

PHILOBIBLON (PENINSULE IBERIQUE : ESPAGNOL, CATALAN, GALICIEN ET PORTUGAIS)

Gemma Avenoza. Universitat de Barcelona – IRCVM¹

PhiloBiblon est une base de données consacrée à l'étude des sources primaires de la littérature médiévale dans la péninsule ibérique. Actuellement, il est formé par trois bases de données sur Internet, dédiées aux littératures médiévales ibériques:

BETA (*Bibliografía Española de Textos Antiguos*) sur les textes castillans (textes aussi en aragonais, léonais, navarroaragonais, etc.)

BITECA (*Bibliografía de Textos Catalans Antics*) sur les textes catalans

BITAGAP (*Bibliografía dos Textos Antigos Galegos e Portugueses*) sur les textes galiciens, portugais et l'ancienne koinè littéraire galicien-portugaise des chansonniers

BETA, BITECA et BITAGAP recueillent des descriptions des manuscrits et des impressions anciennes qui transmettent la littérature médiévale des traditions respectives, ainsi que des informations sur les œuvres, les auteurs, les traducteurs, les anciens propriétaires, les copistes, etc.

Le projet PhiloBiblon fusionne la recherche, que nous pourrions appeler traditionnelle, avec les nouvelles technologies de l'information. Le travail rigoureux des philologues, des codicologues et des historiens est enregistré dans une base de données et diffusé sur internet.

Philibiblon n'est pas un projet né lorsque le concept "Digital Humanities" a commencé à apparaître dans les forums de recherche. Au contraire. Nous travaillons sur cette base de données depuis de nombreuses années.

PhiloBiblon a ses origines dans les années soixante-dix du vingtième siècle et a été une initiative pionnière de ce qu'on appelle désormais «Humanités numériques». Il est apparu comme un complément au *Diccionario del español medieval* qui était préparé à Madison, parce que pour étudier le lexique médiéval, il fallait d'abord avoir accès aux sources

¹ Este trabajo se inscribe en las líneas de investigación de la Red de Excelencia FFI2015-69029-REDT y del proyecto FI2014-55537-C3-1-P financiados por Ministerio de Economía y Competitividad, con aportación de fondos FEDER.

originales, aux manuscrits et aux anciennes impressions. Et, à l'exception de la *Bibliografía* de José Simón Díaz, nous avons très peu d'informations sur la portée du corpus conservé de la littérature médiévale castillane.

John Nitti et les chercheurs qui ont commencé PhiloBiblon ont d'une part un immense fichier de manuscrits et, d'autre part, le défi posé par la nouvelle technologie.

Ils ont opté pour la technologie. Le fichier a été abandonné et ses informations ont été transformées en une base de données qui reproduit initialement la structure d'un fichier manuel. C'était une première étape, mais il était clair que ce n'était pas «tout mettre dans l'ordinateur et voilà». Le programme devrait répondre aux besoins du chercheur et évoluer avec leurs exigences et le rythme des changements technologiques.

Ce premier programme était une base de données simple avec sept champs d'extension limités à cent caractères chacun. Les chercheurs n'ont pas directement accès à la machine, mais ils ont corrigé les listes, ont ajouté des modifications et plus tard les techniciens les ont transférées sur l'ordinateur hôte. Il n'y avait pas encore d'ordinateurs personnels, nous parlons du moment où le plus moderne était l'utilisation de terminaux connectés à l'ordinateur hôte.

Mais tout cela est l'histoire. À la fin des années quatre-vingt du vingtième siècle, Charles Faulhaber a réuni le groupe de chercheurs qui dirige actuellement les bibliographies. Ce groupe de travail initial a été renouvelé et élargi avec l'incorporation de jeunes chercheurs.

C'est ainsi que PhiloBiblon a continué, dirigé par Charles Faulhaber et avec la collaboration technique de John May. La base de données a été modifiée à mesure que la technologie avançait. La suivante étape a été la transformation de la base de données d'origine avec une table unique en une base de données relationnelles exécutée sur DOS. Cette base de données comportait finalement dix tables et près de cent champs.

ANALYTIC

BIBLIOGRAPHY

BIOGRAPHY

COPIES

GEOGRAPHY

INSTITUTIONS

LIBRARIES

MS/ED

UNIFORM_TITLE

SUBJECT

À cette époque, les chercheurs travaillaient déjà avec des ordinateurs personnels. Nous avons trois bibliographies :

BETA (l'ancien BOOST –*Bibliography of Old Spanish Texts*- livre qui avait trois éditions, la dernière en 1983)

BITECA (successeur de BOOCT –*Bibliography of Old Catalan Texts*- livre publié en 1984) et

BITAGAP (qui n'a jamais été publié au format livre).

À cette époque, nous utilisions pour les données un code ASCII [aski] avec seulement 98 caractères :

á = a\, ç = c%, è = e`, n = {I}n{M}

De cette façon, nous pouvions représenter les signes graphiques caractéristiques des langues romanes.

Avant que DOS devienne obsolète, deux CD-ROM avec le contenu de PhiloBiblon ont été publiés et c'était la première publication non conventionnelle des corpus (1992, 1997).

À partir de ce moment, les sciences humaines numériques ont été utilisées non seulement pour le travail interne du projet, mais aussi pour la diffusion des résultats.

L'irruption de WINDOWS signifiait un changement brutal. Les tables ont été transformées, et le nombre de champs a augmenté (actuellement il y a environ 1000 champs. Par contre, nous avons perdu le contrôle sur le programme (maintenant seul le technicien peut le modifier) et il est devenu impossible d'obtenir des rapports imprimés pour les corriger. Avec ce changement, toutes les difficultés ont été compensées par une diffusion de notre travail que nous n'avions jamais rêvé : les tables étaient sur Internet.

De n'importe où, une personne ayant accès à un ordinateur et à Internet peut demander à notre programme où trouver les manuscrits d'une œuvre médiévale hispanique qui l'intéresse, et si les chercheurs de PhiloBiblon les ont déjà examinée, elle trouvera des descriptions de première main des codex.

Aujourd'hui, PhiloBiblon dépasse mille accès par mois et il est utilisé comme source de référence par les chercheurs ainsi que par les bibliothèques (BNE, CRAI-UB RESERVA), les assureurs et les maisons de vente aux enchères. La consultation gratuite et l'absence d'une édition imprimée complète n'a qu'un seul inconvénient: certaines personnes se permettent d'utiliser librement ces données sans indiquer que leur source est l'information offerte par PhiloBiblon.

Organisation de la recherche à PhiloBiblon

Notre travail est structuré en quatre phases.

- 1) Travail sur le terrain dans les bibliothèques et les archives
- 2) Analyse des données collectées
- 3) Introduction de l'information dans la base de données
- 4) Publication d'informations en ligne

Travail sur le terrain

Chacune des bibliographies de PhiloBiblon possède sa propre équipe de chercheurs qui se rendent dans les bibliothèques pour faire les descriptions des originaux directement. Avant chaque «expédition» on rassemble toutes les informations accessibles ainsi que les images en ligne.

Une fois dans la bibliothèque, on travaille avec le questionnaire codicologique du projet sur chaque manuscrit. Ce questionnaire est disponible sur le site de PhiloBiblon dans la section COLLABORATE en trois langues: Espagnol, Catalan et Portugais. Nous avons créé des formulaires pour la description des manuscrits, des imprimés et aussi pour les données biographiques et bibliographiques. Comme vous le voyez, chaque formulaire est accompagné d'un dossier de notes explicatives pour faciliter son utilisation par les chercheurs qui ne connaissent pas le projet.

Le formulaire de collecte de données est le résultat de plus de 30 ans d'expérience dans les bibliothèques et les archives et il est adapté aux possibilités de la base de données. Dans ses premières versions, il avait très peu d'informations sur la matérialité des manuscrits : format, nombre de folios, matériel (papier ou parchemin), présence de décoration et à peine plus ; mais au cours de la dernière décennie, nous avons incorporé un volume très important de données codicologiques, comme la structure des cahiers, la

présence de réclames, l'utilisation de la première ligne, la technique de réglure ou la filigrane du papier.

Nous accordons aussi beaucoup d'attention à l'histoire externe des manuscrits et cherchons à rassembler toute l'information qu'ils fournissent sur leurs anciens possesseurs, les personnes ou les institutions qui les ont manipulés ou gardés, et les vicissitudes qu'ils ont vécu depuis leur copie. Nous avons également ajouté toute l'information que nous recueillons sur les oeuvres. En ce moment, nous reproduisons le début et la fin de toutes ses parties (tables des matières, prologues, textes, etc.) dans une transcription essentiellement paléographique et leur position dans le volume.

Ce travail nécessite la formation des chercheurs en codicologie, en paléographie et en histoire de la littérature et aussi une disponibilité de temps. La plupart des chercheurs de BITECA, pour donner un exemple, sommes professeurs à l'université et nous ne pouvons accéder aux bibliothèques qu'en dehors du temps scolaire.

Analyse des données collectées

Une fois le travail de terrain terminé, tout le matériel est recueilli et des informations sont complétées. Les références sont vérifiées et contrastées avec celles obtenues dans d'autres bibliothèques. Les liens Internet sont vérifiés et on cherche toujours à insérer des liens avec les images numériques des manuscrits présents dans les Web des bibliothèques, avec leurs fichiers descriptifs ou avec des éditions en ligne des textes, s'ils existent.

Introduction de l'information dans la base de données

Une fois nous avons vérifié toutes les informations, elles sont transformées en un format qui permet de les intégrer dans la base de données, car il n'est pas possible de verser directement le contenu des formulaires dans le programme. C'est la partie la moins agréable de l'ensemble du processus, et celui qui a besoin de plus de temps. Prenons en compte, par exemple, qu'une semaine de travail sur le terrain de deux chercheurs peut représenter, entre l'analyse des données et l'introduction d'informations dans le programme, le travail de trois ou quatre mois d'une personne.

Publication sur Internet

Tous les trois à quatre mois, en fonction de la charge de travail des différentes équipes, nous envoyons à Berkeley toutes les informations afin qu'elles soient mises sur le serveur.

Ce travail est assez automatisé et peut prendre au total 24 heures pour toutes les bases de données. Le lendemain de son chargement dans le système, les informations sont disponibles pour tous les utilisateurs.

BETA, BITECA et BITAGAP ont leurs propres équipes de chercheurs et suivent leurs propres critères dans la définition du corpus et la collecte de données, bien que nous utilisions tous la même plate-forme informatique, et que nous partagions la même conception sur l'étude des manuscrits médiévaux.

BETA (3706 MSS)

BITECA (1648 MSS)

BITAGAP (5176 MSS)

Table	Records
Analytic	12,538
Bibliography	6,084
Biography	6,409
Copies	3,497
Geography	1,565
Institutions	635
Library	839
Ms Ed	4,763
Subject	306
Uniform Title	5,052
Total	41,688

Table	Records
Analytic	12,155
Bibliography	5,826
Biography	3,428
Copies	1,468
Geography	788
Institutions	541
Library	415
Ms Ed	1,945
Subject	149
Uniform Title	6,095
Total	32,810

Table	Records
Analytic	40,330
Bibliography	18,062
Biography	20,876
Copies	1,130
Geography	2,658
Institutions	2,119
Library	425
Ms Ed	5,265
Subject	557
Uniform Title	22,402
Total	113,824

Différences de contenu:

BETA (textes castillans): n'inclut pas encore des chansonniers (en attente d'incorporer le contenu du *Cancionero* de Brian Dutton).

BITECA (textes catalans): comprend la poésie.

BITAGAP (textes galicien et portugais): comprend la poésie, offre un registre pour chacune des lois et incorpore progressivement des lettres royales et d'autres textes qui pourraient être considérés plus appropriés au domaine diplomatique qu'au domaine littéraire, mais très utiles pour les chercheurs.

Différences dans le niveau de données des descriptions:

BETA: offre des descriptions plus succinctes, un certain nombre basées sur des informations indirectes et incorpore progressivement des données codicologiques.

BITECA: descriptions plus complètes, de première main et la plupart avec des informations codicologiques complètes.

BITAGAP: descriptions principalement de première main, avec très peu d'informations codicologiques.

Moteur de recherche

Le programme permet la recherche simultanée dans les trois bibliographies ou dans une seule. On peut récupérer les informations à partir de n'importe quel élément: le nom de l'auteur, le traducteur, le destinataire, l'une des formes du titre, la cote du manuscrit, le nom d'une bibliothèque, d'une ville, d'un ancien propriétaire, le début d'un texte, etc.

Pour finir, PHILOBIBLON est un projet ouvert, en transformation continue. Nous sommes ouverts à la collaboration avec d'autres projets similaires ou complémentaires et souhaitons apprendre de toutes leurs expériences. Techniquement, nous cherchons maintenant un moyen de transférer PhiloBiblon vers un environnement de base de données sémantiques et en ce qui concerne ses contenus, nous aimerions avoir plus de collaborateurs, non seulement disposés à faire des descriptions de manuscrits, mais aussi à collaborer dans la phase d'entrée de données dans le programme.

Il convient de noter que BETA met environ quatre ans à rendre accessible le travail sur le terrain alors que BITECA ne met que deux ans.