

RÉALISATION D'UN CDROM/DVD CONTENANT DES DONNÉES DU LANGAGE ORAL ORGANISÉES EN PARCOURS DIDACTIQUES D'INITIATION LINGUISTIQUE A PARTIR DES BASES DE DONNÉES VALIBEL

Introduction

Une des spécificités de l'enseignement de la **linguistique française à l'Université catholique de Louvain** consiste à aborder la langue dans une perspective sociolinguistique¹ et fondée sur l'analyse de corpus. Quelle que soit la sous-discipline enseignée (phonétique, lexicologie, syntaxe, pragmatique), les activités d'apprentissage visent un double objectif (très général) :

- l'appropriation de concepts et de théories linguistiques
- l'acquisition, par les étudiants, d'un savoir-faire (pratique et cognitif) touchant à l'analyse et à l'exploitation de données attestées, le plus souvent des productions orales ou écrites en français².

Du point de vue des enseignants, la mise en œuvre effective d'une initiation à l'analyse de corpus, voire la simple recherche d'exemples de phénomènes langagiers attestés pour rendre les cours plus concrets, demande de grands efforts et consomme beaucoup de temps. En outre, il n'est pas toujours simple, techniquement parlant, de fournir les données sonores et textuelles aux étudiants³, ni de leur donner accès aux logiciels d'analyse linguistique.

1. Objectifs

Le développement d'un **outil multimédia (CD-ROM/ DVD)** fournissant des données linguistiques (enregistrements, transcriptions, analyses acoustiques, concordances lexicales) et des logiciels pour les analyser, vise trois objectifs : (1) fournir à l'enseignant un matériel adéquatement présenté et organisé pour illustrer ses cours à l'aide d'exemples de phénomènes langagiers attestés ; (2) permettre aux étudiants de réaliser des exercices ou des travaux personnels sur des corpus oraux et écrits ; (3) initier les étudiants à l'utilisation de logiciels d'analyse linguistique, voire leur fournir un moyen d'auto-apprentissage de certaines notions linguistiques. Ces trois objectifs sont détaillés ci-dessous.

(1) À l'usage de l'enseignant, ce CD-ROM proposerait divers **types de ressources langagières**, organisées en vue de leur exploitation didactique, comme :

- documents sonores (fichiers sons numérisés de type .wav)
- transcriptions orthographiques d'interactions verbales (quotidiennes, professionnelles, scolaires)
- documents sonores alignés sur une transcription textuelle (orthographique ou phonétique)

¹ Étudier la variation (sociale, géographique, stylistique) du français dans des interactions langagières effectives, pourvues d'enjeux sociaux.

² Le centre de recherche VALIBEL a investi une partie de ses recherches méthodologiques dans le développement de formats adéquats pour constituer une banque de données textuelles orales, qui contient des productions de français en Belgique, mais pas seulement. Cet acquis non négligeable sera réinvesti dans le présent projet.

³ L'analyse acoustique ne tolère pas les formats compressés de type mp3 mais impose l'utilisation de fichiers sons de type .wav. Un enregistrement d'une demi-heure (22050 Hz, 16 bits, mono) « pèse » quelque 80 Mo, ce qui nécessite une très bonne connexion Internet pour être téléchargé à domicile dans un délai raisonnable – et rend l'usage d'un CD-ROM plus attractif.

- transcriptions lemmatisées et étiquetées morphosyntaxiquement (recherche complexe par un concordancier)
- base de données d'informations sociolinguistiques (sous la forme de fiches d'identification) relatives aux locuteurs/scripteurs ayant produit les données (localisation géographique, niveau d'étude, profil socio-professionnel, etc.) et aux contextes d'interaction (nombre de locuteurs, degré de formalité, etc.)
- bases de données (lexicales, syntaxiques, sur l'intonation régionale [« l'accent »], sur la liaison...)

Ces données langagières (textes et sons) et les bases de données associées seront consultables selon différents critères et présentées de manière conviviale.

(2) À l'usage de l'étudiant, les données pourront servir de base à des **exploitations** linguistiques ou sociolinguistiques (exercices ou réalisation de travaux personnels). Elles seront analysables à l'aide de divers **logiciels libres** :

- logiciel d'analyse acoustique (Praat)
- logiciel d'écoute des données sonores (Audacity ; CoolEdit96)
- feuille de calcul (Open Office)
- analyseur morphosyntaxique et concordancier (Unitex)
- étiqueteurs et codeurs pour des phénomènes linguistiques précis (utilitaires développés dans le cadre du projet *Phonologie du Français Contemporain*)
- gestionnaire de bases de données (phpMySQL) permettant de faire des corrélations sociolinguistiques entre les phénomènes langagiers observés et les profils sociaux des locuteurs

(3) En lien avec le second objectif, on développera des **outils didactiques** de trois types :

- des modes d'emplois interactifs (**tutoriaux**) pour initier les étudiants à l'utilisation des logiciels, accompagnés de séquences d'exercices ;
- des **parcours didactiques** pour la réalisation d'un travail en linguistique de corpus. Ces parcours permettront aux étudiants de formuler des hypothèses afin d'exploiter les données à l'aide des outils fournis, en vue d'atteindre les objectifs spécifiques (formulés en termes de compétences) des cours concernés⁴ ;
- éventuellement, on pourra créer certaines séquences pour l'**auto-apprentissage** de notions linguistiques ciblées (l'accentuation, l'intonation, certaines constructions syntaxiques, etc.).

2. Utilisation

Le CD-ROM contiendra des données de français parlé (de Belgique, mais aussi d'autres pays francophones, afin de permettre les analyses contrastives et variationnistes). Ces données ne doivent pas être recueillies : elles sont déjà incluses dans la banque de données textuelles orales VALIBEL⁵. Un nombre suffisant d'enregistrements sonores variés existe, à partir desquels on constituera un échantillon d'enregistrements, représentatif du point de vue des locuteurs et des situations d'interactions (quotidiennes, scolaires, professionnelles).

Le CD-ROM sera **intégré** variablement dans les enseignements :

- là où les données langagières viennent à titre d'exemple, l'usage du CD-ROM ne fera pas l'objet d'une activité d'apprentissage particulière, mais il constituera un moyen de rendre le cours plus concret et plus attractif en l'illustrant avec des phénomènes

⁴ Le point 4 présente un exemple de parcours didactique sur l'accent régional faisant appel à différentes facettes du CD-ROM.

⁵ Dont le CD-ROM constituera par ailleurs une valorisation intéressante et inédite (pédagogique).

authentiques ; les étudiants pourront consulter les exemples chez eux en travaillant sur les cours ;

- là où l'apprentissage de l'utilisation d'un logiciel de traitement du langage fait partie des objectifs du cours (savoir-faire pratique à acquérir), le CD-ROM sera partie intégrante du dispositif d'apprentissage et constituera le support des activités permettant l'acquisition des savoir-faire.

Ce n'est que dans les cours de spécialité et les séminaires (filière linguistique) que la **charge de travail de l'étudiant** sera sensiblement modifiée par le CD-ROM et les activités qu'il permet. Mais cette modification s'accorde avec la réforme de Bologne (où on comptabilise le poids des cours en fonction du « travail-étudiant »), car un tel dispositif permettrait de diminuer le nombre d'heures en présentiel pour certains cours, en permettant l'auto-apprentissage de compétences ciblées, ainsi que des activités d'appropriation des techniques.

3. Aspects didactiques

Une utilisation autonome du CD-ROM par les étudiants nécessite la conception de tutoriaux liés aux logiciels libres utilisés, et de parcours didactiques donnant lieu à différentes activités d'apprentissage (exercices et travaux) prévues par l'enseignant (se référer aux objectifs spécifiques des cours concernés). Les logiciels mis à disposition seront minimalement au nombre de 4 : Praat : analyse acoustique, Intex : analyse syntaxique et concordancier, socioDB : base de données sociolinguistiques, prosoDB : base de données dynamique.

Les tutoriaux consisteront au minimum de textes accompagnés d'images (captures d'écran) ; au mieux, ils seront dynamiques, utilisant par exemple les ressources du logiciel Director.

Chaque tutorial visera à expliquer à l'étudiant comment se servir d'un logiciel ou comment acquérir une technique d'analyse, et à quelles fins. Le CD-ROM contiendra à cet effet des **exercices-types** organisés en fonction des domaines de l'analyse linguistique :

- transcription de données orales : exercices (+ corrigés) permettant d'assimiler les différentes conventions de transcription propres à l'analyse de données orales (par ex. utile pour le séminaire ROM 1171 et pour réaliser un mémoire en linguistique de corpus)
- phonétique et prosodie : exercices pour la segmentation phonétique ; la transcription ; la représentation phonologique et la perception des différents types de tons ;
- lexique : concordancier pour retrouver des formes isolées (mots simples ou complexes) ou des expressions figées, en disposant automatiquement du contexte d'apparition de ces occurrences ;
- syntaxe : portions de textes étiquetées (analyses de type *parts of speech*) et possibilité de construire des graphes simples pour retrouver des formes syntaxiques particulières
- analyse du discours et des interactions verbales : données de différents types interactifs pouvant servir de base à des analyses conversationnelles, de la structuration des échanges, etc. ; informations sur les contextes de ces interactions ; alignement de la transcription avec le son (le cas échéant : avec l'image) pour intégrer les aspects para-verbaux (gestes, regards) aux analyses

Le CD-ROM tel que conçu dans ce projet ne permet pas un auto-apprentissage complet de la part de l'étudiant (il ne remplace pas l'assistance aux cours), car dans cette première phase, on ne prévoit pas de présenter suffisamment de notions théoriques linguistiques de manière exhaustive. On pourrait imaginer, vu la demande qui va dans ce sens, de poursuivre le projet par une deuxième phase qui ajouterait du contenu théorique organisé en vue d'un auto-apprentissage.

4. Exemple de parcours didactique

Thème : Analyse de la variation régionale de l'intonation (les accents) en français.

Objectifs :

Au terme de ce parcours, les étudiants seront capables de

- *distinguer* les différents aspects de la prononciation responsables de la perception d'un accent régional (accent suisse, accent québécois, accent de Liège...) : les aspects segmentaux (qui concernent les sons isolés, par ex. dire *sévère* ou lieu de *sévère*) et supra-segmentaux (rythme, accentuation, intonation, par ex. parler avec un accent traînant ou chantant...)
- *faire* une analyse auditive (à l'oreille) de l'intonation régionale d'une région de la francophonie
- *faire* une analyse acoustique (instrumentale) de certains aspects de l'intonation régionale (allongements, contours intonatifs typiques)
- *décrire* de manière synthétique l'accent d'une région par rapport à la prononciation du français standard (norme et variation)

Ces objectifs sont du niveau d'un cours spécialisé ou d'un séminaire de niveau BAC3 ou licence/master.

Activités d'apprentissage :

Le parcours didactique se présente comme suit (il fonctionne sur la base d'une interaction entre des séances de cours et des périodes de travail individuel ou en groupe avec le CD) :

- initiation à percevoir les accents :
 - les étudiants recherchent sur le CD-ROM des locuteurs qui présentent un accent régional marqué (le formulaire de recherche permet de chercher des locuteurs en fonction de leur âge, origine géographique, niveau d'éducation)
 - ils écoutent des extraits de ces conversations et tentent de manière intuitive de dire ce qui « fait l'accent » d'une personne
 - au cours magistral, on distingue les aspects segmentaux et supra-segmentaux
- analyse auditive + initiation à l'intonation
 - à l'aide d'exercices ciblés, les étudiants s'exercent à percevoir l'intonation : quand est-ce que la voix « monte », « descend » ? quelles syllabes sont accentuées ? Ces exercices seront sur le CD-ROM et pourront être faits par les étudiants autant de fois que nécessaire.
 - au cours, on établit les notions théoriques.
- analyse acoustique
 - une fois les particularités régionales de l'intonation repérées auditivement (et annotées dans les fichiers sons), les étudiants réalisent une analyse acoustique à l'aide du logiciel Praat
 - le CD-ROM, qui fournit le logiciel libre aux étudiants, contient aussi un tutorial interactif leur permettant de s'initier à son utilisation en fonction de certains objectifs précis (mesurer des durées de syllabes, mesurer les différences de hauteur intonative dans un énoncé, etc.)
 - possibilité de faire une représentation graphique de l'intonation (prosogramme)
- rédaction d'un travail de synthèse (en groupe) qui compare l'accent d'une région donnée à celui du français standard

On peut aussi imaginer des parcours plus simples, qui donnent lieu à un travail plus ciblé de la part des étudiants (un seul objectif spécifique au lieu des quatre ici prévus). Les 3 premiers points de ce parcours peuvent eux-mêmes se décomposer en parcours plus petits.

5. Conclusion

Cet outil pédagogique permet une pédagogie active et un apprentissage quasi-autonome qui tient compte de la massification des étudiants et de leur hétérogénéité.