

La concordance bilingue grecque-syriaque des *Discours* de Grégoire de Nazianze

Par

Bastien Kindt, Jean-Claude Haelewyck,
Andrea Schmidt et Nicolas Atas

Université catholique de Louvain

La concordance bilingue grecque-syriaque des œuvres de Grégoire de Nazianze présentée dans ces lignes¹ fait partie d'un ensemble de développements expérimentaux relevant d'un projet plus ambitieux dont l'objectif est double.

1. Il s'agit, d'abord, de rendre accessibles en ligne, via une interface d'interrogation, les données lemmatisées publiées en 1990 et 1991 dans les deux volumes du *Thesaurus Sancti Gregorii Nazianzeni*, la première concordance grecque des *opera omnia* de Grégoire². En utilisant cette interface, le chercheur peut obtenir, par exemple, toutes les *occurrences* du mot grec ποιμήν/*pasteur* attestées dans le *corpus* interrogé. Le résultat de cette requête s'affiche à l'écran sous la forme d'une concordance, comme illustré sous la figure 1³.

¹ Le contenu de cet article a fait l'objet d'une communication présentée par B. Kindt le 20 février 2017 dans le cadre du colloque *Tradurre, tradire, tramandare i Padre Greci nell'Occidente Latino e nell'Oriente Siriaco* – 20-21 janvier 2017, Università degli Studi, Roma Tre.

² MOSSAY *et alii* 1990 ; MOSSAY *et alii* 1991. Sur le *Thesaurus Patrum Graecorum*, une sous-série du *Corpus Christianorum*, cf. COULIE 1996, p. 35-54, et KINDT 2004, p. 213-237.

³ Les figures de cet article sont également disponibles sur le site Internet du projet GREgORI, cf. <https://uclouvain.be/fr/instituts-recherche/incal/ciol/gregori-project.html> (les liens cités dans cet article étaient

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>ποιμήν { N+Com } 9 GRNA Or. 1 7 35 400 D 1 GRNA Or. 1 6 35 400 B 1 GRNA Or. 1 7 35 401 A 10 GRNA Or. 13 1 35 853 A 4 GRNA Or. 13 3 35 856 A 8 GRNA Or. 13 2 35 853 C 6 GRNA Or. 1 7 35 401 A 9 GRNA Or. 13 4 35 856 A 13 GRNA Or. 1 6 35 400 B 3</p> | <p>και ἐπὶ ὕδατος ἀναπαύσεως ἐκτρέφόμενοι γινώσκοντες καλῶς τὸν ποιμένα και γινωσκόμενοι και ἐπόμενοι καλοῦντι ποιμενικῶς και ἐλευθερίως διά τῆς θύρας Καρποφορεῖ μὲν ὑμῖν ὡς ὄρατε ποιμένα· τοῦτο γὰρ ἐλπίζει και εὐχεται και παρ' ὑμῶν αἰτεῖ τῶν ὑπὸ χεῖρα ὁ καλὸς ὁ ἀπὸ τοῦ ἀληθινοῦ και πρῶτου ποιμένος· ὡν εἴη πάντας ἡμῶς και ποιμένας και ποιμνιον ὡς νοσεράς πῶας και θανασίμου πόρρω και νεμομένους και νέμοντας Μανασσῆ τὴν ἐξαγόρευσιν. Δέξασθε λόγον νεόκτιστον ἐπὶ νεοκτίστω ποιμένι· δέξασθε φωνὴν εὐχαριστήριον ἐπὶ τοῖς ὀρωμένοις θαύμασι. Μικροὶ μὲν ἡμεῖς και τοσοῦτον ἡ ῥομφαία τοῦ Θεοῦ κατωθὴ και ἡρεμήσειεν. Οὐτε τῷ ποιμένι τῷ σῶ μέγα τι ἐχαρίσω και σαυτὸν τὰ μέγιστα ἐζημίωσας ἀλλοτριώσας σεαυτὸν τῆς βάλλεν μάχαιραν ἀλλ' εἰρήνην. Οὐκ ἦλθομεν ἐπὶ ἀτιμία τοῦ μεγάλου ποιμένος τοῦ τῆς λαμπρᾶς προκαθεζομένου πόλεως· τίμιον ἴσμεν κεφαλὴν γινωσκόμεν ἁγίον συλαγωγούσης και διασπύσης ἀπὸ τοῦ ἀληθινοῦ και πρῶτου ποιμένος· ὡν εἴη πάντας ἡμῶς και ποιμένας και ποιμνιον ὡς νοσεράς πῶας και θανασίμου Νῦν δὲ ὡ ποιμένων ἄριστε και τελεώτατε δεῦρό μοι και τὸν σὸν ἀπολαμβάνε λαὸν σὺν ἡμῖν τε και πρό τοῦτο γὰρ ἐλπίζει και εὐχεται και παρ' ὑμῶν αἰτεῖ τῶν ὑπὸ χεῖρα ὁ ποιμήν ὁ καλὸς ὁ τίθει τὴν ψυχὴν ὑπὲρ τῶν προβάτων· και διπλοῦν ἀνθ' ἀπλοῦ διδωσιν ὑμῖν</p> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Figure 1. Exemple de concordance en grec : occurrences du lemme ποιμήν, le pasteur

L'utilisateur peut réaliser des recherches basées sur un *lemme*, comme c'est le cas ici avec ποιμήν. Ce type d'interrogation permet d'obtenir la concordance de toutes les *formes* fléchies qui découlent de ce *lemme* (c'est-à-dire ποιμένα, ποιμένας, ποιμένι, ποιμένος, ποιμένων et ποιμήν). Il peut aussi rechercher une *forme*, par exemple ποιμένα. La concordance affiche alors les *occurrences* de la *forme* recherchée, à savoir, ποιμένα. Il peut aussi utiliser les *étiquettes* identifiant la *catégorie morphosyntaxique* d'un lemme, par exemple, toujours pour ποιμήν, « N+Com », l'*étiquette* des noms communs. La concordance affiche dans ce cas toutes les *occurrences* des lemmes catégorisés comme noms communs. Enfin, il est également possible de baser les recherches sur une combinaison de ces différents éléments, afin de produire la concordance de toutes les séquences constituées, par exemple, d'un adjectif (*étiquette* « A ») suivi par une *occurrence* du lemme ποιμήν. La concordance porte alors sur les séquences du type μεγάλου ποιμένος.

2. Il s'agit, ensuite, de mettre en parallèle les données lemmatisées du texte grec (le *texte source*) et celles des versions orientales – en arménien, en géorgien, en arabe et en syriaque⁴ –, ces versions orientales étant les traductions anciennes des œuvres de Grégoire (les *textes cibles*). Pour arriver à ce résultat, deux opérations sont nécessaires ; la *lemmatisation* des

actifs en date du 31 mai 2018). Le lecteur y trouvera aussi les versions complètes des concordances monolingues, en grec ou en syriaque, des *Discours* 1 et 13, ainsi que les concordances bilingues.

⁴ Pour les développements en arménien, cf., par exemple, VAN ELVERDINGHE 2018. En géorgien, cf. COULIE *et alii* 2013, p. 161-201, et PATARIDZE, KINDT 2018. Pour l'arabe, cf. TUERLINCKX 2004, p. 1069-1078. En syriaque, cf. KINDT 2018.

textes des versions orientales, d'une part, et l'alignement des textes sources et des textes cibles, d'autre part.

La lemmatisation des textes des versions orientales est en cours. Ainsi, pour le syriaque, l'interface permet de rechercher, comme dans le cas du grec illustré ci-dessus, les attestations du lemme ܦܫܬܘܪ/pasteur. L'utilisateur peut également formuler des interrogations portant sur une forme ou sur une étiquette relative aux catégories morphosyntaxiques, mais aussi sur la racine sémitique des mots ou sur une séquence de ces différents éléments. Le résultat s'affiche sous la forme d'une concordance, comme illustré sous la figure 2.

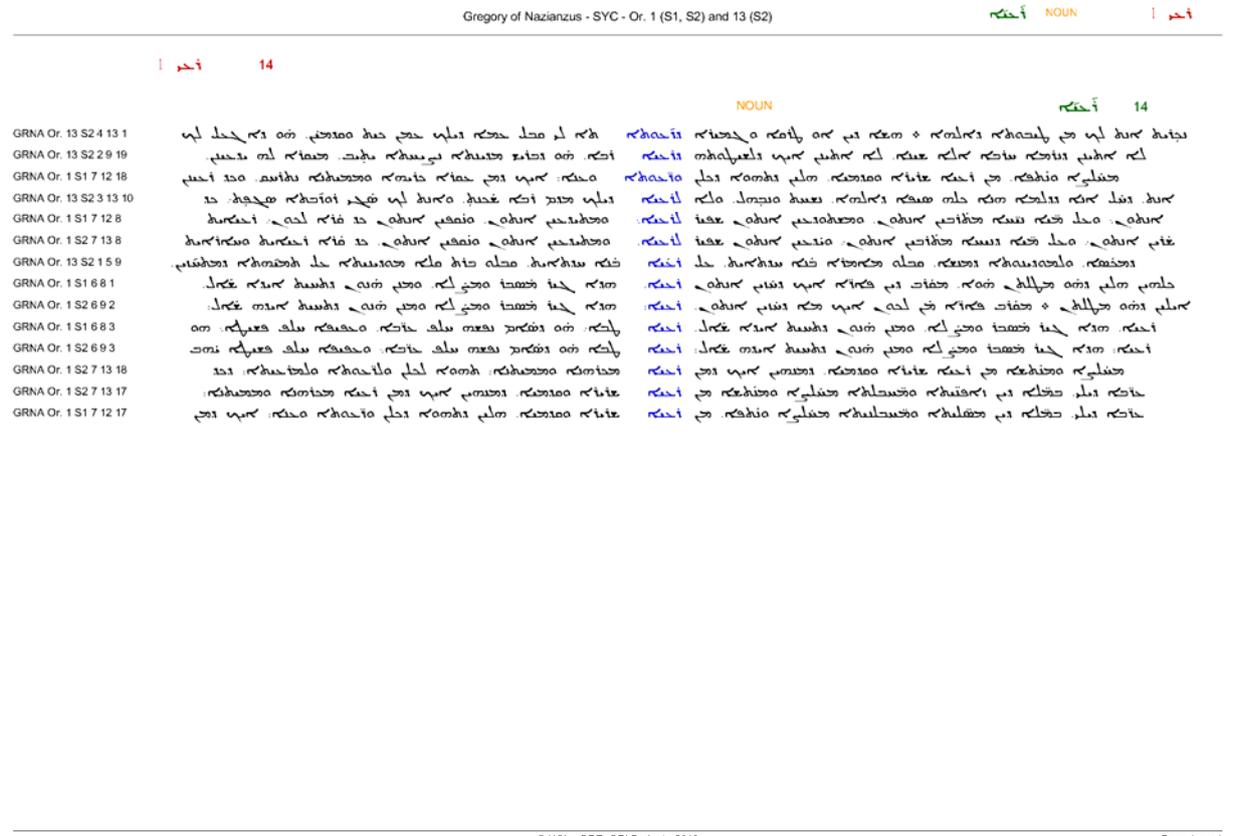


Figure 2. Exemple de concordance en syriaque : occurrences du lemme ܦܫܬܘܪ/pasteur

Cette concordance des occurrences du lemme ܦܫܬܘܪ/pasteur permet de visualiser à l'écran non seulement les formes simples du mot, mais aussi les formes suffixées du type ܦܫܬܘܪܝܢ et ܦܫܬܘܪܝܠ.

L'opération d'alignement, quant à elle, permet d'apparier les mots ou les expressions du texte source aux mots ou expressions qui en sont les traductions dans les textes cibles⁵. Ainsi, l'utilisateur de l'interface peut interroger en même temps le texte grec et une de ses versions. Les résultats s'affichent alors sous la forme de concordances bilingues. Dans la concordance bilingue en grec et en syriaque, la requête basée sur le lemme grec ποιμήν fournit aussi les occurrences syriaques du lemme ܦܫܬܘܪ.

⁵ Sur les notions d'alignement, de texte source et de texte cible, cf. LEFER à paraître. À ce stade des développements, l'alignement est réalisé manuellement à l'aide du logiciel mkAlign développé à l'Université de Paris 3 par Serge Fleury : cf. FLEURY 2012.

L'ambition finale est d'offrir aux chercheurs un outil permettant d'interroger les textes grecs et leurs versions orientales dans un environnement simple d'utilisation, malgré la présence concomitante de sources écrites dans les différentes langues de l'Orient chrétien. Cela impose de mettre au point un outil intégrant de manière efficace l'affichage, sur un même écran de recherche, des différents alphabets utilisés pour écrire ces langues – en l'occurrence, ici, les alphabets latin, grec et syriaque – et des différents sens d'écriture – en l'occurrence horizontal dextroverse (de gauche à droite) pour le grec ou horizontal sinistroverse (de droite à gauche) pour le syriaque.

Ces travaux sont le fruit d'une collaboration entre le projet GREgORI, mené à l'Institut orientaliste de l'Université catholique de Louvain, et le CENTAL, *Centre de traitement automatique du langage*, un laboratoire d'informatique-linguistique de la même université⁶. Les philologues de l'Institut orientaliste conçoivent les *ressources linguistiques* utiles pour le traitement des langues abordées. Les informaticiens du CENTAL développent les *outils informatiques* permettant d'assurer la gestion, la visualisation et l'exploitation de ces données.

Nous exposons ici les résultats du traitement, en grec et en syriaque, des *Discours* 1 et 13 du Nazianzène. Après avoir décrit le *corpus* traité (Section 1), nous présentons les règles d'analyse établies pour la lemmatisation des textes syriaques, à l'instar de ce qui existe pour le grec⁷ (Section 2). Nous décrivons ensuite les concordances et différents outils lexicographiques produits avec les données lemmatisées (Section 3). Nous terminons par une description de la version expérimentale de l'interface en ligne (Section 4). Nous pourrions enfin conclure et ouvrir quelques perspectives (Section 5).

1. Le corpus

Le *corpus* analysé est constitué de deux homélies de Grégoire de Nazianze, les *Discours* 1 et 13. Le *Discours* 1, intitulé *Εἰς τὸ ἅγιον Πάσχα καὶ εἰς τὴν βραδυτῆτα/Sur Pâques et sur les lenteurs de l'auteur*, a été prononcé le jour de Pâques 362, le 31 mars⁸ ; Le *Discours* 13 a pour titre *Εἰς τὴν χειροτονίαν Δοαρῶν ὁμιλία ἐκδοθεῖσα Ἐὐλαλίῳ ἐπισκόπῳ/Discours prononcé à Doara à l'occasion de la consécration de l'évêque Eulalios* et se situe chronologiquement aux alentours de l'année 373⁹.

Pour le grec, les données lemmatisées sont directement reprises du *Thesaurus* cité. Le tableau 1 présente le nombre d'*occurrences*, de *formes* et de *lemmes* dans le *corpus* grec.

| | Occurrences | Formes différentes | Lemmes | Racines |
|--------------------|-------------|--------------------|--------|---------|
| <i>Discours</i> 1 | 777 | 454 | 340 | — |
| <i>Discours</i> 13 | 599 | 383 | 302 | — |
| <i>Total</i> | 1376 | 764 | 559 | — |

Tableau 1. Nombre d'occurrences, de formes différentes et de lemmes dans le corpus grec

⁶ La description et la bibliographie complète du projet GREgORI sont accessibles sur le site Internet cité note 3. Pour le CENTAL, cf. <https://uclouvain.be/fr/instituts-recherche/ilc/cental>. Les auteurs tiennent à adresser leurs plus vifs remerciements à MM. Hubert Naets et Claude Devis, informaticiens au CENTAL.

⁷ KINDT 2004, p. 237-271.

⁸ *Patrologia Graeca*, 35, col. 395a-401b; lemmatisation : Bernard Coulie, Bastien Kindt et Justin Mossay†.

⁹ *Patrologia Graeca*, 35, col. 852a-856c; lemmatisation : Bernard Coulie, Bastien Kindt et Justin Mossay†.

Pour le syriaque, l'analyse porte sur les textes publiés dans le *Corpus Nazianzenum*¹⁰ et a été réalisée *ab ovo*. La version syriaque des homélies de Grégoire est constituée de plusieurs états de traduction¹¹. Le premier, une *Versio antiqua* nommée S1, date du VI^e siècle. Le second relève d'une traduction dite « intermédiaire », une *Versio media* nommée Sm et située chronologiquement après 550. Le troisième, une *Versio nova* nommée S2, est une révision de S1 et de Sm réalisée par Paul d'Édesse au VII^e siècle. Les éditions des *Discours* 1 et 13 reprennent S1, Sm et S2¹². Le tableau 2 présente le nombre d'*occurrences*, de *formes*, de *lemmes* et de *racines* dans le *corpus* syriaque.

| | Occurrences | Formes différentes | Lemmes | Racines |
|-------------------------|-------------|--------------------|--------|---------|
| <i>Discours</i> 1 (S1) | 685 | 475 | 323 | 262 |
| <i>Discours</i> 1 (S2) | 711 | 483 | 326 | 263 |
| <i>Discours</i> 13 (S2) | 560 | 398 | 289 | 254 |
| <i>Total</i> | 1956 | 947 | 562 | 421 |

Tableau 2. Nombre d'*occurrences*, de *formes* différentes, de *lemmes* et de *racines* dans le *corpus* syriaque

Ce *corpus* bilingue réunit un ensemble de données quantitativement modestes. De fait, en grec, les homélies de Grégoire totalisent plus de 216.000 *occurrences* et, en syriaque, le *corpus* des textes déjà publiés dans le *Corpus Nazianzenum* atteint les 103.000 *occurrences*. L'analyse proposée ici pour le syriaque ne prétend donc pas fournir une description globale et aboutie de la langue syriaque, ni de manière générale, ni en ce qui concerne plus particulièrement la langue des différents états des versions syriaques de Grégoire. Le but est, plus modestement, d'offrir une assise expérimentale à une analyse *cohérente* et *stable* des textes, dans le cadre d'un projet plus large de *lemmatisation* et d'*alignement* des *textes sources* et des *textes cibles*. Il s'agit donc de définir, pour la lemmatisation en syriaque, des principes directeurs, comme cela a été fait, *in illo tempore*, pour le grec¹³. Les questions principales portent sur l'intitulé des *lemmes* et des *racines*, et sur la définition des *catégories morphosyntaxiques* dont relèvent les *lemmes*. Ces principes, simples, cohérents et stables – censés décrire sans trahir – pourront servir de normes pour les analyses futures menées tant par les membres du projet GREgORI que par des collaborateurs extérieurs.

2. Ressources linguistiques pour la lemmatisation du syriaque

En grec, la lemmatisation permet d'identifier le *lemme* de chacune des *formes* présentes dans un texte donné. De plus, chaque *lemme* est accompagné d'une indication de sa *catégorie morphosyntaxique* (nom, verbe, adjectif, pronom, etc.), la partie du discours dont le *lemme* relève. L'utilisateur travaillant sur une interface en ligne peut ainsi rechercher dans les textes

¹⁰ Pour le *Discours* 1 : HAELEWYCK 2011 (*Versio Antiqua* [S1], p. 2-12; *Versio nova* [S2], p. 3-13); lemmatisation : Jean-Claude Haelewyck, Bastien Kindt, Andrea Schmidt, avec la collaboration de Nicolas Atas ; *alignement* : Jean-Claude Haelewyck, Bastien Kindt. Pour le *Discours* 13 : SCHMIDT 2002 (*Versio nova* [S2], p. 3-15); lemmatisation : Naima Afif, Jean-Claude Haelewyck, Bastien Kindt, avec la collaboration de Nicolas Atas ; *alignement* : Jean-Claude Haelewyck, Bastien Kindt.

¹¹ Sur les versions syriaque des *Discours* de Grégoire de Nazianze, cf. HAELEWYCK 2017b, p. 75-124.

¹² Les différents *Discours* ne sont pas tous attestés dans ces trois états. Ainsi, pour le *Discours* 13, la version Sm n'est pas attestée.

¹³ Voir la bibliographie citée aux notes 2 et 4.

une *forme*, mais aussi un *lemme* ou encore tous les *lemmes* relevant d'une *catégorie morphosyntaxique* donnée, ou, enfin, une combinaison de ces différents éléments. Les réponses correspondant à ces recherches s'affichent à l'écran, instantanément, sous la forme d'une concordance énumérant la totalité des *occurrences* répondant à la requête formulée par l'utilisateur. Dans ces concordances, chaque *forme* est accompagnée de sa référence dans l'édition utilisée et encadrée d'un large contexte.

La même approche est utilisée pour l'analyse du syriaque, moyennant d'indispensables adaptations dues aux particularités de cette langue par rapport au grec. Deux particularités peuvent être citées d'emblée : l'abondance de *formes* suffixées et préfixées, et la présence d'une unité lexicale placée hiérarchiquement au-dessus du *lemme*, à savoir la *racine* sémitique des mots. Ces éléments du lexique doivent être définis.

2.1. Les formes

Les textes *traités* correspondent aux textes *édités*, considérés comme *textes de base*. Dans la mesure du possible, un *texte de base* est la version numérique d'un texte ayant fait l'objet, comme c'est ici le cas, d'une édition critique. Les *occurrences* sans vocalisation ou avec une vocalisation par lettres grecques (ܝܘܢܝܘܢ ou ܝܘܢܝܘܢ/Jean) ou par points (ܝܘܢܝܘܢ ou ܝܘܢܝܘܢ/roi) conservent la graphie qui est la leur dans le *texte de base*.

Les *occurrences* constituées de plusieurs *formes* – réunissant dans une seule unité graphique plusieurs unités lexicales, y compris les éléments préfixés et suffixés – sont traitées de manière à rendre possible l'identification des différentes *formes* concernées. Ainsi, l'*occurrence* ܝܘܢܝܘܢ/et dans l'air (= grec *καὶ εἰς ἀέρα*) est-elle traitée par programme et réécrite en trois *formes* distinctes ܝܘܢܝܘܢ@ܝܘܢܝܘܢ@ܝܘܢܝܘܢ, chacune séparée par le signe « @ », ce qui rend explicite, dans la suite des traitements, la continuité des *lexèmes* en présence, à savoir : (1) ܝܘܢܝܘܢ, la particule conjonctive ; (2) ܝܘܢܝܘܢ, la particule prépositionnelle ; (3) ܝܘܢܝܘܢ, le nom. La représentation des *lemmes*, des *catégories morphosyntaxiques* et des racines en présence dans cette *occurrence* repose sur le même principe : (1) *lemmes* ܝܘܢܝܘܢ@ܝܘܢܝܘܢ@ܝܘܢܝܘܢ ; (2) *catégories* NOUN@PART@PART ; (3) *racines* ܝܘܢܝܘܢ@ܝܘܢܝܘܢ@ܝܘܢܝܘܢ.

Ce type d'analyse est directement inspiré du traitement des *crases* du grec. Dans une *crase*, deux *lexèmes* sont réunis dans une seule *occurrence*, une seule *unité graphique*. L'*occurrence* *καὶ κεῖνος* est donc réécrite *καὶ@ἐκεῖνος* qui reçoit comme *lemmes* *καὶ@ἐκεῖνος* et comme *analyses morphosyntaxiques* « I+Part@PRO+Dem », une particule (I+Part) suivie d'un démonstratif (PRO+Dem)¹⁴. La différence entre le grec et le syriaque est cependant sensible ; elle tient en deux points essentiels. D'abord, les *formes* préfixées et suffixées du syriaque sont quantitativement beaucoup plus représentées que les *crases* du grec¹⁵. Ensuite, les *crases* du grec sont toujours constituées de deux *formes*, alors qu'en syriaque, le nombre de *formes* réunies en une seule *unité graphique* peut être plus grand. Dans le *corpus* traité, plusieurs *occurrences* réunissent jusqu'à quatre éléments différents, comme par exemple ܝܘܢܝܘܢ/et pour ma pusillanimité (grec *τῆς ἐμῆς δειλίας*), segmenté en ܝܘܢܝܘܢ@ܝܘܢܝܘܢ@ܝܘܢܝܘܢ@ܝܘܢܝܘܢ, séquence identifiant les quatre *formes* en présence, elles-mêmes classées sous les quatre *lemmes* attendus ܝܘܢܝܘܢ@ܝܘܢܝܘܢ@ܝܘܢܝܘܢ@ܝܘܢܝܘܢ, eux-mêmes répertoriés sous les racines ܝܘܢܝܘܢ@ܝܘܢܝܘܢ@ܝܘܢܝܘܢ.

¹⁴ Chaque langue traitée dans le cadre du projet GREgORI à un jeu d'étiquette qui lui est propre. Les étiquettes utilisées pour l'analyse du syriaque peuvent donc être différentes de celle du grec.

¹⁵ En grec, deux *crases* dans le *Discours* 1 (*καὶ κεῖνος* et *καὶ με*) et deux *occurrences* de *καὶ ν* dans le *Discours* 13. Dans le *corpus* syriaque, 778 *occurrences* (sur un total de 1956) sont préfixées ou suffixées.

Il est important de noter que ces modifications ne s'appliquent qu'au texte traité et ne modifient en aucun cas la version originale du texte édité. Dans les concordances et autres outils lexicographiques, les *occurrences* sont reprises telles qu'elles apparaissent dans les textes originaux. Par contre, elles sont reprises sous chacun des *lemmes* dont elles relèvent. Dans une concordance, l'*occurrence* ܝܪܘܘܢ figure donc à la fois sous le *lemme* ܝܪܘܘܢ, sous le *lemme* ܘܘܢ et sous le *lemme* ܝܘܢ.

Certaines *occurrences* considérées comme figées sont analysées comme des *formes* simples. C'est le cas de ܘܘܢܘܢ/seul, seulement (et non ܘܘܢܘܢܘܢ) ou de ܘܘܢܘܢܘܢ/ensuite (et non ܘܘܢܘܢܘܢܘܢ). Ces cas, peu nombreux, seront systématiquement inventoriés dans un document à l'usage des collaborateurs.

2.2. Les lemmes

Le *lemme* peut se définir comme une « forme de départ » ou une « forme de regroupement » dont découlent les *formes* (fléchies) relevant d'un même paradigme ; le *lemme* représente ce paradigme. Il est aussi « une forme neutralisée du discours métalinguistique » servant à la description du lexique d'une langue ; en théorie, mais pas obligatoirement, le *lemme* correspond à une *forme* attestée¹⁶.

Par convention, en syriaque, l'intitulé des *lemmes* est formulé à l'*état emphatique* puisque celui-ci s'est généralisé au détriment de l'*état absolu* : la *forme* ܕܝܘܢ est classée sous le *lemme* ܕܝܘܢ/dieu (*état emphatique*) et non ܝܘܢ (*état absolu*)¹⁷.

Même si en syriaque les noms et les adjectifs relèvent de paradigmes flexionnels identiques, ils sont rationalisés sous deux lemmes différents : nom et adjectif. Ainsi la *forme* ܕܝܘܢܘܢ/maison (en grec οἶκος) est lemmatisée ܕܝܘܢܘܢ, un nom, tandis que la *forme* ܕܝܘܢܘܢܘܢ/domestique, familier (en grec οἰκεῖος) est lemmatisée ܕܝܘܢܘܢܘܢ, un adjectif.

Les *formes* au pluriel ont, par convention, un *lemme* au singulier. Mais les *plurale tantum* conservent un *lemme* au pluriel. Dans l'expression ܕܝܘܢܘܢܘܢ/des cieux (grec τῶν οὐρανῶν), la *forme* ܕܝܘܢܘܢܘܢ a pour *lemme* ܕܝܘܢܘܢܘܢ, un pluriel, marqué par les seyomé.

Les *formes* des composés véritablement lexicalisés en syriaque sont classées sous un *lemme* unique. Ainsi les *formes* du composé ܕܝܘܢܘܢܘܢ/homme (littéralement *fil* d'homme) ne sont pas classées sous les deux *lemmes* ܕܝܘܢܘܢ/fils et ܕܝܘܢܘܢ/homme, mais sous l'unique *lemme* ܕܝܘܢܘܢ, véritablement lexicalisé en syriaque puisqu'en dérivent le verbe dénominal ܕܝܘܢܘܢܘܢ/devenir homme, s'incarner et le substantif ܕܝܘܢܘܢܘܢܘܢ/incarnation¹⁸. Il en va de même pour les *formes* du composé ܕܝܘܢܘܢܘܢ/adversaire classées sous l'unique *lemme* ܕܝܘܢܘܢܘܢ, dont dérivent l'adverbe ܕܝܘܢܘܢܘܢܘܢ/de manière hostile et le substantif ܕܝܘܢܘܢܘܢܘܢ/hostilité. D'autres composés, qui n'ont pas été lexicalisés en syriaque, tels que ܕܝܘܢܘܢܘܢ/toute chose et ܕܝܘܢܘܢܘܢ/toujours, seront classés respectivement sous ܕܝܘܢܘܢ/tout et ܕܝܘܢܘܢ/quelque chose et sous ܕܝܘܢܘܢ/tout et ܕܝܘܢܘܢ/temps.

¹⁶ Pour une définition du *lemme*, cf. KINDT 2004, p. 220-221, § 1-3.

¹⁷ Le *lemme* du mot ܕܝܘܢܘܢ/face (grec πρόσωπον) est formulé à l'*état emphatique* (ܕܝܘܢܘܢ), mais pas celui du mot ܝܪܘܘܢ/air (grec ἀήρ), ce qui obligerait à créer un *lemme* artificiel ܝܪܘܘܢܘܢ. En réalité, les mots grecs qui ne finissent pas en -ο n'ont pas de variations morphologiques pour les états ; cette *forme* ܝܪܘܘܢ peut donc être à l'*état absolu*, construit ou *emphatique*.

¹⁸ La *forme* au pluriel ܕܝܘܢܘܢܘܢ, écrite en deux mots, est analysée comme deux unités lexicales distinctes : le *lemme* ܕܝܘܢ et le *lemme* ܘܘܢܘܢ.

Les formes féminines des adjectifs reçoivent, toujours par convention, un lemme masculin. Ainsi, la forme ܩܘܠܘܢܐ est-elle classée sous le lemme ܩܘܠܘܢܐ/redoutable, terrible ; la forme ܩܘܠܘܢܐ/veuve est classée sous le lemme ܩܘܠܘܢܐ/veuf. En revanche, le mot ܩܘܠܘܢܐ/femelle qui, par nature, n'a pas de masculin, est laissé au féminin.

Les verbes à troisième radicale faible reçoivent une racine à troisième yod, à l'exception des six verbes troisième aleph qui maintiennent l'aleph comme gutturale¹⁹. Pour le lemme de ces verbes à 3^e yod, on ne fait pas la distinction entre verbes d'action (3^e aleph au pe'al parfait) et verbes d'état (en général 3^e yod au pe'al parfait). La forme avec 3^e 'aleph servira donc systématiquement de lemme, ainsi ܐܘܪܝܢܐ et non ܐܘܪܝܢܐ « être pur ».

En cas de graphie multiple, la graphie la plus habituelle sert de lemme. Les formes ܥܘܬܐ et ܥܘܬܐ sont classées sous le lemme unique ܥܘܬܐ ; ܥܘܬܐ et ܥܘܬܐ sous ܥܘܬܐ ; ܥܘܬܐ et ܥܘܬܐ sous ܥܘܬܐ. Quand la statistique ne permet pas de déterminer l'intitulé du lemme, on suit le dictionnaire de référence, en l'occurrence le *Syriac Lexicon* de M. Sokoloff ; les formes ܩܘܠܘܢܐ et ܩܘܠܘܢܐ/arche sont ainsi réunies sous le lemme ܩܘܠܘܢܐ. Il peut arriver que certains mots ne soient pas enregistrés dans les dictionnaires usuels de la langue syriaque ; dans ce cas, le nouveau lemme est créé et consigné dans une liste de mots dits « athesauristus ».

Les lemmes homographes – même si un ou plusieurs des lexèmes qu'ils représentent n'est ou ne sont pas attestés dans le texte traité – sont distingués par les points diacritiques, par la vocalisation par points ou par la *linea occultans*, comme dans les exemples suivants :

ܐܘܪܝܢܐ/ou, ou bien vs ܐܘܪܝܢܐ/ô vs ܐܘܪܝܢܐ/oh !, hélas ! ; ܐܘܪܝܢܐ (pronom personnel indépendant et suffixé) vs ܐܘܪܝܢܐ (pronom démonstratif) ; ܩܘܠܘܢܐ/tribut, impôt vs ܩܘܠܘܢܐ/percepteur ; ܩܘܠܘܢܐ/roi vs ܩܘܠܘܢܐ/conseil ; ܩܘܠܘܢܐ/pleurs vs ܩܘܠܘܢܐ/poule ; ܩܘܠܘܢܐ(interjection virulente) vs ܩܘܠܘܢܐ/alors ; ܩܘܠܘܢܐ/or (métal) vs ܩܘܠܘܢܐ/orfèvre ; ܩܘܠܘܢܐ/certes (l'emprunt au grec μέν) vs ܩܘܠܘܢܐ/qui ?, quel ? (et ses différents emplois) vs ܩܘܠܘܢܐ/à partir de.

Pour assurer un traitement informatique homogène de ces mots, il est indispensable de vérifier que, dans les textes traités, un point-voyelle ne soit pas noté pour un point diacritique, ou inversement, et de respecter le standard d'encodage UNICODE²⁰, tel qu'illustré dans le tableau 3.

| | Exemples | | Codage UNICODE |
|-------------------------------------|----------|--------|----------------|
| Point-voyelle supérieur (MAJ+B) | ܐ | ܐܘܪܝܢܐ | U+073F |
| Point-voyelle inférieur (MAJ+V) | ܐ | ܐܘܪܝܢܐ | U+073C |
| Point-diacritique supérieur (MAJ+P) | ܐ | ܐܘܪܝܢܐ | U+0307 |
| Point-diacritique inférieur (MAJ+M) | ܐ | ܐܘܪܝܢܐ | U+0327 |

Tableau 3. Codage Unicode des points-voyelles et des points-diacritiques

2.3. Les catégories morphosyntaxiques

Chaque lemme reçoit une étiquette indiquant la catégorie morphosyntaxique, la partie du discours, dont il relève. Cette étiquette caractérise le lemme au niveau du lexique en général.

¹⁹ Il s'agit des verbes suivants : ܩܘܠܘܢܐ/consoler, ܩܘܠܘܢܐ ou ܩܘܠܘܢܐ/décorer ou réjouir, ܩܘܠܘܢܐ/rendre impur, ܩܘܠܘܢܐ/orner, ܩܘܠܘܢܐ/souiller et ܩܘܠܘܢܐ/enseigner ; ils sont cités dans DUVAL 1881, § 214.

²⁰ Pour un complément d'information, cf. le site du Consortium Unicode accessible sous l'adresse <http://unicode.org/charts/> ainsi que le guide d'utilisation des polices de caractères Meltho rédigé par George Kiraz et accessible sur le site de la fondation Beth Mardutho sous l'adresse <http://www.bethmardutho.org/index.php/resources/fonts.html>.

Elle ne décrit en aucun cas l'emploi particulier d'un *lemme* dans un contexte précis. Ainsi, dans l'expression suivante :

ܠܝܟܢ ܕܠܝܟܢ ܕܠܝܟܢ/le temps de m'examiner (grec ὅσον ἐμαυτὸν ἐπισκέψασθαι) — Or. I (S1), § 2, l. 1-2 (= HAELEWYCK 2011, p. 4),

la forme ܕܠܝܟܢ/âme – incluse dans l'occurrence ܕܠܝܟܢ – reste classée sous le *lemme* homographe ܕܠܝܟܢ caractérisé comme nom, même si, dans ce contexte, elle rend, en français, l'équivalent d'un pronom réfléchi. De même, dans la phrase suivante :

ܠܝܟܢ ܕܠܝܟܢ ܕܠܝܟܢ ܕܠܝܟܢ ܕܠܝܟܢ/comme Moïse et Jérémie ensuite (grec ὡς Μωϋσῆς καὶ Ἱερεμίας ὕστερον) — Or. I (S2), § 1, l. 9-10 (= HAELEWYCK, p. 5),

la séquence ܕܠܝܟܢ ܕܠܝܟܢ, qui fonctionne comme une conjonction, est analysée comme la suite des *lemmes* ܕܠܝܟܢ, ܕܠܝܟܢ et ܕܠܝܟܢ, caractérisés respectivement comme particule, pronom interrogatif et particule.

Dans l'état actuel du projet, les *étiquettes* distinguent douze *catégories morpho-syntaxiques* : **ADJ**, les adjectifs ; **ADV**, les adverbes ; **CARD**, les déterminants numériques cardinaux ; **NAME**, les noms propres ; **NOUN**, les substantifs ; **ORD**, les déterminants numériques ordinaux ; **PART**, les particules (regroupant les négations, les prépositions et d'autres mots invariables) ; **PRO_dem**, les pronoms démonstratifs ; **PRO_ind**, les pronoms indéfinis ; **PRO_int**, les pronoms interrogatifs ; **PRO_pers**, les pronoms personnels indépendants et suffixés ; et enfin **V1** à **V28**, les verbes. Ces catégories, classées selon l'ordre alphabétique des *étiquettes*, sont explicitées ci-dessous.

2.3.1. ADJ = Les adjectifs

ܠܝܟܢ ܕܠܝܟܢ ܕܠܝܟܢ ܕܠܝܟܢ ܕܠܝܟܢ ܕܠܝܟܢ/les voici pour vous : Abraham le vénérable, ce patriarche, cette tête précieuse et respectée (grec ταῦτα μὲν ὑμῖν ὁ σεμνὸς Ἀβραὰμ οὗτος ὁ πατριάρχης ἡ τιμία κεφαλὴ καὶ αἰδέσιμος) — Or. 1 (S2), § 7, l. 1 (= HAELEWYCK 2011, p. 11)

Les trois adjectifs ܕܠܝܟܢ/vénérable, ܕܠܝܟܢ/précieux et ܕܠܝܟܢ/respecté, respectable reçoivent l'étiquette ADJ²¹.

2.3.2. ADV = Les adverbes

ܠܝܟܢ ܕܠܝܟܢ ܕܠܝܟܢ [ܕܠܝܟܢ]/des paroles qui ne s'effacent pas facilement (grec [λόγους] οὐδὲ ῥαδίως ἀπαλειφομένους) — Or. I, 6 (= p. 10, l. 16) — Or. I (S1), § 6, l. 16 (= HAELEWYCK 2011, p. 10)

Les adverbes reçoivent l'étiquette ADV ; cf., par exemple, ܕܠܝܟܢ/facilement ; ܕܠܝܟܢ/maintenant ; ܕܠܝܟܢ/volontiers, facilement. Le *lemme* ܕܠܝܟܢ/amen (grec ἀμήν) est enregistré comme adverbe.

²¹ De nombreux adjectifs sont à l'origine des participes. C'est le cas des trois adjectifs cités ici : ܕܠܝܟܢ et ܕܠܝܟܢ sont des participes pe'al passifs et ܕܠܝܟܢ un participe pa'el passif. Ils sont pourtant classés sous l'étiquette ADJ. On suit en cela le *Syriac Lexicon* de M. Sokoloff (respectivement aux p. 617, 634 et 755) et le *Dictionnaire syriaque-Français* de L. Costaz (p. 144, 154 et 158). Le *Thesaurus Syriacus* (col. 1624, 1719) classe les deux premiers comme participes, et le troisième (col. 1765) comme adjectif. Le *Compendious Syriac Dictionary* de J. Payne Smith ne tranche pas pour le premier et le troisième en les désignant comme « part. adj. » (p. 196, 218), tandis que le second est vu comme un participe (p. 212).

substantiv, Primäradjektiv) et (2) les noms et adjectifs relevant d'une racine verbale. Dans le premier cas, c'est le nom ou l'adjectif qui est originel ; dans le second, le mot est dérivé d'une racine verbale. La plupart des *noms premiers* relèvent des domaines suivants : la famille (père, frère, etc.), les parties du corps (cœur, pied, etc.), les animaux (taureau, lapin, etc.), les plantes (herbe, ail, etc.), les pierres et les métaux (argile, or, etc.), les concepts temporels (année, éternité, etc.), géographiques (terre, fleuve, etc.), architecturaux (digue, toit, etc.), les ustensiles (corbeille, vase, etc.), et les mots étrangers (palais, paradis, etc.). Ils peuvent eux-mêmes être à l'origine de verbes qui sont alors qualifiés de *verbes dénominatifs*, généralement employés à une *forme* dérivée. La prise en compte de ces *noms premiers* dispense de créer des racines verbales artificielles censées les expliquer. Les *noms premiers* sont ainsi classés sous un *lemme* et une *racine* identiques, toujours formulés à l'état *emphatique*, soit au masculin, soit au féminin. Ainsi le mot ܐܘܪ/père est classé sous les *lemme* et *racine* ܐܘܪ, ܠܘܠܘܢ/chien sous les *lemme* et *racine* ܠܘܠܘܢ (et non ܠܘܠܘܢ), ܐܘܪܘܢ/nourrir sous ܐܘܪܘܢ/nourriture (féminin) et non sous une forme tap'el d'une racine ܐܘܪ, comme conjecturé par le *Compendious Syriac Dictionary* ou le *Thesaurus Syriacus*, ܐܘܪܘܢ/devenir pauvre sous ܐܘܪܘܢ/pauvre, et non sous une forme etmap'al d'une racine ܐܘܪ, comme conjecturé dans les mêmes ouvrages de consultations. Les dictionnaires, en particulier celui de M. Sokoloff, rendent ici de précieux services. Chaque entrée lexicale est suivie entre parenthèses de l'indication de la racine verbale qui a servi à créer le mot ; on sait ainsi que l'adjectif ܐܘܪܘܢ/malade vient de la racine ܐܘܪ ; et sous ܐܘܪ sont notés tous les dérivés de cette racine. Mais, dans bon nombre de cas, le verbe est clairement identifié comme verbe dénominatif, ainsi le verbe ܐܘܪܘܢ/former, façonner (utilisé au pa'el et au etpa'al) a été créé à partir du substantif ܐܘܪܘܢ/image ; c'est alors ce dernier qui apparaît comme racine. Cette façon de procéder, qui peut dérouter l'utilisateur, nous paraît plus respectueuse de l'histoire de la langue. L'utilisateur de la concordance en ligne devra dès lors privilégier la recherche par *lemme* ; ce n'est que dans un second temps, quand il aura identifié la *racine* dont relève mot, qu'il pourra entreprendre une recherche par *racine* pour connaître tous les *lemmes* (et les *formes*) qui en dépendent.

Les racines homographes sont distinguées par un numéro, comme le fait M. Sokoloff. Ainsi, le *lemme* ܐܘܪ est classé sous la racine 4 ܐܘܪ/rester.

Dans les concordances et les listes lexicographiques, les mots empruntés à une autre langue que le syriaque sont accompagnés du signe « ① » (*loanword, emprunt*) : ܐܘܪ < grec ἀήρ/air ; ܐܘܪ < grec εἰκῆ/en vain ; ܐܘܪ < grec γάρ/en effet ; ܐܘܪ < grec ληστής/brigand ; ܐܘܪ < v. pers. *rāza-/mystère. Le tableau 5 dresse la liste des différents emprunts lexicaux rencontrés jusqu'ici dans nos données.

En cas de doute sur l'identification d'un mot comme emprunt, ou sur la langue à laquelle le mot est emprunté, nous suivons M. Sokoloff ou des ouvrages de référence faisant autorité²⁷. C'est par exemple le cas pour la particule ܐܘܪ. Le *Syriac Lexicon* de M. Sokoloff (p. 296) se borne à évoquer l'analogie avec le grec δέ quand la particule est placée en deuxième position dans un membre de phrase. Pour A. Butts, l'analogie est certes sémantique, syntaxique et phonologique, mais les deux mots doivent bien être distingués, puisque le ܐܘܪ syriaque est issu de l'araméen *'iḏayn, et non du grec δέ ; le *lemme* syriaque n'est donc pas caractérisé ici comme un emprunt²⁸. *A contrario*, le *lemme* ܐܘܪ est bien considéré comme un emprunt au grec μέν et sa racine reçoit l'indication « ① ».

²⁷ Sur les emprunts au grec, cf. BUTTS 2016, p. 43-63 ; sur les emprunts aux langues iraniennes, cf. CIANCAGLINI 2008, p. 95-268.

²⁸ BUTTS 2016, p. 174-194.

les informaticiens du CENTAL, permettent d'extraire de ces bases de données les informations nécessaires pour répondre à une requête³⁰. Ces informations sont organisées et affichées selon des mises en pages prédéfinies permettant de les visualiser dans les outils lexicographiques, à savoir des concordances, des index ou des listes alphabétiques ou fréquentielles. Les requêtes peuvent porter sur l'ensemble du texte ou sur une portion du texte. Elles peuvent aussi concerner un ou plusieurs éléments des données lemmatisées, une *forme*, un *lemme*, une *catégorie morphosyntaxique*, une *racine*.

Les pages qui suivent illustrent ces différents types de sortie, ici au format ADOBE PDF : une concordance grecque (figure 4), une concordance syriaque (figure 5), deux concordances bilingues grec-syriaque (figures 6 et 7), une liste bilingue grec-syriaque des emplois du *lemme* θεός (figure 8), une liste bilingue syriaque-grec du *lemme* ܪܘܠܪ (figure 9) et enfin une concordance bilingue syriaque-grec des *lemmes* caractérisés comme noms propres (figure 10).

| Gregory of Nazianzus - GRC - Or. 1 and 13 | | Ἀερῶν {N+Ant} |
|----------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| Ἀερῶν {N+Ant} 1 Or. 1 13 35 396 A 15 | καὶ Ἱερραίας ὑπερον, καὶ προσδραμεῖν ἑτοιμῶς καλοῦντι, ὡς Ἀερῶν τε καὶ Ἡσαΐας, μόνον εὐσεβῶς ἀμάρτερα, τὸ μὲν διὰ τὴν οικεῖαν ἀσθένειαν, τὸ δὲ διὰ | |
| ἀβασάνιστος {A} 1 Or. 13 4 35 856 B 4 | κληρονομεῖς μὴ θαυμάσης· οὐδὲν τῶν μεγάλων ἀδόκιμον οὐδὲν ἀβασάνιστον. Ἐπειτα γὰρ φύσει τοῖς μὲν ταπεινοῖς τὸ ῥάδιον τοῖς δὲ ὑψηλοῖς τὸ δυσκήτηον. | |
| Ἀβερῶν {N+Ant} 1 Or. 13 3 35 853 D 1 | Τὶ φήσῃ ὡ παῖ Δαθάν καὶ Ἀβερῶν καὶ στρατηγὲ ἀσωφρόνιστε ὁ κατὰ Μωυσέως τολμήσας καὶ χεῖρας ἐπαφείς ἡμῖν | |
| Ἀβραάμ {N+Ant} 1 Or. 1 17 35 400 C 8 | Ταῦτα μὲν οὖν ἰμὴν ὁ σεμνὸς Ἀβραάμ οὗτος ὁ πατριάρχης ἢ τῆμα κεφαλὴ καὶ αἰδέσιμος τὸ πάντων τῶν καλῶν καταγώνων ὁ | |
| ἀγαθός {A} 5 Or. 1 13 35 396 A 12 Or. 12 35 397 A 3 Or. 13 4 35 856 A 13 Or. 1 5 35 397 C 6 Or. 1 13 35 396 A 11 | ὡς τάχα γε κρείττων αὐτῆ καὶ τιμιώτερα Θεῷ τῆς ἑτέρων ταχυτήτος· ἀγαθὸν γὰρ καὶ ὑποχωρήσαι Θεῷ τὴ μικρόν, ὡς Μωυσῆς ἐκεῖνος τὸ παλαιόν, καὶ Ἱερραίας ἄνθρωπον, δῶ τῆ καινῆ κτίσει τοῖς κατὰ Θεὸν γεννωμένοις πλάστην ἀγαθὸν καὶ διδάσκαλον Χριστῷ καὶ συννεκροῦμενον προθύμως καὶ συνανιστάμενον. Νῦν δὲ ὡ ποιμένων ἄριστε καὶ τελεώτατε δευρό μοι καὶ τὸν σὸν ἀπολάμβανε λαὸν σὺν ἡμῖν τε καὶ πρὸ ἡμῶν ὄν ἐπειδὴ κάκενος δι' ἡμᾶς ἄνθρωπος. Προσέλαβε τὸ χεῖρον ἵνα δῶ τὸ βέλτιον· ἐπιώχουσε ἵν' ἡμεῖς τῆ ἐκείνου πτωχεῖα πλουτήσωμεν· δουλοῦ μορφῆν ἔλαβεν ἵνα τυραννήσαντες, εἴ τί μοι μέμνησθε τῆς βραδυτήτος, ὡς τάχα γε κρείττων αὐτῆ καὶ τιμιώτερα Θεῷ τῆς ἑτέρων ταχυτήτος· ἀγαθὸν γὰρ καὶ ὑποχωρήσαι Θεῷ τὴ | |
| ἀγαλλίασις {N+Com} 1 Or. 13 4 35 856 B 11 | Ἐ τοῦ θαύματος! Τὸ ἐσπέρας ἠλίσιθη κλαυθμὸς καὶ εἰς τὸ πρωῖ ἀγαλλίασις. Ἐα ληρῆν τοὺς πολεμοῦντας καὶ περιχάσκον ὡς κύνας ὑλακτοῦντας διακενής· | |
| ἀγάπη {N+Com} 1 Or. 1 13 35 396 A 6 | εἴπωμεν, ἀδελφοί, καὶ τοῖς μισοῦσιν ἡμᾶς, μὴ ὅτι τοῖς δι' ἀγάπην τὴ πεποιθήσαν, ἢ πεπονθῶσι· συγχωρήσωμεν πάντα τῆ ἀναστάσει· δῶμεν συγγνώμην | |
| ἄγγελος {N+Com} 1 Or. 13 4 35 856 B 1 | καὶ πρὸ ἡμῶν ὄν ἐνεχειρίσει σοὶ τὸ Πνεῦμα τὸ ἅγιον ὄν προσάγουσιν ἄγγελοι ὄν ὁ βίος καὶ ὁ τρόπος ἐπίστευσε. Εἰ δὲ διὰ πειρασμῶν καὶ κωλυμάτων τὸν θρόνον | |
| ἄγιος {A} 5 Or. 13 4 35 856 C 8 Or. 13 4 35 856 B 15 Or. 13 2 35 853 C 8 Or. 13 4 35 856 A 15 Or. 1 17 35 401 A 5 | σβέσαι καὶ παραστήσαι τῷ Κυρίῳ λαὸν περισούσιον ἔθνους ἄγιον βασιλεῖον ἱεράτευμα ἐν Χριστῷ Ἰησοῦ τῷ Κυρίῳ ἡμῶν ὃ ἡ δοξὰ εἰς τοὺς αἰῶνας. Ἀμην. προσκυνεῖν Θεὸν τὸν Πατέρα Θεὸν τὸν Υἱὸν Θεὸν τὸ Πνεῦμα τὸ ἅγιον ἐν τρισὶν ὑποστάσεσιν ἐν μιᾷ δόξῃ τε καὶ λαμπρότητι. Τὸ ἀπολωλὸς ἐκζητεῖ· τὸ ἀσθενὲς προκαθεζόμενον πόλεως· τίμιον ἴσμεν κεφαλὴν γινώσκον ἄγιον καλοῦμεν καὶ ἀδοκίμοι· μόνον ἔστω φιλότεκνος καὶ πάσης προνοῦν Ἐκκλησίας· λαὸν σὺν ἡμῖν τε καὶ πρὸ ἡμῶν ὄν ἐνεχειρίσει σοὶ τὸ Πνεῦμα τὸ ἅγιον ὄν προσάγουσιν ἄγγελοι ὄν ὁ βίος καὶ ὁ τρόπος ἐπίστευσε. Εἰ δὲ διὰ πειρασμῶν καὶ ἀπὸ μὲν τῆς ὑγιούς πίστεως ἀπαγούσης τῆς εἰς Πατέρα καὶ Υἱὸν καὶ ἄγιον Πνεῦμα τὴν μίαν θεότητά τε καὶ δύναμιν ἧς ἦκουσεν ἀεὶ φωνῆς καὶ ἀκούει τὰ ἐμὰ | |
| ἀδελφός {N+Com} 2 Or. 1 13 35 396 A 5 Or. 13 1 35 852 A 3 | τῆ πανηγύρει, καὶ ἀλλήλους περιπτυσζόμεθα· εἴπωμεν, ἀδελφοί, καὶ τοῖς μισοῦσιν ἡμᾶς, μὴ ὅτι τοῖς δι' ἀγάπην τὴ πεποιθήσαν, ἢ πεπονθῶσι· Δέξασθε τὸν ἡμέτερον λόγον ἀδελφοί· κἂν ἐλάχιστος ἢ κἂν παρὰ πολὺ τῆς ἀξίας λεπτόμενος· πλὴν ὅμως σταθίζον οἶδεν ὁ | |
| ἀδικέω {V} 1 Or. 13 2 35 853 C 8 | πόλεως· τίμιον ἴσμεν κεφαλὴν γινώσκον ἄγιον καλοῦμεν καὶ ἀδοκίμοι· μόνον ἔστω φιλότεκνος καὶ πάσης προνοῦν Ἐκκλησίας. Προσθήκη ἱερέων | |
| ἀδόκιμος {A} 1 | | |

Figure 4. Première page de la concordance lemmatisée du corpus grec

³⁰ Ces outils utilisent, en l'occurrence, SAP Crystal Reports, un logiciel de représentation graphique de données.

4. L'interface en ligne

L'interface en ligne, déjà opérationnelle dans une version expérimentale, permet d'interroger les *corpus* sur un ordinateur personnel. Nous nous limitons ici à décrire l'interface permettant d'interroger le texte grec et la version syriaque (S1) du *Discours 1*. Visuellement, cette interface est volontairement simple. Un premier registre identifie, à gauche, l'auteur (Grégoire de Nazianze) et le texte interrogés (le *Discours 1*). Il présente aussi, à droite, deux boutons : (1) une flèche vers le bas inscrite dans un cercle (figure 11) propose un menu permettant de télécharger des concordances et des outils lexicographiques – document décrits dans la Section 3, ci-dessus – au format ADOBE PDF et (2) la lettre « i » inscrite dans un cercle (figure 12), ouvre une fenêtre contextuelle (pop-up window) indiquant les références bibliographiques précises des éditions suivies (le *texte de base*), identifiant les personnes responsables de la lemmatisation et de l'*alignement* des textes, et signalant les institutions impliquées dans le traitement du *corpus* en question.

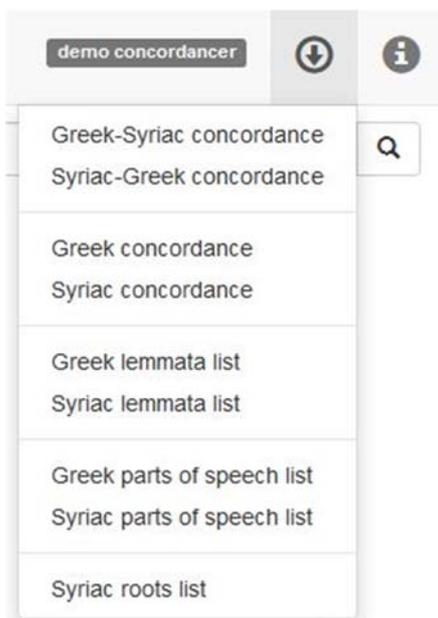


Figure 11. Affichage de la liste des documents téléchargeable au format ADOBE PDF

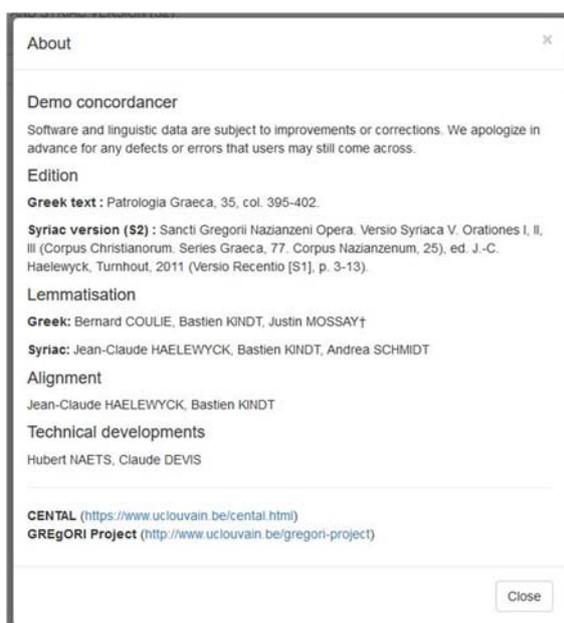


Figure 12. La fenêtre contextuelle décrivant le corpus exploré

Le second registre constitue le champ d’interrogation. À gauche, un menu permet de sélectionner la langue interrogée, le grec du *texte source* ou le syriaque du *texte cible* (ici, S1). La requête peut être une *forme* – ܠܡܠܟܐ (figure 13) –, un *lemme* – ܠܡܠܟܐܐ (figure 14), une *racine* – ܠܡܠܟ (figure 15) –, une *catégorie morphosyntaxique* – par exemple les verbes Etpa‘al (V4) (figure 16) – ou une combinaison de ces trois éléments – par exemple les séquences constituées d’une *forme* d’un *lemme* verbal Pe‘al (« V4 ») suivi d’une *forme* relevant de la *racine* ܠܡܠܟ (figure 17). Les *occurrences* du texte syriaque sont analysées en tenant compte des éléments préfixés et suffixés. Il est possible de formuler une requête cherchant toutes les *occurrences* constituées des particules ܐ et ܕ suivies d’un *lemme* catégorisé comme nom. Cette recherche fournit la concordance des *formes* ܠܡܠܟܐܐ, ܠܡܠܟܐܐܐ, ܠܡܠܟܐܐܐܐ et ܠܡܠܟܐܐܐܐܐܐ (figure 18).

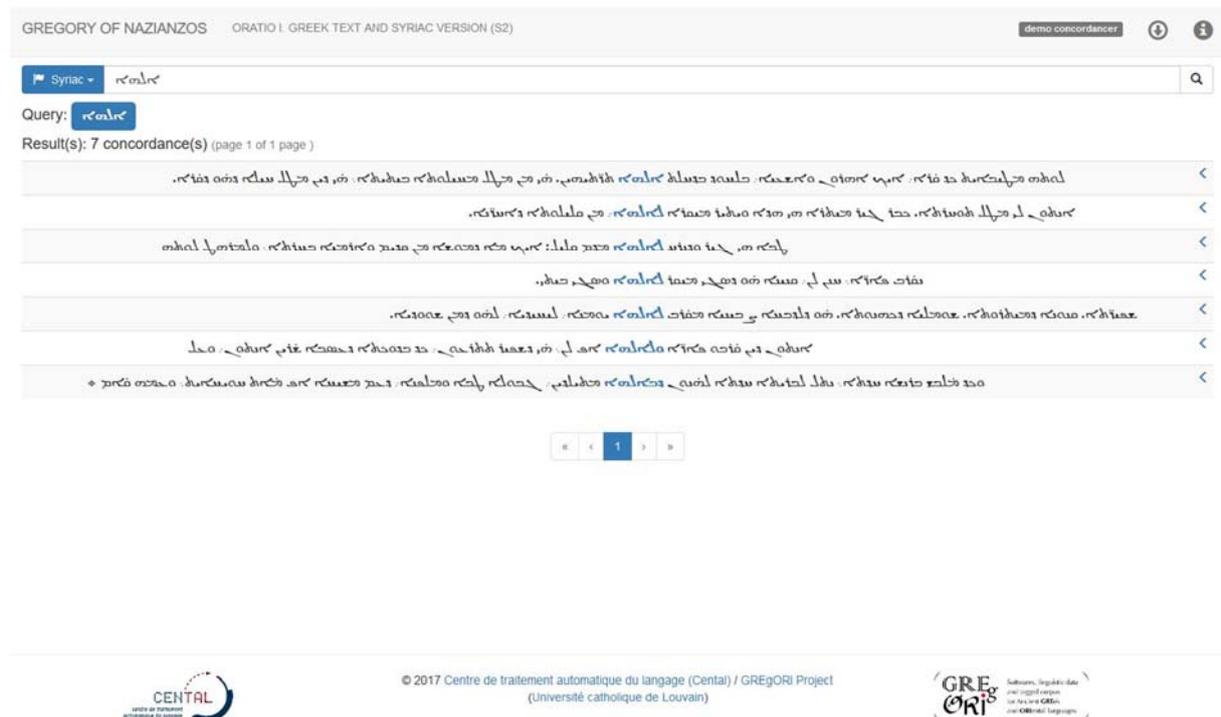


Figure 13. Recherche des occurrences de la forme ܠܡܠܟܐ

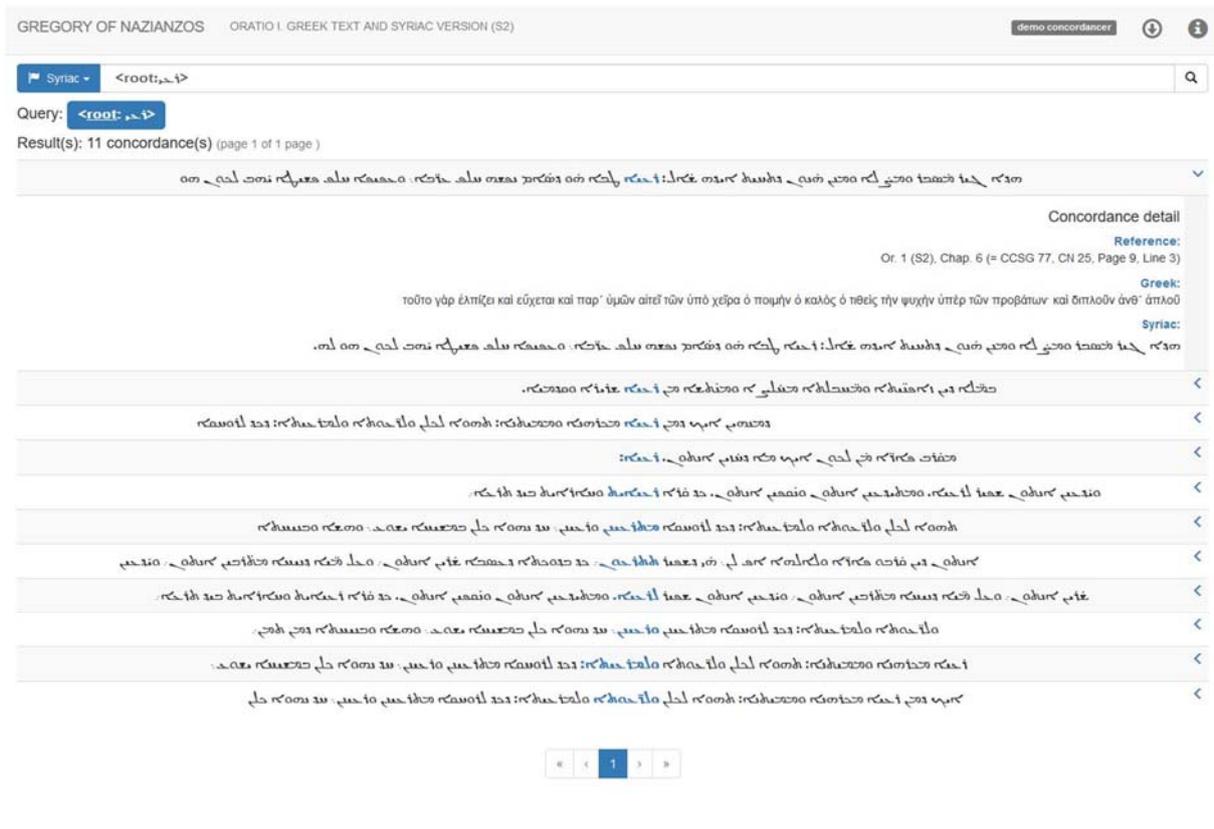


Figure 19. Affichage du contexte étendu (« Concordance details ») d’une occurrence

5. Conclusions et perspectives

D’autres textes sont d’ores et déjà en cours de traitement. A.B. Schmidt et G. Abousamra travaillent à l’édition et à l’analyse lexicale de textes magiques. J.-Cl. Haelewyck et B. Kindt assurent le traitement des trois versions syriaques de l’*Historia Zosimi de Vita Beatorum Rechabitarum*. B. Kindt et N. Atas ont préparé une concordance bilingue (syriaque – anglais) de l’*Oratio Manasseh*, sur base de données lexicales produites par le Peshitta Institute - Eep Talstra Centre for Bible and Computer (Amsterdam). N. Atas a analysé le lexique des chapitres I-X de la *Genèse*, en syriaque et dans leur version en turoyo³¹. A.S. Sembiante, J.-Cl. Haelewyck et B. Kindt ont lemmatisé la version syriaque du premier poème dogmatique de Grégoire de Nazianze ; Y. Arzhanov collabore avec A.B. Schmidt et B. Kindt pour produire la lemmatisation de la version syriaque des *Sentences des philosophes grecs*, un texte qui paraîtra dans le *Corpus Scriptorum Christianorum Orientalium*.

Ces travaux permettront d’éprouver les principes d’analyse présentés dans cet article et contribueront à enrichir quantitativement et qualitativement les corpus et les données linguistiques. Ils serviront aussi à constituer, progressivement, un dictionnaire de lemmatisation dédié à l’analyse de la langue syriaque, comme cela existe déjà pour le grec et d’autres langues de l’Orient chrétien.

³¹ Ce travail a jeté les bases du traitement automatique du turoyo, cf. ATAS 2018. Les concordances syriaque et turoyo de Gen. I-X sont mises à la disposition des chercheurs sur le site du projet GREgORI.

Les interfaces accessibles en ligne offrent un accès aux textes par l'intermédiaire des concordances. Les éléments présents dans ces concordances correspondent à des questions spécifiques posées par l'utilisateur. Ces questions, appelées « requêtes », peuvent porter sur les *racines* (dans le cas des langues sémitiques), sur les *lemmes*, sur les *formes*, sur les *catégories* grammaticales, c'est-à-dire sur toutes les informations linguistiques attachées aux mots présents dans le corpus exploré³². Les requêtes peuvent également être constituées d'une combinaison de ces différents éléments. Une fois la requête lancée, la concordance s'affiche instantanément à l'écran.

Ces concordances sont dites « dynamiques ». Les concordances imprimées sur papier, et diffusées sous forme de livres, sont quant à elle des concordances dites « statiques ». Ces concordances-livres, qui existent depuis longtemps, tant en grec qu'en syriaque, seront amenées à être remplacées par d'autres outils, tels que ceux décrits dans ces lignes. Bien entendu, ces ouvrages ont rendu d'importants services aux chercheurs³³. Mais leur diffusion sur papier (ou, dans certains cas, sur microfiches) constitue une contrainte matérielle importante. Cette contrainte interdit de corriger les données ou de les adapter, d'une part, et offre peu de possibilités de recherche, d'autre part. Les limites physiques de la page du livre obligent les auteurs et les éditeurs à limiter drastiquement la longueur des contextes. Le *pivot*, le mot-clé théoriquement encadré de contextes en amont et en aval, est systématiquement une *forme de mot* et jamais une expression. Il est impossible de faire varier le tri des *racines*, des *lemmes*, des *formes* et des contextes en fonction de l'ordre alphabétique ou de l'ordre de leur apparition dans les textes, selon les besoins de l'utilisateur. Enfin, au niveau théorique, chaque publication repose sur des règles de dépouillement et d'analyse qui lui sont propres³⁴ et, à quelques rares exceptions près³⁵, ces principes demeurent largement implicites³⁶.

Les efforts en matière d'alignement des *textes sources* sur les différents *textes cibles* disponibles ouvrent des perspectives nouvelles. L'automatisation des processus d'analyse permet

³² Il est évidemment possible d'enrichir davantage la nature et le nombre des analyses linguistiques proposées. En grec, par exemple, certains développements utilisent les analyses flexionnelles (cas, genre, nombre, voix, mode, temps, personne, etc.), mais, de manière générale, les travaux du projet GREgORI s'arrêtent volontairement au niveau de l'analyse lexicale.

³³ Sans prétendre ici à l'exhaustivité, on peut citer, pour le grec, la concordance de la *Septante* réalisée par E. Hatch et H.A. Redpath (première édition 1897), ouvrage qui a longtemps fait figure de modèle. Pour les auteurs classiques, il faut mentionner, à titre d'exemple, les concordances publiées par la maison d'édition Olms-Weidmann dans la collection des *Lexika, Indizes, Konkordanzen zur Klassischen Philologie* (cf. <http://www.olms.de/search/result.aspx?SeriesID=002>). Pour les sources patristiques et byzantines, il faut épingler les volumes du *Thesaurus Patrum Graecorum*, une collection de concordances lemmatisées, jadis produites par les promoteurs du projet GREgORI et diffusées par Brepols Publishers (voir la liste des titres parus sur le site du projet, cité note 3 ; dans ce dernier cas, les concordances étaient publiées sur des microfiches). En syriaque, on se doit de citer les ouvrages de AUFRECHT, HURD 1975 ; KIRAZ 1993 ; WINTER 1976 ; BORBONE *et alii* 1997 ; LUND, KIRAZ 2004.

³⁴ Dans la collection des *Lexika, Indizes, Konkordanzen*, diffusée par Olms-Weidmann, le même lexème ἀβουλία reçoit le lemme ἀβουλία dans le volume consacré à Thucydide, mais le lemme ἀβουλίη dans celui consacré à Hérodote. C'est donc la langue des auteurs qui est décrite (ionio-attique pour le premier, ionienne pour le second) et non la langue grecque en général ; sur cette distinction, cf. KINDT 2004, p. 224, § 6. Dans le cadre du projet GREgORI, l'analyse se situe toujours au niveau de la description générale de la langue.

³⁵ Voir, par exemple, la concordance du *Pentateuque* dans la *Peshitta* publiée par P.G. Borbone, J. Cook, K.D. Jenner et D.M. Walter.

³⁶ La *Concordance to the Syriac New Testament* de G.A. Kiraz commence *ex abrupto*, sans aucune explication relative aux textes suivis ou aux principes d'analyse adoptés.

de travailler sur de vastes corpus. Elle procure des résultats tangibles, susceptibles d'être utiles aux recherches dans différents domaines comme l'étude des méthodes de traduction ou la linguistique contrastive, par exemple.

L'accès en ligne permet d'offrir aux chercheurs un corpus destiné à s'accroître au fil du temps. Le *Digital Syriac Corpus* (DSC) vient d'ouvrir la voie en ce domaine³⁷. Les développements réalisés dans le cadre du projet GREgORI s'inscrivent dans la même démarche, mais en tâchant de répondre à trois exigences supplémentaires : (1) les textes sont munis d'un étiquetage lexical (lemmes et racines) et morphosyntaxique (les catégories) abouti ; (2) le résultat de l'interrogation des textes s'affiche sous la forme de concordances à partir desquelles l'utilisateur peut accéder aux textes ; (3) les *textes sources* sont alignés sur leurs *textes cibles*.

L'expérience acquise dans le domaine de l'analyse du syriaque permet aux promoteurs du projet GREgORI, et à leurs collaborateurs, de s'atteler désormais à l'analyse de deux ensembles textuels majeurs, la version syriaque des homélies de Grégoire de Nazianze publiée dans le *Corpus Nazianzenum*, une sous-série du *Corpus Christianorum* publié par Brepols Publishers (103.305 occurrences), et celui des œuvres d'Éphrem de Nisibe publiées par E. Beck dans les volumes du CSCO (301.623 occurrences).

³⁷ Cf. <https://syriacorporus.org>.

BIBLIOGRAPHIE

- Anonyme 1985 : *The concordance to the Peshitta version of the aramaic New Testament*, New Knoxville.
- ARZHANOV 2018 : Y. ARZHANOV, *Syriac Sayings of Greek Philosophers. A Study in Syriac Gnomologia. With Edition and Translation*, Leuven (à paraître dans la série des *Subsidia du Corpus Scriptorum Orientalium Christianorum*).
- ATAS 2018 : N. ATAS, *Édition et étude philologique du manuscrit turoyo Sachau 249 de la Staatsbibliothek de Berlin*, Mémoire de Master en langues et lettres anciennes, orientation orientales, Université catholique de Louvain, Louvain-la-Neuve, 2017-2018 (inédit).
- AUFRECHT, HURD 1975 : W.E. AUFRECHT, J.C. HURD, *A Synoptic Concordance of Aramaic Inscriptions according to H. Donner and W. Röllig* (The International Concordance Library, 1), Missoula.
- BORBONE *et alii* 1997 : P.G. BORBONE, J. COOK, K.D. JENNER, D.M. WALTER, in collaboration with J.A. LUND, M.P. WEITZMAN, *The Pentateuch* (The Old Testament in Syriac according to the Peshitta. Part V. Concordance. Volume 1), Leiden, New York, Cologne.
- BUTTS 2016 : A.M. BUTTS, *Language Change in the Wake of Empire. Syriac in Its Greco-Roman Context* (Linguistic Studies in Ancient West Semitic, 11), Winona Lake.
- CIANCAGLINI 2008 : C.A. CIANCAGLINI, *Iranian Loanwords in Syriac* (Beiträge zur Iranistik, 28), Wiesbaden.
- COSTAZ 2002 : L. COSTAZ, *Dictionnaire Syriaque-Français. Syriac-English Dictionary*, 3^e ed., Beyrouth.
- COULIE 1996 : B. COULIE, « La lemmatisation des textes grecs et byzantins : une approche particulière de la langue et des auteurs », *Byzantion*, 66, p. 35-54.
- COULIE *et alii* 2013 : B. COULIE, B. KINDT, T. PATARIDZE, « Lemmatisation automatique des sources en géorgien ancien », *Le Muséon*, 126, p. 161-201.
- DUVAL 1881 : R. DUVAL, *Traité de grammaire syriaque*, Paris (réimpr. Piscataway, 2010).
- FLEURY 2012 : S. FLEURY, *mkAlign, Manuel d'utilisation*, Université Sorbonne Nouvelle Paris 3 [[pdf](#)].
- HAELEWYCK 1998 : J.-Cl. HAELEWYCK, *Clavis Apocryphorum Veteris Testamenti* (Corpus Christianorum), Turnhout.
- HAELEWYCK 2011 : J.-Cl. HAELEWYCK, *Sancti Gregorii Nazianzeni Opera. Versio Syriaca. V. Orationes I, II, III* (Corpus Christianorum. Series Graeca, 77 ; Corpus Nazianzenum, 25), Turnhout.
- HAELEWYCK 2014 : J.-Cl. HAELEWYCK, « *Historia Zosimi de Vita Beatorum Rechabitarum*. Édition de la version syriaque brève », *Le Muséon*, 127, p. 95-147.
- HAELEWYCK 2015 : J.-Cl. HAELEWYCK, « La version syriaque longue de l'*Historia Zosimi de Vita Beatorum Rechabitarum*. Édition et traduction », *Le Muséon*, 128, p. 295-379.
- HAELEWYCK 2017a : J.-Cl. HAELEWYCK, « *Histoire de Zosime sur la vie des Bienheureux Réchabites*. Les trois recensions syriaques. Édition de la version résumée », *Parole de l'Orient*, 43, p. 175-194.

- HAELEWYCK 2017b : J.-Cl. HAELEWYCK, « Les versions syriaques des Discours de Grégoire de Nazianze : un processus continu de révision », *BABELAO*, 6, p. 75-124 [pdf].
- HATCH, REDPATH 1998 : E. HATCH, H.A. REDPATH, *A Concordance to the Septuagint and the Other Versions of the Old Testament (Including the Apocryphal Books)*, 2^e éd., Grand Rapids.
- KIENAST 2001 : B. KIENAST, *Historische semitische Sprachwissenschaft*, Wiesbaden.
- KINDT 2004 : B. KINDT, « La lemmatisation des sources patristiques et byzantines au service d'une description lexicale du grec ancien. Les principes de formulation des lemmes du Dictionnaire Automatique Grec », *Byzantion*, 74, p. 213-272 [pdf].
- KINDT 2018 : B. KINDT, « Processing Tools for Greek and Other Languages of the Christian Middle East », dans *jdmdh:4184 - Journal of Data Mining & Digital Humanities*, January 6, 2018. Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages [pdf].
- KIRAZ 1993 : G.A. KIRAZ, *A Computer-Generated Concordance to the Syriac New Testament*, 6 vol., Leiden, New York, Cologne.
- KIRAZ 2001 : G.A. KIRAZ, *MELTHO : Syriac OpenType Fonts for Windows XP/2000™ and Windows 95/98/ME™*, Piscataway, 2000-2001 [pdf].
- LEFER à paraître : M.-A. LEFER, « Parallel Corpora », à paraître dans S. GRIES, M. Paquot (éd.), *Practical Handbook of Corpus Linguistics*, New York.
- LIPÍŃSKI 2000 : E. LIPÍŃSKI, *Semitic Languages: Outline of a Comparative Grammar* (Orientalia Lovaniensia Analecta, 80), Leuven.
- LUND 2007 : A.J. LUND, *The Book of the laws of the countries: a dialogue on free will versus fate : a key-word-in-context concordance*, Piscataway.
- LUND, KIRAZ 2004 : J.A. LUND, G.A. KIRAZ, *The Old Syriac Gospel of the Distinct Evangelists: a Key-word-in-context Concordance*, Piscataway.
- MOSCATI 1980 : S. MOSCATI, *An Introduction to the Comparative Grammar of the Semitic Languages. Phonology and Morphology* (Porta Linguarum Orientalium, NS 6), Wiesbaden.
- MOSSAY *et alii* 1990 : J. MOSSAY, CETEDOC, *Thesaurus sancti Gregorii Nazianzeni*, vol. I. *Enumeratio lemmatum, orationes, epistulae, testamentum* (Corpus Christianorum. Thesaurus Patrum Graecorum), Turnhout.
- MOSSAY *et alii* 1991 : J. MOSSAY, B. COULIE, CETEDOC, *Thesaurus sancti Gregorii Nazianzeni*, vol. II. *Enumeratio lemmatum, carmina, christus patiens, vita* (Corpus Christianorum. Thesaurus Patrum Graecorum), Turnhout.
- NÖLDEKE 1898 : Th. NÖLDEKE, *Kurzgefasste syrische Grammatik*, Leipzig (réimpr. Darmstadt, 2015).
- PATARIDZE, KINDT 2018 : T. PATARIDZE, B. KINDT, « Text Alignment in Ancient Greek and Georgian: A Case-Study on the First Homily of Gregory of Nazianzus », dans *jdmdh:4182 - Journal of Data Mining & Digital Humanities*, January 6, 2018. Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages [pdf].
- PAYNE SMITH 1967 : J. PAYNE SMITH, *A Compendious Syriac Dictionary, founded upon the Thesaurus Syriacus of R. Payne Smith*, Oxford, 3^e ed.

- PAYNE SMITH 1901 : R. PAYNE SMITH, *Thesaurus Syriacus*, Oxford, 1879-1901 (réimpr. Hildesheim, New York, 2006).
- SCHMIDT 2002 : A.B. SCHMIDT, *Sancti Gregorii Nazianzeni Opera. Versio Syriaca. II. Orationes XIII, XLI* (Corpus Christianorum. Series Graeca, 47 ; Corpus Nazianzenum, 15), Turnhout.
- SEMBIANTE 2017 : A.S. SEMBIANTE, « Appunti sulla tradizione siriana delle opere di Gregorio Nazianzeno », *Koininia* 10, p. 607-635.
- SOKOLOFF 2009 : M. SOKOLOFF, *A Syriac Lexicon. A Translation from the Latin, Correction, Expansion, and Update of C. Brockelmann's Lexicon Syriacum*, Winona Lake, Piscataway.
- TUERLINCKX 2004 : L. TUERLINCKX, « La lemmatisation de l'arabe non classique », dans A. DISTER, C. FAIRON, G. PURNELLE (éd.), *Le poids des mots. 7^{es} Journées internationales d'Analyse statistique des Données Textuelles, 10-12 mars 2004, Louvain-la-Neuve*, Louvain-la-Neuve, vol. II, p. 1069-1078 [[pdf](#)].
- VAN ELVERDINGHE 2018 : E. VAN ELVERDINGHE, « Recurrent Pattern Modelling in a Corpus of Armenian Manuscript Colophons », dans *jdmdh:4183 – Journal of Data Mining & Digital Humanities*, January 6, 2018, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages [[pdf](#)].
- WINTER 1976 : M.M. WINTER, *A Concordance to the Peshitta Version of Ben Sira* (Monographs of the Peshitta Institute Leiden, 2), Leiden.