

# Le CD-Rom : un nouvel âge pour la recherche? : Étude d'"ADMYTE1", base de textes espagnols médiévaux

Autor(en): **Pellen, René**

Objektyp: **Article**

Zeitschrift: **Revue de linguistique romane**

Band (Jahr): **61 (1997)**

Heft 241-242

PDF erstellt am: **10.08.2024**

Persistenter Link: <https://doi.org/10.5169/seals-399967>

## **Nutzungsbedingungen**

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

## **Haftungsausschluss**

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

LE CD-ROM:  
UN NOUVEL ÂGE POUR LA RECHERCHE?  
ÉTUDE D'«ADMYTE1»,  
BASE DE TEXTES ESPAGNOLS MÉDIÉVAUX

1. *Présentation du disque*

ADMYTE (*Archivo Digital de Manuscritos y Textos Españoles*) est une série ouverte de disques compacts sur l'espagnol médiéval; quatre ont été réalisés jusqu'ici: par ordre chronologique, Admyte1, Admyte0 (le premier programmé), Admyte3, Admyte4<sup>(1)</sup>. L'entreprise, conçue au départ comme une contribution à la commémoration du V<sup>e</sup> Centenaire de la découverte de l'Amérique, a reposé sur l'effort conjoint de la Sociedad Estatal Quinto Centenario, de l'éditeur Micronet, de la Bibliothèque Nationale de Madrid et d'un groupe d'universités espagnoles et nord-américaines (l'*Autónoma* et la *Complutense* de Madrid, Berkeley, Wisconsin-Madison, Toronto). Une abondante littérature ayant déjà présenté le projet dans son ensemble, on se contentera ici d'y renvoyer le lecteur (cf. Bibliographie<sup>(2)</sup>), en s'arrêtant sur les caractéristiques du disque étudié.

Admyte1, dont le contenu détaillé figure dans Marcos Marín 1994 (pp. 226-229), Faulhaber 1995 (pp. 107-110) et dans le Manuel d'utilisation qui accompagne le disque (pp. A1-A15), comprend 61 textes appartenant à des domaines très divers: encyclopédies (Alfonso de la Torre, *Visión delectable*; Bartholomæus Glanville, *De proprietatibus rerum / Propiedades de las cosas* [trad. Fr. Vicente de Burgos]), dictionnaires et grammaires (A. de Palencia, *Universal vocabulario*; Nebrija, *Dictionarium hispano-latinum, Dictionarium latino-hispanicum, Gramática castellana, Introductiones latinæ*; R. Fernández de Santaella, *Vocabulario eclesiástico*), textes juridiques (Alfonso X, *Siete*

---

(1) D'après une information fournie par C. Faulhaber aux membres du groupe de discussion électronique MEDIBER le 26 juin 1996. L'étiquetage des deux derniers sera peut-être modifié. Leur prix n'est pas encore fixé. Pour le contenu, Admyte3 et Admyte4 reproduisent partiellement les deux précédents.

(2) Notamment FAULHABER 1990, 1992, 1995; MARCOS MARÍN 1992 et 1994; Micronet [1992].

*Partidas*; A. Díaz de Montalvo, *Ordenanzas reales*; Ferdinand V et Isabel I, *Capítulos de gobernadores*, etc.), textes scientifiques (Al-Mubasshir ibn Fatik, *Mukhtar al-hikam / Bocados de oro* [trad. anon.]; Nebrija, *Tabla de la diversidad de los días y horas*; B. de Gordonio, *Lilium medicinæ, De ingeniis curandorum morborum / Tratado de las tablas de los diez ingenios...*, *De regimine acutorum morborum / Regimiento de las agudas, Tratado de los niños, Prognostica / Las Pronósticas*; F. Álvarez, *Regimiento contra la peste*; F. López de Villalobos, *Sumario de la medicina*; J. de Ketham, *Fasciculus medicinæ / Compendio de la salud humana*; V. de Taranta, *De epidemia et peste / Tratado de la peste*; Anónimo, *Tratado de la fisonomía*; J. Gutiérrez de Toledo, *Cura de la piedra*; M. Díez de Calatayud, *Llibre del arte de menescalía*), livres de voyages (J. de Mandevilla, *Libro de las maravillas del mundo*; M. Martínez de Ampiés, *Tratado de Roma*; B. von Breydenbach, *Peregrinatio in Terram Sanctam / Viaje siquier peregrinación de la tierra santa* [trad. de M. Martínez de Ampiés]; Marco Polo, *Libro de Marco Polo* [trad. de R. Fernández de Santaella]), chroniques et biographies (D. Rodríguez de Almela, *Valerio de las historias eclesiásticas, Compilación de las batallas campales, Letra sobre los matrimonios, Escritura de cómo y por qué razón no se debe [sic] dividir [...] los reinos y señoríos de España*; D. de Valera, *Crónica de España*; G. Fabricio de Vagad, *Crónica de Aragón*; H. del Pulgar, *Claros varones de Castilla, Letras; Crónica popular del Cid*), traités de chevalerie ou de noblesse (A. de Cartagena, *Doctrinal de los caballeros*; D. de Valera, *Tratado de las armas, Ceremonial de príncipes*; F. Mejía, *Nobiliario vero*), textes poétiques (Í. López de Mendoza, *Bías*; J. del Encina, *Coplas del memento homo, Égloga de Fileno, Testamento de amores*, etc.; A. Montesino, *Coplas hechas sobre la Pasión*; etc.).

Ces textes ont été réunis parce qu'ils ont été imprimés à la fin du XV<sup>e</sup> s. ou au début du XVI<sup>e</sup> (1471-1524). Mais leur rédaction s'échelonne du XIII<sup>e</sup> au XVI<sup>e</sup> et l'on peut se demander si le critère de la date d'impression est ici pertinent en matière de choix, car en dehors d'études sur les incunables les chercheurs ne peuvent guère utiliser ce type de datation. En outre, aucun souci d'équilibre n'a présidé à la représentation des quatre siècles auxquels ont été empruntés les textes: le disque contient 2 textes du XIII<sup>e</sup> (*Bocados de oro* et *Siete Partidas*), 0 du XIV<sup>e</sup>, entre 43 et 50 du XV<sup>e</sup> [surtout de la seconde moitié du siècle], 9 ou 10 du XVI<sup>e</sup>. Il ne s'agit donc pas vraiment d'un corpus au sens littéraire ou linguistique du terme, mais d'un ensemble de textes, de longueur d'ailleurs très inégale (de quelques folios à plusieurs centaines). Les auteurs le présentent comme une «bibliothèque»: «La intención ha sido la de presentar en este disco inicial lo que podría haber sido una biblioteca de un contemporáneo de los navegantes que partieron de Palos rumbo a las Indias [...]. Así se rinde homenaje a los que llevaron la Ley, la imprenta, la universidad y, en

suma, la Latinidad, a las tierras por ellos descubiertas» (Manuel, pp. 6-7). Admyte1, par conséquent, s'adresse en priorité aux chercheurs qui s'intéressent au XV<sup>e</sup> s. et aux débuts de la Renaissance en Espagne; aux médiévistes proprement dits il propose cependant un accès dynamique à deux textes très importants, qu'on ne pouvait explorer jusqu'ici qu'à travers les microfiches de Madison<sup>(3)</sup>.

Une précision doit être ajoutée sur le problème de la datation. Ni dans le disque ni dans les publications le concernant il n'existe, sauf erreur, une liste chronologique des textes. Les seules informations qui soient fournies sur cette question figurent dans les notices qui accompagnent les textes. Pour quelques-uns d'entre eux, notamment les traductions de B. de Gordonio et J. de Mandevilla, la date exacte n'est pas connue. De toute façon, la date qui est indiquée après le titre de l'œuvre est la date d'édition. Il paraîtrait souhaitable que la prochaine édition du disque fournisse d'abord la date de composition – sans qu'il faille la chercher dans le corps de la notice ou après la biographie de l'écrivain – et ensuite seulement la date de l'édition utilisée pour la mémorisation. Il conviendrait aussi de compléter l'information disponible par quelques index comme l'index chronologique récapitulatif, l'index des auteurs, l'index des lieux d'édition.

Tel qu'il est, Admyte1 se présente comme une double base de données (textes et images), exploitable sur tout micro-ordinateur de type PC doté d'un lecteur de CD-ROM et d'un écran VGA; elle fonctionne sous Windows 3.1. Une version Macintosh était initialement prévue, mais le projet a été abandonné. L'exemplaire utilisé comporte un disque (une option permet de choisir la langue de travail dans un large éventail d'idiomes) et un manuel. Le manuel, qui est essentiellement un guide d'utilisation (54 p.), comprend aussi trois appendices: liste des titres (pp. A1-A15), normes de transcription (pp. B1-B7), [description du] glossaire des lemmes et des formes (C1-C5). P. 7 sont cités les responsables scientifiques d'ADMYTE: sous la direction de Francisco Marcos Marín, G. Meiro Martínez, C. B. Faulhaber, Á. Gómez Moreno, A. Martín Santa Olalla, J. Martín Abad, J. J. Nitti, C. Foz Baldiz, F. Calero Sánchez, E. de Lema Garzón, L. M. Martínez Guñales.

## 2. *L'outil de recherche*

L'une des principales innovations d'Admyte1 est à coup sûr l'association de textes et d'images. Les textes ont été saisis (digitalisés) en fonction des

---

(3) *Concordances and Texts of the Royal Scriptorium Manuscripts of Alfonso X*, 1978, 42 microfiches.

normes adoptées par le Séminaire de Madison<sup>(4)</sup>, ce qui permet leur analyse automatique, mais les incunables d'où ils proviennent ont eux-mêmes été mémorisés sous la forme d'images, que l'on peut consulter parallèlement au texte, en affichant le texte et son image dans deux fenêtres juxtaposées. Le disque contient plus de 8 000 pages, correspondant à environ 40 000 000 caractères. Selon les besoins, l'image ou un détail de l'image peut être grossi(e), facilité qui met à tout moment à la disposition du chercheur la source documentaire sous l'angle le plus intéressant, puisqu'il est librement choisi et peut être indéfiniment modifié<sup>(5)</sup>. Lorsqu'il ouvre Admyte1 l'utilisateur a très vite sous les yeux le texte ou l'image de telle ou telle œuvre, les deux ensemble s'il le désire. Après avoir sélectionné son titre dans le sous-menu «Sélection de titre»<sup>(6)</sup>, il lui suffit de recourir à la «Palette d'options» et de cliquer sur les icônes «Transcription» et «Image», puis de faire les ajustements nécessaires pour que les deux documents tiennent dans l'espace de la fenêtre. La page affichée est toujours la première<sup>(7)</sup>. Au moyen de la palette ou du menu Options on peut passer immédiatement à la seconde, puis à la troisième et revenir en arrière sans aucun problème. En revanche, le logiciel d'exploitation ne permet pas de feuilleter le texte, par exemple en sautant  $n$  pages, ni de se positionner à la fin du texte. C'est que sa finalité principale est de repérer les mots ou groupes de mots dans leur contexte: il n'a pas été prévu qu'on puisse consulter les textes sans passer par une interrogation sur les mots.

On regrette souvent ces limites que les concepteurs imposent à l'utilisateur et il serait souhaitable que la prochaine édition du disque étende les possibilités d'exploitation des bases en ajoutant des programmes au logiciel actuel. L'utilisateur a cependant la faculté de télécharger un texte dans un fichier et retrouve alors la liberté que lui donne son traitement de texte ordinaire. Cette générosité tout à fait inhabituelle chez les concepteurs et fabricants de CD-ROM s'explique par le fait qu'ADMYTE est un outil de recherche réalisé pour les chercheurs<sup>(8)</sup>. Elle mérite d'être signalée. Chacun

---

(4) Cf. MACKENZIE, J. G., BURRUS, V., *A Manual of manuscript transcription...*, 4<sup>e</sup> éd., 1992.

(5) Le grossissement peut doubler la taille de l'image; par le procédé inverse de réduction l'image peut être ramenée à un dixième de sa taille.

(6) Les titres sont classés par ordre alphabétique; les noms d'auteur n'apparaissent pas sur cette liste. On ne peut donc pas sélectionner d'un coup toutes les œuvres de l'auteur X, par ex. Diego de Valera, présentes dans la base, comme on peut le faire par «thèmes» (types de texte): encyclopédies, dictionnaires...

(7) Si aucun titre n'a été sélectionné, la page qui s'affiche est la première page du premier texte.

(8) Il va sans dire que la copie que peut effectuer un chercheur est strictement personnelle.

peut ainsi consulter le texte comme il l'entend, voire le soumettre à des programmes lui appartenant.

Sur le disque, l'unité d'affichage et de lecture est la «page», qui correspond à un folio (recto ou verso). Il est possible d'éditer sur imprimante une page quelconque, des pages qui se suivent ou des pages isolées. On peut ainsi se constituer un dossier sur le phénomène ou l'ensemble de phénomènes qu'on étudie. Dans les notices de la Présentation la longueur du texte est indiquée en folios, non en pages; quand on affiche le texte, la référence de folio apparaît en haut de chaque page (elle fait partie du texte) et la référence de page dans la barre de fenêtre, à la droite du titre; il peut y avoir un grand écart entre les deux références, un texte ne commençant pas toujours au premier folio dans un manuscrit (ex. pour les *Letras* de Fernando del Pulgar, qui comprend 91 folios, la page 1 correspond au f° 43v). La mention du nombre de pages dans la Présentation, juste après le numéro du titre (ligne 1), apporterait donc une information utile, de même que l'indication du nombre de caractères occupés par le texte et le nombre de formes occurrentes.

### 2.1. Recherche de mots

Le guide d'utilisation passe directement de la Sélection des titres à la Recherche de mots. La préoccupation des concepteurs est sans ambiguïté. Mais il convient de préciser que cette recherche doit être entendue au sens large et qu'elle pourra porter, au-delà des mots, sur des combinaisons, des séquences, des modèles, grâce à l'emploi de procédures logiques qui seront analysées plus loin. La recherche en théorie la plus simple est la recherche d'un mot (une forme) dans un texte ou dans toute la base. A cet effet, lorsque dans le menu Options il choisit le sous-menu Recherche par mots ou – ce qui revient au même – lorsqu'il clique sur l'icône de la Palette 'recherche de mots', l'utilisateur dispose de plusieurs procédures. Sur la grille d'interrogation qui apparaît à l'écran, le curseur est positionné au début de la case rectangulaire intitulée Recherche (boîte à questions). Il peut alors taper le mot recherché et lancer la recherche avec un retour-chariot (ou en cliquant sur le bouton Chercher figurant à droite, en grisé)<sup>(9)</sup>: le résultat est très rapidement affiché au-dessus de Chercher sous la forme du nombre de pages où se rencontre le mot. Pour visualiser ces pages, il suffit de cliquer sur le deuxième bouton en grisé, Aller à, ou sur l'icône correspondante de la Palette. Le début de la page où figure la première occurrence apparaît immédiatement sur l'écran.

---

(9) Dans le grand rectangle Recherche il s'agit d'indiquer le mot sur lequel doit porter la recherche; le bouton Chercher sert à lancer la recherche.

Mais ici surgissent plusieurs anomalies et, selon le point de vue adopté, plusieurs défauts graves du système adopté. En premier lieu, si le mot ne se trouve pas dans le tiers ou le quart supérieur de la page, il n'est pas affiché dans la fenêtre du texte: il faut le rechercher plus loin en faisant défiler la page. Cette convention de l'affichage par page (sans autre option) ralentit considérablement la consultation. En revanche, le mot, présenté sur fond rouge, se détache parfaitement du texte et se repère sans aucune difficulté.

Autre choix plus déroutant: un décompte fréquentiel reposant sur le nombre de pages où un mot se rencontre. Un mot se répète souvent plusieurs fois dans une même page. Le nombre des «réponses» indiqué à la suite d'une recherche n'a donc aucune signification statistique ni linguistique. Or il ne semble pas que le logiciel d'interrogation permette d'obtenir un renseignement fréquentiel quelconque en dehors du nombre de pages. Cette dernière valeur ne donne au mieux, par conséquent, qu'une idée approximative de l'effectif minimal. Mais en cas d'homonymie elle perd toute signification puisque dans les textes les homonymes ne sont pas distingués. Ainsi, une recherche sur «casa» dans les *Letras* de F. del Pulgar a comme réponse: 12 (pages). Vérification faite, le nombre des occurrences réelles est de 19; mais au f° 84r (p. 82) le mot est le présent du verbe *casar(se)*, ce qui réduit les occurrences du nom (au sing.) à 18. L'absence de désambiguïsation (par un procédé ou par un autre) et l'absence d'indexation grammaticale rendent très difficile la constitution de données, au sens strict du terme: tous les renseignements doivent être vérifiés dans les textes; certes, et ce n'est pas peu, le fil d'Ariane est fourni au chercheur pour qu'il réunisse (s'il a beaucoup de temps) les informations dont il a besoin, mais il lui incombe d'extraire des textes, à l'aide de ce fil conducteur, le matériau qui pour lui ne sera aussi bien que le point de départ de son analyse ou de sa réflexion. Ces limites actuelles du système d'interrogation ne sont que légèrement compensées par la lemmatisation partielle qui sera examinée un peu plus loin.

Il existe une seconde possibilité pour chercher les références d'une forme. Comme les graphies et la morphologie médiévales sont variables, on peut taper la forme à laquelle on pense dans la seconde case rectangulaire de la fenêtre Recherche par mots et provoquer l'affichage, dans la demi-fenêtre inférieure (qui affiche 12 mots), d'un fragment du Glossaire dont le premier mot est la forme tapée ou, si elle n'est pas attestée dans la base, la forme alphabétiquement la plus proche d'elle. Ex. (graphie): on s'interroge sur le vocable *peaje*, qui peut s'écrire avec *g* ou avec *j*; on tape «peage» (ou simplement «pea») et l'on voit s'afficher les 4 formes présentes dans la base: «peage», «peages», «peaje», «peajes»; si l'on sélectionne le texte *Bías contra Fortuna*, de Santillana (c.1448), on vérifie en cliquant successivement sur chacune de ces

formes dans le Glossaire qu'aucune d'elles ne se rencontre dans l'œuvre. Ex. (morphologie): on hésite entre plusieurs variantes pour le substantif dérivé de *viejo*; si l'on tape «veje», au-dessous de «vejectales» dans le Glossaire on lit «vejedad» et «vejes»; «vejedad» revient 6 fois dans la base (2 dans *Bías*), «vejes» 5 fois dans la base (0 dans *Bías*); les deux variantes apparaissent dans les *Siete Partidas* (XIII<sup>e</sup> s.)<sup>(10)</sup>. Toutes ces opérations sont instantanées et donnent donc un accès immédiat aux contextes.

Si la question est plus complexe, par exemple si l'on veut recueillir toutes les formes fléchies d'un verbe, on dispose d'une troisième procédure, qui s'appuie sur l'inventaire des lemmes et des formes. Au lieu de choisir l'option Recherche par mots, on choisit alors l'option Glossaire des Lemmes. La fenêtre qui s'ouvre est divisée en deux grands cadres, dont le premier est surmonté d'une case (Mots), le second d'un titre (Formes des lemmes). Dans la case on tape une forme quelconque du vocable visé, par ex. «robaron»: dans le premier cadre s'affiche alors la partie du Glossaire qui commence à «robaron»; quand on clique sur cette forme dans le cadre, le vocable *robar* s'affiche dans le deuxième cadre, suivi de toutes ses formes fléchies: *robar*, «roba», «robada», «robadas», «robado», etc. On peut, en sélectionnant le vocable ou l'une de ses formes, puis en cliquant sur le bouton Chercher lemme (petite case à fond gris à droite en haut de la fenêtre), savoir si le vocable est peu ou abondamment représenté dans la base, ou dans le texte choisi. Le résultat est, ici encore, presque instantané, ce qui indique que l'algorithme est excellent. Rien n'empêche, évidemment, de limiter la recherche à celles des formes qui coïncident avec le besoin du moment, par ex. aux formes du prétérit de *robar*, dont trois seulement sont attestées dans la base: «robe», «robo», «robaron».

Quand on prend connaissance des contextes, on s'aperçoit cependant qu'on ne peut pas se fier au Glossaire des Lemmes, pour la raison déjà évoquée plus haut: les homographes n'ont pas été discriminés, si bien que «robe» (2 occ. dans la base), prétérit dans le *Doctrinal de los caballeros* d'A. de Cartagena (p. 8, f<sup>o</sup> 4v), est un subjonctif présent dans les *Letras* d'H. del Pulgar (p. 80, f<sup>o</sup> 83r), que «robo» (108 occ.), est le plus souvent un nom. On se rend compte qu'une lemmatisation partielle comme celle qui a été retenue par les auteurs du disque n'offre aucune garantie aux utilisateurs. En fait, la lemmatisation qui a été réalisée (et pour certaines formes seulement – le Manuel assure qu'elle touche la plus grande partie du vocabulaire –) est une lemmatisation abstraite, presque théorique, en tout cas arbitraire: «robo» n'a été lem-

(10) Ce qui montre en passant que la présentation du *DCECH* est inexacte: «*Vejez* [¿S. XIV?, Biblia med. rom., 21.2; Nebr.]; antes se dijo *vejedad* [h. 1250, *Setenario*, f<sup>o</sup> 8v]»...



matisée que sous *robar*; mais il se trouve que le nom, beaucoup plus fréquent que le verbe, n'a pas été, quant à lui, lemmatisé, ce qui fait que cliquer sur «*robos*», dans le Glossaire des lemmes, provoque une recherche sur toutes les formes de *robar*, auxquelles «*robos*» n'appartient même pas<sup>(11)</sup>. On pourrait multiplier les exemples: «*estado*», «*estados*» ont été lemmatisés sous *estar*; mais dans les *Letras* d'H. del Pulgar, les 4 «*estado*» et les 3 «*estados*» sont des noms... En revanche, aucun lemme n'a été introduit pour le nom. De même, «*esta*», «*estas*», «*este*» ont été lemmatisés exclusivement sous *estar*; les occurrences du démonstratif sont donc mêlées à celles du verbe.

Contrairement à ce que pensent les auteurs, une lemmatisation ne saurait être que complète; elle implique une analyse intégrale des contextes, une ventilation des formes aussi exacte que possible et le rejet de l'implicite. Par exemple, quand un texte comme les *Letras* d'H. del Pulgar ne renferme aucun participe passé «*estado(s)*», il est erroné de lemmatiser les deux formes sous *estar* et de laisser de côté le nom, qui seul est représenté dans le texte. Si les contraintes de temps empêchaient de procéder à une véritable lemmatisation, une solution moins mauvaise aurait pu consister à introduire, pour chaque forme, des renvois à tous les lemmes possibles; ainsi, pour «*esta*», la double possibilité *estar* / *este* aurait été signalée, à charge pour l'utilisateur de vérifier dans chaque texte la nature des occurrences concrètes. Cet exemple est particulièrement simple; mais dans des cas plus difficiles la lemmatisation partielle peut entraîner des oublis ou de graves méprises (cf. ci-dessous). De toute façon, dès que la forme atteint une certaine fréquence il devient très fastidieux, voire impossible, de vérifier dans leurs contextes toutes ses occurrences. Ce qui rend indispensable une lemmatisation texte par texte accompagnée d'index exhaustifs qui baliseraient la recherche des formes désirées: toutes les occurrences de «*robo*» N doivent être retrouvées, si c'est «*robo*» N qui fait l'objet de la recherche – et seulement les formes du nom.

On souhaiterait, de même, qu'un sous-programme de concordances soit ajouté au logiciel d'exploitation de la base. Ne serait-ce que pour obtenir des index contextuels de type KWIC et faire l'économie d'une exploration page par page, qui exige un temps de consultation beaucoup trop long, souvent disproportionné par rapport à la moisson finale de données utiles. Ces concordances fourniraient les contextes correspondant exactement aux index fréquents mentionnés plus haut. En l'état actuel de l'informatique il n'est pas

(11) Une petite erreur de programmation sans doute fait que lorsqu'on veut afficher les 4 occurrences de *robar* dans *Bías*, la première, «*robo*» N (p. 9, f° 5r), et la seconde, «*robaua*» (p. 22, f° 11v), n'apparaissent pas sur fond rouge, ce qui oblige à lire tout le texte pour les découvrir.

excessif d'envisager la possibilité d'un passage constant entre un système traditionnel de concordances unilinéaires et un affichage du texte intégral, d'autant qu'avec un écran suffisamment spacieux concordances et texte pourraient apparaître en même temps dans deux fenêtres différentes, la fenêtre texte affichant par exemple le contexte large de la ligne de concordance sélectionnée par le curseur.

## 2.2. *Autres recherches*

Comme dans les traitements ou éditeurs de texte courants il est possible, pour lancer une recherche, de réduire un mot à une séquence de caractères limitée: cf. l'exemple déjà cité de «veje» pour rechercher tous les dérivés nominaux possibles de *viejo*. Mais le logiciel permet à la fois de délimiter plus strictement et d'étendre la recherche à un modèle qui peut coïncider ou non avec un mot. Deux procédés ont été retenus pour y parvenir, qu'expriment les symboles [?] et [\*]. Le premier de ces symboles représente une chaîne de 0 ou 1 caractère (quelconque) s'ajoutant à la séquence alphanumérique tapée; ex. «amarill?» si l'on recherche le masculin et le féminin de l'adjectif. Le second, [\*], représente une chaîne de 0 ou *n* caractères; ex. «reun\*» si l'on cherche tous les mots de la famille de *reunir* et toutes leurs formes fléchies.

En réalité, avec la version des programmes figurant dans le disque utilisé il a été impossible d'obtenir le moindre résultat. Si l'on demande «buen?» sur toute la base, on obtient comme réponse «0», alors que l'interrogation sur «buena» donne «1273 pages» et l'interrogation sur «bueno» «736». La réponse est également négative si l'on fait suivre [?] d'une autre séquence, ex. «buen? era» (alors que «bueno era» figure 6 fois dans *Visión delectable* – la recherche «bueno era» sur ce texte signale la présence de l'expression dans 5 pages –; mais la même recherche sur l'ensemble de la base donne «0»<sup>(12)</sup>. [\*] ne fonctionne pas davantage: si l'on tape «reun\*», la réponse est «0»; pourtant «reunir» existe bien p. 128, f° 64v du *Doctrinal de los caballeros*; si l'on interroge ce texte, la réponse est la même; si l'on fait suivre le radical du verbe du mot qui le suit dans le texte («reun\* en»), le résultat ne change pas. Il arrive même qu'une question ne déclenche aucune recherche. C'est ce qui se produit quand on teste la question «\*metro?» citée dans le Manuel, p. 36, et qui devrait fournir comme réponse le nombre de pages où se rencontrent des mots comme

(12) Le mauvais fonctionnement de ces sous-programmes pourrait être imputable, selon F. MARCOS MARÍN (correspondance privée), à une mauvaise version du logiciel. Il est regrettable, malgré tout, qu'un CD-ROM soit commercialisé avec un logiciel qui n'est que partiellement opérationnel.

«barómetro(s)», «termómetro(s), «pluviómetro(s)» (à noter que la base ne contient aucun de ces mots...).

Comme on vient de le voir, on peut aussi, en théorie, associer dans une question deux mots, qui peuvent être deux modèles, si l'on se sert des deux symboles qui viennent d'être mentionnés. Deux mots: «auian entrado»<sup>(13)</sup>: dans ce cas la réponse est quasi instantanée; cette séquence se rencontre deux fois dans la base, comme le prouve l'affichage des contextes<sup>(14)</sup> (pourquoi cette recherche réussit-elle alors que la recherche sur «bueno era» a échoué?). Deux modèles: «auia\* \*ado»; appliquée au texte *Claros varones de Castilla*, cette recherche, au bout de 8 minutes, n'a pas abouti; or le texte ne contient que 3 occurrences de ce modèle («auian tomado» p. 30, f° 15v, «auian estado» p. 49, f° 25r, «auian resultado» p. 67, f° 34r), que l'on a plus vite fait de recueillir en examinant les 6 occurrences de «auian». Il semble que l'algorithme du programme qui manipule le symbole [\*] – toujours dans la version utilisée – ne soit pas satisfaisant. Une recherche à l'aide de «auia? \*ado» ne donne guère de meilleurs résultats (elle a été arrêtée au bout de 3 minutes, sans avoir abouti).

Le principe est pourtant excellent. Il permet, entre autres, de neutraliser les variations graphiques, ex. si l'on recherche une séquence comme «(la) reina nuestra señora» où «reina» et «señora» sont susceptibles de s'écrire de diverses manières; on pourrait formuler la demande [re\*na nuestra se\*ora]. Les symboles devraient même pouvoir désigner des séquences-mots de 1 caractère ([?] précédé et suivi d'un blanc, en posant que [?] ne pourrait désigner que 1 caractère, non 0) ou de plusieurs caractères ([\*] précédé et suivi d'un blanc, en posant que [\*] ne pourrait dépasser l'espace d'un mot): l'expression retenue pour la demande permettrait alors d'étendre la recherche à des séquences textuelles beaucoup plus larges, ex.: [a \* de] («a fin de acrescentar», *Claros varones*, p. 11, f° 6r; «a suplicacion del principe», *ibid.*, p. 34, f° 17v; «estando a punto debatalla», *ibid.*, p. 41, f° 21r; «venir a manos del injuriado», *ibid.*, p. 50, f° 25v; «a voluntad del que gelo dio», *ibid.*, p. 84, f° 42v; etc.), [s\*n \* \*r] (ex. «sin se mouer», *De proprietatibus rerum*, p. 238, f° 119v, etc.; «syn lo poner», *ibid.*, p. 588, f° 294v; «syn le pedir», *Ordenanzas reales*, p. 103, f° 53v), à quoi l'on pourrait opposer [s\*n \*rse], [s\*n \*rle?], etc. (ex. «syn

(13) Le Manuel, p. 37, cite comme exemple «auia entrado»; il se trouve que cette séquence est inexistante dans la base.

(14) Mais dans la page de la première occurrence [Valerio de las historias eclesiásticas, p. 153, f° 78v] 2 autres «auian» sont affichés en rouge alors qu'il ne sont pas suivis de «entrado»; de même dans la page 114, f° 57v de la *Crónica de España*, correspondant à la seconde occurrence, un autre «auian» apparaît sur fond rouge. V. plus bas les remarques sur la programmation et le traitement des textes.

interponerse», *Universal vocabulario*, p. 225, f° 113r). Pour n'envisager que quelques schémas simples... On ne peut que souhaiter une mise au point rapide des programmes qui ne marchent pas encore parfaitement<sup>(15)</sup> et une extension dans le sens qui vient d'être suggéré.

Mais la panoplie des modalités d'interrogation proposées au chercheur est vaste. Le logiciel utilise notamment deux types d'opérateurs: des opérateurs symboliques et des opérateurs booléens.

L'un des opérateurs symboliques est l'opérateur de distance [*n*], où *n* est un nombre correspondant à un espace maximum en mots entre deux termes cités à gauche et à droite; ex. «tratado .2 rieptos» (Manuel, p. 37), expression par laquelle on recherche «rieptos» à deux mots, au plus, de «tratado». Si on introduit cette question, la réponse est positive (1 ex.): on rencontre en effet «Tratado delos Rieptos» dans le *Tratado de las armas*, p. 1, f° 1r, ce qui est vérifié par l'analyse des diverses occurrences de «rieptos». Mais on constate qu'on obtient la même réponse si l'on interroge à l'aide de l'expression «tratado .1 rieptos», ce qui signifie que «delos» est traité comme un seul mot<sup>(16)</sup>. En fait, «delos» figure dans le Glossaire des formes et si on sélectionne cette unité, on s'aperçoit qu'elle revient dans 5 659 pages de la base! Voilà qui rend aléatoire toute recherche fondée sur la distance, car la préposition et l'article sont aussi graphiés «de los» – il n'est pas possible de savoir combien la base en contient d'occurrences (le résultat est toujours «0»), alors que si l'on traite les textes séparément, le nombre de pages est fourni pour certains (par ex. pour *Bías contra Fortuna*: 2 pages – 3 occ. réelles –); pour d'autres, non (ex. *Bocados de oro*: réponse «0»)<sup>(17)</sup>.

Le traitement de «delos» comme un mot n'explique pas, de toute façon, la réponse «0» qu'on obtient si l'on remplace dans la question première «tratado» par «titulo», que la distance choisie soit 2 ou 1, puisque l'expression

(15) En attendant que Micronet adresse aux acheteurs une version améliorée du logiciel, F. MARCOS suggère de recopier, sous FTP Anonymous, les programmes Admyte.exe, Admyte.ins et Admyte.ayu (ou Admyte.hlp) à l'adresse [lola.lllf.uam.es].

(16) Il est donc inexact de présenter «Tratado de los Rieptos» comme une réponse pour la valeur 2 de la variable *n* (Manuel, p. 37), d'autant que cette séquence, sous la forme citée, est absente de la base.

(17) La même erreur d'analyse se retrouve pour «alos» 3360, «ala» 3613, «alas» 1822 (combien d'occurrences du N *ala* dans ces deux nombres?), «conel» 1 642, «conella» 228, «conela» 2, «conello» 62, «conellos» 344, «conellas» 91, etc., «dela» 6471, «delas» 4206, «dellos» 1899, «dellas» 1057, «dello» 647, «deste» 1461, «desta» 1942, «desto» 910, «daqueste» 5, «daquestos» 4, «daquesta» 19, «daquestas» 6, «daquesto» 8, etc.

«titulo delos rieptos» figure 3 fois dans la base (*Siete Partidas*, p. 743, f° 372r, p. 768, f° 384v; *Tratado de las armas*, p. 17, f° 9v). Cette réponse est d'autant plus surprenante que si, au lieu d'introduire «titulo» comme substitut de «tratado», on recourt à «fablamos», la réponse juste est alors obtenue: «fablamos delos rieptos» se trouve effectivement deux fois dans les *Siete Partidas* (p. 743, f° 372r et p. 745, f° 373r). Impossible, par conséquent, de se fier à l'opérateur de distance dans l'état actuel du logiciel.

Les opérateurs logiques sont au nombre de trois: *ou*, *et*, *sauf* (qui peuvent être formulés soit en espagnol: [.o], [.y], [.no], soit en anglais: [.or], [.and], [.not], soit à l'aide des symboles [+], [&], [!]). Le premier, utilisé avec les deux termes cités dans l'exemple du Manuel et sur l'*Universal vocabulario* d'A. de Palencia, «*medicina .o armas*», fournit un résultat qui paraît cohérent: au total, 52 pages (qui donc renferment soit «*medicina*», soit «*armas*»); si l'on interroge «*medicina*» seul, on obtient «7»; si l'on interroge «*armas*» seul, la réponse est «45»; la somme est exacte et l'on peut déduire de ces renseignements que «*medicina*» et «*armas*» ne se trouvent jamais ensemble dans une même page de texte (ce que vérifie l'examen des occurrences).

Cet opérateur, qui n'est pas en général le plus intéressant des trois, peut s'avérer très précieux; en particulier lorsqu'on veut étudier l'emploi d'hyponymes et d'hypéronymes, de synonymes, d'antonymes, etc. Ex., dans l'*Universal vocabulario* d'A. de Palencia, «*espada .o puñal*» indique 16 pages; mais «*espada*» interrogé isolément en signale 13, et «*puñal*» 5; on sait donc immédiatement que «*espada*» et «*puñal*» se rencontrent à deux reprises dans une même page (p. 241, f° 121r et p. 362, f° 181v), mais que la plupart du temps ils ne sont pas mis en relation directement l'un avec l'autre. Quand ils le sont, ils apparaissent dans le même contexte: p. 241, «*en la vayna do se esconde el puñal o la espada*»; p. 362, «*Gladius. se toma por espada & por puñal*». De même, trois questions brèves, entraînant des réponses immédiates, peuvent éclairer sur les relations, dans ce texte, entre «*enfermo*» et «*sanar*»: «*enfermo*» → 18, «*sanar*» → 8, «*enfermo .o sanar*» → 24; les deux termes interviennent donc de préférence dans des contextes différents; deux fois, néanmoins, ils se trouvent rapprochés: p. 103, f° 52r («*esta aue [Calandrius] si mira al enfermo ha de sanar*») et p. 861, f° 432r («*Sanare. sanar & curar de enfermedad. assi que aya sanidad el enfermo*»)<sup>(18)</sup>. Une recherche sur «*enfermo*» et «*sano*» donne des résultats très voisins: «*enfermo*» → 18, «*sano*» → 16 (mais dans 4 pages il s'agit du verbe latin), «*enfermo .o sano*» → 32 (à corriger donc); ils ne

(18) Il faut néanmoins chercher ces cas d'espèce en faisant défiler les pages une à une sur l'écran.

se trouvent en situation d'antonymes contextuels que p. 409, f° 205r: «o estas sano o enfermo. pero no estas sano: immo de verdad enfermo».

La possibilité de préfixer par [=] les mots qui servent de bornes dans la question en choisissant la forme canonique retenue comme lemme pour les formes lemmatisées ajoute encore à l'utilité de l'opérateur d'union dans des recherches comme celles qui viennent d'être esquissées. Ainsi, «=enfermo» ou «=sano» devraient associer toutes les variables fléchies de chaque lemme. Il se trouve cependant que la lemmatisation a inclus le premier dans les formes d'*enfermar*, le second dans les formes de *sanar* (cf. plus haut les commentaires sur la lemmatisation). La question doit donc être formulée: «=enfermar .o =sanar». La réponse (toujours sur le même texte) fournit l'ensemble des pages où se rencontre l'une des formes au moins de ces deux vocables, «83». Comme «=enfermar» → «35» et «=sanar» → 54 (total: 89), l'intersection entre les deux sous-ensembles se limite à 6 pages (en première approximation). Mais on ne peut aller très loin dans l'interprétation sans dépouiller les occurrences d'«enfermo» et de «sano», qui peuvent correspondre pour l'un et pour l'autre aux formes verbales (PR1, PT3) aussi bien qu'au masculin de l'adjectif. D'autre part, si l'on relève une à une les valeurs (en pages) pour chaque forme fléchie de «=enfermar», on obtient un total de 38, non de 35, comme indiqué dans la réponse du logiciel. La lemmatisation incomplète entraîne, par conséquent, une autre limitation que celles déjà signalées: elle enlève toute garantie aux résultats d'opérations utilisant des symboles comme [=] et [.o] parce qu'on ne sait jamais exactement avec quelles données on opère<sup>(19)</sup>.

L'opérateur d'intersection [.y] offrirait-il une plus grande fiabilité? Pour la raison qui vient d'être évoquée la réponse ne peut être que négative, sauf si l'on opère avec des formes non ambiguës, comme dans l'exemple du Manuel, p. 39: «historia .y España». Mais les 12 pages indiquées ne signalent pas nécessairement des passages où les deux formes choisies sont mises en relation: la

(19) Un autre test a été fait sur le *Lilium medicinae* pour «=enfermar .o =curar». Le résultat affiché est «303»; «=enfermar» → 91, «=curar» → 288 (total: 379). Il y aurait donc, d'après ces chiffres, 76 pages communes. Mais si l'on totalise les valeurs obtenues en interrogeant chaque forme fléchie, on trouve pour «=enfermar» 97, pour «=curar» 568 (total: 665, soit une différence de 286 par rapport au total précédent). L'origine de certaines anomalies est claire: «curallo» 'curarlo' 1, «curenlas» 1, «curenlo» 1, ont été traités comme des mots uniques et non comptabilisés dans les formes de *curar* (contrairement à «curanse» 4, «curarse» 6, «curase» 4, «curense» 21, «curese» 21, si l'on en croit le Glossaire des Lemmes, qui les cite sous *curar*). Mais ils représentent une quantité négligeable dans le total 568. De toute manière le regroupement de «cura» 243 et de «curar» 83 (tous deux présents dans la liste du Glossaire), soit 326, dépasse déjà largement le nombre fourni par le logiciel pour «=curar» (288)...

distance ou la nature du texte peut les séparer sémantiquement l'une de l'autre; c'est ce qui se passe p. 390, f° 195v, de l'*Universal vocabulario*, où les 2 occ. d'«España» et les 4 d'«historia» appartiennent à deux articles différents du dictionnaire, ou encore p. 147, f° 74r, du *Dictionarium hispano-latinum* de Nebrija, p. 170, f° 86r, du *Vocabulario eclesiástico* de R. Fernández de Santaella. Dans un discours suivi, en revanche, la cooccurrence de deux formes revêtues d'une valeur thématique pourrait fournir d'utiles indications et faciliter la découverte éventuelle de réseaux lexico-sémantiques.

Il existe une certaine redondance entre [.y] et [.n]: le second permet simplement de réduire l'écart entre les deux occurrences communes. Étant donné que le résultat de la recherche [.y] est obtenu très rapidement, cet opérateur peut aider néanmoins à un premier débroussaillage. Aussitôt après, [.n], pour une valeur de 10 (ou moins de 10 selon la nature de la recherche) précisera la fréquence des coïncidences en contexte limité – une vérification s'imposant, comme toujours, au niveau des textes pour séparer les juxtapositions aléatoires des mises en relation significantes. Ainsi, pour l'exemple «historia .y España», la valeur 10 de *n* réduit le nombre des pages de 12 à 1 et le passage qui répond à cette condition est en effet intéressant à plus d'un titre. Il se rencontre dans la *Crónica de Aragón*, p. 29, f° 15r: «Ved que rey escogian para su hespaña los nuestros. para del començar: como de rey tan primero de toda su historia. ved si fuera vn poco mejor y mas honrra de españa mirar mejor sus historias». On constate que d'autres occurrences pourront apparaître si l'on modifie l'orthographe d'«España» ou si l'on fait passer «historia» au pluriel...

Effectivement, «historia .y Hespaña» → 8, nombre d'occurrences qui n'est pas négligeable par rapport au premier; en réduisant le champ avec *n* = 10, on obtient 4 contextes, tous situés dans la même *Crónica de Aragón* (pp. 14, 26, 28, 29). Ici, les occurrences voisines sont plus nombreuses qu'avec la variante moderne «España». Comme le premier élément de la question peut également varier, un contrôle général sur les lemmes s'impose; «=historia .y =España» est, malheureusement, une question impossible (réponse: «0») du fait que «historia(s)», «hystoria(s)», «ystoria(s)», «estoria(s)», etc. ont été lemmatisés sous *historiar*; peu importe, dans ces conditions, que «España» → 676 et «=historiar» → 512; «=historia» donnant «0», aucune de ses occurrences ne peut être retrouvée dans les pages où se rencontre «=España». On pourrait sans doute, dans ce cas particulier, contourner l'obstacle en observant que les formes verbales sont très minoritaires dans l'effectif rattaché à «=historiar» («historiar» → 0, «hystoriada» → 2), ce qui permet de les négliger; mais «=España» donne à son tour «0» comme réponse, malgré les 676 pages où est identifiée la forme «España», parce que les diverses variantes d'*España* n'ont pas été lemmatisées (le lemme le plus proche, qui ne couvre pas *España*, est

*español*). L'unique recours est alors d'interroger les formes auxquelles on pense l'une après l'autre. On aura, certes, progressé dans l'analyse de la question, mais sans être certain de n'avoir rien oublié (par ex. «hyspaña» → 9, «hyspañas» → 3).

Enfin, l'opérateur d'exclusion [.no] 'sauf', dans les mêmes limites que le précédent, est susceptible d'être mis à profit<sup>(20)</sup>. Par exemple, «estoria» → 33, «estorias» → 35; «estoria .y estorias» → 1 (*Nobiliario vero*, p. 47, f° 24v<sup>(21)</sup>); «estoria .no estorias» → 32, résultat correct; de même, «estorias .no estoria» → 34. A partir de quoi l'on peut envisager plusieurs vérifications d'ordre à la fois morphologique et historique, pour savoir si les 4 variantes de *historia*, «estoria», «historia», «hystoria» et «ystoria» s'excluent ou s'il arrive qu'elles se rencontrent dans les mêmes textes. La confrontation peut être effectuée rapidement:

– «estoria» → 33	«estoria .no historia» →	33
	«estoria .no hystoria» →	26
	«estoria .no ystoria» →	33
– «historia» → 95	«historia .no estoria» →	95
	«historia .no hystoria» →	80
	«historia .no ystoria» →	91
– «hystoria» → 192	«hystoria .no estoria» →	185
	«hystoria .no historia» →	177
	«hystoria .no ystoria» →	189
– «ystoria» → 43	«ystoria .no estoria» →	43
	«ystoria .no historia» →	39
	«ystoria .no hystoria» →	40

Une valeur identique à gauche et à droite signifie que les formes correspondantes s'excluent totalement; c'est le cas pour «estoria» / «historia» ~ «ystoria», résultat confirmé par l'inversion des termes dans la formule de l'interrogation. L'exclusion n'est pas totale pour «estoria» ~ «hystoria» (7 rencontres), «historia» ~ «hystoria» (15 rencontres), «historia» ~ «ystoria» (4 rencontres), «hystoria» ~ «ystoria» (3 rencontres). La variation «hist-» ~ «hyst-» ~ «yst-» concerne surtout la graphie, mais peut être révélatrice aussi d'une certaine attitude par rapport à l'activité désignée; la variation «est-» ~ «hist-» ajoute à une différence graphique une opposition phonétique. Étant donné la

(20) L'exemple proposé par le Manuel, [Cervantes .no Quijote], induit évidemment une réponse négative: les textes de la base sont tous antérieurs à l'époque de Cervantes.

(21) «Otro si teseo fijo de egeo en el cuento delos rreyes de atenas. Esta segund paresçe en la segunda parte dela estoria general del mundo pero en otras estorias llamase duque de atenas.»



dissymétrie entre la dissociation et la rencontre que met en évidence le test par [.no], il est plus économique, dans le cas présent, d'examiner les résultats du test par [.y], qui indiqueront les œuvres où l'exclusion n'est pas absolue.

En examinant les pages où se produisent les rencontres, on constate que six œuvres seulement font intervenir des variantes dans une même page. Ces textes datent de la fin, voire de l'extrême fin du XV<sup>e</sup> siècle: *Ceremonial de príncipes*, de D. de Valera [a.1474], *Crónica popular del Cid* [fin XV<sup>e</sup>, 1<sup>re</sup> éd. 1498<sup>(22)</sup>], *Peregrinatio in terram sanctam*, de B. de Breidenbach, trad. de M. Martínez de Ampíes [1498], *Tratado de Roma*, de M. Martínez de Ampíes [1498], *Crónica de Aragón*, de G. F. de Vagad [1499], *Vocabulario eclesiástico*, de R. Fernández de Santaella [1499]. Un seul texte, la *Crónica popular del Cid*, utilise dans des contextes voisins la forme médiévale évoluée phonétiquement, «estoria», et la forme savante dans sa variante «hystoria» (les 7 rencontres sont de ce type: pp. 40, 128, 131, 147, 157, 164, 166); tout en mêlant parfois les variantes savantes («historia» ~ «hystoria» pp. 74 et 138; «historia» ~ «ystoria» p. 27; «hystoria» ~ «ystoria» pp. 68 et 97). Cette particularité est peut-être en relation avec les sources ou avec certaines traditions dont l'œuvre a pu s'inspirer; le problème est à étudier, mais il n'est correctement posé que si l'on établit, à l'échelle du texte, les modalités de la variation: «estoria» → 13 («estorias» → 0) / «historia» → 8 («historias» → 0), «hystoria» → 149 («hystorias» → 0), «ystoria» → 6 («ystorias» → 1). Autrement dit, même dans le texte (unique) qui conserve encore des traces de la forme médiévale la plus courante<sup>(23)</sup>, la forme savante occupe presque tout l'espace morphologique et affirme sa prééminence en adoptant la graphie apparemment la plus éloignée de la phonétique, «hy-». Cette information n'est pas seulement utile au phonéticien ou à l'historien de la graphie. Elle peut fournir un cadre efficace de recherche au lexicographe ou à l'historien de la littérature, à l'historiographe... Admyte1 permet de continuer l'enquête à l'aide des diverses procédures qui ont été décrites antérieurement.

D'après le Manuel, p. 38, les opérateurs booléens sont compatibles avec les autres modes d'expression. On peut comprendre qu'ils acceptent soit les chaînes de caractères simples, soit des spécifications à l'aide des autres opérateurs.

Dans la transcription des textes, cependant, le signe «&» a été employé pour transcrire «z». Or [&] est d'autre part le symbole algébrique équivalent

(22) D'après *BOOST 77*, réf. 381.

(23) Cf. Jean ROUDIL, «Index alphabétique des formes de la *Primera Crónica General*», Paris, 1979: «estoria» 924, «hestoria» 1, «estorias» 154 / «istoria» 1, «ystoria» 2, «historia» 2, «hystoria» 3, «istorias» 2, «ystorias» 2, «hystorias» 3.

à [.y]. On devrait pouvoir utiliser des expressions du type «e & &», comme on utilise «e & comio» (qui fonctionne correctement). Mais, et c'est logique puisque le premier terme de l'expression peut comprendre plusieurs mots, les deux «&» ne sont pas identifiables comme des éléments différents, le premier en tant qu'opérateur, le second en tant que donnée ou valeur de la variable [second terme]. Plus largement, l'opérateur [&] n'est pas reconnu lorsqu'on associe l'opérateur de distance, que l'on place ce dernier avant («e .2 & comio») ou après l'opérateur («e & .2 comio»). En réponse le logiciel signale toujours une erreur de syntaxe. Il en est de même si l'on associe dans une expression l'opérateur d'intersection [.y] ou l'opérateur d'exclusion [.no] et l'opérateur de distance [.n], dans des combinaisons comme «e .y .2 comio», «e .2 .y comio», «e .no .2 comio», «e .2 .no comio».

L'un des procédés, en dehors de la programmation, pour résoudre le problème, serait de mettre dans un fichier tous les éléments d'une réponse à une question simple, par ex. à la question «e .y comio»; on pourrait ensuite, sur cette sélection, faire intervenir la question «e .2 comio». Mais il ne semble pas que cette possibilité ait été retenue, voire envisagée, par les concepteurs du logiciel. Et, en effet, avec des réponses qui mobilisent une page entière à chaque fois qu'une occurrence au moins est relevée il faudrait prévoir des espaces très importants pour des questions larges portant sur l'ensemble de la base, à moins de fixer à un nombre *n* de réponses un seuil limite pour ouvrir un fichier de stockage.

Si l'utilisation conjointe des opérateurs symboliques et des opérateurs logiques ne fonctionne pas dans tous les cas, l'association de deux ou de plusieurs opérateurs booléens détermine une recherche et aboutit à des résultats. Ainsi, sachant que les formes du lemme *infancia* (15) sont «infancia» (4), «infançia» (6) et «ynfançia» (5), on peut se demander quelles sont les coïncidences des deux premières avec le verbe *llamar*. L'expression d'une telle opération pourra être «=infancia .y =llamar .no ynfançia»; résultat (vérifié): 5<sup>(24)</sup>. Une autre formulation, puisqu'on sait au départ que les variantes du lemme sont au nombre de trois, serait: «(infancia .o infançia) .y =llamar». A cause de la hiérarchie assignée aux opérateurs ([.y] l'emporte sur [.o] ou [.no], qui se situent au même niveau), l'expression «infancia .o infançia .y =llamar» ne don-

(24) Cf. *Universal vocabulario* d'A. de Palencia, p. 142, f° 71v; p. 281, f° 141r; p. 459, f° 230r; *Vocabulario eclesiástico* de R. Fernández de Santaella, p. 17, f° 9v; *Peregrinatio in terram sanctam*, p. 136, f° 112v (mais «infancia», dans ce passage, est un mot latin; l'occurrence n'est donc pas pertinente si l'on ne veut interroger que le discours en castillan).

Dans deux cas se posent des problèmes d'affichage qui seront examinés plus loin.

nerait pas le même résultat parce que le calcul rechercherait d'abord toutes les coïncidences de «infancia» et de «=llamar» et ajouterait ensuite les occurrences de «infancia», soit  $3 + 4 = 7$ : deux des occurrences de «infancia», qui n'ont pas d'intersection avec *llamar*, seraient alors retenues. Les parenthèses, comme dans les calculs arithmétiques, modifient la priorité et font que la recherche des coïncidences porte sur l'ensemble des occurrences des deux variantes.

A partir de la dernière expression, on peut envisager une recherche où seraient exclues les occurrences de «llama» (du fait de la fréquence relative de «se llama» dans les définitions<sup>(25)</sup>): «(infancia .o infancia) .y (=llamar .no llama)». La réponse est «1», mais elle est erronée, car l'analyse de la réponse à la question sur «=llamar» avait permis, auparavant, de recueillir 4 passages répondant à la nouvelle question. L'explication de cette anomalie est un peu complexe, mais découle de l'état actuel de la base, des options prises pour la lemmatisation et la programmation. Ce qui se passe, c'est que toutes les pages où se rencontre la forme «llama» sont éliminées, sans tenir compte du fait qu'une même page peut renfermer, outre la forme «llama», d'autres formes de *llamar*. La seule page retenue est donc l'unique page où se rencontrent à la fois «infancia» et «se llama» (*Universal vocabulario* d'A. de Palencia, p. 459, f° 230r), à l'exclusion de «llama», tandis que sont rejetées les 3 autres passages qui constitueraient pourtant une bonne réponse à la question: *ibid.*, p. 142, f° 71v («infancia» et «llaman» en plus de «llama»), *Vocabulario eclesiástico* de R. Fernández de Santaella, p. 17, f° 9v («infancia» et «llaman», «llaman», en plus de «llama»), *Peregrinatio in terram sanctam*, p. 136, f° 112v («infancia» [en fait, mot latin] et «llaman», «llamado», en plus de «llama»).

Ces quelques exemples montrent, d'un côté, la souplesse et la puissance du logiciel d'interrogation, de l'autre, les limites à la fiabilité des réponses. Aucun des deux aspects ne doit être surévalué, mais non plus minimisé: dans l'état présent des choses, pour être (presque?) sûr que l'information obtenue est correcte, il est conseillé de la vérifier par des questions complémentaires et par la consultation des contextes; mais il arrive aussi qu'une absence de réponse dissimule l'existence d'information correspondant à ce que l'on recherche. Un détour est souvent nécessaire pour la découvrir et ce détour n'est possible que si les effectifs restent peu importants.

Encore faut-il, pour que l'information soit repérable, que la transcription des textes soit elle-même absolument fiable et que le logiciel, qui doit s'adapter à des conventions très minutieuses, tourne sans ratés dans des conditions parfois difficiles et d'une extrême complexité logique.

(25) Pour la base entière, «=llamar» → 3868, «llama» → 1004.

### 3. L'information dans la base et son utilisation

#### 3.1. Représentation du texte-source (conventions)

Les conventions appliquées dans la transcription respectent les normes fixées par la 4<sup>e</sup> édition du *Manual de transcripción para el «Diccionario de Español Antiguo»* (Madison, 1992) – Manuel, p. B1. La version mémorisée est donc une version aussi proche d'une transcription paléographique que le permettent les ressources de l'informatique. Sont conservés: le découpage des mots et des lignes, la graphie, la ponctuation, l'usage des majuscules de l'original. Les restitutions sont signalées à l'aide de crochets obliques: «sabia me<n>te», «q<ue>rer». Lorsqu'un mot est tronqué entre deux pages (folios), le dernier fragment est remonté à la page précédente et rattaché au premier à l'aide d'un trait d'union, signe utilisé par ailleurs quand un mot est tronqué en fin de ligne (les incunables n'emploient en général aucun signe pour marquer la troncation), ex.: «ente<n>-d<e>r» (*Siete Partidas*, p. 5, f° 3r: «der» est sur le folio 3v), «razo-nes» (*ibid.*, p. 9, f° 5r: «nes» est sur le folio 5v). A l'inverse, quand deux mots sont collés, le transcripteur les sépare quelquefois en introduisant entre eux la marque «[]», ex.: «vn solo[]v<er>dadero dios» (*ibid.*, p. 9, f° 5r), «la[]vsase<n>» (*ibid.*, p. 15, f° 8r)<sup>(26)</sup>. Tout ajout est signalé également par «[]», ex. l'insertion des points oubliés avant ou après les chiffres romains: «Ley [.]xxiii.» (*ibid.*, p. 15, f° 8r), «Las .ix[.] cosas» (*ibid.*, p. 77, f° 39r)<sup>(27)</sup>.

Hormis ces cas d'espèce, aucune intervention critique ne modifie normalement le texte original, si ce n'est la rectification de ce qui peut apparaître comme une erreur matérielle. A ce type appartient le rétablissement de l'ordre des caractères quand deux d'entre eux ont été inversés; deux opérations sont alors signalées: la suppression du premier caractère par «()» et sa restitution au bon endroit par «[]», ex. «T(t)i[t]ulo» cinco» (*ibid.*, p. 27, f° 14r; source: «Ttiulo»), «descom(n)[u](u)[n]ion» (*ibid.*, p. 87, f° 44r; source: «descomnuion»), «m(y)a[y]or mente» (*ibid.*, p. 59, f° 30r; source: «myaor»),

(26) Les prépositions, fréquemment collées, ne sont pourtant pas séparées: «ADios deue o<n>bre temer» (*ibid.*, p. 2, f° 1v), «dela desatar» 'de desatarla' (*ibid.*, p. 9, f° 5r), «ensu poder» (*ibid.*); pas plus que les pronoms enclitiques: «desatarlo» (*ibid.*), «llegale» (*ibid.*). Cf. note 17.

(27) Mais la modification n'est pas systématique: cf. *ibid.* «ti<tulo> iiij. ley .v.» (p. 7, f° 4r), «Ley viij.» (p. 12, f° 6v), «El iiij. caso» (p. 15, f° 8r), «la ley iij titulo xij» (p. 525, f° 263r), «ley xij» (p. 527, f° 264r). De toute façon, la graphie de l'original n'est pas absolument régulière non plus; les nombres sont parfois exprimés par des mots, qui peuvent être suivis ou non d'un point, mais ils ne sont jamais, semble-t-il, précédés par un point: «Ley siete que quiere dezir maestrescuola» (*ibid.*, p. 49, f° 25r), «el titulo treze ley diez & seys.» (*ibid.*, p. 169, f° 85r), «Ley veynte. dela penitencia» (p. 15, f° 8r), «Ley dozena. quales non deuen ser caualleros» (*ibid.*, p. 228, f° 114v).

«ni(u)[n]g(n)[u]n» (*ibid.*). Ailleurs c'est un caractère à l'évidence fautif qui est remplacé: «s(a)[e] faze» (*ibid.*, p. 50, f° 25v; source: «sa faze»), «por bondad d(o)[e] obras» (*ibid.*; source: «do obras»); un fragment redoublé que l'on supprime: «lo en-(en)barga» (*ibid.*, p. 66, f° 33v); un caractère manquant que l'on ajoute: «los calices & la[s] cruces» (*ibid.*, p. 25, f° 13r); une altération de copiste absolument certaine en contexte que l'on corrige: «son des(t)[c]o()mulgados» (*ibid.*, p. 544, f° 272v; source: «son desto mulgados»; contexte: «E demas desto son des(t)[c]o()mulgados estos a tales tan so-la mente por el hecho segund dize en titulo de las descomulgaçiones»)<sup>(28)</sup>.

Dans l'ensemble, et compte tenu des difficultés que pose une transcription minutieuse des incunables, les conventions sont appliquées avec soin. Il conviendrait, malgré tout, de mieux fixer les cas et les modalités d'intervention. Quand le transcritteur sépare deux éléments d'une séquence, il devrait séparer en toutes circonstances les éléments identiques ou analogues. Or on lit d'un côté «otro(s)i» (*ibid.*, p. 397, f° 199r) et d'un autre «otro si» (pp. 17, 29, 41). On lit «a[ ]algund onbre. o a[ ]alguna muger» (p. 394), «a[ ]ora de nona» (p. 22), «con[ ]derecho» (p. 31), «de[ ]fazer» (p. 27), «el sabado de[ ]lazaró» (p. 31), «en[ ]latin» (p. 18), «en[ ]pecado mortal» (p. 19), «por[ ]cubdicia» (p. 25), «aque[ ]l[ ]o[ ]mezillo» (p. 51), «tan[ ]solamente» (p. 191), «fazer[ ]limosnas» (p. 38)... et, par ailleurs, «ael-los se tornaria» (p. 387), «no los deuen vedar aningund cristiano» (p. 20), «puede yr el onbre aconfesarse aotri» (p. 18), «aora de mu(r)e[r]te» (p. 29), «sopena de escomunion» (p. 19), «encontrario» (p. 26), «estan entiniebla» (p. 27), «tangor-do» (*Visión delectable*, p. 114), «de santispiritus» (*Siete Partidas.*, p. 22), «cadaaño» (*Dictionarium hispano-latinum*, p. 34)...

D'une manière générale, dès l'instant où l'on admet le principe d'une modification même minime du texte-source (rétablissement de l'ordre des caractères, ajout ou retrait d'un caractère), il est malaisé de fixer une limite à l'intervention du transcritteur dans la version électronique. Si l'un des buts principaux à atteindre est bien l'identification correcte des formes occurrentes dans le Glossaire et, par voie de conséquence, dans les résultats affichés et dénombrés, il serait indispensable que dans la représentation du texte en mémoire *tous* les mots agglutinés soient séparés. On ne peut admettre, ici «santispiritus» (déjà cité), là «spiritu[ ]santo» (*Siete Partidas*, p. 93), tantôt

(28) Ces retouches, le plus souvent justifiées, relèvent parfois d'une mauvaise lecture du transcritteur. Ex. «p(n)[u]eden ser ordenados» (*ibid.*, p. 53, f° 27r; la source est claire: «pueden»), «si algu(u)[n]d i(n)[u]yzio» (p. 367, f° 184r: «algund», *id.*, cf. le tracé des *n* dans la ligne située peu après: «enel pleyto antel iuez»), «todo qua(u)[n]to troxiere» (p. 385, f° 193r: «quanto», non ambigu).

«cadavno» (*ibid.*, pp. 3, 15, 99, 595), tantôt «cada[]vno» (pp. 148, 149, 151, 161), etc., parce que le logiciel d'analyse n'identifie pas deux éléments distincts dans «santispiritus» ni dans «cadavno» et que le Glossaire, du même coup, intègre de telles séquences comme des formes différentes (cf. ci-dessous). Mais cette segmentation du texte, signalée et parfaitement repérable (par ex. pour des études sur la graphie), ne modifie pas la lettre elle-même. Une inversion est déjà une correction. De là à corriger un texte que l'on pense fautif il n'y a qu'un pas, qu'on peut franchir avec les meilleures intentions du monde (au risque de se tromper), mais qu'on ne franchira pas toujours (en tombant alors dans l'incohérence).

On n'est plus dans la perspective d'une transcription, mais dans celle d'une édition critique traditionnelle, quand on introduit par émendation «alquile[r]es» p. 4 des *Partidas*, là où la source porte «alquilees». Or le texte ne présente nulle part la forme «alquiler»; seule est attestée la forme «alquile» (pp. 404, 4 fois; 405, 6 fois; 629; 631). Le *DCECH* de Corominas-Pascual rappelle cette particularité et remarque: «*Alquilé* se lee todavía en el *Conde Luc.* (ed. Hz. Ureña, 192), en las *Andanzas* de Tafur, en Tallante [...], en Nebr., en Gutiérrez Santaclara y en Ruiz de Alarcón (rimando con *pie*)». Voilà qui donne à penser que l'adjonction du *r* est erronée; mais la possibilité qu'elle le soit n'a même pas été vérifiée dans le texte. Pourtant Nebrija ne précise-t-il pas, dans sa *Gramática castellana* (1492): «*alquile, rabe* [...] tienen acento agudo en la ultima» (p. 154:6-7), «Sacanse los que acaban en *e* aguda, por que sobre el singular reciben esta terminacion *es*, como *el alquile, los alquilees*» (p. 189: 4-6)? D'un autre côté, si l'on estime nécessaire de modifier le pluriel, pourquoi respecter les 12 occurrences du singulier? Cette double inconséquence est à regretter sur tous les plans: non seulement elle introduit dans les relevés (dont le Glossaire) des formes fictives, mais elle montre que ni les transcripteurs, ni leurs correcteurs, qui préparent une exploitation informatique des textes, ne travaillent eux-mêmes avec l'informatique lorsqu'ils établissent leur transcription.

C'est sans doute ce qui explique que, malgré un effort manifeste pour transcrire les textes conformément aux conventions retenues (effort auquel l'utilisateur est extrêmement sensible), tant d'erreurs matérielles subsistent dans la base qu'on a parfois l'impression que la transcription initiale n'a pas été vérifiée.

### 3.2. *Qualité et fiabilité des textes*

Étant donné le volume de la base, il n'était pas possible de procéder à un examen systématique de tous les textes qu'elle contient. Les erreurs qui vont

être signalées ne représentent donc que quelques exemples, glanés au cours de recherches diverses et de vérifications ponctuelles. Ce tour d'horizon, par nature très ingrat, servira cependant aux utilisateurs des textes (déchargés d'Admyte) pour améliorer leur(s) version(s) et à l'équipe scientifique d'Admyte pour supprimer de nombreuses coquilles dans la prochaine édition du disque<sup>(29)</sup>.

Il ne sera ici question que des erreurs matérielles. Les problèmes liés au fonctionnement du logiciel, à l'affichage, à l'utilisation des outils proposés, seront traités plus loin.

Pour des raisons de commodité, puisque la liste de référence de la base est la liste des titres, les erreurs relevées seront citées par ordre alphabétique des titres, le nom de l'auteur figurant entre crochets. Mais le folio sera préféré à la page, afin que les textes déchargés (dépourvus de pagination) puissent être modifiés sans difficulté. La flèche qui suit la parenthèse ouvrante indique la forme ou la séquence correcte, vérifiée dans la source.

– *Bocados* [Al-Mubasshir, trad.], f° 44v: «may ayna» (→ «mas ayna»).

– *Capítulos de gobernadores* [Ferdinand II d'Aragon et Isabel I de Castille], f° 5v: «copia al esciuano» (→ «escriuano»).

– *Ceremonial de príncipes* [D. de Valera], f° 26v: «asoluieresenla tierra» (→ «asoluieres enla»).

– *Crónica de Aragón* [G. F. de Vagad], f° 62v: «que los mosros recibieron» (→ «mosros»).

– *Crón. de Esp.* [D. de Valera], ff. 3r: «un grand Rio llamdado indo» (→ «llamado»), 9v: «las cosas escrias por los autores» (→ «escritas»), 75r: «que no ephasen» (→ «pechassen»), 106v: «asperura» (→ «asperura»).

– *Crónica popular del Cid*, ff. 53r: «auquel nombre» (→ «aquele»), 65v: «E por esta rezon» (→ «razon»), 128v: «apregonopublicamente» (→ «apregono publicamente»).

– *Cuadernos de las leyes nuevas* [Ferdinand d'Aragon et Isabelle de Castille], ff. 4v: «no valgan alos mal-hechores los priuilegios» (→ «priuilegios»), 5r: «detrimen / to» (→ «detrimen-to»).

– *Cura de la piedra...* [J. Gutiérrez de Toledo], ff. 9v: «como andar aualgando» (→ «caualgando»), 25v: «enel regimien / to»<sup>(30)</sup> (→ «regimien-to»), 61r: «seacriado» (→ «sea[ ]criado»).

– *De proprietatibus...* [B. Glanville]: erreur dans le titre figurant dans la liste des titres et dans le titre courant du texte affiché («*De propietatibus*»), forme correcte dans la

(29) Que les concepteurs d'Admyte veuillent bien considérer le relevé et les mises au point qui vont suivre comme une réponse à leur vœu de collaboration.

(30) Dans tous les exemples le signe [/], qui ne figure pas dans le texte, représentera désormais un saut de ligne, de paragraphe ou de page.

Présentation; ff. 14r: «nuestras enfermedade» (→ «enfermedades»), 14v: «del spiritu san / to» (*id.* source; → «san-to»), 98r: «es esponedor» (→ «el esponedor»), 145r: «algu<n>na» (→ «algu<n>a»), 159r: «esta aue es epqueña» (→ «pequeña»), 160r: «ha tan-ta} / ta» (saut de paragraphe, mais répétition du segment «ta»: → effacer «ta»), 161r: «los epqueños estruzes» (→ «pequeños»), 161v: «por las ayuda del iento» (→ «por la ayuda del viento»), 202r: «es mss presto ençendida» (→ «mas»), 202v: «Dize gragorio» (→ «gregorio»), 205r: «algunmente» (→ «algunamente»), 226r: «lo que es ençessario» (→ «necesario»), 241v: «su apeti / to» (→ «apeti-to»), f° 310v: «conueniente al estomado» (→ «estomago»), 317r: «Algunos numeros pares» (→ «Algunos»), 320r: «so la epogdo<n>s» «se llama epogdo<n>s» (→ «epogdo<o>s»: cf. Nebrija, *Dict. hisp.-lat.*, f° 57v: «Epogdoos. i. proporcion sesquioctaua»).

– *Dict. hisp.-lat.* [Nebrija], ff. 4r: «codear por lo q dar del codo» (→ «que»), 9v: «Accretus. aum.» (→ «Accretus. a. um.»), 19r: «Appulsu. us.» (→ «Appulsus.»), 19v: «Arbitratus [...] aquello msmo» (→ «mesmo»), 66r: «Fraudentus [...] aquello msmo» (→ «mesmo»), 20r: «auqella ciudad» (→ «aquella»), 21v: «la subidaçAscensus» (→ «la subida / Ascensus»), 22v: «por la sturia region de españa» (→ «la asturia»), 43r: «Criticula. ae. por las parrilas» (→ «parrillas»), 46v: «por auqebraz» (→ «quebrar»), 57r: «Epitriutus.» (→ «Epitritus.»), 97v: «hija de auqeste» (→ «aqueste»), 97v: «hija de auqeste» (→ «aqueste»), 120r: «Praestolo. as. aaui.» (→ «aui.»), 129r: «escivir alo que otro escrive» (→ «escribir»).

– *Dict. lat.-hisp.* [Nebrija], ff. 19v: «Boslador. {LAT. plumariuis .ij. phygio .onis}» (→ «plumarius», «phrygio»), 55r: «Garguero. lo que garganta o gargavero» (→ «gargavero», sans «}), 75v: «Omiziano de cuaquiera» (→ «cualquiera»), 80v: «erlcaustes» (→ «encaustes»), 87v: «Renombre propio. {LAT. agnoman [...]}}» (→ «agnomen»), 94r: «Sobrino hijo de ermama» (→ «ermana»).

– *Doctrinal de los caballeros* [A. de Cartagena], ff. 1v: «E qu<n>to se deua» (→ «E qua<n>to se deua»), f° 12r: «mien» (texte: → «enlos mien- / bros» → «enlos mien-bros»), 38v: «a vna duen^a que llamaron» (→ «dueña»), 43r: «sen^orial» (→ «señoril»; comp. «señoril» 3 fois f° 33r), 63r: «a solo dios sed<e>ue» (→ «se[[d[e]ue»: la source ne porte aucun tilde), 83v: «alto estillo de espan^a» (→ «españa»).

– *Escritura de cómo y por qué razón...* [D. Rodríguez de Almela], ff. 51v-52r: «may ormente» (→ «may-ormente»: entre les deux pages le mot a éclaté en deux formes: «may» et «ormente», que l'index a mémorisées séparément).

– *Fasciculus medicinæ* [Johannes de Ketham], f° 5r: «muchede<m>bre de humores» (→ «muchedu<m>bre»).

– *Gram. cast.* [Nebrija], f° 32v: «@Otros» (→ «Otros»).

– *Introd. lat.* [Nebrija], ff. 26r: «Secunds a uerborum» (→ «Secunda»), 48r: «Syllepsis [...] hic currs fuit» (→ «currus»).

– *Leyes hechas...* [Ferdinand II d'Aragon et Isabelle de Castille], f° 6v: «&las escip-turas» (→ «escripturas»).

– *Lilium medicinæ* [B. de Gordonio, trad.], ff. 30v: «con esta se-diondez» (→ «fe-diondez»), 180r: «e<m>fermo» (source: «ēfermo», et partout ailleurs «enfermo», avec ou sans restitution: ex. «e<n>fermo» pp. 9, 10, 12, etc.).

– *Llibre del arte de menescalia* [M. Díez de Calatayud], ff. 10r: «dia nio noche» (→ «ni»), 58v: «algunos beurajes» (→ «algunos»).



– *Nobiliario vero* [F. Mejía], ff. 29r: «es aposente-dor» (→ «aposesta-dor»), 54v: «formicaçion» (→ «formicaçio<n>» [sic]), 56r: «las (c)[ç]erimonias» (→ «cerimonias», sans changement: cf. *passim*, graphie conservée), 82v: «que>brantadas» (→ «q<ue>brantadas»).

– *Ordenanzas reales* [A. Díaz de Montalvo], f° 130v: «(a)A(a)ntigua mente» (→ «A(a)ntigua mente»).

– *Prognostica* [B. de Gordonio?, trad.], ff. 25r: «es apreni-ado enel vapor» (→ «apremi-ado»), 37r: «Algunos es enfermedado» (→ «Alguno es enfermado»).

– *Siete Partidas* [Alphonse X], ff. 2v: «en cada vna destas siete partidas mosrtamos» (→ «mostramos»), 3r: «por aprouecharse dellas» (→ «aprouecharse»), 4v: «porçeu mantiene» (→ «porque»), 11r: «para consaagrarlas» (→ «consagrarlas»), 12v: «los xpi<sit>anos» (→ «xpi<sti>anos»), 13v: «que los tiempos no recudan asus fazones» (→ «sazones»), 21r: «algunos vegadas» (→ «algunas»), 23r: «algunod pecados» (→ «algunos»), 26r: «entinc:ion» (→ «entinc:ion»), 37r: «muchhas cosas» (→ «muchas»), 47r: «que lo non aprndistes» (→ «aprendistes»), 90v: «ala merçer dellas» (→ «merçed»), 96v: «si algunod ouiesen» (→ «algunos»), 98r: «lo fijos» (→ «los»), 98v: «Mugeres muchas de otra maneras» (→ «otras»), 114r: «fab(a)lron» (→ «fablaron»), 115v: «Desçeñi el espada» (→ «Desçeñir»), 115v: «qual quier dellas que sae» (→ «sea»), f° 119v: «estodeuiesen» (texte: «esto- / deuiesen fazer» → «esto deuiesen»), 160v: «o de su conçeio» (→ «conçeio»), 163v: «alg(ñ)[u]<n>a derecha razon» (→ «alg(ñ)[un]a», source: «algña»; possible aussi: «alg<u>na», comp. *passim*), 168v: «es fecho el eplazamiento» (→ «enplazamiento»), 169r: «enlos eplazamientos» (→ «enplazamientos»), 172r: «apratadamente» (→ «apartadamente»), 172v: «el epla-zado» (→ «enpla-zado»), 184r: «si algu(u)[n]d i(n)[u]lyzio» (→ «algund»), 184r: «testigos en el plyto» (→ «pleyto»), 184v: «que ouieren ya i(n)[u]lyzio afinado» (→ «ouieron»), 185v: «mi-entra biuiere e poder de su pad(re)» (→ «biuiere<n> en»), 194v: «los preuilgios» (→ «preuilegios»), 196r: «aganer cartas» (→ «aganar»), 200r: «Ley sessente & dos» (→ «sessenta»), 200r: «para paga las debdas» (→ «pagar»), 200r: «carta dela lendida» (→ «vendida»), 200r: «las otras cosas que son de suso dichos» (→ «dichas»), 211v: «& non por abreuiduras» (→ «abreuiaduras»), 238v: «Mas las seeruidunbres» (→ «seruidunbres»), 250v: «casam( )[i]ento» (→ «casamiento»), 272v: «si aaesçiese» (→ «acaesçiese»), 272v: «Ley quinto» (→ «quinta»), 276v: «pedir merçer» (→ «merçed»), 280r: «quieren respnder» (→ «responder»), 304r: «preguntar ni respnder» (→ «responder»), f° 352r: «& tan bien el obispo que esto- / (o)uiese de fazer» (Glos. «estouiese» → «esto ouiese»), f° 356r: «su estir» (→ «vestir»), 365v: «es dicha acisaçion» (→ «acusaçion»), 373r: «es cosa qu torna» (→ «que»), 377v: «presos e ençerraodas» (→ «ençerrados»), 385v: «algmna cosa mueble» (→ «alguna»), 396r: «en alg(n)[u]na calle» (→ «alguna»).

– *Tratado de los niños...* [B. de Gordonio?, trad.], Present.: «18 de abril de 01495» (→ «1495»), ff. 9v: «Capitulo .xxv. de la escoriacion delas piernas» (→ «escoriacion»; cf. f° 12r: «dela escoriacion delas piernas»), 10r: «seanmedianeras» (→ «sean medianeras»), 12v: «poco vale para aprouerchar» (→ «aprouerchar»).

– *Univ. vocab.* [A. de Palencia], ff. 8v: «Adoniu<m> [...] co<m>pone se de diametro catalectico» (→ «dimetro»), 12v: «sedixo alcides» (→ «se dixo»), 19r: «son enfermidades» (→ «enfermedades»), 20v: «enfermadad enla garganta» (→ «enfermedad»), 23v: «presente la enfermedad» (→ «enfermedad»), 32v: «Arrogare es apprpriar» (→ «appropriar»), 52v: «subrutilat daque fulgorem» (→ «datque»), 54r: «paareçe» (→ «pareçe»), 64r: «do se

apo(<n>)se[<n>]ta<n>» (→ «apo(<n>)se[n]ta<n>», source «apōsetā»), 64r: «p<ar>a» (→ «p[ar]a», source «pa»), 98v: «dela enfermdad» (→ «enfermedad»), 101v: «appelatur» (→ «appellatur»), 103r: «festo pmpeyo» (→ «pompeyo»), 109r: «Dens. del genero msculino» (→ «masculino»), 130r: «libro secun-ldo» (→ «secun-do»), 132r: «Emorogia. es enfermdad» (→ «enfermedad»), 136r: «la va con trembor» (→ «la vna con trembror»), 179r: «Geometria [...] el que con stabidora» (→ «sabidora»), 179r: «& los egiptianos» (→ «egiptianos»), 180r: «gemere faciat aum qui gerat» (→ «eum»), 193r: «Hiems [...] se escrue» (→ «escrue»), 195v: «Dicta autem gree historia» (→ «gree»), 246v: «afirma<n> algu(n)[u]<n>os» (→ «algu<n>os»), 221v: «falco: enfermo» (→ «flaco»), 233r: «amenti equalitate» (→ «equalitate»), 501r: «Idem etiam inferum ap-pallatur» (→ «ap-pellatur»), 294r: «tubal sue inuentor» (→ «fue»), 317v: «la autroridad» (→ «autoridad»), 331r: «orthogra-pho es quien a derechas scriue» (→ «escrue»), 372r: «para el epi-rito» (→ «el spi-rito»), 399r: «Logica diuiditur in dialecticam epididicam. & sophisticam» (→ «epidicticam»; cf. *ibid.*, f° 136r: «Epidictica. scientia demonstratiua philosophia»), 399r: «doctri-nalem» (→ «doctrinalem»), 414r: «despues del termor» (→ «temor»), 417v: «pa<ar>tes» (→ «p<ar>tes»), 417v: «idet amissam sapientiam» (→ «idest»), 448r: «asee-ianza dela espina» (→ «a seme-ianza»), 458v: «segud dizen algu-nos» (→ «segund»), 480r: «& hay los compuesto & retransitios» (→ «compuestos»), 503r: «Tongillatim. idet» (→ «idest»), 505r: «see tiran a ma-no» (→ «se»), 534v: «secun-sum pronomina» (→ «secun-dum»), 548r: «se se[']a(<n>)lan» (source: «se senālan → «se sen[~]a(~)lan»), 538r: «sedund suso diximos» (→ «segund»).

– *Valerio*... [D. Rz. de Almela], ff. 100r: «la caualleria del Rey nauchdono-sor» (→ «nabuchdono-sor»), 158r: «rasgo sus vertiduras» (→ «vestiduras»).

– *Visión delectable* [A. de la Torre], ff. 3r: «fua criado» (→ «fue»), 3r: «los sontidos forynsecos» (→ «sentidos»), «los mordederes embidiosos» (→ «mordedores»), «la abitalbe parte» (→ «abitabile»), «aupa entrado» (→ «auya»), «et defecus leuiores» (→ «defectus»), 3v: «E ser: agrupdas» (→ «arguydas»), «delos clarissimos bates» (→ «clarissimos»), «aunque ercoles non suele fallado» (→ «non fuese fallado»), «de su propero Reyno» (→ «prospero»), «las gentes tenin» (→ «tenian»), «et exedientes las vulgares» (→ «excedientes»), «lo proçedian de los pechos» (→ «le»), 25r: «Antes lo consçe» (→ «conosçe»), 63r: «to vi-essen cargo» (*id.* source; → «to( )ui-essen»), 68r: «es muy seblante» (→ «senblante»).

– *Vocabulario eclesiástico* [R. Fernández de Santaella], Present.: «febrereo de 1499» (→ «febrero»), «la misma lengua La» (→ «lengua. La»), f° 23r: «algunlogar» (→ «algun logar»).

Bien que ces erreurs ne représentent, ensemble, qu'une longueur de texte négligeable par rapport à l'énorme masse de texte mémorisé, elles suffisent à jeter le doute dans l'esprit de l'utilisateur sur la qualité de l'information première. En fait, à chaque fois qu'on interroge la base on découvre de nouvelles erreurs de transcription; il n'est pas toujours possible d'interrompre son travail pour en prendre note. Mais si on les rapproche des innombrables occurrences des formes agglutinées, il est difficile d'être jamais sûr des résultats obtenus. Cette incertitude est aggravée, dans la version examinée, par certaines imperfections du logiciel – dont l'amélioration est annoncée pour la prochaine édition – et un certain manque d'harmonisation entre les opérations de saisie et

de traitement. On peut distinguer, pour clarifier l'exposé, les problèmes concernant la représentation de phénomènes particuliers, les problèmes d'affichage et les problèmes liés au Glossaire. En réalité, dans le fonctionnement concret de la base ces problèmes, comme on le verra, interfèrent sans cesse.

### 3.3. Utilisation de l'information

#### 3.3.1. Représentation de quelques phénomènes particuliers

On lit dans le Manuel, p. B5: «<sup>^</sup> circunflejo únicamente puede aparecer entre corchetes indicando intercalación del copista o entre paréntesis indicando supresión del copista». On rencontre pourtant fréquemment l'accent circonflexe entre deux caractères dans le *Doctrinal de caballeros* (d'A. de Cartagena), sans cette valeur particulière que lui ont attribuée les conventions de saisie: ex. «sen<sup>^</sup>or», p. 4, «duen<sup>^</sup>a», p. 76. Vérification faite, on s'aperçoit que ces séquences sont simplement de mauvaises transcriptions de la source, où figurent «señor» et «dueña». Ce qui est fâcheux, c'est, d'une part, que le Glossaire ne fournisse que «senor», «duena» (une question sur «sen<sup>^</sup>or» donne, de toute façon, la réponse 0), d'autre part, que le transcripteur ait généralisé son système propre, de sorte que «señor», «señora» et leurs dérivés présentent «-n-» dans le Glossaire et «-n<sup>^</sup>-» tout au long du texte, soit dans 57 pages pour «señor», dans 4 pour «señores», 1 pour «señora», 1 pour «señoria», 1 pour «señorial» (erroné pour «señoril»), 4 pour «señorío»; à quoi il faut ajouter les formes s'écrivant normalement avec ñ (2 «compaña», 2 «compañera», 1 «conpaña», 5 «conpañia», 1 «estrañas», 4 «tamaña», 3 «tamaño», etc.).

La barre qu'emploie Nebrija au-dessus de *x* et de *ch* dans sa *Gramática* pour indiquer la valeur phonétique qu'avaient ces graphèmes en castillan est représentée, sur le disque, par la séquence «(~)», ex. «lebrix(~)a», «circunflex(~)o», «monarch(~)ia», «much(~)edumbre». Si l'on veut rechercher ces formes à partir du Glossaire, il faut négliger les parenthèses, ex. repérer d'abord «lebrix~a». Mais le codage semble avoir éliminé la possibilité de recourir au lemme comme moyen d'interrogation. Si l'on met à part «lebrix~a», qui n'a pas été lemmatisé, raison pour laquelle la réponse à «=nebrija» est «0», les trois autres formes, qui ont été lemmatisées, devraient apparaître dans la liste des formes de chaque vocable; or la réponse est toujours «0». Tout se passe donc comme si les formes en «-x(~)-» et en «-ch(~)-» n'existaient pas dans la base. La séquence de codage «(~)», du reste, ne figure ni dans la liste des signes spéciaux présentée dans le Manuel, pp. B4-5, ni dans celle du disque (Normes de transcription). Comme les parenthèses servent, par ailleurs, à représenter la suppression d'un élément, il est probable que le traitement de ces dernières a perturbé le traitement des premières (voir encore, plus bas, le problème posé par les parenthèses renfermant un blanc).

Enfin, l'astérisque a été retenu pour signaler la reconstruction d'un texte illisible (Manuel, p. B5). On le rencontre pourtant dans des passages où la source est claire, ex. *Siete Partidas*, f° 194v: «[\*E] los preuilgios» [*sic*] (l'incunable porte simplement «Z»), «han fuerça [\*de le]y», «[\*ca pr]e-uilegio» (source: «Ca»), «[\*asi co]m<m>o». La variante «pr]e-uilegio» est correctement reconnue dans le texte, mais non acceptée comme forme d'interrogation: elle est neutralisée dans «preuilegio», de même que toutes les variantes tronquées. Quant à «[\*E]», la convention de transcription pour «Z» étant «&», le remplacement de ce signe par une séquence de restitution ne trouve aucune explication satisfaisante; si la version utilisée par le transcripateur était moins nette que l'image du texte présentée dans la base, la leçon aurait dû être «[\*&]»; mais le choix de [\*] pour exprimer une restitution est un choix malheureux puisque [\*] constitue, d'un autre côté, un symbole de substitution dans une question exprimée à l'aide d'une forme incomplète (cf. ci-dessus). En tout cas, il serait indispensable que l'utilisateur puisse, en fonction de ses préoccupations, retrouver toutes les séquences qui ont été plus ou moins reconstruites dans les passages réputés illisibles. D'une manière générale, que toutes les formes possédant une caractéristique particulière soient considérées comme des variantes distinctes et recueillies comme telles dans le Glossaire.

A l'inverse il manque dans le système un signe permettant de reconnaître les terminaisons, qui, dans la base actuelle, sont traitées comme des formes à part entière. Ainsi, «ta», présent dans 429 pages, correspond, entre autres, à la terminaison de participes ou adjectifs, dont le féminin est noté de façon abrégée, dans les dictionnaires ou les grammaires, à la suite de la forme masculine, citée intégralement. Ex., «abominatus. ta. tum.» (*Univ. vocab.*, p. 5, f° 3r). Faute d'un codage adéquat, le Glossaire se gonfle d'un nombre élevé de formes fictives, qui de surcroît ne correspondent pas toujours au même phénomène (cf. ci-dessous pour les formes éclatées): les noms, les adjectifs, les verbes sont susceptibles de recevoir un cortège d'abréviations. Sans vouloir être exhaustif, on relèvera «a» 7841 (il faudrait distinguer ici la préposition – latine et castillane –, le verbe, les abréviations du féminin, les erreurs de saisie), «ae» 559, «us» 406, «i» 1441, «ii» 207, «ij» 2178, «us» 406, «um» 635, «tus» 313, «tum» 447, «ctus» 25, «cta» 69, «ctum» 102, «cte» 5, «sus» 4723, «sa» 197, «sum» 406, «aris» 447, «atis» 421, «itis» 71; «as» 788, «aui» 506, «is» 689, «iui» 77, «dus» 6, «da» 1100, «dum» 234, «dam» 6, «dae» 4, «di» 229<sup>(31)</sup>. D'autres exemples de cas particuliers proviennent de la syllabation de certaines formes:

(31) Quelques-unes de ces séquences peuvent être, comme «a», des mots autonomes («da», «di», «sus», «sum», «tus»). Mais pour chacune d'elles un certain nombre d'occurrences, qu'il n'est pas possible d'évaluer, correspondent à des abréviations.

cf. «an. to. nio.» (*Gramática cast.*, p. 109, f° 55r) et des mentions d'affixes dans les grammaires: cf. «ano» [«toledano»], «es» [«frances»], «eño» [«extremeño»] (*ibid.*, p. 61, f° 31r). Syllabes et morphèmes sont traités à nouveau comme des formes différentes.

### 3.3.2. Identification et affichage

Les fragments qui viennent d'être cités peuvent également résulter soit d'une erreur de saisie, soit d'une erreur d'algorithme. L'erreur de saisie la plus fréquente est l'oubli du trait d'union en fin de ligne dans les cas de troncation. Les programmes de traitement ne peuvent alors rétablir le lien entre l'élément final de la ligne supérieure et l'élément initial de la ligne inférieure. Ex. «ca / ta» (*Visión delectable*, p. 109, f° 55r), «dulçu / ra» (*ibid.*, p. 140, f° 70v). Ces deux formes ne seront pas prises en compte dans une recherche sur «cata» ou «dulçura», puisque seuls existent dans les textes et le Glossaire leurs éléments éclatés. En revanche, l'une des occurrences de «ca» sera celle de «ca-ta», et l'une de celle de «ra» (qui en compte 152) celle de «dulçura» («dulçu», forme fictive, précédant «dulçura» dans l'index). On ne peut envisager d'autre remède qu'une révision plus attentive des textes avant leur introduction dans la base. Comme fréquemment aucun signe de liaison n'est employé dans les incunables, le plus simple est de recourir au signe conventionnel (les oublis sont extrêmement nombreux). Mais il faut, dans ce cas, que l'algorithme prenne en compte toutes les césures. Or on observe dans ce domaine un fonctionnement qui paraît plutôt aléatoire. Ainsi, divers «ento» et «to» procèdent de césures effectives mais négligées: «mouimi-ento» (*De proprietatibus*, p. 114, f° 57v), «seguimi-ento» (*Cuadernos...*, p. 12, f° 6v), «to-dos» (*Dict. lat.-hisp.*, p. 2, f° 1v), «de to-} / do» (*Capítulos de gobernadores*, p. 13, ff. 6v-7r; marque de fin de rubrique plus changement de page), «del qual text-to» (*Cura de la piedra*, p. 61, f° 31r); la conséquence de ce dysfonctionnement est la création de nouvelles formes fictives «mouimi», «seguimi», «tex», etc. Mais parfois une forme est correctement identifiée dans le Glossaire, par ex. «enuieçimien-to» (*Univ. vocab.*, p. 1046, f° 524v) et pourtant susceptible d'être analysée deux fois: l'une sous l'entrée «enuieçimiento» (l'affichage est bon), l'autre sous l'entrée «to» (et l'affichage se limite à la finale du mot).

Il semble, de même, que l'algorithme de traitement ne soit pas entièrement satisfaisant quand une forme comprend des parenthèses ou lorsque deux formes, collées dans la source, sont séparées par des crochets. Les parenthèses encadrent normalement le caractère qui doit être supprimé. Ce caractère peut être un blanc, quand il s'agit de regrouper deux éléments d'une même forme séparés par un espace: «segu( )nd» (*Siete partidas*, p. 674, f° 337v; source:

«segu nd»), «muche( )dumbre» (*Univ. vocab.*, p. 912, f° 457v; source: «muche dumbre»). Dans les deux cas, le logiciel, ignorant les parenthèses, éclate les formes et le Glossaire s'enrichit de «segu», «nd», «muche» et «dumbre» (à l'affichage: «segu(», «)nd», «muche(», «)dumbre»). De même on ne peut atteindre «pad( )re» à partir de «padre» dans les *Siete partidas* (p. 370, f° 185v): le Glossaire a recueilli «pad» et «re». On pourrait multiplier les exemples. Parfois l'effacement porte sur un autre caractère que le blanc ou sur plusieurs caractères. Le résultat est le même: «es(&)-tonçe» (*Siete partidas*, p. 235, f° 118r) ou «co-(co)stunbre» (*ibid.*, p. 8, f° 4v) éclatent en «es», «tonçe», «co», une autre bizarrerie s'ajoutant aux précédentes dans le dernier cas: le Glossaire comporte et «costunbre» et «cocostunbre» comme doublons pour retrouver la forme corrigée; mais, paradoxalement, lorsque la séquence s'affiche sur fond rouge, ce n'est pas «co-( )stunbre» qui est souligné, mais «(co)stunbre»; autrement dit, l'élément correct est évacué en faveur de l'élément à effacer...

Des perturbations inverses surgissent lorsque les marques «[]» sont employées pour séparer deux mots collés dans la source: s'il manque un espace entre les deux crochets, les deux mots continuent à n'en former qu'un, les crochets n'ayant pas en eux-mêmes valeur de séparateurs puisqu'ils servent souvent à noter qu'un caractère est ajouté et doit être pris en compte. C'est pourquoi «se[]faze» (*Nobiliario vero*, p. 160, f° 82r), «por[]trazon» (*ibid.*), «sea[]fina» (*ibid.*, p. 75, f° 38v), qui figurent dans le Glossaire comme formes d'un seul tenant, ne peuvent être retrouvées et affichées que globalement<sup>(32)</sup>. Il en est de même, parfois, quand parenthèses et crochets apparaissent dans une séquence ininterrompue: «segund[]a()costunbran» (*Siete partidas*, p. 80, f° 80v), «aquel[]o()mezillo» (*ibid.*, p. 51, f° 26r); aucun des quatre éléments ne sera identifié si on lance une question sur «segund», «acostunbran», «aquel» ou «omezillo».

Ces problèmes ne sont certainement pas difficiles à résoudre. Il suffirait d'harmoniser représentation et programmation; en décidant, par exemple, que les parenthèses doivent renfermer au minimum un caractère, même si ce caractère est un blanc<sup>(33)</sup>, et que le contenu – à effacer – doit être testé; que les crochets également enferment, soit une lettre soit un blanc, que dans ce dernier cas, la séquence qui suit est à interpréter comme une nouvelle

(32) Le résultat est comparable au résultat obtenu si aucune marque ne sépare plusieurs formes: comp. «delas[]partes» (*Leyes hechas...*, p. 5, f° 3r) et «alasvezes» (*Siete Partidas*, p. 67, f° 34r, p. 72, f° 36v) ou «delasper-sonas» (*Doctrinal de los caballeros*, p. 48, f° 24v); chaque séquence figure dans le Glossaire comme une forme distincte.

(33) A moins que l'on n'estime que si le caractère à effacer est un espace, les parenthèses peuvent se suivre immédiatement...

forme<sup>(34)</sup>. La cohérence voudrait aussi que lorsqu'on ajoute un tiret à une forme qui n'en comporte pas dans la source (en cas de troncature en fin de ligne), on le signale entre crochets, comme on le fait (par convention) pour tous les autres caractères ajoutés: «tab[-]la», non «tab-la» (*Nobiliario vero*, p. 2), «contie[-]ne», non «contie-ne» (*ibid.*) – car certains incunables signalent la liaison dans la graphie.

L'affichage pose, d'autre part, quelques problèmes spécifiques, mais en relation étroite avec la logique de l'algorithme. Il arrive, par ex., bien que le phénomène se produise rarement, que le cadre rouge soit légèrement décalé par rapport à la forme soulignée: cf. «infançia» (*Visión delectable*, p. 32, *Univ. vocab.*, p. 142), «dios» (*Visión d.*, pp. 14, 19), où le tracé droit du dernier caractère, en blanc, est doublé en noir hors du rectangle rouge. Ailleurs le décalage a la largeur d'un caractère, si bien que le blanc précédant la forme disparaît et que le dernier caractère est doublé en noir hors du rectangle rouge: «dedios» pour «de dios» (*Visión d.*, p. 43), «dedi-i-os» pour «de di-os» (*ibid.*, pp. 48, 51), «delonbree» pour «del onbre» (*Siete partidas*, p. 154), «delon-bre» (*ibid.*, p. 9: le «l» est presque écrasé par le «o»), «alon-bre» (*ibid.*, p. 4), «lamugerr» (*De proprietatibus*, pp. 112, 276, 400). Il peut même atteindre un espace de deux caractères («dasiria/a/» pour «de asiria/», *De proprietatibus*, p. 365). Par endroits, enfin, des marques extérieures à la séquence soulignée, comme l'accolade de rubrique ou des crochets, sont intégrés au rectangle de soulignage: «}Díos» (*Visión delectable*, p. 103), «[^onbre]]» (*Introd. lat.*, p. 19).

Plus gênante est l'absence d'affichage. C'est le cas, évidemment, pour les formes éclatées (cf. ci-dessus); mais quand on fait confiance au système, il est normal d'attendre toutes les occurrences de «conçei» si l'on pose une question sur ce mot. Or «conç( )ei» (*Siete Partidas*, p. 383) sera ignoré, ses deux éléments étant séparés dans l'analyse (et le Glossaire ne connaît que «conç» et «ei»). Mais même quand la séquence ne comporte aucune coupure l'affichage peut ne pas avoir lieu; ex. «cadavez» (*Ordenanzas sobre los escribanos*, p. 11) ouvre bien la page où se trouve le syntagme; mais ce dernier, transcrit «cada[^ ]vez», n'est pas souligné. La marque de transcription mentionne ici une interpolation de copiste (source: «cadalvez»), que le logiciel ne traite pas jusqu'au bout. L'ennui c'est que la séquence est également retrouvée si l'on part de «cada» (elle change un peu à l'affichage: «cada[^ ^]vez», où seul «cada[^ ^» est affiché), mais non si l'on part de «vez». Autre exemple, qui met aussi en jeu des marques de codage: «iota» n'apparaît pas en rouge à l'écran

(34) A moins que l'on ne considère que lorsque deux crochets se suivent ils équivalent à l'ajout d'un séparateur entre formes à distinguer...

dans la page 4 des *Cuadernos de las leyes nuevas*, bien que la forme existe et que la page soit ouverte à partir du Glossaire («iota»). La forme se trouve dans le texte entre crochets: «{GL. {LAT. [^2iota]}}»; elle correspond à une glose marginale, en latin, et se voit donc traitée comme un ajout, signalé par deux rubriques enchâssées. Enfin, l'emploi de l'accent grave pour coder la présence d'un caractère suscrit entraîne la suppression du soulignage, alors que la page s'affiche: cf. «quales» (*Siete partidas*, pp. 25, 27, 87, 88, 98, etc.: texte «qua`les»). Il semble que ce soit toujours le cas, comp. dans les *Siete partidas* «qui`stiones» (p. 2), «pri`mera» (*ibid.*), «sacra`mentos» (p. 9).

L'affichage n'est pas obtenu non plus si, par hasard ou volontairement, on interroge successivement deux formes qui se situent dans la même page. Quand on interroge la seconde, la page reste ouverte et seule la première forme est soulignée. Par exemple, après examen des premières occurrences de «sacramentos» dans les *Siete partidas*, p. 1, on interroge «trinidad», qui se rencontre également p. 1: l'affichage ne se modifie pas; si l'on veut qu'il fasse apparaître en rouge la forme suivante, il faut, avant d'appeler cette dernière, afficher la page suivante de «sacramentos» ou des *Partidas*. Cette anomalie empêche, par exemple, de vérifier, par rapport à «sacramentos», les contextes de «trinidad» dans la même page 1.

Mais l'absence de soulignage peut alterner avec un excès d'affichage. C'est ainsi que lorsqu'on demande une séquence de deux formes (ex. «primero que»), outre les occurrences du modèle, toutes les occurrences isolées et de «primero» et de «que» sont soulignées, ce qui perturbe la consultation et risque de masquer l'une des occurrences pertinentes si l'une ou l'autre des deux formes revient avec une haute fréquence (ce qui se passe pour «que»). On aboutit même à des réponses paradoxales; par ex., à une question sur «bueno era» on lira comme réponse, p. 130 de la *Visión delectable*, «era bueno»: la séquence d'origine n'a pas été retournée, mais chacun de ses termes apparaissant dans le contexte cité et chacun étant retenu comme réponse correcte (indépendamment de l'autre), la relation d'ordre se trouve neutralisée. L'affichage présente le même défaut quand on utilise l'opérateur [.y] («primero .y que»); autrement dit, au niveau de l'affichage le fonctionnement induit par l'opérateur [.y] est celui qui devrait correspondre exclusivement à [.o] («primero .o que»)<sup>(35)</sup>.

(35) La différence entre les nombres (254 avec [.y] dans les *Siete partidas*, 824 avec [.o]) prouve que le fonctionnement logique du programme est correct; mais l'affichage, qui doit être déclenché par deux opérations successives, n'est pas cohérent, pour [.y], avec la question posée.



### 3.3.3. Glossaire

Le Glossaire est une pièce maîtresse du système puisqu'il sert très souvent à lancer une recherche. C'est lui, d'autre part, qui enregistre toutes les erreurs d'analyse. En fait, il faut distinguer le Glossaire des Formes (ou GF) et le Glossaire des Lemmes (ou GL, qui ne regroupe que les formes lemmatisées et permet, sous un lemme, d'afficher la liste des formes lemmatisées correspondant à ce lemme).

GF, comme on l'a indiqué plus haut, renferme plus de 47000 unités (Manuel, p. 26). Il est présenté en majuscules (et seules les majuscules sont utilisées dans les questions). L'ordre suivi est l'ordre alphabétique, à quelques exceptions près. Tout d'abord, donc, il ne tient pas compte des majuscules (si bien que la même forme «dios» sert à retrouver les occurrences de «dios» et de «Dios») – faut-il rappeler que l'écriture médiévale ne suivait pas en la matière les normes actuelles? Mais un chercheur peut désirer savoir comment sont utilisées les majuscules dans un texte. La base ne lui en donne pas le moyen pour l'instant. Or le problème n'est pas trivial. Certaines formes identiques peuvent être des homomorphes de lemmes différents, l'un nom commun, l'autre nom propre. Ce dernier, dans la perspective rigoureuse des auteurs eux-mêmes, devrait être marqué. La graphie permettrait alors, malgré les majuscules, de distinguer «domingo» et «Domingo», comme le fait Nebrija dans son *Dict. hisp.-lat.*, p. 104. Rien ne marque non plus les mots étrangers, qui abondent dans la base, notamment les mots latins, au point que certains passages de l'index ne comprennent que des mots latins: «nocendo», «nocendum», «nocendus», «nocens», «nocente», «nocentes», «nocentia», «noceo», «nocere», «noces», «nocet», «nocetur». Les homomorphes du latin et du castillan ou d'une autre langue sont donc confondus (cf. ci-dessus).

D'un autre côté, l'ordre des caractères n'est pas toujours celui qu'on attendrait (sans se préoccuper des codes informatiques): ç suit z (ex. «morzillo» → «morçielago»), ñ suit ç («caçurro» → «caña»). Mais ce qui entraîne le plus de distorsions dans l'ordre des formes est le double parti pris de ne pas supprimer les caractères qui, dans le texte, sont mis entre parenthèses et à la fois d'effacer les parenthèses. Comme les crochets subissent le même sort, GF regorge de formes surprenantes, parfois indéchiffrables: «nnningun», «caoen» 'caen' (dans les textes: «(n)N(n)ingun» [*Ordenanzas reales*, p. 468; sans erreur de transcription: «(n)Ningun»], «ca(o)[e]n» [*Univ. vocab.*, p. 812]). Étant donné que GL permet de rapprocher les formes qui se trouvent éloignées par leur graphie, on renouvellera le vœu de voir apparaître dans le Glossaire complet toutes les formes avec leurs signes particuliers, ce qui, entre autres choses, concentrera l'attention sur les formes inattendues qui, elles, ne sont que des

erreurs de transcription. Par ex., les variantes de «ninguna» sont, d'après GF: «ninguna», «ningnuna», «ningua», «ninguana», «ningunda», «ninguna», «ningunna», «ninguuna», «niunguna», «njnguna», «nninguna», «nnynguna», «nyn-guna» (soit 13, sauf oubli). Plusieurs sont des formes modifiées: «ning(n)[u]na» (*Siete partidas*, pp. 33, 135, 791, 811), «ningu(u)[n]a» (*Ordenanzas reales*, p. 456; *Partidas*, p. 95), «ni(u)[n]guna» (*Partidas*, p. 546), «ni(u)[n]guna» (*Partidas*, p. 190), «(n)Ninguna» (*Ordenanzas*, pp. 463, 495), «(n)NYnguna» (*ibid.*, p. 123). Parmi les autres on relève des erreurs diverses: «ninguna» → «ningu-na» (*Visión delectable*, p. 54), → «ni[n]guna» (*Ordenanzas reales*, p. 103), → «ni[n]guna» (*Prognostica*, p. 2); «ningua» → «ninguna» (*Nobiliario vero*, p. 110); «ninguana» → «ninguna» (*Visión del.*, pp. 13, 51); «ningunda» → «ninguna» (*Partidas*, p. 185); «ningunna» → «ninguna» (*De proprietatibus*, p. 47<sup>(36)</sup>, *Crón. pop. del Cid*, p. 201). Les formes présentes dans GF, si l'on tenait compte des variations graphiques et des rectifications, seraient: «ning(n)[u]na», «ninguna», «ninguna», «ningu-na», «ninguna», «ni[n]guna», «ni[n]guna», «ningu(u)[n]a», «ni(u)[n]guna», «ni(u)[n]guna», «njnguna», «(n)Ninguna», «(n)NYnguna» (13 formes). La liste n'en serait pas allongée et l'information gagnerait en précision. A l'utilisateur, selon ses besoins, d'utiliser ces formes, soit en fonction de leur graphie, soit en fonction de leur phonétique.

Il serait d'autant plus nécessaire de disposer de toutes les variantes dans GF qu'elles ne se rencontrent pas toutes dans GL, sous le vocable *ninguno*: manquent à la liste «ningua», «ningunda», «ningunna», «ninguuna», «niunguna», «nninguna», «nnynguna» (la moitié des formes). Quelle qu'en soit la raison, la différence entre les deux glossaires est difficile à justifier.

D'une manière générale, GL ne retient que les formes qui ont été lemmatisées. Ce qui implique, *a contrario*, que toutes les formes qui s'y trouvent l'ont été. Si l'on compare les deux listes initiales, on constate, en effet, que tout le début de GF a disparu: «256v» - «@otros» (18 formes)<sup>(37)</sup>. GL commence donc à «a» et très vite on relève de nouvelles absences; après «aa»: «aaantigua», «aabad», «aabemafa», ce qui ne laisse pas de surprendre, puisqu'il existe un lemme *antiguo* et un lemme *abad*. Mais les formes «aaantigua» et «aabad» n'apparaissent pas sous ces lemmes. Pourquoi les a-t-on laissées dans GF? De

(36) A signaler: décalage d'un caractère vers la gauche de la séquence soulignée.

(37) Ces 18 formes n'ont pas leur place dans GF: elles proviennent, les unes de remarques du transcripteur (qui ne devraient pas être mêlées au texte lors du traitement), les autres d'erreurs: «2onzas», «4peipeque». Les formes non lemmatisées auraient pu servir aux auteurs à vérifier leur vocabulaire. Elles les auraient grandement aidés à corriger leurs textes.

fait ce sont des erreurs<sup>(38)</sup>, mais ce n'est pas en tant qu'erreurs qu'elles ont été supprimées. On observe dans GL un phénomène absolument symétrique: entre «aa» et «aacaesce» on rencontre «aabruz»; si l'on clique sur «aabruz», le vocable *aabruz* apparaît au-dessus du tableau d'affichage inférieur et, dans le tableau, les deux formes correspondantes, «aabruz» et «aambruz»; mais si l'on interroge sur ces formes, la réponse est obstinément «0»; dans le tableau inférieur de GF elles n'apparaissent pas davantage. Faute de pouvoir les identifier par ailleurs, on en est donc réduit à considérer le vocable et les deux formes comme des références vides, parasites. Il est aisé de renouveler cette observation: «aabruz» dans GL est suivi par «acaesce» et «acaesçe», mais dans GF seule figure la forme «acaesçe»; la forme «aacatad» de GF ne se retrouve pas dans GL, qui pourtant comporte bien le vocable *acatar*. Les deux glossaires semblent refléter deux étapes différentes dans le traitement du vocabulaire.

Mais l'examen de GL et la comparaison avec GF soulève à nouveau avec une extrême acuité le problème de la méthode et des critères retenus pour la lemmatisation. Dès le début de GL on découvre que «contece», «contecen», etc. ont été lemmatisées sous *acontecer*, mais que «abaxa» l'a été sous *bajar* (alors que «abaxamiento» l'a été sous *abajamiento*)... A supposer que les variantes simples et les variantes préfixées soient employées de façon quasiment aléatoire (ce qui n'est pas le cas en général, leurs effectifs en font foi), on ne voit pas au nom de quel critère on choisirait comme lemme tantôt la forme préfixée, tantôt la forme simple; encore moins pourquoi, dans une même famille lexicale, on modifierait ce choix quand on passe du verbe au nom dérivé du verbe. Pourtant, ce ne sont là que broutilles par rapport à d'autres choix. On se demande vraiment quel jeu de hasard a pu classer les formes latines «abduce», «abducere» sous *aducir* (ce qui double la bévue morphologique d'un contresens sémantique), «asna», «asnas», «asno», «asnos» sous *asnar*, «abellana» sous *avellanar*, «baxo» [le plus souvent adj. ou adv., ou segment de «a baxo», «de baxo»] sous *bajar* (mais cf. ci-dessus «estas» [démonstr.] sous *estar*), «creades» sous *crear* (malgré son unique occurrence, non analysée: «non creades a malos», *Valerio*, p. 177 → *creer*), «aseguir» sous *conseguir* («no quiso venir aseguir el pleito», *Partidas*, p. 434; «aque venga aseguir / se acusacion», *ibid.*, p. 734 → «a[]seguir»), «mi», «mia», «mio», etc. sous *miar*<sup>(39)</sup>...

(38) «aaantigua»: *Ordenanzas reales*, p. 256, texte: «(a)A(a)ntigua mente», source: «Aantigua» → «A(a)ntigua»; «aabad», *Crón. pop. del Cid*, p. 159, source: «abbad» → «abbad».

(39) *Miar*, représenté dans la liste des formes de GL, ne figure pas dans GF, donc dans la base. D'autres formes de GL sont ainsi absentes de GF: «asenta» s. v. *asentar*, «miados» s. v. *miar*, «midos», s. v. *medir*...

Trois exemples résumeront les aberrations d'un travail réalisé, dirait-on, sans recours au texte et dans la plus grande précipitation. Toutes les variantes de l'adverbe *así* ont été lemmatisées sous l'infinitif *asir* («asi», «assi», «assy», «asy», «ansi», «anssi», «anssy», «ansy», «asin», «aasi», «aasy»<sup>(40)</sup>). La forme «abien» a été rattachée au verbe *haber*, alors que la consultation de quelques exemples aurait montré que la plupart du temps «abien» devait être lu «a bien»<sup>(41)</sup>: «abien vista de los alcaldes» (*Partidas*, p. 145), «abien vista de onbre [sic] buenos» (*ibid.*, p. 148), «de lo leuar abien» (*ibid.*, p. 154), «& viene abien quando es verdadera» (*ibid.*, p. 166). Enfin, les adverbes en *-mente* ne sont lemmatisés sous la forme adverbiale que dans le cas où ils sont écrits dans les textes en un seul mot, cas le moins fréquent. GL, par voie de conséquence, abonde en formes d'adjectifs qui ne sont que le premier segment d'adverbes de manière et «mente» apparaît isolément dans 3477 pages<sup>(42)</sup>!

Quand un lemme existe, deux types d'erreur reviennent avec une particulière insistance: ou bien une forme manque, voire plusieurs, ou bien une pseudo-forme s'ajoute aux formes correctes. Une lacune provient souvent d'une graphie déviante (à cause du caractère non exhaustif de la lemmatisation) ou d'une mauvaise segmentation du texte numérisé: s. v. *absolución*, «asolucion» (*Tratado de las armas*, p. 5); s. v. *absoluto*, «asoluto» (*Es amor en quien...*, p. 8); s. v. *absolver*, «asoluieres» («asoluieres en la tierra», *Ceremonial de príncipes*, p. 4), «asoluiuo» («asoluiuo», *Crón. pop. del Cid*, p. 212); s. v. *asignar*; «asina» [PR3] (*De proprietatibus*, p. 404, *Cura de la piedra*, p. 57), «asinado» (*Ordenanzas reales*, p. 187), «asinar» (*ibid.*, p. 44)<sup>(43)</sup>; s. v. *asir*, «asirse» (*Nobiliario vero*, p. 100); s. v. *aspecto*, «aspetto» (transposant le lat. «aspectus»,

(40) Curieusement, dans le cadre inférieur de GL, les formes du lemme ne sont pas citées selon un ordre alphabétique uniforme, mais par séries juxtaposées procédant peut-être de tris successifs; ainsi «aa-» suit une autre série en «a-». Cette présentation ne facilite pas les vérifications sur les formes ni la globalisation des données. À noter que certaines des occurrences des formes citées comprennent [*a + sí*] et sont donc à lire «a[*]*si» (*Visión delectable*, p. 2, f° 1v, p. 42, f° 21v), «a[*]*ssi» (*ibid.*, p. 109, f° 54v, p. 123, f° 62r), etc.

(41) En fait, après une vérification complète, on sait que «abien», qui figure dans 21 pages, ne recouvre aucune occ. verbale. L'imparfait courant dans la base, à la 3<sup>e</sup> pers. du plur., est «auian» 446; «auien» ne se rencontre que dans 8 pages.

(42) Plusieurs cas d'espèce peuvent être relevés, ex. «abi-erta mente» (*Partidas*, p. 155), «abierta / mente» (*Visión delectable*, p. 145), «abierta o clara-mente» (*Vocab. ecle-siástico*, p. 267), «abierta & claramente» (*ibid.*, p. 286).

(43) Il est d'autant plus regrettable qu'«asina» ait été écarté de la lemmatisation qu'il s'agit d'un mot polyvalent: outre la forme du présent d'*asignar*, c'est un latinisme fréquent, fém. d'«asinus» (*Univ. vocab.*, pp. 99, 580; *Dict. hisp.-lat.*, p. 42; avec la valeur métaphorique de 'carne', *Univ. vocab.*, p. 69), et aussi un nom de ville, «ciudad de achaia» (*Dict. hisp.-lat.*, p. 42).

*Univ. vocab.*, pp. 479, 488, 503); s. v. *aspereza*, «asperesa» (*De proprietatibus*, pp. 479, 488, 503); s. v. *azotar*, «asotados» (*ibid.*, p. 602)<sup>(44)</sup>... Quant aux formes parasites qui viennent allonger les listes, la plupart sont, soit des erreurs (cf. les variantes d'«alguna» ci-dessus), soit des retombées de la non-séparation des formes agglutinées: s. v. *sabor*, «asabor» (*Partidas*, p. 290, → «a[sabor]»); s. v. *saludar*, «salud», «salut», «asalud» (*Partidas*, p. 409: «[venir] asalud de su anima»; *Doctrinal de los caballeros*, p. 124: «[las cosas] conplideras asalud de su anima» → «a[salud]»); s. v. *servidumbre*, «aseruidunbre» (*Partidas*, p. 473: «tornar aseruidunbre», p. 506: «tornasen aseruidunbre» → «a[seruidunbre]»)...

Un dernier problème mérite d'être évoqué: les formes étrangères n'ont pas été lemmatisées. On comprend que les auteurs, pour économiser du temps, et peut-être aussi faute de collaborateurs latinistes, aient restreint la lemmatisation au vocabulaire castillan. Mais ce choix n'est pas très heureux, étant donné que dans la conscience linguistique des lettrés du XV<sup>e</sup> s. le castillan et le latin (au moins) coexistaient et s'influençaient mutuellement. Par ailleurs, une meilleure connaissance du latin médiéval est une urgence pour les historiens du castillan s'ils veulent mieux comprendre les modèles linguistiques que proposaient Mena, Alonso de Palencia, ou Nebrija. La lemmatisation des mots étrangers pourrait servir également à bien d'autres historiens non linguistes. Le choix opéré est d'autant plus regrettable que nombre de formes sont homomorphes en latin et en castillan. Or une forme virtuellement castillane n'est pas nécessairement documentée, dans la base, en contexte castillan (cf. «noctiluca» 'luciérnaga'). D'un autre côté, les formes ambiguës qui ont été lemmatisées véhiculent à la fois des occurrences latines et des occurrences castillanes, puisque la transcription ne les distingue pas (ex. «aspera», cast. *passim*, lat. fém. *Univ. vocab.*, pp. 70, 132; lat. n. *ibid.*, pp. 160, 179, 233; «deliberatiua», cast. *Lilium*, pp. 232, 234, lat. fém. *ibid.*, p. 331, n. *ibid.*, p. 37). Lorsque ce cas se présente, l'utilisateur du disque n'a d'autre ressource qu'un contrôle systématique, occurrence par occurrence<sup>(45)</sup>.

#### 4. Bilan

De cet examen attentif d'Admyte1 il ressort que l'enthousiasme que suscite la possibilité d'interroger à tout moment et très rapidement une base de

(44) Il a été vérifié que ces formes ne résultaient pas d'erreurs de transcription.

(45) Parfois une forme latine se trouve égarée parmi les formes romanes, ex. «astris» s. v. *astro*, «matrimonium» s. v. *matrimoniar* (ici se superposent deux types de confusion), «meridianum» s. v. *meridiano*.

61 textes d'espagnol médiéval doit être tempéré par la prudence que conseille un ensemble impressionnant de défauts ou de négligences. Certes, si l'on a besoin surtout de témoignages sur les mots, on sera moins exigeant sur la fidélité des transcriptions: la base peut, en effet, proposer 50 exemples là où l'on en désirait 5; dans ce cas, peu importe que sur les 50 il y en ait 5 ou 10 d'erronés, ou qu'à côté des 50 il y en ait 10 d'oubliés: l'attente sera largement satisfaite. Mais si l'on tient à savoir exactement combien d'occurrences de la forme F contient la base, ou combien de formes du vocable V y sont attestées, alors on se heurte, en l'état actuel des choses, à une impossibilité d'autant plus irréductible qu'elle se confond avec une probabilité d'erreur qui, sans s'appliquer concrètement à chaque phénomène étudié, jette un doute sur chaque réponse et même, qui pis est, sur chaque absence de réponse. Car on n'est sûr que des réponses vérifiées sur les textes, non des réponses elles-mêmes; encore faut-il que le système ait fourni un objet de vérification; lorsque la réponse est «0», on n'est jamais certain que la forme cherchée n'est pas restée collée à une forme qui la précède, ou qu'elle n'a pas été disloquée par un passage à la ligne non marqué ou non correctement traité.

Malgré ces réserves, ces limites, Admyte1 est un prodigieux instrument de travail, car les *bonnes* réponses qu'il fournit n'auraient pu être obtenues sans lui; il suffit, pour s'en convaincre, de comparer la situation du chercheur avant sa parution et les possibilités actuelles. Si la présente étude critique a voulu attirer l'attention, comme il se doit, sur les faiblesses de l'ouvrage, elle s'est efforcée aussi (et en priorité) de montrer les qualités techniques de la base de données et tous les services qu'elle pouvait rendre. Ampleur de la documentation textuelle, simplicité et rapidité d'accès, prouesses du logiciel qui, en marche normale, identifie «ninguna» dans «ni<n>guna», «ningu<n>a», «nin-guna» ou «ni[n]guna», lisibilité des pages affichées (la taille des caractères est presque un peu grande), clarté du soulignage sur fond rouge, possibilité de recourir à tout instant à l'image de l'incunable, de la réduire, d'en choisir seulement une partie... Voilà des qualités extrêmement séduisantes, et qui sont le résultat d'un travail considérable.

C'est au vu de ces qualités que l'utilisateur se fait exigeant; dans un instrument de travail élaboré surtout à l'intention de spécialistes il n'y a pas de place pour les erreurs dites matérielles. Les fautes de transcription devraient tendre vers zéro et si les textes avaient été soigneusement révisés la plupart des fautes grossières qui ont été signalées auraient disparu. A ce sujet, plusieurs remarques. Il n'est pas possible de corriger une base; chaque texte aurait dû être mis au point avant son introduction dans la base. Mais pour mettre un texte au point, quand on dispose de l'informatique, il est indispensable de s'aider de l'informatique, ce qui ne paraît pas avoir été fait. L'étude minutieuse de

l'index des formes aurait tout de suite révélé un grand nombre d'erreurs ou d'incohérences. Pour le traitement de certains hapax, on aurait pu interroger d'autres membres de l'équipe susceptibles d'avoir rencontré les mêmes problèmes. Par ailleurs l'informatique permet d'opérer certaines vérifications de manière systématique, par exemple en recherchant, à partir d'une faute *x*, toutes les fautes similaires (tous les «aa», tous les «uu», tous les «(n)N(n)»...). Le seul moyen de réduire à la fois le temps de préparation et le nombre des erreurs est d'améliorer la méthode de travail. En un sens, les textes actuels n'auraient jamais dû obtenir *l'imprimatur*! Mais comme ils servent malgré tout, il appartient à la communauté des chercheurs de collaborer à leur mise au point progressive (tel est aussi le vœu des auteurs du disque). Cette étude a été rédigée dans cet esprit. Il appartient à tous également d'expérimenter le logiciel, le fonctionnement des outils de recherche, des opérateurs. Lorsqu'un problème est bien cerné il est possible, en général, de lui trouver une solution.

Admyte1 pose, cependant, quelques grands problèmes qui valent, plus ou moins, pour tous les CD-ROM réalisés jusqu'à ce jour. Dans la mesure où le disque compact sert de support à un nombre toujours croissant d'ouvrages de référence et d'instruments de recherche<sup>(46)</sup>, il est souhaitable que se développe conjointement une méthodologie du CD-ROM. L'expérience d'Admyte permet de formuler quelques principes qui vaudraient au moins pour tous les projets analogues.

En premier lieu, la matière même du disque appelle une réflexion préliminaire sur les objectifs: quel type de base, pour quel public, pour quels types de recherches? Il est probable que l'examen de ces questions aurait suggéré de réunir exclusivement des textes de la seconde moitié du XV<sup>e</sup> s. pour constituer un ensemble de textes liés en synchronie, de préférence à un choix hybride où le XIV<sup>e</sup> est totalement absent. Si le public visé est plutôt un public universitaire (dans sa diversité), et non le grand public, un corpus équilibré répondrait

---

(46) Sur les disques susceptibles d'intéresser le chercheur des Humanités, et plus particulièrement le médiéviste, cf., dans *Le Médiéviste et l'Ordinateur*: René PELLEN, «Les CD-ROM pour médiévistes: premiers éléments d'une discographie», n° 28, 1993, pp. 13-18; «Les CD-ROM pour médiévistes: première mise à jour», n° 30, 1994, pp. 40-42; Claire PANIJEL, «Documents électroniques pour latinistes», n° 33, 1996, pp. 49-51. A quoi l'on ajoutera, pour l'espagnol, les deux disques récemment publiés par le Hispanic Seminary of Medieval Studies: Lloyd KASTEN, John NITTI, Wilhelmina JONXIS-HENKEMANS, *The Electronic texts and concordances of the prose works of Alonso X, El Sabio* (1997); John NITTI, Lloyd KASTEN, *The Electronic texts and concordances of medieval Navarro-Aragonese manuscripts*, *id.* Ainsi que le *Diccionario de la Real Academia Española* (1995), le *Diccionario de uso del español* de M. MOLINER (1996) et, bientôt, le *DCECH* de COROMINAS-PASCUAL.

sans doute mieux à son attente qu'une juxtaposition de documents, même si plusieurs genres ont été retenus dans le disque actuel. Par corpus équilibré on pourrait entendre un choix de textes qui accorderait leur juste place à tous les types de sources, littéraires et non-littéraires, et qui éviterait de surreprésenter un genre plutôt qu'un autre. À cet égard, la richesse d'Admyte1 en textes juridiques, en récits de voyage, en traductions, en dictionnaires et encyclopédies mérite d'être soulignée: c'est la première fois que des textes comme les livres de médecine les plus utilisés au Moyen Âge ou les ouvrages linguistiques de Nebrija et d'Alonso de Palencia deviennent accessibles sans difficulté. La contrepartie (dans ce premier disque, mais d'autres pourront le compléter), c'est la trop faible représentation des textes littéraires (Rodríguez del Padrón, D. de San Pedro, *Celestina*), en particulier de la poésie (Mena, Santillana, Manrique) et du théâtre (Encina).

Une fois résolus les problèmes du public et du contenu, on ne peut éluder le problème essentiel des fonctionnalités du disque, des modes d'exploitation de l'information mémorisée. Si l'on opte (au moins dans un premier temps) pour des recherches reposant le plus souvent sur des mots, il s'ensuit que lors de la transcription et, plus tard, lors de la programmation, chaque mot doit être facilement identifiable, quelle que soit la complexité du contexte. Il n'est plus possible, dès lors, de laisser les mots soudés les uns aux autres ou des parties de mots disjointes (quel que soit le système matériel qu'on adopte pour représenter les opérations du transcripteur). De la rigueur avec laquelle ont été analysés et saisis les textes et de la cohérence entre les conventions de saisie et l'algorithme de traitement dépend la qualité des résultats. Mais pour apprécier cette qualité, encore faut-il se donner le temps de procéder à d'amples vérifications, non seulement dans les documents de sortie (index divers), mais au niveau des réponses fournies par le logiciel d'exploitation. On ne saurait imputer au logiciel des réponses aberrantes qui s'expliquent par les insuffisances ou les erreurs des textes (ainsi la présence dans GF de formes comme «subidaçAscensus» [*Dict. hisp.-lat.*, p. 42] → «subida / Ascensus»), ni aux textes les échecs dans l'identification des séquences dont seul le logiciel est responsable (ex. l'interprétation du trait d'union, tantôt comme un élément de liaison, tantôt comme un séparateur<sup>(47)</sup>). Peut-être conviendrait-il, pour tester le prototype, de recourir à des utilisateurs extérieurs à l'équipe, selon une pratique courante dans certaines disciplines. Avant la fabrication du disque défi-

---

(47) Il faudrait, pour en être sûr, connaître la réalité informatique des textes, un blanc en fin de ligne n'apparaissant pas à l'écran. Mais on pourrait citer d'autres cas-types, comme la non-séparation des formes isolées dans les textes par «[]» («la[]vsasen» [*Partidas*, p. 15]), le classement non-alphabétique des formes dans GL, etc.



nitif et la mise sur le marché, des retouches (éventuellement importantes) pourraient être apportées à tous les éléments vitaux de la base. On éviterait aussi, en s'entourant de toutes ces précautions, d'éditer un produit trop rapidement obsolète.

Pour en revenir à Admyte<sup>1</sup>, on résumera pour clore ce tour d'horizon diverses suggestions qui ont été faites au fil de l'analyse. Les premières porteraient sur les instruments de travail. Il serait bon que l'index des titres rappelle le nom de l'auteur et la date du texte (non de l'édition), que dans la Présentation de chaque titre figure une discussion sur la date (quand la date est incertaine), quelques renseignements sur l'auteur (permettant de le situer) et sur les particularités de la source utilisée; que des éléments de la Présentation puissent être recopiés dans le presse-papiers, puis dans un fichier. La longueur du texte en mots apparaîtrait au début de l'article (à droite). Dans le Glossaire des Formes serait mentionné, en face de chaque forme, l'effectif en occurrences, non en pages; toutes les marques de transcription seraient conservées, chaque forme étant citée telle qu'elle se rencontre dans les textes. La sélection d'un titre entraînerait la réduction du vocabulaire affiché dans GF au seul vocabulaire du texte sélectionné. Le Glossaire des Lemmes communiquerait l'effectif du lemme et donnerait la liste de toutes les formes répertoriées par ordre alphabétique strict, avec l'effectif de chacune. Si GL ne lemmatisait pas tout le vocabulaire de la base, chaque lemme regrouperait au moins toutes les formes fléchies qui lui correspondent. Il est clair qu'avec une lemmatisation partielle une comparaison exhaustive des formes lemmatisées et des formes non-lemmatisées reste le seul moyen de ne pas oublier des formes séparées par l'alphabet ou la graphie.

En ce qui concerne la lemmatisation proprement dite, dont on souhaiterait qu'elle soit intégrale, elle ne saurait s'effectuer dans l'abstrait: l'analyse détaillée des textes en est la condition *sine qua non*. Mais un tel travail est extrêmement long et exige, pour un seul texte de moyenne importance, des années d'effort – ou la constitution d'équipes nombreuses appliquant les mêmes conventions. Il est illusoire, en tout cas, de vouloir en faire l'économie ou de le confier à des «sous-traitants»: pour certaines recherches (par ex. les recherches lexico-sémantiques ou statistiques), l'utilité de la base dépend directement de la qualité de la lemmatisation. Quelle confiance accorder à un glossaire qui lemmatise «mi» sous *miar*, «abducere» [lat.] sous *aducir*, «maduros» sous *madurar*? Mais l'absence de codage pour distinguer les formes étrangères des formes castillanes interdira, tant qu'elle sera maintenue, la lemmatisation de *tous* les termes castillans et donc une connaissance précise des structures lexicales des textes.

Pour laisser à la consultation toute la liberté dont elle a besoin, l'utilisateur aimerait aussi pouvoir choisir le passage du texte qu'il veut consulter; sauter, selon le cas, de la première à la n<sup>ième</sup> occurrence et revenir à la première sans relancer la procédure d'interrogation; il préférerait, d'autre part, que l'affichage en contexte de la forme recherchée présente toujours la forme au milieu du texte cité, l'affichage de la page entière (ou d'un contexte plus large) demeurant optionnel; que la mention de folio suive, dans le titre courant, la mention de la page; il aurait enfin besoin de concordances (de type traditionnel) lui permettant de vérifier très rapidement les collocations du mot étudié, de repérer la phraséologie, les motifs, les particularités d'écriture. Car la comparaison implique la vision presque simultanée de contextes, ou pour le moins la possibilité de passer très rapidement d'un contexte à un autre, de revenir en arrière, etc. Le défilement successif de pages entières est l'un des inconvénients majeurs du système actuel d'Admyte, surtout lorsqu'on garde l'option image attachée au texte, ce qui ralentit considérablement le défilement des pages.

Peut-on ajouter que certaines fonctions usuelles des traitements ou éditeurs de texte les plus rudimentaires amélioreraient grandement la situation de l'utilisateur? Par exemple, la fonction Recherche, qui permet de retrouver dans un texte toute séquence incluant les codages employés, comme les parenthèses ou les crochets que GF et GL éliminent? Le logiciel d'Admyte a neutralisé les marques et les opérateurs symboliques ou logiques ne leur sont pas applicables. Tel chercheur pourrait cependant vouloir comparer dans un texte donné les variantes graphiques de «ninguna» («ninguna», «ninguna», etc.), vérifier les interventions «éditoriales» du transcripteur (cf. l'exemple de «alquile[r]es» cité en 3.1).

D'autres procédures d'exploration seraient également les bienvenues. Par exemple, des index fréquentiels et des index inverses, par ordre des formes, par ordre des vocables, pour la base entière ou pour un ou plusieurs textes donnés. Sans parler d'hyperliens qui autoriseraient les sauts hypertextes, notamment d'une forme aux passages qui la définissent, d'une occurrence au lemme correspondant dans GL<sup>(48)</sup>. Mais de telles procédures impliquent comme préalable un long travail d'analyse linguistique et de balisage: elles pourraient s'inscrire parmi les projets à moyen ou long terme. C'est, malgré tout, la mise au point de ces nouvelles procédures, associée à une structuration à plusieurs niveaux de l'information textuelle, qui permettra un jour

---

(48) Cf. les possibilités qu'offrent des dictionnaires comme le *DRAE* ou l'*OED*. Mais l'utilisation d'hyperliens est beaucoup plus aisée dans des dictionnaires que dans des textes aux structures floues.

au CD-ROM de modifier radicalement les conditions de la recherche<sup>(49)</sup>. Il a surtout transformé jusqu'ici les conditions matérielles de l'activité de recherche (accès permanent, sur un poste individuel, pour un coût relativement modeste, à des volumes d'information considérables; multiplication des sources disponibles; possibilité, par les réseaux, d'élargir cet accès...). En réalisant le saut qualitatif que suggère l'étude d'Admyte1, il pourrait modifier, plus profondément, le paysage intellectuel et méthodologique de la recherche. Il est en passe de devenir un nouvel usuel; c'est vrai. Mais pour prendre le relais des anciens, il doit innover. Aux concepteurs de prendre la mesure de cette exigence, en temps, en rigueur, en imagination...

CNRS, URA 998, C.É.S.C.M.  
Université de Poitiers.

René PELLEN

### Bibliographie

*ADMYTE1* [*Archivo digital de manuscritos y textos españoles*, vol. I], Madrid, Micronet, 1993, 1 CD-ROM [61 textes, publiés pour la plupart à la fin du XV<sup>e</sup> s.], 54 p., 3 append. A1-C5 [Abrév.: Admyte1].

*Bibliography of Old Spanish Texts (Literary texts, edition 2)*, comp. par Anthony CÁRDENAS, Jean Gilkinson MACKENZIE, John NITTI et Ellen ANDERSON, Madison, The Hispanic Seminary of Medieval Studies, 1977, X-128 p. (Bibliographical Series; 1.) [Abrév.: BOOST 77.]

*Concordances and texts of the royal scriptorium manuscripts of Alfonso X, El Sabio*, éd. de Lloyd KASTEN et John NITTI, Madison, The Hispanic Seminary of Medieval Studies, 1978, 42 microfiches.

*The Electronic texts and concordances of medieval Navarro-Aragonese manuscripts* [CD-ROM], prép. par John NITTI et Lloyd KASTEN, Madison, The Hispanic Seminary of Medieval Studies, 1997.

---

(49) La constitution de corpus médiévaux pose inévitablement le problème de la nature du texte avec lequel on veut travailler (cf. 3.1): la transcription doit-elle conserver la version de la source, sans en corriger le moindre détail?; doit-elle intégrer les résultats d'une étude éditoriale, c'est-à-dire critique, du texte? L'informatique devrait permettre, désormais, de ne plus mélanger les genres et d'offrir, d'un côté, une transcription stricte, d'un autre, une version critique – l'étendue et la nature des interventions étant définies par la philosophie du transcripateur en matière d'édition critique!... Pour les utilisateurs qui ne sont pas familiers de la langue médiévale, on pourrait concevoir une option «texte modernisé» qui faciliterait la lecture littéraire. La structuration de l'information des bases textuelles comprend, schématiquement: la mise en forme informatique du texte par rapport aux objectifs et aux modes de traitement retenus (texte représenté en machine [trm], distinct du simple [mrf] 'machine readable form'), des index (des formes, des lemmes), des index contextuels. Ces niveaux, fonctionnellement liés, seraient accessibles de façon autonome.

- The Electronic texts and concordances of the prose works of Alonso X, El Sabio* [CD-ROM], prép. par Lloyd KASTEN, John NITTI, Wilhelmina JONXIS-HENKEMANS, Madison, The Hispanic Seminary of Medieval Studies, 1997.
- FAULHABER, Charles B., MARCOS MARÍN, Francisco, «ADMYTE, Archivo digital de Manuscritos y Textos Españoles», *La Corónica*, vol. 18:2, 1990, pp. 131-145.
- FAULHABER, Charles B., «La Conservación y utilización de textos en el futuro inmediato: ADMYTE, el archivo digital de manuscritos y textos españoles», *Hispania*, t. 75, 1992, pp. 1009-1024.
- FAULHABER, Charles B., «An Up-date on ADMYTE for 1995», *La Corónica*, vol. 23:2, 1995, pp. 104-110.
- MACKENZIE, Jean Gilkinson, BURRUS, Victoria, *A Manual of manuscript transcription for the «Dictionary of the Old Spanish Language»*, 4<sup>e</sup> éd., Madison, The Hispanic Seminary of Medieval Studies, 1992, 6-176 p., 33 ill. [Trad. esp. d'Aurora MARTÍN DE SANTA OLLA SÁNCHEZ, *Manual de transcripción para el «Diccionario de Español Antiguo»*.]
- MARCOS MARÍN, Francisco, *Informática y humanidades*, Madrid, Gredos, 1994, 816 p. (Manuales.) [Sur ADMYTE, v. pp. 179-229.]
- MARCOS MARÍN, Francisco, FAULHABER, Charles B., «La Conservación y utilización de textos en el futuro inmediato: ADMYTE, el archivo digital de manuscritos y textos españoles», *Hispania*, t. 75, 1992, pp. 1009-1024.
- MICRONET, S. A., «ADMYTE», Madrid, [1992], 8 p.
- MOLINER, María, *Diccionario de uso del español*, éd. sur CD-ROM, Madrid, Gredos, 1996.
- Oxford English Dictionary* [CD-ROM], Oxford University Press, 1992 [version pour PC; 2<sup>e</sup> éd., pour Macintosh, 1993].
- PANJEL, Claire, «Documents électroniques pour latinistes», *Le Médiéviste et l'Ordinateur*, n° 33, 1996, pp. 49-51.
- PELLEN, René, «Les CD-ROM pour médiévistes: premiers éléments d'une discographie», *Le Médiéviste et l'Ordinateur*, n° 28, 1993, pp. 13-18.
- PELLEN, René, «Les CD-ROM pour médiévistes: première mise à jour», *Le Médiéviste et l'Ordinateur*, n° 30, 1994, pp. 40-42.
- REAL ACADEMIA ESPAÑOLA, *Diccionario de la lengua española*, 21<sup>e</sup> éd., éd. sur CD-ROM, Madrid, Espasa-Calpe, 1995.
- ROUDIL, Jean, «Index alphabétique des formes de la *Primera Crónica General*», *Cahiers de Linguistique Hispanique Médiévale*, n° 4 bis, 1979, pp. 205-365.

