



PERSPECTIVES OFFERTES PAR UN CORPUS ÉLECTRONIQUE DE GLOSES SUR PRISCIEN

FRANCK CINATO
EPHE PARIS

Résumé

Les perspectives de recherches offertes par tout corpus électronique de texte dépendent entièrement des choix techniques qui ont présidé à leur élaboration. Dans le cas d'une édition de gloses, les réflexions préliminaires à la constitution de corpus montrent que le philologue ne peut faire l'économie des aspects techniques qui aboutiront à l'encodage des données. En raison des particularités du matériel à exploiter — les gloses —, les aspects à prendre en compte dépassent largement le cadre textuel strict, tant il est important, sinon essentiel, d'ajouter quantité d'informations métalinguistiques et factuelles. Cette contribution tente de faire le point sur l'ensemble des informations se révélant décisives dans le cadre d'un projet de corpus électronique de gloses.

Abstract

The research perspectives offered by any electronic text corpus depend entirely on the technical choices that led to their development. In the case of an edition of glosses, preliminary discussions to corpus creation show that the philologist can not deal without the technical aspects that will lead to the encoding of data. Because of the material's special nature to manage - the glosses -, aspects to consider go far beyond strictly textual, it is as important if not essential, to add much more meta-linguistic and factual information. This contribution attempts to review all decisive information needed in the context of electronic corpus of glosses project.

La préparation des journées d'études *Sodalitas* organisées par Alessandro Garcea a été l'occasion d'amorcer une réflexion autour des éditions électroniques d'un matériel textuel particulier : les gloses¹. Celles reposant sur l'*Ars grammatica* de Priscien ont été l'objet de ma thèse² : de ce fait les problématiques inhérentes aux gloses seront abordées de manières restrictives à ce sujet, bien que *de facto* elles peuvent aisément s'appliquer à d'autres corpus importants de même nature, notamment les collections de gloses bibliques³ et les glossaires⁴.

Il s'agit d'abord de décrire les spécificités de données glossographiques, pour qu'ensuite les stratégies appliquées lors de leur constitution en un corpus électronique puissent en tenir compte adéquatement. C'est donc l'occasion ici – à défaut de proposer des solutions toutes faites – de cerner précisément les besoins et les problèmes d'encodage liés à cette documentation très spécifique.

1. Les gloses

1.1. Définition

Avant tout, il convient de définir brièvement en quoi consistent les gloses. Hors considérations matérielles ou historiques, elles sont des outils de travail polyvalents et en tant que tels, elles composent un ensemble de stratégie visant à la transmission de savoirs⁵.

¹ La reprise intégrale du texte m'a permis d'apporter des suggestions plus précises aux problématiques qui seront évoquées. Je remercie ici Padraic Moran pour ses suggestions, et ses documents de travail sur les gloses de *St-Gall 904* à la source du site (<http://www.stgallpriscian.ie/>).

² Thèse de l'EPHE (4^e sect.) : « *Glose de Prisciano*, Travaux de maîtres carolingiens sur l'*Ars* de Priscien, Gloses interlinéaires, marginales et collectées », soutenue en novembre 2010.

³ Voir VACIAGO (éd) 2004 et EISENHUT 2010. Voir également sur ce domaine l'entreprise www.glossae.net initiée par N. Bériou, M. Burghart et M. Morard.

⁴ Les glossaires médiévaux n'ont rien de commun avec les dictionnaires modernes, et pour lesquels il serait une grave erreur de chercher à leur appliquer les balisages prédéfinis par la TEI.

⁵ Voir par exemple, trois grammaires glosées de Priscien en ligne : St-Gall, SB, 903, copié vers 800 en Italie (Vérone ?) (<http://www.e-codices.unifr.ch/en/preview/csg/0903>) ; St-Gall, SB, 903, copié vers 850 en Irlande (?) (<http://www.e-codices.unifr.ch/en/preview/csg/0904>) ; Köln, Erzbischöfliche Diözesan- und Dombibliothek, 200, copié dans le centre de l'Allemagne dans le second tier du IX^e siècle (http://www.ceec.uni-koeln.de/ceecggi/kleioc/0010/exec/pagemed/%22kn280200_001.jpg%22/segment/%22body%22).

Les divers *marginalia* qui composent le péri-texte de Priscien, ne remplissent pas tous les mêmes fonctions. Ainsi, une annotation, qui de manière formelle s'apparenterait à une glose, ne répond peut-être pas nécessairement à la définition stricte. Avant tout, il faut être averti que sous le vocable « glose » se cache en réalité un « couple indissociable » : un mot à expliquer, le lemme et son explication, la glose proprement dite (voir ci-dessous 2.3.5). Or, il s'agit déjà d'un glissement du sens original, car l'étymologie du mot *glossa* renvoie précisément à ce que nous appelons le lemme, c'est-à-dire l'élément qui demande une explication. Dans ma thèse, j'ai proposé cette définition du phénomène :

Glose manuscrite : toute augmentation péri-textuelle qui précise ou diversifie l'information contenue dans un texte principal.

On comprend ainsi qu'une annotation ne se rapportant pas au texte par le truchement du lemme sera exclue de la famille des gloses, car relevant des notes socio-historiques (invocation, colophons, etc.) ou des *marginalia* fonctionnels (réclames, signatures etc.), tout comme les *ex-libris*⁶, etc. En revanche, certaines variantes de lectures seront à verser dans le corpus des gloses, dans la mesure où elles apportent de l'information pour l'établissement du texte. Les gloses servent donc de véhicule à une importante variété d'explications, elles peuvent porter sur un mot en particulier et expliciter sa nature grammaticale, son étymologie etc. ; ou bien elles servent à discuter des doctrines grammaticales exposées par le grammairien. Une distinction typologique permet immédiatement de cerner les niveaux d'interventions des gloses (voir ci-dessous 2.3.6.1). Elles remplissent un double objectif en fonction de l'auditoire : adressées à l'élève, elles servent de support pédagogique ; à l'intention des maîtres, elles contiennent des discussions scientifiques destinées à d'autres maîtres.

1.2. Collections de gloses

Les gloses d'un témoin forment un ensemble que nous appellerons par convention une collection. Bien qu'unique à certains égards, elle partage inmanquablement des points communs avec une ou plusieurs autres collections. Mais par « collection », faut-il comprendre l'ensemble des gloses d'un seul témoin manuscrit ? Il s'avère que les choses ne sont pas aussi simples. En effet, Si les grammaires glosées de Priscien en sont les témoins privilégiés, elles ne sont pas les seules à transmettre des gloses sur ce grammairien. Certains manuscrits proposent des collections où les gloses ont été copiées bout à bout sans leur texte d'appuis. Ces *glossae collectae* ont à leur tour été recopié avec pour résultat qu'une même collection peut être connue par plusieurs témoins manuscrits. Parmi eux, certains mêlent plusieurs collections.

⁶ Voir TURA 2005.

1.3. Typologie et transmission

L'approche typologique des gloses, au-delà d'un simple classement, permet de mettre en évidence les préoccupations des glossateurs, ainsi que la spécificité de leurs enseignements. Je n'entrerai pas plus ici dans les détails en ce qui concerne la diversité typologique du contenu des gloses sur les grammairiens (voir annexe 2), mais un dernier aspect classificateur doit encore être évoqué. Il s'agit des informations « externes », liées à leur transmission, aspects proprement paléographiques, qui se révèlent d'une importance capitale dans le cadre de l'étude des gloses. La distinction des écritures détermine des éléments essentiels à l'analyse. Les premiers qui viennent à l'esprit ont trait à la localisation et la datation des mains : un même manuscrit a souvent été étudié en différents lieux et à différentes époques. Il faut donc – pour rendre compte de la perspective diachronique au sein de la collection – établir une chronologie relative des mains qui ont participé à la formation du peri-texte. Cette étape n'est possible que dans le cas des sources de type Γ défini plus bas (voir 2.3.1.). Lorsque la collection de gloses émane d'un autre type de source, en général, les gloses ont été recopiées d'une même main réduisant à néant la diachronie sous-jacente. Ainsi, la datation proposée sera la même pour toutes les entrées de cette collection, en gardant à l'esprit qu'il s'agit là d'une l'indication servant de *terminus* et non de la date de création de l'explication. Nous touchons à l'un des points souvent cause de confusions : les glossateurs rédigent de nouvelles gloses moins souvent qu'ils n'en copient d'anciennes. Les compétences attendues sont très différentes, puisque dans le premier cas seul, il sera permis de désigner un glossateur comme auteur de gloses.

Après avoir caractérisé à grands traits la glose et sa transmission, il faut s'entendre sur le sens à donner à la locution « corpus électronique de gloses ». Il ne saurait être question d'envisager le corpus en question comme un ensemble unique de toutes les gloses se rapportant à Priscien, au risque de constituer un énorme fourre-tout sans cohérence.

2. Les enjeux des corpus électroniques : structure du corpus

2.1. Composants

L'architecture d'un hypothétique « corpus électronique de gloses » ne peut reposer sur un ensemble homogène de texte, comme c'est le cas pour le *CGL* hyper-texte (*Corpus Grammaticorum Latinorum*)⁷. En raison des spécificités du matériel textuel, les briques essentielles constituant le corpus devront être organisées selon une hiérarchie où les témoins manuscrits seront subsumés aux

⁷ Voir A. GARCEA, *CGL* <http://kaali.linguist.jussieu.fr/CGL/>

collections transmises. C'est pourquoi le corpus devrait être structuré en fonction, d'une part, des œuvres de Priscien et, d'autre part, selon les sources des collections de gloses (types). Certaines informations figureront en toute lettre dans le fichier source ; d'autres devront être générées automatiquement par l'interface en fonction des requêtes des utilisateurs. Si l'utilisation du langage XML TEI (Text Encoding Initiative)⁸ s'avère le mieux adapté en vue de l'exploitation des données, nous ne ferons que très peu de suggestions techniques (en terme d'éléments et d'attributs à employer lors de l'encodage), voire aucune dans la plupart des cas. Les remarques exprimées ici seront surtout d'ordre théorique, dans le but d'alimenter la réflexion sur le sujet, plutôt que de la clôre. Les informations à encoder étant de natures différentes, il convient de tenir compte d'une première distinction générale : a) le texte glossographique proprement dit, qui constitue le contenu fondamental ; b) les informations complémentaires qui lui sont associées (systèmes de références et données "cachées", etc.).

2.2. Technologie adaptée

Le contenu fondamental des collections – lemmes et gloses – devra être encodé de manière à permettre la mise en application de technologies visant à faciliter l'exploitation scientifique des données. Sous le mot technologie, il faut entendre les outils informatiques permettant d'opérer – grâce à des automatisations – diverses manipulations et traitements des informations. La mise en œuvre de ces technologies se justifie d'autant plus facilement dans le cas des gloses que leur volume important les rend peu manipulables sans cela. Pour se faire une idée du volume en question, il suffit de quelques chiffres : la grammaire du manuscrit de St-Gall 904 (siglé *G*) porte plus de 9000 gloses ; les gloses collectées du Vatican, Reg. lat. 1650 comptent plus de 4000 gloses couvrant seulement la moitié de l'*Ars*. Ce qui correspond déjà, en ne tenant compte que de ces deux sources de gloses, à plus de 13000 entrées, soit presque l'équivalent du nombre total des citations d'auteurs présentes dans tous les textes des grammairiens latins retenus pour le *CGL*. Avec une dizaine de collections seulement, le nombre d'entrées du corpus pourrait dépasser les 100 000. Il est possible d'entrevoir à ce stade quelles incidences pourront avoir les choix techniques, tant en terme de conception que de matériel physique. Si l'on écarte le dernier point, hors de propos ici, les enjeux initiaux sont liés à la *structure*, à la *sélection* et à l'*encodage* des données constituant le corpus. Ce corpus de gloses ouvrira autant de possibilités d'exploitations que le contenu des métadonnées sera organisé pour tenir compte des spécificités intrinsèques des gloses. La qualité des résultats et les possibilités qui seront offertes par le corpus dépendront de la richesse et de la variété des informations retenues : elles constitueront autant de critères de sériations et d'outils d'analyse.

⁸ Voir <http://tei-c.org>

2.3. Organisation des Métadonnées

Résumons les informations évoquées en les hiérarchisant :

1. Collections de gloses
2. Auteur (ici Priscien)
3. Œuvres grammaticales (*Ars grammatica*, etc.)
4. Témoins manuscrits
5. Lemmes
- 6.1 Gloses / type, etc.
- 6.2 Gloses / main, etc.
7. Sources

2.3.1. Collections de gloses

Elles constituent les briques élémentaires du corpus de données, strictes équivalent des textes composant le *CGL*, mais contrairement à une œuvre littéraire au sens strict, elles se présentent sous trois faciès particuliers en rapport avec leur mode de transmission. Chaque type nécessitera d'apporter des ajustements particuliers au système de référence, puisque l'identification et le référencement des gloses dépendent des collections qui les transmettent.

– Type Γ .

Grammaire glosée (ou extraits de la grammaire): collection propre à un témoin manuscrit unique où les gloses se trouvent à proximité immédiate du texte (intégral ou non) ; elles résultent souvent des travaux de glossateurs qui se sont succédés durant une période plus ou moins longue.

– Type γ .

Recueil de *glossae collectae* : ces collections peuvent avoir bénéficié d'une tradition manuscrite, impliquant la possibilité de témoins multiples ; les gloses se succèdent les unes à la suite des autres, dans l'ordre du texte dont le rôle est tenu par les lemmes seuls ; généralement copié d'une seule main (diachronie « écrasée ») et proposant une orientation thématique.

– Type Δ .

Gloses citées : cet ensemble constitue en réalité un groupe factice dans lequel sont rassemblées les gloses qui apparaissent dans d'autres contextes (principalement commentaires ou glossaires) ; s'agissant de témoins indirects de telle ou telle tradition de gloses, le contexte de la citation se superpose au contexte priscienique, nécessitant une double référence (au texte de Priscien et à celui où se trouve la citation). – Le groupe des gloses entrées dans le texte de Priscien à date ancienne pourrait prendre la forme d'une collection de ce type.

L'exemple de l'Encadré 1 montre que des gloses sur *Lar* se retrouvent dans les 3 types de sources. Elles peuvent être apparentées, indépendamment de leur type de source et de leur collection d'origine (cf. *D* et *W_a*).

Encadré 1. Gloses sur LAR

Δ	→	Lar (Prisc. 2, 149.11) <i>Ars Bern.</i> (113, 29) ...lar, id est ignis... (115, 22) Item las, quod et lar dicitur, id est ignis, laris facit.
a. Libri Prisciani		
Γ	→	<i>Mil.</i> id est <i>tene</i> (frg. Milan, Bibl. Ambr. cod. A 138 sup.) <i>GK</i> ignis <i>E</i> id est <i>tan</i> /m.a/ ignis ; domus uel turris uel ignis /m.b/ <i>D</i> Lar, quando significat ΚΑΤΟΙΚΙΑΙΟΝ (id est habitaculum deorum), 'laris' facit genitium ; sin autem imperatorem Veientorum 'Lartis', quem mactauit Cossus et testis est Ouidius dicens : «Larte ferox cesus Cossus opima tulit». Lar domus mulierum, quia ignis in ea semper custodiebatur. Nam 'lar' ignis a Greco ΚΑΤΟΙΚΕΟ, id est uro.
b. Glossae collectae		
Υ	→	<i>V</i> domus, uel ignis. <i>FS</i> focus <i>D₁</i> focus et domus <i>N</i> ΛΑΡ : ignis uel domus (...) ΚΑΤΟΙΚΙΑΙΟΝ : habitatorium. ΚΑΤΟΙΚΩ : habito. ΚΑΤΟΙΚΟΣ : habitator. (<i>Miller p. 126</i>) <i>W_β</i> ignis <i>W_α</i> domus mulierum quod in ea custodiebatur ignis. Nam lar 'ignis' ΚΑΤΟΙΚΕΟ (id est uro) dicitur. Lar, quando significat ΚΑΤΟΙΚΙΑΙΟΝ (id est habitaculum deorum), 'laris' facit genitium ; sin autem imperatorem Veientorum 'Lartis', quem mactauit Cossus et testis Ouidius in Epigrammatis sicut dixit : «Larte ferox caesus Cossus opima tulit». Lippius in IIII ab Urbe condita : «Larte columno rege Veientum».

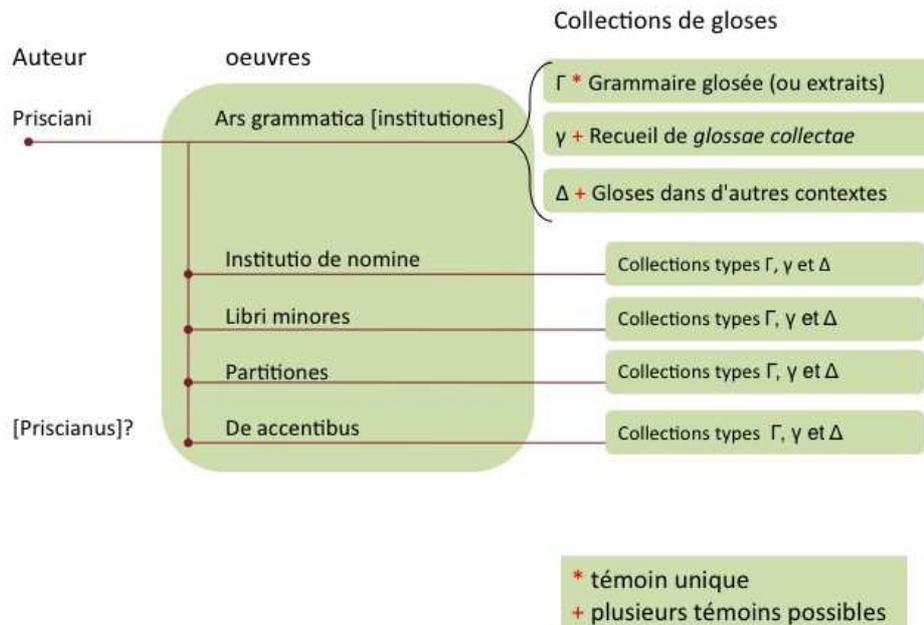
2.3.2. Priscien

L'objectif du corpus de gloses sur ce grammairien est d'offrir un outil d'analyse des données textuelles – et factuelles – de manière à permettre l'étude de la réception du grammairien. Dans une perspective historique, les données rassemblées dans le corpus ne doivent pas être attachées aux textes de Priscien, c'est-à-dire que physiquement les données devraient être autonomes. Affirmation qui semble aller de soi, mais qui posera une cascade de problèmes techniques inhérents à ce choix quand il sera question de mettre en place une interface (moteur de recherche et interface de parcours textuel).

2.3.3. Œuvres

Il est bien entendu, que l'analyse des échanges transversaux entre les collections de gloses au sein des œuvres de Priscien devra tenir compte de tous ses écrits grammaticaux (voir Encadré 2).

Encadré 2. Structure d'ensemble



Or, comme rien n'est simple en ce qui concerne les gloses, il faut définir plusieurs niveaux de structure dans chaque œuvre : une structure physique et une autre logique ou intellectuelle (voir l'exemple du *CGL* dans A. Garcea, F. Cinato et C. Plancq 2010, p. 388) :

Physique : structure de l'édition moderne servant de référence

Intellectuelle : structure définie par l'auteur

Les deux informations étant liées, la problématique est de pouvoir identifier des segments textuels (lemme) de manière automatique et uniforme dans toutes les collections. En d'autre terme, ce genre de données ne devra pas figurer dans le texte (contenu fondamental), mais devraient en constituer la structure (en tant que meta-donnée complémentaire). Par exemple, dans toutes les collections, le terme *Symplegas* portera la référence :

Prisc. *Ars* IV.13 (2, 124, 23)

= le mot se trouve dans la 13e section du livre 4 de l'*Ars* de Priscien, à la ligne 23 de la page 124 du second volume des *GL*.

2.3.4. *Manuscrits*

Les manuscrits sont les supports matériels des données textuelles, ils assument le rôle de témoins. En tant que source primaire, un certain nombre d'informations externes relatives à l'histoire particulière de chaque manuscrit présente une réelle pertinence dans le cadre de l'étude des gloses. À titre d'exemple parmi les informations qu'il est possible de tirer des manuscrits, il est nécessaire de faire la distinction entre les lieux de productions, de résidence et de conservation, déterminante pour suivre le parcours des gloses ; où et quand elles apparaissent pour la première fois, puis dans quelles circonstances elles ont été connues ailleurs. Ces informations seront présentées dans l'en-tête de la collection et pourront ainsi être exploitées ultérieurement.

2.3.5. *Micro-structure : lemme / glose*

Le lemme est le mot (ou les mots) sur lequel repose l'explication transmise par la glose. Ainsi, il est essentiel de délimiter leurs aires respectives à l'aide de balises qui les distingueront au sein des chaînes de caractères⁹. Chacun comportera un numéro d'entrée indiquant sa position dans la collection.

Ex.: n° ... **Simplagas** nymphæ.

Au sein des collections, les lemmes représentent des portions du texte de l'auteur : la lemmatisation ne pose pas de problème dans les collections de type γ , car elle a été effectuée par le créateur de la collection tel qu'il apparaît sur le témoin ; en revanche, elle nécessite une intervention éditoriale dans le cas de collection de type Γ , dans la mesure où elle demande d'interpréter le système mis en œuvre par le glossateur et de déterminer l'étendue du lemme. Intervention capitale, dans certaines conditions, pour restituer l'intelligibilité de l'explication, puisque le lecteur du manuscrit avait accès à l'ensemble du texte.

Notons que le lemme peut comporter plusieurs mots et que sa graphie peut rencontrer des différences avec celles mise en avant dans l'édition imprimée. Ce constat anodin s'avère d'une importance cruciale en termes de recherche d'une *chaîne de caractères*. À chaque graphie particulière devra être associée la graphie retenue dans l'édition de référence. Il va sans dire que l'orthographe des manuscrits devant être respectée scrupuleusement, le moteur de recherche devra apporter en complément de résultats, les formes approchantes. Enfin, le lemme peut correspondre à des paroles rapportées, c'est-à-dire une citation. Le lemme,

⁹ Lemme et glose nécessitant parfois des annotations critiques, l'élément <term> devra encadrer le mot support de l'explication, qui à son tour recevra la balise <gloss>. Ces deux éléments <term> et <gloss> pourront éventuellement comporter des éléments <lem> permettant d'appeler des apparats critiques (<app>). — L'attribut @part permet d'identifier les types de problèmes affectant le groupe, notamment en indiquant au minimum si l'une ou l'autre des deux parties manque (par omission du lemme ou de la glose suite à un problème textuel, etc.).

qui n'émane pas véritablement de Priscien, devra donc bénéficier d'une troisième référence, celle de l'auteur de la citation apportée par Priscien. C'est le cas de l'exemple précédant : le nom *Symplegas* appartient à une citation de Lucain (*Phars.* 2, 718).

NB. Il serait souhaitable d'indiquer aussi la classe grammaticale du lemme, donnée métalinguistique capitale permettant de mettre en corrélation types de glose et de lemmes. Cette typologie devra être calquée sur les distinctions en usage chez les *grammatici Latini* (*partes orationis* et leur sous-groupes ; voir Annexe 1)¹⁰.

2.3.6. Gloses : systèmes de références et informations complémentaires

Les explications sur le texte de Priscien, les gloses, dépendent d'un contexte manuscrit précis, qu'il conviendra impérativement de décrire, sous peine de perte d'informations essentielles. Dès le départ, elles sont distinguées par leur appartenance à l'un ou l'autre type de collection (Γ ou γ). Et si le type γ se contente d'une référence relativement simple (sigle du manuscrit, folio, ligne et n° d'ordre), le type Γ demande une information supplémentaire : la localisation de la glose dans l'aire de la page, en l'occurrence, si la glose est interlinéaire ou marginale et dans ce dernier cas, on devra en préciser l'emplacement.

La glose *nympha* se rencontre sur deux témoins de types différents : Γ , une grammaire glossée (*J* = Cologne, Erzb. Dioz. Bibl., 200) qui entretient des relations avec la collection γ des gloses collectées *V*.

<i>V</i>	27va20b	4.153	Simplagas : <i>nympha</i> .
<i>J</i>	29v6 b	4. ...	Simplagas \ <i>nympha</i> / (m.1)

NB. Les signes typographiques */* qui encadrent la glose indiquent qu'elle se trouve dans l'interligne, au-dessus du lemme.

Deux points réclament encore l'attention : l'imbrication de lemmes dans les gloses et la problématique des signes de renvois. Pour signaler une glose marginale et la lier avec le lemme d'un texte, les glossateurs ont eu recours à deux stratégies : soit ils établissaient le lien entre lemme et glose par un signe graphique faisant office d'appel de notre (signe de renvoi) ; soit ils répétaient le lemme en marge (fréquemment en capitale ou dans un module différent du reste de la glose). Il n'est pas rare non plus de voir le lemme répété à plusieurs reprises, au sein même de la glose : techniquement, cela se traduit par le fait que le balisage des

¹⁰ Par exemple, grâce à un attribut @type, qui se traduit dans le cas de notre exemple (un nom propre : *nomen proprium*, code "NP") en rapport avec la mythologie gréco-romaine (code "NPd"): <term n"... " type="NPd">.

gloses autorise la présence de lemme « secondaire » en son sein¹¹, tandis que la réciproque devra être interdite¹².

La reprise d'un lemme, objet de l'explication donnée par la glose, doit être clairement distinguée de la possibilité pour la glose de citer Priscien ou tout autre auteur. Ce sera à l'éditeur de définir le statut des éléments textuels à encoder et, en vertu de son devoir de lemmatisation évoquée plus haut, il devra délimiter dans la glose ce qui est de l'ordre du lemme, de la paraphrase ou de la citation.

2.3.6.1. Typologie

Le classement du contenu des explications, comme outils d'analyse rend un service immédiat au chercheur. Divers grilles typologiques¹³ peuvent servir de modèle pour l'établissement d'une taxonomie nécessaire au « marquage » des types¹⁴.

Un très bel exemple de mise en œuvre de la taxonomie de Hofman est donné par l'édition électronique de P. Moran. Il a pris en compte les trois volets essentiels : évidemment, les gloses ; mais aussi le texte de Priscien et les manuscrits d'où les gloses sont tirées. Son site permet de naviguer dans ces trois documents en faisant des liens vers le texte du *CGL* et vers le manuscrit numérisé sur le site *e-codices* (ms. de St-Gall).

La glose *nympha* explique par une catégorisation simple un nom propre d'un pseudo-personnage mythologique « Symplegas (est le nom) d'une nymphe »¹⁵. Elle correspond dans la grille typologique retenue au type 54 (glose explicative encyclopédique) du sous-type 41 (catégorisation simple, par un mot (à valeur hyperonymique), portant sur la mythologie), soit le code 54.41¹⁶.

Une même glose peut avoir été composée à partir de plusieurs types de constituants ; pour rendre compte de ce phénomène, elle devra subir une décomposition (voir 2.3.7). En dernier lieu, il faudra appliquer un classement métalinguistique explicitant la langue utilisée des glossateurs, latin ou grec, mais aussi, les vernaculaires, celtique, germanique par exemple, qui se rencontrent dans les gloses carolingiennes sur Priscien.

¹¹ Selon la séquence : <term> ... </term> <gloss> ...<term> ... </term> ... </gloss>.

¹² Séquence interdite <term> ... <gloss> ... </gloss> ... </term> <gloss> ... </gloss> ! La séquence devrait être celle-ci : <term> ... </term> : <gloss> ... </gloss> . <term> ... </term> : <gloss> ... </gloss>.

¹³ WIELAND 1983 ; HOFMAN 1996. Voir Annexe 2.

¹⁴ L'élément <gloss> pourrait ainsi être caractérisée par un attribut @type complété par @subtype.

¹⁵ *Symplegas* est en réalité le nom d'écueils du Pont-Euxin (voir HYG. *Fab.* 19, 4, éd. Marshall, 1993 ; et OV. *Met.* 15, 338). Il s'agit peut-être d'une glose déplacée, car deux vers plus haut en 716 se trouve chez Lucain le mot *Cyane* (Κυάνη), qui est bien le nom d'une Nymphe ; on doit postuler ici des scolies à Lucain comme intermédiaire aux gloses à Priscien.

¹⁶ En TEI : <gloss n="" loc="" type="54" subtype="41">nympha</gloss>.

2.3.6.2. Mains des copistes et des glossateurs

Bien qu'il faille prendre toutes les précautions nécessaires pour analyser ces données, il apparaît essentiel, pour l'établissement de la diachronie au sein des collections, de signaler les changements de scripteurs¹⁷. Il convient d'abord de distinguer mains de copistes et de glossateurs. Par convention, on attribuera un chiffre aux copistes (mains 1, 2, 3, etc.) et une lettre aux annotateurs (mains A, B, C, etc.). En fonction du type de collection, les éléments du groupe lexical (lemme et glose) n'émaneront pas des mêmes personnages. En effet, le scripteur de gloses collectées (collection γ) est considéré comme un copiste pouvant potentiellement avoir été glosé, tandis que dans les collections de gloses *in situ*, (Γ), le lemme uniquement est de la main du copiste (responsable du texte principal), alors que l'explication qui s'y rapporte est l'œuvre du glossateur¹⁸. Il faudra établir une chronologie relative des mains, en prenant garde d'établir une corrélation entre numération des mains et chronologie de leurs interventions, c'est-à-dire non pas en terme quantitatif ou ordre d'apparition, comme il est parfois d'usage quand il s'agit de décrire l'histoire de la copie d'un texte. La succession des copistes (1, 2, 3 etc.) et des glossateurs (A, B, C etc.) signifie que A a écrit avant B et 1 avant 2, etc., et pas que A est responsable de la majorité des gloses ou qu'il s'agit de la première écriture apparaissant sur le manuscrit. Il pourrait résulter des incohérences ou des méprises à attribuer le code 'A' à un glossateur du XIV^e siècle, par exemple, alors que l'écriture du glossateur 'B' serait du IX^e siècle. Sachant que le péri-texte est produit par accumulation, il n'est pas rare de voir un lemme déjà glosé recevoir un complément d'information de la part d'un second glossateur. De nouveau, ce constat implique un fractionnement du balisage des informations.

2.3.7. Sources des gloses

La problématique des sources ne se satisfait pas d'une référence unique pour la simple raison que les gloses procèdent la plupart du temps par accréation et que leur constitution est souvent complexe (voir 2.3.6.1). Les problématiques évoquées au point précédant autour de la question des mains, quand elles sont mises en perspective avec celles des sources, permettent de caractériser deux formes d'accumulation, l'une verticale, l'autre horizontale. En effet, les gloses sont très souvent un agglomérat que le crible de la recherche des sources et des parallèles ne fait que confirmer : pour rendre compte de ces phénomènes, l'éditeur devra retenir comme élément minimal d'analyse, non pas la glose dans

¹⁷ D'une part, la personnalité d'une main n'implique pas que le scripteur d'une explication en soit l'auteur et d'autre part la datation absolue d'une glose au moyen de son support (écriture datable) ne s'applique pas automatiquement à l'explication qu'elle transmet.

¹⁸ À l'exception des cas où le copiste du texte s'avère être aussi le glossateur.

sonensemble, mais des fractions de cette glose. Prenons l'exemple d'un lemme qui apparaît plusieurs fois chez Priscien :

Gl. n° 4.244 (V 28rb19), sur Prisc. au livre IV.23 (*GL* 2, 131, 3)

Cardo : locus est in quo ostium uertitur¹, masculi<ni> generis².

Gl. n° 6.77 (V 33va7), sur Prisc. au livre VI (*GL* 2, 206, 10)

Cardo : ima pars ianuae¹, uel locus in quo ostium uertitur et semper mouetur².

—

Gl. n° 4.244.1 et 6.77.2 viennent toutes deux du *Liber glossarum*, celle sur le livre IV donnant une leçon plus brève ; la source du *Lib. gl.* est Isidore de Séville (*Lib. gl.* CA 671 < Isid. *Et.* 15, 15, 77).

Gl. n° 6.77.1 vient aussi du *Lib. gl.*, mais émane à l'origine d'un autre glossaire ou de gloses collectées de provenance indéterminée (*Lib. gl.* CA 667 : de glossis).

Enfin, la *gl.* n° 4.244.2, typiquement grammaticale, semble tirée du *de dubiis nominibus*, (éd. Glorie, 1968, 768, 191) ... *cardo generis masculini* ..., mais à tout aussi bien pu être déduit du contexte de Priscien, par exemple en *GL* 2, 145, 23 ; bien que Priscien dise ailleurs *hic et haec cardo*¹⁹.

Lorsque les gloses sont considérées par tronçons, la juxtaposition des éléments qui les composent apparaît clairement. Ainsi les sources constitueront un second niveau de référence, appelant un second appareil, distinct de celui comportant les notes critiques. On notera qu'il existe une corrélation entre le « découpage » typologique et la mise en évidence des sources. Dans certains cas, il sera aussi possible de rendre compte des outils de travail en fonction des mains de glossateurs²⁰.

¹⁹ Cf. PRISCIEN V.42 (*GL* 2, 169, 6-9) *Sciendum tamen, quod uetustissimi in multis, ut diximus, supra dictarum terminationum inueniuntur confudisse genera, nulla significationis differentia coacti, sed sola auctoritate, ut hic et haec aspergo, (...) cardo (...).* – VI.15 (*GL* 2, 206, 9-13) *Excipiuntur haec, quae in genetiuo non seruant o : hic cardo cardinis (quidam tamen ueterum etiam feminino genere hoc protulerunt. Gracchus in Atalanta : « o grata cardo, regium egressum indicans » (...).*

²⁰ Les mots-outils (par exemple les connecteurs *uel, siue, idest*, etc.) pourront bénéficier d'un balisage métalinguistique mettant en relief les procédés de composition.

4. Perspectives : questions d'inter-opérabilités

4.1. Lien avec les corpus de texte existants

Après avoir évoqué la nature spécifique des informations liées aux gloses et la manière dont il faudrait les encoder dans la TEI, voyons en quoi leur exploitation peut se révéler pertinente pour l'étude de la réception d'un auteur (Priscien). Un des premiers objectifs à atteindre serait la possibilité de rattacher les données des corpus de gloses, à celles du *CGL*. Le traitement automatique des celles-ci, lors de l'interrogation des textes, permet d'envisager des développements futurs qui offriraient non seulement des lectures du texte avec différents appareils paramétrables, mais aussi d'en suivre la réception à travers les gloses médiévales. Ainsi le lecteur de Priscien du XXI^e siècle pourrait afficher, en parallèle au texte de l'*Ars*, les gloses carolingiennes de tel ou tel témoin manuscrit, les commentaires à la charnière du XI-XII^e siècle, comme les *Glosulae* et les *Notae Dunlmenses*, voire encore les gloses de Guillaume de Conches et celles de Pierre Hélie.

4.2. Autres bases de données

Les textes et les péri-textes constituent des données que l'on qualifie de purement « linguistique », qu'il serait possible d'étoffer avec des informations « para-linguistiques ». Il faudrait imaginer une inter-opérabilité des sources d'informations de manière à croiser des données de natures différentes : pensons seulement aux notices et fac-similés numériques de manuscrits qui se trouvent sur des serveurs différents (*Bnf Archives et Manuscrits*, *e-codices* des bibliothèques Suisse ou *Manuscripta medievalia* pour les bibliothèques allemandes, etc.)²¹. Le site de P. Moran encore une fois est précurseur à cet égard, car il permet de visualiser non seulement les images du manuscrit, mais aussi les pages correspondantes à la recherche dans l'édition de Priscien (Hertz, *GL* 2-3) et dans l'édition princeps (mais partielle) des gloses du *Thesaurus palaeohibernicus*. Un autre volet encore pourrait porter sur la littérature secondaire qui se trouve déjà mise en ligne (cf. *Gallica*, *Persée*, etc.)²². Les références en ligne étant de plus en plus nombreuses, un moteur de recherche idéal pourrait « moissonner » le web en parallèle des bases de données textuelles natives (propre à l'outil) et proposer des

²¹ <http://archivesetmanuscrits.bnf.fr/>
<http://www.cesg.unifr.ch/fr/index.htm>
<http://www.manuscripta-mediaevalia.de/#/4>

²² http://www.bnf.fr/fr/collections_et_services/bibliotheques_numeriques_gallica.html
<http://www.persee.fr/web/guest/home>

résultats, au moyen de filtres, en rapport avec les sources, la bibliographie (cf. par exemple le lien entre *CGL* et *BGL*) et même des dictionnaires (cf. *Perseus*)²³.

5. Intérêt

Pour terminer, car il ne saurait être question de conclure un sujet dont le chantier commence à peine, évoquons, en terme de centres d'intérêts, les retentissements que pourrait avoir un corpus de gloses à Priscien.

Comme outil d'analyse textuelle, le corpus intéresse l'histoire des théories linguistiques au Moyen Âge, mais aussi l'histoire de l'enseignement. Il donnerait accès à un matériel particulièrement difficile à consulter autrement et permettrait d'effectuer des corrélations entre les préoccupations des maîtres et le contexte factuel (historique et matériel) dans lequel elles s'inséraient.

Comme source documentaire, intervenant dans une démarche heuristique, le corpus ouvrirait des perspectives inédites à l'étude de la réception du texte et des transferts de connaissances. Concernant plus particulièrement l'histoire intellectuelle, l'étude des traces scolaires de première main que sont les gloses, proposera des panoramas variés en fonction de tranches chronologiques spécifiques – notamment des relations entre *scriptoria* à l'époque carolingienne et post carolingienne ou des courants doctrinaux dans le contexte des universités – qui se trouveront étayés par un fondement solide constitué de l'ensemble des détails extraits des gloses.

Le but le plus immédiat en vue de l'exploitation scientifique d'un tel corpus, sera de dresser une cartographie globale de l'histoire de la constitution des gloses : diffusion et répartition géographique du fonds commun et mise en évidence de la chronologie des innovations. Quand d'autres corpus pourront être exploités, comme celui des *glossae biblicae*, il sera possible d'analyser les échanges entre champs d'études – dans ce cas précis, arts libéraux et études bibliques – et ainsi d'évaluer l'influence du travail des *grammatici* dans la constitution de la culture intellectuelle médiévale.

Les enjeux autour des données textuelles numériques exploitables scientifiquement par des philologues ne résident plus uniquement dans leur nature électronique, mais d'une manière plus subtile, dans la structure de l'encodage permettant d'attacher aux données brutes des informations supplémentaires dans le but d'offrir différentes possibilités d'exploitation (recoupements automatisés sur plusieurs niveaux typologiques, lectures paramétrables, etc.). Un des défis majeurs que devront relever les philologues et plus généralement les historiens des textes, sera de développer des outils électroniques – sur le principe de moteurs

²³ <http://www.perseus.tufts.edu/hopper/>

de recherche configurables – permettant l’inter-opérabilité de bases de données hétérogènes.

Or, pour que de tels outils ne soient pas de « simples » *google-philology* et ainsi éviter une accumulation linéaire basée sur des critères de pertinence, l’information devra avoir subi une structuration qui offrirait des lectures variés, favorisant les recoupements selon les types d’informations retenus et permettant la corrélation des données textuelles issues des textes, des gloses et des glossaires avec celles apportant des descriptions précises des sources et de leurs contextes. Il s’agit là de contraintes difficiles à surmonter, qui réclament dans certain cas une étape supplémentaire délicate : l’encodage des données.

L’objectif de cette contribution était d’attirer l’attention des philologues sur ce point : l’étape d’encodage n’est pas à considérer comme un aspect mineur, ou du moins simplement « technique » qu’il faudrait régler en fin de parcours, comme on prépare une publication ; tout au contraire, il s’agit d’un enjeu central, au cœur de la réflexion méthodologique, se trouvant donc en amont du projet de recherche visant à récolter et exploiter des données numériques.

SOURCES PRIMAIRES

ANONYMVS BOBIENSIS, *La grammatica dell’A.B., con un’appendice carisiana*, éd. M. De Nonno, Roma, 1982.

CHARISII *Artis Grammaticae Libri V*, éd. K. Barwick, Leipzig, 1964².

DIOMEDIS *Ars*, éd. H. Keil, *Grammatici Latini*, 1, Leipzig, 1857.

DOSITHÉE, *Grammaire latine*, éd. G. Bonnet, Paris, 2005.

Glossae biblicae, éd. P. Vaciago, Turnhout, 2004 (CCCM 189A-B).

VICTORINI *Ars Grammatica*, éd. I. Mariotti, Firenze, 1967.

BIBLIOGRAPHIE

BARWICK K. 1922, *Remmius Palaemon und die römische ars grammatica*, Leipzig.

BONNET G. 2000, « Charisius et Dosithée, Reflets de Cominien », *Revue de Philologie*, p. 7-16.

- EISENHUT H. 2010, *The online edition of the glosses in the Codex Sangallensis 621. New approaches to medieval sources in consequence of electronically available manuscripts*. (http://lib.stanford.edu/files/100114_PPP_Eisenhut%20smaller.pdf).
- GARCEA A. – CINATO F. – PLANCQ C. 2010, « *Corpus Grammaticorum Latinorum*. Un projet de traitement informatique autour des grammairiens latins », *Rivista di Cultura Classica e Medievale* 52/2, p. 377-400.
- GRÉVISSE M. 1993¹³, *Le bon usage*, Bruxelles.
- HADOT P. 1971, *Marius Victorinus, Recherches sur sa vie et ses œuvres*, Paris.
- HOFMAN R. 1996, *The Sankt Gall Priscian Commentary. Part 1. Vol. 1: Introduction; Book 1-5. Vol. 2: Translation and Commentary. Indices*, Munster, 1996.
- LAMBERT P.-Y. 1987, « Les signes de renvois dans le Priscien de Saint-Gall », *Études Celtiques* 24, p. 217-238.
- MAZHUGA V. I. 2001, « Le chapitre sur les verbes défectifs latins dans le manuel de Charisius : sa source principale et sa réexploitation par les auteurs postérieurs » (en russe), *Hyperboreus* 7, p. 283-294.
- 2006, « Les grammairiens latins sur la forme verbale *pepigi* », *Hyperboreus* 12, p. 251-260.
- TURA A. 2005, « Essai sur les marginalia en tant que pratique et documents », in D. Jacquart et Ch. Burnett (éds.), *Scientia in Margine. Etudes sur les marginalia dans les manuscrits scientifiques du Moyen âge à la Renaissance*, Genève, p. 261-387.
- URÍA VARELA J. 2000, « Textual Criticism and Source Study in Ancient Latin Grammar : Charisius *Ars Grammatica* 46.5 Barwick : *Excerpta Bobiensia, Grammatici Latini* 1.547.35 Keil », *Classical Philology* 95, p. 61-71.
- WIELAND G. R. 1983, *The Latin Glosses on Arator and Prudentius in Cambridge University Library MS Gg. 5. 35*, Toronto.

ANNEXE 1

Typologie des lemmes

<i>partes orationis</i>	<i>code</i>		<i>code</i>	
<i>nomen</i>	NC	nom commun concret		
			NCo	objet (manufacturé, bâtiment, etc.)
			NCa	animé (êtres vivants, parties du corps, etc.)
			NCm	minéraux
			NCv	végétaux
			NCp	nom des liens de parenté
	NP	nom propre		
			NPd	divinités, ou touchant à la mythologie
			NPp	personne
			NPg	peuple, etc.
			NPI	lieu (cité, mer, île, montagne, etc.)
			NPf	fleuve
			NPn	phénomènes naturels ; objets célestes, etc. (vents, étoiles, etc.)
	NA	nom abstrait		
			NAi	institution, coutume, etc.
			NAm	poids et mesures
			NAt	technique (grammaire, etc.)
<i>pronomen</i>	PN	pronom		
<i>uerbum</i>	V	verbe		
<i>aduerbium</i>	AV	adverbe		
<i>participium</i>	A	adjectif et participe		
<i>coniunctio</i>	C	conjonction		
<i>praepositio</i>	P	préposition		
<i>interiectio</i>	I	interjection		
	S	syntagme		
			SN	synt. nominal
			SP	synt. prépositionnel
			SV	synt. verbal
	Pro	proposition complète		

ANNEXE 2

Typologie augmentée

Les notes (ou astérisques) signalent les modifications apportées à la formulation de R. Hofman (1996, p. 83-95). Dans cette typologie, les types 1 à 7 désignent les familles, tandis que les sous-groupes (11, 12, etc.) décrivent plus précisément la nature des explications.

- 1 Prosodique
 - 11 Macron indiquant une syllabe longue²⁴
 - 12 Commentaire sur la quantité des syllabes
 - 13 Commentaire prosodique élaboré
- 2 Lexicale
 - 211 Traduction vernaculaire*
 - 2111 – d'une glose
 - 2112 – d'un lemme
 - 212 Synonyme
 - 213 Négation d'un antonyme
 - 214 Hyponyme expliqué par un hyperonyme
 - 215 Définition littérale²⁵
 - 221 Traduction du grec au latin (ou l'inverse)
 - 222 Changement de préfixe
 - 223 Adjectif glosé par un nom décliné*
 - 23 Différences (*differentiae*)
 - 24 Dérivation lexicale*
 - 25 Décomposition d'un mot*
- 3 Morphologique
 - 31 Grammaire du *nomen*
 - 311 Morphologie (déclinaison)
 - 3111 Nominatif
 - 3112 Vocatif
 - 3113 Génitif
 - 3114 Datif
 - 3115 Accusatif
 - 3116 Ablatif
 - 312 Nombre
 - 313 Genre et autres points morphologiques
 - 32 Grammaire du *pronom*
 - 321 Spécifie le pronom (en général)
 - 322 Pronom relatif
 - 3221 Antécédent d'un pronom relatif (spécifique)
 - 3222 Antécédent d'un pronom relatif repris d'une glose précédente
 - 3223 Interprétation ou équivalent lexicale de l'antécédent d'un pronom relatif

²⁴ HOFMAN (1996, p. 84) : « *length mark indicating syllable length in poetic quotations* ».

²⁵ HOFMAN (1996, p. 85) : « *definition given by an entire sentence* ».

- 3224 Pronom relatif sans antécédent identifiable
- 323 Autres pronoms
- 3231 Pronom démonstratif pour un autre
- 3232 Explicite le contexte d'un pronom
- 3233 Substitution du pronom avec changement de cas
- 33 Grammaire du verbe et participe
- 331 Morphologie verbale (générale)
- 332 Morphologie du *tempus*
- 333 Morphologie du *modus*
- 334 Morphologie du participe et du gérondif
- 34 Indication d'adverbe
- 35 Explicite une conjonction
- 36 Explicite une préposition
- 37 Explicite une interjection
- 4 Syntaxique
 - 41 Signe de construction syntaxique
 - 4111 Corrélation entre adjectif/préposition et substantif
 - 4112 Corrélation du complément du nom (?)
 - 4121 Corrélation entre un pronom relatif et son antécédent
 - 4122 Corrélation entre un autre pronom et un nom
 - 4123 Corrélation entre un verbe et son sujet
 - 413 Rétablissement d'un ordre de lecture
 - 4141 Corrélation des éléments de structure d'une proposition
 - 4142 Corrélation entre deux propositions
 - 4143 Mot supplétif à un oubli de l'auteur
 - 4144 Corrélation entre l'explication et citations
 - 4145 Autres mises en corrélation
 - 42 Mot(s) explicitant la construction
 - 421 Identifie le locuteur ou une citation
 - 422 Supplétives
 - 4221 Suppléée par un nom ou pronom au nominatif
 - 4222 – par un nom ou pronom au génitif
 - 4223 – par un nom ou pronom au datif
 - 4224 – par un nom ou pronom à l'accusatif
 - 4225 – par un nom ou pronom au vocatif (?)
 - 4226 – par un adverbe
 - 4227 – par un adjectif
 - 4228 – par un verbe ou une courte proposition
 - 4229 – par une conjonction ou une préposition
 - 423 Modifie un adjectif (?)
 - 424 Ajout (adit-glosses)
- 5 Explicative
 - 51 Figures de styles
 - 521 Commentaire exégétique (= 541 Hofman)
 - 5211 Paraphrase ou glose tirée de Priscien
 - 5212 Indication (spécification) portée sur le titre d'une oeuvre
 - 522 « *Quia*-glosses » = commentaire explicatif introduit par *quia* (= 542 Hofman)
 - 523 Éclaire les ambiguïtés du texte (= 543 Hofman)
 - 53 Étymologique (= 55 Hofman)
 - 54 Encyclopédique (= 56 Hofman)
 - 541 Géographique (= 561 Hofman)
 - 542 Objet inaccoutumé (= 562 Hofman)
 - 543 Nom propre (= 563 Hofman)
 - 544 Mythologique ou littérature latine (= 572 Hofman)
 - 551 Confrontation entre diverses sources (= 571 Hofman)

- 6 Ecdotique
- 611 Atéthèse médiévale
 - 612 Correction critique du texte ou Correction rédactionnelle
 - 62 Collation et variantes (= 58 Hofman)
 - 631 Titulation marginale (postérieure ou non à la copie) ; Mise en « exergue » intertitre, sous-titre, mot, etc.
 - 632 Résumé du contenu (= 52 Hofman)
 - 64 Références croisées (ou interne)²⁶ (= 53 Hofman)
 - 641 Mot ou signe de renvoi (à une explication)
 - 6411 Glose '*similiter*'
 - 6421 Signe graphique (de renvoi)
 - 6422 Renvoi verbal (explicité en toute lettre)
 - 65 Source (=57 Hofman)
- 7 Informations diverses sur le contexte socio-historique (= 59 Hofman)
- 71 Colophons
 - 72 Suscriptions et souscriptions
 - 73 Doxologique (formule de prière)
 - 731 Invocations
 - 732 Imprécation (parfois se confond avec l'ex-libris)
 - 74 Pétitions de suffrages
 - 75 Autres mentions (jussion, oblation collation, authentification, commémorative, etc.)
 - 76 Autres observations (quotidienne, etc.)

© Eruditio Antiqua 2011

ISSN 2105-0791

www.eruditio-antiqua.mom.fr

eruditio-antiqua@mom.fr

Image : © Kunsthistorisches Museum, Vienna

²⁶ Cf. LAMBERT 1987.