



HAL
open science

Ressources pour l'étude des appellations d'oeuvres visuelles de l'Antiquité classique : corpus, dictionnaires et outil de reconnaissance automatique

Aurore Lessieux, Anne-Violaine Szabados, Iris Eshkol-Taravella, Marlene Nazarian-Trochet

► To cite this version:

Aurore Lessieux, Anne-Violaine Szabados, Iris Eshkol-Taravella, Marlene Nazarian-Trochet. Ressources pour l'étude des appellations d'oeuvres visuelles de l'Antiquité classique: corpus, dictionnaires et outil de reconnaissance automatique. Terminologie & Ontologie: Théories et Applications. Actes de la conférence TOTh 2022, Université Savoie Mont Blanc, 2 & 3 juin 2022, Presses Universitaires Savoie Mont Blanc, pp.51-73, 2023, Terminologica, 978-2-37741-085-9. hal-04368609

HAL Id: hal-04368609

<https://hal.science/hal-04368609v1>

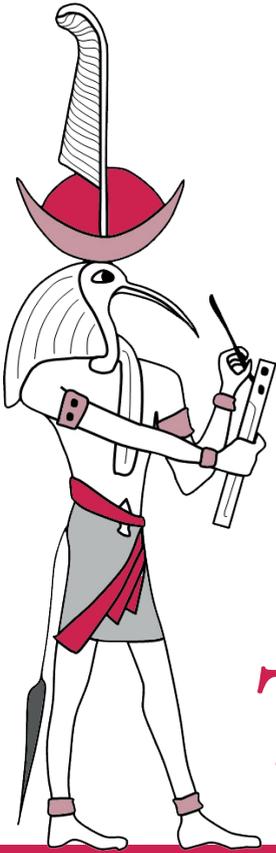
Submitted on 1 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



TOTh 2022

Terminologie & Ontologie: Théories et Applications

Terminologie & Ontologie: Théories et Applications

Actes de la conférence

TOTh 2022

Université Savoie Mont Blanc

2 & 3 juin 2022



Les ouvrages TOTh précédents sont disponibles:

- sur le site de l'Université Savoie Mont Blanc (btk.univ-smb.fr/livres)
- sur le site du Comptoir des Presses d'Universités (www.lcdpu.fr)
- ou auprès de: contact@toth.condillac.org

Éditeur: Presses Universitaires Savoie Mont Blanc
27 rue Marcoz
BP 1104
73011 CHAMBÉRY CEDEX
www.univ-smb.fr

Réalisation: C. Brun, C. Roche, M. Papadopoulou
Collection «Terminologica»
ISBN: 978-2-37741-085-9
ISSN: 2607-5008
Dépôt légal: juin 2023

Terminologie & Ontologie: Théories et Applications



Actes de la conférence

TOTH 2022

Université Savoie Mont Blanc

2 & 3 juin 2022

<http://toth.condillac.org>

avec le soutien de:

- Université Savoie Mont Blanc
- École d'ingénieurs Polytech Annecy Chambéry
- Ministère de la Culture. Ce projet est soutenu financièrement par le Ministère de la Culture – Délégation Générale à la Langue Française et aux Langues de France



Presses Universitaires Savoie Mont Blanc
Collection « Terminologica »

Comité scientifique

Président du Comité scientifique: Christophe Roche

Comité de pilotage

Rute Costa	Universidade Nova de Lisboa
Humbley John	Université Paris 7
Kockaert Hendrik	University of Leuven
Christophe Roche	Université Savoie Mont Blanc

Comité de programme 2022

Le comité de programme est constitué chaque année à partir du comité scientifique de TOTh en fonction des soumissions reçues. La composition du comité scientifique est accessible à l'adresse suivante: <http://toth.condillac.org/committees>

Amparo Alcina	Universitat Jaume I – Spain
Xiaomi An	Renmin University – China
Albina Auksoariute	Institute of the Lithuanian Language – Lithuania
Bruno Bachimont	Université Technologie de Compiègne – France
Jean-Paul Barthès	Université Technologie de Compiègne – France
Christopher Brewster	Maastricht University – Netherlands
Danielle Candel	CNRS, Université Paris Diderot – France
Sylviane Cardey	Université de Franche-Comté – France
Stéphane Chaudiron	Université de Lille 3 – France
Rute Costa	Universidade NOVA de Lisboa – Portugal
Eric De La Clergery	INRIA – France
Dardo De Vecchi	Kedge Business School – France
Thierry Declerck	DFKI – Germany
Valérie Delavigne	Université Paris 3 – France
Sylvie Desprès	Université Paris 13 – France
Juan Carlos Diaz Vasquez	EAFIT University – Colombia
Pamela Faber	Universidad de Granada – Spain
Christiane Fellbaum	Princeton University – USA
Cécile Frérot	Université Stendhal Grenoble 3 – France
Iolanda Galanes	Universidade de Vigo – Spain
Teodora Ghiviriga	Alexandru Ioan Cuza University – Romania
Rufus Gouws	University of Stellenbosch – South Africa
Frédérique Henry	ATILF – France
John Humbley	Université Paris 7 – France
Yangli Jia	University of Liaocheng – China
Kyo Kageura	University of Tokyo – Japan
Barbara Karsch	BIK Terminology – USA
Hendrik Kockaert	University of Leuven – Belgium
Héba Lecocq	Université Sorbonne nouvelle – France
Hélène Ledouble	Université de Toulon – France

Patrick Leroyer	Aarhus University – Denmark
Georg Löckinger	University of Applied Sciences Upper Austria – Austria
Rodolfo Maslias	TermCoord, European Parlement – Luxembourg
Candida Jaci de Sousa Melo	Universidade Federal do Rio Grande do Norte – Brazil
Fidelma Ní Ghallchobhair	Foras na Gaeilge, Irish-Language Body – Ireland
António Pareja Lora	Universidad Complutense de Madrid – Spain
Sandrine Peraldi	University College Dublin – Ireland
Silvia Piccini	Italian National Research Council – Italy
Suzanne Pinson	Université Paris Dauphine – France
Maria Pozzi	el colegio de méxico – Mexico
Bihua Qiu	China National Committee for Terms in Sciences and Technologies – China
Jean Quirion	Université d'Ottawa – Canada
Renato Reinau	Université de Genève – Switzerland
Christophe Roche	Université Savoie Mont Blanc – France
Mathieu Roche	CIRAD – France
Micaela Rossi	Università degli studi di Genova – Italy
Klaus-Dirk Schmitz	Cologne University – Germany
Marcus Spies	University of Munich – Germany
Frieda Steurs	University of Leuven – Belgium
Anne Theissen	Université de Strasbourg – France
Philippe Thoiron	Université Lyon 2 – France
Kara Warburton	City University of Hong Kong – China
Maria Teresa Zanola	Università Cattolica del Sacro Cuore – Italy

Avant-propos



La seizième édition de la conférence TOTh s'est tenue, selon une habitude bien établie, les jeudi et vendredi de la première semaine de juin. Pour la deuxième année consécutive, TOTh s'est déroulée à la fois en présentiel et à distance. Ainsi, nous avons pu renouer avec les plaisirs qu'offrent le présentiel et les moments de sociabilité comme le dîner de TOTh dans la ville historique de Chambéry, tout en permettant à un plus grand nombre de suivre et de participer à nos travaux.

Notre collègue Platon Pétridis, Professeur d'Archéologie Byzantine à l'Université d'Athènes, a ouvert la Conférence TOTh 2022 sur le thème de la classification et appellation des objets dans les contextes archéologiques et pour qui «une nouvelle approche en matière de terminologie et de taxinomie dans les sciences archéologiques s'avère donc indispensable». Un sujet proche de nos préoccupations portant sur l'organisation des objets et leur dénomination.

Quelques chiffres donneront une idée de cet événement initié en 2007.

- Quatorze communications ont été retenues pour publication sur les 17 présentées après une sélection rigoureuse par un comité de programme international issu de 24 pays. La diversité des sujets abordés, tant théoriques que pratiques, illustre la richesse et le dynamisme de notre discipline.
- Sur les 87 personnes enregistrées, plus de 50 ont suivi de manière assidue les présentations. 18 pays étaient représentés : Albanie, Allemagne, Argentine, Chine, Chypre, Congo, Espagne, Estonie, États-Unis, France, Grèce, Italie, Lettonie, Lituanie, Niger, Pays Bas, Portugal, Suisse.

Les actes sont publiés aux Presses Universitaires Savoie Mont Blanc. Ceux des années précédentes sont accessibles à partir du site de la conférence (<http://toth.condillac.org/>) et des Presses Universitaires Savoie

Mont Blanc (https://btk.univ-smb.fr/livres/?fwp_collections_revues=-terminologica).

Je terminerai en remerciant le Ministère de la Culture, et plus précisément la Délégation Générale à la Langue Française et aux Langues de France, l'Université Savoie Mont Blanc et l'École Polytech Annecy-Chambéry pour leur support et leur aide financière à l'organisation de la conférence et à la publication des actes.

Christophe Roche
Président du Comité scientifique

SOMMAIRE

CONFÉRENCE D'OUVERTURE	13
<hr/>	
Des Chiffres et des Lettres? Classification et appellation des objets dans les contextes archéologiques	
Platon Petridis	15
ARTICLES	29
<hr/>	
Enjeux pour la mise en réseau et l'analyse des connaissances archéologiques	
Guillaume Reich, Sébastien Durost, Jean-Pierre Girard <i>avec la collaboration d'Éric Lacombe et de Miled Rousset</i>	31
Ressources pour l'étude des appellations d'œuvres visuelles de l'Antiquité classique : corpus, dictionnaires et outil de reconnaissance automatique	
Aurore Lessieux, Anne-Violaine Szabados, Iris Eshkol-Taravella, Marlène Nazarian	51
Developing Training Corpora for Automatic Extraction of Cybersecurity Terminology	
Sigita Rackevičienė, Andrius Utkā	75
Le vocabulaire des manuels francophones de psycholinguistique	
Jacques François	97
Does NMT make the Human Factor in Terminology Extraction and Translation Obsolete?	
Tamara Christmann, Mihaela Vela	119

Terminologie(s) et territorialité : la description des bières artisanales en français et en italien	139
Nicla Mercurio	
TimeInfo: a Semantic Annotation Framework for Temporal Information in Scientific Papers	161
Salah Yahiaoui and Iana Atanassova	
Evaluation of Machine Translated Artificial Intelligence Terminology with Respect to Fluency and Adequacy	175
Urtė Kvyklytė and Jurgita Mikelionienė	
Technological taxonomies for hypernym and hyponym retrieval in patent texts	195
You Zuo, Yixuan Li, Alma Parias García, Kim Gerdes	
A low-cost method for text summarization	217
Adrian Vogel-Fernandez, Pablo Calleja, Mariano Rico	
Ressources lexicales et terminologiques pour les langues en danger : enjeux, défis et méthodes	225
Antonia Cristinoi	
Porting the “One Size Fits All” Model for Terminology into a Linked Data Compatible Format	245
Giorgio Maria Di Nunzio, Federica Vezzani, Thierry Declerck, Patricia Martín-Chozas	
Harmonising concepts and terms for the development of knowledge. A study on the need for names and the necessity for harmonisation in the 19th century	259
Beatrice Ragazzini	
Implementation of terms and their grammatical category in an ontology. The ONTODIC model	281
Amparo Alcina	

Ressources pour l'étude des appellations d'œuvres visuelles de l'Antiquité classique : corpus, dictionnaires et outil de reconnaissance automatique

Aurore Lessieux*, Anne-Violaine Szabados**, Iris Eshkol-Taravella*,
Marlène Nazarian**

* Modèles, Dynamiques, Corpus (MoDyCo) UMR7114

Université Paris Nanterre, France

ieshkolt@parisnanterre.fr

aurorelessieux@parisnanterre.fr

** Archéologies et Sciences de l'Antiquité (ArScAn) UMR7041

CNRS, LIMC, France

marlene.nazarian@cnrs.fr

anne-violaine.szabados@cnrs.fr

Résumé. Le projet pluridisciplinaire MonumentAL a pour objectif de repérer et répertorier les appellations d'œuvres d'art de l'Antiquité classique en utilisant les méthodes du TAL et d'en étudier l'élaboration et l'évolution dans différents types de publications. Il repose sur une collaboration étroite entre historiens de l'art (LIMC), linguistes-TAListes (MoDyCo), conservateurs et bibliothécaires (BnF). Le corpus est constitué de textes en français publiés du XVIII^e au XXI^e siècle. Le traitement proposé implique plusieurs étapes : sélection du corpus textuel d'étude, élaboration d'une typologie des appellations, constitution d'un corpus annoté manuellement par les experts du domaine, création et réutilisation de dictionnaires de termes, création d'un nouveau thésaurus des noms d'œuvres d'art et développement d'un outil de reconnaissance automatique des appellations fondé sur des méthodes symboliques.

1. Introduction

Les appellations des œuvres visuelles de l'Antiquité classique sont au cœur de MonumentAL – Monuments antiques et Traitement Automatique de la Langue¹ – un projet qui vise à les révéler, les répertorier et les étudier par les historiens d'art et les archéologues dans une perspective diachronique. Compte tenu du grand nombre d'objets figurés concernés et de la multiplicité de leurs dénominations, ainsi que de la volumétrie élevée des sources textuelles et de la variété de leur mise en pages, des méthodes et des outils relevant du TAL ont été privilégiés afin d'extraire les appellations d'un corpus de textes en français produits du XVIII^e au XXI^e s.

Cette recherche pluridisciplinaire, qui repose sur une collaboration étroite entre des linguistes-TAListes (MoDyCo / Modèles Dynamique Corpus), des historiens d'art, des archéologues et des philologues (LIMC / Lexicon Iconographicum Mythologiae Classicae) et des conservateurs et bibliothécaires de la Bibliothèque nationale de France (BnF), a permis, d'établir des patrons linguistiques spécifiques aux appellations d'œuvres visuelles et de thèmes iconographiques afin de les reconnaître automatiquement et de constituer des vocabulaires français dédiés au TAL et des thésaurus réutilisables pour d'autres besoins relatifs aux données du patrimoine culturel.

Cet article présente le processus mis en place pour développer le module de la détection automatique des appellations d'œuvres d'art en français et les résultats de la première phase de leur reconnaissance automatique. L'accent est mis sur les problématiques liées aux vocabulaires. Le traitement proposé implique plusieurs étapes : la constitution du corpus textuel d'étude, l'élaboration d'une typologie des appellations et des patrons linguistiques correspondants, l'utilisation et l'enrichissement de vocabulaires, la création d'un corpus de référence annoté

1 MonumentAL : <https://heurist.huma-num.fr/heurist/?db=MonumentAL&web-site>. Ce travail a été réalisé dans le cadre du labex Les passés dans le présent et a bénéficié de l'aide de l'État géré 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>).

manuellement par les experts du domaine, la reconnaissance automatique des appellations et son exploitation.

2. Problématiques et objectifs

Les textes, produits de l'Antiquité à nos jours, restituent des noms donnés aux objets figurés antiques relevant du monde classique et aux images que ces œuvres véhiculent. Ces noms émanent davantage d'une tradition naissant des discours, des commentaires et des échanges d'idées sur les œuvres visuelles, du besoin d'une appellation pour les citer plutôt que d'un dialogue entre l'artiste et le spectateur. En effet, jusqu'au début du XX^e s., les artistes titrent rarement leurs créations, contrairement à aujourd'hui². D'après les informations sur les artefacts figurés qui nous sont parvenues grâce aux sources textuelles, archéologiques et aux objets conservés, rares sont les images antiques pour lesquelles on peut supposer qu'elles ont été titrées par leur créateur et la majorité des désignations formulées dans les publications en français sont des créations postérieures à celles de l'œuvre matérielle.

La dénomination d'un objet visuel antique est un processus complexe qui dépend de multiples paramètres parmi lesquels le contexte d'écriture, le profil de l'auteur et le lectorat visé, le type de publication et d'œuvre visuelle ainsi que le sujet de l'image. À partir de ce sujet iconographique sont élaborés la plupart des titres d'art selon des types de formulations qui sont déjà présents dans les textes antiques, par exemple *Persée délivrant Andromède*, *Aphrodite de Cnide* ou dans une version plus courte la *Cnidienne* ou plus longue *statue de l'Aphrodite de Cnide de Praxitèle*.

Dans cet article, le mot 'appellation' a été privilégié car il est moins restreint que 'titre', plus en adéquation avec les appellations thématiques d'image comme celles citées dans les catalogues de céramique peinte. Il correspond aussi mieux au fait qu'un nom d'œuvre peut évoquer non pas un seul mais plusieurs référents matériels, plusieurs images et même des concepts, comme *Aphrodite de Cnide*, qui est à la

2 Sur le titre d'œuvre visuelle : Prioux (2011), Prioux (s. d.), Biasi (2012).

Ressources pour l'étude des appellations d'œuvres visuelles de l'Antiquité classique : corpus, dictionnaires et outil de reconnaissance automatique

fois le titre d'une statue grecque célèbre de Praxitèle aujourd'hui perdue, le nom donné à ses nombreuses copies antiques ou plus récentes et celui de son type iconographique statuaire maintes fois reproduit et énoncé.

Des plateformes Web proposent la récupération et l'analyse statistique de corpus textuels³ et de plus en plus de publications sont librement accessibles en ligne, grâce aux politiques institutionnelles actuelles favorisant l'ouverture des données. Ainsi, la *Revue archéologique* par exemple comprend 209 fascicules anciens ocrés, couvrant un siècle de parution (1844-1851), qui totalisent 60 952 vues/pages, disponibles sur Gallica⁴. Dans ce contexte, les méthodes et techniques du TAL répondent à deux objectifs du projet MonumentAL : le repérage des titres pour en étudier l'évolution sur plusieurs siècles, et leur récupération afin de constituer ou d'enrichir des vocabulaires et des référentiels exploitables pour de l'annotation de documents, de la saisie normalisée des métadonnées et données. Ces titres sont bien spécifiques au domaine car ils contiennent de nombreux mots relevant des cultures antiques et dérivant du latin ou du grec ancien, comme les épicleses. Quelques thésaurus proposent des titres d'œuvres d'art antique mais ils sont souvent généralistes, peu fournis ou plus en adéquation avec l'art renaissance ou classique européen qu'antique⁵. Les nouvelles données repérées grâce aux moyens du TAL permettront de créer un nouveau thésaurus réutilisable des titres et des sujets iconographiques bien adapté aux œuvres antiques, ŒUVRE. Associé aux données du corpus numérique d'objets figurés LIMC-icon⁶, donc directement aux notices sur les objets figurés, il est l'un des enjeux de ce projet.

3 Par ex. ISTEEX (www.istex.fr/), Gargantext (gargantext.org/) ou, pour des corpus de textes antiques en grec et latin, Hyperbase (hyperbase.unice.fr/hyperbase/).

4 <https://gallica.bnf.fr/ark:/12148/cb32856350w/date&rk=21459;2>

5 Par ex. PACTOLS (<https://pactols.frantiq.fr/opentheso/>); Iconclass (<https://iconclass.org/>). Cultural Objects Name Authority, Getty Iconography Authority : www.getty.edu/research/tools/vocabularies.

6 Site Web LIMC-France : www.limc-france.fr.

3. Le corpus textuel

Les appellations d'œuvres visuelles se sont normalisées au fil du temps, passant d'expressions plus détaillées ou plus explicatives à des segments textuels plus courts et formalisés. Un objet a pu être désigné sous plusieurs noms bien différents sous l'effet du parcours historique de l'œuvre, des connaissances du moment ou des différents contextes d'une transmission textuelle du savoir à visées plus ou moins érudites ou marquée par la mise en place de nouvelles disciplines savantes, à partir XVIII^e s. L'étude diachronique s'avère pertinente et le corpus textuel envisagé, traitant d'histoire de l'art et d'archéologie, couvre une période de publication allant du XVIII^e au XXI^e s. afin de prendre en compte la diversité des appellations et leur évolution. L'instance Zotero du projet recense déjà plus d'un millier de références de publications en français sélectionnées principalement sur Gallica, pour garantir un accès libre et durable aux sources, ainsi que sur Persée, OpenEdition, Cairn, Hal⁷ et Wikipédia pour des périodiques spécialisés et des textes récents. Cette sélection, qui n'est pas close, réunit un panel varié représentatif des écrits destinés à des spécialistes du domaine, des amateurs ou des néophytes :

- essai (monographie, chapitre ou recueil d'articles), manuel,
- traduction de textes antiques grecs et latins,
- catalogue et guide de collection (musée, vente, etc.) structurés ou pas,
- guide ou récit de voyage,
- article et fascicule de périodique,
- notice d'encyclopédie, lexique (lexiques de sculpteurs), index.

Les versions de travail sont en format.txt et issues, pour la plupart, d'une océrisation qui perd la mise en forme des caractères et la mise en page, souvent porteuses de sens dans les ouvrages d'histoire de l'art (titres d'œuvres et thématiques en italique, catalogues semi structurés),

7 gallica.bnf.fr; www.persee.fr; www.openedition.org; www.cairn.info; hal.archives-ouvertes.fr

Ressources pour l'étude des appellations d'œuvres visuelles de l'Antiquité classique : corpus, dictionnaires et outil de reconnaissance automatique

et qui génère des scories: erreurs de caractères et de ponctuations, mauvaise récupération des autres alphabets (du grec dans un texte en français). Le parti pris est d'assumer ces changements pour répondre à l'objectif de la veille informationnelle. Les traductions françaises de textes antiques grecs et latins, la plupart publiées au XIX^e s., sont plutôt récupérées sur des sites dédiés aux corpus antiques, tels que Remacle (remacle.org), qui fournissent des textes corrigés.

Un sous corpus d'une trentaine de publications a été sélectionné et utilisé dans une première étape consistant en une annotation manuelle pour établir la typologie des appellations et pour repérer les mots spécifiques au domaine qui les composent afin d'élaborer les patrons linguistiques et les dictionnaires nécessaires au développement d'un outil de reconnaissance automatique des appellations d'œuvre. Afin d'atteindre la représentativité des données, c'est-à-dire, pour rendre compte des différentes pratiques d'appellation, le choix couvre le panel du lectorat et la couverture temporelle (quatorze écrits de spécialistes en histoire de l'art et seize d'amateurs) pour un total de trente publications, 591 711 tokens. Les textes sur la statuaire et les pierres gravées (intailles et camées), puis ceux sur la mosaïque, la peinture pariétale, les vases peints et des ouvrages mêlant les aires culturelles ont été choisis pour tester l'efficacité de l'approche sur des domaines pour lesquels les traditions de rédaction des spécialistes varient. Des traductions de textes antiques, en particulier des livres 34-36 de l'*Histoire naturelle* de Pline, ont fourni une bonne partie des titres de tableaux peints.

4. La typologie des appellations

Grâce à une première étape d'analyse manuelle du corpus, associée à des réunions assurant le partage de compétences et de connaissances entre historiens d'art et linguistes, une typologie des appellations d'œuvres visuelles a pu être établie et servir de modèle pour l'annotation manuelle. Trois catégories d'appellations y sont distinguées : l'Appellation Structurée, le Thème et l'Appellation Courte.

4.1. L'Appellation Structurée

L'Appellation Structurée' (balise d'encodage: AS) doit suffire à reconnaître une image et un objet connus ou identifiés, un sujet ou un objet matériel figurés. Les différentes structures de ce groupe sont régulières et récurrentes. Soit elles sont focalisées sur le personnage principal figuré tel un personnage antique divin, mythique ou historique (*Zeus olympien de Phidias, Gladiateur mourant du Capitole*), soit elles associent le type d'objet avec le sujet (*statue de Silène*), le nom de l'artiste (*peinture d'Apelle, Dinos de Sophilos*), de la collection (*Tête Kaufmann*). Ces appellations réunissent plusieurs des composants suivants: le type de l'artefact ou de la représentation, son matériau, le nom du personnage représenté, celui de l'artiste, du collectionneur ou de la collection, un lieu. Un thème iconographique, relevant du deuxième type (Thème), peut se substituer au personnage représenté. Par exemple :

- *statue de l'Aphrodite de Cnide de Praxitèle*: type d'artefact + personnage + lieu + artiste,
- *l'Hercule Farnèse*: personnage + collection,
- *l'Aurige de Delphes*: personnage générique + lieu,
- *le Vase Médicis*: type d'artefact + collection,
- *une Vénus en marbre*: personnage + matériau,
- *copie du Supplice de Marsyas*: type de représentation + Thème.

4.2. Le Thème

'Thème' (TM) correspond aux thèmes iconographiques qui sont exprimés par une brève description de la scène figurée, le nom d'un événement, d'un mythe (*Hercule combattant le lion de Némée*). On y distingue principalement des événements liés aux personnages antiques, notamment de la vie comme une naissance, un mariage ou une mort (*Naissance d'Athéna, Mariage de Pélée et Thétis, Suicide d'Ajax*), les combats, les rituels, les processions, les banquets (*Bataille contre les Perses, scène de libation, Sacrifice d'Iphigénie, Jugement de Pâris, Pompe du Mégabyze*) ; des descriptions du personnage principal

Ressources pour l'étude des appellations d'œuvres visuelles de l'Antiquité classique : corpus, dictionnaires et outil de reconnaissance automatique

et de sa place par rapport à un objet (*Vénus couchée dans une barque, Amour avec la dépouille du lion de Némée*) ou à un animal (*Hercule sur le sanglier d'Erymanthe*), son attitude (*Hercule au repos*) ou son costume (*Antigone en armure*); des actions accomplies ou subies par des personnages (*Enlèvement de Perséphone, Dionysos portant Bacchus enfant, Hercule dompté par l'Amour*). Les énumérations sont dans ce groupe et certains thèmes sont exprimés par un seul mot, comme la *Nekyia* (invocation du défunt). Un Thème peut être introduit par un verbe relatif à la description iconographique ou à la création artistique (*représenter, sculpter...*) ou inséré dans une AS en tant que sujet représenté par l'œuvre. Par exemple :

- *Persée délivrant Andromède; Andromède délivrée par Persée*: personnage + verbe + personnage (proposition participiale),
- *Apothéose d'Hercule; Départ de guerrier*: événement / action (thème principal) + nom propre (personnage) / personnage générique,
- *Hercule au repos; Bacchus sur un char*: personnage / thème + préposition + personnage / thème secondaire,
- *Éros et Psyché; Jupiter, Junon et Minerve*: personnage + personnage...,
- une *Amazonomachie* (bataille contre les Amazones): un thème principal.

4.3. L'Appellation Courte

L'Appellation Courte (AC) est limitée au nom du personnage précédé d'un déterminant (*le Laocoon, cette Vénus, la Samothrace*) ou inséré entre des bornes le délimitant strictement et sans ambiguïté, comme de la ponctuation (*– Apollon.*). Cette forme succincte était notamment utilisée par les auteurs du XVIII^e et du début du XIX^e siècle pour les statues célèbres, à une époque où peu de sculptures antiques étaient connues et publiées. Elle peut trouver racine dans les échanges des érudits sur les œuvres: les auteurs citaient les mêmes statues et *le Torse* suffisait à reconnaître le *Torse du Belvédère*, tout comme *le Sauroctone* (tueur de lézard) l'*Apollon sauroctone*. Quelques statues

sont aussi nommées par un seul mot, depuis l'Antiquité comme le *Diadumène* (celui qui ceint sa tête du bandeau), ou par un surnom plus récent (*Pasquin* ; *Idolino*). Un thème iconographique en un mot reste dans Thème pour être inclus dans le dictionnaire exploité pour le repérage automatique de TM.

5. L'annotation manuelle

La Figure 1, montre un exemple d'annotation manuelle effectuée sous le logiciel Glozz sur un extrait de catalogue de collections (Reinach 1895). L'appellation *groupe du Laocoon* (jaune) est une Appellation Structurée, *Triomphe de Pompée* (bleu) est un Thème, et *un Laocoon* (orange) est une Appellation Courte.

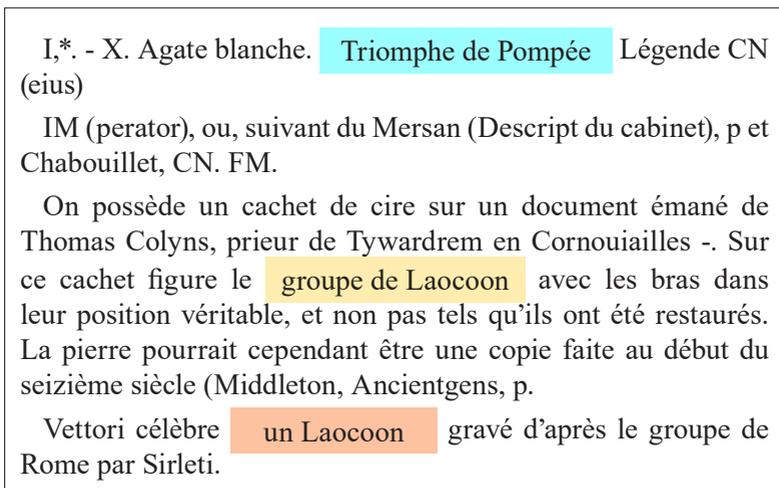


FIG. 1 – Annotation avec Glozz d'un extrait de catalogue (Reinach 1895, 101).

Les trente textes représentatifs sélectionnés d'une manière aléatoire dans le corpus textuel ont été annotés manuellement selon

Ressources pour l'étude des appellations d'œuvres visuelles de l'Antiquité classique : corpus, dictionnaires et outil de reconnaissance automatique

cette typologie par trois historiennes de l'art avec l'application Glozz (Widlöcher *et al.* 2009). Ce corpus annoté comprend 1 701 Appellations Structurées (AS), 1 277 Thèmes (TM) et 640 Appellations Courtes (AC). L'évaluation de l'annotation manuelle a été faite sur le corpus annoté par deux annotateurs en utilisant le kappa de Cohen (1960). Le kappa obtenu pour les AS est de 0,93, il est de 0,84 pour les TM et de 0,88 pour les AS. Selon l'échelle de Landis & Koch (1977), cet accord est considéré comme excellent.

L'Appellation Structurée a le meilleur score inter annotateurs car il s'agit de la catégorie dont les constructions sont les plus identifiables : elles sont le plus souvent formées à partir du personnage représenté ou du type d'objet. L'association avec le nom de l'artiste, du collectionneur ou de la collection est aussi facteur d'amélioration.

La catégorie Thème, qui comprend les thèmes et descriptions iconographiques, est plus difficile à annoter car elle se confond facilement avec une partie de description complète et détaillée d'une scène d'œuvre d'art figurée. L'identifier et la différencier du segment de description, sans le recours à la mise en page et à la mise en forme, nécessite une solide connaissance des expressions traditionnellement employées et du domaine artistique. '*Amour captif devant une image de Némésis*' est interprétable de deux manières : comme une seule représentation (TM) ou comme deux œuvres, un *Amour captif* (TM) placé devant une *image de Némésis* (AS). Selon le niveau de connaissance de l'annotateur ce segment textuel peut être annoté selon un ou deux types.

Il apparaît, d'après l'analyse des appellations repérées, que le premier type de formulation (AS), bien adapté aux scènes simples, à un personnage, est plus fréquemment utilisé pour nommer les statues et les objets mis en relation avec leur contexte de création ou leur conservation (*Zeus de Phidias*; *Hercule Farnèse*). Dans les textes annotés portant sur le domaine de la sculpture, 68% des appellations sont des AS et 19% sont des TM. La structure régulière des AS réunit surtout des termes relevant de la culture antique (personnes, types d'objet) et de l'histoire de l'art (collections, matériaux). Le deuxième type de formulation paraît employé plutôt pour des scènes plus complexes ou avec

plusieurs personnages, comme les sujets peints sur la céramique, la peinture de tableau – aujourd’hui perdue mais connue grâce aux textes antiques qui témoignent de l’ancienneté de cette tradition d’appellation – ou la peinture pariétale. Dans les textes annotés portant sur le domaine de la céramique et de la peinture, 52% des appellations sont des TM et 35% sont des AS. Même si la variété des mots utilisés dans les TM est limitée par la tradition et le vocabulaire du domaine, la présence de mots et de verbes généralistes peut générer une confusion entre ces titres et un extrait de la description détaillée de l’image, qui est une forme d’expression courante dans les publications des historiens de l’art. Or, il s’agit de rechercher l’appellation d’une œuvre ou d’une image et non un segment d’une description longue.

Ce corpus annoté a servi de corpus de référence pour le développement et l’évaluation d’un outil de reconnaissance automatique des appellations fondé sur des méthodes symboliques. La reconnaissance s’appuie sur la création de patrons syntaxiques utilisant un ensemble de vocabulaires construits pour le projet à partir de lexiques existants et du lexique nouveau repéré durant l’annotation manuelle.

6. L’annotation automatique

6.1. La méthode choisie

Les appellations d’œuvres d’art peuvent être considérées comme les entités nommées car elles se réfèrent ‘à une entité unique et de manière autonome dans le corpus’ (Ehrmann 2008). La reconnaissance des entités nommées (REN) est une tâche fréquente dans le TAL qui exploite des méthodes et techniques variées : les méthodes symboliques (Díez Platas *et al.* 2020 ; Collin et Guerraz 2015), l’apprentissage supervisé de surface (Brandesen *et al.* 2020) et profond (Brandesen *et al.* 2021).

Pour ce projet, le choix s’est porté sur l’utilisation des méthodes symboliques pour plusieurs raisons. La récurrence repérée dans les composants et la structure interne des appellations d’œuvres et des thèmes iconographiques de l’art classique dans les textes grecs (Prioux s. d.) est aussi dans les textes en français. De plus, dans une visée plu-

Ressources pour l'étude des appellations d'œuvres visuelles de l'Antiquité classique : corpus, dictionnaires et outil de reconnaissance automatique

ridisciplinaire, l'emploi d'une méthode opaque comme l'apprentissage profond aurait été un frein dans la réutilisation de l'outil par un public non TAListe. Or, la familiarisation des historiens de l'art, participants du projet, avec les approches et les outils numériques TAL abordables est un attendu du projet. La capacité de l'outil sélectionné à proposer une représentation claire et visuelle du processus de reconnaissance automatique, son accessibilité et sa prise en main simple et rapide ont été des critères décisifs dans le choix de la méthode et de l'outil dans MonumentAL. D'autre part, en considérant la taille du corpus à annoter (591 711 tokens), le choix d'une méthode symbolique a semblé le plus pertinent. Le développement de l'outil de reconnaissance automatique des appellations est donc fondé sur des méthodes symboliques.

6.2. L'outil d'annotation automatique

La méthode proposée par cette recherche s'appuie sur la création de patrons syntaxiques utilisant les dictionnaires⁸ créés spécialement pour le projet et formalisés sous Unitex (Paumier, 2009). Les patrons Unitex ont été développés sur 40% du corpus annoté manuellement, puis testés sur 30% du corpus et enfin évalués sur les 30% restants.

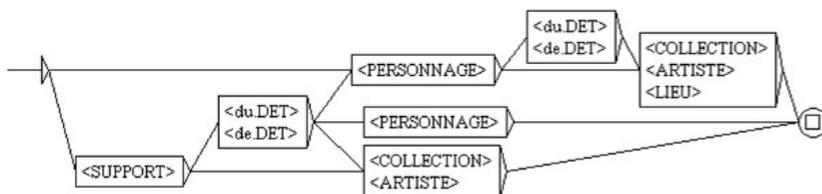


FIG. 2 – Graphe Unitex (simplifié) pour la reconnaissance d'Appellation Structurée.

8 Par convention dans cet article, un dictionnaire préparé pour Unitex est nommé dictionnaire, pour le différencier d'un thésaurus, et il est écrit en petites capitales.

Les patrons syntaxiques de reconnaissance des appellations d'œuvre dépendent des termes présents dans les dictionnaires. Ainsi, les patrons syntaxiques tels que PERSONNAGE + ARTISTE (*Zeus de Phidias*) ou SUPPORT+ PERSONNAGE + COLLECTION (*statue de Mercure du musée du Vatican*) reconnaissent une appellation d'œuvre uniquement si les unités lexicales la composant sont présentes dans les dictionnaires.

Pour améliorer la précision de certains patrons syntaxiques, une sélection de termes spécifiques est nécessaire. À titre d'exemple, le dictionnaire PERSONNAGE contient aussi bien des noms de divinité, que des noms de fonctions et des gentilés. Son utilisation en l'état pour une reconnaissance d'Appellation Courte (déterminant + PERSONNAGE) produit trop de bruit : là où *l'Apollon* est quasiment assuré d'avoir pour référent une œuvre visuelle, *l'empereur*, *l'Amazone* ou *les Romains* font plus probablement référence à des individus. Afin d'améliorer la précision de la reconnaissance d'Appellation Courte, il est ainsi nécessaire d'utiliser un sous-groupe du dictionnaire PERSONNAGE qui contient uniquement des individus dont le nom propre correspond à un personnage distinct et qui n'est pas générique.

6.3. Évaluation de la reconnaissance automatique

Les résultats de la reconnaissance automatique par types d'appellation sont présentés dans le Tableau 2.

Type d'appellation	Précision	Rappel	F-mesure
Appellation Structurée	0,72	0,84	0,78
Thème	0,85	0,72	0,78
Appellation Courte	0,87	0,84	0,89

TAB. 2 – Évaluation de l'outil de reconnaissance des appellations d'œuvres.

Ressources pour l'étude des appellations d'œuvres visuelles de l'Antiquité classique : corpus, dictionnaires et outil de reconnaissance automatique

L'Appellation Courte est la catégorie la mieux reconnue avec un score de F-mesure de 0.89. Les Thèmes obtiennent une F-mesure de 0,78 due à leur ambiguïté avec les descriptions détaillées de scène, une forme d'expression spécifique des textes sur l'art. Enfin, les Appellations Structurées sont reconnues également avec une F-mesure de 0,78 qui peut être expliquée entre autres par l'ambiguïté de certaines appellations pouvant avoir plusieurs référents (*Zeus Olympien* peut se référer à la divinité grecque ou à la statue de Phidias).

On peut mettre en relation ces résultats (F-mesure comprise entre 0,78 et 0,89) avec des travaux sur des données comparables. Même si des données traitées ne sont pas identiques, des méthodologies symboliques y sont également utilisées afin de reconnaître des entités nommées dans les textes médiévaux espagnols (Díez Platas *et al.* 2020) pour lesquels la F-mesure obtenue est entre 0,74 et 0,87, ou pour les titres de films (Collin et Guerraz 2015) où la F-mesure est de 0,63.

En comparaison avec ces travaux, les résultats obtenus par cette recherche (tableau 2) sont meilleurs. Pour le domaine des humanités numériques, Brandsen *et al.* utilisent d'autres techniques comme l'apprentissage automatique de surface (F-mesure de 0,70) (Brandsen *et al.* 2020) et l'apprentissage profond (F-mesure moyenne de 0,735) (Brandsen *et al.* 2021). Ces résultats sont comparables à ceux obtenus en utilisant les méthodes symboliques dans cette phase du projet.

7. Dictionnaires et thésaurus

La phase d'annotation manuelle d'un extrait du corpus (40% du corpus) a démontré une récurrence dans les composants construisant les appellations d'œuvres et les thèmes iconographiques. Le bon fonctionnement de la reconnaissance automatique repose sur l'utilisation de dictionnaires dédiés réunissant des termes actuels ou anciens, des termes spécifiques à l'histoire de l'art ou à la culture antique, présentant donc de nombreuses variations. Celles-ci dérivent notamment de la francisation et de la translittération irrégulières des termes issus du latin et du grec ancien (Ikaros / Icaros / Ikarus / Icarus / Icare). L'utilisation de thésaurus préexistants était un préalable à ce projet à

cause du volume important de ce vocabulaire particulier. Plusieurs des composants des appellations trouvent des correspondances étroites avec plusieurs notions clés fréquemment convoquées dans les données relatives au patrimoine culturel : type d'objet, personnage représenté, agents de la création (artiste, commanditaire) et de la conservation (collectionneur, collection et musée), lieu. On les retrouve à la fois dans des modélisations fondées sur des standards d'interopérabilité, ou des ontologies comme le CIDOC CRM⁹, et dans des thésaurus et référentiels du domaine.

Dans le projet, les dictionnaires de plusieurs composants ont d'abord été générés à partir du thésaurus TheA (Thésaurus-Antiquité du LIMC), dont l'élaboration a été faite d'après des corpus d'objets antiques figurés (Szabados 2014). Il s'agit des dictionnaires ARTISTE, MATÉRIAU, SUPPORT (types d'objet), PERSONNAGE, LIEU et COLLECTION. L'annotation manuelle a montré la nécessité de créer deux nouveaux dictionnaires : THÈME et ADJECTIF spécifiques au domaine. Cette phase et la reconnaissance automatique a permis d'enrichir TheA avec de nouveaux termes, y compris des variantes, des synonymes et des cacographies (mots erronés par l'océrisation). La génération des dictionnaires (en liste simple) à partir de ce thésaurus est facilitée par un script python et il est apparu plus judicieux de maintenir ce processus car TheA est compatible avec la norme SKOS des thésaurus¹⁰ qui permet de différencier la forme préférentielle (normalisée) du terme (skos:prefLabel : Aphrodite Anadyomène), de ses variantes (skos:altLabel : Kypris / Cypris / Cythérée Anadyomène) et de ses cacographies (placées dans skos:hiddenLabel : Aphrodite anadyomene / Anadyomône), et d'associer des alignements, pour ce projet avec le référentiel wikidata.

Au final, le processus de reconnaissance automatique permet aussi, par la récolte rendue possible des appellations, de créer un nouveau thé-

9 Schéma montrant les classes de haut niveau, auxquelles on peut ajouter *E57 Material* pour les matériaux : Bekiari *et al.* 2022, 34-37, fig. 1 ; un exemple pour le Laocoon : fig. 2.

10 Simple Knowledge Organization System : <https://www.w3.org/2004/02/skos/>

Ressources pour l'étude des appellations d'œuvres visuelles de l'Antiquité classique : corpus, dictionnaires et outil de reconnaissance automatique

saurus des titres et appellations thématiques d'œuvres de l'art antique classique, ŒUVRE, l'un des attendus de MonumentAL.

Le Tableau 1 présente la quantité de termes dans les sept dictionnaires principaux et dans ŒUVRE en juin 2022. Le nombre de termes de TheA alors non repérés dans le lot de textes traités a été évalué pour les ensembles les plus denses (ARTISTE, COLLECTION, LIEU et PERSONNAGE). La colonne 'total' correspond au nombre de termes par dictionnaire utilisés dans le processus de reconnaissance automatique ; le total des termes des dictionnaires est de 17 335.

	prefLabel	altLabel et hiddenLabel	termes TheA non repérés	total
artiste	549	291	958	1 798
collection	641	628	2 791	4 060
lieu	398	206	3 989	4 593
matériau	827	9	(non évalué)	836
personnage	1 306	1 774	1 356	4 436
support	252	135	(non évalué)	387
thème	392	258	-	650
œuvre	575	(non évalué)	-	575

TAB. 1 – Quantification des termes dans les dictionnaires principaux (juin 2022).

Les premiers résultats de la reconnaissance automatique ont montré que presque tous ces dictionnaires principaux devaient être subdivisés pour créer de nouveaux dictionnaires afin de nuancer le fonctionnement de la reconnaissance en affinant la répartition des termes dans les patrons linguistiques. Il s'agit d'atteindre un équilibre entre les notions exprimées, la nature des mots et leur place dans les types d'appellation. Les sous-groupes sont distincts dans le thésaurus (TheA et ses

micro-thésaurus) et le restent dans les quinze dictionnaires finaux. Les spécificités et les subdivisions principales des dictionnaires sont :

- **PERSONNAGE** représenté en tant que sujet d'une œuvre visuelle antique :

→ subdivision noms propres / noms communs,

→ subdivision personnages mythiques / historiques,

→ subdivision nom propre d'individu distinct / nom générique ou de groupe :

- distinctions entre : divinité, héros, être et peuple mythiques / hybride / animal légendaire (*Pégase*) / personnification (*Eirene* ou la *Paix*, le *Nil*),
- distinctions entre : personnage générique (*Satyre.s*, *griffon.s*) / groupe (*titan.s*, *Muse.s*) / gentilé (*Amazone.s*),
- distinctions entre : personnage historique (*Périclès*) / animal réel (*cheval*) / fonction (*prêtre*) / état (*enfant*).

- **SUPPORT** et type de l'image, distinctions entre :

- type d'objet portant une figuration (*bas-relief*, *cratère en calice*, *pélikè*),
- nom de matériau utilisé pour désigner un objet (un *bronze*, un *marbre*),
- type d'image et notions de représentation (*copie*, *imitation*, *tirage*, etc.) : ce groupe reste ici car il occupe la même place que le type d'objet dans l'AS.

- **ARTISTE ; COLLECTION ; LIEU** : exploités de la même façon pour l'AS, ils distinguent traditionnellement l'œuvre quand ils sont associés au **PERSONNAGE**.

- **ARTISTE** rassemble les noms d'artistes antiques (Phidias → : *Zeus de Phidias*), dont les peintres de vase selon la tradition de nommage (manière du Peintre d'Achille ; reverse-group of Ferrara T 463¹¹),

11 Récupération à partir de TheA de versions anglaises des noms à cause de la prépondérance de la tradition anglophone enracinée par J. D. Beazley, A.D. Tren-

Ressources pour l'étude des appellations d'œuvres visuelles de l'Antiquité classique : corpus, dictionnaires et outil de reconnaissance automatique

- **LIEU** correspond aux lieux de création et d'usage antiques, de trouvaille ou de conservation de l'œuvre figurée (Annaba / Bône / Hippone / Hippo Régius; Cos / Kos; le Palatin; Amyclées / Amyklai → *trône d'Amyclées*),
- **COLLECTION** est fondé sur la notion de regroupement d'œuvres ou d'images (une collection correspond à un ensemble). On y retrouve :
 - nom de collectionneur (Médicis¹² → *Vénus de Médicis*),
 - nom de personne morale, comme un musée ou une institution (Musée du Prado → *Diadumène du Prado*),
 - édifice antique porteur d'un décor architectural (Parthénon → *Frise / métopes du Parthénon*) ou contenant des œuvres visuelles (Lesché des Cnidiens, portique d'Octavie, tombeau de Mausole)¹³.
- **THÈME** réunit en trois groupes le lexique courant des thèmes iconographiques :
 - terme qui désigner un thème à lui seul (*psychostasie, bacchante*),
 - nom commun sur lequel l'appellation est focalisée (*apothéose...*),
 - nom commun utilisé comme complément (*strigile, couronne, lierre*).
- **ADJECTIF** contient les adjectifs associés aux composants SUPPORT, MATÉRIAU, PERSONNAGE et THÈME. On y distingue les mots du domaine (*phidiesque*: en lien avec Phidias) et des épiclèses qui sont des épithètes précisant notamment des attributions, des particularités ou des lieux liés aux divinités antiques. Les épi-

dall et A. Cambitoglou.

12 La famille Médicis, cf. www.wikidata.org/wiki/Q170022.

13 Placer les édifices ici résout une porosité dans l'association support (objet/monument) + personnage. Et, après constat issu de l'annotation, les termes 'tombe' et 'tombeau' sont dans ce groupe pour l'antique, contrairement à l'usage pour les courants artistiques à partir du médiéval.

clèses sont plus ou moins francisées (*Artémis Éphésienne / Artémis Ephesia*).

8. Constats, améliorations et perspectives

Depuis l'évaluation, des améliorations ont été apportées dans les dictionnaires et les patrons linguistiques pour optimiser les résultats de la reconnaissance automatique. Un enrichissement de LIEU, notamment, a fortement diminué le bruit généré par des associations ambiguës non anticipées, comme celle associant un nom de lieu au mot 'marbre' qui est utilisé à la fois pour nommer un matériau et un type d'objet (*'Marbres' Elgin* pour les décors du Parthénon au British Museum). De nombreux noms de marbre (matériau) sont distingués par le nom du lieu de la carrière d'exploitation : *marbre de Luni*, *marbre de Paros*. Ces termes étaient identifiés lors de la reconnaissance comme des objets d'art et non des matériaux car la conception des patrons linguistiques du projet permet de repérer des nouveaux mots commençant par une majuscule (pour trouver de nouvelles appellations). Les résultats se sont améliorés en enrichissant le dictionnaire LIEU car un segment comme 'marbre de Luni' est identifié comme un matériau (Luni alors reconnu comme lieu) et ne devient exploitable, pour repérer une appellation d'œuvre, qu'associé à un autre composant pertinent du patron linguistique. Il est donc envisagé de tester d'autres enrichissements de LIEU à partir de gazetteers sur le monde antique, comme Pleiades ou PACTOLS, ou plus généralistes (Geonames ; TGN Getty¹⁴) mais avec un risque de distorsion lié à un accroissement des homonymes entre les dictionnaires LIEU, COLLECTION, PERSONNAGE.

Les bases posées lors de cette première phase du projet forment un socle solide pour l'élargissement de la reconnaissance et de l'annotation automatiques à d'autres formes d'appellations, et même au repérage d'objets cités sans titre propre, en adaptant les patrons linguistiques. La prise en compte des ponctuations servant de limites extrêmes pour l'Appellation Courte (– *Apollon*.) a réduit le silence imposé par la perte de

14 pleiades.stoa.org; www.geonames.org/; www.getty.edu/research/tools/vocabularies.

Ressources pour l'étude des appellations d'œuvres visuelles de l'Antiquité classique : corpus, dictionnaires et outil de reconnaissance automatique

la mise en page. Le projet DataCatalogue (Scheithauer 2022) porte justement sur la transformation des pdf des catalogues de vente d'œuvres d'art en fichier XML-TEI restituant le découpage particulier du texte en notices semi-structurées (comme dans un catalogue d'exposition). L'exploitation de la mise en page pourra être testée pour les trois types d'appellations sur les fichiers qui sont déjà mis à disposition : le titre succinct et la description de chaque notice seront directement ciblés grâce aux balises TEI spécifiques qui les délimitent. Pour des analyses statistiques, on pourrait avancer l'hypothèse que des *Aphrodite de Cnide* citées dans les titres de notice d'un catalogue de vente ancien seraient forcément des copies et non l'original.

Cette approche peut apporter, dans ce cas, une part de la réponse au problème de la distinction entre la citation d'une œuvre matérielle originale (*Aphrodite de Cnide de Praxitèle*), de ses dérivés (copie de l'*Aphrodite de Cnide*) et du modèle iconographique conceptuel (type de l'*Aphrodite de Cnide*), problème qui n'est pas complètement résolu par la prise en compte des mots comme 'copie' dans SUPPORT ou la présence du nom de l'artiste dans le titre car la formule succincte *Aphrodite de Cnide* est employée pour les trois. Leur distinction pose aussi la question du choix du terme dans un thésaurus pour la forme normalisée des titres et celle des types iconographiques. Est-il judicieux dans ŒUVRE (compatible SKOS) de choisir l'appellation la plus distinctive pour la forme normalisée des appellations (prefLabel:Aphrodite de Cnide de Praxitèle) alors qu'*Aphrodite de Cnide* est la variante la plus usitée du nom de la statue originale et la version choisie pour le 'Label' de wikidata¹⁵? L'enjeu est d'éviter d'identifier une réplique de même nom comme original de Praxitèle au moment d'une indexation à partir d'un thésaurus pour ce domaine. Une partie de la réponse est dans les habitudes d'écriture des auteurs, que peuvent faire ressortir l'approche TAL et les analyses textuelles.

15 www.wikidata.org/wiki/Q618535. Wikidata est de plus en plus la référence pour les alignements.

9. Conclusion

Le foisonnement, en nombre et en variations, des titres d'œuvres visuelles classiques, difficile à maîtriser, freine l'étude de l'évolution de ces appellations et complique le choix de celles qui peuvent prendre le rôle de référence. Pour répondre à cet objectif du projet MonumentAL, cette collaboration réunissant les historiens de l'art et les linguistes informaticiens, a privilégié une approche de TAL qui repose sur l'exploitation de dictionnaires et de thésaurus du domaine préexistants. Le processus d'annotation automatique avec des méthodes symboliques obtient, pour les désignations d'œuvres d'art antique classique dans un corpus français relevant de l'histoire de l'art, une F-mesure variant entre 0,78 et 0,89. Grâce à ces résultats engageants, un travail équivalent a été commencé au printemps 2022 sur les corpus textuels antiques en latin et en grec ancien. Des tests sur des ouvrages denses (de centaines de pages) et des séries (fascicules de revue) ont depuis démontré la rapidité de l'application et du traitement des graphes et la prise en main facile de ces outils.

La récolte des titres, des composants et des termes des dictionnaires, qui porte ses fruits avec la création du thésaurus ŒUVRE, ouvre des perspectives vers de l'analyse textuelle pour l'étude de l'évolution chronologique des appellations – notamment autour des termes caractéristiques d'un thème iconographique ou du constat d'un remplacement des noms de personnages grecs par leur version latine à certaines périodes –, pour repérer les appellations candidates comme terme de référence, ou pour évaluer la répartition des modes d'appellation selon la typologie des publications et les lectorats visés, ou encore la fortune des œuvres dans les publications ou leur citation dans la littérature.

Références

Bekiaris, Chryssoula, George Bruseker, Martin Doerr, Christian-Emil Ore, Stephen Stead, et Athanasios Velios, éd. 2022. "Definition of the CIDOC Conceptual Reference Model". <https://cidoc-crm.org/Version/version-7.2.1>.

Ressources pour l'étude des appellations d'œuvres visuelles de l'Antiquité classique : corpus, dictionnaires et outil de reconnaissance automatique

- Biasi, Pierre-Marc de, Marianne Jakobi, et Ségolène Le Men, éd. 2012. *La Fabrique du titre, Nommer les œuvres d'art*. Paris : CNRS éditions.
- Brandsen, Alex, Suzan Verberne, Karsten Lambers, et Milco Wansleeben. 2020. "Creating a Dataset for Named Entity Recognition in the Archaeology Domain". In *Proceedings of the 12th Conference on Language Resources and Evaluation*, 4573-77. Marseille. <http://www.lrec-conf.org/proceedings/lrec2020/index.html#4573>.
- Collin, Olivier, et Aleksandra Guerraz. 2015. "Classification d'entités nommées de type 'film' ». In *Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles*, 364-370. Caen. talnarchives.atala.org/TALN/TALN-2015/taln-2015-court-008.pdf.
- Díez Platas, M^a Luisa, Salvador Ros Muñoz, Elena González-Blanco, Pablo Ruiz Fabo, et Elena Álvarez Mellado. 2021. "Medieval Spanish (12th–15th Centuries) Named Entity Recognition and Attribute Annotation System Based on Contextual Information". *Journal of the Association for Information Science and Technology* 72 (2): 224-38. doi:10.1002/asi.24399.
- Ehrmann, Maud. 2008. "Les Entités Nommées, de la linguistique au TAL : Statut théorique et méthodes de désambiguïsation". Thèses, Université Paris Diderot. <https://hal.archives-ouvertes.fr/tel-01639190>.
- Landis, Richard J., et Gary G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data". *Biometrics* 33 (1): 159-74.
- Michon, Étienne. 1918. *Catalogue sommaire des marbres antiques*. Paris : Musée du Louvre..
- Paumier, Sébastien, Takuya Nakamura, et Voyatzi Stavroula. 2009. "UNITEX, a Corpus Processing System with Multi-Lingual Linguistic Resources". In *eLexicography in the 21st century: new challenges, new applications (eLEX'09)*, 173-75. France.
- Prioux, Évelyne. 2011. « Images de la statuaire archaïque dans les Aitia de Callimaque », *Aitia. Regards sur la culture hellénistique au XXI^e siècle*, n° 1. ENS Éditions. doi:10.4000/aitia.74.
- Prioux, Évelyne. s. d. (à paraître) "Les titres et désignations antiques des œuvres d'art célèbres". In *Vocabulaire des collectionneurs de*

- l'Antiquité à la fin du XIX^e siècle. Aix-en-Provence*, édité par Pedro Duarte et Florence Le Bars-Tosi, 31 p. PUP/CPAF.
- Reinach, Salomon. 1895. *Pierres gravées des collections Marlborough et d'Orléans, des recueils d'Eckhel, Gori, Levesque de Gravelle, Mariette, Millin, Stosch*. Paris : Firmin-Didot.
- Scheithauer, Hugo, Laurent Romary, Frédérique Duyrat, et Federico Nurra. 2022. "DataCatalogue : présentation du projet". <https://hal.inria.fr/hal-03618381>.
- Szabados, Anne-Violaine. 2015. "From the LIMC Vocabulary to LOD. Current and expected Uses of the Multilingual Thesaurus TheA". In *Information Technologies for Epigraphy and Cultural Heritage. First EAGLE International Conference*, édité par Silvia Orlandi, Raffaella Santucci, Vittore Casarosa, et Pietro Maria Liuzo, 51-67. Paris : Sapienza Università Editrice. <https://www.eagle-network.eu/wp-content/uploads/2015/01/Paris-Conference-Proceedings.pdf>.
- Widlöcher, Antoine, et Yann Mathet. 2012. "The Glozz platform : a corpus annotation and mining tool". In *Proceedings of the ACM Symposium on Document Engineering, Sept 2012. Paris*, 171-80. Paris. <https://hal.archives-ouvertes.fr/hal-01023774>.

Abstract

The multidisciplinary project MonumentAL aims to identify and classify the titles and appellations given to ancient works of art by using NLP and text mining methods. This project involves a close collaboration between art historians (LIMC), NLP researchers (MoDyCo), curators and librarians (BnF). This presentation is focused on a specific part of the project, namely the analysis of texts written in French and published between the 18th and the 21st century. Our procedure involves several stages: selection of the textual corpus of study, elaboration of a typology of appellations, constitution of an annotated corpus (annotated manually by experts in the field), creation of reusable dictionaries of terms, of a new thesaurus of artworks name and development of a tool for automatic recognition of appellations based on symbolic methods.

