Καμβύσου Περσῶν βασιλέως· ὁ δὲ Καμβύσης οὗτος τοῦ Περσειδῶν γένους ἦν· οἱ δὲ Περσεῖδαι ἀπὸ Περσέως κληῖονται· μητὴρ δὲ ὁμολογεῖται Μανδάνης γενέσθαι· ἡ δὲ Μανδάνη αὕτη Ἀστυάγους ἦν θυγάτηρ τοῦ Μήδων γενομένου βασιλέως.
(Xén., *Cyrop.*, I, 2, 1)

Cyrus, dit-on, eut pour père Cambyse, roi des Perses. Ce Cambyse était de la famille des Perséides, dont le nom vient de Persée. On s'accorde à lui donner pour mère Mandane. Cette Mandane était la fille d'Astyage, roi des Mèdes²⁵.

Ce n'est évidemment pas le sens même des mots qui est révélateur de la réminiscence détectée ici mais bien l'association de ces termes dans un contexte précis et à une place particulière au début de ces deux œuvres. Si évidente soit-elle, seule une lecture très attentive des œuvres permet de découvrir cette allusion à Xénophon chez Arrien. Or il a fallu attendre la publication en 1980 du premier volume du commentaire d'Albert B. Bosworth sur l'*Anabase* d'Arrien pour la mettre au jour.

23. A. B. BOSWORTH, *op. cit.* (n. 11), vol. 1, p. 45.

24. Le texte utilisé est celui édité par A. G. ROOS, Leipzig, Teubner, 1967. La traduction française est personnelle.

25. Texte et traduction : Xénophon, *Cyropédie, livres I-II*, éd. et trad. M. BIZOS, Paris, « Les Belles Lettres », 1971.

4. Un répertoire de fiches : méthode et objectifs

L'objectif final du projet vise à constituer un répertoire le plus complet possible des réminiscences potentielles découvertes grâce au dépouillement de la littérature spécialisée et grâce aux nouveaux résultats apportés par le logiciel. Pour celles du premier type, nous avons décidé, en raison de notre souci d'exhaustivité, de reprendre tous les passages considérés par les spécialistes comme une réutilisation littéraire même si notre commentaire conduit ensuite à une conclusion mitigée, voire opposée. Toutes les références ont alors été réunies dans l'ordre des textes dans deux index inverses : Arrien - Xénophon ; Xénophon - Arrien (Annexe 6).

Chaque reprise textuelle fera l'objet d'une fiche (les résultats de la recherche livresque sont aujourd'hui en cours d'étude – on en compte actuellement 261) reprenant les éléments suivants : la réminiscence, son niveau de reprise (le nombre de termes repris par Arrien), les références de chacune de ses occurrences chez les deux auteurs avec leur contexte, les extraits du texte grec, leur traduction et la bibliographie (s'il y a lieu) et enfin et surtout un commentaire critique en deux volets (résumés des commentaires antérieurs et nos nouveaux éclairages).

Mener parallèlement le dépouillement et les améliorations du logiciel nous a permis de comparer les résultats de la première méthode de recherche avec la deuxième. Ainsi, en confrontant les deux *Cynégétiques*, *ReminiFinder* trouve deux réminiscences (Arrien, 30, 2 et Xénophon, 7, 3 ; Arrien, 2, 2-3 et Xénophon, 5, 29) potentielles de niveau cinq déjà repérées par les spécialistes et en propose une encore non identifiée (Arrien, 5, 10 et Xénophon, 4, 1). Ce répertoire poursuit donc un but très concret et utilitaire dans le développement même du programme mais se justifie avant tout pour lui-même.

En effet, une fois la base de données augmentée des résultats du programme et une fois chaque fiche analysée, elle constituera un fondement solide et fiable pour réévaluer notre analyse, notre expertise et donc notre connaissance du système de reprise textuelle de Xénophon chez Arrien. Elle pourra ensuite servir de base à de futures recherches destinées à mieux saisir le fonctionnement de création littéraire à l'époque impériale, subtil équilibre entre inspiration et innovation.

5. En guise de conclusion

Cet ambitieux projet alliant philologie et informatique apporte beaucoup à l'étude de l'œuvre d'Arrien et du concept de réutilisation littéraire puisqu'il élabore le premier répertoire le plus complet possible des réminiscences potentielles (celles détectées préalablement et les nouvelles ob-

tenues avec *ReminiFinder*) de Xénophon chez l’auteur de Nicomédie. Toutefois, la méthode utilisée pour parvenir à cet objectif est elle-même un objet d’étude dans le domaine de la philologie numérique. En effet, la création d’un logiciel spécifique dans un champ d’étude encore assez peu exploité par voie informatique comme la littérature grecque pose question au spécialiste. Comme nous l’avons décrit, certaines interrogations ont été soulevées (lemmatisation spécifique, limitation du « bruit »...) et le seront encore au cours du développement du programme ; les solutions définies ou celles à venir²⁶ font partie intégrante du processus d’analyse scientifique.

Néanmoins, le recours à de tels instruments même très performants peut constituer aussi un obstacle pour le philologue car, en plus des différences de langage et de méthode entre la philologie et l’informatique, une certaine dépendance se crée lorsqu’il faut mettre en place les améliorations et les mises à jour du logiciel. C’est pourquoi les prochains outils à créer²⁷ viseront à une plus grande autonomie de l’utilisateur vis-à-vis des informaticiens.

Avec ce projet, nous sommes donc confrontés à la fois à des questions philologiques et informatiques qu’il convient de résoudre parallèlement pour optimiser au mieux les performances de *ReminiFinder*. L’objectif final de ce volet de notre entreprise scientifique est que le logiciel soit capable de repérer automatiquement et le plus efficacement possible les réminiscences potentielles de n’importe quel corpus afin d’étudier d’autres auteurs qu’Arrien et Xénophon²⁸.

Marc VANDERSMISSEN
F.R.S.-FNRS - Université de Liège

Koen VANHAEGENDOREN
Chargé de cours en langues et littératures classiques
Université de Liège
Place du 20-Août, 7 - Bât. A1
B - 4000 Liège

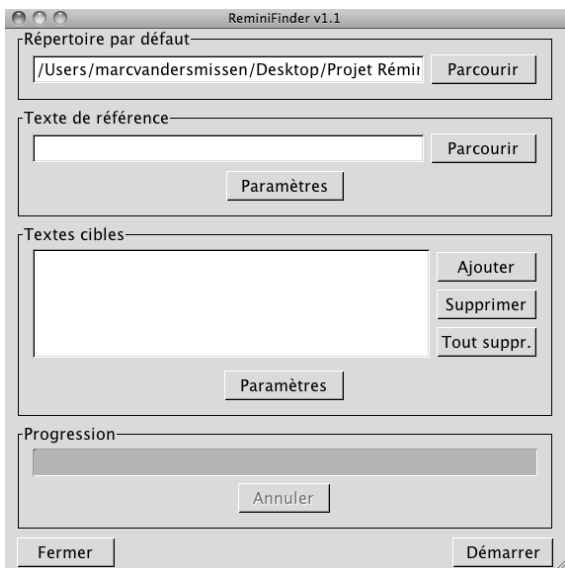
26. Nous tenons à remercier une fois encore Stephan Guralnik, Paul Pietquin, l’équipe du LASLA et du CIPL (notamment Gérard Purnelle, Laurent Simon et Benoît Morimont) : sans eux, la continuité du développement du programme n’aurait pas été possible.

27. À titre d’exemple citons le lecteur des fichiers du LASLA.

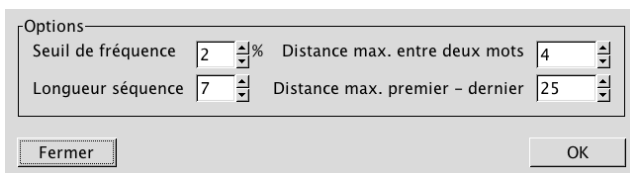
28. La discussion que nous avons pu avoir suite à la présentation de ce projet à Rome, le 20 octobre 2011, a manifesté l’intérêt que voient plusieurs des participants à élargir les applications de *ReminiFinder* à d’autres langues, telles que le français, le latin classique et le latin médiéval.

Annexes

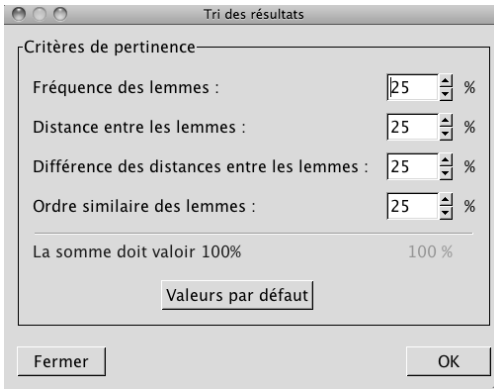
1. Interface d'accueil



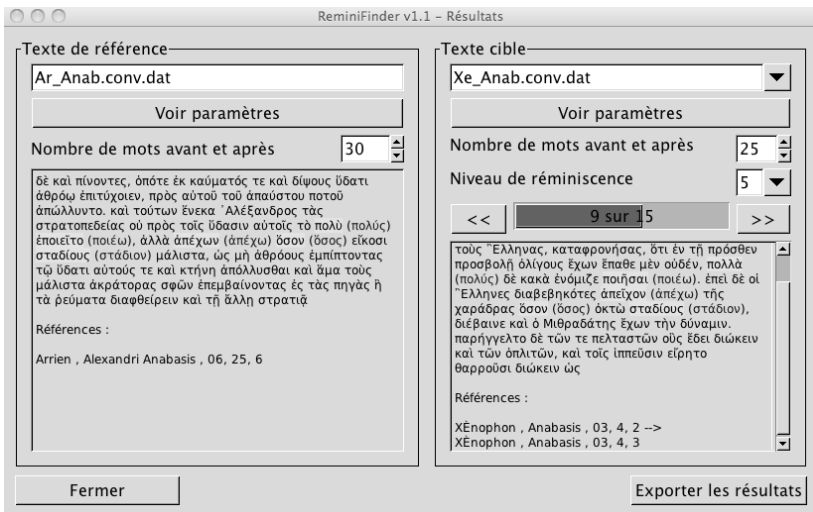
2. Options de recherche



3. Critères de pertinence dans le classement des résultats



4. Fenêtre de présentation des résultats



5. Paramètres de recherche : catégories grammaticales et liste des lemmes exclus

Paramètres du texte de référence

Catégories grammaticales à conserver

<input type="checkbox"/> Article	<input checked="" type="checkbox"/> Verbe	<input type="checkbox"/> Interjection
<input checked="" type="checkbox"/> Substantif	<input checked="" type="checkbox"/> Adverbe	<input type="checkbox"/> Particule
<input checked="" type="checkbox"/> Adjectif	<input type="checkbox"/> Préposition	<input type="checkbox"/> Variable
<input type="checkbox"/> Numéral	<input type="checkbox"/> Conj. Coord.	<input type="checkbox"/> Auxiliaire "être"
<input type="checkbox"/> Adjectif-Pronom	<input type="checkbox"/> Conj. Subord.	

Emplacement d'un fichier d'exclusions

6. Un exemple de l'index Arrien - Xénophon

ARRIEN	XÉNOPHON	Étudié dans
<i>Anabase d'Alexandre</i>		
I, 1, 1. => λέγεται δῆ	<i>Cyrop.</i> , I, 2, 1.	BOSWORTH (1980, vol. I, p. 45).
I, 1, 5. => ἐν ἀριστερῇ/δεξιῇ ἔχων	<i>Anab.</i> , I, 5, 1 ; II, 4, 28 ; IV, 3, 16. <i>Hell.</i> , VI, 5, 27.	TONNET (1988, p. 249 et vol. II, p. 184, n. 149).
I, 1, 12. => οἱ τοξόται [...] τοὺς προεκθέοντας ἀνέστελλον	<i>Anab.</i> , V, 4, 23. => τούτους ἀνέστελλον οἱ τοξόται	TONNET (1988, p. 245 et vol. II, 182, n. 128).
I, 1, 12. => οὐκέτι ἐδέξαντο ἀλλὰ [...] κατὰ τοῦ ὄρους ἔφυγον	<i>Anab.</i> , VI, 5, 27. => οὐκέτι ἐδέξαντο οἱ πολέμοι ἀλλὰ ἔφυγον	TONNET (1988, p. 246).
I, 2, 2. => ὄνομα τῇ πόλει ἦν	<i>Anab.</i> , III, 4, 10.	TONNET (1988, p. 249 et vol. II, 184, n. 150).
I, 2, 4. => νάπος	<i>Anab.</i> , VI, 5, 12; 5, 13; 5, 18; 5, 19; 5, 20; 5, 22; 5, 31. <i>Cyrop.</i> , I, 6, 43.	TONNET (1988, p. 264).
I, 2, 5. => νάπη	<i>Anab.</i> , IV, 5, 15; 5, 18; V, 2, 31. <i>Cynég.</i> , IX, 11. <i>Hell.</i> , V, 4, 44.	TONNET (1988, p. 264).