

SYSTÈMES INTERACTIFS ET ÉTUDE DES LANGUES ET CIVILISATIONS ANCIENNES

François CHARPIN

Les techniques de programmation et de gestion des données qu'utilisent l'Intelligence Artificielle et les outils Hypertexte ou Hypermédia permettent un renouvellement complet des instruments scientifiques et pédagogiques utilisés dans le domaine des langues et civilisations anciennes.

Il ne s'agit plus d'engranger des masses de documents bruts : textes au kilomètre, documents de toute nature, figurés ou non... sans se soucier de leur lecture, de leur origine, de leur interprétation, de leur validité dans un corpus ; il n'est pas acceptable que, pour des raisons de facilité, on considère le mot seulement comme une séquence de lettres, la reproduction d'une mosaïque seulement comme une belle image, la visualisation d'un plan seulement comme une maquette touristique... Un travail scientifique se doit d'indiquer les références des objets étudiés, d'en proposer une interprétation, de préciser les réseaux de connaissances qu'ils sont susceptibles d'impliquer.

L'informatisation des données antiques utilise donc toutes les ressources qu'offrent les traitements automatiques liés aux outils de formalisation, d'ordre linguistique, mais aussi sémantique et relationnel. Soutenue par les Universités PARIS-7 et LILLE-3, l'équipe de l'UA 1423 (Centre de Recherche en Documentation : Langues et Civilisations de l'Antiquité) s'est spécialisée dans ce type d'étude.

OUTILS DE CONSULTATION

Un document est toujours lié à un contexte qui en permet la lecture.

Un énoncé grec ou un énoncé latin perd une grande partie de sa signification s'il n'est pas situé dans des unités clairement indexées, accompagné de ses références bibliographiques (édition de base), de l'apparat critique, d'une ou de plusieurs traductions qui, au regard de

l'original antique, fournissent l'interprétation de l'éditeur. Aucune monnaie ne doit être présentée sans qu'en soient donnés le numéro d'identification, la collection, la datation, le métal, la dénomination, l'origine, l'autorité émettrice, la description de l'avvers et du revers, les inscriptions, les éléments bibliographiques, les reproductions... Aucun plan de bâtiment ou de ville ne doit être fourni sans qu'apparaissent le nom de l'éditeur et les documents compulsés. Aucun vase n'est cité sans indication de sa collection, de sa forme, de son type, de son peintre, de sa décoration, de son contexte, de sa date ou de sa période, sans la notation des inscriptions, des dimensions, de la première publication et de la bibliographie de l'objet ; l'ensemble est éventuellement complété par une collection d'images, elles-mêmes précisément étiquetées (objet décrit, place, orientation...).

Dans cette perspective, le document n'apparaît jamais comme une configuration accidentelle d'éléments qui se manifestent à l'intérieur d'un fichier, mais comme le point de rencontre d'une série de fichiers cohérents. L'équipe, dans le cadre du programme PERSEUS qui porte sur le grec (dirigé par l'Université Harvard) et dans le cadre du programme ITER qui porte sur le latin (et dont elle a la responsabilité), souscrit entièrement à ces exigences de présentation. Pour que l'environnement informatique soit plus confortable, des fichiers complémentaires ont été ajoutés pour renfermer les annotations et commentaires personnels de l'utilisateur.

OUTILS LINGUISTIQUES

Au niveau du texte, l'un des éléments essentiels du traitement automatique est la détermination des unités linguistiques. Dans une formulation rudimentaire on admet qu'un énoncé se découpe en chapitres, un chapitre en paragraphes, un paragraphe en phrases, une phrase en syntagmes, un syntagme en morphèmes. L'identification de ces différents éléments est impliquée par toute enquête qui se veut scientifique. Plusieurs auteurs ont cru qu'ils pouvaient faire l'économie de ce travail ; on peut assurément considérer un mot comme une chaîne de caractères et faire repérer toutes ses occurrences à l'intérieur d'un ouvrage ; il faut cependant constater qu'une telle pratique ne donne aucune des garanties de simplicité et d'exhaustivité que l'on est en droit d'attendre d'un automate ; pour ne prendre qu'un exemple, s'il s'agit de trouver les références du verbe ALLER dans *A la recherche du temps perdu*, il faudra demander successivement toutes les formes qui

commencent par AL, par AIL, par IR, par VA... A supposer qu'aucun paradigme ne soit oublié, le lecteur disposera certainement des formes du verbes ALLER, mais aussi de milliers d'autres dont il n'a que faire.

L'équipe a construit des instruments pour traiter linguistiquement l'énoncé ; elle a mis au point des analyseurs morphologiques du latin et du grec (systèmes experts AULA) ; des outils lexicographiques qui prennent comme unité le chapitre, le paragraphe, la phrase et qui sont capables de rechercher des termes contigus ou non (programmes RECH) ; des outils métriques qui analysent automatiquement l'hexamètre latin et l'hexamètre grec (LATIN3 et GREC3) ; des outils syntaxiques qui, utilisant les données scrupuleusement établies par le Laboratoire d'Analyse Statistique de l'Université de Liège, précisent des règles de distribution des unités intonées (CAESEDIT), des syntagmes sujet et objet, qui permettent, pour le latin comme pour le grec, de mettre au point des règles d'ordre dans la répartition des constituants de l'énoncé. De multiples enquêtes sont menées sur la notion de profondeur, sur les contraintes liées à la voix, au temps, au mode, à la détermination... Ainsi s'élabore une véritable grammaire de la performance.

Cette expérience enrichit les bases de données ITER et PERSEUS. Dans ce dernier logiciel, il suffit de sélectionner un mot dans le texte et d'actionner le bouton adéquat, pour obtenir toutes les analyses et tous les lemmes correspondant à la forme choisie. Ainsi le vocable *αποδεχθεντα* est analysé comme participe d'*αποδεχομαι* et d'*αποδεικνυμι* au nominatif, au vocatif, à l'accusatif pluriels neutres. La lemmatisation est automatique. Le travail lexicographique en est considérablement facilité. En sélectionnant le lemme et en activant la touche de définition, l'automate fait apparaître sur l'écran l'article du dictionnaire de Liddell and Scott ; il est donc possible de contrôler le sens et l'emploi d'un vocable dans un contexte donné. A partir du lemme, la machine énumère toutes les références (classées par cas, par oeuvre...) dans un auteur donné. Evidemment, chaque référence, à son tour, permet d'appeler le texte lui-même où la citation est présentée en caractères ombrés en regard de la traduction.

La grande originalité, brevetée par le système PERSEUS, est dans l'utilisation du dictionnaire de Liddell and Scott comme langage pivot : toutes les définitions anglaises ont été systématiquement classées en fonction du lemme qui les introduit. Il devient donc possible d'énumérer tous les mots grecs qui, dans l'article que leur consacre le dictionnaire, font intervenir un vocable donné ; par exemple, les 164 termes qui impliquent « to speak », les 27 termes qui impliquent « freedom »... Ces

listes permettent un balayage comparatif de tous les auteurs connus du système, et évidemment le recours aux listes de références et de citations qui leur sont associables. Un tel traitement permet d'espérer la reconnaissance exhaustive de tous les concepts présents dans une oeuvre.

OUTILS RELATIONNELS

L'informatisation des données portant sur les civilisations grecque et romaine implique que sur un même support se trouvent enregistrés des fichiers très nombreux, de nature différente et provenant de collections disparates ; c'est ainsi que la base VELA renferme des textes, des analyses, des règles de grammaire, des règles sémantiques ; le système PERSEUS contient un précis d'histoire ancienne, une bibliothèque d'auteurs, un Atlas, des documents portant sur l'art et l'archéologie, un inventaire d'outils et références, une encyclopédie... Pour ne prendre qu'un exemple, il peut faire afficher, presque instantanément, le plan du Parthénon, une monnaie représentant une victoire ailée, une statue du musée de Delphes, ou encore la photo d'Athènes obtenue par satellite... Si une connaissance globale du monde antique est imaginable, il faut bien admettre que de telles procédures permettent d'en approcher ; tout le problème est de définir les liens qui doivent être établis à l'intérieur d'un même fichier et entre des fichiers différents.

Au niveau le plus élémentaire, les différents éléments d'un même fichier ont été conçus pour former des ensembles cohérents ; dans VELA, toutes les règles syntaxiques sont affichées avec le même format ; dans PERSEUS, les schémas d'un même édifice sont présentés à la même échelle et dans la même orientation ; ainsi le développement d'un même site archéologique (par exemple, le Sanctuaire d'Apollon à Delphes) peut être suivi sur l'écran selon ses modifications chronologiques (600 avant J-C, 586-548 avant J-C, 550-500 avant J-C, 500-450 avant J-C).

Par le jeu des boutons associés au langage Hypertalk, tout mot d'une version latine appelle immédiatement la fenêtre d'analyse qui lui correspond ; toute intervention sur un bouton placé dans un plan à grande échelle du système PERSEUS appelle la diapositive numérisée qui a été saisie selon l'angle de la prise de vue indiqué sur le plan. Ainsi il devient possible d'entreprendre une visite du site et de contrôler par l'image l'organisation et la disposition des principaux témoins archéologiques évoqués par les plans et par les notices descriptives.

Un certain nombre d'outils lexicaux permettent de rechercher un vocable anglais ou grec à travers la totalité des documents (textes et documents). C'est ainsi que, dans PERSEUS, une recherche sur le nom propre CRETE donne toutes les références de ce mot dans l'encyclopédie, dans Hérodote, dans les fichiers archéologiques... D'autres outils, plus complexes, font appel à des mots-clefs ; dans PERSEUS, l'index des bâtiments architecturaux permet d'identifier tous les sites qui présentent un autel, des fortifications, des bains... ; l'index des monnaies, toutes les pièces qui possèdent une même origine, ou qui appartiennent à une même collection... On peut accéder aux collections d'objets grâce à un programme qui sélectionne le type (sculpture, vase ou monnaie), le motif représenté (bâtiments, personnages légendaires, divinités, animaux, habillement...) et, dans ce motif, une spécification plus précise ; par exemple, pour les vases décorés de scènes mythologiques, le système propose Achille et Ajax au jeu, Achille et Hector, Achille et Memnon, Achille et Penthésilée...

L'équipe utilise les techniques hyper-texte et hyper-media initiées par HYPERCARD. Ainsi, dans PERSEUS, en sélectionnant à l'intérieur de la traduction le nom d'Argos et en appelant le lien « Atlas géographique », le lecteur du livre 1 d'Hérodote provoque l'affichage du Péloponnèse et l'insertion de cette ville sur la carte. En manoeuvrant la souris sur les mots soulignés dans le précis d'histoire grecque, il déclenche des fenêtres d'explications et de commentaires, l'appel d'une monnaie, d'un vase, ou d'un site qui complètent l'exposé. Toute une recherche est actuellement entreprise sur la notion de réseau pour développer ce type d'utilisation : il appartient aux spécialistes de construire les liaisons interactives qu'ils entendent établir entre les différents domaines de leur savoir. L'Antiquiste, et lui seul, est capable de définir les procédures autorisant une lecture non-linéaire des documents. Cela suppose un renouvellement complet de la discipline et que s'estompent les distinctions absurdes que les traditions académiques ont établies entre « littéraires », « historiens » et « archéologues ».

On peut aisément concevoir les facilités extraordinaires qu'un tel instrument apporte au chercheur et à l'enseignant. Il peut, à tout moment, convoquer sur l'écran les documents figurés qui sont associables à un texte ; autour de l'Iliade, il peut reconstituer tout un environnement d'illustrations et de documents archéologiques. Des réalisations telles qu'ITER ou PERSEUS ont encore une extension limitée ; mais leur intérêt fondamental est de rechercher l'exhaustivité ; ils accueilleront, si possible, la totalité des textes, la totalité des monuments figurés, les

cartes de tous les sites ; ils seront complétés par les inscriptions, par les collections de papyrus... La stratégie d'une telle entreprise implique des améliorations constantes et une constante re-interprétation des données.

François CHARPIN
Université PARIS 7 - UA 1423 CNRS

BIBLIOGRAPHIE

- ITER Unité Associée 1423 (Centre de Recherche en Documentation) du CNRS. Université de Lille 3 (G. Losfeld).
Université Paris 7 (F. Charpin).
- LASLA Université de Liège, Place du 20-août 32, B-4000 Liège.
- PERSEUS The President and Fellows of Harvard College and the
Annenberg/CPB Project. Harvard University.