

# L'ORDINATEUR DANS L'ENSEIGNEMENT DE LA LANGUE ET DE LA LINGUISTIQUE HISPANIQUE

Javier SANCHEZ

## 1 - LES MÉTHODES D'ENSEIGNEMENT DE LA LANGUE

C'est dans l'apprentissage des langues que le développement de l'informatique et surtout des systèmes multimédias semble le plus prometteur car il offre de nouvelles méthodes d'enseignement interactives individualisées et/ou de groupe. Ainsi, grâce aux outils multimédias, en particulier le couplage des CD-ROM et des cartes sons, on a vu paraître des logiciels qui mettent à la disposition des étudiants, textes, sons et images (même animées). L'auto-apprentissage et l'enseignement gagnent en qualité puisque la carte vocale et le CD-ROM permettent d'intervenir dans les deux niveaux d'apprentissage : l'oral et l'écrit. De plus, le renouvellement des méthodes traditionnelles, souvent considérées par les élèves comme fastidieuses, a permis un nouvel élan pédagogique très positif, aussi bien pour les enseignants que pour les « enseignés », grâce à ces nouveaux supports interactifs qui permettent d'élaborer des programmes plus structurés et conviviaux.

En effet, les progrès réalisés sur la numérisation du texte, du son et des images, grâce à l'interactivité proposée avec des commandes simples, ont apporté une meilleure aide à l'apprentissage de la grammaire aussi bien pour l'oral que pour l'écrit. Il est vrai que depuis longtemps le texte ou les images ainsi que les enregistrements sonores ou audiovisuels sont utilisés par les professeurs de langues vivantes. Mais l'originalité des outils multimédias est de pouvoir accéder directement et simultanément aux diverses composantes de l'apprentissage des langues (textes, sons et images), en proposant dans les logiciels des simulations de situations de

communication les plus proches du fonctionnement réel de la langue. Ainsi, par exemple, l'élève devant son ordinateur et équipé d'un casque et d'un micro peut suivre les instructions (questions ou exercices oraux ou écrits), réaliser des exercices de grammaire, corriger ses fautes et consulter les règles à la demande. Cette interactivité et cette convivialité mettent en puissance le travail de l'élève par rapport à l'apprentissage traditionnel (livres, dictionnaires, etc.). C'est-à-dire que l'étudiant est plongé dans un univers attractif dont il fait partie et dans lequel il joue un rôle important.

Pour cela, ce type de logiciels affiche à l'écran des textes ou des images et l'on demande oralement et/ou par écrit d'exécuter des consignes ou des exercices ; l'étudiant se voit ainsi conduit, par une interactivité dont le but est de mener le plus efficacement possible le déroulement de l'apprentissage, jusqu'à la fin de la leçon. C'est ainsi que l'association texte-image-son apporte en langue vivante une aide réelle et pratique dans l'apprentissage de la prononciation, de l'orthographe, du lexique et de la grammaire dans la mesure où il devient facile d'écouter et d'enregistrer sa voix (afin de la comparer), de lire les retranscriptions, de réaliser des exercices et de consulter les règles de grammaire qui les gèrent.

Nous donnerons un exemple en langue espagnole : le logiciel Langotec <sup>1</sup> s'adresse à ceux qui désirent apprendre à communiquer en espagnol et il est notamment utilisé, par certains enseignants, dans les lycées. C'est une méthode accessible aux vrais débutants conçue comme support d'un apprentissage autogéré et pouvant s'intégrer dans les cours de langue. Sa présentation et son fonctionnement rendent ce logiciel particulièrement facile et agréable à utiliser.

La durée de travail, entre 80 et 120 heures, dépend de l'utilisation totale ou partielle de toutes ses fonctions, ainsi que du niveau de l'élève. Le cours, divisé en 30 leçons, permet un parcours libre et l'élève peut à tout moment s'interrompre pour recommencer la leçon ou en refaire une partie.

Le logiciel aborde toute la combinatoire de base de la phrase simple et

---

<sup>1</sup> LANGOTEC, A.C.F. Didacticiels, 15 rue Antoine Thomas - 94200 Ivry sur Seine.

quelques éléments de la phrase complexe (les relatifs et le présent du subjonctif). Le noyau verbal occupe une place très importante car; en effet, la phrase espagnole se construit autour du verbe. Dans ce premier niveau sont traités l'indicatif (présent, passé simple, parfait, imparfait et futur), le présent du subjonctif, et l'impératif. Une attention particulière a été consacrée à la présentation systématique et détaillée des aspects les plus importants de l'irrégularité des verbes espagnols. *Haber, tener*, et surtout *ser y estar* sont traités de façon plus approfondie.

Le vocabulaire introduit peut constituer la base d'une communication variée dans la conversation de tous les jours (la famille, la maison, la ville, l'école, le travail, la banque, les transports, les voyages, la santé, le sport, etc.) et il contient environ 1500 mots (plus de 1000 noms, 300 verbes, 100 adjectifs et 80 adverbes) choisis parmi les plus fréquemment utilisés.

La présentation de tout le matériel pédagogique est strictement graduelle, aussi bien pour le Lexique que pour la Morphologie et la Syntaxe. Cette gradation rigoureuse et systématique fait de Langotec une méthode accessible aux vrais débutants. Basée sur l'étude de la structure de la phrase espagnole et des éléments de grammaire qui sont toujours introduits dans leur contexte.

Considérant la langue comme un système structuré et non comme une liste de mots ou de phrases, ce programme permet de ne pas faire apprendre par coeur à l'élève des expressions toutes faites, et lui fournit les ressources nécessaires pour qu'il puisse construire par lui-même les phrases. Chaque page de présentation du matériel pédagogique met en éveil chez l'apprenant sa façon d'inférer, généraliser, mettre en relation, répéter mentalement, mémoriser, afin d'activer les automatismes linguistiques de base : écouter, parler, lire et écrire. En effet, la répétition, la pratique et l'exercice constituent des points essentiels dans l'acquisition des automatismes. C'est pourquoi le logiciel propose à chaque étape une série d'exercices de lecture et d'écriture. Cette répétition n'est pas mécanique : le type d'exercice proposé permet de tester la compréhension de l'élève (structure, lexique et contexte) et il peut à tout moment revoir les écrans de présentation du matériel pédagogique.

premières tentatives l'orientation et l'assistance nécessaires pour résoudre les tâches qui lui sont proposées. Pour tous les mots de la leçon, l'élève peut consulter le dictionnaire intégré, c'est-à-dire, en tapant le mot inconnu (l'infinitif pour les verbes, le masculin singulier pour les noms et les adjectifs) il obtient immédiatement la traduction. Un autre intérêt est que l'élève peut aussi consulter la grammaire et cette fonction est à la fois un rappel en français de la terminologie grammaticale et un résumé des points de grammaire traités dans la méthode. Elle offre aussi des indications sur l'utilisation de certaines tournures et expressions idiomatiques.

Plusieurs possibilités d'écoute sont possibles, tant pour les exemples que pour les exercices. L'élève peut choisir d'entendre toute la page, ou bien chaque phrase séparément. A tout moment il peut arrêter le son, réécouter ou recommencer. Pour réentendre une phrase particulière dans un texte ou un mot dans une liste, il suffit de placer le curseur sur la phrase choisie et d'appuyer sur "Entrée". Pour toutes les phrases sonorisées, l'apprenant a la possibilité de s'enregistrer. Il peut se réécouter et comparer sa prononciation avec celle du modèle. Ce type d'exercice développe la capacité d'autocontrôle linguistique. Pour activer la fonction "Laboratoire de langues", il suffit d'appuyer sur une touche, l'élève entend le modèle et répète, puis il réentend le modèle et immédiatement après, sa propre voix. La convivialité de ces fonctions permet à l'élève de refaire ces exercices très facilement et de motiver ainsi l'apprentissage.

## 2. L'INFORMATIQUE POUR L'ÉTUDE LINGUISTIQUE

Outre l'aspect pédagogique de la langue espagnole à un niveau élémentaire, à partir des méthodes multimédias dont nous venons de parler, l'informatique a également beaucoup influencé les enseignements de plus haut niveau en linguistique hispanique, en particulier à l'université, où l'on a pu développer des filières spécialisées - allant jusqu'au doctorat - centrées sur les applications en « industries de la langue ». Nous en présenterons ici quelques aspects.

En effet, les principes méthodologico-informatiques utilisés par les étudiants de licence mais surtout de Maîtrise, DEA et Doctorat, ont permis d'élaborer de meilleures stratégies pour les études linguistiques sur grands corpus. Si une des nos principales préoccupations est d'accéder le plus parfaitement possible aux systèmes de relation qui sous-tendent la création du langage (morphosyntaxe, lexicque, typologie des discours, etc..) au plan linguistique, l'analyse devient véritablement difficile quand on veut explorer le texte de grande dimension, non seulement du point de vue structurel, mais également du point de vue syntaxique. L'étudiant-chercheur trouve dans les outils informatiques de véritables aides à l'étude et à l'exploration des textes enregistrés et stockés dans des banques de données textuelles.

C'est pourquoi, les propositions méthodologiques en linguistique-informatique ou d'analyse textuelle par ordinateur dont nous faisons état ici ont suivi et continuent à suivre les évolutions technologiques les plus récentes qui nous ont permis de progresser de façon significative dans l'élaboration d'une méthodologie pour l'étude des textes. Ainsi, les étudiants accèdent plus facilement à l'exploration et à l'apprentissage des lexicques et des structures fondamentales de la langue espagnole, et les plus avancés (Maîtrise, DEA, Doctorats) peuvent réaliser des recherches centrées sur l'analyse des corpus de grande dimension en double contextualité : structurelle et syntaxique.

C'est ainsi que nous nous sommes acheminés vers des procédures permettant, par alternance d'outils informatiques, d'effectuer des études textuelles de plus en plus performantes dont nous résumons ici les différentes étapes :

- 1) lecture scanner et reconnaissance automatique de caractères des textes à étudier pour la constitution ou l'amplification des banques de données textuelles sur support magnétique (disquettes ou CD-ROM) ;
- 2) utilisation des récents logiciels lexicométriques pour la création des listes alphabétiques (ordre alphabétique) ou hiérarchique (ordre de fréquence) ;
- 3) création des bases de données textuelles à partir de l'élaboration d'un logiciel de conversion de listes dans une de nos dernières recherches ;

- 4) utilisation des logiciels d'analyse statistique pour l'étude des typologies des discours (lexiques et structures spécifiques).

C'est ainsi qu'à partir de cet ensemble opérationnel, et en s'inscrivant dans la tradition lexicologique à travers l'utilisation des logiciels de lexicométrie et d'analyse statistico-lexicale, nos procédures tentent de mettre à profit du mieux possible les progrès technologiques liés aux évolutions les plus actuelles de la micro-informatique qu'il faut constamment mettre à jour.

Cet ensemble structuré d'outils a permis une lecture progressive et dynamique des textes et a rendu de plus en plus performante et conviviale l'étude des lexiques et des structures grammaticales des corpus enregistrés. Les perspectives de ces procédures qui avaient comme but, à l'origine, d'étendre au texte les limites de la linguistique traditionnelle généralement centrée sur l'étude de l'énoncé, ont pris une nouvelle dimension actuellement dans le cadre des "industries de la langue". En effet, les capacités de ces méthodes pour l'examen exhaustif et systématique des formes dans les textes ainsi que pour la constitution de bases de connaissances sont tout à fait adaptées à l'élaboration des dictionnaires à partir des bases de données. D'autre part, ces recherches devraient contribuer, nous l'espérons, aux applications en traitement automatique de la langue et de la traduction et par conséquent à l'enseignement assisté par ordinateur.

## **2.1. Elaboration des dictionnaires électroniques**

Il existe deux normes utilisées pour la constitution de dictionnaires : la norme prescriptive et la norme statistique. La norme prescriptive fait référence à tout ce qui a été défini comme obligatoire pour l'ensemble des êtres parlants d'une communauté linguistique alors que la norme statistique est de plus en plus appliquée en lexicographie parce qu'elle permet une représentation assez objective de l'usage du vocabulaire. En effet, on entend aujourd'hui par norme statistique celle qui est vérifiée statistiquement et qui est la plus étendue et acceptée en général par les membres d'une communauté linguistique. Elle a comme caractéristique

essentielle de représenter une coupe transversale synchronique de la langue, qui réunit à la fois ses influences historiques et les normes prescriptives acceptées. Il est très important de souligner ces deux types d'approches lexicographiques, car nous pouvons constater que nos applications en linguistique-informatique, et en particulier nos procédures descriptives et d'analyses statistiques, sont parfaitement bien adaptées à la constitution des dictionnaires électroniques sur CD-ROM selon la norme statistique qui est la plus courante actuellement et que la plupart des ouvrages lexicographiques prennent en compte pour la sélection des entrées des dictionnaires afin d'atteindre la meilleure description de la langue.

Ces dictionnaires pourraient fournir des informations comme par exemple :

- 1) l'usage des formes dans les différents types de discours (type de récit, dialogue, poésie, références aux auteurs, etc.) ;
- 2) les différences diatopiques, c'est-à-dire celles qui sont liées au lieu (variantes régionales, nationales, etc.). Les différences entre l'espagnol péninsulaire et celui d'Amérique en sont la meilleure illustration.
- 3) la différenciation diastratique relative aux différences établies dans les lexiques des divers groupes sociaux existant actuellement (langue orale).
- 4) la différenciation diaphasique qui fait référence aux styles linguistiques existants dans une langue, et employés dans certains contextes, tels que les langages : populaire, standard, familial, protocolaire, scientifique, etc.
- 5) les technicisms et les terminologies spécifiques aux différents secteurs : économique, commercial, politique, journalistique, scientifique etc., doivent être également pris en compte, ainsi que les néologismes.

## **2.2. Traitement automatique de la langue et de la traduction**

Il existe également un deuxième pôle de travaux qui concerne actuellement un champ scientifique et technologique reconnu : le traite

ment automatique des langues et de la traduction visant à améliorer les stratégies pédagogiques.

Des rencontres récentes (Congrès de la Sociedad Española para el Procesamiento del Lenguaje Natural, 16-18 septembre 1992, Université de Grenade) ont permis de dégager, outre la traduction et la lexicographie computationnelle, d'autres axes d'investigation tels que la gestion documentaire, le traitement de l'écriture et des documents écrits, et les interfaces de communication en langue naturelle entre l'homme et la machine. Ainsi, de plus en plus, des équipes de recherches se constituent autour de ces différents secteurs d'application. Nous citerons donc quelques projets en cours dans ce domaine. A titre d'exemple, aux Etats Unis, il existe un logiciel de reconnaissance de la parole pour la langue allemande intitulé "Dragondictate" élaboré par A. Demedts (Dragon Systems Inc., Boston, U.S.A.). De même, un système d'extraction de contenu textuel, construit par Current Institut of Mathematical Sciences (New York University) a été lancé sur le marché des industries de la langue. En France, J. Vergne du Laboratoire Informatique de l'Université de Caen, a développé une grammaire par blocs qui met en oeuvre une stratégie d'analyse séquentielle permettant de traiter le français. En Espagne, une équipe de l'Université de Grenade travaille sur la reconnaissance automatique de la parole en collaboration avec le Département d'Electronique et de Technologie Informatique de cette même université. Par ailleurs, les applications de l'analyse statistique de V. Sabido (Département de philologie espagnole de l'Université de Grenade) et de F. Garcia (Département d'Algèbre de l'Université de Grenade) ont permis d'amorcer la création d'un logiciel de désambiguïsation morphologique italo-espagnol au Département des Langues de l'Université de Malaga. Enfin, à Barcelone, grâce au partenariat entre facultés et entreprises, l'Université de Barcelone et le Centre de Technologie de la Langue de l'Université de Séville ont associé leurs compétences afin de construire un dictionnaire catalan des synonymes.

Mais nous ne sommes pas encore capables actuellement, en ce qui concerne la traduction et le traitement automatiques de la langue espagnole, de proposer de véritables outils performants et exhaustifs. Nous pouvons donc dire que ce champ d'investigation est loin d'être épuisé et qu'il a encore devant lui de nombreuses années de recherches.

Mais nous constatons déjà que les dernières rencontres internationales mettent en évidence que les industries de la langue se nourris

sent des avancées fondamentales de la linguistique-informatique. Cela confirme donc que "les laboratoires universitaires et les filières d'enseignement, sans délaisser la voie de la théorie, se positionnent favorablement dans celle de l'opérateur". Dans ce domaine, nos travaux actuels se caractérisent par l'intégration des différentes composantes de la grammaire (morphologie, syntaxe, sémantique, typologie des discours et des textes) dans un nouvel univers linguistico-informatique multidimensionnel capable de les mettre en relation et dont le but est de construire progressivement une grammaire opératoire pouvant servir à l'élaboration d'outils multimédias de plus en plus performants.

Ces recherches devraient apporter aux "industries de la langue" et aux enseignements universitaires de nouvelles perspectives d'application. En effet, dans ce cadre, l'élaboration des bases de connaissances comportant les différentes possibilités morpho-syntaxico-typologiques des formes permettrait d'alimenter des thésaurus utilisables dans l'enseignement multimédia et dans la Traduction Assistée par Ordinateur (T.A.O.). Il s'agit donc de proposer, à partir d'une réflexion sur les principes de synthèse, un ensemble linguistico-informatique conjointement adapté pour la création des dictionnaires et l'élaboration progressive des grammaires textuelles, destiné à faciliter les différentes applications dans les "industries de la langue" : assistantat dans la traduction, dans l'apprentissage de la prononciation, et dans l'acquisition du lexique et de la grammaire.

Javier SANCHEZ,  
Université de Limoges  
Département d'Espagnol