

## LE TAL POUR LES APPELLATIONS D'ŒUVRES FIGURÉES DE L'ANTIQUITÉ CLASSIQUE: ÉVOLUTION DES RESSOURCES NUMÉRIQUES DU PROJET MONUMENTAL

### 1. INTRODUCTION

Dans l'environnement numérique, le titre de l'œuvre d'art figurée trouve sa place dans la recherche d'information, que celle-ci soit une requête formulée par l'internaute ou un dialogue machine-machine. Ce nom joue un rôle dans l'identification même de l'œuvre ou du sujet de l'image, notamment dans l'affichage des réponses, dans le set d'informations dédié à l'œuvre et donc, le cas échéant, dans ses métadonnées documentaires. Citer une œuvre de façon distinctive suppose aussi de la désigner sous un titre, une appellation univoque.

Le programme MonumenTAL/Monuments antiques et Traitement Automatique de la Langue<sup>1</sup> porte précisément sur les titres et désignations des œuvres de l'art figuré de l'Antiquité classique, qu'il aborde à partir des textes et en se fondant sur des pratiques relevant du Traitement Automatique de la Langue (TAL), pour aborder l'étude de l'élaboration et de l'évolution des modes de formulation de ces désignations dans une approche diachronique. Fondé sur une collaboration interdisciplinaire réunissant les historiens de l'art et philologues de l'équipe française du LIMC (UMR7041 Archéologies et Sciences de l'Antiquité/ArScAn) – spécialistes de l'iconographie classique – les linguistes-talistes de l'UMR7114 Modèles Dynamique Corpus/MoDyCo, et les conservateurs et bibliothécaires de la Bibliothèque nationale de France (BnF), ce projet commencé en 2020 vise à développer des moyens pour la reconnaissance et l'annotation automatiques des noms d'œuvres d'art visuel dans les textes et à expérimenter des pratiques dans la perspective de l'Intelligence Artificielle. Par sa thématique, il est également étroitement lié à l'enrichissement des ressources en ligne LIMC-France et de son thésaurus TheA/Thesaurus-Antiquité (<http://www.limc-france.fr/>). Cet article présente une synthèse, sous l'angle de vue de l'historien de l'art et archéologue, centrée sur l'outil de reconnaissance et d'annotation textuelles automatiques réalisé de façon collégiale au sein du programme (LESSIEUX *et al.* 2022, 2023), en l'état actuel de son évolution, puis quelques premiers essais d'utilisation des données produites pour des analyses, de la fouille de texte ou de l'apprentissage profond.

<sup>1</sup> MonumenTAL (<https://heuristic.huma-num.fr/heuristic/?db=MonumenTAL&website&id=186>) a été réalisé dans le cadre du labex Les passés dans le présent et a bénéficié de l'aide de l'État gérée par l'ANR au titre du programme Investissements d'avenir portant la réf. ANR-11-LABX-0026-01 (<http://passes-present.eu/fr/monumental-monuments-antiques-et-traitement-automatique-de-la-langue-44335>).

## 2. SPÉCIFICITÉS DU NOM DES ŒUVRES FIGURÉES ANTIQUES

Les données du web, de plus en plus massives grâce à l'accroissement des ressources numériques nouvelles comme celles des bases de données, ou à la mise en ligne sous forme numérique de publications nativement numériques ou numérisées, y compris les plus anciennes, restituent un large éventail de noms attribués aux œuvres d'art, leurs différents titres et les variantes de ceux-ci. En revanche, des thésaurus du domaine sont encore souvent limités pour les titres d'œuvres antiques ou répondent plutôt aux besoins des historiens de l'art européen plus récent. Dans ce contexte, un moyen de repérer les œuvres antiques par leur titre, à l'aide d'un thésaurus dédié ou par des processus automatisés ou relevant des techniques de l'Intelligence Artificielle, serait bienvenu.

L'attribution d'un titre distinctif, voire unique, comme le ferait aujourd'hui le créateur de l'œuvre d'art figurée, est une pratique relativement récente, qui remonte aux XVIII<sup>e</sup>-XIX<sup>e</sup> s. (DE BIASI *et al.* 2012), alors que pour les siècles précédents et notamment pour la période antique, les témoignages qui nous sont parvenus proviennent quasiment de textes dans lesquels les œuvres sont citées selon des expressions variables créées par les auteurs ou la tradition (PRIOUX à paraître). Faute d'un processus de désignation imposant un nom unique, autre que l'attribution d'un numéro d'inventaire par la collection en charge de l'objet (MINISTÈRE DE LA CULTURE 2020), la plupart des œuvres figurées de l'Antiquité classique sont connues sous plusieurs appellations différentes élaborées et attribuées au fil du temps, selon les pratiques et le contexte d'écriture ou d'usage du moment, certaines appellations seulement se stabilisant.

De plus, un même nom d'œuvre connue peut aussi être employé pour désigner d'autres objets figurés. Le phénomène des copies et répliques d'époque romaine, en particulier, et les créations plus récentes inspirées par les sculptures antiques redécouvertes depuis la période moderne, correspondent à une reprise de types iconographiques. Dans la pratique, ces derniers sont nommés comme l'œuvre-archétype, ce qui fragilise encore davantage l'unicité souhaitée de la relation "titre-œuvre". Par exemple, d'innombrables copies de la statue de Praxitèle l'*Aphrodite de Cnide* sont désignées de la même façon, comme le fait F. Cumont qui titre ainsi une notice de catalogue de musée avant de préciser à la page suivante qu'il s'agit d'une réplique de la statue de Praxitèle (CUMONT 1913, 15-16 n. 10: Fig. 1). La répétition, pendant des siècles, de compositions et de schémas visuels a aussi contribué à instaurer des noms de thèmes iconographiques et de sujets narratifs, par exemple *Héraclès combattant le lion*, comme titres de multiples images ou objets figurés. Cette reprise de motifs, caractéristique de l'art antique classique, complique encore la distinction entre les artefacts.

Face à cette diversité dans les usages, le terme "appellation" a été préféré par le projet au mot "titre", ressenti comme trop restrictif par rapport au

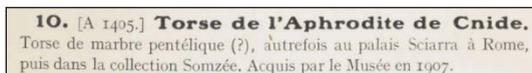


Fig. 1 – Extrait de CUMONT 1913, p.15.

souhait de prendre en compte les thèmes iconographiques si fréquemment utilisés pour nommer des images, des scènes, de même que les signalements un peu flous d'artefacts figurés comme *statue de Poséidon* qui ont un fort potentiel pour l'interopérabilité des données du patrimoine antique. Les appellations thématiques ont été traitées de façon expérimentale car leur formulation peut aussi correspondre à un segment textuel dans une description détaillée d'image sans que ce soit pour autant un titre, une confusion qui générerait du bruit parmi les résultats souhaités.

### 3. UN CORPUS TEXTUEL SPÉCIALISÉ POUR UNE APPROCHE DIACHRONIQUE

Les créations figurées produites dans les mondes anciens grecs, romains ou étrusques ont ainsi été citées de diverses façons dans les textes, de l'Antiquité à nos jours, par des mentions simples témoignant de leur existence ou comme sujet d'étude ou élément d'élaboration de théories relatives aux cultures visuelles ou matérielles. Pour repérer cette variété, développer les moyens de la reconnaissance et de l'annotation automatiques, et poser les bases de l'étude de l'évolution des formulations, le corpus textuel français choisi cible des écrits d'historiens et théoriciens de l'art antique et d'archéologues et couvre la période allant du XVIII<sup>e</sup> s. à nos jours. Il inclut aussi des traductions de textes antiques originellement rédigés en grec ancien ou en latin. Les publications consacrées à la sculpture ont été utilisées en premier puis le corpus a été progressivement élargi aux textes sur les gemmes gravées, sur la peinture, la mosaïque et la céramique.

Les soixante-quinze publications en français sélectionnées pour les développements et les tests totalisent 883.740 tokens (un token correspond à un mot, une unité, et la tokenisation à un découpage des phrases en mots) et s'inscrivent dans la typologie suivante:

- Des essais, des manuels, des traductions de textes antiques grecs ou latins,
- Des catalogues de collection ou de vente (textes avec des notices structurées), des guides de musées et des récits de voyage,
- Des articles et des fascicules de périodique,
- Des notices d'encyclopédie,
- Des listes, des lexiques, des index.

Des œuvres visuelles antiques sont déjà citées dans des ouvrages des XVI<sup>e</sup> et XVII<sup>e</sup> s. mais les publications du corpus textuel ne sont pas antérieures

au XVIII<sup>e</sup> s. car les processus de TAL du projet utilisent les textes en version TXT; or, la fiabilité de l'océrisation des ouvrages originaux anciens est trop diminuée par la qualité d'impression irrégulière et les typographies anciennes. Des textes présentant un taux raisonnable d'erreurs d'océrisation ont été pris en compte car cette situation correspond à l'une des utilisations envisagées pour l'outil de TAL: le repérage des œuvres dans les publications disponibles en ligne, y compris les plus volumineuses. La mise en page et la casse, le cas échéant porteuses de sens pour les titres – comme l'emplacement dans une notice de catalogue d'art semi-structuré ou le titre en italique ou en gras – ne sont malheureusement pas conservées à l'identique par le processus d'océrisation et n'ont donc pas pu être exploitées.

Afin de faciliter la reproductibilité des procédures réalisées, les publications en open access ou en open data ont été privilégiées et sélectionnées principalement sur Gallica (<https://gallica.bnf.fr/>) et, pour les plus récentes, sur OpenEdition (<https://www.openedition.org/>), Persée (<https://www.persee.fr/>), Hal (<https://hal.science/>), Wikipédia et Wikisource (<https://fr.wikipedia.org> et <https://fr.wikisource.org/>), ces sites de ressources en ligne proposant l'accès sur le long terme à un large panel de publications. Les fichiers TXT sont ceux téléchargés sur Gallica; ceux des autres sources sont générés sur le moment. L'ensemble des fichiers TXT, y compris les TXT originaux, les données produites par le projet et sa documentation sont déposés sur un espace GitLab de l'IR\* Huma-Num (<https://gitlab.huma-num.fr/monumental/monumental>). Une instance Zotero a été créée pour gérer les références des publications utilisées, des métadonnées créées pour le protocole d'analyse (type et période de publication, thématique, lectorat, état d'océrisation, etc.), ainsi qu'un identifiant univoque propre à MonumenTAL assurant le sourçage, à chaque étape du projet, de toutes les versions modifiées du fichier TXT original et des segments textuels qui en sont extraits.

#### 4. CHOIX DE LA MÉTHODE DE TRAITEMENT AUTOMATIQUE DE LA LANGUE

Deux types de formulations des appellations d'œuvres visuelles ont d'abord été envisagés: une suite traditionnelle, plus ou moins longue, combinant des noms et des déterminants, comme l'*Aphrodite de Cnide* ou la *fameuse statue en marbre de l'Aphrodite de Cnide de Praxitèle*, et le thème iconographique ou sujet narratif, par exemple *Héraclès combattant le lion* ou exprimé en un seul mot spécifique au domaine comme *Amazonomachie* (combat avec des Amazones).

Après le mode de formulation, le deuxième constat est celui de la présence récurrente de noms de personnages représentés et d'acteurs du monde de l'art (artiste, collectionneur, commanditaire), ainsi que celui de notions omniprésentes dans les données, les ontologies (notamment le CIDOC CRM: <https://>

cidoc-crm.org/) et les thésaurus relatifs au patrimoine culturel matériel, telles le type d'artefact, le matériau, le lieu, la collection: *statue en marbre de l'Aphrodite de Cnide de Praxitèle* = type d'artefact + matériau + personnage figuré + lieu + artiste. Beaucoup de ces termes sont propres au domaine et à la période artistiques traités et sont donc absents des dictionnaires standards français utilisés habituellement par les outils numériques de TAL. Certains mots dérivent du grec ancien ou du latin et ont été francisés dans les textes selon des principes de transcription irréguliers (par exemple pour Icare: Ikaros, Icaros, Ikarus, Icarus).

Compte tenu de la récurrence de quelques types de formulation et de l'omniprésence de vocabulaires spécialisés, la méthode de TAL dite symbolique, et la recherche d'entités nommées (NOUVEL *et al.* 2016), ont été privilégiées pour repérer et annoter automatiquement les appellations d'œuvres figurées (LESSIEUX *et al.* 2022, § 6). Pour l'application de ces approches à d'autres thématiques: COLLIN, GUERRAZ 2015; DÍEZ PLATAS *et al.* 2021). Il s'agit de faire interagir des patrons linguistiques (qui sont des modélisations des formulations textuelles), sous forme de graphes numériques, avec des dictionnaires (qui se présentent dans le cas présent comme des listes simples de termes). Pour MonumentAL, ces dictionnaires réunissent un vocabulaire spécifique complétant le dictionnaire standard français. Comparé à l'utilisation d'un thésaurus de titres d'œuvres, dont le contenu reste figé, la méthode de TAL choisie permet un repérage dynamique des formulations, c'est-à-dire quel que soit l'ordre ou la forme des composants de la formulation, les choix de transcription ou de graphie des mots, ou encore les éventuels changements d'identification des personnages et d'interprétation des scènes figurées.

## 5. CERNER LES FORMULATIONS MODÈLES: DE L'ANNOTATION MANUELLE À L'AUTOMATISATION

Le processus d'élaboration de cet outil de repérage et d'annotation a suivi deux phases: la constitution de la typologie des appellations et du vocabulaire à prendre en compte, puis la modélisation des graphes numériques et l'élaboration des dictionnaires pour le TAL.

Lors de la première phase, une annotation manuelle a été effectuée, en concertation, par plusieurs historiens de l'art et des linguistes-talistes sur une trentaine de publications en fichiers TXT ayant subi, via un script Python, une transformation préalable limitée à la suppression des sauts de ligne et espaces en surplus, et à la normalisation de caractères (tirets; guillemets). Les chiffres, la différenciation des majuscules et des minuscules, et des mots considérés comme vides par d'autres projets de TAL ont été conservés car estimés pertinents. Le logiciel gratuit d'annotation Glozz (<http://www.glozz.org/>; WIDLÖCHER, MATHET 2012; <http://explorationdecorpus.corpusecrits.huma-num.fr/glozz/>) a été choisi pour sa facilité d'utilisation (une annotation sous forme de

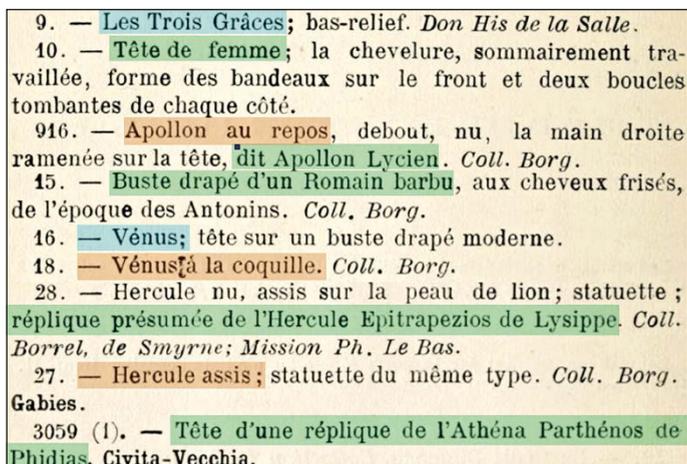


Fig. 2 – Extrait surligné de la p. 2 de MUSÉES NATIONAUX, LOUVRE (MICHON) 1922.

surlignages colorés selon les balises paramétrées par rapport aux besoins du projet) et d'extraction des segments annotés (appellations d'œuvres; mots du vocabulaire). Une "convention d'annotation", fixant les règles d'annotation décidées pour le projet, a été établie. La fiabilité de l'annotation a été évaluée (LESSIEUX *et al.* 2022, § 5) par le calcul de l'accord inter-annotateur (mesure de la similarité des décisions prises par différents annotateurs de type kappa de Cohen,  $\kappa$ : [https://fr.wikipedia.org/wiki/Kappa\\_de\\_Cohen](https://fr.wikipedia.org/wiki/Kappa_de_Cohen), consulté le 8.1.2024).

Cette phase a confirmé les deux types presentis – la suite traditionnelle de déterminants et de noms et l'expression du thème – qui trouvent un écho dans l'analyse des titres dans les sources antiques présentée par É. Prioux (PRIoux à paraître). L'annotation manuelle a aussi révélé un mode de formulation, très fréquent, consistant en une réduction de la suite traditionnelle au seul nom du personnage représenté précédé d'un déterminant, ainsi que des constructions du type *Vénus dite de Médicis*, ou *Aphrodite Callipyge, dite Vénus belle-fesse*.

Une page d'un catalogue du Louvre (MUSÉES NATIONAUX, LOUVRE (MICHON) 1922, 2: <https://gallica.bnf.fr/ark:/12148/bpt6k6237020f/f22.item>) donne quelques exemples pour la statuaire (Fig. 2).

## 6. EXPLOITER LES EXPRESSIONS SPÉCIFIQUES POUR ENRICHIR LA TYPOLOGIE DES APPELLATIONS EN FRANÇAIS

Le patron linguistique a été réalisé en version numérique avec le logiciel gratuit d'analyse de corpus Unitex (<https://unitexgramlab.org/fr>; PAUMIER *et al.* 2009). Il réunit quarante-six graphes (Fig. 3) dont un, principal

(1Appellation\_all.grf), qui appelle les autres graphes (sous-graphes) selon une architecture en cascade.

Sous l'effet de la modélisation en graphe numérique, qui a induit des adaptations par rapport à la phase d'annotation manuelle, la typologie des appellations a évolué et a été enrichie depuis la présentation faite à la Conférence internationale TOTH 2022 (LESSIEUX *et al.* 2023, 56-59). Les types principaux finalement retenus sont l'Appellation Courte (balise d'annotation AC) et l'Appellation Structurée (AS). Pour correspondre à la cible de MonumentAL – les désignations d'œuvres – le Thème (TM) est compris dans l'AS et n'est reconnu en tant que titre que sous certaines conditions. Enfin, deux types de formulation supplémentaires relèvent de L'Appellation Structurée: l'Appellation Enrichie (AE) et l'Appellation Introduite (AX).

L'interaction étroite des dictionnaires spécifiques avec les graphes, qui les appellent, et les premiers résultats d'annotation automatique ont également fait évoluer le jeu de dictionnaires par rapport à sa présentation en 2022 (LESSIEUX *et al.* 2023, 64-69): pour ces modifications, voir *infra* § 7. Dans cet article, par commodité, les mots composant les noms de la typologie des appellations commencent par une majuscule et sont abrégés comme leurs balises d'annotation, les noms de dictionnaire sont en majuscules. Dans les schémas de graphes Unitex, un nom de dictionnaire est en majuscules, sans surlignage et entre des chevrons; un nom de graphe est surligné en gris; une balise d'annotation est en gras et entre des chevrons.

Normalization	AS_SUPP2.grf	LIEU_all.grf	SUPPORT_all.grf
Preprocessing	AX_VRBArt.grf	LIEU_all_VIRG.grf	TM.grf
1Appellation_all.grf	BORNE.grf	MATERIAU_all.grf	TM_AS.grf
AC.grf	COLL_all.grf	MATERIAU_all_VIRG.grf	TMF.grf
AE_DIT.grf	COLL_all1.grf	OEUVRE.grf	TMF_APP.grf
APP.grf	DET_art.grf	PARENTHESE.grf	TMP.grf
ARTISTE_all.grf	DET_de.grf	PERS_AUTEUR_all.grf	TMP_APP.grf
AS_all.grf	DET_le.grf	PERSO_all.grf	TMS.grf
AS_IND.grf	ENUM.grf	PERSO_all_ADJ.grf	VIRG.grf
AS_SUJ.grf	ENUM_ADJ.grf	PERSO2_all.grf	VIRG_tt.grf
AS_SUJ2.grf	import_glozz.grf	PERSO2_all_ADJ.grf	VRB_Art.grf
AS_SUPP.grf	INV.grf	PREP_place.grf	Z_PLUS.grf

Fig. 3 – Liste des graphes Unitex pour les appellations d'œuvres de MonumentAL.

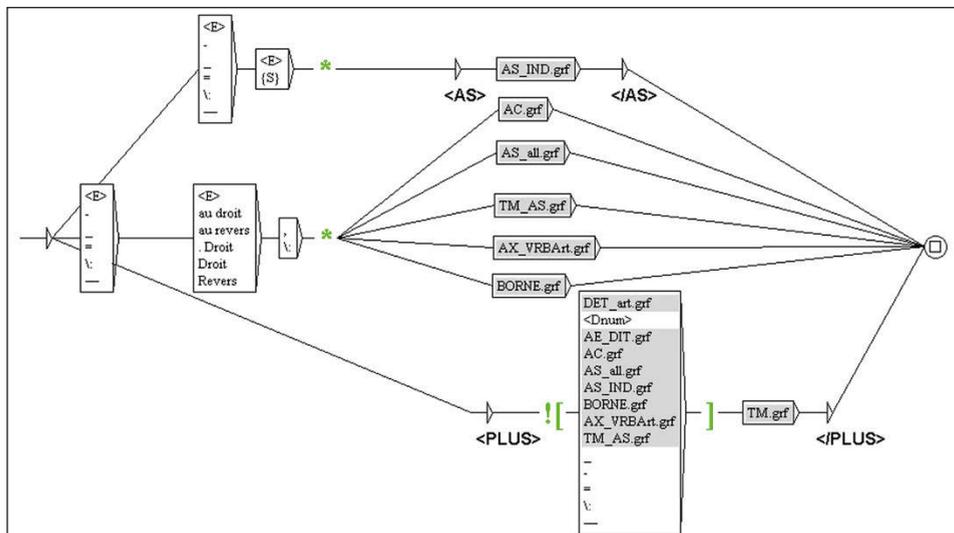


Fig. 4 – Schéma du graphe Unitex APP pour la récupération des titres selon la typologie MonumentTAL.

La Fig. 4 présente, dans le second niveau du graphe Unitex, les sous-graphes de la typologie appelés par le sous-graphe APP.grf.

### 6.1 L'Appellation Courte (AC)

L'AC (Fig. 2: surlignage bleu) est réduite au nom du personnage précédé d'un déterminant (n. 9: « *Les Trois Grâces* »). En français, cette combinaison suffit dans la plupart des cas à confirmer la matérialité de la représentation. Par exemple la forme réduite *la Vénus* suppose dans la plupart des cas un objet tandis que le seul nom *Vénus* peut correspondre à l'entité divine sans qu'il s'agisse d'une représentation figurée. Cette forme est employée comme rappel dans le texte (*la Médicis*: pour *la Vénus de Médicis*) ou pour des œuvres très célèbres (*le Laocoon*; *l'Apollon* [du Belvédère]; *la Cnidienne*; *la Samothrace*). Elle fonctionne particulièrement bien avec des noms propres d'individus mais manque de pertinence avec les noms de groupe, les noms génériques, qui sont donc écartés. Les *hermès* (piliers hermaïques) et leurs variantes, comme les « Hermathènes » (JAUCOURT 1765), qui sont à la fois objet et représentation figurée, et « les trois Grâces », expression employée surtout pour des œuvres visuelles, sont acceptés.

Les dénominations des personnages individualisés ou génériques (groupes, fonctions...) sont récupérées si elles sont délimitées par un jeu pertinent de ponctuations, comme dans une notice de catalogue (n. 16: « – *Vénus*; » ou « . *Satyre* . », etc.).

## 6.2 L'Appellation Structurée (AS), l'Appellation Enrichie (AE), l'Appellation Introduite par un verbe (AX)

L'AS (surlignage vert) commence par un personnage représenté ou un support d'image/type d'objet (n. 10; n. 15: « *Buste drapé d'un Romain barbu* ») complété, pour former une suite, d'au moins l'un des composants permutable et facultatifs correspondant aux grandes notions fréquemment convoquées pour le patrimoine matériel (n. 28; n. 3059: « *Tête d'une réplique de l'Athéna Parthénos de Phidias* »):

- Personnage: personnage représenté ou artiste,
- Type d'objet, de support ou de représentation (par ex. réplique, copie, image, etc.),
- Matériau,
- Lieu de création, d'utilisation, de découverte ou de conservation,
- Collection, collectionneur, édifice antique.

Les signalements des œuvres dans un index, dont l'ordre des composants peut être inversé (« *Milo, Vénus de* »), sont recherchés (Fig. 4: graphe AS\_IND).

L'AE (Appellation Enrichie), courante dans le corpus, est introduite par un composant comme “, dit”, “soi-disant” ou “dénommé” qui signale son emploi en titre d'œuvre (n. 916: « *dit Apollon Lycien* »). Elle peut avoir plusieurs structurations: composant + composant (*Vénus dite du Capitole*), composant + appellation (*Vénus dite Vénus du Capitole*) ou appellation + appellation (*Vénus du Capitole, dite Vénus Capitoline; Apollon Sauroctone appelé Apollon de Cleveland*). Pour simplifier à la fois l'évaluation de ses occurrences et l'architecture des graphes, l'AE est balisée au sein de l'AS, comme l'un de ses composants et non plus comme un type indépendant.

L'AX (Appellation Introduite) est la désignation d'une œuvre introduite par un verbe spécifique au domaine (sculpter, peindre...) ou très pertinent (Fig. 4: graphe AX\_VRBArt). Le verbe “représenter” est de loin le plus employé et ses résultats sont si bons que cette formulation a été retenue. Toutes les conjugaisons sont admises, y compris des formes anciennes telles que “représentoit” pour l'imparfait. Les verbes comme “dédier”, “ériger”, “consacrer” [... une statue à une divinité], ambigus, ont généré trop de bruit et ont été écartés. Par contre, l'approche diachronique a fait ressortir des usages anciens récurrents qui ont été retenus, comme un emploi du verbe “offrir” équivalent à “montrer”, “représenter”: « *offrant <AX> Diane sous les attributs d'Hécate</AX>* » (TIECK 1844, 4 n. 26). Le segment textuel annoté AX correspond au titre, à une AS.

La modélisation de l'AS permet aussi de repérer de façon dynamique des appellations traditionnelles dont une partie des mots est hors dictionnaire comme les noms de défunts inscrits sur les monuments funéraires, par ex. « *Stèle funéraire de Q. Clodius* » (MUSÉES NATIONAUX, LOUVRE (MICHON)

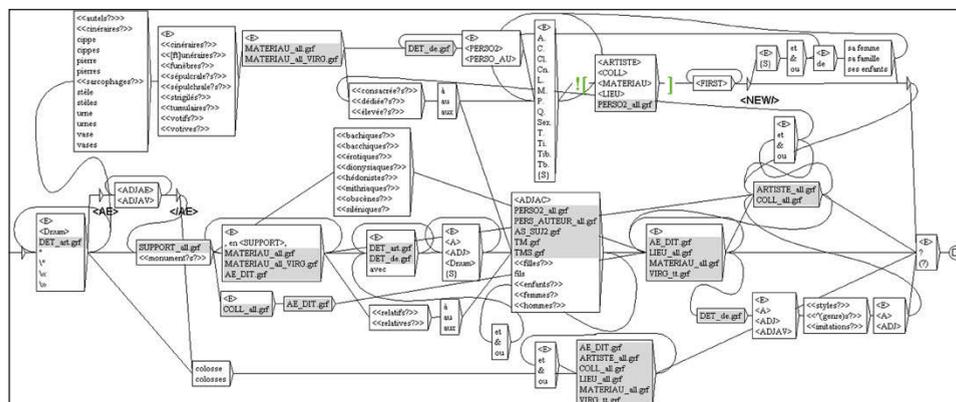


Fig. 5 – Graphe Unitex pour les Appellations Structurées de MonumentAL commençant par un type d’objet.

1922, 109 n. 2001). La partie supérieure du graphe AS\_SUPP (Fig. 5) montre la modélisation de ce repérage:

Type d’objet (obligatoirement l’un des mots listés dans le premier panneau) + adjectif (facultatif ou de la liste) + matériau (facultatif) + déterminant de type “de”/“des” (obligatoire) + des initiales courantes de noms latins (facultatif) + des mots commençant par une majuscule (<FIRST>, obligatoire) + “et de sa famille” / “femme” / “enfants” (facultatif et avec un retour vers la reconstruction dynamique du nom).

<FIRST> est une expression conventionnelle, pour les mots commençant par une majuscule, permettant de retrouver nos noms propres non prévus. La balise <NEW/> signale les termes annotés mais absents des dictionnaires afin de faciliter, le cas échéant, l’enrichissement des dictionnaires.

### 6.3 Le Thème (TM)

Le TM (surlignage orange) est formulé en proposition participiale (participes passé ou présent) commençant par un personnage (*Persée délivrant Andromède / Andromède délivrée par Persée*) ou un mot exprimant un événement, une action, une position, un état... (*Naissance d’Athéna*). Il peut aussi s’agir d’un personnage associé à un objet (*Apollon à l’Omphalos, Dionysos sur un bige*). Des mots seuls, dérivant des langues anciennes et dont l’usage en titre est courant (*Amazonomachie; Nekyia*), ou des expressions forgées par la tradition sont aussi recherchés (n. 916: « *Apollon au repos* »).

Pour éviter la sélection d’un groupe TM dans le cœur d’une description détaillée de l’image, les appellations thématiques sont sélectionnées si elles

sont incluses dans une AS (par exemple: <AS>statue d'<TM>Apollon menaçant un lézard rampant sur un tronc</TM></AS>). Si elles sont délimitées par les mêmes jeux de ponctuations que ceux des AC (n. 18: « – Vénus à la coquille. »; n. 27: « – Hercule assis. »), si un nom de thème traditionnel est employé seul (*Nekya*; etc.) ou si les formulations incluent un matériau – ce dernier attestant la matérialité de l'œuvre (*Supplice de Marsyas en marbre*) – elles sont balisées en tant que TM et AS (TM\_AS). Les autres cas sont repérés mais non retenus comme titre d'œuvre (Fig. 4: balise <PLUS>). Cette solution est certes peu satisfaisante car elle écarte des segments textuels pertinents mais elle répond aux objectifs de MonumenTAL; elle limite le bruit généré par la sélection d'un segment textuel au sein d'une description et qui ne serait pas utilisé en désignation d'œuvre. Une solution pourrait être apportée, pour les catalogues d'objets figurés, par le projet DataCatalogue (INHA, INRIA, BnF: SCHEITHAUER *et al.* 2023) qui travaille au balisage automatique en XML-TEI des différents éléments structurés d'une page et d'une notice de catalogue de vente d'art, dont la zone de titre de la scène figurée ou de l'artefact (entraînement de modèles de segmentation automatique des différentes zones d'information).

#### 6.4 Les adjectifs et les épiclèses sont-ils marqueurs d'un type d'appellation ?

La présence fréquente dans l'AS d'adjectifs qualifiant la scène ou le personnage (“statue de Junon assise”) a remis en cause la séparation distincte des annotations en AS et TM proposée à l'issue de la phase d'annotation manuelle et qui a été appliquée lors du calcul des résultats de la reconnaissance automatique (F-mesure: LESSIEUX *et al.* 2023, 63-64). Finalement, la modélisation de MonumenTAL privilégie, pour le français, la structuration du segment textuel: “statue de Junon assise” correspond à une AS (type d'objet + personnage). C'est encore plus marqué dans le cas des épiclèses, qui qualifient les divinités et qui sont particulièrement utilisées pour désigner les statues de culte. L'épiclèse, qui suit le nom de la divinité, peut exprimer par exemple un lieu (“Artémis Ephesia”: d'Éphèse), une attribution (“Apollon Alexikakos”: qui éloigne le mal), un aspect ou une action (“Isis Pelagia”: à la voile). La composante lieu correspond à une AS et l'action à un TM. Si la répartition en appellation dite « Structurée » et en appellation thématique est ressentie comme naturelle du point de vue de l'historien de l'art, elle s'avère délicate à modéliser et implique de séparer et répartir en amont les adjectifs et les épiclèses selon leur sens. Or, la recherche des adjectifs a été modifiée après juin 2022 pour profiter des possibilités offertes par le TAL: la recherche de tous les adjectifs du dictionnaire français par défaut a été préférée (<A> dans les graphes) et elle est complétée par un dictionnaire d'adjectifs spécifiques au domaine (dont les épiclèses) et pour des analyses qualitatives envisagées. Ce choix n'a pas dégradé le résultat.

Finalement, dans la modélisation, ce sont plutôt le début de la séquence textuelle et son mode de structuration qui différencient l'AS et le TM: l'AS commence par un support/type d'objet ou le personnage représenté, la composante TM commence plutôt par un mot exprimant un événement (naissance) ou un personnage décrit.

### 6.5 Proposer une utilisation modulable des graphes

Le graphe MonumentAL réunit quarante-six (sous-)graphes dont six correspondant aux composants principaux des appellations. La fonctionnalité d'appel de sous-graphes par Unitex a été exploitée pour permettre des recherches indépendantes de celle des appellations d'œuvres, à partir de ces six (sous-)graphes consacrés aux notions fréquentes du patrimoine culturel: ARTISTE\_all, COLL\_all (pour les collections), LIEU\_all, MATERIAU\_all, PERSO\_all (pour les personnages figurés), SUPPORT\_all. Ces graphes ont été conçus pour trouver des expressions allant de la plus simple à la plus riche et complexe.

La Fig. 6 montre SUPPORT\_all.grf. Ce patron balise des expressions énumérant plusieurs types d'objets ou de représentations ("fragment d'une copie de statue"), le cas échéant avec des adjectifs, des mots correspondant à la fois à des types d'objets et des matériaux ("un bronze"; "des marbres") ou encore des types d'objets nommés en lien avec un lieu, par l'appel du graphe LIEU\_all.grf.

## 7. LES DICTIONNAIRES ET LE THÉSAURUS OEUVRE: SOURCES ET ÉVOLUTION

### 7.1 Les dictionnaires pour l'outil de TAL

Le repérage des appellations d'œuvres et la reconnaissance d'entités nommées par Unitex reposent en grande partie sur l'emploi du dictionnaire français standard, proposé par défaut par l'application, et celui du jeu de dictionnaires spécifiques ajouté à Unitex. Ce jeu, qui a été détaillé dans la présentation de 2022 (LESSIEUX *et al.* 2023, 64-69), a évolué vers une simplification du vocabulaire et une plus grande exploitation des dynamiques apportées par les patrons linguistiques numériques. À l'exception d'ADJECTIF et de COLLECTION, les dictionnaires sont gérés dans TheA/Thesaurus-Antiquité (sur TheA: SZABADOS 2015) sous forme de thésaurus compatibles avec la norme SKOS (SKOS 2009), pour préserver notamment le multilinguisme, l'organisation du vocabulaire en termes préférés, alternatifs et cachés (skos:prefLabel; skos:altLabel; skos:hiddenLabel), les alignements (skos:exactMatch; skos:closeMatch), les regroupements de termes fondés sur des relations hiérarchiques (termes génériques/spécifiques: skos:broader skos:narrower) ou des classifications par facettes (SKOS 2009, § 9: Concept

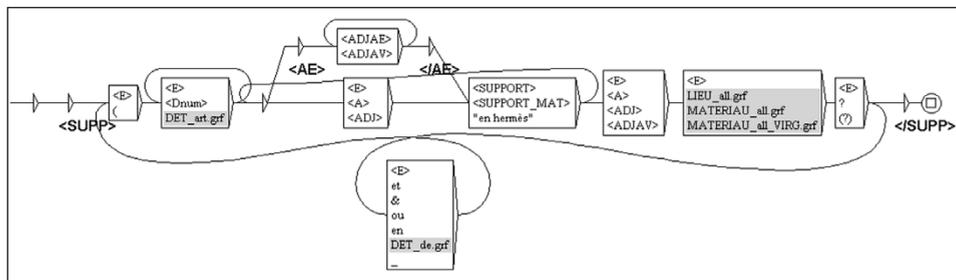


Fig. 6 – Graphe Unitex SUPPORT\_all pour les types d'objets et de représentations.

ex voto, .SUPPORT	est, être.V+z1:P3s
ex-voto, .SUPPORT	est, .N+z1:ma
ex=voto, .SUPPORT	Est, .ADJ
exaleiptron, .SUPPORT	dit, dire.V+z1:P3s
exemplaire, .SUPPORT	dire, .V
exemplaires, .SUPPORT	un, .DET
exvoto, .SUPPORT	une, .DET
fibule, .SUPPORT	
fibules, .SUPPORT	
figuration, .SUPPORT	
figurations, .SUPPORT	
figure, .SUPPORT	
figures, .SUPPORT	

Fig. 7 – Extraits des deux dictionnaires DELA de MonumentAL, SUPPORT et MTAL.

Collections). Ces regroupements permettent la création, via un notebook Python, de dictionnaires plus fins qui optimisent le fonctionnement des graphes. À partir des huit ensembles principaux, listés ci-dessous, sont générés vingt-six dictionnaires pour Unitex, qui prennent la forme de listes de termes (Fig. 7, gauche), de notions (pour une réflexion sur “concept” ou “notion” dans ce contexte: ROCHE 2008).

ADJECTIF, OEUVRE ont été créés pour le projet. Les dictionnaires suivants proviennent du thésaurus TheA du LIMC et ont été enrichis grâce à MonumentAL:

- ARTISTE: noms d'artistes, d'ateliers,
- COLLECTION: noms de collections ou de collectionneurs, de commanditaires; noms d'édifices antiques. Ces derniers sont inclus dans ce groupe dans la mesure où ils peuvent rassembler des œuvres visuelles, comme une collection (*Frise du Parthénon*),

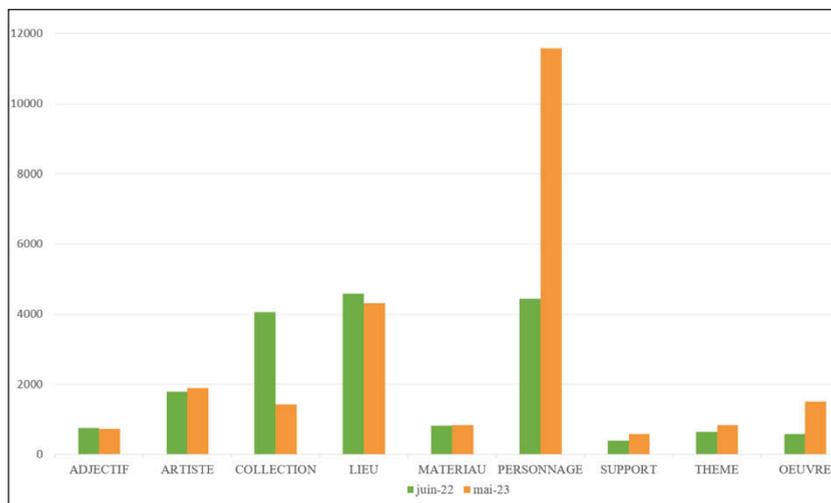


Fig. 8 – Quantifications des termes dans les dictionnaires MonumentAL en 2022 (à gauche en vert) et 2023 (à droite en orange).

- LIEU: lieux de création, d’usage, de découverte ou de conservation de l’objet figuré,
- MATERIAU matériau(x) composant un objet figuré,
- PERSONNAGE: êtres animés représentés sur des images, répartis en noms de personnages mythologiques ou historiques (individus/groupes; noms propres/noms génériques), gentilés, fonctions (prêtre, servante...), états (enfant, vieillard...), animaux,
- SUPPORT: types d’objet ou supports de représentation figurée. Ce groupe est notamment subdivisé en type d’objet et en vocabulaire de représentation (image, copie, réplique...) car ce dernier prend le même rôle dans les graphes,
- THEME: objets inanimés (char, bouclier, pomme...), noms d’événement, d’activité (naissance, Amazonomachie, chasse...).

L’ensemble totalise 23030 termes en mai 2023, ce qui fait une augmentation de 27% par rapport à juin 2022 (Fig. 8). L’augmentation forte de PERSONNAGE s’explique par le regroupement et l’ajout de personnages historiques, d’animaux et de variantes (ce groupe devient celui des êtres animés) et la baisse de COLLECTION par son remaniement pour exploiter la dynamique offerte par le TAL: les composants des noms de collections sont séparés en quatre types (quatre sous-dictionnaires) puis réorganisés par le graphe COLL\_all (collection). ADJECTIF, qui stagne, a été entièrement repensé et modifié pour produire désormais des sous-dictionnaires correspondant aux adjectifs spécialisés (par ex. “praxitélien” pour ce qui est

relatif à l'artiste Praxitèle), aux épiclèses, à des lots d'adjectifs balisés pour des analyses de l'évolution des formulations (en début des titres d'œuvres: adjectifs laudatifs, péjoratifs ou du type “pseudo”, “soi-disant” ou “prétendu” qui expriment dans ce contexte un changement d'identification du sujet représenté: *pseudo-Germanicus*). LIEU pourrait être enrichi par le contenu de gazetteer comme le TGN Getty (<https://www.getty.edu/research/tools/vocabularies/tgn/>), Pleiades (<https://pleiades.stoa.org/>), GeoNames (<https://www.geonames.org/>) ou encore de Pactols\_Lieux (<https://pactols.frantiq.fr/opentheso/?idt=th17>) mais ce serait au risque de l'introduction d'ambiguïtés, par exemple entre des noms de villes actuelles et des noms de personnifications de villes antiques, ce qui impliquerait un temps supplémentaire pour la vérification humaine sur laquelle sont fondés le maintien de la cohérence au sein de TheA et celui de ses alignements avec Wikidata.

Deux dictionnaires particuliers ont encore été ajoutés.

PERS\_AUTEUR, qui est une liste d'auteurs antiques, sert à éviter de baliser des noms d'auteurs composés comme “Apollonios de Rhodes”, qui peut correspondre à la fois à un auteur antique et à un PERSONNAGE + LIEU. Ceci n'empêche pas l'annotation du segment textuel plus long “buste d'Apollonios de Rhodes” qui est repéré grâce à “buste” (SUPPORT) et au paramétrage choisi dans Unitex pour privilégier l'annotation des formulations trouvées les plus longues.

Les dictionnaires d'Unitex utilisent le formalisme DELA; la séquence d'informations grammaticales, sémantiques et flexionnelles peut être modifiée au besoin pour optimiser l'annotation (PAUMIER *et al.* 2021, 41-46). C'est ce que fait le dictionnaire MTAL (Fig. 7, droite), qui limite les emplois de termes porteurs d'ambiguïtés par rapport au dictionnaire français par défaut, et donc le bruit comme celui généré par “est” qui est par défaut un nom (N), un adjectif (par exemple pour *Fronton est [oriental] du Parthénon*) et le verbe “être” à la troisième personne du singulier du présent.

## 7.2 Le thésaurus OEUVRE

L'un des objectifs de MonumentAL est de constituer et proposer un thésaurus des titres d'œuvres figurées de l'Antiquité classique car les thésaurus du domaine rassemblent peu de titres – Pactols réunit une cinquantaine de titres d'objets figurés de diverses aires culturelles (<https://pactols.frantiq.fr/opentheso/index.xhtml>: voir œuvres artistiques nommées) – ou correspondent davantage aux créations visuelles de l'art européen de la Renaissance et des siècles suivants, comme CONA du Getty (Cultural Objects Names Authority) et IconClass (<https://iconclass.org/>). Ainsi, dans CONA, la requête “Venus” ramène trente-six réponses dont quatre seulement concernent des statues antiques: seule la *Vénus de Milo* y a un titre distinctif, quant à *Vénus de Médicis* (*Medici Venus*), il s'agit là d'une copie du XVIII<sup>e</sup> s. (<http://vocab.getty>.

edu/page/cona/700008153); pour “Aphrodite” il y a deux statues antiques sur quinze réponses.

L'idée est de créer ce thésaurus à partir de la recherche par le TAL dans les publications anciennes et récentes – cette dernière fournissant les variantes de titres utiles pour l'interopérabilité et orientant vers la version qui pourrait servir de référence – puis de l'enrichir via les saisies d'études d'iconographie dans le corpus numérique d'objets antiques LIMC-icon (LIMC-FRANCE). Le thésaurus OEUVRE a été réalisé et compte déjà 1521 termes, variantes comprises, dont 762 termes de référence. Compatible avec SKOS, il est géré dans la nouvelle version de l'application LIMC-France.

Grâce à ce nouveau vocabulaire, une adaptation particulière du graphe des appellations a été ajoutée pour retrouver des titres d'œuvres figurées qui échappent à la typologie ou dont des composants sont trop rares pour entrer dans les dictionnaires, ou dérivés d'une autre langue, comme la *Fanciulla d'Anzio* (Jeune fille d'Anzio); de même les titres comme *le Diadumène* (celui qui se ceint d'un bandeau) qui désignent en un mot le personnage selon son action. Le graphe principal (1Appellation\_all.grf) annote les termes d'OEUVRE (balise OE) qui n'ont pas été déjà traités par le graphe de la typologie des appellations (APP.grf).

## 8. PERSPECTIVES: L'EXPLOITATION DES MOYENS DU TAL POUR DES ANALYSES OU DES APPRENTISSAGES

Les résultats de l'annotation automatique, corrigés manuellement pour constituer le “corpus de référence” français, ont commencé à être exploités pour des analyses des formulations ou/et contextuelles et pour expérimenter des apprentissages de type profonds (deep learning) dans la perspective de l'Intelligence artificielle.

Le “corpus de référence” français annoté (soixante-quinze publications; 883.740 tokens) totalise 15.414 annotations relevant de la typologie et d'OEUVRE (Tab. 1):

Types	AC	AS	AE	AX	TM	OE
Annotations	1176	7228	371	2700	3631	308

Tab. 1 – Nombre des annotations par type dans le corpus de référence en français.

Une expérimentation d'apprentissage profond a commencé fin 2023 via un Transformer (modèle Bert) avec le corpus de référence limité, grâce à un script Python, à ses annotations AC et AS mais le temps de calcul et l'accès à un serveur de calcul ralentit la progression des résultats malgré un début encourageant.

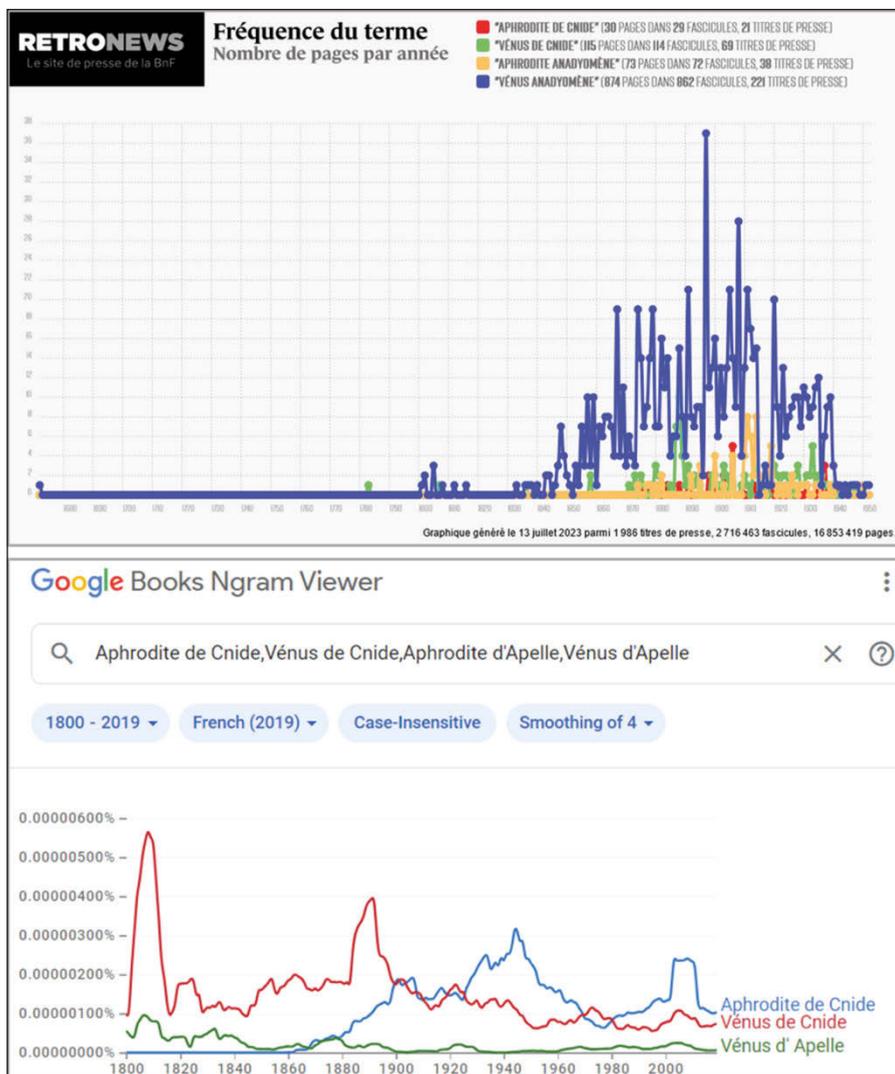


Fig. 9 – En haut: fréquence d’Aphrodite ou de Vénus d’Apelle ou de Praxitèle sur RetroNews (1680-1940). En bas: fréquence d’Aphrodite ou de Vénus d’Apelle ou de Praxitèle (de Cnide), sur Books NGram Viewer (1800-2019).

Précédemment, des analyses ont été menées à partir de toutes les appellations annotées extraites du corpus de référence. Parmi les problématiques explorées, voici l’une des plus simples: l’emploi de noms provenant du latin pour nommer les personnages représentés, même s’ils sont grecs.



L'hypothèse à confirmer est la prédominance de l'emploi de la version latine jusqu'au XIX<sup>e</sup> s., que l'on constate sur une analyse de la fréquence des occurrences de représentations d'Aphrodite (ou Vénus) par Apelle ou Praxitèle dans la collection de presse française de la BnF (période allant de 1680 à 1940), mise en ligne sur RetroNews (<https://www.retronews.fr/>). Elle montre "Vénus de Praxitèle" (bleu foncé) nettement plus citée (Fig. 9, haut). La même analyse sur Books Ngram Viewer (1800-2019) montre une inversion dans la fréquence des noms grecs et latins à partir de la fin du XIX<sup>e</sup> s. (Fig. 9, bas).

L'analyse de la fréquence des noms grecs ou latins, par quart de siècle (du XIX<sup>e</sup> au XXI<sup>e</sup> s., sauf les publications de la deuxième moitié du XX<sup>e</sup> s., encore sous droits) appliquée directement aux contenus des appellations annotées extraites du corpus de référence MonumenTAL montre une évolution similaire mais plus tardive pour les noms Aphrodite/Vénus, Athéna/Minerve, Artémis/Diane, Héraklès/Héraclès/Hercule, Zeus/Jupiter, groupés sur la Fig. 10a, séparés sur la Fig. 10b.

Enfin, une expérience d'analyse fondée sur la fouille de texte a été commencée fin 2023 sur la plateforme institutionnelle GarganText (<https://gargantext.org/>), dont certaines fonctionnalités sont encore en cours de finalisation. GarganText reçoit les textes, les segmente, génère des dictionnaires de termes paramétrables en ligne ou en reçoit par téléchargement, puis affiche des résultats sous forme de datavisualisation. Quelques textes du corpus (TXT non annotés) traitant des représentations d'Aphrodite ou Vénus qui y ont été déposés pour une analyse, ainsi qu'une liste de termes, donnent un résultat qui montre des regroupements autour de notions liées à l'objet support d'image (à gauche: Vénus, statues, groupe, figure [équivalent à statue], marbre), ou au phénomène des copies de statues (en bas à droite: Aphrodite, sculpture, copie, imitation, original, type), ou encore à l'évocation de l'artiste (en haut: rapprochement de Praxitèle avec antiques, image, nue [comme son Aphrodite]) (Fig. 11).

## 9. CONCLUSION

Les objectifs du projet tels que la production de moyens de TAL pour le repérage et l'annotation automatiques des titres d'œuvres figurées antiques ont été atteints et ont permis d'élaborer le thésaurus OEUVRE. Son enrichissement par des titres d'images mythologiques sera assuré via celui du corpus d'iconographie de la mythologie LIMC-France pour garantir sa qualité. Il s'agit bien là des titres attribués à des œuvres précises. L'évolution du projet a permis de déplacer le curseur vers les appellations génériques traditionnelles forgées pour désigner des représentations, par exemple thématiques ou narratives. La production de données qui en résulte fait envisager

la création, dans LIMC-France, d'un vocabulaire – un référentiel – des types et des thèmes iconographiques, parallèle à OEUVRE, ce qui repose d'ailleurs le problème de leurs appellations communes et de leur différenciation, notamment pour l'interopérabilité des données et les alignements. Titre d'œuvre et type iconographique ne recouvrent pas la même notion. La marge de manœuvre offerte par les méthodes de TAL, qui a aussi permis d'inclure des désignations simples correspondant à des signalements d'artefacts figurés, même sans identité particulière ("statuette de musicienne"), ouvre la voie vers une utilisation pour de la veille informationnelle ou encore pour de l'annotation ou de l'indexation automatisée de corpus d'images, sur la base de la reconnaissance automatique de formes visuelles: l'indexation des documents visuels de LIMC-icon par TheA (LIMC-France), et donc par OEUVRE ou par les notions composant une appellation, peut contribuer à fournir un corpus visuel de référence et à exploiter des analyses des composants pour tenter de déterminer des sujets narratifs récurrents présentés sur des images. Il est aussi possible, comme en témoignent les publications du corpus de référence mêlant les artefacts produits dans des contextes différents, tels des catalogues de musée, d'exploiter la même typologie des appellations, le cas échéant en adaptant le vocabulaire, pour repérer des créations visuelles d'autres aires et périodes culturelles.

ANNE-VIOLAINE SZABADOS  
CNRS, Archéologies et Sciences de l'Antiquité (ArScAn – UMR7041)  
Équipe ESPRI-LIMC, France  
anne-violaine.szabados@cnrs.fr

## BIBLIOGRAPHIE

- DE BIASI P.-M., JAKOBI M., LE MEN S. (eds.) 2012, *La Fabrique du titre. Nommer les œuvres d'art*, Paris, CNRS éditions.
- COLLIN O., GUERRAZ A. 2015, *Classification d'entités nommées de type «film»*, in *Actes de la 22<sup>e</sup> conférence sur le Traitement Automatique des Langues Naturelles*, Caen, Association pour le Traitement Automatique des Langues, 364-370.
- CUMONT F. 1913, *Catalogue des sculptures et inscriptions antiques (monuments lapidaires) des Musées Royaux du Cinquantenaire*, Deuxième éd., Direction des Musées royaux d'art et d'histoire (<https://gallica.bnf.fr/ark:/12148/bpt6k64650701>).
- DÍEZ PLATAS M.L., ROS MUÑOZ S., GONZÁLEZ-BLANCO E., RUIZ FABO P., ÁLVAREZ MELLADO E. 2021, *Medieval Spanish (12<sup>th</sup>-15<sup>th</sup> centuries) named entity recognition and attribute annotation system based on contextual information*, «Journal of the Association for Information Science and Technology», 72, 2, 224-238 (<https://doi.org/10.1002/asi.24399>).
- JAUCOURT L. 1765, *HERMATHENE*, in D. DIDEROT, J. LE ROND D'ALEMBERT (eds.), *L'Encyclopédie* ([https://fr.wikisource.org/wiki/L%E2%80%99Encyclop%C3%A9die/1re\\_%C3%A9dition/HERMATHENE](https://fr.wikisource.org/wiki/L%E2%80%99Encyclop%C3%A9die/1re_%C3%A9dition/HERMATHENE)).
- LESSIEUX A., ESHKOL-TARAVELLA I., SZABADOS A.-V., NAZARIAN M. 2022, *Reconnaissance automatique des appellations d'œuvres visuelles antiques*, in Y. ESTÈVE, T. JIMÉNEZ, T.

- PARCOLLET, M. ZANON BOITO (eds.), *Traitement Automatique des Langues Naturelles*, Avignon, ATALA, 36-44 (<https://hal.archives-ouvertes.fr/hal-03701470>).
- LESSIEUX A., SZABADOS A.-V., ESHKOL-TARAVELLA I., NAZARIAN-TROCHET M. 2023, *Resources pour l'étude des appellations d'œuvres visuelles de l'Antiquité classique: corpus, dictionnaires et outil de reconnaissance automatique*, in *Terminologie & Ontologie: Théories et Applications. Actes de la conférence TOTh 2022 (Université Savoie Mont Blanc 2022)*, Terminologica Presses Universitaires Savoie Mont Blanc, 51-73 (<https://hal.science/hal-04368609>).
- LIMC 1981-1997, *Lexicon Iconographicum Mythologiae Classicae (LIMC)*, Zürich, München, Düsseldorf, Artemis & Winkler Verlag.
- LIMC-FRANCE, Corpus numérique LIMC-icon, site web LIMC-France et thésaurus TheA (<http://www.limc-france.fr/>).
- MINISTÈRE DE LA CULTURE 2020, *Informatisation réglementaire des collections d'un musée de France*, 15 Décembre 2020 (<https://www.culture.gouv.fr/Thematiques/Musees/Pour-les-professionnels/Conserver-et-gerer-les-collections/Informatiser-les-collections-d-un-musee-de-France/Informatisation-reglementaire-des-collections-d-un-musee-de-France> [consultation: 26.12.2023]).
- MUSÉES NATIONAUX, LOUVRE (MICHON É.) 1922, *Catalogue sommaire des marbres antiques / Musée national du Louvre, Département des antiquités grecques et romaines*, Paris, Musée National (<https://gallica.bnf.fr/ark:/12148/bpt6k6237020f>).
- NOUVEL D., EHRMANN M., ROSSET S. 2016, *Named Entities for Computational Linguistics*, Wiley.
- PAUMIER S., GUENTHNER F., LAPORTE É., MALCHOK F., MARSCHNER C., MATINEAU C., MARTINEZ C., MAUREL D., NAGEL S., NEME A. et al. 2021, *UNITEX 3.3. Manuel d'utilisation*, Université Paris-Est Marne-la-Vallée (<https://hal.science/hal-03589598/file/Unitex-GramLab-3.3-usermanual-fr.pdf>).
- PAUMIER S., NAKAMURA T., VOYATZI S. 2009, *UNITEX, a Corpus Processing System with Multi-Lingual Linguistic Resources*, in *eLexicography in the 21<sup>st</sup> Century: New Challenges, New Applications (eLEX'09)*, France, 173-175 (<https://hal.science/hal-00621564>).
- PRIOUX É. à paraître, *Les titres et désignations antiques des œuvres d'art célèbres*, in *Actes du colloque MonumenTAL (Nanterre 2023)*.
- REINACH A. 1921, *Textes grecs et latins relatifs à l'histoire de la peinture ancienne: recueil Milliet. T. I, Paris, C. Klincksieck* (<https://gallica.bnf.fr/ark:/12148/bpt6k11622612>).
- ROCHE C. 2008, *Faut-il revisiter les Principes terminologiques?*, in *TOTh 2008. Actes de la deuxième conférence TOTh (Annecy 2008)*, Annecy, 53-72.
- SCHIEITHAUER H., BÉNIÈRE S., MOREUX J.-P., ROMARY L. 2023, *DataCatalogue: rétro-structuration automatique des catalogues de vente*, Webinaire Culture-Inria, Ministère de la Culture (<https://hal.science/hal-04360229/document>).
- SKOS 2009, *SKOS Simple Knowledge Organization System Reference. W3C Recommendation 18 August 2009* (<https://www.w3.org/TR/skos-reference/>).
- SZABADOS A.-V. 2015, *From the LIMC Vocabulary to LOD. Current and expected uses of the Multilingual Thesaurus TheA*, in S. ORLANDI, R. SANTUCCI, V. CASAROSA, P.M. LIUZZO (eds.), *Information Technologies for Epigraphy and Cultural Heritage. Proceedings of the First EAGLE International Conference*, Collana Convegni 26, Roma, Sapienza Università Editrice, 51-67 (<https://www.eagle-network.eu/wp-content/uploads/2015/01/Paris-Conference-Proceedings.pdf>).
- TIECK F. 1844, *Notice des sculptures antiques*, Berlin, Imprimerie de Moeser et Kühn (<https://gallica.bnf.fr/ark:/12148/bpt6k64241091>).
- WIDLÖCHER A., MATHET Y. 2012, *The Glozz platform: A corpus annotation and mining tool*, in *Proceedings of the ACM Symposium on Document Engineering (Paris 2012)*, Paris, 171-180 (<https://hal.archives-ouvertes.fr/hal-01023774>).

ABSTRACT

The NLP tools for the automatic recognition and annotation of titles of figurative artworks from The Classical World, developed by the MonumentAL project, have evolved through the digital modeling of linguistic patterns. These have helped to broaden the focus from the titles of specific artworks to the naming of generic iconographic types, and to add old and recent expressions specific to art historians and archaeologists. Thanks to this work based on a diachronic approach, a thesaurus of artwork titles (OEUVRE) gathering reference terms, variants and cacographies has been created and is now linked to the online LIMC-France database (corpus of Ancient artworks). The text corpus (Gold standard), from the 18th to the 21st century, and its annotations can now be exploited for statistical analysis or deep learning experimentation.