

écri+, un dispositif en ligne d'évaluation, de formation et de certification des compétences écrites en français : le cas de la citation, entre formation et recherche

Gabriella DE LUCA
UPEC, Ceditec (France)
gabriella-regina.de-luca-silva-moreira@u-pec.fr
<http://orcid.org/0000-0003-3324-4469>

Sarah DE VOGÜÉ
Université Paris Nanterre Modyco (France)
devogue@parisnanterre.fr
<https://orcid.org/0000-0003-2822-8739>

Julie LEFEBVRE
Université Paris Nanterre Modyco (France)
jlefebvre@parisnanterre.fr
<https://orcid.org/0000-0002-2971-8070>

Frédérique SITRI (Auteur correspondant)
UPEC, Ceditec (France)
Frederique.sitri@u-pec.fr
<https://orcid.org/0000-0002-4177-7962>

Résumé : Dans cet article, nous présentons les différents aspects d'un dispositif de remédiation de l'écrit en ligne pour les étudiants, élaboré dans le cadre d'un projet regroupant quinze universités françaises (projet écri+ (<https://ecriplus.fr/>), dont l'une des instigatrices est Sarah De Vogüé, qui en coordonne l'une des actions). Prenant comme exemple le traitement du discours rapporté et de la citation, activité emblématique de l'écrit universitaire, nous montrerons comment interagissent, dans la conception des tests et des activités, 1) le choix d'un modèle opératoire, 2) la constitution et l'analyse de corpus d'apprenants et de corpus d'experts, et 3) les contraintes liées au support numériques.

Mots clefs : littéracie universitaire ; discours ; recherche appliquée ; français

English:

écri+, an online system for evaluation, training and certification of written skills in french: the case of reported speech, between formation and research

Abstract: In this contribution, we describe different aspects of an online writing remediation system for students, developed within the framework of a project involving fifteen French universities (the écri+ project (<https://ecriplus.fr/>), which was initiated by a group of scholars including Sarah De Vogüé, who coordinates one of its actions). Taking the example of the treatment given for reported speech and quotation, which constitutes emblematic activity for academic writing, we will show how interact in the designing of tests and activities, 1) the choice of a linguistic model, 2) the constitution and analysis of two corpora of writings, one from learner writers, one from expert writers, and 3) the constraints linked to the digital medium.

Keywords: academic literacy ; discourse ; applied research ; french



Español:

écri+, un sistema en línea para la evaluación, la formación y la certificación de las competencias escritas en francés. El caso de la citación: entre la formación y la investigación

Resumen: En este artículo se presentan los diferentes aspectos de un sistema de “remediación” de la escritura en línea para estudiantes, desarrollado en el marco de un proyecto en el que participan quince universidades francesas: el proyecto *écri+* (<https://ecriplus.fr/>), del que una de las autoras, Sarah De Vogüé, es impulsora a la vez que coordina una de las acciones. Tomando como ejemplo el tratamiento del discurso referido y de la citación, actividad emblemática de la escritura universitaria, mostraremos cómo, en el diseño de las pruebas y de las actividades, interactúan 1) la elección de un modelo operativo, 2) la constitución y el análisis de los corpus (de aprendices y de expertos), y 3) las limitaciones ligadas al soporte digital.

Palabras clave: literacidad universitaria; discurso; investigación aplicada; francés

Català:

écri+, un sistema en línia per a l'avaluació, la formació i la certificació de les competències escrites en francès. El cas de la citació: entre la formació i la recerca

Resum: En aquest article presentem els diferents aspectes d'un sistema de “remediació” de l'escriptura en línia per a estudiants, desenvolupat en el marc d'un projecte en el qual participen quinze universitats franceses: el projecte *écri+* (<https://ecriplus.fr/>), del qual una de les autores, Sarah De Vogüé, n'és impulsora, alhora que en coordina una de les accions. Prenent com a exemple el tractament del discurs referit i de la citació, activitat emblemàtica de l'escriptura universitària, mostrarem de quina manera, en el disseny de les proves i de les activitats, interactuen 1) l'elecció d'un model operatiu, 2) la constitució i l'anàlisi dels corpus d'aprenents i dels corpus (d'aprenents i d'experts), i 3) les limitacions lligades al suport digital.

Paraules clau: literacitat universitària; discurs; investigació aplicada; francès

Introduction

Sous une forme ou une autre, la maîtrise de l'écrit paraît requise à tous les niveaux de qualification du monde professionnel contemporain (notamment Rot et al., 2014), qu'il s'agisse de simples coches dans des feuilles de service (Boutet, 2014), des échanges de messages, des curriculum vitae, des comptes-rendus, des annonces ou des notes de synthèse, sans compter tous les genres de ce que l'on peut décrire comme une écriture experte de haut niveau (Bouquet, 2009). L'écrit étant vecteur de savoir partageable (voir Goody (1979), et pour les implications dans le champ scolaire, Reuter (2006)), sa maîtrise est bien évidemment attendue aussi dans les activités scolaires et plus spécifiquement, pour les adultes, dans les études supérieures, ou dans la formation tout au long de la vie.

Le réseau de compétences mobilisé est particulièrement complexe : impliquant de manière indissociable compréhension, interprétation, planification, production et révision/reformulation, il convoque chacune des strates en jeu dans la fabrication de discours, depuis le système de conditionnement des réalisations de phonèmes, jusqu'aux modalités qui régissent la structuration des textes, en passant par les principes de découpage syllabique, les jeux de morphèmes, les réseaux lexicaux, la construction des phrases et les divers modes de déploiement de points de vue différenciés dans les discours élaborés. Aucun de ces niveaux n'est propre à l'écrit, l'ensemble, agencement de textes compris, étant caractéristique des langages humains en général (voir Benveniste, 1974). Pourtant, l'écriture des langues et l'agencement de discours écrits ne cessent de faire difficulté, pour des locuteurs qui par ailleurs maîtrisent aussi bien qu'il est possible la langue utilisée pour formuler ces discours à l'oral.

La difficulté est générale et chacun en fait l'expérience, mais elle touche de manière plus radicale une partie



des candidats scripteurs, les dysfonctionnements concernant, tout du moins en France, un nombre important de personnes qui sont en difficult3 dans leurs 3tudes comme dans leurs activit3s professionnelles 3 cause de d3faillances du c3t3 de l'3crit. Les causes de ces dysfonctionnements sont plurielles. Elles peuvent 3tre partiellement mises en relation avec des carences du syst3me scolaire fran3ais, dont les faibles r3sultats dans les 3valuations PISA (<https://www2.compareyourcountry.org/pisa/country/fra?lg=en>) t3moignent des probl3mes rencontr3s pour arriver 3 former l'ensemble de la population, notamment dans le domaine de l'3crit. Sont sans doute en cause des probl3mes de moyens mais peut-3tre aussi une forme d'incompr3hension de ce qui peut faire difficult3 dans l'3crit, voire m3me de ce qu'est l'3crit dans ses diff3rentes dimensions et ses diff3rents modes de fonctionnement. Sur ces questions, les travaux de recherche sont nombreux pour une meilleure compr3hension des enjeux qui sont attach3s 3 l'enseignement scolaire d'une part, pour une meilleure appr3hension de l'3crit lui-m3me et des enjeux qu'il recouvre d'autre part. Ils sont moins nombreux pour ce qui concerne l'3crit des adultes, que ce soit dans le monde du travail, ou au moment des 3tudes sup3rieures (voir cependant Boch et Frier, 2015 ; Delcambre et Lahanier-Reuter, 2012 ; Laborde-Milaa et al., 2015). Surtout, si le temps d3volu 3 l'apprentissage de l'3crit est proportionnellement tr3s important 3 l'3cole, le temps pass3 sur la question de l'3crit dans les 3tudes sup3rieures, notamment 3 l'universit3, est g3n3ralement r3duit, ou limit3 3 l'apprentissage d'3crits de sp3cialit3. De plus en plus cependant, dans les 3tablissements fran3ais d'enseignement sup3rieur, des enseignements de ma3trise de l'3crit sont mis en place, au titre de la formation dite transversale¹.

Parall3lement, des ressources en ligne plus ou moins payantes se multiplient, signes du besoin ressenti dans le domaine aussi bien par les institutions qui y recourent, que par les personnes concern3es, qui y voient un moyen de r3soudre au moins partiellement leurs difficult3s. Ces ressources, qui peuvent 3tre tr3s lucratives, se centrent g3n3ralement sur l'orthographe (y compris lorsqu'elles annoncent de la grammaire, volontiers r3duite 3 la question des accords et de l'orthographe grammaticale), avec maintenant quelques incursions dans le domaine du lexique, des genres de texte et des registres de langue. Pour la plupart, elles ne sont pas produites par des sp3cialistes du langage, malgr3 une tendance r3cente 3 revendiquer une implication de la recherche notamment en sciences du langage².

C'est dans ce contexte que le projet 3cri+ (<https://ecriplus.fr/>) s'est mis en place en 2017 (PIA NCU n3 ANR-17-NCUN-0015), projet auquel les auteurs ont activement contribu3.

Dans une premi3re partie nous pr3senterons les sp3cificit3s du dispositif 3cri+, qui se base sur des donn3es produites par les 3tudiants pour produire tests et activit3s p3dagogiques. Dans une seconde partie, nous nous

¹ Voir la version consolid3e au 28 f3vrier 2017 de l'arr3t3 du 1er ao3t relatif 3 la Licence faisant 3tat du besoin de faire acqu3rir 3 l'3tudiant « des comp3tences transversales ou g3n3riques, telles que l'aptitude (...) 3 l'expression 3crite et orale (...)» (<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000024457754>).

² Noter l'exception du ccdmd qu3b3cois (<http://www.ccdmd.qc.ca/fr/>) en place depuis de nombreuses ann3es ; l'Universit3 Ouverte des Humanit3s (UOH <http://uoh.fr/front/>), qui est une institution charg3e de produire, faire produire et diffuser des ressources num3riques dans le champ des sciences humaines y donne acc3s depuis longtemps.



concentrerons sur la question de la citation : nous pr3senterons les principes qui ont guid3 l'3laboration des s3quences p3dagogiques d3di3es 3 la r3m3diation des difficult3s en question en insistant sur l'articulation entre le choix du mod3le et l'observation des corpus d'3crits d'apprenants et d'3crits d'experts, et sur les contraintes li3es au support num3rique. On montrera pour terminer comment les d3veloppements en cours conduisent 3 prendre en compte les genres de discours et 3 s'appuyer sur les outils du TAL (Traitement Automatique des Langues) pour affiner la description des donn3es (th3se en cours de Gabriella de Luca, 3labor3e dans la continuit3 de De Luca, 2021). On verra tout au long de cet article que le dispositif 3cri+ vise non seulement 3 produire des outils didactiques mais 3galement 3 faire avancer la recherche dans le domaine de l'3tude de la langue et des discours.

1. Le dispositif 3cri+ : principes et mise en 3uvre

Le projet 3cri+, financ3 par l'Agence Nationale de la Recherche, r3unit autour de l'Universit3 Ouverte des Humanit3s (UOH) un consortium de 15 3tablissements d'enseignement sup3rieur, avec pour objectif sur dix ans (2017-2027) la construction d'un dispositif national d'3valuation, de formation et de certification des comp3tences 3crites en fran3ais. 3labor3 dans la perspective d'un public de langue maternelle fran3aise (jeunes adultes dont le fran3ais est la L1), 3cri+ pourrait sans doute, au prix de quelques adaptations, servir 3 un public de L2 de niveau B1. Ce dispositif, en cours de construction, comprend plusieurs volets : a) le d3veloppement d'une plateforme d'3valuation et de formation en libre acc3s (<http://app.tests.ecriplus.fr>) ; b) l'3laboration de ressources en ligne ; c) la mise en place d'une certification nationale ; d) la constitution d'un r3seau d'accompagnement et de partage de bonnes pratiques entre les 3tablissements d'enseignement sup3rieur qui d3ploient ou pr3voient de d3ployer des formations 3 l'3crit, en pr3sentiel ou 3 distance.

Hormis le fait qu'il est national et public, le dispositif pr3sente plusieurs particularit3s notables. Tout d'abord, loin de se concentrer sur l'orthographe, il mobilise l'ensemble des domaines impliqu3s dans le champ de l'3crit. Ensuite, il concerne sp3cifiquement les jeunes adultes et il s'adresse de fait 3 ces derniers, et aux difficult3s qu'ils rencontrent. Pour cela, il s'appuie sur des erreurs ou maladroites effectivement produites, au lieu de d3cider a priori des probl3mes 3 traiter, ou de reprendre 3 l'identique les th3matiques trait3es 3 l'3cole 3l3mentaire (tables de conjugaison, r3gles d'accord, etc.). Enfin, une derni3re particularit3 est qu'au lieu de consid3rer que les difficult3s de l'3crit sont connues et n'ont plus qu'3 3tre enseign3es correctement, 3cri+ choisit de partir des donn3es rassembl3es (erreurs et maladroites d'une part, repr3sentations et m3thodes d'autre part, genres de textes et mod3les enfin) pour essayer de comprendre d'abord ce qui fait difficult3, pour avancer ce faisant aussi dans la compr3hension de ce qu'est l'3crit et de ce que peut mobiliser une progression dans la ma3trise de l'3crit. De ce point de vue, le projet est autant un projet de formation qu'un projet de recherche dont l'objectif est d'utiliser les donn3es recueillies pour faire avancer l'3tat des connaissances actuelles sur l'3crit, sur les modalit3s d'apprentissage, mais aussi sur les langues et



sur les compétences que tout un chacun peut posséder relativement à ces langues et à leurs usages. Le choix des modèles linguistiques et la constitution de corpus d'apprenants sont donc à la base du projet écrit+, tout autant que la réflexion sur l'outil numérique.

1.1 Des données aux modules

Le dispositif écrit+, avec ses tests de positionnement et ses modules de remédiation, s'appuie sur un référentiel ad hoc, dont la structure a été élaborée au fur à mesure d'analyse des erreurs diagnostiquées et des besoins constatés.

Ce référentiel se décline en huit niveaux de compétences (Fondamental 1 et 2, Avancé 1 et 2, Spécialisé 1 et 2, Expert 1 et 2) et quatre grands domaines (Mots, Phrases, Textes, Discours). Chaque domaine recouvre à son tour quatre compétences ; par exemple dans le domaine du Discours, on a *Analyser et structurer des contenus*, *Jouer avec les sous-entendus*, *Déterminer les points de vue adoptés* qui nous intéresse plus précisément ici, et, plus lié à l'orthographe, *Maîtriser les marques de conjugaison*. Chaque compétence regroupe quant à elle une douzaine de thématiques différentes ; par exemple la compétence *Déterminer les points de vue adoptés* regroupe notamment diverses thématiques liées à la citation, sur lesquelles nous revenons dans la deuxième partie.

Les thématiques pouvant se déployer sur les huit niveaux de compétence, elles peuvent inclure chacune jusqu'à huit types de problèmes de niveaux différents, certains relevant de la compréhension, d'autres plutôt de la rédaction. Ainsi, la thématique se rapportant au Discours Direct comprend autant d'items qu'on a pu identifier de difficultés dans son utilisation (voir Tableau 1). Chaque item apparaît dans le référentiel avec son libellé et un code. Au total, avec ses compétences, ses thématiques et ses niveaux, le référentiel peut recouvrir jusqu'à 16x12x8 items. On en donne un extrait dans le tableau qui suit, avec les items des quatre premiers niveaux pour les quatre thématiques du discours indirect, des formes en *selon*, du discours direct et des îlots textuels. On voit que chaque item est doté d'un code (ex. D-PVs0), qui indique le domaine dont il relève (ici D pour le domaine du Discours), la compétence qu'il développe (ici PVs pour *Déterminer les points de vue adoptés*).

Tableau 1

Extrait du référentiel sur les modes de citation : items des niveaux 1 à 4.

En orange les problèmes de compréhension, en bleu clair les problèmes de formulation, en bleu turquoise les problèmes de ponctuation

Introduire le discours indirect D-PVs0	Pronoms dans le discours indirect avec discours rapportant exclusivement P3 D-PVs0_2	Pronoms dans le discours indirect avec P1 ou P2 dans discours rapportant D-PVs0_3	Pronoms dans discours indirect quand plusieurs P3 D-PVs0_4
	Syntaxe de formes comme <i>Selon</i> , <i>D'après</i> , <i>Comme le V X</i> D-PVs2_2	Interprétation de formes comme <i>Selon</i> , <i>D'après</i> , <i>Comme le V X</i> D-PVs2	Conditions sémantiques d'emploi de formes comme <i>Selon</i> , <i>D'après</i> , <i>Comme le V X</i> D-PVs2_4

Syntaxe de la citation directe : introduire un Discours Direct autonome D-PVs5_2	ponctuation des citations directes : Discours Direct autonome D-PVs5	Identifier dans un texte le segment qui reformule la citation. D-PVs7_3	Choisir une reformulation qui convienne pour accompagner une citation D-PVs6_2
ponctuation pour intégrer les îlots textuels D-PVs5_3	intégration des questions comme îlots textuels D-PVs5_4	personnes dans les îlots textuels D-PVs4	personnes et temps dans les îlots textuels D-PVs4_2

1.2 Tests, tutoriels, modules

Pour construire les tests de positionnement, l'équipe des concepteurs, encadrée par Sarah De Vogüé, a produit pour chaque type de problème une demi-douzaine de questions de contenu équivalent. Ces questions ont été élaborées à partir de dysfonctionnements attestés sur les données qui sont recueillies. A chaque question sont associés une explication et un tutoriel, que l'utilisateur peut consulter à volonté, qu'il ait répondu correctement ou pas à la question posée ; explication et tutoriel sont ce qui permet à l'utilisateur de se former au fur à mesure qu'il progresse dans le test

Les tests se déroulent en suivant un algorithme adaptatif ; l'utilisateur doit répondre à des questions pour chacune des thématiques de la compétence qu'il a choisi de travailler, mais le niveau des questions qui lui sont posées est fonction du niveau acquis et donc des résultats obtenus pour les questions précédentes. Un extrait est présenté dans la Figure 1 ci-dessous :

Figure 1

Plateforme écrit+test ; Extrait d'une question sur l'insertion des citations.

En haut la thématique, le niveau atteint, et la position de la question dans la série en cours

écrit+ Déterminer les points de vue adoptés

NIVEAU 5 GAGNÉ!
Déterminer les points de vue adoptés

1 2 3 4

Lisez attentivement l'extrait suivant. A l'endroit marqué [X] va être insérée une citation. Quatre solutions sont proposées. Quelle citation vous semble s'insérer pertinemment dans l'extrait ?

- *Les dernières forêts primaires tropicales ne constituent plus que 5 % des forêts de la planète alors qu'elles recèlent 75 % des réserves de biodiversité. Francis Hallé, biologiste et botaniste, est convaincu qu'on ne pourra sauver ce qu'il en reste qu'en réévaluant le statut du règne végétal qui a été trop longtemps soumis au modèle de l'animal prédateur : [X] C'est ne rien comprendre à ce monde dont nous dépendons puisque, à travers la photosynthèse, le végétal nous permet de respirer en nous fournissant l'oxygène. (D'après Philosophie Magazine, n°94, 11/2015)*

a. « L'animal est un prédateur qui se nourrit en faisant la course à ses proies et en tournant le dos à ses excréments. La plante capte l'énergie en permanence sans avoir à agresser des proies. Elle se nourrit de la lumière solaire qui est inépuisable et stocke son énergie pour croître en hauteur. »

À terme, une partie au moins des tutoriels prendra la forme d'une vid3o de deux minutes³, dans laquelle l'utilisateur est mis à contribution, d'abord au travers d'une brève réactivation de savoirs acquis, ensuite à l'occasion d'un problème à résoudre où il se voit demandé quelle méthode il utiliserait pour le résoudre ; la vid3o parcourt alors plusieurs méthodes qui ont pu être mentionnées par d'autres utilisateurs, et en explicite une en particulier qui est en même temps illustrée. Le tutoriel comprend enfin des renvois à des ressources en ligne, dont celles produites dans le cadre du projet pour le site *J'améliore mon français écrit*, qui comprend une série de modules sur lesquels nous revenons dans la deuxième partie de cet article. L'intérêt d'utiliser la vid3o pour ces tutoriels est que cela permet de s'adresser à des utilisateurs qui sont susceptibles d'avoir des difficultés pour lire et comprendre l'écrit, puisque c'est à ce type de difficultés que s'adresse le dispositif écrit+ dans son ensemble. Les difficultés sont potentiellement démultipliées dans la mesure où l'écrit des explications est à visée métalinguistique. En doublant les explications par la voix et en les introduisant progressivement, on peut mieux accompagner l'utilisateur dans sa lecture.

Que ce soit dans le test lui-même, dans les tutoriels ou dans les modules, on part toujours non pas de règles qu'il s'agirait d'appliquer, mais de problèmes à résoudre, problèmes à l'occasion desquels on va pouvoir mettre en évidence des principes de fonctionnement et des méthodes de résolution. Derrière ce choix, il y a des raisons d'ordre pédagogique, les apprentissages étant plus efficaces quand inscrits dans une perspective pratique de résolution de problèmes. Il y a aussi des raisons relatives à la conception que l'on peut avoir de l'écrit, donné parfois comme un tissu de normes à respecter, quand on soutient qu'il renvoie à une pratique, à des savoir-faire et à des problèmes à résoudre.

2. La citation dans les écrits des étudiants

Savoir citer autrui fait partie des compétences qu'il est nécessaire de posséder lorsque l'on parle ou que l'on écrit ; quel que soit le discours, et quel que soit le support pour ce discours, l'on est forcément amené d'une façon ou d'une autre à devoir citer des sources (documents, auteurs, etc.) d'une part, à convoquer des points de vue pluriels d'autre part pour mettre en perspective faits et positions. Cette compétence n'est certes pas propre à l'écrit, mais elle prend à l'écrit des formes particulières. En outre, comme pour toutes les autres compétences linguistiques, il est fréquent que quelqu'un maîtrisant par ailleurs aussi bien qu'il est possible cette compétence à l'oral, en vienne à perdre ses moyens à l'écrit jusqu'à ne plus savoir comment, quoi, ou pourquoi citer.

La situation à l'écrit est d'autant plus délicate pour l'apprenti scripteur qu'elle se trouve prise de fait dans un réseau d'injonctions particulièrement déstabilisant, quand il est à la fois attendu de chacun qu'il cite et sache citer, qu'il ne s'en tienne pas à citer mais puisse aussi au moins commenter voire s'approprier le discours cité, qu'il ne

³ Élaborées par une équipe coordonnée par l'Université de Clermont-Ferrand, avec l'Université de Côte d'Azur et l'Université de Paris Nanterre.

tombe pas pour autant dans le plagiat ou l'emprunt d'idées, fortement stigmatisés, et donc qu'il sache aussi « citer ses sources », en suivant pour ce faire des normes strictes, mais qu'en même temps il ne se dispense surtout pas de reprendre éléments de contenu et éléments de langage introduits par les enseignants dans les cours sur lesquels il s'appuie, voire par les institutions ou les entreprises dans lesquelles il s'inscrit (voir la fameuse culture d'entreprise), ces reprises-là se faisant sans citation mais au contraire en les intégrant et en se les appropriant.

Cependant, indépendamment de ces injonctions hétéroclites, la mobilisation de discours autres pose de fait des problèmes qui sont constitutifs de la parole et des usages qu'on en fait ; dans toute parole, il faut se positionner par rapport à autrui ; toute parole se nourrit des paroles d'autrui, mais aussi des paroles possibles, et encore du silence des faits intrinsèquement résistants à la parole. A l'écrit, ces problèmes sont exacerbés, par la déconnection subjective qu'impose l'écrit, avec ces mots et textes détachés de soi et laissés au loisir et à la sagacité d'un éventuel lecteur ; comment se positionner quand on a le sentiment de ne plus être aux commandes de la production du sens ? Ces difficultés sont plus exacerbées encore dans un contexte de formation, où est en jeu la constitution même du sujet et d'une voix ou d'une expertise qui lui soit propre, au milieu de toutes les voix et toutes les expertises qui vont les nourrir.

2.1 Un dispositif basé sur des corpus d'apprenants et sur un modèle opératoire

Les recherches qui, dans le champ des littéracies universitaires, abordent la question du discours rapporté et de la citation, le font généralement sous l'angle des fonctions argumentatives et textuelles de la citation, et peu sous l'angle des formes de langue en jeu. De même, les manuels ou ressources à destination des étudiants soit se focalisent sur les normes universitaires de gestion des citations (citations courtes ou longues) et des références bibliographiques (par exemple Cislaru et al., 2020) ou bien, quand des exercices sont proposés, ils sont basés sur des modèles descriptifs inadéquats face aux difficultés rencontrées par les étudiants (voir entre autres Callet, 2012).

La démarche adoptée pour le traitement de la citation présente la double caractéristique d'une part de s'appuyer sur un modèle novateur, et d'autre part de prendre appui sur un double corpus, de copies d'étudiants et de textes d'experts (cf. Garnier et al., 2015). En effet les modèles traditionnels de description du discours rapporté, tels qu'ils apparaissent dans les grammaires scolaires et dans les manuels du secondaire, soulèvent un certain nombre de difficultés :

- Le champ du discours rapporté est souvent limité aux trois formes que sont le discours direct (DD), le discours indirect (DI) et le discours indirect libre (DIL), ce qui laisse de côté les modalisations introduites par un circonstant (*en selon X, pour X, d'après X ou comme le dit/explique/rappelle/montre/... X*) et les îlots textuels (*Le général Coligny conseille au roi une « politique de tolérance »*) – qui sont pourtant des ressources indispensables dans la rédaction d'un écrit universitaire.

- Ces trois formes sont généralement réduites à un panel de réalisations restreint, essentiellement issu d'un corpus romanesque opposant « parole du narrateur » et « parole du personnage ».
- Les manipulations (exercices) sont fréquemment centrées sur le passage du DD au DI et inversement, avec focalisation sur les modifications touchant les embrayeurs et les temps selon le modèle narratif prégnant. Comme le souligne Combettes (1989) l'approche développée dans les manuels repose sur un postulat théorique non explicité, celui de la « dérivation » des autres modes à partir du DD, jugé « premier » :

Dans la plupart des ouvrages de grammaire [...] le DR [discours rapporté] est rattaché ordinairement aux chapitres qui traitent de la subordination, de la concordance des temps, de la personne etc. Cette façon de décrire le DR recouvre un point de vue qui semble aller de soi et qui se trouve rarement remis en question ; les différentes variétés du DR, le discours direct (DD), le discours indirect (DI), le discours indirect libre (DIL) sont considérés comme liés les uns aux autres par des transformations ; moyennant quelques modifications, il serait possible de « passer » d'un énoncé de DD à un énoncé de DI, etc. [...] (Combettes, 1989, p. 111)

Outre ses difficultés intrinsèques, ce modèle, et les activités de transformation auxquels il donne lieu, ne rend pas compte des difficultés des étudiants ni non plus, d'ailleurs des pratiques des scripteurs experts. C'est ce que montre l'observation du corpus exploratoire sur lesquels on s'est appuyé pour élaborer les modules de remédiation et les questions du test de positionnement. Ce double corpus comprend (voir Boch et Grossmann, 2002, pour une approche analogue) :

- des copies d'étudiants.

Le projet présente la particularité d'être basé sur le relevé et l'analyse des « dysfonctionnements » effectivement observés dans des copies d'étudiants de différentes disciplines. Plusieurs centaines de copies de sciences du langage et d'histoire, et quelques dizaines de mémoires ont ainsi été dépouillés « à la main », la fréquence des phénomènes étant ici évaluée de façon intuitive. A ce stade, il s'agit d'un corpus exploratoire à visée didactique, qui ne répond pas à des critères de scientificité. Nous verrons dans la deuxième partie comment la suite du projet implique la constitution d'un corpus intégrant de façon raisonnée la diversité des disciplines et des genres.

- un corpus d'écrits « experts » (cf. sur ce point Boch & Rinck, 2015).

Constitué là encore de façon exploratoire, il contient des articles scientifiques de différentes disciplines des Sciences Humaines et Sociales (histoire, sociologie...) recueillis « à la main » et donnant lieu à une exploration qualitative.

L'analyse des copies fait ainsi apparaître différents éléments. Il apparaît tout d'abord qu'un grand nombre de difficultés sont de nature syntaxique, et touchent à l'insertion du segment entre guillemets de discours direct,

absence de ponctuation, difficultés dans l'emploi de « comme le dit X »⁴, emploi de « selon X » comme introducteur de discours direct⁵. Pour ces dernières expressions, les difficultés sont en particulier liées à leur portée différenciée selon leur position dans la phrase (position initiale ou non). Plus spécifiquement, un nombre important de dysfonctionnements concerne l'emploi des embrayeurs et déictiques, touchant à la confusion entre le discours direct, autonome syntaxiquement et énonciativement, et l'« îlot textuel », intégré syntaxiquement⁶.

On note également des problèmes de lexique, en particulier dans l'emploi des verbes de parole, soit par absence de précision lexicale, soit par des erreurs de catégorisation des actes de parole représentés ; là se situent en réalité les principales difficultés liées à l'emploi du discours indirect.

Au final, l'intégration du segment entre guillemets du discours direct semble constituer un des problèmes majeurs que rencontrent les étudiants. Or cette difficulté est en rapport avec la nature autonome de ce mode, tel que posé par le modèle sur lequel nous nous appuyons.

L'observation du corpus de textes d'experts a permis par ailleurs de faire apparaître la diversité des modes d'introduction du segment entre guillemets de discours direct. On note en particulier qu'à côté de la configuration « classique », recensée par les grammairiens, dans laquelle ce segment est complément d'un verbe de dire, il arrive fréquemment qu'il soit simplement précédé ou suivi par une ou plusieurs phrases qui en donnent le contenu sans qu'il soit « introduit » au sens syntaxique du terme⁷.

L'ensemble de ces constatations est en accord avec le modèle développé par Authier-Revuz (2020). Redéfinissant le discours rapporté, à la suite de Voloshinov, comme « du discours *sur* et *dans* le discours », ce modèle considère l'ensemble des phénomènes linguistiques impliqués par cette articulation de deux actes d'énonciation, à savoir : 1) les opérations métalangagières mise en jeu ; reformulation, autonymie, catégorisation; 2) les différentes modalités d'articulation des deux actes, selon que le discours représenté est l'objet ou la source du dire; 3) les modalités d'articulation des deux ancrages énonciatifs; 4) la double recontextualisation impliquée par cette représentation. Sont ainsi identifiés trois traits différentiels dont la combinaison produit cinq modes : le discours direct, le discours indirect, le Discours Indirect Libre-Bivocal, la modalisation en assertion seconde, la modalisation autonymique d'emprunt. Ces modes, définis à un niveau abstrait, correspondent à un éventail de formes dont l'actualisation en discours donne lieu à des effets de sens différents selon les genres. L'intérêt du modèle, qui s'inscrit

⁴ Dans un exemple comme *Le public est aussi un être d'émotion comme le montre Paul Claudel dans L'échange « il pleure et il rit »*, on ne sait pas sur quoi porte l'expression « comme le montre Paul Claudel ».

⁵ Exemple : *Selon Françoise Picot « Pourquoi un texte et pas seulement des phrases isolées? »*.

⁶ Dans la phrase *Jean-Sébastien Bach n'avait pas attendu Beethoven pour « faire entrer dans ses fugues un élément poétique »* il faut remplacer *mes* par *ses*.

⁷ *En donnant à voir les situations réelles des assistés, l'ethnographie a permis de rompre avec le discours officiel et médiatique. « Les ethnologues ont déconstruit le discours hégémonique sur la restructuration du welfare, en confrontant les idéologies dominantes à ce que l'on appelle les réalités de l'appauvrissement »* (Morgen, Maskovsky, 2003, p. 325) (Vincent Dubois « Ethnographie l'action publique », *Gouvernement et action publique* 1/2012 (n° 1), p. 83-101). Les éléments en gras indiquent les relations de synonymie-paraphrase entre la séquence qui précède le segment entre guillemets et celui-ci ; le segment entre guillemets vient « exemplifier » ce qui précède.

contre le modèle « dérivationniste » traditionnel, est multiple. Sur le plan des formes, il permet de rendre compte des formes de modalisation (et pas seulement du triptyque DD/DI/DIL), de la diversité des réalisations propres à chaque mode, et enfin de l'irréductible hétérogénéité du DD, analysé comme un segment autonome. Sur le plan textuel, il permet de comprendre comment la façon de représenter les dires des auteurs configure en discours une « image » de l'énonciateur.

Sur la base de l'analyse d'erreurs qui vient d'être présentée, nous avons conçu un programme de remédiation consacré à la maîtrise des formes de discours rapporté⁸ susceptibles d'être pratiquées par un public d'étudiants de licence. Ce programme s'inscrit dans la section « Discours »⁹ du site « J'améliore mon français écrit – Modules de français pour les études supérieures »¹⁰ hébergé par l'Université Paul-Valéry Montpellier 3. Intitulé « Citations », il est composé de quatre modules ; « Citations avec guillemets » (discours direct et modalisation autonymique d'emprunt), « Lexique de la citation » (discours direct et discours indirect), « Autres outils pour citer » (modalisation en assertion seconde) et « Synthèse ». Dans les lignes qui suivent, nous décrirons deux aspects majeurs de l'élaboration de ce dispositif, le premier lié à l'utilisation d'un logiciel et à ses contraintes pour la confection du support de remédiation, le second concernant le choix d'un lexique adapté à des lecteurs étudiants non spécialistes du champ théorique de la RDA. Nous analyserons ensuite quelques exercices proposés. Ce faisant, nous mettrons l'accent sur les contraintes et les apports du numérique dans la fabrication de ce matériel de remédiation.

2.2 Penser la remédiation avec le logiciel Scenari : la granularité comme ressource

Les quelques éléments que nous mentionnerons ici relèvent du vaste champ de réflexion consacré à l'incidence du choix des outils et des supports d'écriture, en particulier numériques, sur la production de matériel didactique. La rédaction du texte des quatre modules proposés a en effet été menée dans le cadre posé par Scenari, logiciel libre et gratuit destiné à la « conception et [à] la modélisation de documents »¹¹, notamment dans le domaine de l'enseignement, de la formation et de l'ingénierie pédagogique. La figure 2 montre un extrait du module que ce logiciel a permis d'élaborer, avec dans le bloc de gauche le sommaire et à droite le début de la fiche de cours, avec les trois propriétés données comme définitives du discours direct.

⁸ Nous utiliserons désormais la dénomination « Représentation du Discours Autre », soit RDA.

⁹ Outre « Discours », les autres sections sont les suivantes ; « Mots », « Phrase », « Texte ». Ces quatre domaines sont couverts par trente modules de remédiation au total.

¹⁰ http://uoh.univ-montp3.fr/j_ameliore_ma_maitrise_du_francais/portail/

¹¹ <http://scenari.org/>

Figure 2

J'ameliore mon franais 6crit : extrait du module « Citations avec guillemets »

Qu'est-ce que le discours direct ? **Fondamental**

- Dans les textes acad6miques, le segment au discours direct **est la plupart du temps entre guillemets** (sauf exception : voir fiche **Ponctuation du Discours Direct**).
- Le segment au discours direct est **autonome** par rapport aux mots de celui qui r6dige : il est pr6sent6 comme un « bloc » import6 tel quel dans ce qu'on appellera le « discours citant ».
- Les marques de 1^{re} et 2^{me} personne (*je, tu, nous, vous, mon, ton, notre, votre* ...), les temps verbaux, les indicateurs de temps et de lieu (*aujourd'hui, maintenant, ici, ...*) sont ceux de la situation dans laquelle celui que l'on cite produit son discours. C'est en lisant ce qui pr6c6de ou ce qui suit le segment entre guillemets que l'on peut identifier correctement a quoi renvoient les marques de personne et les temps.

2.2.1 Structuration des modules

Dans un objectif d'homog6n6isation de l'ensemble du site, chaque module respecte une m6me trame : une partie inaugurale « Objectifs » pr6sentant les connaissances vis6es en relation avec l'activit6 d'6criture universitaire (par exemple conna4tre les « diff6rentes faons d'int6grer une citation entre guillemets dans votre texte ») ; une « Introduction » annonant le plan et rappelant les probl6mes rencontr6s par l'6tudiant (« Comment introduire dans mon propre texte un segment de phrase venant d'ailleurs ? ») ; une « Conclusion » faisant le point sur les connaissances acquises ; et un d6veloppement. Chaque partie du d6veloppement d6bute par un exercice guid6 (« a vos stylos ») sensibilisant au probl6me trait6, puis des fiches de cours (par exemple, « Le discours direct ; comment l'introduire dans un texte ? ») alternent avec des exercices (par exemple, « Introduire le discours direct » et « Ponctuer au discours direct ») permettant a l'6tudiant d'6valuer son acquisition des connaissances.

Les contraintes r6dactionnelles li6es au choix du logiciel ont 6t6 plus sensibles d6s lors que le texte 6tait plus long, en l'occurrence dans les fiches de cours. Il a ainsi fallu adapter le contenu au mode d'organisation textuelle impos6 par le logiciel. De faon a rompre l'uniformit6 visuelle du seul niveau « texte » (pr6sent6, par d6faut, sur fond blanc), trois niveaux de granularit6 suppl6mentaires (« Fondamental », « Attention » et « Exemple ») ont 6t6 retenus. Cette diff6renciation, d'abord mat6rielle (chacun de ces niveaux 6tant associ6 a un fond de couleur distinct), nous a invit6es a remanier la hi6rarchisation des 6l6ments de contenu du cours en fonction de l'hypoth6se d'une lecture du texte tabulaire et s6lective (et non lin6aire et exhaustive). Dans l'optique du d6veloppement de supports de rem6diation en ligne, et de faon a affiner la hi6rarchisation des contenus de cours, il nous semble qu'il serait n6cessaire de s'appuyer sur des 6tudes pr6cises des modalit6s de lecture sur 6cran (d'ordinateur ou de t6l6phone) en usage aupr6s d'un public d'6tudiants de licence.

2.2.2 Choisir une pr3sentation et un lexique adapt3s : entre th3orie et usage situ3

La publication sur Internet du support de rem3diation induit une diffusion large. La question du choix d'un lexique adapt3 3 une masse d'3tudiants non sp3cialistes est ainsi apparue comme centrale lors de l'3laboration des modules. Nous avons recherch3 des solutions permettant de concilier les propositions th3oriques et terminologiques du cadre innovant de la « Repr3sentation du Discours Autre » (voir 2.1.) avec les besoins des 3tudiants non sp3cialistes. Il s'agissait en effet non pas de leur donner acc3s 3 une connaissance g3n3rale du ph3nom3ne de la RDA, mais de leur fournir une aide relative 3 l'une des pratiques de la RDA, li3e en l'occurrence 3 la litt3racie universitaire et suppos3e identifi3e par l'3tudiant sous le nom de « citation ».

Pour parvenir 3 concilier ces objectifs et ces terminologies distinctes, nous avons privil3gi3 un d3coupage en modules en fonction des marques (guillemets, lexique, pr3positions) amen3es 3 3tre manipul3es par l'3tudiant et non en fonction des diff3rents modes de RDA. Un m3me mode peut ainsi 3tre trait3 dans deux modules. Il en va ainsi par exemple du discours direct qui est abord3 3 la fois dans le module « Citations avec guillemets », pour ce qui est du traitement de la partie autonome de la forme, et dans le module « Lexique de la citation », pour ce qui est du choix des verbes ou noms susceptibles de cat3goriser l'acte d'3nonciation repr3sent3.

La primeur donn3e 3 l'approche de l'3tudiant en qu3te d'outils de rem3diation s'est 3galement traduite dans la formulation des titres. Nous avons ainsi privil3gi3 le point de vue de l'3tudiant dans les titres de premier niveau qui ne contiennent pas de terminologie sp3cialis3e dans l'3tude du discours rapport3 (« Citations avec guillemets », « Lexique de la citation », « Autres outils pour citer »). Lorsque le nom d'un mode de RDA 3tait connu du fait de son usage au cours des 3tudes secondaires, ainsi de « discours direct », ou lorsque le nom du ph3nom3ne 3tait imag3 et susceptible d'aider 3 la compr3hension du ph3nom3ne, ainsi de « 3lot textuel », nous l'avons fait figurer dans les titres de second niveau et dans le d3veloppement.

2.2.3 Les exercices : entra3ner la comp3tence m3talangag3re de l'3tudiant au niveau du texte

Le logiciel Scenari n'autorisant pas la correction automatique de productions textuelles longues, le programme d'exercices a 3t3 con3u pour entra3ner la comp3tence m3talangag3re de l'3tudiant sp3cifiquement sur les formes de RDA. L'hypoth3se sous-jacente 3 ce choix est que l'aptitude 3 juger de la qualit3 d'une forme de RDA lors de l'activit3 de lecture contribuerait 3 la ma3trise desdites formes dans le cadre de l'activit3 d'3criture. Sur cette base, nous avons choisi les types d'exercices propos3s par Scenari les mieux 3 m3me d'exercer cette comp3tence m3talangag3re 3 la lecture.

Des questionnaires à choix multiples invitent ainsi à juger de la bonne formation d'une forme de RDA en fonction des éléments de contenu dispensés dans le cours¹². Les questions portent sur des extraits de texte recueillis dans des copies d'étudiants ou dans des écrits académiques. L'exercice à trou (plusieurs items étant proposés pour remplir un espace blanc) nous a également semblé intéressant dans la mesure où il permet d'associer l'entraînement de l'activité métalangagière à la mise en situation de production écrite¹³.

Enfin, nous avons opté pour l'exercice reposant sur l'utilisation du curseur sur un texte¹⁴ car il permet à la fois d'être attentif aux marques permettant de repérer la présence d'un acte d'énonciation représenté et d'apprécier la dimension textuelle du discours rapporté, une forme n'étant alors plus considérée isolément, mais dans sa relation aux autres formes présentes dans le texte. L'apprentissage de cette dimension textuelle de la RDA, favorisée par l'outil numérique, nous semble la plus originale et gagnerait à être développée, l'étudiant étant en effet amené, dans sa pratique, à produire des textes et non uniquement des phrases.

3. Perspectives : prendre en compte les genres et les disciplines

Pour affiner la description des dysfonctionnements et améliorer le programme de remédiation proposé au travers du dispositif qui a été présenté dans la partie précédente, la constitution d'un corpus raisonné qui prenait en compte la variété des disciplines et des genres de discours s'est avérée nécessaire. Nous présentons ici le travail de constitution d'un corpus de copies d'examen, ainsi que les procédés de numérisation et de transcription de ces données, qui ont eu pour but d'identifier les dysfonctionnements concernant les formes de RDA. Nous aborderons ensuite la nécessité de recourir à l'annotation du corpus pour le traitement automatique de textes, en raison du grand nombre de données.

3.1 De la collecte à la numérisation et à l'analyse du corpus

Pour établir un corpus raisonné, nous avons recueilli 895 copies d'examen dans les archives de l'Université Paris Nanterre, après avoir dépouillé 17 boîtes d'archives au contenu très varié, soit environ 7000 copies d'examen de Licence 1 à Master 2 des filières suivantes : Arts du Spectacle, Humanités, Philosophie, Économie-Gestion, Histoire, Administration Économique et Sociale, Histoire de l'Art et de l'Archéologie et Management des Événements et des

¹² Exemple de consigne de QCM: «Les extraits suivants contiennent des îlot textuels. Certains extraits ne respectent pas les règles décrites dans la fiche. Les problèmes peuvent concerner l'emploi des pronoms ou la ponctuation. Cochez les extraits qui vous semblent corrects».

¹³ La consigne de cet exercice est la suivante: «Dans les extraits suivants, on vous propose plusieurs verbes dont le verbe *décrire* pour introduire la citation. Choisissez le verbe qui convient le mieux; ce n'est pas toujours *décrire*. En revanche, il peut y avoir plusieurs bonnes réponses».

¹⁴ La consigne de cet exercice est la suivante: «Le texte suivant est une critique de deux ouvrages sur la justice. En faisant passer votre curseur sur le texte, identifiez les mots, verbes ou noms, qui introduisent le discours des auteurs cités».

Loisirs Sportifs. Mis à part les copies, les boîtes d'archives peuvent aussi contenir les consignes de chaque épreuve et la liste d'émargement qui atteste de la présence de l'étudiant le jour de l'examen.

Ce corpus est considéré comme représentatif des usages de la RDA dans la copie d'examen à l'université pour la fin des années 2010. Par ailleurs, il n'est pas exhaustif, étant donné que chaque unité de formation et de recherche (UFR) de l'Université Paris Nanterre conserve ses propres archives de copies d'examen au sein des secrétariats qui, dès lors qu'ils manquent de place, contactent le service des archives pour donner ces boîtes de documents qui seront généralement conservées un an avant d'être détruites¹⁵. Ce corpus prend en compte la variété des disciplines, des genres de discours ainsi que des niveaux d'études.

En raison de contraintes temporelles, sur les 895 copies numérisées, seules 250 ont pu être analysées. Comme il s'agit d'écrits manuscrits, la reproduction photographique de ces textes a servi de base à une transcription linéarisée (Testenoire, 2017) des formes de RDA employées de manière erronée.

Comme nous l'avons vu dans la première partie, un référentiel a été élaboré en matière de compréhension et de production textuelle, divisé par domaines (discours, phrase, texte, mot) et par champs de compétences (marques de personne, propositions relatives, relations entre constituants, etc.). Nous nous sommes concentrées sur les erreurs de la catégorie « discours », qui incluent le discours direct, le discours indirect, les formes « Selon X, ... » et les îlots textuels. Par ailleurs, après le classement des erreurs repérées dans les copies, il fallait identifier le genre de discours de chaque production écrite pour retrouver ses caractéristiques principales. Enfin, on relevait le lexique spécifique de la discipline et aussi le lexique universitaire¹⁶.

En ce qui concerne la maîtrise de l'usage des formes de RDA, les résultats confirment les observations menées sur le corpus exploratoire : les erreurs portent majoritairement sur l'emploi du discours direct. En général, les étudiants ajoutent des phrases entre guillemets sans les introduire ou les commenter comme on peut voir dans cet exemple tiré d'une copie d'examen d'un étudiant de L1 en Arts du Spectacle :

Aussi, Diderot théorise la présence du quatrième sur la scène de théâtre. L'attention de l'auteur n'est alors plus portée sur la salle et les attentes du spectateur, mais bien sur la scène, les personnages et les relations entre eux. Le but n'est plus seulement de plaire au public. « Diderot avait déjà théorisé le cinéma » Eisenstein. Cela change beaucoup le théâtre et son esthétique, notamment sur le jeu de l'acteur.

Les genres de discours les plus fréquents dans les copies recueillies sont la dissertation, la synthèse et le commentaire (de texte, de document, d'œuvre d'art, de film, etc.). Il est important de remarquer que les caractéristiques de chaque genre varient en fonction des disciplines, mais aussi en fonction des enseignants. Ainsi, une dissertation en Histoire, par exemple, comporte des sous-titres dans chaque partie du texte, ce qui n'est pas le cas en

¹⁵ Ces informations sont issues d'un entretien avec Marion Trousselle, responsable des archives de l'Université Paris Nanterre.

¹⁶ Voir la distinction faite par Beck et al. (2012) entre étage 1 du lexique élémentaire, étage 2 du lexique académique transdisciplinaire, et étage 3 des lexiques disciplinaires.

Arts du Spectacle ou en Administration Économique et Sociale. Par ailleurs, à partir des commentaires des correcteurs, on a repéré qu'en Histoire l'usage du futur historique est interdit. De la même manière, une synthèse peut avoir des caractéristiques différentes au sein de la même discipline, car chaque enseignant donne des consignes distinctes, ce qui peut représenter une grande difficulté pour les étudiants puisque les genres ne sont pas bien définis et par conséquent pas bien maîtrisés.

En tout état de cause, l'examen de ces données montre que la maîtrise de l'écriture est une compétence qui se développe effectivement tout du long de la formation universitaire¹⁷, puisque on a vérifié que les erreurs commencent à diminuer sensiblement en 3eme année de licence et en master, notamment.

3.2 L'annotation pour le traitement automatique des données

Étant donné qu'il est impossible d'analyser manuellement toutes les copies d'examen collectées, l'annotation du corpus s'est révélée indispensable pour avoir recours au traitement automatique des données, qui permettra une description et une classification plus fine des formes de discours rapporté par genres discursifs. Le traitement automatique des textes auquel nous ferons appel sera mené sur la base d'une annotation réalisée à l'aide de logiciels adaptés – comme le logiciel Inception et la plateforme d'annotation et d'exploration de corpus en ligne Sarramanka, développée par Sarra El Ayari, ingénieure du laboratoire Structures Formelles du Langage (CNRS et Université Paris 8).

L'indexation automatique permettra ainsi de corrélér plus finement la fréquence des différentes formes aux variables que sont la discipline, le niveau d'études et le genre de discours. Ces éléments donneront des pistes d'analyse permettant de réfléchir à une catégorisation des formes difficilement repérables de RDA, les formes non marquées. Pour vérifier leur distribution et leur fréquence, toutes les formes seront relevées, ce qui nous amènera à nous interroger sur les formes considérées comme non normées et aussi sur les formes sanctionnées par l'enseignant et/ou le correcteur.

En comprenant mieux les caractéristiques du support copie d'examen à partir de l'annotation du corpus, nous pourrons mieux connaître les spécificités de la citation, cette pratique emblématique du discours universitaire, et identifier en parallèle les difficultés des étudiants. A partir de ces constats, on pourra proposer des remédiations plus adaptées aux besoins de chaque discipline dans le cas des formes considérées comme erronées.

¹⁷ Voir Boch et Grossmann (2002) et Boré (2004).

Conclusion

Nous avons voulu montrer dans cet article, à partir de l'expérience menée dans le cadre du projet écri+, un exemple précis d'interaction entre les possibilités et ressources fournies par l'outil numérique, la recherche en linguistique et la réflexion didactique et pédagogique. Nous avons particulièrement insisté sur la place des données, via la constitution de corpus de différentes natures, corpus dont la constitution et l'exploitation mettent également en œuvre des outils d'annotation et de traitement automatique. Nous avons enfin montré comment, au fil de l'avancement du projet, la question des genres de discours avait pu être prise en compte, au plus près des conditions de production des textes.

Bibliographie

- Authier-Revuz, Jacqueline (2020). *La représentation du discours autre. Principes pour une description*. De Gruyter. <http://doi.org/10.1515/9783110641226>
- Beck, Isabel L. ; McKeown, Margaret G. ; Kucan, Linda (2013). *Bringing Words to Life : Robust Vocabulary Instruction* (2e éd.). Guilford Press.
- Benveniste, Émile (1974). Les niveaux de l'analyse linguistique. À *Problèmes de linguistique générale* (vol. 2) (p. 119-131). Gallimard.
- Boch, Françoise ; Grossmann, Francis (2002). Se référer au discours d'autrui : quelques éléments de comparaison entre experts et néophytes. *ENJEUX*, 54, 41-51.
- Boch, Françoise ; Frier, Catherine (eds.) (2015). *Écrire dans l'enseignement supérieur*. ELLUG.
- Boch, Françoise ; Rinck, Fanny (2015). Anaphores démonstratives dans les écrits d'étudiants de Master : comparaison avec les pratiques expertes. *Linx*, 72, 131-150.
- Boré, Catherine (2004). Discours rapportés dans les brouillons d'élèves : vrai dialogisme pour une polyphonie à construire. *Pratiques : linguistique, littérature, didactique*, 123-124, 143-169.
- Bouquet, Bénédicte (2009). Diversité et enjeux des écrits professionnels. *Vie sociale*, 2, 81-93. <http://doi.org/10.3917/vsoc.092.0081>
- Boutet, Josiane (2014). Les écrits au travail. À Isabelle Laborde-Milla, Sylvie Plane, Rinck Fanny et Sitri Frédérique (éds.), *La formation aux écrits professionnels : des écrits en situation de travail aux dispositifs de formation* (p. 17-27). EME éditions.
- Callet, Stéphanie (2012). *Les discours direct et indirect. Règles, exercices et corrigés*. De Boeck.
- Cislaru, Georgeta ; Claudel, Chantal ; Vlad, Marina (2020) (4eme éd.). *L'écrit universitaire en pratique*. De Boeck.
- Combettes, Bernard (1989). Discours rapporté et énonciation : trois approches différentes. *Pratiques*, 64, 111-122. <http://doi.org/10.3406/prati.1989.1601>
- Delcambre, Isabelle ; Lahanier-Reuter, Dominique (éds.) (2012). Littéracies universitaires : nouvelles perspectives. *Pratiques. Linguistique, littérature, didactique*, 153-154. <http://doi.org/10.4000/pratiques.1902>
- De Luca, Gabriella (2021). *La citation dans les copies d'examen à l'université ; pistes pour une étude de la représentation du discours autre dans les copies d'examen de licence* [Mémoire de master non publié]. Université Paris Nanterre.
- Garnier, Sylvie ; Rinck, Fanny ; Sitri, Frédérique ; et De Vogüé, Sarah (2015). Former à l'écrit universitaire, un

terrain pour la linguistique ? *Linx*, 72, 1-9.

Goody, Jack (1977) (1979 pour la traduction franaise). *La raison graphique*. Minuit.

Heiden, Serge, Magu3, Jean-Pierre, et Pincemin, B3n3dicte. (2010). TXM : Une plateforme logicielle open-source pour la textom3trie – conception et d3veloppement.  Sergio Bolasco, Isabella Chiari et Luca Giuliano (eds.), *Proceedings of 10th International Conference on the Statistical Analysis of Textual Data - JADT 2010* (vol. 2) (p. 1021-1032). Edizioni Universitarie di Lettere Economia Diritto. http://halshs.archives-ouvertes.fr/docs/00/54/97/79/PDF/Heiden_al_jadt2010.pdf

Laborde-Milaa, Isabelle ; Plane, Sylvie ; Rinck, Fanny ; Sitri, Fr3d3rique. (2015). *La formation aux 3crits professionnels : des 3crits en situation de travail aux dispositifs de formation. Le discours et la langue*. EME 3ditions.

Reuter, Yves (2006). A propos des usages de Goody en didactique. 3l3ments d'analyse et de discussion. *Pratiques*, 131-132, 131-154. <http://doi.org/10.3406/prati.2006.2124>

Rot, Gwenaelle ; Borzeix, Annie ; Demazi3re, Didier (2014). Ce que les 3crits font au travail. *Sociologie du travail*, 56(1), 4-15. <http://doi:10.1016/j.soctra.2013.12.022>

Testenoire, Pierre-Yves (2017). Transcrire des 3crits scolaires : entre philologie et g3n3tique textuelle. *Corpus*, 16. <http://doi.org/10.4000/corpus.2762>

