

D'un corpus d'évaluation diagnostique à l'élaboration d'un corpus d'enseignement, accompagnement des déplacements intellectuels et identitaires de futurs enseignants de français

Elalouf, Marie-Laure

Laboratoire ÉMA

Cergy-Paris Université

Pour citer cet article :

Elalouf M.-L. (2024). D'un corpus d'évaluation diagnostique à l'élaboration d'un corpus d'enseignement, accompagnement des déplacements intellectuels et identitaires de futurs enseignants de français. *Scolagram* n°11, novembre 2024. Enjeux des corpus en didactique de la grammaire pour la formation des enseignants. https://scolagram.u-cergy.fr/index.php/content_page/item/1356-d-un-corpus-d-evaluation-diagnostic-a-l-elaboration-d-un-corpus-d-enseignement-accompagnement-des-deplacements-intellectuels-et-identitaires-de-futurs-enseignants-de-francais

Résumé : Cet article expose comment les conceptions de futurs enseignements de Lettres ont été prises en compte dans l'élaboration d'un cours à distance, visant leur formation grammaticale et didactique. L'analyse des discussions grammaticales générées par une situation-problème permet un diagnostic tandis que des priorités se dégagent de l'examen des écrits universitaires produits par des étudiants suivis de la préprofessionnalisation à leur prise de fonction. Ces conclusions nourrissent un scénario de formation qui travaille conjointement les dimensions institutionnelle, cognitive, médiative, personnelle et sociale de la professionnalité. Fourni in extenso en annexes, le cours à distance portant sur l'utilisation des grands corpus et la conception de corpus d'enseignement est commenté. Le retour réflexif sur celui-ci met au jour la nécessité d'explicitier et de justifier les critères d'élaboration des corpus d'enseignement et de formation.

Mots clés : formation des enseignants, conceptions initiales, activité métalinguistique, corpus d'enseignement, corpus de formation, CAPES de Lettres, mastérisation, enseignement à distance.

L'appel à contribution de *Scolagram* offrant la possibilité de présenter des dispositifs didactiques commentés, j'ai choisi dans cet article de faire un retour critique sur la façon dont j'ai abordé l'élaboration de corpus pour l'étude de la langue en formation des enseignants du second degré entre 1991 et 2021. Les annexes auxquelles le propos se rapporte sont des extraits d'un scénario d'enseignement à distance, lui-même précédé et accompagné par des enseignements en présentiel (ANNEXES 1 à 4).

Rouvrir ses propres archives ébranle les allant-de-soi. Quels choix ont été dictés par le contexte sans recul suffisant et quels autres ont été plus réfléchis ? En essayant de ne pas céder à l'illusion rétrospective, je reviendrai sur la rédaction d'un cours de 1^e année de master MEEF2 Lettres¹ pour l'enseignement à distance traitant de la question des corpus dans deux unités d'enseignement (dorénavant UE) complémentaires. Il s'agira de présenter ces corpus conçus pour la formation initiale des enseignants en expliquant ce qui a présidé à leur élaboration (1), le contexte dans lequel ils ont été mis en œuvre et les scénarios de formation dans lesquels ils se sont inscrits (2). Cette présentation permettra d'ouvrir une discussion sur le statut des corpus en linguistique et en didactique (3). À titre provisoire, nous définirons le corpus comme un ensemble de données langagières pouvant servir d'échantillon du langage selon certaines conditions que l'article cherche à répertorier et caractériser.

1. La genèse des corpus de formation proposés

La préparation à l'épreuve de grammaire et stylistique et à l'épreuve sur dossier du CAPES² m'a conduite dans les années 1990 à être attentive à l'expérience scolaire mobilisée par les étudiants, source de représentations, d'automatismes, de formules et d'exemples prototypiques, et à m'interroger sur les résistances qu'elle suscitait à la compréhension de démarches issues des recherches linguistiques et à la description des usages langagiers. Cette question m'a semblé nécessiter des enquêtes (Elalouf, Benoit et Tomassone, 1996 ; Elalouf, 2023) et au-delà des avis déclarés, des recherches sur la base de corpus. Je montrerai ce qu'éclaire la comparaison de corpus d'écrits d'élèves et d'étudiants sur une même tâche (1.1), le suivi longitudinal d'étudiants jusqu'à leur prise de fonction (1.2.) et j'expliquerai comment cela m'a permis d'élaborer un corpus diagnostique pour le recueil de leurs conceptions (1.3).

1.1. Convergences entre corpus d'élèves et d'étudiants

C'est en réunissant des collections d'écrits d'élèves et d'étudiants à des fins de recherche que s'est nourrie la réflexion sur la constitution de corpus pour la formation des

¹ MEEF2 : Métiers de l'enseignement, de l'éducation et de la formation second degré. Il s'agit ici de la formation des étudiants de Lettres modernes qui enseigneront le français au collège et au lycée général ou technologique en France.

² Entre 1991 et 2008, les étudiants préparaient après la licence un concours d'enseignement en 1^e année d'IUFM avant de suivre une année de stage en 2^e année. À côté des épreuves académiques du CAPES (littérature, grammaire et stylistique), l'épreuve sur dossier avait une visée de préprofessionnalisation (Elalouf, 2023).

enseignants et l'enseignement. L'étude de corpus scolaires et universitaires permet en effet de remonter aux besoins de formation des futurs enseignants et de mesurer l'écart entre une formation cloisonnant littérature, linguistique, didactique et les exigences d'une lecture-interprétation co-construite avec les élèves en tenant compte de leurs hypothèses initiales. Pour préparer la rencontre entre une classe et la matérialité d'un texte ouvert, on attendrait que le futur enseignant soit en mesure de « construire une interprétation défendable d'un texte en conjuguant des observations relevant de différents niveaux d'analyse sans penser dans un rapport d'exclusion étude lexicale, grammaire de phrase et linguistique textuelle » (Elalouf et Tomassone, 2001, p. 396).

Or, en proposant la lecture d'un même poème, accompagné d'une contextualisation historique et culturelle, à une classe de seconde et à des étudiants préparant le CAPES, Roberte Tomassone et moi avons été frappées de constater, en comparant leurs textes, combien leur compréhension du genre scolaire *commentaire de texte* était informée par des représentations convergentes de la langue : une approche référentielle qui privilégie le relevé des noms et néglige celui des verbes ou des connecteurs, une pratique citationnelle peu attentive aux découpages syntaxiques, une absence de prise en compte de la phrase comme instance de production du sens, une relative insensibilité aux enchaînements et ruptures thématiques qui configurent les plans énonciatifs ainsi qu'aux indices de la « voix auctoriale » (op. cit. p. 402). Ce que les uns et les autres ont retenu de l'enseignement grammatical ne semble pas pouvoir être convoqué pour élaborer la signification d'un texte.

1.2. Enseignements tirés du suivi longitudinal de futurs enseignants

On est en droit d'objecter qu'un corpus recueilli dans le cadre de la préparation à une épreuve très codifiée offre une photographie biaisée des compétences linguistiques et métalinguistiques de futurs enseignants. Les recherches sur les littératies universitaires (Dabène et Reuter, 1998 ; Defays, Maréchal et Mélon, 2000 ; Laborde-Milaa, Boch et Reuter, 2004) m'ont invitée à adopter un empan plus large, dépassant l'identification d'habiletés linguistiques isolées ou de compétences transversales disjointes d'un champ disciplinaire et des genres discursifs qui le constituent. C'est ainsi que des étudiants de Lettres ont été suivis entre 1997 et 2007, depuis le module de sensibilisation aux métiers de l'enseignement (en 2^e année de Licence) jusqu'à leur prise de fonction avec, au départ, deux hypothèses :

- (1) Il y a une évolution des représentations et une recomposition des connaissances sur le langage, la langue et les discours à mesure que la perspective de l'enseignement se rapproche.
- (2) Mais il y a des points de résistance dont certains sont communs à tous les futurs enseignants de français et d'autres liés aux particularités de leur itinéraire. (Elalouf, 2007, 2009).

L'ensemble des écrits académiques et professionnalisants de ces étudiants a été soumis à une analyse de discours afin d'étudier, à la faveur des récurrences spécifiques à chaque corpus, comment évoluent les représentations sur le langage, la langue, comment se manifestent leurs compétences langagières en réception et en production, quels obstacles

et quelles dynamiques sont repérables dans le développement des compétences métalinguistiques.

En confrontant les analyses menées sur chaque suivi d'étudiant pour dégager des régularités, il n'a pas été possible de valider la première hypothèse. Ce sont plutôt les caractéristiques d'étapes clés du cursus, diversement investies et mises en relation, qui permettent de rendre compte de la nature des transformations : le premier écrit long pour les étudiants ayant fait une maîtrise, la préparation de l'épreuve de français moderne avec ses limites et le mémoire professionnel³. En revanche, la seconde hypothèse a pu être confirmée. Des points de résistance sont apparus comme autant de freins à la recomposition de connaissances grammaticales antérieures et des apports universitaires : la fossilisation (Trévisse, 2001) des premières descriptions grammaticales introduites à l'école primaire et rarement reprises dans la formation ultérieure comme la définition du mot ou de la phrase ; la prégnance d'une approche linéaire de la phrase et du texte occultant leur organisation ; l'illusion d'une transparence terminologique donnant accès à la description du fait de langue concerné « comme son nom l'indique » ; le malentendu sur les manipulations souvent assimilées aux moyens mnémotechniques introduits en début de scolarité.

L'étude de ce grand corpus universitaire (Elalouf, 2007, 2009) m'a permis d'anticiper sur les résistances à travailler en priorité dans différents contextes : cours de prérentrée pour des étudiants s'inscrivant au concours d'enseignement (CAPES ou agrégation) sans formation grammaticale, cours de didactique pendant la première mastérisation (2010-2013).

1.3. Conception d'un corpus diagnostic

J'ai ainsi expérimenté pendant plusieurs années un corpus forgé, proposé en début de formation. Conçu pour recueillir et travailler les conceptions initiales des étudiants, il avait une consigne très ouverte :

Comparez ces deux énoncés et faites toutes les remarques que cette comparaison vous inspire.

1-Tout à coup, on a vu une sorte de serpent.

2-Quelqu'un crut soudain apercevoir un serpent à sonnette.

L'exercice permettait de travailler la notion fossilisée de *mot vs unité lexicale, unité grammaticale*, de voir comment était appréhendée la relation *forme/sens*, de constater quel niveau d'analyse de la langue était privilégié ou occulté. Après un temps de réflexion individuelle, les étudiants confrontaient par groupe de 4 leurs remarques, notaient leurs points d'accord et de désaccord, puis un seul groupe présentait ses observations mais

³ La recherche a été menée avant la mastérisation : les étudiants préparaient le CAPES après la Licence, mais certains pouvaient avoir fait une maîtrise avant. L'épreuve de français moderne comportait des questions de grammaire sur un texte littéraire postérieur à 1500 et une analyse stylistique. Le mémoire professionnel était fait pendant l'année de stage, généralement encadré par des formateurs du second degré.

l'ensemble des étudiants étaient invités à les discuter. Il reste peu de traces de ce travail, l'essentiel étant que les étudiants sortent du corset de l'étiquetage et s'autorisent à entrer dans une argumentation. J'avais en effet constaté dans l'étude précédente que des sauts qualitatifs se produisaient quand les étudiants faisaient feu de tout bois pour décrire et discuter une construction résistant aux outils d'analyse dont ils disposaient.

En dépit de l'absence d'archives de ces interactions, il subsiste la mémoire des remarques les plus fréquentes. La première portait généralement sur le niveau ou registre de langue, selon la terminologie qu'avaient rencontrée les étudiants, et permettait d'ouvrir la discussion autour des oppositions suivantes : courant/soutenu ? familier/soutenu ? oral/écrit ? récit conversationnel/genre narratif ? Venaient ensuite des remarques sur le sens sous l'angle du lexique : *on retrouve le même mot, il y a des synonymes, ça veut dire à peu près la même chose*. Mais suffit-il d'employer des synonymes pour que les deux énoncés soient à peu près équivalents ? L'idée que les deux phrases étaient construites de la même façon n'émergeait pas immédiatement. Une fois énoncée, il s'agissait de savoir comment faire pour vérifier l'hypothèse, ce qui supposait de mobiliser une manipulation non pas de façon mécanique mais dans un but euristique. Le travail sur la commutation permettait de poser des équivalences entre des mots, entrant dans les mêmes classes grammaticales, constitués d'un ou plusieurs mots graphiques, partageant certaines propriétés morphosyntaxiques (Beumanoir-Secq, Cogis et Elalouf, 2010).

La question la plus délicate à traiter était celle de la différence de sens entre les deux énoncés : cela supposait de caractériser deux formes de modalisation *une sorte de + nom* et *croire + verbe à l'infinitif* peu décrites dans les grammaires et jamais mises en relation ; cela conduisait aussi à comprendre que les syntagmes *une sorte de serpent* et *un serpent à sonnette* bien qu'apparemment construits sur le même modèle relevaient de deux constructions différentes, ayant toutes deux pour noyau le nom *serpent*. Ce travail, calibré pour rythmer une séance d'introduction et poser des jalons pour des approfondissements ultérieurs, a gagné à être proposé plusieurs années consécutives, la mémoire des discussions antérieures permettant une ouverture plus grande à des analyses imprévues et une confiance affermie dans la capacité des étudiants eux-mêmes à apporter des arguments dans le cadre ainsi posé.

2. Contexte institutionnel et choix de formation

Initié lorsque le CAPES se préparait en 1^e année d'IUFM, aménagé pendant la première mastérisation (2008-2013) qui a déporté le concours en fin de master, le travail va se poursuivre dans un autre format avec la seconde mastérisation (2014-2020). En effet, les premiers signes du reflux des candidatures se manifestent déjà et pour élargir l'offre et maintenir la formation à l'université de Cergy-Pontoise, il a été décidé de la proposer à distance. Je présenterai d'abord les modifications de maquette induites par celles du concours (2.1.) puis les scénarios de formation conçus dans le cadre d'une UE d'initiation à la didactique du français (2.2) et d'une nouvelle UE de didactique de la grammaire (2.3). Les cours présentés ici et fournis en annexe sont un complément au premier chapitre de l'UE « Histoire et épistémologie de la discipline *français* » et aux deux derniers chapitres de l'UE « Didactique de la grammaire pour l'écrit ».

2.1. La seconde mastérisation

Avec l'institution des ESPÉ⁴, le concours est ramené en fin de M1 ; il a une dimension professionnalisante afin de préparer les lauréats à un stage en responsabilité à mi-temps l'année suivante. Dans ce cadre, l'épreuve de grammaire et stylistique⁵ évolue : le quart de la note est réservé à une question de « mise en perspective des savoirs grammaticaux » qui s'appuie sur les deux textes littéraires proposés (ancien français et français moderne) ainsi que sur des extraits de manuels. Il s'agit de présenter les éléments d'une réflexion pédagogique sur une question grammaticale dans la perspective de l'enseignement à un niveau donné. Chacune des deux épreuves d'oral – l'explication de textes et l'épreuve sur dossier – prévoit une réflexion sur l'adaptation de l'enseignement à un public et à un contexte professionnel réel. La traduction dans les maquettes universitaires a consisté, dans l'académie de Versailles, à maintenir l'UE « Histoire et épistémologie de la discipline *français* » en première année de master et à proposer une UE de didactique de la grammaire pour l'écrit aux premier et second semestres de M1 et une UE de didactique de la grammaire pour l'oral au second semestre de M1.

2.2. Concevoir un scénario de formation pour réfléchir à l'hétérogénéité de la discipline *français*

Avec ses 4 séances synchrones et la supervision du travail asynchrone (pour un total de 12h TD), l'UE « Histoire et épistémologie de la discipline *français* » ne permet qu'une initiation. Le premier chapitre pose des éléments de définition sur ce qu'est une discipline scolaire, en s'appuyant sur l'article « discipline scolaire » du *Dictionnaire des concepts fondamentaux des didactiques* (Reuter et al., 2007) illustré d'exemples pris aux programmes et à des manuels. Il est scandé par des questions qui invitent l'étudiant à convoquer son expérience scolaire comme élève mais aussi comme assistant d'éducation, intervenant en soutien scolaire, stagiaire en observation selon les situations. Ces questions débouchent sur des consignes pour les échanges asynchrones qui précèdent la séance synchrone.

| Date | Lecture préalable | Travail individuel préalable | Travail collaboratif préalable | Travail pendant la connexion |
|-------|--------------------|--|--------------------------------|---|
| 23/09 | Lire le chapitre 1 | Retour réflexif sur votre expérience : court texte à envoyer à tous sur le forum avant la connexion synchrone. | Mise en situation | Confrontation des analyses proposées dans les deux activités préalables |

Tableau 1 : Extrait du plan de travail envoyé aux étudiants en début de semestre

⁴ Loi du 8 juillet 2013 d'orientation et de programmation pour la refondation de l'École de la République.

⁵ Elle est portée à 6h avec un coefficient égal à celui de la composition française. Cette disposition est supprimée en 2020.

Travail individuel préalable : Après avoir lu le chapitre 1, et notamment le point 3. 3. *une discipline traversée de tensions*,

- vous essaieriez d'expliquer en quelques lignes à quelle(s) tension(s) vous avez été le plus sensible tant qu'élève, en donnant un ou plusieurs exemples précis.
- vous évoquerez en quelques lignes un souvenir scolaire précis : en cours de français (précisez la classe et la dominante), vous avez proposé un raisonnement valide de votre point de vue mais qui ne correspondait pas aux attentes du professeur. Analysez le malentendu en vous plaçant successivement du point de vue de l'élève que vous avez été et de l'enseignant que vous souhaitez devenir.

Tableau 2 : Consigne de travail individuel

Les écrits réflexifs individuels avaient pour but de faire réfléchir aux différents régimes de rationalité de la discipline, comme source possible de malentendus. Ils pouvaient être mis en relation avec le travail collaboratif mené par groupe de 2 à 4 étudiants.

Travail collaboratif préalable : *Cette mise en situation a pour but de vous sensibiliser aux différents types de raisonnements que vous mettez en œuvre quand vous faites des observations sur la langue et aux niveaux d'analyse dont ils relèvent.*

- Observez ce couple de phrases.
 1. *Tout à coup, on a vu une sorte de serpent.*
 2. *Quelqu'un crut soudain apercevoir un serpent à sonnette.*
- Mettez en relation des segments de la phrase 1 avec des segments de la phrase 2.

>>> Quelles remarques pouvez-vous faire ? De quel niveau d'analyse relèvent-elles ?
- Comparez ces deux énoncés dans leur globalité.

>>> Quelles remarques pouvez-vous faire ? De quel niveau d'analyse relèvent-elles ?
- 1^{ère} étape : Recherche en groupe.
- 2^e étape : Pendant la connexion synchrone, confrontation de vos résultats et mise en évidence des savoirs mis en jeu et des procédures utilisées.
- 3^e étape : Une récapitulation qui sera donnée ultérieurement pour prendre en compte vos remarques et questions.

Tableau 3 : Consigne de travail collaboratif

Passer d'un dialogue grammatical en présentiel à l'élaboration d'un scénario pour la distance ne se fait pas sans tâtonnements. Selon les années et les groupes, il a été plus ou moins facile de s'appuyer sur un travail préalable de tous. La nécessité s'est fait sentir de cadrer davantage le travail asynchrone alors que la dynamique de la discussion grammaticale en présentiel reposait sur une consigne ouverte. L'autre moyen de soutenir l'engagement des étudiants a été d'inscrire le travail sur corpus dans une temporalité, la première et la seconde étape permettant un diagnostic et conduisant dans une troisième étape à un apport de connaissances contextualisé à différentes situations d'enseignement : l'évaluation d'une production écrite, la réflexion sur la norme selon les genres textuels, l'activité grammaticale. C'est ce volet qui figure en annexe (ANNEXE 1a). Il s'agit là de reprendre la notion de malentendu évoquée dans les souvenirs scolaires de

façon affective pour commencer à clarifier les différents registres épistémologiques sur lesquels peut jouer l'enseignant de français en veillant à s'ajuster à ceux des élèves. Les apports linguistiques sont contextualisés dans cette perspective (ANNEXE 1b). La feuille de route proposée en évaluation de cette UE invite à prolonger cette réflexion (ANNEXE 2).

2.3. Concevoir un scénario de formation pour réfléchir à l'utilisation possible de grands corpus et à la conception de corpus d'enseignement

2.3.1. L'élaboration du cours

Le cours de didactique de la grammaire pour l'écrit est construit selon les mêmes principes :

Travail individuel préalable : Un cours est proposé à votre lecture avant la connexion. Il s'agira d'en prendre connaissance en annotant les passages qui de votre point de vue, demandent éclaircissement. Vous serez invité, à chaque étape de la réflexion, à faire par vous-même un court exercice ou une petite recherche que vous ferez seul pour la première connexion et que vous mènerez de façon collaborative pour les suivantes. Pour chaque connexion, ayez à portée de main la *Grammaire méthodique du français*, ainsi qu'un dictionnaire.

Tableau 4 : Extrait du plan de travail envoyé aux étudiants en début de semestre

Les deux derniers chapitres de ce cours figurent en annexe. Ils sont étroitement liés : les apports de connaissances et les exercices proposés dans le chapitre 4 *Observer un corpus et le compléter dans la perspective de l'enseignement* (ANNEXE 3) conduisent à des mises en situation développées dans le chapitre 5 *Exploitation d'un corpus dans la perspective de l'enseignement* (ANNEXE 4).

En voici le sommaire :

Chapitre 4 : observer un corpus et le compléter dans la perspective de l'enseignement

1. Les données de la description grammaticale

- a) Lecture d'un extrait de la *Grammaire méthodique du français*
- b) Définition et exemples

2. Dégager des régularités à partir de textes relevant de genres et de registres différents

- a) premier extrait : Charles FOURIER, 1830, *Le Nouveau monde industriel ou l'Agriculture combinée*
- b) deuxième extrait : Extrait du corpus de *l'Est Républicain*, 1999-2003
- c) troisième extrait : Revue *Sciences humaines*, septembre 2013,
- d) quatrième extrait : *20 minutes*, 19 septembre 2013, interview de Tony Parker

- Relevez les occurrences du morphème « on ».
- À partir de la comparaison des contextes dans lesquels il est employé, confrontez vos observations avec celles que vous avez faites précédemment sur les plans

- morphologique
- syntaxique
- sémantique
- énonciatif.

3. Mise en situation

Constituez de courts corpus permettant de faire observer à des élèves en plusieurs étapes les caractéristiques du pronom « on »

- en classe de 6^e, en relation avec l'étude des formes de phrase (négative/ positive),
- en classe de 4^e, en relation avec les accords sujet-verbe complexes,
- en classe de 3^e ou de 2^e, en relation avec l'initiation à l'énonciation et à l'écrit argumentatif.

Chapitre 5 : exploitation d'un corpus dans la perspective de l'enseignement

1. Des corpus différents selon l'âge des élèves, les objectifs d'apprentissage visés et les modalités de travail retenues

Exemples de corpus constitués selon des besoins identifiés pour travailler différents niveaux d'analyse de la langue.

2. De la constitution du corpus à la préparation de la séance de langue

Mise en perspective du corpus dans la séance et la progression.

3. Constituer un corpus d'observation pour faire appréhender les caractéristiques du pronom « on »

3.1. Les caractéristiques du pronom « on » : le savoir du professeur

3.2. Un corpus en classe de 6^e

3.3. Un corpus en classe de 4^e

3.4. Un corpus en classe de 2^e

Séance 1

Étape 1 : étude de la situation d'énonciation

Étape 2 : quelle fonction ont, dans l'argumentation, les passages où le locuteur s'efface ?

Séance 2 : identification des propriétés de « on »

Séance 3 : transposition à l'interprétation d'un passage clé du *Discours sur l'inégalité*

Tableau 5 : Sommaire des deux derniers chapitres du cours à distance « Didactique de la grammaire pour l'écrit ».

2.3.1. Le choix de commencer par la lecture d'une grammaire de référence : chapitre 4-1 du cours

Pour introduire la notion de *corpus* qui, dans son acception technique, n'est pas connue des étudiants, il a paru important de les aider à problématiser la relation entre grammaire et données de la description grammaticale en clarifiant les différentes acceptions du mot *grammaire*, notamment la distinction entre grammaire immanente vs grammaire-description, en identifiant les différentes branches de la grammaire auxquelles font appel les différentes épreuves du concours et en explicitant le rôle des recueils empiriques dans la mise au jour des régularités ainsi que les limites de ces recueils en termes de représentativité. La réponse aux questions posées exige une (re)lecture de l'introduction de la *Grammaire méthodique du français*, lecture critique accompagnée sans laquelle les étudiants se replient sur des grammaires tout public ou des manuels (ANNEXE 3a). Les

exemples de corpus proposés dans la section suivante (ANNEXE 3b) attirent l'attention sur le fait que la quantité d'occurrences est une caractéristique nouvelle des corpus électroniques, mais qu'elle ne suffit pas à déterminer le domaine de validité des régularités observées : la diversification des recueils et leur contextualisation selon l'appartenance générique des textes sont une exigence épistémologique qui est mise à l'épreuve dans la seconde partie du chapitre.

2.3.2. Étudier un même fait de langue dans des textes relevant de genres différents : chapitre 4-2 du cours

Quatre extraits sont proposés, relevant de différents genres, pour lesquels le fonctionnement du pronom indéfini *on* est à étudier. Les consignes sont assez proches de celles que les étudiants rencontrent dans l'épreuve de grammaire, repérage et description, à ceci près que les textes soumis à leur examen ne sont pas littéraires. Quatre niveaux d'analyse sont donnés pour obliger les étudiants à faire une lecture stratifiée des définitions dans les grammaires consultées et permettre les confrontations. Il est en effet envisagé que dans le travail collaboratif, ils se répartissent les textes à étudier et confrontent leurs résultats avant de déposer sur le forum une synthèse qui est discutée lors de la connexion synchrone. Le travail vise à faire apparaître un noyau de propriétés communes sur les plans morphologique, syntaxique et sémantique et des variations sur le plan énonciatif à rapporter au genre des textes et aux types de discours qui s'y déploient. La synthèse des observations est reprise dans le chapitre 5 (3.1 Les caractéristiques du pronom « on » : Le savoir du professeur : ANNEXE 4c). Le plan textuel y est introduit en complément après avoir poussé la discussion sur l'extrait 1 (*on verra...on pourrait*) : est-ce que deux *on* consécutifs désignent nécessairement les mêmes personnes ? Qu'est-ce qui distingue le fonctionnement de *on* de la reprise anaphorique par *il* ?

2.3.3. Aborder un même fait de langue à trois niveaux différents : chapitre 4-3 du cours

La mise en situation proposée en fin de chapitre 4 est travaillée entre les deux connexions synchrones (ANNEXE 3d). Il s'agit de faire réfléchir à la possibilité d'aborder un même fait de langue à trois niveaux scolaires différents, avec des objectifs différents exigeant chacun un corpus spécifique. Pour ce travail collaboratif, chaque groupe choisissant un niveau, les étudiants disposent de la lecture des points 1 et 2 du chapitre 5 (ANNEXE 4a, ANNEXE 4b) mais le point 3 (Le savoir du professeur : ANNEXE 4c) est proposé à l'issue de la séance synchrone pendant laquelle ils ont proposé et discuté leurs corpus.

2.3.4. Des outils pour concevoir un corpus et l'exploiter en classe : chapitre 5-1 et 5-2 du cours

Dans la première section du chapitre 5, il est apparu nécessaire de passer en revue les différents niveaux d'analyse de la langue et de montrer la cohérence entre un corpus et une tâche relevant de chacun de ces niveaux (ANNEXE 4a). La seconde section vise à aider les étudiants à se projeter dans l'élaboration et la conduite d'une séance d'étude de la langue à un moment où ils sont en stage de pratique accompagnée (ANNEXE 4b).

2.3.5. Des mises en œuvre possibles : chapitre 5-3 du cours

La section 3 présente des réponses possibles aux trois mises en situation.

En 6^e (ANNEXE 4d), l'objectif étant de traiter une difficulté liée au passage de l'oral à l'écrit, un corpus de phrases dictées permet de placer les élèves devant le problème de transcription. Il s'agit d'un corpus de phrases forgées mettant en contraste des séquences homophones qui s'opposent sur le plan de la polarité (ex. *on a/on n'a pas*). Un corpus complémentaire d'erreurs d'élèves et d'adultes aide à déterminer les contextes critiques. Un exercice de vérification combine consolidation et apport culturel. Les questions au futur professeur aident à chaque étape à se projeter dans la mise en œuvre possible de la séance.

En 4^e (ANNEXE 4e), le travail sur *on* est le point de départ d'une réflexion sur les accords induits par un pronom singulier sur le plan morphologique mais dont le sémantisme peut comporter l'idée de pluralité. Les différents corpus proposés élargissent le champ des observations en montrant que ce conflit entre morphologie et sémantique se rencontre dans d'autres classes de mots variables : les noms collectifs, certains déterminants de quantité.

La proposition pour la classe de seconde (ANNEXE 4f) a été entièrement rédigée pour répondre aux interrogations des étudiants qui observaient très peu de séances d'étude de la langue au lycée et peinaient à faire la distinction entre vérification des connaissances supposées acquises au collège et réflexion sur le fonctionnement des discours. À la différence des corpus précédents, celui-ci est constitué d'extraits d'un texte à visée argumentative. Dans la première séance, les élèves sont amenés à mobiliser leurs connaissances grammaticales pour distinguer les plans énonciatifs. Ces observations sont le point de départ d'activités visant la conceptualisation des propriétés de *on* dans la séance 2. Le réinvestissement proposé dans la 3^e séance lie étroitement, comme dans la première séance, l'emploi du pronom *on* et la conduite de l'argumentation.

En mettant en vis-à-vis le premier corpus diagnostic (1.3) et l'accompagnement à la conception de corpus d'enseignement (2.3), j'ai tenté de retracer un projet global de formation dans lequel « la dialogalité à savoir l'échange (en présence ou différé) entre personnes (enseignant et/ ou pairs [...] [fonctionne] comme relais du dialogue de soi à soi » (Gerlaud et Elalouf, 2023). La confrontation des futurs enseignants à des corpus y stimule l'activité métalinguistique, autorise la prise de risque intellectuel, exige une appropriation critique des grammaires et contribue ainsi à modéliser l'activité du futur professeur en travaillant conjointement les composantes cognitive et médiative de celle-ci. Mais cette dynamique appelle d'autres articulations : les retours réflexifs proposés aux étudiants sur leur expérience d'élève et leurs premières expériences d'enseignement associés à une mise en perspective historique des programmes permet de travailler simultanément les composantes personnelles et institutionnelle/sociale. Sachant que les pratiques des enseignants sont rapidement stables et cohérentes et leurs marges de manœuvre limitées (Robert et Rogalski, 2002), la prise en compte interactive de chacune de ces composantes dès le début de la formation est une préoccupation majeure.

3. Corpus linguistiques, corpus didactiques : polysémie trompeuse ou possible convergence ?

Sous le terme de *corpus*, les étudiants ont rencontré en lisant ce cours différentes configurations de données linguistiques. Ils ont également été appelés à manipuler des énoncés sans que leur statut soit mentionné. Ce défaut de clarté cognitive apparaît à la relecture. La première observation proposée aux étudiants est appelée *couple de phrases* à juste titre, puisqu'il s'agit d'énoncés forgés et non d'énoncés attestés. En revanche, parmi les exemples de corpus d'enseignement donnés dans le chapitre 5, il y a des énoncés forgés à côté d'énoncés attestés. Il serait nécessaire d'explicitier dans quels cas il est préférable d'avoir recours aux uns ou aux autres et pourquoi. Par ailleurs, le lien entre les grands corpus présentés au chapitre 4 et ceux proposés pour l'enseignement n'est pas clarifié. Dans le chapitre 4, les quatre extraits proposés pour l'observation de *on* sont issus le premier de *Frantext*, le second du corpus de *l'Est Républicain*, le troisième du corpus de la revue *Sciences humaines* et le dernier d'un corpus de presse écrite. Mais le geste qui consiste à prélever dans des corpus existants des extraits soumis à l'observation en formation n'est pas commenté. Il court ainsi le risque d'être naturalisé comme la lecture d'extraits dans la culture scolaire. Au regard de ces insuffisances, il a semblé nécessaire de revenir aux fondements épistémologiques des linguistiques de corpus pour examiner à quelles conditions l'usage du terme *corpus* pour la formation et l'enseignement de la langue peut être retenu. Nous examinerons quelle place est accordée à l'usage (3.1), comment l'objet d'étude est constitué (3.2), ce qui caractérise l'approche empirique des données langagières (3.3.) et la conception du langage sous-jacente (3.4).

3.1. La centralité de l'usage

Pourquoi lorsqu'on est linguiste choisir de travailler sur des corpus plutôt que par introspection ? s'interroge Marie-Paule Jacques :

La linguistique de corpus prend sens dans la réintroduction de la question de l'usage, elle amène à situer, c'est-à-dire à replacer les phénomènes observés et décrits dans un contexte. Les possibilités du système sont alors saisies à travers leurs réalisations effectives, les textes (Jacques, 2005 : 29).

Si la question d'une réintroduction de l'usage se pose, c'est que celui-ci est occulté par la linguistique introspective centrée sur la compétence du locuteur-auditeur idéal (Chomsky, 1965). Avec l'usage, c'est la dimension sociale de l'activité langagière qui est réintroduite et la relation entre langage et langue qui est repensée, cette dernière étant appréhendée comme une abstraction – « les possibilités du système » – dégagée à partir des « réalisations effectives ». La question de l'usage est également centrale pour le didacticien qui se trouve placé devant un choix : présenter seulement des corpus normés, décontextualisés pour l'étude de la langue, au risque d'une absence de transfert dans les pratiques en production et en réception, comme c'est unanimement déploré, ou faire aussi travailler les élèves sur leurs propres productions, avec leurs spécificités, leur rapport variable à la norme, ce qui demande de faire appel aux démarches de la linguistique de

corpus.

3.2. Constituer l'objet d'étude

Si l'on définit le corpus à la suite de Sinclair (traduit par Habert *et al.*, 1997 : 11) comme un ensemble de données langagières attestées « sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon du langage », cela suppose que le linguiste soit en mesure de justifier les choix qu'il opère. Il mobilise pour cela sa connaissance de la langue mais doit pouvoir aussi la remettre en question en se confrontant aux données. Lorsque l'enseignant constitue un corpus pour l'étude de la langue, il choisit des segments de discours actualisant certaines propriétés relatives à un niveau d'analyse de la langue pour les faire observer. La nécessité d'explicitier les critères qui ont présidé à la constitution du corpus s'impose également, mais une difficulté supplémentaire réside dans le nombre des critères à prendre en compte et dans le caractère dynamique de cette prise en compte : critères linguistiques (quelle(s) propriété(s), à quel niveau d'analyse), didactiques (connaissances préalables, progression, formes de travail pédagogique), culturels (familiarité avec les énoncés traités et capacité à les secondariser), ergonomiques (taille du corpus, temps pour le traiter). On a pu constater par exemple que le travail sur des couples d'énoncés était efficace pour faire comprendre une opposition (par exemple celle entre aspect non accompli et accompli traitée en ANNEXE 4a « Si l'observation porte sur la syntaxe »), mais qu'elle pouvait conduire à une approche simplificatrice des faits de langue, binaire, quand elle n'était pas complétée par d'autres corpus, textuels.

3.3. Une approche empirique

La linguistique de corpus adopte une démarche empirique qui s'oppose à la démarche expérimentale de la linguistique introspective. Cette distinction mérite d'être précisée pour comprendre ce que la démarche inductive préconisée par la grammaire rénovée doit à l'une et à l'autre. Une démarche empirique n'est pas a-théorique, elle diffère la théorisation en la subordonnant à l'examen des données, sans ignorer que le chercheur les aborde avec ses préconceptions, ses connaissances, même s'il s'efforce de les contrôler. En cela, elle se distingue d'une approche expérimentale qui a pour but de valider ou de falsifier une théorie :

Tester l'acceptabilité d'une séquence, c'est procéder à une expérience. La construction d'exemples et de contre-exemples constitue l'activité expérimentale du linguiste qui vérifie la théorie de certains phénomènes (Gaston Gross, 1975 : 20).

Qu'en est-il pour des élèves qui abordent la description grammaticale ? Eux non plus ne sont jamais vierges de préconceptions et de connaissances partielles, mais n'ont généralement pas les moyens de les contrôler, ce qui suppose le relais de l'enseignant. À quelles conditions peuvent-ils découvrir certaines régularités de fonctionnement ?

Eric Genevay (1996 : 57-58) qui revendique, dans la rénovation de l'enseignement en

Suisse romande, « la démarche inductive la plus authentique possible » l'exemplifie avec les expansions du nom. Le maître amène les élèves à produire une liste dans laquelle il retient « un échantillonnage représentant les diverses constructions possibles : *un jean délavé, une casquette à carreaux, une chemise qui ne se repasse pas...* ». Les élèves ont produit un corpus. Peut-être a-t-il déjà subi des modifications au moment de sa transcription collective, sans que celles-ci aient été explicitées. Puis la constitution de l'échantillon conduit à écarter certaines formes : le maître peut choisir de laisser de côté les adjectifs antéposés pour les aborder dans un second temps comme le préconise Genevay. Il reste alors dans la logique de la linguistique de corpus en envisageant la construction progressive des propriétés des constituants de l'expansion du nom. Il peut aussi choisir d'écarter des formes qui résistent à son analyse comme *une casquette Nike, un truc super, une soupe beurk*. Il n'est plus alors dans la logique d'un examen de corpus qui suppose de rendre compte de *toutes* les données, à moins qu'il envisage de réserver l'examen de ces formes à un autre moment où il abordera leurs conditions de production et des familles de paraphrases reliées à différentes situations d'énonciation et différents rapports à la norme.

À partir de l'échantillon sélectionné, l'observation « consiste alors à examiner les résultats provoqués sur les unités proposées, par une série de manipulations dont les plus fréquentes sont des essais de suppression, de déplacement ou de remplacement. Par exemple, dans les groupes du nom retenus ci-dessus, on remarque qu'il est possible de supprimer ce qui suit le nom » (op. cit.). Recourir aux méthodes de la linguistique distributionnelle reste dans la logique d'une approche de corpus dans la mesure où les groupes réduits sont attestables : rechercher des exemples dans des corpus peut consister en une vérification. En revanche, d'autres manipulations qui relèvent davantage d'une approche expérimentale sont plus discutables. Les élèves sont ainsi invités à supprimer le déterminant pour conclure que le déterminant et le nom sont les constituants obligatoires du groupe nominal alors même que le nom *carreaux* est employé sans déterminant dans *une casquette à carreaux*. L'observation est ici surdéterminée par la théorisation visée, trop puissante, mais porteuse de la force de la tradition.

Inviter les élèves à recourir aux opérations qui sont constitutives de la production langagière, dont attestent les corpus de génétique textuelle – ajouter, supprimer, remplacer, déplacer – leur permet d'observer la langue avec les outils dont ils disposent ; leur demander de produire des énoncés agrammaticaux pour valider ou invalider un raisonnement exige une grande circonspection. Il reste autour des manipulations de nombreux malentendus à travailler.

3.4. Une conception du langage, de son acquisition et de son enseignement

Si l'on considère l'action langagière comme une activité sociale, un instrument de schématisation de l'expérience nécessaire au développement du psychisme par la production et la réception de textes écrits et oraux, il y a lieu de penser qu'« une démarche didactique *idéale* consisterait à faire débiter l'enseignement de la langue par des activités de lecture et de production de textes, et à articuler ensuite à cette démarche

initiale des activités d'inférence puis de codification des régularités observables dans le corpus de textes ainsi mobilisé » (Bronckart, 1996, 2022 : 67).

Mais l'idée que la langue serait première et les textes une mise en fonctionnement seconde reste profondément ancrée dans les représentations des acteurs (enseignants, mais aussi parents et élèves) ; elle continue de structurer les pratiques scolaires. Le compromis proposé par Bronckart est de mener des activités de codification et d'inférences sur la base de « corpus de phrases ou d'énoncés *triés* aboutissant à une maîtrise des principales notions et règles du système de la langue » et parallèlement des activités de sensibilisation au fonctionnement des textes « sur la base d'un corpus de textes empiriques eux aussi *triés* ». On note la récurrence du verbe *trier*, souligné par l'italique. Mais les critères de tri restent à préciser. Il est seulement indiqué qu'ils doivent être compatibles avec une visée considérée comme un pari :

les concepts construits dans le cadre de l'enseignement grammatical pourront, à terme, constituer des instruments généralisables aux activités de production et d'interprétation des textes (Bronckart, 2022 : 68).

Quels critères de tri permettront d'avancer sur cette voie ? Des tris pour planifier une progression par enrichissement progressif des notions et mise en relation dans des réseaux conceptuels ? Des tris pour réserver des temps d'examen aux imprévus, aux cas limites ? Quelle place les productions orales et écrites des élèves tiendront-elles dans ces corpus ? Comment seront sollicités les outils de grammatisation (dictionnaires, grammaires) qui reposent explicitement sur des corpus plus larges, la navigation sur internet qui donne accès à des corpus non triés ? Comment seront sollicitées les linguistiques de corpus dont la démarche n'oppose pas structuration phrastique et textuelle ? Ce sont autant de questions posées aujourd'hui à la recherche en didactique de la langue et à la linguistique de corpus.

Conclusion

Cet article tente de montrer que le recueil de grands corpus scolaires et universitaires constitue une ressource à partir de laquelle il est possible d'ajuster différents corpus pour la formation. Les circonstances des multiples réformes de la formation n'ont pas laissé le loisir de recueillir des traces autres que mémorielles de l'évaluation diagnostic. Mais peut-être l'article suscitera-t-il de nouveaux recueils. Le retour réflexif sur ces scénarios a mis au jour un défaut d'explicitation et fait percevoir en creux un besoin accru de justification, tant pour les corpus de formation que pour les corpus d'enseignement qu'ils sont censés aider à produire. Ces justifications sont d'autant plus complexes qu'elles nécessitent de prendre en compte de façon dynamique les différentes composantes cognitive, médiative, personnelle, institutionnelle et sociale de l'activité enseignante.

Sur le plan institutionnel, un verrou majeur subsiste, qui a des conséquences sur le couplage des composantes cognitive et médiative dans la conduite de la classe. Il est dénoncé depuis plusieurs décennies par Bernard Combettes :

tout se passe comme si la finalité de l'enseignement – quelle que soit d'ailleurs la

méthode pédagogique adoptée – était l'acquisition de catégories du système de la langue et comme si la connaissance de ces catégories entraînait une plus grande maîtrise dans les activités langagières. Cette insistance sur les éléments a le grand inconvénient de laisser au second plan les constructions syntaxiques, qui deviennent des sortes d'« applications », de réalisation des natures et fonctions, alors que le mouvement inverse apparaît comme plus pertinent, ou pour le moins « naturel » : les constructions devraient être considérées comme la donnée première, les unités n'étant que des outils pour l'observation raisonnée des structures (Combettes, 2009).

Un constat similaire « d'inconsistance scientifique et d'inefficacité opératoire » conduit Dan Van Raemdonck et ses collègues (2012: 18) à poser d'abord le « mécanisme prototypique d'apport d'information à un support ». L'intérêt porté par la linguistique de corpus aux constructions se traduit aujourd'hui dans l'analyse linguistique des productions d'élèves (Masseron, 2005, 2011, Cappeau et Roubaud, 2018 ; Elalouf et Roubaud, 2016 ; Roubaud et Elalouf, 2020 ; Koulibali, 2023). Gageons que ces travaux contribueront à baliser des cheminements du langage à la langue en partant de la pratique langagière effective des élèves.

Bibliographie

- Beaumanoir-Secq, M. Cogis, D. et Elalouf, M.-L. (2010) : « Pour un usage raisonné et progressif de la commutation en classe », *Repères 41*, Paris, 47-70.
- Bronckart, J.-P. (1996, 2022, 2^e édition). *Activité langagière, textes et discours : pour un interactionisme socio-discursif*. Limoges, Lambert-Lucas.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
- Dabène, M. et Reuter, Y. (dir.) (1998). *Pratiques de l'écrit et modes d'accès au savoir dans l'enseignement universitaire. Lidil 17*. Grenoble : ELLUG.
- Defays, J.-M., Maréchal, M. et Mélon, S. (2000). *La maîtrise du français : Du niveau secondaire au niveau supérieur*. Bruxelles : De Boeck.
- Elalouf, M.-L., Benoit, J.-P., et Tomassone R. (1996) : *Enseigner le français. La formation linguistique des professeurs de Lettres en question*. Amsterdam, De Werelt, Travaux et recherches en linguistique appliquée.
- Elalouf, M.-L. et Tomassone, R. (2001) : « Observer et interpréter : le mot, la phrase et le texte dans les stratégies des élèves et des futurs professeurs », in Garcia-debanc, C, Confais, J.-P. et Grandaty, M. (coord.) *Quelles grammaires enseigner à l'école et au collège ? Discours, genres, texte, phrase* (395-408). Delagrave-CRDP Midi-Pyrénées.
- Elalouf, M.-L., (2007) : Habilitation à diriger des recherches « Apprentissage de l'écrit et réflexion métalinguistique, étude qualitative de corpus scolaires et universitaires », sous la direction de Bernard Combettes, Nancy 2.

- Elalouf, M.-L., (2009) : « De la grammaire enseignée à l'enseignement de la grammaire : diversité des parcours de PLC2 Lettres et points de résistance », in Goigoux, R., Ria, L. et Toczec-Capelle, M.C., *Les parcours de formation des enseignants débutants* (189-203). Presses de l'Université Blaise Pascal.
- Elalouf, M.-L et Roubaud, M.- N. (2016) : « Un couple de verbes antinomiques à l'épreuve de la syntaxe et de l'usage. Le cas des verbes : *se souvenir* et *oublier* », *Scolagram 2, La syntaxe du verbe, usage et enseignement*. En ligne : <http://scolagram.u-cergy.fr>
- Elalouf, M.-L., (2023). La formation des enseignants de français, La place de la langue dans son évolution (1990-2022). *Le Français aujourd'hui*, 221, 63-77.
- Genevay, E. (1996). S'il vous plaît...Invente-moi une grammaire. Dans S. G. Chartrand (dir.) (2^e édition) *Pour un nouvel enseignement de la grammaire*, 53-84. Montréal, Éditions Logiques.
- Gerlaud, B. et Elalouf, M.-L (2023). Annoter au lycée, un geste professionnel partageable ? *Recherches 79*, pp. 65-88.
- Gross M. (1975). *Méthodes en syntaxe : régime des constructions complétives*. Paris : Hermann.
- Habert B., Nazarenko A. et Salem A. (1997). *Les linguistiques de corpus*. Paris : Armand Colin.
- Jacques M.-P. (2005). Pourquoi une linguistique de corpus ? In G. Williams (Éd.), *La linguistique de corpus*, (21-30). Rennes : Presses Universitaires de Rennes.
- Koulibali, Y. (2023). *Les constructions des verbes hyperfréquents dans des corpus d'écrits scolaires : contribution d'une étude textométrique à la connaissance des usages des verbes de parole et d'états mentaux par des collégiens (6^e et 5^e) en éducation prioritaire*. Thèse de doctorat de sciences du langage de CY Cergy Paris Université sous la direction de Marie-Laure Elalouf.
- Laborde-Milaa, I. Boch, F. et Reuter, Y. (dir.) (2004). *Pratiques 121-122*. Metz.
- Masseron, C. (2005). Indicateurs langagiers et stratégies scripturales, du discours à la langue. *Pratiques*, 125-126, 205-249.
- Masseron, C. (2011). L'analyse linguistique des écrits scolaires », *Pratiques*, 149-150, 129-162. DOI : <https://doi.org/10.4000/pratiques.1725>
- Reuter, Y (éd.), Cohen-Azria, C., Daunay, B., Delcambre, I. et Lahanier-Reuter, D. (2007). *Dictionnaire des concepts fondamentaux des didactiques*. Bruxelles, De Boeck.
- Robert, A. et Roglaski, J. (2002). Le système complexe et cohérent des pratiques des enseignants de mathématiques : une double approche. *Revue canadienne de l'enseignement des sciences des mathématiques et des technologies*, Vol 4, 505-528.
- Roubaud, M.-N. et Elalouf, M.-L (2020). S'approprier les constructions des verbes *se souvenir* et *oublier* : de la prise en compte de la parole des élèves allophones (11-



15 ans), *Études de linguistique appliquée*, pp. 163-179.

Trévisse, A. (2001). Métadiscours explicatifs ou métalangue de bois ? Quelques exemples tirés d'exposés de candidats à l'agrégation interne d'anglais. In Actes du Colloque International *Métalangue et terminologie linguistique* (739-750). Université Stendhal - Grenoble III, *Orbis/ Supplementa*, Tome 17, Peeters : Louvain.

Van Raemdonck, D., Daille, M. et Meinertzhausen, L. (2012). *Le sens grammatical*. Bruxelles : Peter Lang.