

PROJET DE BASE DE DONNEES ALIENTO

Prof. Marie-Christine Bornes Varol
Dept. Etudes Hébraïques
INALCO – Paris & CERMOM
UMR8099, Langues, Musiques, Sociétés
varol@noos.fr

Prof. Marie-Sol Ortola
Dept. Espagnol
Université Nancy 2
Dir. Equipe ROMANIA
ortola@univ-nancy2.fr

Jean-Daniel Gronoff
Ingénieur informaticien
Association PetrusRamus.org
jd.gronoff@free.fr

Résumé du projet ALIENTO

1. Contexte scientifique et objectifs du projet
2. Description du projet, méthodologie
3. Résultats attendus

Au IX^e siècle, la riche tradition arabe de l'*adab* parvient en Espagne, elle y rencontre la tradition sapientielle juive de la littérature midrashique. De nouveaux recueils sont composés, des œuvres originales produites aux X^e et XI^e siècles et à partir du début du XII^e siècle, les recueils d'*exempla*, de sagesse et de dits des philosophes sont traduits en hébreu, en latin, en langues romanes. De nouvelles compilations se créent pour servir la prédication chrétienne, incluant des proverbes populaires, des maximes et sentences tirées d'auteurs latins et grecs, de philosophes, de la patristique et des morales d'*exempla*. Cet héritage complexe se retrouve en bonne part dans l'immense littérature parémiologique espagnole qui s'épanouit aux XVI^e et XVII^e siècles et dans les proverbiars espagnols, judéo-espagnols et maghrébins contemporains. Jusqu'à présent, ces énoncés brefs qui s'échangent n'ont pas fait l'objet d'une étude d'ensemble qui retrace leur circulation à travers les différentes langues (savantes et vernaculaires) utilisées dans la péninsule ibérique. Nous nous proposons de combler ce vide en constituant un outil de recherche efficace permettant non seulement d'identifier les énoncés échangés d'une culture à une autre, d'une langue à une autre, mais aussi d'informer sur les sources de ces énoncés et sur leur postérité.

Le projet réunit des chercheurs membres de diverses équipes scientifiques dont trois sont lorraines (ROMANIA [Nancy2], HISCANT [Nancy2] et CELTED [METZ]) et une parisienne (CERMOM [Inalco]). Il vise à la constitution d'une base de données textuelles (énoncés sapientiels) qui doit permettre de comprendre les sources et les cheminements des corpus sapientiels dans une aire géographique et une époque données ; leur impact et leur diffusion ; les modifications que les cultures de réception opèrent ; les circuits de transmission des unités ; les liens que les unités entretiennent entre elles à l'intérieur d'une culture donnée ; et les liens des unités avec les corpus sapientiaux et exemplaires.

Il s'agit d'un projet collaboratif impliquant un ingénieur de recherche dont la fonction sera d'imaginer des procédures pour faire converger par calcul des domaines de connaissance diversifiés, de vérifier leur validité, nous permettant de faire l'analyse linguistique et interculturelle d'un corpus multilingue (arabe, hébreu, grec, araméen, latin, judéo-espagnol, espagnol...) et de développer un programme d'interrogation multilingue (français, espagnol et anglais) de la base de données. L'option retenue est celle du balisage des unités (énoncés sapientiels) dans les textes saisis ou numérisés. La base de données textuelles (1) sera reliée à (2) une base de données bibliographique ; (3) une base de données biographiques sur les auteurs, compilateurs & traducteurs; (4) une base de données sur les œuvres (compilant les données issues des études et éditions critiques); (5) une base lexicale par langue, renvoyant aux équivalences possibles de termes apparentés avec leurs différentes graphies.

ALIENTO Project -abstract

1. Scientific context and aims of the project
2. Description of the project, methodology
3. Expected goals

In the IXth century, the rich Arabic tradition of *adâb* finds its way to Spain where it meets the Jewish sapiential tradition of the midrashic literature. New collections are created, including original works in the Xth and XIth centuries and at the beginning of the XIIth century, *exempla*, wisdom and philosophers collections are translated into Hebrew, Latin, and Romance languages. There are also new anthologies used in Christian predication, including popular proverbs, maxims and sentences from Latin and Greek authors and philosophers, from patristic texts and moral *exempla*. Much of this complex heritage can be found in the extensive Spanish paremiologic literature which is at its highest in the XVIth and XVIIth centuries and in current Spanish, Judeo-Spanish and Maghrebian collections of proverbs. Until now, these short statements exchanged from culture to culture have not been studied comprehensively with respect to their sources and their posterity, through the different (learned and vernacular) languages used in the Iberian peninsula. Our aim is to fill this gap and to build an efficient research tool which will be used not only to identify the statements being exchanged between cultures, between languages, but also to provide information about the sources of these sapiential statements and their posterity.

Researchers involved in the project belong to different scientific teams: three are from Lorraine (ROMANIA [Nancy2], HISCANT [Nancy2] and CELTED [Metz]) and one from Paris (CERMOM [Inalco]. The objective of the project is to build a textual database (sentential items units) which will help understand the sources and paths of sapientials corpuses in a given geographical area and specified period; their impact and diffusion; the resulting changes in the receiving cultures; the transmission paths of the units; the connections between units within a given culture; and between units and sapiential and exemplary corpuses.

This is a collaborative project in which will participate a research engineer whose task will be to design computational procedures and to verify their validity in order to get domains with diversified knowledge to converge, and to develop a multi-language (French, Spanish and English) programme used to query the database so we will be able to make linguistic and intercultural analysis of a multilingual (Arabic, Hebrew, Greek, Aramaic, Latin, Judeo-Spanish, Spanish...) corpus.

The option we have chosen is to encode units (sapiential statements) in the documents we have numerised or typed. The textual data base (1) will be connected to (2) a bibliographic data base, (3) a biographic data base concerning authors, compilers and translators, (4) a data base regarding the works (compiling data from monographies and critical editions), (5) a lexical data base in which each language appears highlighting the possible equivalences of related terms with their different spellings.

Resumen del proyecto ALIENTO

- 1) Contexto científico y finalidad
- 2) Descripción del proyecto
- 3) Resultados esperados

En el siglo IX, la fecunda tradición árabe del *adab* llega a España donde se encuentra con la tradición sapiencial judía de la literatura midrásica. Se componen nuevas colecciones, se redactan obras originales en los siglos X y XI y, a partir de principios del XII, se traducen las colecciones de *exempla*, de sabiduría y dichos de filósofos al hebreo, al latín, a las lenguas románicas. Se escriben nuevas compilaciones utilizadas por la Iglesia para la predicación cristiana, que incluyen proverbios, refranes, máximas y sentencias sacadas de autores latinos y griegos, de textos filosóficos, de la patrística, y de las lecciones de *exempla*. Esta compleja herencia viene recogida en gran parte en la inmensa literatura paremiológica española que florece en los siglos XVI y XVII y en los refraneros españoles, judeo-españoles y magrebíes contemporáneos. Hasta el momento, estos enunciados breves que se intercambian desde hace siglos no han sido objeto de un estudio de conjunto que reconstruya su circulación a través de los diferentes idiomas cultos y vernaculares que se usaron en la Península ibérica. Nos proponemos colmar esta laguna constituyendo una herramienta de investigación eficaz que permita no sólo identificar los enunciados que pasaron de una cultura a otra o de un idioma a otro, sino también informar sobre las fuentes de estos enunciados y su posteridad.

El proyecto reúne investigadores de diversos grupos de investigación entre los que tres proceden de Lorena (Romania, Hiscant, Celted) y uno de París (Cermom). El objetivo del proyecto es la elaboración de una base de datos textuales (enunciados sapienciales) que permitirá entender qué fuentes tuvieron y qué camino siguieron los corpora sapienciales en una área geográfica y una época determinadas ; cuáles fueron su impacto y difusión ; qué modificaciones se debieron a las culturas de recepción ; qué caminos de transmisión recorrieron las unidades ; qué lazos mantienen estas unidades entre sí dentro de una cultura determinada ; qué lazos tienen con los corpora sapienciales y ejemplares.

Se trata de un proyecto colaborativo que incluye a un ingeniero informático cuya función es idear procedimientos que calculen la convergencia de los campos diversificados de conocimiento, indaguen sobre su validez y nos permitan hacer el análisis lingüístico e intercultural de un corpus multilingüe (árabe, hebreo, griego, arameo, latín, judeo-español, español) y desarrollar un programa de interrogación multilingüe (francés, español e inglés) de la base de datos. La opción elegida es la transcripción informática de los textos o su numerización y su codificación (etiquetaje de los “enunciados sapienciales”). La base de datos textuales (1) se enlazará con (2) una base de datos bibliográficos; (3) una base de datos biográficos de autores, compiladores y traductores; (4) una base sobre las obras que compile los datos sacados de los estudios y las ediciones críticas; (5) una base léxica (por idioma) que remita a las variantes terminológicas y gráficas.

סיכום הפרוייקט סיכום הפרוייקט

1. הפרוייקט תוכן מדעי ומטרות
2. תיאור הפרוייקט ומתודולוגיה
3. צפי תוצאות

את דרכה לספרד ופגשה שם במסורת במאה התשיעית לספירה מצאה המסורת הערבית העשירה של האפד' עשרה חוברים אוספים חדשים ובהם גם יצירות-במאות העשירית והאחת החוכמה היהודית של הספרות המדרשית. תורגמו לעברית, ללטינית שרה, ספרות מוסר, דברי חכמה וכתבי הגות פילוסופיים-מקוריות. החל במאה השנים ולשפות הרומאניות. כן קובצו פתגמים עממיים ואמרות שמקורן בכותבים רומאים ויוונים, בפילוסופים, בכתבים על כלולהזאת נכבד מן הירושה המורכבת אבות הכנסייה ובסיפורי מוסר, כדי לשמש את הנבואה הנוצרית. חלק עשרה וכן אצל אוספי-עשרה והשבע- הפרמיולוגית הספרדית העשירה שהגיעה לשיא פריחתה במאות הששספרות ספרדים או צפון אפריקניים עכשוויים. חיבורים קצרים אלה, תוצר מובהק של המפגש בין-פתגמים ספרדים, יהודים מתחקה אחר מסעם של חיבורים אלה בין השפות התרבותיות, לא נחקרו עד היום; עד כה לא נערך מחקר מקיף ה (השונות) המלומדות או הוורנקולריות) ששימשו בחצי האי האיברי. אנו מציעים למלא חלל זה וליצור כלי מחקרי את יעיל שיאפשר לא רק לזהות את מעברם של החיבורים מתרבות לתרבות ומשפה לשפה, אלא גם להאיר מקורותיהם ואת גלגוליהם המאוחרים.

ROMANIA הפרוייקט הוא מפגש של חוקרים הפועלים בצוותים מדעיים שונים: שלושה מחבל לורן (CERMOM [Inalco] (ואחד מפריס) CELTED [METZ] וְ HISCANT [Nancy2] , [Nancy2] מטרתו היא בניית בסיס נתונים טקסטואלי) של דברי החוכמה) שיאפשר הבנה מעמיקה יותר של התפתחות החוכמה ושל מקורותיהם במרחב גיאוגרפי ובזמן נתון, יתחקה אחר השפעתם ואופן הפצתם, אחר-קורפוס השינויים שערכה התרבות שקלטה אותם, המסלולים בהם נמסרו, הקשר בין היחידות בתוך תרבות נתונה, והקשרים בינם לבין קורפוס החוכמה וקורפוס ספרות המוסר.

תף גם מהנדס מחקר. תפקידו הוא לדמות פרוצדורות ממחשבות כדי לרכז את תחומי בפרוייקט שיתופי זה משת לשוני זה-תרבותי של קורפוס רב-המידע השונים ולבדוק את מהימנותם. פרוצדורות אלה יאפשרו ניתוח בלשני ובין ית לחקר מסד לשון-ספרדית וספרדית... (ופיתוח תוכנה רב-) עברית, יוונית, ארמית, לטינית, יהודית הנתונים) צרפתית, ספרדית ואנגלית)

דרך הפעולה שבה בחרנו היא סימון היחידות) דברי החוכמה) בתוך טקסטים מוקלדים או סרוקים. מסד הנתונים למסד נתונים ביוגרפי על המחברים, מחברי האוספים; הטקסטואלי) 1) יקושר למסד נתונים ביבליוגרפי) 2) למסד (5); למסד נתונים על היצירות) המשלב מידע הלקוח ממחקרים ומהדורות מדעיות) (4); והמתרגמים) 3) שעל ההבדלים באיותן נתונים על אוצר המילים בכל שפה, ועל הקשר בין מילים מקבילות או בנות אותה משפחה, הוא אף יוכל להצביע.

ملخص المشروع

1. السياق العلمي والأهداف
2. المنهجية وصف للمشروع
3. النتائج المتوقعة

في القرن التاسع الميلادي ، بلغ الأدب العربي الثري إسبانيا ، حيث التقى بالثقافة اليهودية للحكم الموجودة في الأدب المديني... فألفت مجموعات نصوص جديدة وابتكرت أعمال في القرنين العاشر والحادي عشر وانطلقاً من القرن الثاني عشر، تراجعت مجموعات الأمثال م وأقوال الفلاسفة إلى العبرية واللواتينية واللغات الأوروبية الدارجة. والحك وأنشئت تصنيفات جديدة من أجل الوعظ المسيحيين، أُدرجت فيها بعض الأمثال الشعبية وبعض الأقوال والعبر المقتطفة من أعمال الكتاب اللواتينيين والإغريقيين والفلاسفة جزءاً كبيراً من هذا التراث المتعدد الأصول في أدب الأمثال والنصوص الباطنية والعبارة. ونجد إسبانيا الوافر الذي ازدهر في القرنين السادس عشر والسابع عشر ، وفي مجموعات الأمثال الإسبانية والإسبانية اليهودية والمغربية المعاصرة.

أما تصنيف مدى تنقلها وحتمى الآن ، لم تتوضع هذه المقولات القصيرة المتبادلة موضع دراسة عبر اللغات المختلفة (الفصاحة من هنا والدارجة) الممتعة في شبه الجزيرة الإسبانية. فتريدم لهذا النقص، نقدم مشروعاً هدفه إيجاد أدوات بحث فعالة لتتصدر على تعريف بحث بمعلومات حول الأقوال المتبادلة من ثقافة إلى أخرى ومن لغة إلى أخرى، بل، تزود أيضاً البمصائر هذه الأقوال والكتيب التي استفادت منها في ما بعد.

انضم إلى المشروع باحثون من فرق علمية مختلفة، ثلاثة من هنا من إقليم اللورين (رومانيا وفريسي باريسي) سيروم [المعهد¹ أنانسي 2] ، هيسكانت [أنانسي 2] [وسالتيدي] مدينة ماتز (.) يسعى المشروع إلى إقامة قاعدة² اللغات والثقافات الشرقية ، إينالكو [الوطني لدراسة معطيات لنصوص (حكيم) وظيفتها الإمداد بمعلومات حول المصائر وحول الكتب التي اقتبست من هذه النصوص في إطار جغرافي وزمني معيّن وحول تأثيرها ومدى انتشارها وحول التغيرات التي المضيئة " وحول طرق تنقل الوحدات وحول الارتباطات بين الوحدات داخل قامتها " والثقافات ثقافية معينة وحول الارتباطات بين الوحدات ومجموعات نصوص الحكم والأمثال.

وسيشارك في هذا المشروع الجماعي من ندس بحث ستكون مهمته، من جهة، تصوّر برامج حيته وتمكننا من القيام بتحويل مجموعة لاستخراج مديني معرفية مختلفة والتحقق من نصوص متعددة اللغات) بالعربية والعبرية والإغريقية والأرامية واللواتينية واليهودية الإسبانية والإسبانية (من حيث اللغة وما بين الثقافات ، ومن جهة أخرى، تطوير برنامجية والإنكليزية). واخترت لاستطلاع قاعدة المعطيات في لغات متعددة (الفرنسية والإسبانية القيام بموضوع إشارات للوحدات) الحكم (بعد نسخ النصوص أو ترقيمه). وستكون قاعدة المعطيات للنصوص (1) (ربوطة) (2) بقاعدة معطيات للمراجع (3) بقاعدة معطيات لسير الكتب بقطفة من المصنفين والمترجمين (و) (4) بقاعدة معطيات للأعمال) تجمّع بين المعطيات الم الدراسات والطبعات المحققة (و) (5) قاعدة مفردات لكل لغة تدلّ على المصطلحات المترادفة أو المتقاربة وعلى مختلف أشكالها الخطية.

¹ Lorraine (ROMANIA [Nancy2], HISCANT [Nancy2] et CELTED[Metz])

² CERMOM [Inalco, Paris]

