

D'un modèle pour l'édition électronique de documents judéo-espagnols à une collaboration pluridisciplinaire et internationale

*Soufiane ROUISSI et Ana STULIC-ETCHEVERS
Université Michel de Montaigne Bordeaux 3*

Cet ouvrage collectif s'inscrit dans la lignée de nos travaux conduits en vue du développement de l'édition électronique des documents judéo-espagnols à des fins de recherche. Ce projet, soutenu par l'équipe de recherche AMERIBER (EA 3656) de l'Université Michel de Montaigne Bordeaux 3, a été articulé au croisement de nos thèmes de recherche respectifs s'appuyant sur nos activités scientifiques sur les corpus documentaires¹ d'une part et le document numérique² d'autre part. Depuis 2005, nous abordons dans nos travaux ces deux dimensions dans une perspective pluridisciplinaire entre la linguistique de corpus et les sciences de l'information et de la communication. Le point de départ reposait sur la volonté de construire un dispositif technologique qui permettrait de rendre accessibles aux chercheurs les documents en judéo-espagnol produits au fil du temps dans les différentes contrées où la population séfarade s'est installée.

1. La constitution du corpus pour la réalisation de la thèse de doctorat d'Ana Stulic-Etchevers, *La grammaire de loke et siendo (ke) en judéo-espagnol des Balkans*, soutenue en 2007 sous la direction de Nadine Ly, Université Michel de Montaigne-Bordeaux 3, a nécessité un travail de recensement et d'analyse des sources écrites séfarades de différentes époques, ainsi que la numérisation d'un certain nombre de ces sources. Voir les chapitres 6 et 7 de la partie II, p. 90-123.

2. La thèse de doctorat en sciences de l'information et de la communication de Soufiane Rouissi, soutenue à Bordeaux en 2004, porte sur les questions liées à la normalisation pour la production de documents numériques.

La plupart des écrits en judéo-espagnol antérieurs au XX^e siècle sont rédigés en caractères hébreux. Plus précisément, il s'agit d'une adaptation de l'écriture hébraïque qui ne dénote que très partiellement les voyelles et qui a également connu sa propre évolution sans passer par une standardisation institutionnelle, ce qui rend sa transcription très délicate. Cela constitue aussi une difficulté quand il s'agit d'identifier correctement les documents, car pour pouvoir les lire, il est indispensable de maîtriser à la fois la langue espagnole (et les particularités de l'espagnol « séfarade ») et l'écriture hébraïque. C'est pourquoi la tradition écrite judéo-espagnole, à l'exception des cercles intellectuels séfarades, reste très peu connue de nos jours.

Pendant, de nombreux documents originaux, manuscrits, livres et autres imprimés, ont été conservés. Le nombre de livres en judéo-espagnol imprimés entre 1492 et l'époque moderne est estimé à plusieurs milliers³. Il faut également prendre en compte les publications périodiques, très nombreuses à partir du XIX^e siècle, qui n'ont pas été encore entièrement répertoriées, ainsi que des manuscrits de nature diverse dont le nombre exact est impossible à déterminer. Le travail dans le domaine de la description linguistique, mené durant les premières décennies du XX^e siècle, a également abouti à un certain nombre de transcriptions phonétiques et d'enregistrements sonores. Bien que fondées sur une réalité orale, ces transcriptions phonétiques doivent aussi être analysées comme cas spécifiques des sources écrites.

Du point de vue de la linguistique générale, le judéo-espagnol est très intéressant car il offre de nombreuses possibilités d'analyse comparative et historique. Malheureusement, dans de nombreux pays où il a été maintenu actif pendant des siècles, son usage a diminué considérablement depuis le début du XX^e siècle. Plusieurs facteurs expliquent les raisons de ce déclin. Dans de nombreuses villes des Balkans où la culture judéo-espagnole possède une tradition longue de plusieurs siècles, le judéo-espagnol a été progressivement remplacé par d'autres langues. Ce phénomène a été accéléré à la suite de la mort tragique d'un très grand nombre de locuteurs lors de la Deuxième Guerre mondiale⁴.

3. M. Molho, *Literatura sefardita de Oriente*, Madrid/Barcelona, Instituto Arias Montano, CSIC, 1960, p.145 ; A. Ben-Ur, « Ladino in Print : Towards a comprehensive bibliography », *Jewish History* 16, 2002, p. 322, note 4.

4. Pour les témoignages voir, par exemple, le livre de Salvador Santa Puche, *Libro de los testimonios: los sefardíes y el Holocausto*, vol. I, Sephardi Federation of Palm Beach County,

La culture séfarade n'a jamais été unilingue. À l'hébreu, langue de la religion et de la tradition écrite, et au judéo-espagnol véhiculaire, les Séfarades ont ajouté au cours du temps d'autres langues : bulgare, français, grec, roumain, serbe, turc, etc. La présence de ces langues non-traditionnelles dans les écrits, notamment dans la presse, devient remarquable durant la première moitié du XX^e siècle. Ces textes nous renseignent sur d'importants aspects de la vie séfarade de l'époque. Ils témoignent – bien que parfois négativement – de l'histoire externe du judéo-espagnol et offrent des vestiges précieux de la continuité de la culture séfarade.

Par conséquent, le travail de conservation, de numérisation et de traitement des sources originales peut largement contribuer à la préservation des connaissances sur la culture séfarade et, plus spécifiquement, sur le judéo-espagnol. L'un de nos objectifs repose sur le développement d'un corpus numérique afin de rendre les documents séfarades accessibles aux chercheurs grâce aux technologies de l'information et de la communication (TIC). Les nouvelles possibilités dues à l'avènement de l'ère numérique se caractérisent notamment par une accessibilité plus grande et évidente sur les sources tout en potentialisant une communication entre les membres d'une communauté scientifique qui est à la fois peu nombreuse et géographiquement dispersée. Nos travaux, dès leur origine, portent donc en eux deux préoccupations principales : opérationnaliser l'accès aux documents et observer les usages qui en découlent.

Partant de nos propres expériences de numérisation des textes judéo-espagnols en écriture hébraïque, notamment différents documents imprimés au XIX^e siècle à Belgrade et conservés par le Musée Historique Juif (Belgrade), nous avons procédé tout d'abord à l'élaboration d'un modèle conceptuel visant l'articulation d'une édition numérique destinée à des usages dans un contexte de recherche. Un premier modèle qui prend en compte les problèmes généraux d'édition de documents à des fins scientifiques a été présenté lors du colloque international *Lesser Used Languages & Computer Linguistics* organisé par EURAC Research qui s'est tenu à Bolzano (Italie) les 27 et 28 octobre 2005⁵. Dans un format plus détaillé, en janvier 2006, dans

Barcelone, 2003.

5. Soufiane Rouissi, Ana Stulic, « Annotation of documents for electronic edition of Judeo-Spanish texts: problems and solutions » dans Isabella Ties (éd.), *Lesser Used Languages & Computer Linguistics*, Proceedings of the Lesser Used Languages and Computer Linguistics

le cadre du séminaire du GRIAL⁶, nous avons présenté ce projet qui a été inclus en tant que projet émergent dans le programme quadriennal 2007-2010 de l'EA 3656 AMERIBER-Université Michel de Montaigne Bordeaux 3, sous le titre : « Projet de recherche pour le développement de l'édition électronique des documents judéo-espagnols ».

Afin de confronter notre modèle conceptuel aux pratiques réelles et/ou déclarées par les chercheurs en matière d'activités scientifiques sur les corpus documentaires, nous avons procédé, en mai et juin 2006, à la mise en place d'une enquête qui visait à recueillir l'avis des membres de communautés scientifiques de différentes disciplines confrontées au travail sur des corpus documentaires (principalement dans les domaines de l'analyse textuelle, de la linguistique de corpus et des sciences de l'information et de la communication) sur les thèmes principaux comme le travail de recherche scientifique en général dans le monde numérique, le recours aux logiciels, le travail à plusieurs et le partage de résultats. Cette enquête, gérée en ligne par un formulaire présenté en trois langues (français, anglais et espagnol) nous a permis de toucher un nombre important de chercheurs dans une dimension internationale. Elle a notamment été consultée par près d'un millier de chercheurs et près d'un quart d'entre eux ont répondu au questionnaire proposé. C'est ainsi que nous avons dénombré 245 questionnaires complétés par des chercheurs de différents pays : Espagne, France, États-Unis, Royaume Uni, Roumanie, Canada, Allemagne, Belgique, Maroc, Venezuela, etc. L'intérêt suscité par cette enquête est révélé par un nombre important de participants comme nous venons de le souligner ainsi que par la quantité des commentaires fournis et par les nombreux messages supplémentaires souvent accompagnés de demandes de compléments d'information. Nous avons été confortés dans nos hypothèses de départ : le développement du numérique doit permettre une *mise en discussion à plusieurs* des textes dans une communauté de chercheurs peu nombreuse et géographiquement dispersée. C'est bien dans cette direction que nous avons mené nos travaux depuis leur origine. Parmi les principaux résultats de cette enquête, nous retenons que 58,8 % des chercheurs ayant répondu à nos questions se déclaraient prêts à

Conference, Bolzano, 27th-28th October 2005, EURAC research, Bolzano, 2006, p. 265-280.

6. Groupe Interdisciplinaire d'Analyse Littérale (GRIAL), Université Michel de Montaigne Bordeaux 3.

partager et à échanger sur leurs travaux dans le cadre de l'étude scientifique des textes. Cette volonté déclarée de partage des travaux devrait, selon les participants à notre étude, être soumise à certaines conditions comme le recours à une charte de travail où les droits et devoirs de chacun seraient respectés, ou encore selon des principes de réciprocité et de respect des droits d'auteur.

Nos réflexions fondées sur l'analyse des résultats de l'enquête, ainsi que sur leurs répercussions sur la vision du modèle tel que nous l'avions envisagé ont été présentées dans une communication intitulée « Annotation en mode collaboratif au service de l'étude de documents anciens dans un contexte numérique » lors du colloque international *Digital Humanities 2006* qui s'est tenu à La Sorbonne à Paris du 5 au 8 juillet 2006. Le retour d'information que nous avons obtenu à l'occasion de cette communication devant un public international de chercheurs de disciplines et d'horizons très différents mais ayant en commun le travail sur les corpus documentaires au sein des sciences humaines, nous a encore plus convaincu que l'édition des documents judéo-espagnols devait être placée dans une approche documentaire large qui favoriserait la réutilisation des ressources et le travail en mode collaboratif.

C'est ainsi que nos propres analyses, enrichies par les différents échanges avec les collègues qui travaillent sur le judéo-espagnol dans d'autres centres universitaires et scientifiques, ainsi que les résultats de l'enquête, nous ont permis d'être à l'origine, au cours des années 2007 et 2008, de l'élaboration d'un prototype et de la création d'un réseau de collaborations externes.

Lors du colloque international *Corpus Diacrónicos en Lenguas Iberorromances* qui s'est déroulé du 25 au 27 octobre 2007 à Palma de Majorque (Iles Baléares) nous avons pu ainsi commenter les résultats et discuter des fonctionnalités qui seraient attendues par des chercheurs concernés par ces mêmes questions. La communication qui a été présentée lors de ce colloque a été suivie par la rédaction d'un chapitre dans un ouvrage collectif dirigé par Andrés Enrique Arias de l'Université des Iles Baléares⁷.

Dans la continuité, le bilan du projet, les principales caractéristiques du prototype, ainsi que l'analyse des difficultés rencontrées ont été présentés

7. Ana Stulic-Etchevers, Soufiane Rouissi, « Pensando un corpus en modo colaborativo: hacia el prototipo del corpus judeoespañol digital », dans Andrés Enrique Arias (éd.), *Diacronía de las lenguas iberorromances: nuevas perspectivas desde la lingüística de corpus*, Madrid / Frankfurt am Main, Iberoamericana / Vervuert, 2009, p. 117-134.

lors de la journée d'étude *Éditions électroniques* organisée par le GRIAL en avril 2008. Ce travail a été publié dans un ouvrage collectif coordonné par Federico Bravo⁸.

Nous exposons ici quelques grands principes que nous avons suivis dans notre travail de modélisation. Compte tenu des problèmes spécifiques liés au traitement des textes judéo-espagnols et leur organisation dans des corpus, tout en tenant compte de l'approche théorique sur le document numérique et des conditions de mise en usage en particulier dans un mode collaboratif, nous avons travaillé sur un modèle qui devrait répondre aux besoins identifiés. Nous nous sommes concentrés sur la définition des besoins, sans faire de choix trop contraignants concernant la future mise en œuvre d'un environnement numérique capable de servir nos activités de recherche. En ce sens, notre contribution se situe sur un plan analytique préliminaire à une réalisation concrète. Pour atteindre l'objectif principal, qui est la constitution d'un corpus documentaire, le modèle nécessite d'être conçu comme une bibliothèque numérique de documents qui sont décrits suivant des spécifications suffisamment fines, mais aussi ouvertes et interopérables. La dimension collaborative doit prendre en considération la gestion des utilisateurs. Quelques exigences que le dispositif devrait satisfaire ont déjà été identifiées dans la phase préliminaire. Nous pouvons les énoncer ainsi :

- le document source doit être accessible, en version translittérée⁹ (texte numérique), et, idéalement, se présenter aussi comme une collection de fichiers d'images ;
- la version transcrite est donnée comme le point de départ de la discussion ou de l'analyse ;
- les annotations descriptives sont présentées sous forme de métadonnées (élaborées conformément à la norme Dublin Core, désormais largement acceptée) ;

8. Ana Stulic-Etchevers, Soufiane Rouissi, « Questionnements autour de la mise en place d'un corpus numérique judéo-espagnol », *Enjeux et perspectives de l'édition électronique*, Textes réunis et présentés par Federico Bravo, Bordeaux, Presses Universitaires de Bordeaux, Collection de la Maison des Pays Ibériques, 2010, p. 31-48.

9. Nous définissons la translittération comme la substitution de chaque signe du document original par un signe graphique (ou une valeur numérique) ayant la valeur et la distribution équivalentes à celles que l'on peut observer dans le système du document original.

- l'utilisateur autorisé peut ajouter des annotations libres ;
- l'utilisateur autorisé peut exporter le résultat de son travail en faisant le choix de récupérer ou non les annotations faites par lui-même.

Le dispositif technologique envisagé pour la publication des documents judéo-espagnols doit prévoir la numérisation des documents originaux de façon à permettre une pluralité de représentations accessibles aux chercheurs : 1) le format image, 2) la numérisation qui suit les normes rigoureuses d'édition paléographique, ainsi que 3) les différentes transcriptions normalisées qui incluent un plus grand degré d'interprétation.

Dans un souci de réutilisation des ressources, c'est-à-dire pour faciliter leur localisation et leur identification, nous avons défini, sur tous les niveaux de représentation, un protocole de description qui emploie les normes et les standards largement utilisés et reconnus pour des usages dans un contexte numérique. D'autre part, nous prévoyons que le chercheur puisse agir sur les versions proposées, par l'ajout de ses propres annotations (normalisées ou libres) à tous les niveaux de représentation, par la correction ou la modification du système de transcription proposé. Le résultat du travail individuel sur les documents devrait par la suite être sauvegardé et exporté dans un des formats standardisés pour les besoins personnels, et/ou, si le chercheur-utilisateur le juge pertinent, les résultats obtenus pourraient aussi être publiés et mis en discussion à l'intérieur de notre dispositif.

Le modèle de corpus ainsi défini doit servir la communauté scientifique concernée, mais l'établissement d'un protocole de description particulièrement destiné aux documents séfarades nécessite l'expertise de cette même communauté. Tout en poursuivant nos investigations, les différents contacts noués au cours du temps nous ont conduits fin 2009 à organiser un colloque international avec le soutien de l'Équipe d'Accueil AMERIBER de l'Université Michel de Montaigne Bordeaux 3 et en partenariat avec la Faculté de Philologie de l'Université de Belgrade et le Centre Culturel Français de Belgrade¹⁰. Le colloque international intitulé « Sources judéo-espagnoles des Balkans dans les recherches scientifiques »¹¹, qui s'est tenu les 15 et 16 décembre 2009 à Belgrade (Serbie), était consacré à l'exposition des

10. Cette institution, rattachée à l'Ambassade de France à Belgrade, depuis 2011 porte le nom de l'Institut français de Serbie.

11. Colloque organisé par Ana Stulic-Etchevers, Soufiane Rouissi pour l'Université Michel de Montaigne Bordeaux 3 et Jelena Filipović pour l'Université de Belgrade.

travaux de recherches qui concernaient, d'une part, l'approche documentaire des sources séfarades, c'est-à-dire, leur recensement et description, et d'autre part, leur analyse dans les sciences humaines (linguistique, littérature, histoire, sociologie, etc.). Autour de cette problématique commune aux différentes disciplines, nous souhaitons réunir les travaux de recherche qui concernent les documents séfarades de la région des Balkans.

Cette rencontre nous a permis d'élargir les champs scientifiques impliqués par notre modèle de corpus et d'inclure les dimensions nouvelles dans la description de l'héritage documentaire séfarade, comme la pluralité des langues et des systèmes d'écritures qui le caractérisent, les différents types de contenus laissés en marge de la recherche scientifique (publicités, œuvres rabbiniques moins connues, notes manuscrites de nature diverse, etc.) et la diversité d'approches scientifiques dans l'analyse des sources séfarades. Nous réunissons dans cet ouvrage des travaux de chercheurs qui s'intéressent à ces questions. Cet ouvrage s'articule ainsi autour de trois parties pour présenter, selon différents points de vue, ces travaux menés sur des sources séfarades issues de différentes zones géographiques.

La première partie soulève les problèmes liés au recensement des sources, à la nature de celles-ci et à leur classification. Les travaux s'interrogent à la fois sur les caractéristiques formelles des documents (systèmes d'écriture, modes de production) et sur les conditions historiques dans lesquelles les publications ont vu le jour et ont été diffusées. Karlo Budor traite la question du polymorphisme et des systèmes d'écriture des documents judéo-espagnols des Balkans, alors que Michael Studemund-Halévy s'intéresse à l'apparition de l'alphabet cyrillique dans les écrits séfarades. Ivana Vučina Simović propose une classification des sources écrites séfarades dans une perspective sociolinguistique. La presse et les périodiques font l'objet des autres contributions de cette première partie. Krinka Vidaković Petrov étudie la presse séfarade de Belgrade et Sarajevo de 1888 à 1941. Dimitrije Pešić présente le magazine *Hashalom* de Belgrade et Sofia (de 1903 à 1906). Nicole Abravanel décrit et contextualise *Le Judaïsme sépharadi* de Paris à Londres (de 1932 à 1966).

Dans la deuxième partie, les travaux sont orientés vers l'analyse du contenu des sources dans une diversité d'approches. Les travaux de Paloma Díaz-Mas portent sur les annonces publicitaires dans les écrits séfarades (de 1880 à 1930). Drita Tutunović présente les collections yougoslaves de

proverbes judéo-espagnols. Katja Šmid met en lumière les informations à caractère historique contenues dans les œuvres rabbiniques de Eli'ezer Papo de Sarajevo (XIX^e siècle). Aitor García Moreno nous fait découvrir l'usage particulier des gloses dans *Luzero de la pasensia* (dans les années 1880 en Roumanie). Gorana Zečević-Krnetić décrit la forme *fin* en judéo-espagnol balkanique. Željko Jovanović, dans une approche comparative, étudie le thème de l'inceste paternel dans le *romancero* séfarade.

La troisième partie s'intéresse plus particulièrement à la préservation et à la valorisation des textes avec Jelena Filipović qui rappelle l'importance des textes authentiques pour la revitalisation du judéo-espagnol. Les applications informatiques sont décrites par Duško Vitas et Cvetana Krstev à travers la question des méthodes d'alignement de textes. Javier Pueyo présente une proposition d'alignement automatisé des sources pour le corpus biblique séfarade en se basant sur la méthode traditionnelle de traduction de textes bibliques.

Les textes¹² réunis dans ce livre répondent ainsi à la volonté de découvrir la diversité des sources séfarades écrites, la pluralité des approches scientifiques à l'œuvre dans leur étude et la pluridisciplinarité devenue indispensable dans leur traitement numérique à des fins de recherche.

12. Tous les textes rédigés initialement en espagnol, anglais ou serbe ont été traduits et adaptés en langue française.