



L'anglais de spécialité : des structures linguistiques à l'apprentissage d'une langue fortement codée

Laura Hartwell

► **To cite this version:**

Laura Hartwell. L'anglais de spécialité : des structures linguistiques à l'apprentissage d'une langue fortement codée . Linguistique. École Doctorale Langues littératures et sciences humaines de l'UGA, 2016. <tel-01333475>

HAL Id: tel-01333475

<https://hal.archives-ouvertes.fr/tel-01333475>

Submitted on 17 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

RAPPORT DE SYNTHÈSE

en vue d'obtenir

**L'HABILITATION À DIRIGER DES
RECHERCHES**

UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Linguistique anglaise et sciences du langage**

Arrêté ministériel : 7 août 2006

Présenté par

Laura M. HARTWELL

Garante : Agnès Tutin

Garant : John Osborne

préparé au sein du **Laboratoire LIDILEM**

dans l'**École Doctorale Langues littératures et sciences
humaines**

**L'anglais de spécialité :
des structures linguistiques à
l'apprentissage d'une langue
fortement codée, Volume I**

Rapport de synthèse soutenu publiquement le
14 avril 2016 devant le jury composé de :

Mme. Shirley CARTER-THOMAS

PU à l'Institut Mines-Télécom, Rapporteure,

Mme. Marie-Christine DEYRICH

PU (*émérite*) à l'Université Montesquieu, Rapporteure,

Mme. Kjersti FLØTTUM

PU à l'Université de Bergen, Présidente,

M. François MANIEZ

PU à l'Université de Lyon 2, Rapporteur,

M. John OSBORNE

PU à l'Université Savoie Mont-Blanc, Garant,

Mme. Agnès TUTIN

PU à l'Université de Grenoble-Alpes, Garante.



Il semble aujourd'hui crucial de dessiner une stratégie à court et moyen terme qui permette de repenser notre façon d'enseigner et de former. L'accès à l'information s'est radicalement transformé dans les dernières années du fait des outils numériques et des objets communicants, porteurs de la construction des connaissances. Les méthodes de travail dans les entreprises ont également été bouleversées. Pourtant, notre façon d'enseigner reste largement en décalage avec ces révolutions. Il faut reconnaître l'engagement des équipes enseignantes dans cette transformation, cette révolution médiatisée que nous voulons mettre en œuvre.

(P. Lévy, Président de l'UJF, 2015)

Le langage est une législation, la langue en est le code. Nous ne voyons pas le pouvoir qui est dans la langue, parce que nous oublions que toute langue est un classement, et que tout classement est oppressif : ordo veut dire à la fois répartition et commination. [...] Dans notre langue française (ce sont là des exemples grossiers), je suis astreint à me poser d'abord en sujet, avant d'énoncer l'action qui ne sera plus dès lors que mon attribut : ce que je fais n'est que la conséquence et la consécution de ce que je suis ; de la même manière, je suis obligé de toujours choisir entre le masculin et le féminin, le neutre ou le complexe me sont interdits ; de même encore, je suis obligé de marquer mon rapport à l'autre en recourant soit au tu, soit au vous : le suspens affectif ou social m'est refusé. [...] Parler, et à plus forte raison discourir, ce n'est pas communiquer, comme on le répète trop souvent, c'est assujettir : toute la langue est une réaction généralisée. [...] Cette tricherie salutaire, cette esquive, ce leurre magnifique, qui permet d'entendre la langue hors pouvoir, dans la splendeur d'une révolution permanente du langage, je l'appelle pour ma part : littérature.

(R. Barthes, Professeur au Collège de France, 1977).

Remerciements

J'exprime d'abord toute ma gratitude à Agnès Tutin et à John Osborne pour avoir accepté généreusement ce travail d'encadrement, d'exigence et de conseil. Leur réactivité, leurs connaissances et leur générosité ont été précieuses.

Je remercie sincèrement la présidente, l'ensemble des rapporteurs et les membres du jury qui m'ont fait l'honneur de consacrer leur temps et de partager leur sagesse. Je remercie tout particulièrement François Maniez pour ses commentaires approfondis.

Merci aux membres du laboratoire LIDILEM, surtout de l'axe 1, qui m'apporte un cadre de travail chaleureux, créatif et productif. Je remercie également l'équipe dynamique du Service des langues à l'UGA et Emmanuelle Esperança-Rodier pour son œil de lynx.

J'éprouve une profonde gratitude pour Nicole Décuré qui m'a accompagnée avec franchise pendant toutes ces années de recherche.

Enfin, j'émets une pensée affectueuse pour ma famille qui m'a tant soutenue dans cette démarche chronophage.

TABLE DES MATIÈRES

LISTE DES ABRÉVIATIONS	11
PRÉSENTATION DU DOSSIER	13
LISTE DE FIGURES	15
LISTE DE TABLEAUX	17
<i>PARCOURS DE RECHERCHE</i>	19
<i>ORIENTATIONS DE LA NOTE DE SYNTHÈSE</i>	22
<i>UN FIL DIRECTEUR</i>	23
1. PREMIÈRE PARTIE : APPROCHES THÉORIQUES, REPÈRES HISTORIQUES ET ÉPISTÉMOLOGIQUES	25
1.1 TROIS SPHÈRES	25
1.1.1. <i>L'ANGLAIS DE SPÉCIALITÉ</i>	25
1.1.2. <i>UNE APPROCHE FONDÉE SUR LA LINGUISTIQUE DE CORPUS</i>	28
1.1.3. <i>L'APPROCHE SOCIO-CONSTRUCTIVISTE DE LA DIDACTIQUE</i>	30
1.1.4. <i>TROIS SPHÈRES INDISSOCIABLES</i>	31
1.2. LES DÉBUTS ET LES NOTIONS CLÉS DE LA LINGUISTIQUE DE CORPUS	34
1.2.1. <i>RÉPÈRES HISTORIQUES</i>	34
1.2.1.1. L'APPORT DU SURVEY OF ENGLISH USAGE	35
1.2.1.2. L'APPORT DE L'OSTI REPORT	36
1.2.1.3. UNE TRADITION DE GRANDS CORPUS GÉNÉRAUX	37
1.2.2. <i>QUALITÉS COMBINATOIRES</i>	39
1.2.2.1. VERS UNE DÉFINITION DE LA COLLOCATION	39
1.2.2.2. VERS UNE DÉFINITION DE LA COLLIGATION	41
1.2.2.3. AUTRES APPROCHES ET TYPES D'ASSOCIATIONS	42
1.2.2.4. APPLICATIONS DIDACTIQUES	43
1.2.3. <i>LE GENRE AU SEIN D'UNE COMMUNAUTÉ DE DISCOURS</i>	44
1.2.3.1. DÉBUT SYMBOLIQUE DE L'ANGLAIS DE SPÉCIALITÉ	44
1.2.3.2. COMMUNAUTÉ DE DISCOURS, GENRE ET BUT RHÉTORIQUE	45
1.2.4. <i>SITUATION SOCIO-RHÉTORIQUE ET DIDACTIQUE</i>	46
1.2.4.1. TÂCHES EN CONTEXTE D'APPRENTISSAGE	46
1.2.4.2. LES COLLOCATIONS ET L'APPRENTISSAGE	47
1.3. TROIS PHÉNOMÈNES LINGUISTIQUES DE L'ANGLAIS DES SCIENCES	49
1.3.1. <i>CERNER LA CAUSALITÉ</i>	49
1.3.1.1. VERBES DE CAUSALITÉ	49
1.3.1.2. CONSTRUCTIONS CAUSATIVES	51
1.3.1.3. CAUSALITÉ EXPRIMÉE PAR LOCUTION	51
1.3.2. <i>LA MÉTAPHORE GRAMMATICALE</i>	52
1.3.2.1. APPROCHE SOCIO-FONCTIONNELLE	52

1.3.2.2.	ANGLAIS DES SCIENCES	52
1.3.2.3.	ÉLÉMENTS DES GROUPES NOMINAUX	53
1.3.2.4.	GLISSEMENT GRAMMATICAL	56
1.3.3.	<i>PRÉSENCE AUCTORIALE ET POSITIONNEMENT</i>	57
1.3.3.1.	SUBJECTIVITÉ	57
1.3.3.2.	L'APPORT DE HYLAND	57
1.3.3.3.	LA PRÉSENCE AUCTORIALE EXPRIMÉE PAR UN PRONOM	58
1.3.3.4.	PORTÉE CULTURELLE ET DIDACTIQUE	59
1.3.3.5.	DIVERSITÉS D'APPROCHES	59
	RÉSUMÉ DE LA PREMIÈRE PARTIE	60
	2. DEUXIÈME PARTIE : ANALYSE DES PRATIQUES, PROPOSITIONS DIDACTIQUES ET MOYENS D'ÉVALUATION	61
	2.1. CADRES INSTITUTIONNELS DE L'ENSEIGNEMENT DES LANGUES	61
2.1.1.	<i>L'ANALYSE DU DISCOURS AU SERVICE DES OBJECTIFS ÉDUCATIFS</i>	61
2.1.1.1.	LE CONSEIL DE L'EUROPE	61
2.1.1.2.	LE RÔLE DE L'ÉTAYAGE	62
2.1.1.3.	L'ANALYSE DE L'ACCESSIBILITÉ	63
2.1.1.4.	LES COMPÉTENCES CATALOGUÉES DANS LE CADRE EUROPÉEN COMMUN DE RÉFÉRENCE POUR LES LANGUES	64
2.1.1.5.	UNE RÉPONSE FRANÇAISE	64
2.1.2.	<i>LA MODÉLISATION DU PROCESSUS D'ENSEIGNEMENT</i>	65
2.1.2.1.	DÉBAT AUTOUR DES THÉORIES DE L'ACQUISITION DES LANGUES ÉTRANGÈRES	67
2.1.2.2.	UN MODÈLE DE RÉLATION D'AGENCE	68
2.1.2.3.	UN MODÈLE D'ERGONOMIE DIDACTIQUE	70
2.1.3.	<i>UN MODÈLE SYSTÉMIQUE DE L'ENSEIGNEMENT</i>	72
2.1.3.1.	PROJET INTERCULTUREL MÉDIATISÉ PAR UN WIKI COLLABORATIF	74
2.1.3.2.	DISPOSITIF DE SUPPORT EN LIGNE : ENGLISH BOOSTER COURSE FOR MASTER'S DEGREE STUDENTS (EBCM)	78
2.1.3.3.	DISPOSITIF BASÉ À PARTIR DE CORPUS : CORPUS4ALL	82
	2.2. RESSOURCES ET PROPOSITIONS DIDACTIQUES	85
2.2.1.	<i>LE LEXIQUE ACADÉMIQUE ET SCIENTIFIQUE</i>	85
2.2.1.1.	L'INTRODUCTION DE LISTES LEXICALES ACADÉMIQUES	85
2.2.1.2.	LA RÉPRÉSENTATION DES SCIENCES	87
2.2.1.3.	RÉPONDRE AUX BESOINS LEXICAUX DES ÉTUDIANT-ES	87
2.2.1.4.	L'APPORT DES DICTIONNAIRES	88
2.2.1.5.	L'APPORT DES CORPUS À DESTINATION DES NON LINGUISTES	90
2.2.2.	<i>ASPECTS DIDACTIQUES</i>	91
2.2.2.1.	USAGE-BASED GRAMMAR	91
2.2.2.2.	ÉQUILIBRE ENTRE CONTENU ET ACCESSIBILITÉ AU CORPUS	92
2.2.3.	<i>LE CONTEXTE INITIAL DU PROJET ENIGMA PLUS</i>	93
2.2.3.1.	L'INTÉRÊT DE LA RECONSTRUCTION DE TEXTE	93
2.2.3.2.	PROJET DE L'ÉTABLISSEMENT	95
2.2.3.3.	ANALYSE DES PRATIQUES	95
2.2.3.4.	AJUSTER LES PARAMÈTRES EN FONCTION DES PRATIQUES	95

2.2.4.	<i>ANALYSE DES PRATIQUES EN ENVIRONNEMENT MÉDIATISÉ</i>	96
2.2.4.1.	PRATIQUES INDIVIDUALISÉES	96
2.2.4.2.	PRENDRE EN COMPTE L'ASPECT SOCIAL	98
2.3.	ÉVALUATION DES COMPÉTENCES	99
2.3.1.	<i>LA PRESSION DE LA CERTIFICATION</i>	99
2.3.1.1.	ADAPTION AUX BESOINS PROFESSIONNELS	99
2.3.1.2.	PROPOSITION D'ÉVALUATION SUR MESURE	100
2.3.2.	<i>LA COMPRÉHENSION DE DOCUMENTS ORAUX</i>	100
2.3.2.1.	ADÉQUATION DU CONTENU ET DE LA POPULATION	100
2.3.2.2.	UNE APPROCHE PAR LES TÂCHES	101
2.3.2.3.	L'AUTENTICITÉ DES TÂCHES	102
2.3.3.	<i>LA PRISE DE NOTES</i>	103
2.3.3.1.	DEMANDES COGNITIVES	103
2.3.3.2.	GRILLE D'ÉVALUATION DANS UN CONTEXTE HÉTÉROGÈNE	103
2.3.3.3.	CORRÉLATION DES RÉSULTATS	105
2.3.3.4.	ÉTAYAGE MÉDIATISÉ	106
2.3.3.5.	PORTÉE TRANSDISCIPLINAIRE	109
2.3.4.	<i>LE DÉCALAGE ENTRE BESOINS ET CERTIFICATION</i>	109
2.3.4.1.	EFFET DE L'OBLIGATION DE CERTIFICATION	109
2.3.4.2.	MANQUE DE RECHERCHE CRITIQUE	110
2.3.4.3.	EFFET DE RETOUR	111
2.3.4.4.	REPRÉSENTATIONS SOCIÉTALES MISES À MAL	112
RÉSUMÉ DE LA DEUXIÈME PARTIE		112
3.	TROISIÈME PARTIE : ANALYSE DU DISCOURS DES EXPERT.E.S	113
3.1.	MÉTHODES ET MOYENS D'ANALYSE DU DISCOURS DES EXPERT.E.S	114
3.1.1.	<i>UNE APPROCHE FONDÉE SUR DES CORPUS SPÉCIALISÉS</i>	114
3.1.1.1.	JOHN SWALES CONFERENCE CORPUS	114
3.1.1.2.	CORPUS DE L'ÉTUDE INTERDISCIPLINAIRE ET INTERLINGUISTIQUE DU DISCOURS ACADÉMIQUE	115
3.1.1.3.	LE CORPUS D'IDENTITÉ CULTURELLE DANS LA PROSE ACADÉMIQUE	116
3.1.1.4.	CORPUS SCIENTEXT	116
3.1.2.	<i>LA PUISSANCE DES OUTILS : TXM, PLATEFORME D'ANALYSE TEXTUELLE</i>	118
3.1.2.1.	L'IMPORTANCE DES OUTILS	118
3.1.2.2.	UNE APPROCHE QUANTITATIVE	119
3.1.2.3.	LES AUXILIAIRES MODAUX EN LINGUISTIQUE DE L'ORAL	119
3.1.2.4.	LIEN ENTRE L'ANGLAIS GÉNÉRAL ET L'ANGLAIS DE SPÉCIALITÉ	121
3.1.2.5.	FRÉQUENCE DE COOCCURRENTS	123
3.1.2.6.	CALCULER LA SIGNIFICATION STATISTIQUE	123
3.1.3.	<i>L'INTÉRÊT DE L'ANNOTATION SYNTAXIQUE DE SCIENTEXT</i>	124
3.1.3.1.	UNE VISION GLOBALE GRÂCE AUX STATISTIQUES	125
3.1.3.2.	LA QUALITÉ LINGUISTIQUE DU CORPUS	128
3.1.3.3.	ÉVALUER L'EXISTENCE D'ADVERBE ENTRE <i>TO</i> ET UN VERBE (<i>SPLIT INFINITIVES</i>)	129
3.1.3.4.	INVESTIR DANS UN CORPUS	129
3.1.3.5.	RICHESSSE ET PORTÉE DIDACTIQUE DE LA PHRASÉOLOGIE TRANSLINGUISTIQUE	130

3.2. QUELQUES NOTIONS LINGUISTIQUES CONSTITUTIVES DE L'ANGLAIS DES SCIENCES	131
3.2.1. CAUSALITÉ OU SUBJECTIVITÉ : TROIS ADVERBES D'ARTICULATION	132
3.2.1.1. CONTRASTES D'EMPLOI	132
3.2.1.2. CONTRASTES DE FRÉQUENCES	133
3.2.1.3. PATRONS LEXICO-GRAMMATICAUX	134
3.2.1.4. CONTRASTES SÉMANTIQUES	135
3.2.1.5. CONTRASTES DE VARIATIONS LEXICALES	136
3.1.2.7. PORTÉE DIDACTIQUE	137
3.2.2. SUBJECTIVITÉ : HOPE ET WISH	138
3.2.2.1. CONTRIBUTION AU DÉBAT AUTOUR DU MODALISATEUR <i>HOPEFULLY</i>	138
3.2.2.2. DIFFÉRENCES SÉMANTIQUES DE <i>HOPE</i> ET <i>WISH</i>	139
3.2.2.3. RÉSULTATS DES REQUÊTES DE <i>HOPE</i> ET <i>WISH</i>	139
3.1.3.6. CONTRASTES CONTEXTUELS	140
3.1.3.7. QUALITÉ SUBJECTIVE	141
3.2.3. NEUTRALITÉ OU QUALITÉS INTERPERSONNELLES	142
3.2.3.1. DIFFÉRENCES ENTRE LE GENRE ÉCRIT ET ORAL EN ANGLAIS DES SCIENCES	142
3.2.2.2. PORTÉE DIDACTIQUE	142
3.2.2.3. RÉSULTATS D'UNE ANALYSE DE FRÉQUENCE LEXICALE	143
3.2.2.4. CONCLUSIONS AUTOUR DE <i>THAT</i> SELON LE GENRE	143
3.2.3.5. FORTE PRÉSENCE DE <i>SO</i> À L'ORAL	144
3.2.3.6. RECOURS AUX PRONOMS PERSONNELS	146
3.2.3.7. ASPECTS PRAGMATIQUES	147
RÉSUMÉ DE LA TROISIÈME PARTIE	148
PERSPECTIVES	150
CONCLUSION GÉNÉRALE	152
INDEX DES TRAVAUX	154
INDEX DES AUTEURS	156
LISTE DES ANNEXES	159
BIBLIOGRAPHIES	174
<i>RÉFÉRENCES AUX TRAVAUX PERSONNELS</i>	174
<i>PILOTAGE DE PROJETS DE PÉDAGOGIE NUMÉRIQUE</i>	176
<i>CORPUS ET OUTILS RÉFÉRENCÉS</i>	177
<i>BIBLIOGRAPHIE</i>	178
<i>CURRICULUM VITAE</i>	ERREUR ! SIGNET NON DÉFINI.

LISTE DES ABRÉVIATIONS

ALSIC : Apprentissage des Langues et Systèmes d'Information et de Communication (revue)
ALMT : Apprentissage des langues médiatisé par les technologies
ALTE : Association of language testers in Europe
ANOVA : Analysis of variance
ANR : Agence nationale de la recherche
APS : Activités physiques et sportives
ASC : Apprentissage sur corpus
AWL : Academic Word List
BAWE : British Academic Written English
BNC : British National Corpus
CEL : Cambridge English Lexicon
CAPES : Certificat d'aptitude au professorat de l'enseignement du second degré
CECRL : Cadre européen commun de référence pour les langues
CEVU : Conseil des études et de la vie universitaire
CLAWS : Constituent likelihood automatic word-tagging system
CLES : Certificat en langues de l'enseignement supérieur
CIA : Contrastive interlanguage analysis
CNU : Conseil national des universités
COBUILD : Collins Birmingham University International Language Database
COCA : Corpus of contemporary American English
DEUG : Diplôme d'études universitaires générales
EBCM : English booster course for Master's degree students
EIIDA : Étude interdisciplinaire et interlinguistique du discours académique
EMEA : European medicines agency
ENS : École normale supérieure
ESPE : École supérieure du professorat et de l'éducation
FASP : Fiction à substrat professionnel
GERAS : Groupe d'étude et de recherche en anglais de spécialité
GSL : General service list
HDR : Habilitation à diriger les recherches
ICAR : Laboratoire Interactions, corpus, apprentissages, représentations
IDEX : Initiatives d'excellence
ITDD : Ingénierie de la traçabilité et du développement durable
IMRD : Introduction, méthodes, résultats et discussion
IUFM : Institut universitaire de la formation des maîtres
JSCC : John Swales conference corpus
KIAP : Identité culturelle dans la prose académique
KWIC : Keyword in context
LAIRDIL : Laboratoire interuniversitaire de recherche en didactique des langues
LACES : Laboratoire culture, éducation, sociétés
LANSAD : Langues pour spécialistes d'autres disciplines
LEST : Loisirs, environnement, sport et tourisme

LIDILEM : Laboratoire de linguistique et didactique des langues étrangères et maternelles
LIG : Laboratoire d'informatique de Grenoble
LSG : Langue scientifique générale
LOCNESS : Louvain Corpus of Native English Essays
MCSE : Minimum Competence in Scientific English (manuel)
MI : Mutual information
MOOC : Massive open online course
MW : Merriam-Webster dictionary
MWLD : Merriam-Webster learner's dictionary
OADL : Oxford advanced American dictionary for learners of English
OED : Oxford English Dictionary
OOPT : Oxford online placement test
OSTI : Office for scientific and technical information
PMI : Pointwise mutual information
SAES : Société des anglicistes de l'enseignement supérieur
SIG : Special interest group
SKELL : Sketch engine for language learning
SUP : Service interuniversitaire de pédagogie
TAL : Traitement automatique des langues
TEI : Text encoding initiative
TICE : Technologies de l'information et de la communication de l'éducation
TOEFL : Test of English as a foreign language
TOEIC : Test of English for international communication
UCREL : University centre for computer corpus research on language
UJF : Université Joseph Fourier
UGA : Université Grenoble Alpes
UNR-RA : Université numérique en région Rhône-Alpes
XML : Extensible markup language
XAIRA : XML aware indexing and retrieval architecture

PRÉSENTATION DU DOSSIER

Ce dossier, élaboré en vue de l'obtention de l'habilitation à diriger des recherches, est composé de deux volumes.

Le volume I comprend une synthèse et un approfondissement des travaux de recherche effectués depuis ma soutenance de thèse (2009) autour de l'anglais de spécialité et de la didactique des langues. Le présent dossier incorpore un double système de référence bibliographique déjà adopté par de nombreux chercheurs. Un format [HART##a] renvoie à la bibliographie personnelle et un format (Auteur, année) renvoie à toutes les références externes. À la fin du Volume I se trouvent deux index, un index des travaux personnels et un index des références extérieures par auteur. Ensuite vous trouverez une liste des projets que j'ai pilotés, les corpus et outils cités, les bibliographies et enfin, un curriculum vitae détaillé.

Le Volume II comprend l'ensemble de mes travaux par type de publication. Chaque publication est précédée par la page de couverture de l'ouvrage et suivie par une table des matières, comme le conseille la SAES (Société des Anglicistes de l'Enseignement Supérieur). Un rappel bibliographique avec indications de références internes à ces deux volumes est proposé afin de faciliter la lecture.

LISTE DE FIGURES

Figure 1 : Modèle des interfaces entre les trois sphères : anglais de spécialité, didactique des langues et linguistique de corpus	33
Figure 2 : Décomposition de la collocation <i>freshly baked potatoes</i> : le collocatif est une collocation, Tutin, 2010	40
Figure 3 : Adaptation du "general drift" vers la stabilité de Halliday (2004 : 42)	56
Figure 4 : <i>Representations of a mediational structure</i> de Blin et Jalkanen (2014 : 157)	69
Figure 5 : Modèle d'ergonomie didactique du tutorat de langue en ligne de Bertin (2015), Bertin <i>et al.</i> (2010 : 27).....	71
Figure 6 : Modle de trois sphères du contexte d'apprentissage institutionnel	73
Figure 7 : Modèle d'un cours collectif médiatisé.....	75
Figure 8 : Saisie d'écran de la page d'accueil du dispositif de Wikispaces.....	77
Figure 9 : Image postée sur le Wikispaces.....	78
Figure 10 : Saisie d'écran du English Booster Course for Master's degree students	80
Figure 11 : Modèle d'un dispositif d'aide en ligne, l'EBCM	81
Figure 12 : Modèle du dispositif Corpus4All.....	83
Figure 13 : Saisie d'écran du projet UNR-RA Corpus4All en cours de construction.....	84
Figure 14 : Saisie d'écran de Scientext Enigma Plus	94
Figure 15 : Grille d'évaluation pour la prise de notes en Master2 en anglais de spécialité	104
Figure 16 : Saisie d'écran du dispositif numérique d'aide à la prise de notes	107
Figure 17 : Extrait de <i>Notable Science</i> : Carte de concepts	108
Figure 18 : Résultats d'une requête sur TXM : auxiliaires de modalité (JSCC)	120
Figure 19 : Saisie d'écran des résultats d'une requête de cooccurrences de <i>can/may/would</i> (JSCC)	122
Figure 20 : Saisie d'écran de résultats "KWIC" de Scientext.....	125
Figure 21 : Saisie d'écran d'une requête sur Scientext, mode libre	126
Figure 22 : Saisie d'écran des résultats statistiques d'une requête sur Scientext de [we + Verbe].....	127
Figure 23 : Saisie d'écran de Scientext Enigma [we + HYPOTHESIZE].....	131

LISTE DE TABLEAUX

Tableau 1 : Extrait des données : adjectifs/noms de Maniez (2012)	41
Tableau 2 : Extraits de phrases prototypiques (Swales, 1971) et (Swales & Feak, 2004)	47
Tableau 3 : Verbes de causalité (action ou évènement) dans les résumés du corpus Scientext	50
Tableau 4 : Extrait de Scientext [a <i>explains</i> b].....	55
Tableau 5 : Taux de couverture des 4 sous-corpus par le GSL et l'AWL en pourcentages (Coxhead, 2000 : 224)	86
Tableau 6 : Comparaison de ressources disponibles sur quatre sites de dictionnaires.	89
Tableau 7 : Assiduité des tâches de l'EBCM	98
Tableau 8 : Scores d'un test normalisé (OOPT) et la note en prise de notes.....	106
Tableau 9 : Calculs statistiques de trois modaux et prénoms du JSCC.....	124
Tableau 10 : Fréquence de chaque adverbe et caractéristiques des verbes associés.	133
Tableau 11 : Verbes les plus fréquents selon l'adverbe et leur conjugaison	134
Tableau 12 : Les huit verbes en <i>-ing</i> les plus fréquents après <i>thus</i> et <i>thereby</i>	137
Tableau 13 : Nombre d'occurrences par catégorie grammaticale et par contexte	141
Tableau 14 : Dix mots les plus fréquents à l'oral et à l'écrit du corpus EIIDA, anglais scientifique.....	143
Tableau 15 : <i>So</i> dans le sous-corpus EIIDA d'anglais scientifique	145

PARCOURS DE RECHERCHE

Ce dossier de synthèse a pour objet d'étude l'anglais de spécialité avec une forte orientation vers l'anglais des sciences sous l'angle de deux approches épistémologiques, la linguistique de corpus et parallèlement la didactique des langues. La problématique qui sous-tend ce dossier de synthèse est d'identifier les entrées pertinentes à l'enseignement de l'anglais de spécialité à partir de conclusions tirées d'analyses linguistiques effectuées à partir de données authentiques issues de la linguistique de corpus. L'ensemble de mes travaux de recherche traite de cette problématique, en parallèle avec quatorze ans d'enseignement, dont plus de dix ans d'enseignement d'anglais auprès d'étudiant.e.s en sciences et de projets didactiques conçus à partir de bases théoriques de la didactique des langues.

Mon inscription en HDR (Habilitation à Diriger les Recherches) est le fruit d'un cheminement relativement homogène fait de rencontres nationales et internationales, autour de la recherche et de l'enseignement en anglistique et en didactique des langues. Depuis 2003, j'assure des travaux dirigés d'anglais de spécialité auprès d'un public LANSAD (Langues pour Spécialistes d'Autres Disciplines) dans des parcours scientifiques de l'UJF (Université Joseph Fourier). Les interrogations formulées en début de ce parcours ont été traitées dans ma thèse *L'image dans l'autoformation et l'évaluation de la compréhension de l'oral dans le domaine de l'anglais de spécialité* (2009). Suite à ma qualification aux sections 7 et 11 du CNU (Conseil national des universités), j'ai intégré comme MCF le laboratoire LIDILEM à l'Université de Stendhal, Grenoble 3 en 2011, tout en continuant mon enseignement auprès des scientifiques à l'UJF. Au 1 janvier 2016, ces deux universités et l'Université Pierre Mendès France fusionnent au sein de l'Université Grenoble Alpes (UGA).

Alors que je préparais un *Bachelor of Arts en Littérature et Culture Française* à l'Université de San Diego (États-Unis d'Amérique), j'ai assuré les travaux pratiques en

français langue étrangère auprès d'autres étudiant.e.s dans le même établissement. Cette première confrontation avec les difficultés d'apprentissage me fit ressentir la nécessité de placer l'humain au centre de l'enseignement. Ma première publication internationale avec comité de sélection fut tirée de mon mémoire de licence sur la représentation de l'impact des études sur la vie des femmes dans une sélection de la littérature francophone [HART03a].

Suite à ces expériences, j'ai éprouvé le besoin d'accroître mes bases théoriques avant de commencer une carrière dans l'enseignement. J'ai donc entrepris un *Master of Education (Program of Teaching and Learning: Specialization in Literacy)* à l'Université de San Diego où j'ai abordé les grandes théories de l'acquisition des langues étrangères. Pendant cette période, je fus aussi assistante dans un collège « d'immersion » anglais/français. Sélectionnée comme assistante de recherche à l'université (*Graduate Student Fellowship*), j'ai participé à la mise en œuvre d'une enquête autour de la poésie en classe primaire. Mon mémoire de master traitait de l'évaluation en langues étrangères dans l'enseignement supérieur.

Dès l'obtention du CAPES (Certificat d'aptitude au professorat de l'enseignement du second degré) en 2005, j'ai demandé à Nicole Décuré, Directrice du LAIRDIL (Laboratoire Interuniversitaire de Recherche en Didactique des Langues) à l'Université de Toulouse de bien vouloir encadrer mes recherches de thèse au sujet de l'impact de la présence de l'image lors de l'écoute de documents audiovisuels, un domaine particulièrement riche en supports visuels. Encore stagiaire au Service des Langues à l'UJF, j'ai pris en main la coordination de deux projets en TICE (Technologies de l'information et de la communication de l'éducation) : le renouvellement d'une plateforme de reconstruction de texte conçu par Jonathan Upjohn dans les années 1980 et la conception et mise en place d'une aide numérique à la prise de notes. Ces deux projets continuent à être la cible d'une recherche-action de ma part, afin que la recherche irrigue les projets pédagogiques et les objectifs institutionnels autour de l'anglais de spécialité, dont la validation du niveau « indépendant » (B2) en programmes de master dans le cadre des directives européennes (Fries-Verdeil, 2009) [HART11c, HART13a] (cf. 2.3.1).

Lors de mon parcours, j'ai assumé des responsabilités importantes, dont la direction du Service des Langues à l'UJF (une cinquantaine d'enseignant.e.s assurant plus de 9000 heures de travaux dirigés par an), le pilotage de projets TICE, la préparation aux certificats de langue et enfin, la conception de grilles d'évaluation. Suite à ma qualification aux sections 7 et 11, j'ai intégré, comme MCF, le laboratoire LIDILEM à l'Université de Stendhal, Grenoble 3 en 2011.

Au sein de l'axe 1 (description linguistique, Corpus et TAL) du LIDILEM, j'ai trouvé un cadre extrêmement favorable à la collaboration avec des collègues en linguistique et notamment en linguistique de corpus. Au plan de la recherche, mon intégration au laboratoire LIDILEM représente à la fois une ouverture et un défi. De nouveaux projets avec des collègues passionné.e.s par la linguistique, une lecture intensive et un très grand nombre de requêtes sur corpus m'ont permis de relever ce défi. En ce qui concerne l'animation de la recherche, j'encadre en cotutelle une thèse en didactique de l'anglais, je réponds à des appels de financement et je participe à des revues scientifiques et à l'organisation d'animation de la recherche. Ces expériences m'ont amenée à éprouver le besoin de faire un inventaire approfondi des objectifs, des interrogations, des fils conducteurs et des acquis de mon parcours afin de montrer mes capacités à diriger des travaux de recherche.

En ce qui concerne les projets financés, j'ai bénéficié en 2011-2012, avec Marie-Paule Jacques, d'un financement institutionnel sur le projet « Structure discursive de l'article de recherche » dont la thématique de la voix auctoriale dans l'article de recherche a donné lieu à des publications [HART12a, HART13c, HART14b] et a irrigué les projets TICE, dont *Scientext Enigma* (cf. 2.2.3-4). Ceci m'a donné une grande liberté de recherche, propice à l'approfondissement des techniques et des méthodologies liées à la linguistique de corpus. L'établissement de nouveaux échanges scientifiques m'a été extrêmement bénéfique. J'ai participé à l'organisation de rencontres intitulées « La phraséologie : des collocations aux séquences figées » avec l'axe 1 du LIDILEM. L'évolution du corpus *Scientext* est un projet du laboratoire LIDILEM auquel je participe depuis mon intégration au laboratoire en 2011.

Enfin, dans le cadre du co-encadrement (50%) de la thèse de Sayena Molaie avec Françoise Raby : « Implementing Gardner's Theory of Multiple Intelligences in the Teaching of English as a Second Language » et depuis 2015 l'encadrement de la thèse (100%) d'Irina Mukhlynina : « Influence de la présentation multimodale d'un contenu textuel écrit autour d'une vidéo sur la compréhension de l'oral du texte audio-visuel en langue étrangère (français, anglais, russe) dans le contexte d'apprentissage non formel », je soutiens la recherche de jeunes chercheuses.

Conserver des échanges scientifiques avec des chercheurs anglophones est essentiel pour une prise de recul disciplinaire. Dans ce sens, des collaborations diverses à l'Université de San Diego et la participation aux vidéoconférences autour de la didactique des langues organisées par le LAIRDIL (2011, 2012, 2013), l'organisation avec Cristelle Cavalla de la journée d'étude « Valoriser et développer les outils autour des corpus dans une perspective didactique » sur le Campus Sud-Rhône-Alpes à Valence (UGA) depuis 2014 me permettent de approfondir ma recherche en didactique de l'anglais langue étrangère. De nouvelles rencontres au sein de l'association européenne d'EuroCALL, en télécommunication avec Bin Zou (Xi'an Jiaotong-Liverpool University) (*cf.* 2.1.3.4.) et en linguistique de corpus, dont Marie-Paule Jacques et Fanny Rinck (LIDILEM) et Emmanuelle Esperança-Rodier du LIG (Laboratoire d'informatique de Grenoble) au projet IDEX Étude interdisciplinaire et interlinguistique du discours académique (EIIDA), me permettent une belle ouverture théorique, méthodologique et interdisciplinaire (*cf.* 3.1.1.3).

ORIENTATIONS DE LA NOTE DE SYNTHÈSE

Mes préoccupations, liées aux besoins linguistiques en anglais de spécialité des apprenant.e.s et l'analyse linguistique du discours des expert.e.s, m'ont conduite à suivre trois pistes entrecroisées allant du théorique au plus pratique :

Première partie : une réflexion sur les diverses approches théoriques et méthodologiques de l'analyse linguistique du discours de spécialité, notamment à partir de corpus ;

Deuxième partie : une étude des pratiques, des compétences langagières des apprenant.e.s, ainsi que des ressources didactiques ;

Troisième partie : une analyse linguistique du discours des expert.e.s¹ en anglais de spécialité.

Au fil de ce dossier de synthèse, plusieurs problématiques liées à ces trois thèmes seront abordées. Au lieu d'une organisation chronologique, j'adopte ici une construction thématique, car mon itinéraire de recherche se caractérise par une forte harmonie thématique. Pour faciliter la lecture, le dossier sera partagé en trois parties, allant du théorique, aux problématiques plus appliquées, actuelles et à venir.

Ma recherche s'appuie sur les interrogations relevées lors de la réflexion sur ma propre expérience d'enseignement, mes questionnements autour de la langue générale, de la langue de spécialité ainsi que dans le cadre de la direction d'une équipe d'enseignants ou de projets spécifiques. Dans cette synthèse, je m'efforce de transcender les considérations spécifiques afin d'aborder plus largement les questions théoriques et épistémologiques autour de l'anglais de spécialité.

UN FIL DIRECTEUR

L'exploration des théories et des méthodologies permettant d'étayer l'acquisition des pratiques langagières adaptées aux besoins de l'apprenant.e est le fil directeur de mes travaux. J'ai d'abord enseigné l'anglais dans des collèges et des lycées aux États-Unis et en France, mais j'ai trouvé ma voie dans l'enseignement supérieur.

¹ Dans ce dossier, le terme *apprenant.e* réfère à toute personne inscrite dans un cadre institutionnel d'enseignement. Par *expert.e*, je réfère, sauf précision, à des spécialistes de leur discipline et s'exprimant en anglais, que la personne soit native (*native speaker*) ou non native (*non native speaker*).

Mes vingt-six publications depuis mon premier poste d'assistante de recherche sont évoquées tout au long de ce dossier de synthèse. Au fil des pages, je défends l'idée que dans le système éducatif français, nous ne prêtons pas assez attention à l'anglais sur objectifs spécifiques, notamment en anglais des sciences. Cela se reflète dans la difficulté systématique à recruter des enseignant.e.s et des enseignant.e.s-chercheurs ayant les bases solides en linguistique et en didactique pour enseigner auprès d'un public LANSAD ou de mener des recherches dans ce domaine. Cette lacune est soulignée dans le Bulletin du SAES en mars 2007 : « Les Écoles Doctorales devraient aider les futurs candidats à postuler à des emplois dans l'Enseignement supérieur, en proposant une ouverture à d'autres formations (de type Langue de spécialité) » (2007 : 39).

À mon sens, les lacunes autour de l'anglais de spécialité se creusent dès la licence. Les programmes en LEA (Langues étrangères appliquées), où nous trouvons la plupart des cours autour des langues de spécialité, ne sont pas orientés vers la recherche, comme cela est le cas pour les programmes en LLCE (Langues, Littératures et Civilisations étrangères) où les étudiant.e.s préparent les concours de l'agrégation. Pour ces concours, nous trouvons des thématiques de la littérature (par exemple, l'œuvre de *Shakespeare*) et de la civilisation (*i.e.* la Grande Famine en Irlande, les années Roosevelt aux États-Unis) et même de la linguistique : la négation (2015) et la référence à l'avenir (2013). De ce fait, les programmes de master sont peu compatibles avec un intérêt et une préparation théorique et méthodologique aux langues de spécialité. Pourtant, il s'agit d'un domaine à forte demande dans l'enseignement supérieur : 14-26% de recrutement de MCF et 6-15% de tous les recrutements de postes de second degré sont sur des profils d'anglais de spécialité avec une affectation dans l'enseignement supérieur (Braud, Millot, Sarré & Wozniak, 2015). Seule une recherche autour du discours spécialisé, autour des moyens d'analyse linguistique sur corpus et autour des considérations didactiques, nous permettra de répondre efficacement à la demande croissante de formation en langues de spécialité dans l'enseignement supérieur.

1. PREMIÈRE PARTIE : APPROCHES THÉORIQUES, REPÈRES HISTORIQUES ET ÉPISTÉMOLOGIQUES

Dans cette première partie, j'apporte tout d'abord des précisions sur trois notions essentielles dans mes travaux de recherche autour du discours de spécialité : l'anglais de spécialité, les corpus, et la didactique des langues. J'esquisserai ensuite les grands repères historiques et épistémologiques des travaux internationaux en linguistique de corpus et en anglais de spécialité qui ont influencé ma recherche. Pour finir, je présenterai trois phénomènes linguistiques qui sont centraux en anglais des sciences : la causalité, la métaphore grammaticale et enfin, le positionnement des chercheurs par rapport aux pairs.

1.1 TROIS SPHÈRES

1.1.1. L'ANGLAIS DE SPÉCIALITÉ

Ce dossier de synthèse s'inscrit dans le champ de l'anglais de spécialité, particulièrement l'anglais des sciences, afin « de contribuer au développement et à l'amélioration de l'enseignement de l'anglais de spécialité » (GERAS, 2007), but mis en avant par le GERAS (Groupe d'Étude et de Recherche en Anglais de Spécialité), société savante rattachée à la SAES (Société des Anglicistes de l'Enseignement Supérieur). L'anglais des sciences, tout comme l'anglais du droit ou l'anglais du marketing font partie de l'anglais de spécialité pratiqué par une communauté de discours spécifique. J'adopte, comme point de départ la définition de Perrin (1995) selon laquelle l'anglais de spécialité peut être défini comme étant :

[...] l'utilisation, à des fins bien particulières, dans un contexte bien délimité, le plus souvent correspondant à une pratique professionnelle, des constituants habituels de toute langue : lexicale, certes, avec évidemment, concentration de termes relevant du domaine, et syntaxe, avec sans doute prédominance de certains tours récurrents (Perrin, 1995 : 15).

Le concept de contexte est central et nous le retrouvons dans la définition de Petit (2010) d'un domaine spécialisé :

tout secteur de la société constitué autour et en vue de l'exercice d'une activité principale qui, par sa nature, sa finalité et ses modalités particulières ainsi que par les compétences particulières qu'elle met en jeu chez ses acteurs, définit la place reconnaissable de ce secteur au sein de la société et d'un ensemble de ses autres secteurs et détermine sa composition et son organisation spécifiques (Petit, 2010 : sans pagination).

Deyrich (2004) introduit la notion dynamique de « paysage professionnel » prenant ainsi en compte les qualités culturelles et subjectives de nos regards sur les contextes.

L'intérêt pour l'anglais de spécialité apparaît en France dans les années 1970, en partie à cause des besoins importants dans l'enseignement supérieur. L'identification de l'anglais de spécialité, selon Bertin (2009 : 3), nécessite de « définir un champ d'étude et de recherche susceptible de fonder un enseignement universitaire de qualité », par l'identification d'un objet langagier, les méthodes de recherches adaptées afin de mieux l'enseigner (Bertin, 2009 : 3). Dans cette optique, je considère ces trois aspects (objet de recherche, méthodes d'analyse adaptées et conclusions relatives à l'enseignement) comme étant les trois versants de la recherche en anglais de spécialité.

Il me semble néanmoins que la perspective un peu différente proposée par la notion d'*English for Academic Purposes* (EAP) permet d'incorporer des éléments complémentaires. Le journal du même nom, l'EAP le décrit ainsi :

A wide range of linguistic, applied linguistic and educational topics may be treated from the perspective of English for academic purposes; these include: classroom language, teaching methodology, teacher education, assessment of language, needs analysis, materials development and evaluation, discourse analysis, acquisition studies in EAP contexts, research writing and speaking at all academic levels, the sociopolitics of English in academic uses and language planning (Journal of English for Academic Purposes (site web), 2015).

Dans cette perspective, l'objet d'étude inclut des sujets tels que le discours des contextes d'apprentissages et d'évaluation, l'expression écrite et orale et à tous ses niveaux de compétence, la recherche ou bien la planification linguistique. Je pense que ces aspects ont toute leur place afin d'offrir une vision complète de la langue, de son usage et de son enseignement. Au sein de la conception française de l'anglais de spécialité, il existe aussi des domaines spécialisés, comme par exemple l'anglais du tourisme ou l'anglais culinaire, qui sont moins associés à l'anglais académique souligné par la notion de l'EAP.

L'anglais des sciences exactes, souvent le sujet de mes recherches, possède aussi des particularités, qui ne sont pas forcément partagées par d'autres types d'anglais de spécialité. L'analyse du discours des sciences révèle des similitudes au plan organisationnel et discursif. Dans l'article de recherche en sciences exactes, le format IMRD (Introduction, Méthodes, Résultats et Discussion) impose des contraintes de rédaction, qui restent très codées. Ces similitudes discursives répétées facilitent la compréhension en permettant d'aborder plus facilement des notions scientifiques complexes.

En anglais académique, nous trouvons aussi au niveau de la phrase, des expressions polylexicales récurrentes, à fonction référentielle (*point of view, take into account*), discursive (*we will begin by, after reviewing*) ou interpersonnelle (*it is necessary to, we believe*) ou des routines à schéma abstrait (*we have adapted here the model of X ; on this point, we disagree with X*). Ces expressions récurrentes guident la compréhension et leur familiarité permet de soulager l'attention cognitive en laissant place au contenu scientifique du texte.

Dans ma propre recherche linguistique, j'ai d'abord analysé les verbes, qui évoquent souvent des actes d'observation et de manipulation. Sans surprise, les verbes de causalité d'action ou d'évènement sont les plus récurrents parmi les cinquante verbes les plus fréquents dans les résumés d'article du corpus Scientext [HART13b]. Une autre nécessité discursive des scientifiques est la conceptualisation des processus, pour laquelle la métaphore grammaticale est particulièrement puissante. Enfin, la prise de

position à l'écrit est sujette à la modalité phrastique ou la « précaution oratoire » (*hedging*), souvent sous forme de verbes ou d'adverbes de modalité, relatifs aux autres membres de la communauté scientifique.

L'anglais scientifique se divise en plusieurs genres écrits et oraux, dont l'article ou la note de recherche, l'étude de cas, la présentation orale à une conférence ou au sein du laboratoire [HART13d, HART11d]. Une *hybridité discursive* se crée entre discours de spécialiste et discours ordinaire, qui se manifeste lors des points de contact entre experts, professionnels et citoyens « ordinaires » [HART14c]. Il existe ainsi des genres parallèles, comme le courriel professionnel ou la vulgarisation scientifique, dont l'étude dépasse les confins du présent dossier. L'anglais scientifique reste un domaine à multiples facettes allant d'une analyse purement linguistique d'un texte donné à la planification de programmes d'études basée sur les besoins professionnels d'une communauté précise.

J'adopte la position selon laquelle l'approche sur corpus permet de dégager des éléments linguistiques authentiques, sans pour autant constituer une théorie. Un corpus offre « un réservoir d'exemples, une manifestation authentique des possibilités de la langue ou d'un discours, mais n'est pas en tant que tel 'la langue' » (Tutin, 2010). La linguistique de corpus permet d'évaluer, au-delà du simple relevé de lexique ou de motifs lexico-syntaxiques par exemple, la qualité interpersonnelle et dialogique du positionnement en discours scientifique (Fløttum, Dahl & Kinn, 2006a ; Grossmann & Tutin, 2014).

1.1.2. UNE APPROCHÉ FONDÉE SUR LA LINGUISTIQUE DE CORPUS

Ce dossier de synthèse prend appui sur une définition qui restreint la notion de *corpus* à « une collection conséquente de textes authentiques dans un format électronique, conçue pour être représentative d'une variété de langage » (Boulton, 2012 : 272). Effectuer une requête sur un ensemble de documents en ligne via un moteur de recherche comme Google, bien que nous y retrouvions des textes authentiques, ne s'inscrit pas dans le domaine de la linguistique de corpus, car l'ensemble des documents n'est pas conçu pour représenter ou comprendre une variété de langue. Dans mes

études, je travaille essentiellement à partir du corpus « Scientext » (cf. 3.1.3.), qui comporte une très grande quantité de textes du genre « article de recherche ». Mon approche par l'analyse du discours s'inscrit dans la tradition fondée par Firth (cf. 1.2.1.), Swales (cf. 1.2.3.) et Hyland (cf. 1.2.4.), complétée par les travaux socio-fonctionnels de Halliday (cf. 1.3.2.).

La définition des paramètres du corpus est d'une importance capitale, par la nature des textes, leur traitement et les capacités des outils d'exploitation. Une des premières considérations lors de la constitution d'un corpus est l'objectif d'exploration. Un corpus de grande taille, comme le corpus de quelque 7 500 articles de recherche en anglais de Scientext (35 244 378 mots), permet d'identifier les expressions polylexicales récurrentes à fonction référentielle, discursive ou interpersonnelle d'un domaine aussi vaste que l'anglais écrit des sciences expérimentales. Le volume important d'unités lexicales de Scientext permet de faire la différence entre un élément isolé, mais présent par hasard et des constructions peu fréquentes, mais particulières à un domaine, par exemple autour de la causalité, phénomène essentiel en anglais des sciences. En revanche, un corpus plus petit, comme celui d'environ 180 textes, créé dans le cadre du projet EIIIDA (cf. 3.1.1.3.), permet d'examiner les différences pertinentes entre l'anglais, l'espagnol et le français dans deux disciplines différentes (sciences exactes et linguistique) et/ou entre l'oralité et la rédaction écrite.

Un corpus peut être constitué, par exemple, d'articles scientifiques à la fois actuels et plus anciens, dans la perspective d'une étude diachronique (cf. Banks 2010), ou de transcriptions de présentations orales en français et en anglais, afin d'étudier les différences translinguistiques dans le même genre. Une étude peut aussi être transgenre lorsqu'elle traite, par exemple, des transcriptions de présentations orales et des articles de recherche. Elle peut être à visée socio-linguistique, par exemple, lorsque l'analyse porte sur le type des thèmes abordés ou les différences d'expression par genre femme/homme.

Les études de corpus permettent de cerner des collocations ou expressions polylexicales d'intérêt linguistique et d'identifier des lexèmes ou motifs d'intérêt didactique, qui sont

particulièrement fréquents ou mal maîtrisés par les apprenant.e.s, par exemple *the purpose of the current study* ou *it is presently unknown*, pour lesquelles le terme français *actuel* peut être source d'interférence. Grâce à des corpus multidisciplinaires, nous pouvons identifier des expressions académiques peu fréquentes dans un contexte général, telles que *play a role*, *this study suggests that*, *this holds true*. Seul un corpus constitué de textes issus de la communauté de discours en question permet de cibler ses particularités, dont le lexique, la syntaxe, des patrons et, plus largement, les aspects discursifs dont l'organisation interne ou le positionnement des auteurs.

1.1.3. L'APPROCHE SOCIO-CONSTRUCTIVISTE DE LA DIDACTIQUE

L'approche socio-constructiviste de la didactique est une troisième dominante de ma recherche. J'adopte, depuis mes études en Master of Education, un cadre théorique pour la didactique basé sur une vision socio-constructiviste de l'apprentissage, sans pour autant nier l'importance des aspects cognitifs, motivationnels et affectifs. Selon cette perspective socio-constructiviste, l'apprentissage a lieu lors d'interactions sociales, en fonction des objets (outils ou signes) disponibles (Erickson 1984 ; Erickson 1996 ; Vygotski, 1978). L'apprentissage a lieu lorsque l'interaction sociale et les outils disponibles permettent à l'apprenant.e d'accéder à un niveau de développement supérieur à ses propres capacités. Vygotski nomme cet espace de possible apprentissage « la zone proximale de développement » qui est :

[...] la distance entre le niveau de développement actuel, tel qu'on peut le déterminer à travers la façon dont l'enfant résout les problèmes seul et le niveau de développement potentiel, tel qu'on peut le déterminer à travers la façon dont l'enfant résout les problèmes lorsqu'il est assisté d'un adulte ou collabore avec d'autres enfants plus avancés (Vygotski, 1976).

Dans le cas de l'apprentissage d'une langue étrangère, Krashen (1982 ; 1983) reste emblématique pour avoir introduit le concept de *Natural Approach*, dont l'objectif fut de transposer le développement de la langue maternelle à la classe de langue seconde. Il mettait en avant la notion de *comprehensible input* (ensemble des données pouvant être comprises), ce qui recentrait l'attention sur l'importance du sens des énoncés. L'importance du sens avait été précédemment noyée dans l'approche comportementaliste (stimulus / réponse) de Skinner (1954, 1958). Je souscris à la

théorie selon laquelle l'apprenant.e a besoin de contacts répétés avec des éléments compréhensibles de la langue d'un niveau légèrement supérieur à son niveau actuel ($i = 1$) et qui correspondent à ses objectifs communicationnels afin de pouvoir évoluer dans un contexte social cible. Les éléments compréhensibles peuvent s'étendre d'une unité lexicale à un livre entier, mais doivent être choisis en fonction de l'objectif communicationnel. Nous pouvons critiquer chez Krashen l'idée qu'une langue seconde s'apprend d'une manière similaire à la langue maternelle, notamment l'importance des transferts positifs et des interférences depuis la langue maternelle ainsi que des stratégies. De plus, il est difficile d'établir véritablement le niveau i (l'interlangue) d'un.e apprenant.e, ainsi que le niveau $i + 1$ (e.g. l'input comportant des connaissances ou des structures que l'apprenant.e doit être prêt.e à acquérir en fonction de l'interlangue), mais les travaux sur corpus d'apprenant.e.s peuvent apporter des réponses.

Il existe de nombreux travaux à partir de corpus constitués de textes produits par des apprenant.e.s natifs et natives dans un cadre académique, dont le corpus Scientext d'écrits universitaires en anglais langue étrangère, comportant 1 020 146 mots, le BAWE (British Academic Written English) comportant 6 506 995 mots et le LOCNESS (Louvain Corpus of Native English Essays) comportant 324 304 mots. Ce type de corpus permet de mettre en exergue, par exemple, l'interférence de la langue maternelle sur la langue cible (Boulton & Tyne, 2014 ; Osborne, 2008 ; Osborne, 2011). Nous trouvons aussi en CIA (*Contrastive interlanguage analysis*) des comparaisons entre la langue cible exprimée par des non natifs, particulièrement des populations d'origine linguistique différente (cf. Granger, 1996 ; Paquot, 2008). À l'heure actuelle, la notion de « reference language varieties » permet de prendre en compte des variables telles que les tâches, les dialectes et les objectifs linguistiques (Granger, 2015).

1.1.4. TROIS SPHÈRES INDISSOCIABLES

Je propose une modélisation de ces trois sphères de ma recherche : anglais de spécialité, linguistique de corpus et didactique des langues (cf. Figure 1). La sphère supérieure du modèle représente l'anglais de spécialité, en d'autres termes, l'anglais pratiqué par une communauté discursive ayant des objectifs, des théories, des méthodologies et des

productions en commun (cf. 1.2.3.6.) dans un paysage professionnel donné (Deyrich, 2004). Ces productions discursives peuvent être divisées par genre. Je reprends la notion de Bakhtine (1986/2001) selon laquelle le *genre* réfère à des productions de communication issues de situations typiques et autour de thèmes typiques. Selon les circonstances, ces situations et ces thèmes impactent les relations entre les mots et le sens créé :

Genres correspond to typical situations of speech communication, typical themes, and consequently, almost to particular contracts between the *meanings* of words and actual concrete reality under certain typical circumstances (Bakhtine, 1986/2001 : 129).

Au sein de chaque genre, nous trouvons une plus forte présence de certains énoncés selon le genre, genre qui évolue au sein d'une communauté de discours. L'imbrication de ces trois niveaux (énoncé, genre, communauté) est modélisée dans la sphère supérieure de la Figure 1.

En linguistique de corpus, les collocations (cf. 1.2) ont souvent servi de point de départ lors de l'exploration de textes, comme schématisé dans la sphère en bas à droite de la Figure 1. La linguistique de corpus sur textes spécialisés en anglais est représentée ici par l'aire « A ».

La didactique des langues est représentée par la sphère en bas à gauche de la Figure 1. Historiquement, l'acquisition de champs lexicaux est un des fondements de l'enseignement de l'anglais de spécialité (cf. 2.2.1.1.). La répétition de ces champs lexicaux s'opère lors de micro- et macro-tâches et a pour objectif de permettre à l'apprenant.e de mettre en pratique les connaissances et savoirs linguistiques nécessaires à la communication dans un contexte précis. Les études à partir de corpus d'apprenant.e.s sont un point de rencontre entre la didactique et la linguistique (cf. l'intersection d'union B). Lorsque les connaissances et les savoir-faire en sciences et en langues d'un.e étudiant.e sont assez avancés, il ou elle peut rentrer dans la communauté de discours de sa discipline (cf. l'aire « C » de la Figure 1).

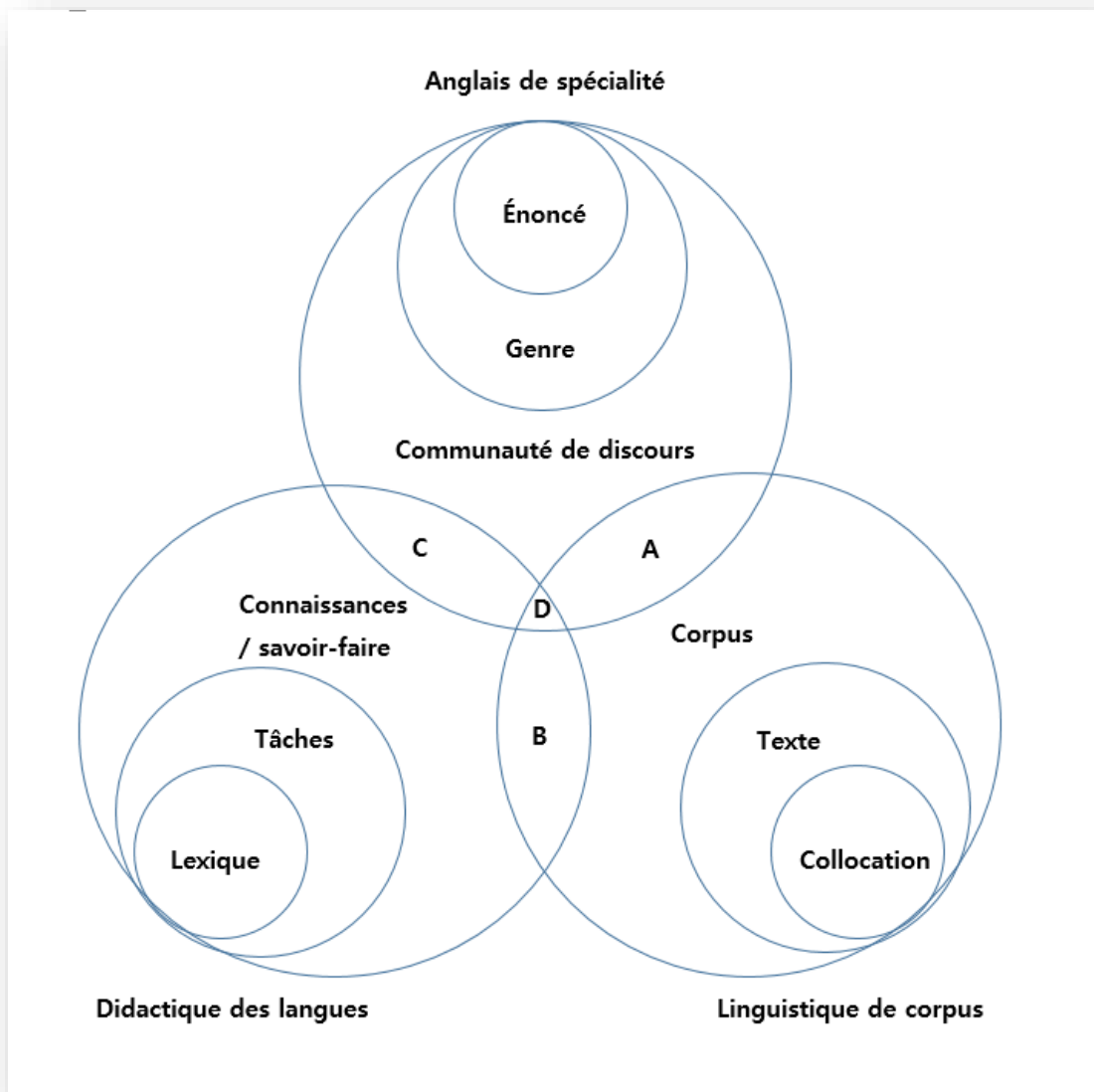


Figure 1 : Modèle des interfaces entre les trois sphères : anglais de spécialité, didactique des langues et linguistique de corpus

Dans ma pratique, les trois sphères sont indissociables, car c'est à travers les analyses de la linguistique de corpus que je bâtis ma réflexion théorique et méthodologique en anglais de spécialité, ceci afin d'étayer l'intégration de l'étudiant.e dans sa communauté disciplinaire anglophone. Je tente d'approfondir les connaissances en anglais de spécialité à travers le filtre de la linguistique de corpus dans un objectif d'applications en didactique, comme représenté par l'aire « D ».

1.2. LES DÉBUTS ET LES NOTIONS CLÉS DE LA LINGUISTIQUE DE CORPUS

Il me semble essentiel de replacer la linguistique sur corpus dans son cadre historique et théorique, afin de me positionner dans cette tradition relativement récente.

En 1933, H. E. Palmer publia un rapport détaillé, *l'Interim Report on English Collocations* qui posa des bases théoriques autour de la notion de la collocation, qu'il décrit comme étant « *a succession of two or more words that must be learnt as an integral whole and not pieced together from its component parts* » (Palmer, 1933 : page de couverture). Cela remet directement en cause le concept pédagogique de listes de mots les plus fréquents, car les collocations ont un effet sémantique. Par exemple, lorsque le mot isolé *crack* signifie en général une *faille*, *to crack a joke* signifie *raconter une blague*. Comme beaucoup de lexiques académiques sont basées sur la fréquence, les mots polysémiques risquent d'être surreprésentés (*e.g. set, check*). La proposition de Palmer déplace l'attention depuis les unités individuelles vers des couples d'unités en collocation (*cf. Williams et Million (2012 : 4)*).

1.2.1. RÉPÈRES HISTORIQUES

Dans *Modes of Meaning* (1951/1957), Firth propose une approche innovante de la linguistique descriptive qui prend en compte plusieurs niveaux de création de sens, y compris le contexte social, la syntaxe, le vocabulaire, la phonologie et la phonétique. Surtout, il met en évidence le caractère performatif et contextualisé du sens : le sens est créé par la « collocation » d'un mot qui apporte une abstraction au niveau syntagmatique. Il soutient que la collocation d'un mot fait partie de son sens et que le sens généré dépendra de la forme littéraire ou du genre. En d'autres termes, le mot *crack* (*cf. 1.3.2.*) est polysémique, mais il n'y a guère de confusion concernant son sens lorsqu'il apparaît dans un texte au sujet du pétrole ou de la dépression nerveuse.

La notion de collocation est associée à la dimension syntagmatique : « *Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words* » (1951/1957 : 196). Le système de collocation a pour effet de modifier le sens des mots qui sont en collocation.

Cette opération de création de sens ne dépend pas de la syntaxe. Nous pouvons reprendre l'exemple de *crack* et *crack a joke*. Une approche qui se limiterait aux sens primaires des unités lexicales isolées ne serait pas complète, car le sens créé par collocation serait occulté. Cette approche de l'importance du contexte est aussi reflétée dans sa citation connue « you shall know a word by the company it keeps » (1951/1957 : 11).

Firth évoque aussi, de manière un peu ambiguë, la colligation. Avec ce terme, il fait référence à des relations grammaticales entre des catégories de discours, puisque les relations grammaticales portent sur des catégories, et non des unités lexicales. La construction de base de phrase étant : [Sujet + Verbe], voire [Sujet + Verbe + Complément]. Grâce aux avancées technologiques, les linguistes de corpus ont pu travailler sur des corpus de plus en plus grands, mais en gardant des approches parfois contrastées sur l'importance de la syntaxe dans la création de sens et dans la description de la langue. Je défends également la position que nous comprenons un mot en fonction du contexte et des autres mots qui l'entourent.

1.2.1.1. L'APPORT DU SURVEY OF ENGLISH USAGE

Dans les années 1950, le *Survey of English Usage* de Randolph Quirk est le premier projet de collection de textes assemblés ayant pour objectif une analyse empirique de la langue anglaise. Il comporte 200 textes d'environ 5 000 mots chacun, d'anglais britannique écrit et oral. Combiné avec le Lancaster-Oslo/Bergen corpus et le Brown University Corpus, il a notamment donné lieu au très complet *Comprehensive Grammar of English Language* (Quirk, Greenbaum, Leech & Svartvik, 1985) riche de quelques 1 792 pages de grammaire descriptive de la langue anglaise à destination des linguistes. Un tel ouvrage a mis en évidence l'intérêt de cette approche et des méthodologies de la linguistique de corpus, mais l'intégration des applications directes didactiques reste problématique à cause de la quantité très importante de données et de la complexité. Les applications indirectes, à travers les dictionnaires (cf. 2.2.1.4.), les ressources (cf. 2.2.1.5.) et les approches didactiques (cf. 1.2.2.4.) sont en revanche d'un grand intérêt.

1.2.1.2 L'APPORT DE L'OSTI REPORT

Dans les années 1970, Daley, Jones et Sinclair (2004)² amorcent des recherches empiriques d'envergure à partir de corpus spécialisé, publiées trois décennies plus tard dans l'ouvrage *English collocation studies: The OSTI report*. Le rapport empirique de l'Office for Scientific and Technical Information (Osti) releva des notions clés, dont celles de la terminologie, le registre, les collocations et leurs « patrons linéaires », des éléments lexicaux et des méthodologies statistiques. Grâce à cette recherche, des chercheurs comme Michael Halliday et John Sinclair ont pu bénéficier des progrès technologiques et informatiques, afin de vérifier leurs propres hypothèses concernant la langue. Ce travail phare a lancé à la fois les fondements méthodologiques en linguistique de corpus, mais aussi le débat fascinant autour de l'importance des éléments lexicaux et syntaxiques des énoncés.

Ils ont souligné la notion de distance moyenne d'écart (*span*), en d'autres termes, l'absence ou la quantité de mots entre le pivot (le mot principalement concerné) et ses éventuelles collocations dans le contexte gauche ou le contexte droit du pivot (Sinclair, Jones & Daley, 2004). Ils ont montré que la fréquence de certains mots dépend du type de texte, par exemple, *the* est le mot le plus fréquent dans leurs deux corpus de transcriptions de l'oral et de textes scientifiques, mais le pronom personnel *I* est classé deuxième dans leur corpus de transcriptions, mais classé 241^{ie} dans leur corpus de textes scientifiques. Il y a donc des différences selon le type de document (*ibid* : 58). Ceci ouvre la voie théorique à des études sur les genres (*cf.* 1.2.3.). Surtout, il y a par leurs travaux une effervescence méthodologique et une valorisation de la notion de l'anglais académique et de spécialité.

² L'OSTI report (2004) est resté un document de travail interne pendant des décennies jusqu'à sa première publication en 2004.

Sinclair fut aussi « Founding Editor in Chief » du COBUILD (Collins Birmingham University International Language Database), source pour de nombreux matériaux de références en langues. Il a travaillé sur le COBUILD corpus, qui donna lieu par la suite à la fois au corpus Bank of English (4,5 milliard de mots) d'anglais général et à de nombreux dictionnaires, dont le *Collins COBUILD English Language Dictionary* (1987), ainsi que des livres de grammaire et une série de ressources en anglais de spécialité, dans des domaines aussi variés que l'industrie du pétrole et du gaz, la comptabilité ou les assurances (Collins Learning Catalogue, 2007-2013). Selon Stubbs (2007), le projet original du COBUILD fut peu utilisable par les apprenant.e.s car les exemples étaient trop complexes, mais la méthodologie et l'approche du projet COBUILD furent essentielles à la naissance de l'approche moderne de la lexicographie (Stubbs, 2007 : 7). Nous voyons de nouveau qu'une rigoureuse analyse linguistique de corpus a un effet souvent indirect en didactique, via la création de ressources.

1.2.1.3. UNE TRADITION DE GRANDS CORPUS GÉNÉRAUX

Nous trouvons dans le monde anglophone une tradition d'analyse à partir de corpus spécialisés, mais aussi un travail très important de constitution de très grands corpus librement accessibles pour la recherche et pour les usages des apprenant.e.s. À l'heure actuelle, de nombreux corpus sont disponibles en ligne, les plus connus en anglais étant le BNC (British National Corpus) et le COCA (Corpus of Contemporary American English). Le BNC contient 100 millions de mots produits entre les années 1970 et 1994 à partir de documents britanniques de fiction, journaux régionaux et nationaux, revues populaires, des articles et essais académiques, mais aussi des transcriptions de sources orales, dont des réunions d'affaires ou gouvernementales, ou des émissions et appels radio. Le BNC est annoté par catégorie grammaticale (*parts of speech*) via CLAWS (Constituent Likelihood Automatic Word-tagging System).

Comme pour d'autres équipes travaillant sur corpus, le travail continu de l'équipe de l'UCREL (University Centre for Computer Corpus Research on Language) sur le plan informatique a permis d'arriver à un taux de justesse situé entre 96-97%, tombant à 1,5% d'erreur pour les catégories grammaticales principales, avec seulement 3,3% des ambiguïtés non résolues (UCREL, en ligne). En d'autres termes, l'annotation

automatique des catégories grammaticales a des résultats fiables, seul un petit pourcentage pose un problème d'ambiguïté. Ce taux de fiabilité est la norme pour la plupart des étiqueteurs, comme nous le trouvons au sein de Scientext (cf. 3.1.3.).

Sur le site du BNC, il y a la possibilité d'opérer une simple requête de 50 occurrences d'un mot ou d'une séquence lexicale, dont les restrictions par catégorie grammaticale. L'évolution du BNC est intéressante, car si le BNC contribua largement à l'intérêt de la recherche sur corpus à son époque, vingt ans après, les outils d'accès sont, de nos jours, dépassés. Le corpus et l'outil d'exploitation XAIRA (XML Aware Indexing and Retrieval Architecture) peut néanmoins être téléchargé, ce qui permet d'accéder à des recherches plus avancées, comme celles disponibles sur le COCA. Le BNC peut être téléchargé gratuitement et reste accessible sur d'autres sites.

Le COCA, plus récent, fut constitué entre 1990 et 2012 de 450 millions de mots issus, à parts égales, de cinq catégories : des programmes télévisés et radiophoniques ; des œuvres de fiction publiées dans des journaux littéraires ou enfantins, ainsi que des chapitres de romans ou des scénarios cinématographiques, des textes d'une centaine de revues populaires, les diverses sections d'une dizaine de journaux et enfin, les textes d'une centaine de revues académiques avec évaluation par les pairs dans un large éventail de domaines. L'interface permet d'effectuer des requêtes par mot, séquence de mots, avec ou sans joker (permettant de remplacer, par exemple une lettre ou des lettres, dont *give** peut représenter *give, given, gives, giver*) ou catégorie grammaticale. Le mode « collocation » permet de choisir l'éloignement entre le pivot et la collocation qui peut se trouver entre 1 et 9 occurrences (e.g. mots-formes ou signes de ponctuation) à gauche ou à droite de ce pivot. Sur le COCA, les résultats sont automatiquement indiqués par catégorie grammaticale, mais sans indiquer la relation syntaxique. Les résultats d'une requête en mode KWIC (*Keyword in context*) indiquent la catégorie grammaticale des quatre mots à gauche et à droite du pivot via un système de surbrillance colorée. Le corpus académique comporte 21 170 textes écrits, dont 3 208

(15,2%) en sciences et technologies, ce qui est relativement peu en comparaison avec les autres disciplines académiques³.

Ces deux grands corpus ont impulsé une véritable dynamique dans la façon d’aborder la linguistique de l’anglais général, oral et écrit, mais aussi de l’anglais académique et de spécialité. Dans ce sens, sans faire un travail sur ces deux corpus, mes recherches bénéficient de ce climat fécond et des connaissances approfondies en linguistique de corpus.

1.2.2. QUALITÉS COMBINATOIRES

La reconnaissance qu’un mot a une attraction plus forte pour certains mots et que la qualité combinatoire a un sens particulier, date de plusieurs siècles (Stubbs, 2009 : 15-16 ; Tutin, 2010).

1.2.2.1 VERS UNE DÉFINITION DE LA COLLOCATION

Tutin (2010) considère que les collocations sont « les associations mémorisées et préconstruites mettant en jeu (prototypiquement) deux éléments linguistiques sémantiquement pleins » (Tutin, 2010 : 57). Ces éléments entretiennent une relation syntaxique où au moins un élément garde son sens habituel dans une relation, qui peut être plus ou moins productive et prédictible. Les séquences à plus de deux items peuvent généralement être décomposées puisqu’une collocation (*freshly baked*) peut

³ Il me semble qu’il pourrait y avoir une catégorisation par discipline des textes plus rigoureuse que celle utilisée pour le COCA. À mon sens, il y a quatre revues qui ne devraient pas se trouver dans la catégorie des sciences et technologies : *Technology Teacher* et *Journal of Environmental Education* sont davantage orientées vers le domaine de l’éducation, la revue *Armed Forces et Society* relève plutôt du domaine des sciences politiques et la revue *PSA (Photographic Society of America)* du domaine des humanités. Ces considérations modifient le poids des catégories disciplinaires représentées dans le COCA. J’identifie 3 208 textes relevant des disciplines des sciences et technologies (15,2%), alors que le COCA estime qu’il y en a 4 176 (19,7%) avec ces quatre revues supplémentaires. De même, *Physical Educator* dépend de la discipline des sciences de l’éducation et non du domaine des sciences sociales, enfin, le *Journal of Psychology* dépend des sciences sociales et non de la philosophie / religion.

être le collocatif dans une collocation (avec *potatoes*, *muffins* ou *bread*), comme illustré dans la Figure 2. Par exemple, pour la collocation transparente *terribly expensive*, le collocatif *terribly* est productif dans le sens qu’il peut se combiner à d’autres bases (*late*, *sorry*).

Collocation <i>freshly baked potatoes</i>		
collocatif (= collocation)		Base
<i>freshly baked</i>		
collocatif	base	<i>potatoes</i>
<i>freshly</i>	<i>baked</i>	

Figure 2 : Décomposition de la collocation *freshly baked potatoes* : le collocatif est une collocation, Tutin, 2010

Il faut noter l’importance, dans le cadre didactique de la notion que « les éléments de la collocation apparaissent de façon récurrente et sont mémorisés par le locuteur. Ces suites de mots *prêts à parler* participent à la cohésion textuelle et sont souvent étroitement liées à un genre textuel » (*ibid* : 59). Nous retrouvons dans cette définition la théorie du *priming* (amorçage) en d’autres termes, l’idée qu’un mot appelle un autre mot, par effet de mémoire (Hoey, 2005).

Maniez (2012) propose une étude sur la modification d’un nom en anglais médical, qui comporte deux, voire trois façons de modifier un nom : pré-modification par un nom (*heart murmur*), complément (*murmur of the heart*) et prémodification par un adjectif (*cardiac murmur*). À partir d’une étude du corpus EMEA (European Medicines Agency), il constate, par exemple, que certains noms sont plus fréquents en collocation, avec l’adjectif *hepatic* que d’autres qui sont prémodifiés par le nom *liver*, tel que montré dans le Tableau 1 (2012 : 92-93). Il estime que la consultation de corpus spécialisés (par exemple, l’anglais médical) et généralistes (dont le British National Corpus) est importante pour les apprenant.e.s, et qu’au-delà de la fréquence statistique, le contexte

joue un rôle dans les choix discursifs. J’adopte cette position, qui met en relief l’intérêt des corpus pour identifier la finesse des pratiques linguistiques.

Tableau 1 : Extrait des données : adjectifs / noms de Maniez (2012)

Fréquence avec l’adjectif <i>hepatic</i>	Fréquence avec le nom de pré-modification <i>liver</i>	Nom modifié
3528	103	impairment
513	14	insufficiency
282	119	dysfunction
615	1164	function
80	615	disease
169	615	enzyme(s)

Gledhill (2000) propose une approche à partir des mots grammaticaux (*to, this*) à très haute fréquence, avec une attention à la section de l’article de recherche, par exemple, médicale. Il met en avant le fait que la distribution des mots lexicaux (e.g. *patients, effect*) est plus affectée par la section d’article que les mots grammaticaux (Gledhill, 1997). Dans leur discussion du terme *mot*, Granger et Paquot (2008) soulignent que les prépositions complexes (*due to*), les adverbes (*in fact*) et les conjonctions (*even if*) sont parfois considérés comme des catégories mineures ou des unités phraséologiques. Je défends la position que les collocations sont composées de deux unités lexicales : *look up* dans la phrase *look up at the sky* serait une collocation, mais dans *look up a word in the dictionary*, il s’agit d’un verbe à particule.

1.2.2.2 VERS UNE DÉFINITION DE LA COLLIGATION

Par contraste avec la collocation, la colligation introduit les catégories grammaticales (au-delà d’une relation syntaxique) au sein des schémas. Selon Sinclair (2003) :

Colligation is similar to collocation in that they both concern the cooccurrence of linguistic features in a text. Colligation is the occurrence of a grammatical class or structural pattern with another one, or with a word or phrase. “Negative”, “possessive” and “modal” are the kinds of largely grammatical categories that figure in colligation (Sinclair, 2003 : 145, cité dans Legallois, 2012 : 39).

Pour Legallois (2012), la colligation ne devrait pas être réduite à la grammaire, par exemple, comme évoquée dans la construction [Verbe + Adverbe]. Pour lui, la colligation peut être considérée comme « une corrélation étroite – statistiquement mesurable et pertinente – entre une forme lexicale et une catégorie fonctionnelle », par exemple, la négation ou la propriété. J'estime également que les colligations comportent une unité lexicale et une catégorie fonctionnelle.

1.2.2.3. AUTRES APPROCHES ET TYPES D'ASSOCIATIONS

En plus de la collocation et la colligation, les approches de la linguistique de corpus portent souvent sur l'analyse de séquences polylexicales, dont les typologies sont nombreuses. L'approche de Hunston et Francis (2000) définit des *patterns*, ou *patrons* en français, comme étant toutes les unités lexicales et structures qui sont régulièrement associées avec une unité lexicale et qui contribuent au sens (2000 : 37). La méthode employée par Hunston (2011) est essentiellement l'étude d'une concordance, afin d'identifier des similitudes. Particulièrement intéressante est sa proposition selon laquelle la linguistique de corpus peut identifier des séquences sémantiques (*by the fact that, to the fact that, despite the fact that*) et que les études :

do not tell us what language is like, or even what the English used in the New Scientist is like. Rather, they tell us what things are often said, and that is a very different question. Increasingly, however, the corpora are being interrogated for what their component texts say, as well as how they say it, and semantic sequences offer a way of expressing that information (Hunston, 2011 : 170).

En d'autres termes, la linguistique de corpus permet de révéler des patrons dans une revue en particulier, mais aussi de déterminer des sujets et des modes d'expression.

L'approche de Biber et Conrad (*cf.* Biber, Conrad & Reppen, 1998 ; Biber, Conrad & Cortes, 2004) s'appuie fortement sur la notion de fréquence et introduit la notion de *lexical bundles*. De manière générale, le seuil de fréquence d'un *lexical bundle* est d'au moins quatre mots en continu et sa fréquence est d'au moins 40 occurrences pour un million de mots et ceci dans au moins cinq textes différents du corpus (Biber, 1996 : 134). Les *lexical bundles* sont beaucoup plus fréquents dans le registre oral du discours

de la classe, à fonction référentielle, à fonction organisationnelle ou pour prendre une position (*stance*) (*ibid* : 147).

Enfin, l'approche de *priming* (*amorçage*) de Hoey (2005) s'intéresse aux effets du contact avec des séquences sur la « *mental concordance* » de chaque être humain. Comme Hoey le dit lui-même, les études linguistiques de corpus permettraient de révéler des amorçages partagés, car le corpus personnel d'un individu reste, selon lui, « *irretrievable, unstudyable and unique* »⁴ (2005 : 14). Il me semble néanmoins, que les travaux sur les corpus d'apprenant.e.s et les corpus d'expert.e.s permettent d'avoir une meilleure compréhension de l'acquisition des associations lexicales fréquentes.

1.2.2.4 APPLICATIONS DIDACTIQUES

En didactique des langues, il faut faire la différence entre l'analyse de l'interlangue (*cf.* 1.1.3.), la création des ressources (*cf.* 1.2.1.2.) et l'utilisation directe par les apprenant.es (*cf.* 2.2.2.2). En utilisation directe, Kennedy (2008) émet l'hypothèse que la phraséologie n'était pas mise en avant en contexte d'apprentissage pour plusieurs raisons : l'attente de nouveaux moyens informatiques en parallèle avec une mise en avant des théories et pratiques en didactique des langues autour de l'accent sur le sens (« *message-focused* ») au lieu d'un accent autour de la forme. Les technologies du TAL permettent assez facilement l'identification de constructions, mais sans cohérence sémantique. Des constructions comme [*at the ___ of the ___*] (*at the end of the day*) sont peu motivantes (Kennedy, 2008 : 37) et peu accessibles pour l'apprenant.e non-linguiste.

Des pistes intéressantes émanent de Pecman (2005), qui propose un dictionnaire bilingue (français / anglais) phraséologique en « langue scientifique générale » ayant une approche onomasiologique et de Williams et Millon (2014), qui proposent un

⁴ Ce qui peut être traduit par : « impossible à récupérer, impossible à étudier et unique ».

dictionnaire d'anglais « scientifique de patrons de verbes avec des classes conceptuelles qui vont refléter l'utilisation de ces verbes » en s'appuyant notamment sur la représentation visuelle de réseaux collocationnels (Williams & Million, 2014 : 65) et le projet TermITH auquel participent des membres du LIDILEM dans l'objectif d'établir une liste du lexique scientifique transdisciplinaire et d'améliorer l'extraction automatique d'information (Hatier, Yan, Tran, Tutin, & Jacques, soumis).

La notion de fréquence permet, dans un premier temps, d'identifier les structures incontournables d'un texte ou d'une section d'un texte. À mon sens, l'observation de motifs pertinents permet à l'apprenant.e d'accéder à une quantité importante de « *input* » (ensemble des données) afin de saisir une notion lexico-grammaticale correspondant à l'objectif communicatif, mais la façon de présenter des données dans le cadre d'un objectif linguistique précis reste à déterminer. Les qualités sémantiques des données et leur rapport avec un objectif discursif sont des critères essentiels.

1.2.3. LE GENRE AU SEIN D'UNE COMMUNAUTÉ DE DISCOURS

Les travaux de Swales sont ceux qui m'ont le plus influencée, car ils replacent le texte dans son contexte d'origine. Swales est essentiellement reconnu pour l'intérêt qu'il a porté aux notions de la *communauté de discours* et de *genre* (1990/2004), mais aussi pour avoir valorisé et approfondi le domaine de l'English for Academic Purposes (1984) (cf. 1.1.1). Ses positions, auxquelles j'adhère complètement, me semblent toutes pertinentes pour irriguer la recherche et les pratiques en didactique des langues étrangères.

1.2.3.1. DÉBUT SYMBOLIQUE DE L'ANGLAIS DE SPÉCIALITÉ

Selon Swales (1984), le début symbolique du concept d'anglais de spécialité se trouve dans l'article de Charles Barber (1962) *Some Measurable Characteristics of Modern Scientific Prose*, dans lequel l'auteur nous livre les pourcentages de fréquence des temps des verbes, notamment le présent simple actif (64%), le présent passif (25%), mais aussi la quasi-absence du BE + V-ing (0,6%). Swales (1984) s'est appuyé en partie sur ces données pour ne pas enseigner la construction BE + V-ing en cours d'anglais auprès des scientifiques, proposition mal reçue par sa hiérarchie, qui prévoyait que l'ensemble des

temps soit abordé en dehors de toute analyse des besoins linguistiques. Par cette démarche, Swales rapproche les pratiques en classe des pratiques linguistiques de la communauté de discours des experts.

1.2.3.2. COMMUNAUTÉ DE DISCOURS, GENRE ET BUT RHÉTORIQUE

Selon Swales (1990/2004), il y a six caractéristiques permettant de définir une communauté de discours. Une communauté de discours :

- 1) has a broadly agreed set of common goals, but not necessarily a common field of study;
- 2) has mechanisms of intercommunication among its members for actual communication, for example meetings, telecommunications, conversations, etc.;
- 3) has participatory mechanisms primarily to provide information and feedback according to the discourse community goals;
- 4) possesses and utilizes one genre or more in the communicative furtherance of its aims;
- 5) has acquired some specific lexis;
- 6) has a threshold level of members, both novice and experts, with suitable degrees of relevant content and discursive expertise. (Version abrégée de Swales, 1990/2004 : 24-27).

Nous voyons qu'une communauté se manifeste à travers plusieurs genres discursifs au plan de la communication. Citant Maingueneau (1991 : 213), Carter-Thomas (2009) définit les *genres* comme étant :

des discours 'rapportables à des lieux d'énonciation déterminés'. On parlerait ainsi d'un éditorial et d'un article scientifique primaire respectivement comme des genres du discours journalistique et du discours scientifique. [...Le] point essentiel concerne la question de finalité ou de but rhétorique (Carter-Thomas, 2009 : 136-137).

Elle précise qu'une attention aux « pratiques sociales » fonde la notion de discours spécialisé, car « tout en tenant compte de l'ancrage disciplinaire ou cognitif d'un genre, ce qui distingue un genre scientifique d'un autre est surtout cette question d'un but ou d'une fin communicationnelle commune » (Carter-Thomas, 2009 : 138). Cette conception du genre, fortement liée à des pratiques et à un objectif rhétorique, a des répercussions au plan didactique, surtout lors de l'enseignement sur tâches.

1.2.4. SITUATION SOCIO-RHÉTORIQUE ET DIDACTIQUE

1.2.4.1. TÂCHES EN CONTEXTE D'APPRENTISSAGE

Selon Swales (1990/2004), une tâche doit avoir une connexion avec une situation socio-rhétorique possible :

one of a set of differentiated, sequenceable goal-directed activities drawing upon a range of cognitive and communicative procedures relatable to the acquisition of pre-genre and genre skills appropriate to a foreseen or emerging sociorhetorical situation (Swales, 1990/2004 : 76).

Nous pouvons aussi considérer la salle de classe comme étant un contexte sociorhétorique. Lorsqu'il s'agit d'un groupe de niveau avancé travaillant sur sa propre discipline, il y a davantage de rapprochement avec un contexte socio-rhétorique d'expert.e.s, c'est-à-dire une communauté de discours (Swales, 1990/2004 : 76), comme nous avons vu dans l'interface triptyque dans l'intersection entre contexte didactique des langues et anglais de spécialité (cf. l'intersection d'union C, Figure 1, page 33).

L'anglais des sciences est une sous-catégorie de l'anglais de spécialité et reste fortement lié à l'anglais académique tout en ayant des similitudes avec d'autres langues, par exemple l'expression d'hypothèse ou de citation. Au plan linguistique, il peut être caractérisé par une plus grande fréquence de certaines unités lexicales, collocations et constructions lexico-grammaticales, qui correspondent à des objectifs discursifs spécifiques ou au genre d'expression (présentation orale, article de recherche).

En effet, Swales propose un enseignement basé sur les besoins socio-rhétoriques de la communauté de discours de l'apprenant.e, par exemple des phrases descriptives comportant une clause en *-ing*. Il s'agit d'une chaîne d'évènements, pour laquelle une série de connecteurs (*and, so*) ne serait pas souhaitable. Dans le Tableau 2, l'exemple 1 est tiré de *Writing Scientific English* (Swales, 1971, cité dans Swales, 1984 : 15) et l'exemple 2 d'*Academic Writing for Graduate Students* (Swales & Feak, 2004 : 103). Les deux sont des exemples d'une structure utile pour décrire les étapes d'un processus et généralement le résultat et ses causes.

Tableau 2 : Extraits de phrases prototypiques (Swales, 1971) et (Swales & Feak, 2004)

a	Subordinate clause	1) When the piston is drawn sharply upwards, 2) When the ABS controller senses that a wheel is about to lock up,
<i>explains</i>	Main clause	1) the air below the piston rises, 2) it automatically changes the pressure in the brake lines of the car to prevent the lockup
b	-ing clause	1) thus causing the pressure to fall (1971). 2) thus resulting in maximum brake performance (2013).

Nous remarquons dans le Tableau 2, la présence de groupes nominaux (*ABS controller, brake lines, brake performance*), qui sont des manifestations de la nominalisation en anglais des sciences. Le phénomène de nominalisation s'intègre dans le concept plus vaste de métaphore grammaticale, que nous traiterons par la suite (cf. 1.3.2.4.).

1.2.4.2. LES COLLOCATIONS ET L'APPRENTISSAGE

Je rejoins Tutin (2010) pour penser que « l'approche sur corpus est plus méthodologique et empirique que théorique » (Tutin, 2010 : 4) offrant une manifestation de langue, dans l'objectif d'y puiser les détails qui constituent la langue, dont les collocations. Les outils d'interrogation de corpus sont d'ordre lexical ou grammatical, ce qui permet d'identifier les motifs. Ce qui m'intéresse dans la notion de collocation et de colligation est l'expression sémantique, perspective portée par d'autres chercheurs. Dans le cadre de son étude phraséologique de la LSG (Langue scientifique générale), Pecman (2005) propose un dictionnaire bilingue où la consultation d'un lemme, tel que *possibilité*, renvoie à des synonymes (*accès, capacité...*), des collocations (*avoir les capacités, capacité remarquable...*) et des exemples en langue cible (*Hyperpolarized 129Xe and 3He have tremendous potential to increase...*), mettant ainsi en avant la quête de sens dans un contexte scientifique. Ses travaux à partir d'un corpus divisé par section IMRAD, pourraient permettre à l'apprenant.e d'explorer les cooccurrences, par exemple à partir du verbe PROVIDE :

to provide (additional/accurate/qualitative/detailed/all the) information on/to; to provide (general/complete/conceptual/compelling) explanation for; to provide (correct/crude) indication of; to provide (additional/clear) insight into (Pecman, 2012 : sans pagination).

J'adopte la position selon laquelle les collocations sont pertinentes au plan linguistique en anglais des sciences et en didactique, car représentatives du mode de fonctionnement langagier naturel à l'être humain, et ceci bien plus que la mémorisation de vocabulaire ou de règles de grammaire.

Comme Cavalla (2010) et Cavalla et Loiseau (2014), je pense qu'il est pertinent au plan didactique de présenter des collocations selon leurs variations formelles observées. À partir de son étude d'écrits universitaires en français sur Scientext, elle remarque que la forme neutre *présenter des résultats* n'est que très rarement présente dans le corpus, alors que la forme nominalisée *la présentation des résultats* est beaucoup plus présente.

En d'autres termes, dès son origine, la linguistique de corpus a proposé à la fois une méthodologie pour examiner la langue authentique, mais elle a surtout attiré l'attention sur la création de sens lors de la combinaison d'unités lexicales. En linguistique de corpus, il existe plusieurs courants, dont une approche par le lexique (où les collocations sont au cœur de l'analyse), une approche où la syntaxe domine (cf. Quirk *et al.*, 1985) ainsi que l'approche socio-fonctionnelle de Halliday (1997/2004) (cf. 1.3.2.). Je me suis d'abord intéressée à l'approche par *patterns*, c'est-à-dire des « comportements colligational et collocationnel du lexique » identifiés par l'observation contextualisée d'un corpus de grande taille (Legallois & François, 2006 : 40).

Pour cela, je m'appuie sur la recherche francophone et anglo-saxonne en didactique des langues et en linguistique de corpus, notamment je prends comme élément linguistique de base les patrons lexico-grammaticaux, mis en avant notamment par Hunston et Francis (1996) et Biber *et al.* (1998). J'adopte la définition de Legallois (2006) selon laquelle un *pattern* est « une configuration syntaxique, un patron distributionnel qu'intègrent certaines classes sémantiquement homogènes de mots. La dimension de ce patron est indépendante de la notion de syntagme » (Legallois, 2006 : 33), un *motif* étant un enchaînement qui peut comporter des éléments non contigus, contrairement aux patrons.

Autour des notions discursives, les travaux sur la subjectivité, dont ceux de Hyland (2005, 2012), de Fløttum (2007, 2010) et du laboratoire LIDILEM (Laboratoire de

Linguistique et de Didactique des Langues Etrangères et Maternelles) inspirent ma recherche. Aujourd’hui, j’adopte de plus en plus une approche à partir d’expressions polylexicales, par exemple, celles que j’ai incorporées dans la plateforme Corpus4All (cf. 2.1.3.6.) :

If you look closely at...
On the right/left, you can see that...
Let me now show you...
Let’s consider this now in more detail...

Ces expressions me semblent plus accessibles dans le cadre didactique des langues de spécialité auprès d’un public LANSAD, que des patrons comme [*If* + PRON + LOOK + ADV + *at*] ou [*on* + *the* + left/right + PRON + SEE + *that*]. Ce public a besoin de mémoriser et de mettre en pratique des expressions fréquentes dans leur domaine, sans pour autant être en contact avec une grande quantité d’*input* (cf. 1.1.3).

1.3. TROIS PHÉNOMÈNES LINGUISTIQUES DE L’ANGLAIS DES SCIENCES

Dans cette section, je traite de trois phénomènes discursifs qui sont récurrents en anglais des sciences : la causalité, la nominalisation et le positionnement. Ces trois phénomènes offrent d’abord une mise en contexte des théories, approches et outils que nous venons de voir afin d’illustrer leurs intérêts et applications. Surtout, ils soulignent trois phénomènes constitutifs de l’anglais des sciences et montrent donc tout l’intérêt à aborder l’anglais de spécialité comme étant une variété de l’anglais ayant des spécificités linguistiques et discursives, qui ne se limitent pas à un simple jargon technique.

1.3.1. CERNER LA CAUSALITÉ

1.3.1.1. VERBES DE CAUSALITÉ

La causalité est incontournable en anglais des sciences exactes, car l’observation de cause à effet est récurrente. Dans leur étude de la langue française, Bak Sienkiewicz et Novakova (2013) ont trouvé que la causalité est exprimée par une riche collection de verbes, dont des verbes neutres (*entraîner*), intensifs (*augmenter*), morphologiques

(favoriser) et des construction (faire + Verbe), ceci dans toutes les sections du texte scientifique, mais davantage en sciences expérimentales. Le manuel *Minimum Competence in Scientific English* y consacre un chapitre entier et propose pour l'évoquer les verbes et verbes à particule, dont : CAUSE, RESULT IN, LEAD TO, BE RESPONSIBLE FOR, BRING ABOUT, GIVE RISE TO, TRIGGER, SPARK (Blattes, Jans & Upjohn, 2003 : 82).

Dans mes propres recherches sur le résumé de l'article de recherche [HART13b], des requêtes dévoilent les verbes de causalité, d'action ou d'évènement les plus fréquents, comme indiqué dans le Tableau 3. Nous pouvons observer que parmi les cinquante verbes les plus fréquents de ce corpus, treize évoquent une relation causale, ce qui souligne l'importance de la causalité dans les textes scientifiques.

Tableau 3 : Verbes de causalité (action ou évènement) dans les résumés du corpus Scientext

Lemme	Type	Position par fréquence	Quantité
increase	intensif	4	758
reduce	intensif	9	445
decrease	intensif	16	328
affect	neutre	19	277
improve	intensif	21	260
allow	neutre	24	250
cause	neutre	26	244
inhibit	intensif	34	198
activate	neutre	34	198
make	matrice	39	178
enhance	intensif	41	162
induce	neutre	42	159
lead	neutre	46	152

1.3.1.2. CONSTRUCTIONS CAUSATIVES

Selon Payne (2008 : 181), la majorité des constructions causatives en anglais sont de type analytique, en d'autres termes, elles comportent un « verbe matrice » (*to make*), dont le complément fait référence au résultat. Dans les constructions lexicales, sans l'addition d'un opérateur/verbe matrice, il existe trois types morphosyntaxiques : sans modification du verbe (*boil*), avec modification idiosyncrasique (un autre terme) (*to fall/to fell*) ou à partir d'autres verbes matrices (*to be born/to give birth*). En d'autres termes, *she boils water* est une construction causative, mais *water boils* ne l'est pas. Les constructions causatives analytiques (avec un verbe matrice) sont peu utilisées lorsque le PATIENT de la causalité est inanimé ou sans agent : *the scientist boiled the water* est préférable à *the scientist caused the water to boil* ou *the scientist made the water boil*.

Selon Gross (2012), la causalité peut porter sur trois types sémantiques du prédicat : un évènement (*la tempête a causé des dégâts*), une action (*cela l'a fait partir*) ou un état (*l'explication a rendu ce texte plus explicite*). En français, FAIRE est un verbe fréquent de causalité, comme MAKE en anglais. Dans cette optique, Gilquin (2010) a mené une étude, à partir notamment du BNC, autour des spécificités des constructions syntaxiques de causalité comportant un des verbes CAUSE, GET, HAVE et MAKE, ce qui souligne les aspects linguistiques communs à l'anglais général et à l'anglais des sciences.

1.3.1.3. CAUSALITÉ EXPRIMÉE PAR LOCUTION

La construction d'une phrase de nature causale peut impacter la mise en relief des informations qu'elle véhicule. Prenons deux exemples contrastés de cause à effet :

- 1) La pluie a causé des dégâts.
- 2) Il y a eu des dégâts à cause de la pluie.

Dans l'exemple 1, le relateur de cause est un prédicat conjugué (*a causé*), le sujet (*pluie*) est interprété comme la cause, alors que l'objet (*dégâts*) est la conséquence. Cette phrase peut aussi porter la négation. La conséquence (*dégâts*) est thématisée dans le deuxième exemple où le verbe perd son actualisation au profit de la locution *à cause de*, mettant l'accent sur la pluie. Le Tableau 3 montre que les verbes de causalité (INCREASE,

REDUCE...) sont d'ordre transdisciplinaire, ce qui est un avantage au plan didactique. Par contraste, nous verrons que les processus spécifiques à la discipline sont présents dans les groupes nominaux liés par les verbes de causalité, comme repris par la phrase prototypique de Halliday [a *explains* b].

1.3.2. LA MÉTAPHORE GRAMMATICALE

1.3.2.1. APPROCHE SOCIO-FONCTIONNELLE

L'approche socio-fonctionnelle de Halliday mérite ce titre grâce à l'incorporation de la dimension sociale dans le sillage de la tradition linguistique du fonctionnalisme (Halliday 2003 ; Halliday, 1997/2004). Selon lui, les trois métafonctions (idéationnelle, interpersonnelle, textuelle) permettent de saisir au plan théorique, la relation entre les formes internes du langage et les pratiques langagières de la vie réelle. Les métafonctions servent à aller d'un système de probabilités (autour des choix) vers l'expression, dans un contexte précis. Au niveau de la phrase en anglais général et de spécialité, le « thème » (ce dont on parle), en début de phrase, possède un degré de « dynamisme communicatif » moindre, que la fin de phrase ou « rhème » (ce que l'on dit du thème), qui introduit généralement une information nouvelle. Cette structuration apporte une cohésion textuelle qui facilite sa lecture (Carter-Thomas, 1994 : 64). Halliday a aussi beaucoup apporté aux connaissances de l'anglais des sciences.

1.3.2.2. ANGLAIS DES SCIENCES

Dans son analyse de l'anglais des sciences, Halliday (2003) dresse l'écart entre le monde quotidien des objets et les événements du monde scientifique dans lequel la langue permet la théorisation :

Where the grammar of daily life presents the world as a mix of things and going-on, of order and disorder, stability and flux, the elaborated grammar of science reconstrues it as a world of things: it holds the world still, symbolically, while it is observed and measured – and also experimented with and theorized about (Halliday, 2003 : 421).

À mon sens, cette notion est celle qui sépare fondamentalement l'anglais des sciences de l'anglais général. Comprendre l'idée que la grammaire des sciences est une grammaire qui permet de « fixer » un mouvement en objet afin de l'analyser au point

d'en tirer des théories, c'est comprendre le fonctionnement même de l'anglais des sciences. Alors que l'approche de Swales est surtout une ouverture vers le contexte d'expression, Halliday nous fait rentrer dans le rouage du texte. Sur le plan didactique, l'approche de Swales offre un intérêt direct, celle de Halliday des connaissances métalinguistiques essentielles.

1.3.2.3. ÉLÉMENTS DES GROUPES NOMINAUX

Sur le plan grammatical, l'anglais des sciences comporte quelques traits saillants tels que les structures nominales complexes. La langue anglaise est particulièrement riche en moyens productifs pour nominaliser une action ($V \rightarrow N$ ACTION désignée par V), qui peuvent entraîner ou non une modification morphologique. Les plus fréquents étant (Payne, 2008 : 230) :

to talk \rightarrow a talk (pas de modification)

to talk \rightarrow a talk over (ajout d'une préposition)

to talk \rightarrow talking (modification morphologique)

to talk \rightarrow a talker (modification morphologique)

($V \rightarrow N$ Patient of V), to talk \rightarrow a talk

($V \rightarrow N$ Patient of V), to tutor \rightarrow a tutee (modification morphologique).

($V \rightarrow N$ INSTRUMENT for V), to grind \rightarrow a grinder (modification morphologique)

($V \rightarrow N$ Manner of V), to throw a ball that curves (balle tournante) \rightarrow a curve

Les groupes nominaux peuvent combiner diverses formes grammaticales : substantifs (*androgen, processes, marker*), adjectifs (*responsive, cellular*), *V-ed* (*assisted*) et *V-ing* (*breeding*), comme dans ces exemples tirés de Scientext :

- *androgen responsive cellular processes* [processus de type cellulaire pendant lesquels il y a une réaction/réponse à l'androgène]

- *marker assisted breeding* [croisement (pour l'élevage) qui se fait à l'aide de marquages]

Ces exemples montrent que le groupe nominal peut symboliser un processus (*processes, breeding*) qui est nominalisé. La tendance vers la stabilisation incarnée par la métaphore grammaticale ne se limite pas à la nominalisation, mais peut toucher à l'ensemble des catégories grammaticales de la phrase.

Comme Swales (cf. 1.2.3.), Halliday (1997/2004) souligne aussi le rôle essentiel des verbes dans ses travaux sur la structure prototypique en anglais des sciences [a *explains* b]. Selon Halliday (1997/2004), la construction prototypique [a *explains* b] comporte (a) et (b), qui sont fréquemment des groupes nominaux. Parmi les verbes les plus fréquents dans le corpus Scientext, figurent SHOW, IDENTIFY et SUGGEST, verbes qui sont, à mon sens, au cœur de cette construction prototypique de l'argumentation scientifique [HART13b]. Je reprends la proposition de Halliday (2006), selon laquelle cette structure prototypique englobe aussi la notion de causalité, comme lors de son exemple :

Proposition de Halliday (2006)

a	The driver's overrapid downhill driving of the bus
explains	caused
b	brake failure.

Dans cette phrase, les deux groupes nominaux (a) et (b) sont des actions (*to drive / conduire* et *to fail / lâcher*) réinterprétées comme des objets (*driving / failure*) et liées par le verbe (*explains*) qui établit une relation logico-sémantique entre les deux clauses (*to cause/provoquer*). Cette phrase pourrait être reformulée :

Reformulation de la proposition de Halliday (2006)

a	The driver was driving the bus down the hill too quickly,
explains	which caused
b	the brakes to fail.

Le nom verbal *driving* de la proposition de Halliday est remplacé dans la reformulation par BE + V-ing (*was driving*), le groupe nominal *brake failure* est remplacé par un sujet et l'infinitif. L'adverbe *downhill* se trouve dans le complément circonstanciel (*down the hill*) et l'adjectif *overrapid* par *too quickly*. Dans cette deuxième formulation, il y a un éloignement des objets concrets vers des actions ou des qualités dynamiques.

Dans la construction prototypique [a *explains* b], les groupes nominaux sont reliés entre eux, par exemple, l'extrait de Scientext détaillé dans le Tableau 4.

Tableau 4 : Extrait de Scientext [a explains b]

a	groupe nominal	...loss of the endogenous insulin supply
explains	groupe verbal	therefore results in
b	groupe nominal	exogenous insulin dependency.

Dans cet exemple trois éléments s'enchaînent : (a) un groupe nominal (*loss of the endogenous insulin supply*), un groupe verbal (*results in*) et (b) un deuxième groupe nominal (*exogenous insulin dependency*). Le groupe verbal *results in*, mais aussi les verbes PROVIDE, MADE, INCREASE, SUGGEST (cf. Tableau 3) créent une relation logique entre deux groupes nominaux afin de discerner la cause et l'effet, une étape clé dans la construction du raisonnement scientifique.

La présence importante de groupes nominaux en anglais des sciences est une manifestation, parmi d'autres, de la tendance à « stabiliser », au niveau grammatical, des phénomènes scientifiques :

[There is] a drift towards the concrete, whereby each element is reconstrued⁵ in the guise of one that lies further towards the pole of stability and persistence through time. Thus, entities are more stable than qualities, and qualities than processes; while logical semantic relators like 'and', 'but', 'then', 'so', are the least stable – and hence the most complex – of all. (Halliday, 2004 : 41-43).

Dans la Figure 3, nous voyons que toutes les catégories grammaticales sont concernées par cette tendance (*drift*) vers une forme plus stable. Les conjonctions, telles que *but*, *and*, *then*, *so* étant sur le versant le plus instable, et les entités, comme *surface* ou *cracks*

⁵ Halliday (2004) emploie le verbe anglais « reconstrue ». Le Larousse (en ligne) propose pour la traduction en français de *to reconstrue* : interpréter, expliquer, analyser ou décomposer.

étant les plus stables. Le phénomène de métaphore grammaticale se manifeste ici, par exemple, dans la transposition de la qualité *cracked* en entité *cracks*.

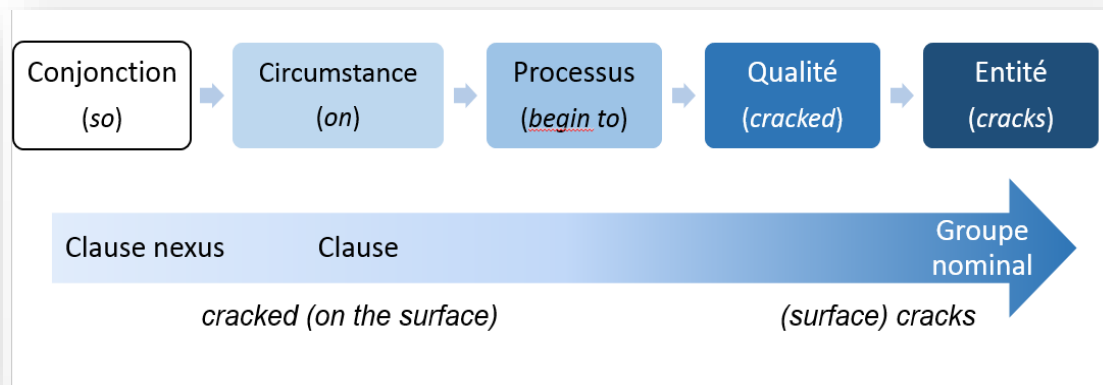


Figure 3 : Adaptation du "general drift" vers la stabilité de Halliday (2004 : 42)

1.3.2.4. GLISSEMENT GRAMMATICAL

La nominalisation est emblématique d'un glissement grammatical vers la stabilité, appelé *métaphore grammaticale*. Halliday (2004), qui a conceptualisé cette notion, met en avant le fort potentiel de la langue à théoriser en créant des phénomènes virtuels sur le plan sémiotique :

Grammatical metaphor creates virtual phenomena – virtual entities, virtual processes – which exist solely on the semiotic plane; this makes them extremely powerful abstract tools for thinking with. Thus what grammatical metaphor does is to increase the power that language has for theorizing (Halliday 2004 : xvii, gras d'origine).

Ce glissement n'est pas métaphorique dans le sens sémantique (voire traditionnel) du terme, mais évoque une notion de transposition grammaticale. Par exemple, l'expression verbale comportant un complément circonstanciel *cracks on the surface* peut être transformée, en *surface cracks*. Cette transposition ôte la préposition *on* qui dénote une instabilité physique et temporelle. Au contraire, la formulation *surface cracks* évoque un objet (*failles*) intégré à l'élément *surface* en dehors de toute circonstance de temps ou de place. Grâce à la transposition grammaticale métaphorique, les failles « sont tenues immobiles, symboliquement, pouvant être observées et mesurées », pour reprendre les mots de Halliday (*supra*). De la même

façon, *damage to the DNA* peut être nominalisé en *DNA damage*. Alors que la préposition *to* évoque un mouvement ou une action (*le dégât causé à l'ADN*), *DNA damage* réoriente le sens vers un objet (*dégât*) intégré à l'objet DNA.

1.3.3. PRÉSENCE AUCTORIALE ET POSITIONNEMENT

1.3.3.1. SUBJECTIVITÉ

Les sciences et les sciences exactes en particulier ont été longtemps considérées comme étant objectives. Cette notion a été désavouée par des études linguistiques transdisciplinaires sur la présence auctoriale et sur le positionnement. L'analyse à partir de corpus permet de cerner l'opération délicate de positionnement, essentielle dans le contexte d'une communauté de discours. L'importance didactique pour des apprenant.e.s LANSAD m'a incitée à entreprendre des recherches sur ce sujet.

1.3.3.2. L'APPORT DE HYLAND

Hyland a effectué un travail considérable autour du genre et de l'écrit académique à travers les disciplines (2002, 2006), portant son attention sur les différences entre discours des expert.e.s et des apprenant.e.s (2004). Prenant une approche interpersonnelle, il défend la position que :

in pursuing their personal and disciplinary goals, writers seek to create a recognizable social world through rhetorical choices which allow them to conduct interpersonal negotiations and balance claims for the significance, originality and plausibility of their work against the convictions and expectations of their readers (Hyland, 2005 : 176).

Cet équilibre se construit à travers les notions dialogiques de *point de vue (stance)* et d'*engagement*. *Point de vue* réfère à la « voix » ou à la marque d'autorité de l'auteur.e lors de la transmission, plus ou moins prononcée, de jugements et d'opinions en respect des normes sociales. Hyland et Sancho-Guinda (2012) se réfèrent à Gray et Biber (2012), en définissant le terme *stance* comme « a continuum of evaluative meaning which varies along two axes: one epistemic and interpersonal (i.e. from feelings and attitudes to a status of knowledge) and the other linguistic (i.e. from lexis to grammar) » (Sancho-Guinda & Hyland, 2012 : 4). Cette voix est exprimée par l'utilisation de la modalité phrastique (*possible, might*) ou au contraire de l'amplification (*boosters*) (*clearly, obviously, demonstrate*), de marques d'attitude (*agree, unfortunately, hopefully,*

remarkable) et enfin, d'autoréférence (Hyland, 2004 ; Hyland, 2005 ; Hyland 2006 ; Hyland & Sancho Guinda, 2012). L'*engagement* est la reconnaissance discursive des lecteurs et lectrices, les impliquant et les guidant dans l'argumentation et les interprétations en s'appuyant sur des pronoms y faisant référence, des directives (e.g. *consider, note, let, see, it is necessary to*), des questions, des connaissances partagées et des apartés personnels (Hyland, 2001 : 218).

1.3.3.3. LA PRÉSENCE AUCTORIALE EXPRIMÉE PAR UN PRONOM

L'utilisation du pronom est liée à la discipline (Gjesdal, 2013) et à la culture (Fløttum *et al.*, 2006a). En anglais, Hyland (2001) n'a trouvé aucune référence à *I* dans les textes de biologie, mais 24 occurrences de *we* par 10 000 mots ; cela contraste de façon significative avec les publications en philosophie pour lesquelles il y avait 35,6 occurrences d'*I* par 10 000 mots, mais seulement 1,4 occurrence de *we*. Cette prédilection pour le pluriel en biologie est un résultat du fait que les publications sont des productions collectives, contrairement à la philosophie. Swales et Feak (2004) ont trouvé que les physiciens et les chimistes ont plus tendance à adopter l'utilisation du pronom *we*, que dans les sciences sociales (Swales & Feak, 2004 : 284-285).

D'autres études examinent le verbe rattaché au pronom. À partir des données Scientext, Grossmann (2013) note que le pronom *on* est d'avantage utilisé que *nous* en collocation avec des verbes d'affirmation, surtout lors de la discussion des données linguistiques. Tutin (2010) a constaté que les verbes français évoquant une forte opinion (PENSER, ADMETTRE) sont employés avec des pronoms inclusifs, tandis que les verbes d'intention ou de choix (MONTRER, CONCLURE) étaient liés à l'autoréférence exclusive.

Selon Hyland (1996), l'utilisation de verbes de modalité et la reconnaissance de lacunes permettent d'adoucir la critique des pairs, mais l'emploi de « sujets » personnels (e.g. *our results do not support the latter observation, in our hands...*) serait le moyen le plus efficace d'atténuer une critique. En anglais des sciences, Banks (1993) trouve que les pronoms personnels sont quatre fois plus fréquents avec des verbes exprimant une « précaution oratoire » (*hedge*). Selon Banks (1993), cette collocation permet de réduire sa force, un avis qui n'est pas partagé par Hyland qui estime qu'il s'agit d'une acceptation

publique qui a pour effet d'atténuer l'affirmation, et donc d'augmenter la force de la modalité phrastique (Hyland, 1996 : 447), position confirmée en français à partir d'une étude sur Scientext par Tutin (2010), qui constate que lors d'une dimension argumentative et polémique, le pronom *on* est privilégié, les verbes sont modalisés, ce qui atténue la force de la position de l'auteur.e.

1.3.3.4. *PORTÉE CULTURELLE ET DIDACTIQUE*

Au plan didactique, une meilleure connaissance de l'emploi des pronoms est importante, car la voix passive est souvent mise en avant de manière générale afin d'éviter l'utilisation de pronoms (*cf.* Henderson, 2013). Pourtant, Tarone, Dwyer, Gillette & Icke (1981) écrivent :

Writers of astrophysics journal papers tend to use the first person plural active WE form to indicate points in the logical development of the argument where they have made a unique procedural choice; the passive seems to be used when the authors are simply following established or standard procedure, as in using accepted equations or describing what logically follows from their earlier procedural choice (Tarone *et al.*, 1981, cité dans Swales, 1984 : 13).

L'expression du positionnement est influencée par la culture. Il y a donc un fort intérêt didactique à en comprendre les contours, car une simple traduction vers l'anglais risque de ne pas être adaptée. À l'avenir, je souhaite approfondir cette question, qui ne peut être confirmée sans une exploration rigoureuse de nombreuses données.

1.3.3.5. *DIVERSITÉS D'APPROCHES*

Au plan théorique, il y a à la fois une approche lexicale autour des collocations que j'ai développée jusqu'ici et une approche grammaticale incarnée par Halliday. Ce que je retire des travaux de Halliday est son concept de métaphore grammaticale. À mon sens, il s'agit du phénomène linguistique le plus important de l'anglais des sciences exactes, parce que ce phénomène impacte l'ensemble des éléments d'une phrase et caractérise le discours en sciences exactes. J'é mets l'hypothèse que la métaphore grammaticale se retrouve dans d'autres genres de l'anglais de spécialité, par exemple dans « l'article de recherche en psychologie », mais resterait absente du genre « plaidoirie en anglais du droit ». Je souhaite ouvrir ma recherche à cette question transdisciplinaire à l'avenir. La

difficulté d'une telle recherche est la complexité grammaticale qui s'opère à tous les niveaux de la phrase, rendant une analyse linguistique à partir de corpus problématique.

Halliday explique la métaphore grammaticale à partir de quelques exemples. Dans mes propres recherches, l'analyse linguistique d'une grande quantité de verbes de causalité est une façon de commencer cette exploration à partir de corpus.

RÉSUMÉ DE LA PREMIÈRE PARTIE

Cette première partie pose le cadre théorique de ce dossier de synthèse. Elle retrace les grandes étapes de l'évolution du domaine de la linguistique de corpus et son articulation avec l'objet d'étude : l'anglais des sciences. Dès les travaux de Palmer (1933) l'attention portée à la mémoire humaine et à la didactique est centrale, puisqu'il explique qu'une collocation est une succession de deux mots ou plus qui doivent être appris ensemble (cité dans Stubbs 2009 : 17, Williams & Million 2012). Dans l'approche de Firth, l'importance du contexte et de la sémantique dans l'analyse linguistique est renforcée. Avec le rapport d'OSTI, la méthodologie progresse, la notion de genre se dessine et le débat autour de l'importance des éléments lexicaux s'amorce.

Swales confirme la notion de genre et met en avant la notion d'une communauté de discours, les deux notions étant déterminantes au plan linguistique. Il prône aussi un enseignement de l'anglais scientifique basé sur des tâches et à partir des besoins socio-rhétoriques des apprenant.e.s. L'expression de la causalité est centrale au discours scientifique, tout comme la métaphore grammaticale, qui peut permettre au plan linguistique la stabilité d'un processus sous forme d'objet. Enfin, le positionnement discursif est largement analysé grâce aux corpus, particulièrement la citation, la « précaution oratoire » (*hedge*), l'engagement, le point de vue (*stance*) et le rôle des auteur.e.s (Boch & Rinck, 2010 ; Hyland, 1996 ; Hyland, 2001 ; Hyland, 2005 ; Fløttum, Kinn & Dahl, 2006b ; Fløttum, Dahl, Kinn, Gjesdal & Vold, 2007 ; Fløttum, 2010 ; Grossmann & Tutin, 2014).

2. DEUXIÈME PARTIE : ANALYSE DES PRATIQUES, PROPOSITIONS DIDACTIQUES ET MOYENS D'ÉVALUATION

2.1. CADRES INSTITUTIONNELS DE L'ENSEIGNEMENT DES LANGUES

2.1.1. L'ANALYSE DU DISCOURS AU SERVICE DES OBJECTIFS ÉDUCATIFS

Dans le cadre de la didactique en général, l'analyse du discours permet de déconstruire un contexte d'apprentissage afin de comprendre les éléments et leur progression. Selon Cazden (2001) le langage du programme d'études, le langage de régulation (*contrôle*) et de l'identité personnelle correspondent aux trois catégories fonctionnelles de langage :

- the communication of propositional information (also termed the referential, cognitive or ideational function),
- the establishment and maintenance of social relationships,
- the expression of the speaker's identity and attitudes (Cazden, 2001 : 3).

Dans le cadre éducatif, il me semble essentiel, en effet, d'évaluer le langage du contexte et de ne pas se limiter à un descriptif des objectifs ou des programmes.

2.1.1.1. *LE CONSEIL DE L'EUROPE*

La notion de pratiques des apprenant.e.s évoque à la fois des productions langagières et des capacités de communication. Il s'agit de compétences de communication englobant des « connaissances linguistiques et pragmatiques que l'apprenant a intériorisées et qui lui permettent de comprendre et de produire des messages » (Narcy-Combes, M.-F., 2005 : 146). L'importance des compétences en langues est fortement défendue par le Conseil de l'Europe, notamment dans le cadre de la mobilité

internationale. Au-delà des intérêts immédiats d'une meilleure communication, le Conseil de l'Europe s'appuie depuis des décennies sur l'enseignement des langues dans la lutte contre la xénophobie et l'ultranationalisme [HART11c].

Dans le cadre de la construction moderne de l'Europe, l'objectif annoncé n'est pas une normalisation de l'enseignement des langues, mais la promotion du respect des identités et de la diversité à travers le plurilinguisme et l'étude des littératures nationales et régionales ainsi que l'indépendance de la réflexion, du jugement et des savoir-faire sociaux (Conseil de l'Europe 2001 : 56). Dans ce sens, j'adhère à cette affirmation de Gee (1999) selon laquelle :

Your own Discourse grid is the limit of your understanding and it is the fundamental job of education to give people bigger and better Discourse maps, ones that reflect the working of Discourses throughout society, the world, and history in relationship to each other and to the learner (Gee, 1999 : 23).

2.1.1.2. LE RÔLE DE L'ÉTAYAGE

L'approche par la linguistique de corpus et l'analyse du discours permet d'identifier à la fois la langue des apprenant.e.s et la langue des expert.e.s. Au sujet de la notion d'étayage, Bruner (1985/2012), propose que l'acquisition ne passe pas par de simples modèles linguistiques, mais par un format transactionnel où des *hints (suggestions)* et *props (accessoires)* permettent à l'apprenant.e de mieux saisir en contexte, les sens et les intentions. Bruner évoquait l'acquisition enfantine du langage, mais la notion d'étayage se retrouve dans de multiples contextes où l'accès au sens est essentiel. Dans le cas de l'enseignement de langue de spécialité, Deyrich (2001) explique que le nécessaire recours à la documentation du domaine de spécialité modifie l'étayage lors des étapes de la simplification et de la médiation didactique.

La difficulté de la compréhension du langage, des concepts et de la culture des textes à caractère spécialisé est illustrée par Isani (2009) dans sa discussion sur l'accessibilité à la FASP (fiction à substrat professionnel). Comme elle, j'adopte la notion que la triangulation auteur-texte-lecteur est un phénomène dynamique qui dépend des stratégies d'étayage mises en place par l'auteur et de l'évolution des connaissances des

lecteurs et lectrices dans un substrat donné [HART11e]. En d'autres termes, l'auteur reste attentif aux éventuelles lacunes de compréhension, et peut tenter de les combler, par exemple en introduisant des explications de texte. Ainsi, les lecteurs et lectrices évoluent sur le plan conceptuel.

2.1.1.3. L'ANALYSE DE L'ACCESSIBILITÉ

La notion d'accessibilité aux supports numériques induit aussi le concept de triangulation, surtout lorsque l'enseignant.e est absent.e. Cazade (1995) s'interroge au sujet des utilisateurs/trices :

Pourquoi certains ne comprennent-ils pas, n'assimilent-ils pas, n'enregistrent-ils pas, même quand ils passent du temps sur un logiciel, qui n'est pas nécessairement mauvais d'ailleurs ? Ils n'arrivent pas à fixer ce qu'ils ont pourtant vu, parfois lu, entendu ou même écrit. Pourquoi encore certains points passent-ils et d'autres moins ? (Cazade, 1995 : 3).

Après l'inventaire des critères techniques (livre de bord, effet zoom) d'accès à un « environnement hypertexte » bâti à partir de minicorpus, Cazade aborde la notion de transmission, c'est-à-dire la vérification de l'apprentissage, mettant l'accent sur un apprentissage actif et « les questionnements visant à structurer la connaissance et à préciser l'interaction de ses composantes » (*ibid*, 10). Ceci est une question récurrente, fruit de la multiplication des dispositifs numériques d'apprentissage de langues étrangères.

Pour y répondre, l'analyse du discours est une méthode de recherche fréquemment employée en didactique des langues étrangères en présentiel et en environnement informatisé, car elle permet d'examiner finement les réactions et les différents aspects de la production linguistique de l'apprenant.e, dont la quantité, la répétition, la reformulation, les manquements, la progression, la justesse et la corrélation avec le langage cible de l'enseignant.e et du dispositif (Chapelle, 1997 ; Chapelle, 1998 ; Mackey & Gass, 2005). En d'autres termes, l'exploration de dispositifs d'apprentissage devrait aborder la qualité du langage et du contenu proposé en direction de l'apprenant.e, ainsi que le retour linguistique ou conceptuel des apprenant.e.s, afin de mieux comprendre la transmission ou l'intégration qui s'est réellement opérée. Il faut dépasser une approche

descriptive des objectifs ou des éléments constitutifs du dispositif ou une simple synthèse des commentaires des utilisateurs/trices, pour examiner les véritables pratiques des utilisateurs/trices à la lumière des objectifs linguistiques.

C'est dans cette approche systématique qu'ont été conçus les projets explorés dans cette deuxième partie. Je retrace ici mes recherches autour de deux projets pédagogiques, *Enigma Plus* et *Notable Science*, en compréhension de l'oral en anglais scientifique, ainsi que l'évaluation de la prise de notes, une pratique authentique lors de la compréhension de l'oral.

2.1.1.4. LES COMPÉTENCES CATALOGUÉES DANS LE CADRE EUROPÉEN COMMUN DE RÉFÉRENCE POUR LES LANGUES

Manifestation de la volonté du Conseil de l'Europe, l'omniprésent CECRL (Cadre européen commun de référence pour les langues) est accompagné d'une pléthore de documents annexes développant les moyens de le lier aux modes d'évaluation des langues étrangères et régionales dans les établissements primaires, secondaires et supérieurs. Loin d'être le fruit d'une étude empirique systématique, le CECRL est basé sur un inventaire des critères recueillis auprès des enseignant.e.s à travers plusieurs pays et contextes d'Europe. Dans ce sens, il ne reflète pas une approche théorique de l'acquisition du langage, ni une approche conceptuelle des spécifications de l'évaluation, mais reste centré sur le niveau de capacité linguistique opéré dans certaines situations, notamment académiques ou professionnelles (Conseil de l'Europe, 2003 ; Council of Europe, 2001 ; Davidson & Fulcher, 2007 ; Jones & Saville, 2009). Au plan théorique, de nombreuses études nous éclairent sur les concepts d'utilité, de validité et de fiabilité, concepts centraux à l'évaluation des compétences de l'anglais général comme de l'anglais de spécialité (Bachman, 2005 ; Bachman & Palmer, 1996 ; Douglas, 2000 ; Fulcher, 2009 ; McNamara, 2000).

2.1.1.5. UNE RÉPONSE FRANÇAISE

En France, l'officialisation du mouvement européen pour l'insertion professionnelle et la mobilité fut l'arrêté du 25 avril 2002 stipulant que le diplôme de master « ne peut être délivré qu'après validation de l'aptitude à maîtriser au moins une langue étrangère »

(arrêté du 25 avril 2002 relatif au diplôme national de master - NOR : MENS0200982A). En conséquence, les programmes universitaires de l'enseignement en LANSAD sont souvent orientés vers des intérêts académiques et professionnels [HART11c]. L'annulation de cet arrêté en 2014 n'a pas eu forcément d'impact dans les établissements de l'enseignement supérieur où une politique en langue étrangère a été mise en place au fil des années.

À l'UJF, j'ai fortement contribué à la mise place d'une solution universitaire de validation des compétences en anglais de spécialité (cf. Annexe I). Comme indiqué en annexe de diplôme de master, ce dispositif est composé de trois « Épreuves de niveau 'indépendant' en anglais de spécialité (B2 du CECRL) », qui peuvent être acquis ou non acquis indépendamment des autres épreuves :

Compréhension de documents écrits à caractère scientifique
Présentation orale en lien avec sa discipline scientifique
Prise de notes à partir de documents oraux à caractère scientifique

Dans ce cadre, mes premières recherches ont ciblé la compétence de l'anglais oral scientifique et les pratiques réelles d'utilisation de supports numériques [HART05a, HART07a, HART07b]. La Direction du Service des Langues de l'UJF m'a confié dès 2006 la mission de porter le projet de renouvellement de la plateforme multimodale *Enigma Plus* en collaboration avec son concepteur, Jonathan Upjohn. Ceci marqua le début de plus d'une décennie de réflexion autour de l'apport de la pédagogie numérique à l'apprentissage, mais aussi des lacunes dues notamment au manque d'interaction sociale. Une modélisation du contexte de l'apprentissage offre la possibilité d'explorer ces questions en profondeur.

2.1.2. LA MODÉLISATION DU PROCESSUS D'ENSEIGNEMENT

Dans mes travaux, je défends l'hypothèse qu'il existe un lien fort entre la qualité des projets pédagogiques et une prise de recul permettant d'évaluer les pratiques. Selon Nancy-Combes, J.-P. (2009), l'enseignement :

correspond à la définition d'un curriculum, d'objectifs, de contenus et de tâches, et aussi à l'organisation de dispositifs d'apprentissage, la gestion de ces dispositifs, des groupes et des individus. Il s'agit d'activités sociales. Il est logique que des normes et des attentes

soient définies au niveau du code qui sera privilégié (anglais US, mandarin, arabe standard, à tel niveau de compétences, etc.) (Narcy-Combes, J.-P., 2009 : en ligne).

Pour cette analyse heuristique, j'adopte une vision systémique de ces éléments, en d'autres termes, une approche globale et interdisciplinaire qui prend en compte les effets des interactions, souvent non-linéaires, autour des contenus et actions organisés dans un objectif précis au sein d'un programme de cours. L'apprentissage n'est pas toujours le miroir des objectifs de l'enseignement. Selon Narcy-Combes, J.-P. (2009), dans l'apprentissage, « *l'input* sera reçu en fonction des caractéristiques affectives, culturelles et cognitives de l'apprenant et sera reconstruit en une interprétation relativement différente des objectifs » (Narcy-Combes, J.-P., 2009 : sans pagination). J'adopte sa position selon laquelle il est important dans toute analyse didactique d'examiner les tâches et les artéfacts réellement produits.

La recherche en anglais de spécialité a nourri mon enseignement ainsi que ma réflexion autour de la notion de « transmission » lors de la mise en place de plateformes d'apprentissage d'anglais. Une analyse des pratiques des utilisateurs/trices révèle la complexité de la multimodalité, qui modifie la réception des informations par l'apprenant.e selon le média. Puis, un travail sur la prise de notes a permis de nouveau une réflexion sur la multimodalité, puisque l'apprenant.e gère à la fois les voies visuelles, auditives et la production écrite. Dans ce sens, la division actuelle par compétences (écrit ou oral, compréhension ou expression, voire interaction) est parfois artificielle. Pourtant, ces notions facilitent une prise en compte de la distance qui sépare l'interlangue des apprenant.e.s dans l'enseignement supérieur et le discours des expert.e.s.

La modélisation permet d'apporter une synthèse des éléments et des relations entre les éléments constitutifs d'un système d'apprentissage. Les éléments peuvent être des objets, des productions langagières, des personnes, des processus, des concepts et des théories. Les modèles reflètent une ou des théories conceptuelles et permettent aussi d'identifier des éléments et d'éventuelles lacunes et de prendre, en conséquence, des décisions éclairées en fonction du contexte.

2.1.2.1. *DÉBAT AUTOUR DES THÉORIES DE L'ACQUISITION DES LANGUES ÉTRANGÈRES*

Avant d'aborder la modélisation, il semble nécessaire de traiter brièvement de la question de l'apprentissage. Selon Matthey et Véronique (2004), il existe trois courants principaux concernant l'acquisition des langues étrangères : la grammaire générative chomskyenne, le fonctionnalisme et l'interactionnisme, cependant :

Ces ensembles de travaux [en acquisition des langues étrangères], en faible synergie ou qui s'ignorent superbement, sont traversés et structurés par des controverses scientifiques entre approches compréhensives et perspectives explicatives, entre représentations concurrentes des activités cognitives (localisées à l'intérieur de l'individu ou dans la structuration sociale de ses activités avec autrui) et par les polémiques entre théories linguistiques (qui s'opposent notamment sur les conceptions philosophiques de l'inné et de l'acquis, du rationalisme et de l'empirisme). [D'autres] illustrent bien ces courants philosophiquement et épistémologiquement opposés avec, d'un côté, le rationalisme de la théorie de la Grammaire Universelle (GU) et de l'autre un empirisme pragmatique mettant au centre de la connaissance scientifique les aspects socioculturels des compétences linguistiques (Matthey & Véronique, 2004 : 203).

Il existe donc un débat important autour des théories de l'acquisition des langues étrangères. Ce qui semble certain, c'est que le processus cognitif de l'apprentissage demande à être en contact avec une grande quantité d'*input* oral et écrit, et dont certains aspects doivent être mémorisés afin de former des automatismes en dépit de l'influence de la langue maternelle. Comme je l'ai évoqué dans ma thèse (2009), ce processus cognitif est influencé par de nombreux facteurs, dont la motivation (Raby & Narcy-Combes, J.-P., 2009), l'affect, l'attention (Ellis, 2001), l'entraînement (Hilton, 2005), les objectifs, les connaissances métacognitives et leur emploi, les stratégies d'apprentissage et les savoir-faire (Dörnyei, 2006 ; O'Malley & Chamot, 1995) ainsi que par la pratique de la communication ou encore le contexte social et physique.

En d'autres termes, bien qu'il existe des controverses autour de l'acquisition des langues, certains aspects saillants ont été identifiés, notamment l'importance de l'exposition à la langue cible et l'importance de prendre en compte l'individu dans un contexte socio-rhétorique.

2.1.2.2. UN MODÈLE DE RÉLATION D'AGENCE

Selon Blin et Jalkanen (2014), la qualité linéaire des *learning design methodology* (méthodologie de la conception de l'apprentissage) et *design-based research* (recherche fondée sur la conception de modèles) ne donne pas la possibilité aux participant.e.s de proposer des changements (*agency*). Ils suggèrent par contre que la *cultural historical activity theory* offre un espace d'équilibre entre les conceptions du duo « enseigner » et « apprendre », ce qui ouvre la porte à des modifications du système par les étudiant.e.s ou les enseignant.e.s. Le modèle d'un projet de classe, sous forme de triangle et à partir des trois axes clés (les outils et les artéfacts, les règles et enfin, les divisions du travail) est présenté dans la Figure 4 (Blin & Jalkanen, 2014 : 157). Ils prennent comme exemple un projet collaboratif et en ligne qui consistait en l'organisation d'un évènement pour lever des fonds et pour éveiller les consciences pour une association caritative aidant des personnes sans domicile fixe.

Le sommet du triangle représente la partie visible de l'iceberg, puisqu'il s'agit des outils et des artéfacts, parmi lesquels les consignes, les connaissances de base, les modèles de collecte de fonds et toutes sortes de technologies (*cf.* Figure 4). Ils font référence à Engeström (2008), qui nomme les trois éléments de la ligne inférieure le « curriculum caché ». En bas à gauche, se trouvent les règles explicites ou implicites autour des emplois du temps, de l'évaluation et des modes d'interaction comportementale. Centrée en bas, est représentée la communauté, dont les étudiant.e.s, les équipes enseignantes et administratives et l'association caritative. Et enfin, à droite la répartition du travail entre les étudiant.e.s, les enseignants ou la communauté. Sur la ligne médiane horizontale, nous trouvons les équipes d'étudiant.e.s ainsi que l'objectif du cours. Enfin, la production du cours est repoussée en dehors du triangle et vers le haut, car il s'agit d'un produit visible. L'existence de contradictions au sein de ce système dynamique est conçue comme source de développement.

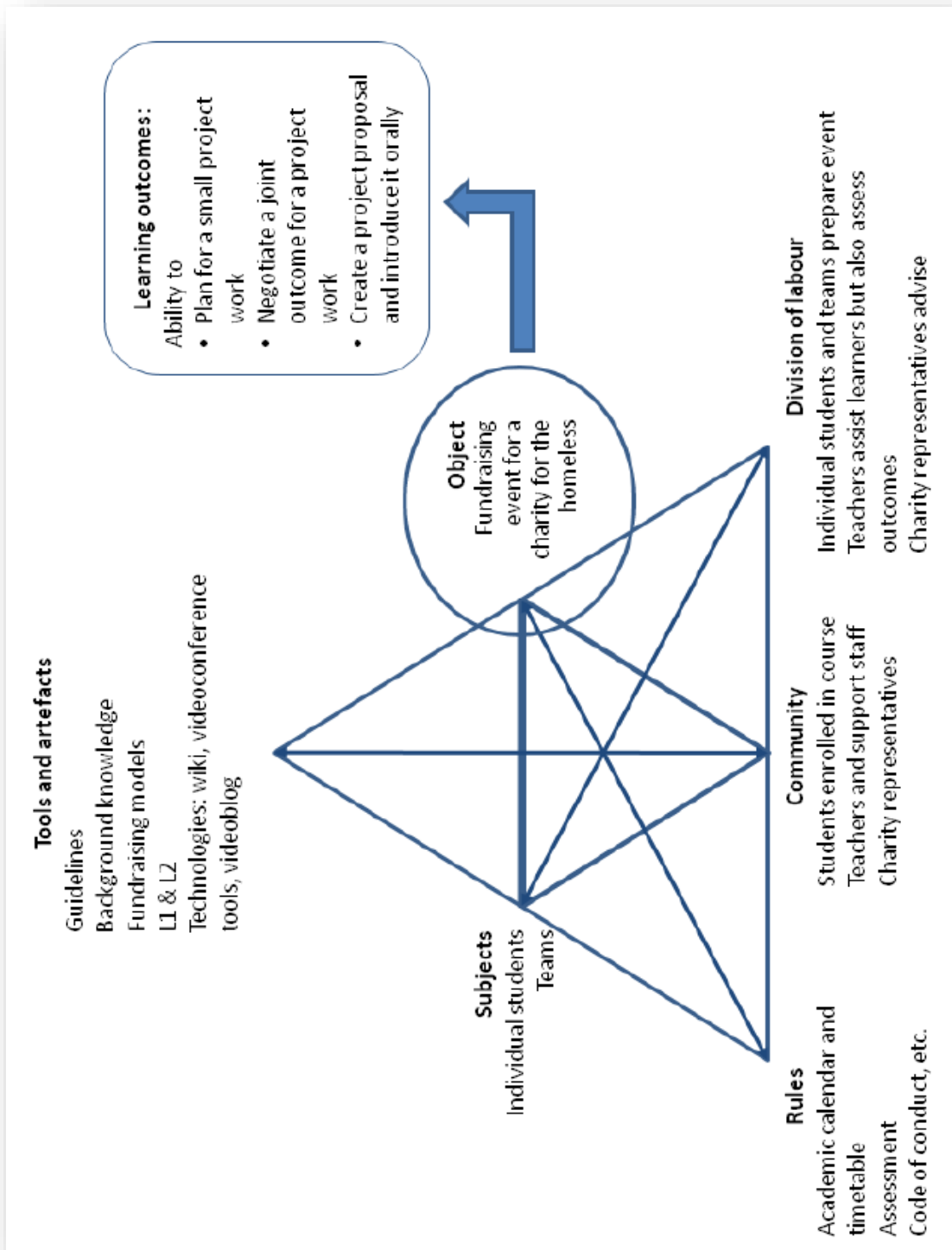


Figure 4 : Representations of a mediational structure de Blin et Jalkanen (2014 : 157)

2.1.2.3. UN MODÈLE D'ERGONOMIE DIDACTIQUE

Par contraste, Bertin (2001, 2014) et Bertin, Gravé et Narcy-Combes, J.-P. (2010) proposent un modèle holistique d'enseignement par tâches et médiatisé par l'informatique et ceci, sous la forme d'un ovale dans lequel les théories du langage et de l'acquisition de langue étrangère sont explicitement notées. Ces aspects influent sur la mise en place des macro- et micro-tâches au centre du modèle. Le modèle d'ergonomie didactique présenté dans la Figure 5 est basé sur une approche socio-constructiviste de l'apprentissage dans lequel l'élément « technologie » est intégré au triangle classique langue/culture + enseignant.e + apprenant.e et. La notion de contexte est placée en haut de la figure, montrant ainsi son influence sur l'ensemble. Le jeu de flèches grises (l'organisation de l'interaction gérée par l'enseignant.e ou type d'interaction avec la langue), pointillées (interaction directe de l'apprenant.e), à trait (relations avec le contexte) et à plusieurs nuances de gras et à tête unique ou à double tête relie les éléments constitutifs, encadrés de traits gras (en rouge dans le texte d'origine) et des éléments particulièrement liés sur fond gris.

Il y a un travail soigné autour de l'interface technologique entre l'enseignant.e et l'apprenant.e, en possible présence des pairs (*peers*). Enfin, la présence de tuteur et du monitoring est aussi représentée. Sur ce dernier aspect, Guichon (2011) propose une modélisation concise des compétences nécessaires à l'activité du tutorat synchrone autour des trois types de régulation : socio-affective, pédagogique et techno-sémio-pédagogique. Le modèle proposé évite de cataloguer les détails afin que l'activité ne soit pas perçue comme l'addition de compétences disjointes (*cf.* Annexe II).

Deyrich (2007) propose un modèle « expérientiel » dans le cadre de la formation des maîtres, le point de départ de la mise en place de tâches d'acquisition lexicale venant de l'expérience en langue cible. Puis, intervient la distanciation par observation réflexive, le recours aux principes théoriques avant la prise de décisions didactiques et de transformations pédagogiques. Cette démarche met l'accent sur les aspects humains (perspectives de l'enseignant.e et des apprenant.e.s), tout en s'appuyant sur les données lexicales et la théorie d'acquisition du langage (*cf.* Annexe III).

Selon Bertin (2015) un modèle n'a pas pour objectif d'être normatif, mais doit servir à dévoiler la complexité et à engager une interrogation sur les pratiques, par exemple sur l'utilisation du lecteur MP3 en cours de langue ou des différents types de Mooc (Massive open online course). Son cadre global évite un émiettement sur un aspect ponctuel, et tend vers une cohérence théorique.

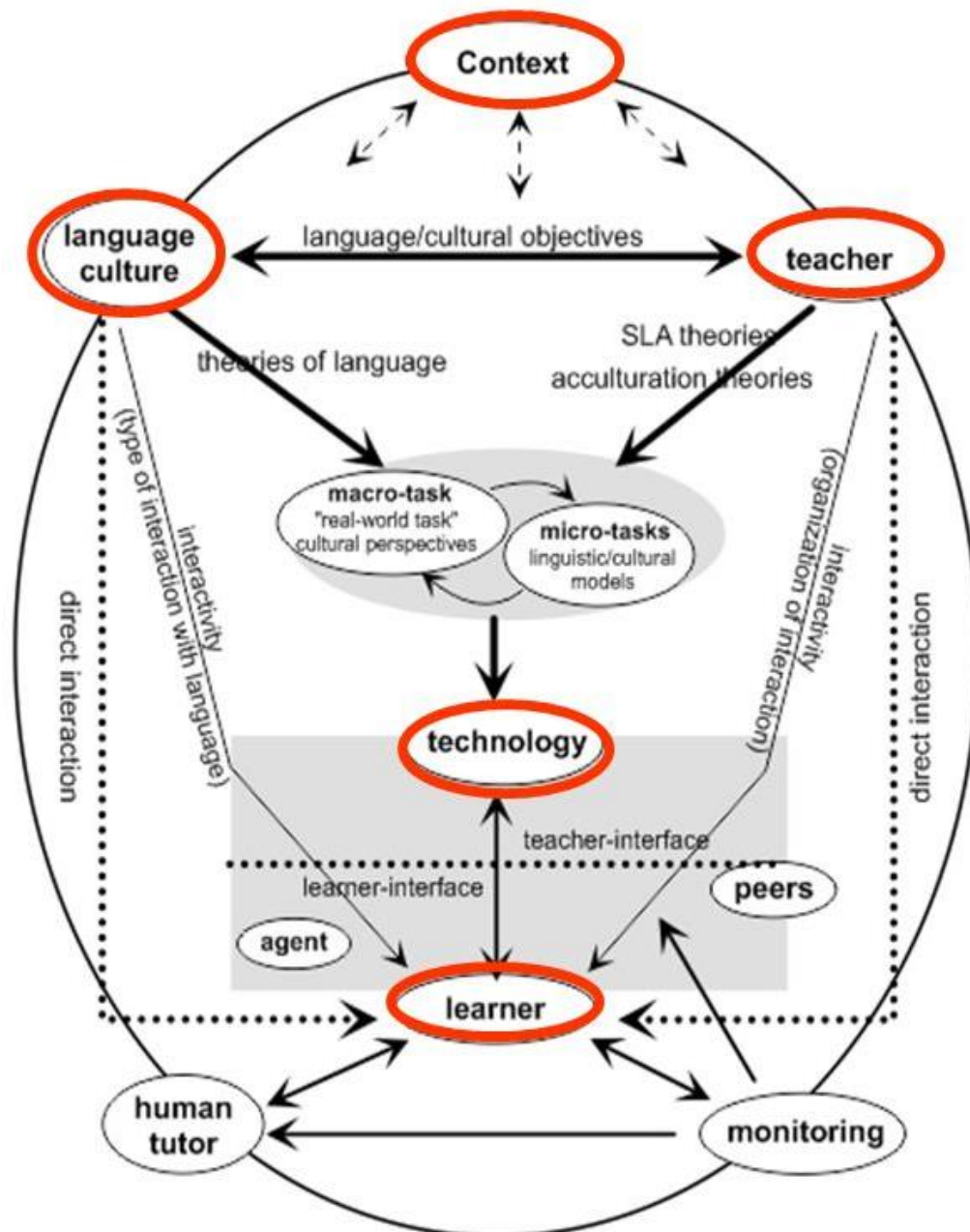


Figure 5 : Modèle d'ergonomie didactique du tutorat de langue en ligne de Bertin (2015), Bertin *et al.* (2010 : 27).

2.1.3. UN MODÈLE SYSTÉMIQUE DE L'ENSEIGNEMENT

Au fil des années, j'ai conçu un modèle systémique de l'expérience de l'enseignement/apprentissage dans le cadre institutionnel qui représente une macro-vision des éléments, comme illustrée dans la Figure 6. Ce modèle reprend les trois notions que nous avons déjà abordées dans la discussion autour de la linguistique, notamment : la sphère idéationnelle (le contenu en langues et cultures), la sphère interpersonnelle (l'ensemble des acteurs et des actrices) et la sphère structurelle, cette dernière remplace la notion textuelle trouvée dans le discours de Halliday sur la linguistique (cf. 1.3.2.).

Les trois sphères sont entourées par des lignes discontinues afin de représenter la fluidité des contours. Enfin, l'importance de la recherche est soulignée dans chaque sphère, car c'est grâce à la recherche que nous comprenons les objets et les processus du contexte d'apprentissage et d'enseignement. Ce modèle diffère aussi de l'approche de Cazden (2001) (cf. 2.1.1.), car à la place de la sphère structurelle, elle propose la représentation de l'expression de l'identité et les attitudes de l'apprenant.e (Cazden, 2001 : 3). Dans le cas de l'anglais de spécialité, la capacité à s'intégrer dans une communauté de discours (cf. 1.2.3.6.), correspond à cette notion d'identité et se trouve dans le tableau sous le terme *besoins professionnels* [HART13a].

Contrairement au modèle de Blin et Jalkanen (2014), les tâches (cf. 1.2.4.1.) et artefacts sont au centre du modèle, car plus qu'un « produit » du contexte d'apprentissage, les tâches et artefacts, qui ne se limitent pas aux évaluations, devraient être au centre d'une analyse heuristique de ce même contexte, comme dans le modèle de Deyrich (2011) (cf. Annexe III). Comme pour les autres modèles, nous y retrouvons à la fois l'individu apprenant, ainsi que l'ensemble des étudiant.e.s, mais ici, un poids plus important est octroyé à la sphère structurelle.

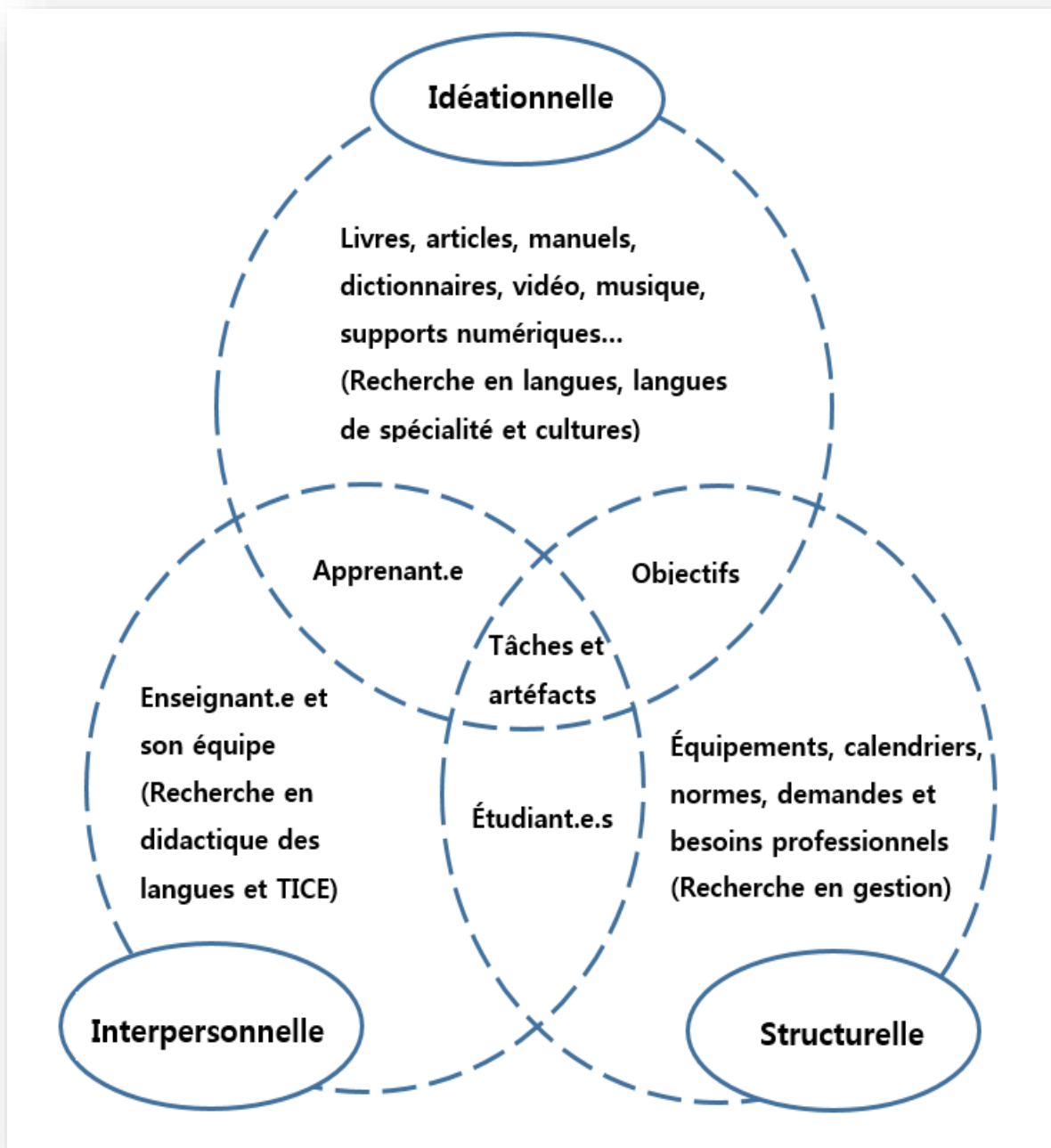


Figure 6 : Modèle de trois sphères du contexte d'apprentissage institutionnel

À mon sens, il ne faut pas sous-estimer l'impact des conditions physiques de travail, notamment l'équipement informatique, les emplois du temps, les contraintes des normes, dont le CECRL, qui dominant la scène de l'enseignement des langues étrangères actuellement en France ou encore les demandes des entreprises, comme par exemple, le critère de certification en langues lors du recrutement (*cf.* 2.3.1). L'incorporation de

cette sphère structurelle entraîne le fait que l'équipe (enseignante, administrative, technologique) soit mise en avant. Cela souligne aussi l'importance des relations entre les personnels, qui peut se manifester, par exemple, par la réactivité du service des ressources documentaires ou la qualité du suivi informatique.

Afin d'illustrer mes recherches et travaux dans ce domaine, je propose trois études de cas. Dans le premier cas, il s'agit d'un cours collaboratif d'anglais entre la Chine et la France. Le deuxième cas est l'étude d'un dispositif de soutien en ligne pour les étudiants en master en grande difficulté linguistique. Enfin, le troisième cas conceptualise une ressource d'anglais des sciences à destination des étudiant.e.s de niveau avancé.

2.1.3.1. PROJET INTERCULTUREL MÉDIATISÉ PAR UN WIKI COLLABORATIF

En 2013, Bin Zou de Xi'an Jiaotong-Liverpool University et moi-même avons mis en place un cours collaboratif entre ses quarante étudiant.e.s de licence en gestion des affaires en Chine et mes vingt-cinq étudiant.e.s de licence de management du sport en France [HART13d, HART14b]. L'objectif étant d'offrir une ouverture culturelle et un contexte de communication authentique avec des pairs ayant des intérêts en commun. Afin de réduire le stress lié à l'expression en public, nous avons introduit progressivement les moyens de communication, en commençant par l'écrit via une plateforme Wikispaces, et terminant par la communication vocale via Skype en fin de semestre. Chaque groupe avait un objectif spécifique sur le plan linguistique. En Chine, l'accent était autour du discours de la gestion et la préparation de voyages en Europe ; en France, le discours d'une offre de voyage touristique ciblant une population chinoise (*cf.* Figure 7).

Ce projet a demandé une préparation et une progression importante en moyens informatiques. Le premier wiki, créé en France n'était pas accessible depuis la Chine et nous avons adopté Wikispaces. Il a fallu expliquer aux douze groupes de travail, les aspects techniques de Wikispaces, mais aussi Animoto (service internet permettant de créer gratuitement des clips vidéo de courte durée à partir de supports visuels et d'enregistrements audio), Audacity (pour l'enregistrement de la voix et le traitement du document audio) et Skype (pour la communication et l'enregistrement de la voix).

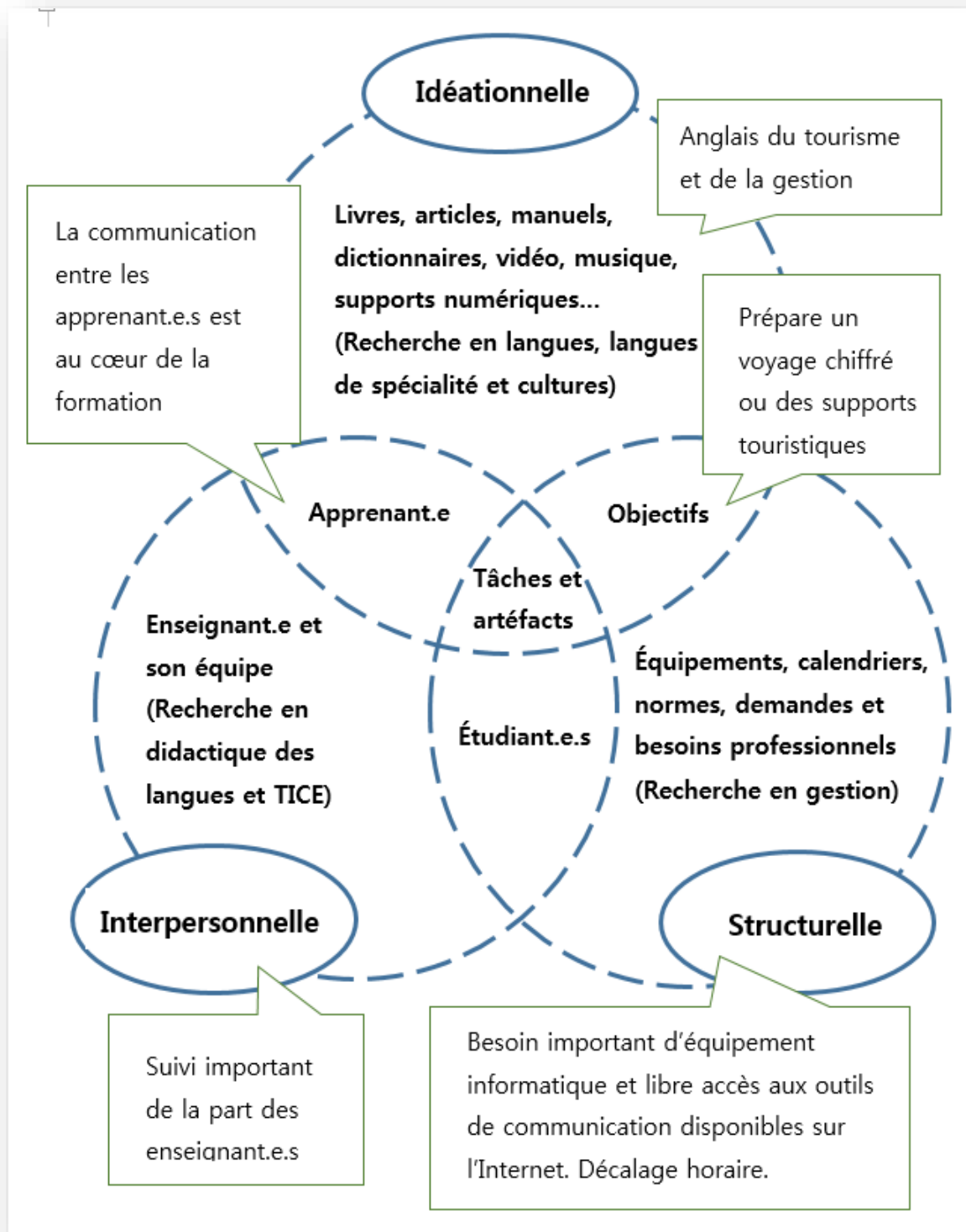


Figure 7 : Modèle d'un cours collectif médiatisé

En Chine, il y avait un meilleur support informatique et le groupe chinois était, de manière générale, mieux équipé que le groupe français. Dans mon établissement, l'utilisation de Skype est proscrite pour des raisons de sécurité. Le manque de personnel informatique a pour résultat qu'aucun autre dispositif similaire n'est proposé et qu'à l'époque, il y n'avait pas de suivi informatique de l'équipement du laboratoire des langues. Estimant que le projet ne comportait pas de risques de sécurité pour l'établissement ou la nation, comme cela peut être le cas lors des discussions autour de l'industrie nucléaire ou de la bactériologie, nous avons adopté cet outil en classe. Le manque d'antivirus sur chaque poste était certainement plus problématique, mais ceci est vrai pour tous les travaux effectués dans le laboratoire.

Au plan de l'interpersonnalité, nous avons anticipé l'anxiété due à la perspective de parler avec des personnes inconnues. La communication par l'image lors des clips vidéo créés via Animoto a permis d'établir des liens affectifs et une ouverture culturelle [HART11d]. En fait, la technologie, comme Audacity et Skype, permet d'enregistrer sa voix et de l'entendre en classe ou en privé, possibilités que des étudiant.e.s ont déclaré apprécier lors de l'enquête en fin de semestre [HART13d, HART14b].

Le projet a aussi nécessité un engagement important de la part de l'équipe enseignante, en contradiction avec un contexte institutionnel où les heures sont comptées. Au plan structurel, il a fallu une organisation rigoureuse et visible des groupes et des emplois du temps (*cf.* Figure 8). Le manque de temps est le point de tension dans ce type de contexte multinational à objectifs variables.

Dans la sphère idéationnelle, la mise en place des outils informatiques a rendu les étudiant.e.s responsables d'une grande partie du contenu culturel du cours. Les dispositifs Wikispaces et Animoto ont notamment permis le partage des images et des photographies personnelles, dont à titre d'exemple, l'image de la Figure 9.



Figure 8 : Saisie d'écran de la page d'accueil du dispositif de Wikispaces



Figure 9 : Image postée sur le Wikispaces

Le mélange de textes, d'images personnelles ou tirées de l'Internet, d'enregistrements audio et audiovisuels a constitué un environnement multimodal, contribution productive, car l'imagerie visuelle ou mentale peut être exploitée en classe de langues (Décuré, 1994 ; Décuré, 1996 ; Décuré, 2000). Décuré prône une attention à l'esthétique et à la participation personnelle en soulignant que les photographies utilisées en cours de langues doivent être « belles, insolites, frappantes, car autrement on n'a pas envie de les regarder, donc pas envie de jouer, donc pas envie de parler » (Décuré 1990 : 48).

À cause du décalage horaire, les étudiant.e.s en Chine s'apprêtaient souvent à dîner et il y a eu de nombreuses discussions autour des deux types de cuisine, aussi accompagnées d'images. Ce partage a participé à la découverte culturelle, mais a aussi eu pour effet de créer des liens affectifs, qui sont porteurs d'une plus grande aisance à l'oral. Dans l'enquête en fin de semestre, nous trouvons des affirmations, telles que « *Chinese correspondents were not shy and were very happy to talk to us* » ou « *[I am] a little less afraid of speaking in English* ». Je résumerai en disant que les aspects humains furent le point fort de cette expérience parfois chronophage.

2.1.3.2. DISPOSITIF DE SUPPORT EN LIGNE : ENGLISH BOOSTER COURSE FOR MASTER'S DEGREE STUDENTS (EBCM)

Le deuxième cas concerne un dispositif de soutien en ligne, « English Booster Course for Master's Degree students » (EBCM), pour les étudiant.e.s en deuxième année de master

en parallèle avec des cours en présentiel dans les parcours d'activités physiques et sportives, biologie, chimie, génie mécanique, informatique, mathématiques, physique. L'objectif du programme en master de l'établissement est construit autour la validation de trois compétences en anglais de spécialité (cf. 2.1.1.5.) :

- 1) Compréhension de documents écrits à caractère scientifique
- 2) Présentation orale en lien avec sa discipline scientifique
- 3) Prise de notes à partir de documents oraux à caractère scientifique

L'établissement a financé les travaux d'une équipe de trois enseignants et trois ingénieurs en informatique que j'ai pilotée (cf. Figure 10). Le parcours, créé sur la plateforme de Chamilo, est divisé en trois sections (*Back to basics, Consolidating skills, Final steps*). Les deux premières sections proposent des aides à un examen de compréhension de documents écrits à caractère scientifique. Toutes les sections proposent, de manière progressive, des aides à la préparation aux épreuves de la présentation orale et de la prise de notes (cf. 2.3.2-3).

Comme souligné dans le modèle de la Figure 11, il y a une forte influence institutionnelle, en particulier venant du Conseil de l'Europe. En réponse à cela, l'article 6 de l'arrêté du 25 avril 2002 et l'article 16 de l'arrêté du 22 janvier 2014 relatifs au diplôme national de master, affirment que « le diplôme de master ne peut être délivré qu'après validation de l'aptitude à maîtriser au moins une langue vivante étrangère. » Contrairement aux écoles d'ingénieur, l'établissement ne souhaitait pas refuser le diplôme de master en cas d'échec. Il a été proposé donc un soutien en ligne obligatoire pour les étudiant.e.s n'ayant pas validé les trois compétences en anglais (cf. 2.1.1.5.).

Un premier soutien en ligne fut établi dès 2009. Au fil des années, il apparaissait que le parcours en ligne posait deux problèmes structurels : son déroulement à la fin de l'année académique alors que beaucoup d'étudiants sont en stage, et l'objectif langagier qui était la rédaction d'un CV et d'une lettre de motivation en anglais. J'ai obtenu le déplacement du dispositif en début de semestre, en parallèle avec les cours en présentiel. Puis, au plan idéationnel, il y a eu un déplacement vers l'anglais des sciences en soutien aux travaux des cours et en préparation aux épreuves.

Chamilo
E-Learning & Collaboration Software

Page d'accueil Mes cours Blog enseignant Aug 2016 English Booster Course for Master's 2015-2016

Hartwell Laura

Vue prof

Welcome to the English Booster Course for Master's degree students!

L'ensemble des activités sont prévues de manière progressive afin de vous aider à réussir dans votre cours en présentiel. Ce parcours ne remplace pas le travail demandé par votre enseignant.e et ne valide pas le niveau "indépendant" (B2).

Une amélioration des trois compétences (compréhension de l'écrit, prise de notes et présentation orale) sont visées par les différentes activités, sachant que elles sont complémentaires et que vous pouvez refaire toutes les exercices tant que vous le souhaitez. Nous vous conseillons de faire deux ou trois activités par semaine afin de profiter pleinement des ressources disponibles.

Les groupes de niveau A1 et A2 feront trois activités d'écriture sur le Wiki qui doivent être terminées aux dates précisées dans l'agenda. Vous aurez un retour personnalisé auquel vous devez répondre concernant les productions écrites. En dehors de l'écriture sur le Wiki (activité tutorée), votre progression sur le semestre est évaluée par la plateforme de manière automatique. (Le groupe composé d'étudiants de niveau B1, non évalués et en difficulté lors de l'examen en M1 n'aura accès aux activités d'écriture sur le Wiki, qu'en cas de non validation des compétences pendant le semestre).

N'hésitez pas à participer au forum !

-) *Activities marked with an * can be included in your note-taking or self-learning assignment of your regular English class.*

Unit 1 Back to basics

Oral presentations and present forms, Reading comprehension, Listening to Enigmas*, Comparatives and superlatives for oral presentations, Reading comprehension, General Keywords*, Wiki time* (25-75 words).

Unit 2 Consolidating skills

Future forms for oral presentations, Reading comprehension, Modality for oral presentations, Note-taking (Conseils du SUP), Past forms for oral presentations, Reading comprehension, Keywords for science*, Compounds and possessives for oral presentations, Plagiarism, Wiki time (75-125 words)

Unit 3 Final steps

Passive forms for oral presentations, Note-Taking (Notable Science), Conditionals for oral presentations, Infinitives and -ing for oral presentations, Pronouncing difficult keywords, Wiki time* (100-150 words).

Description de cours
Documents
Forum

Agenda
Parcours

Figure 10 : Saisie d'écran du English Booster Course for Master's degree students

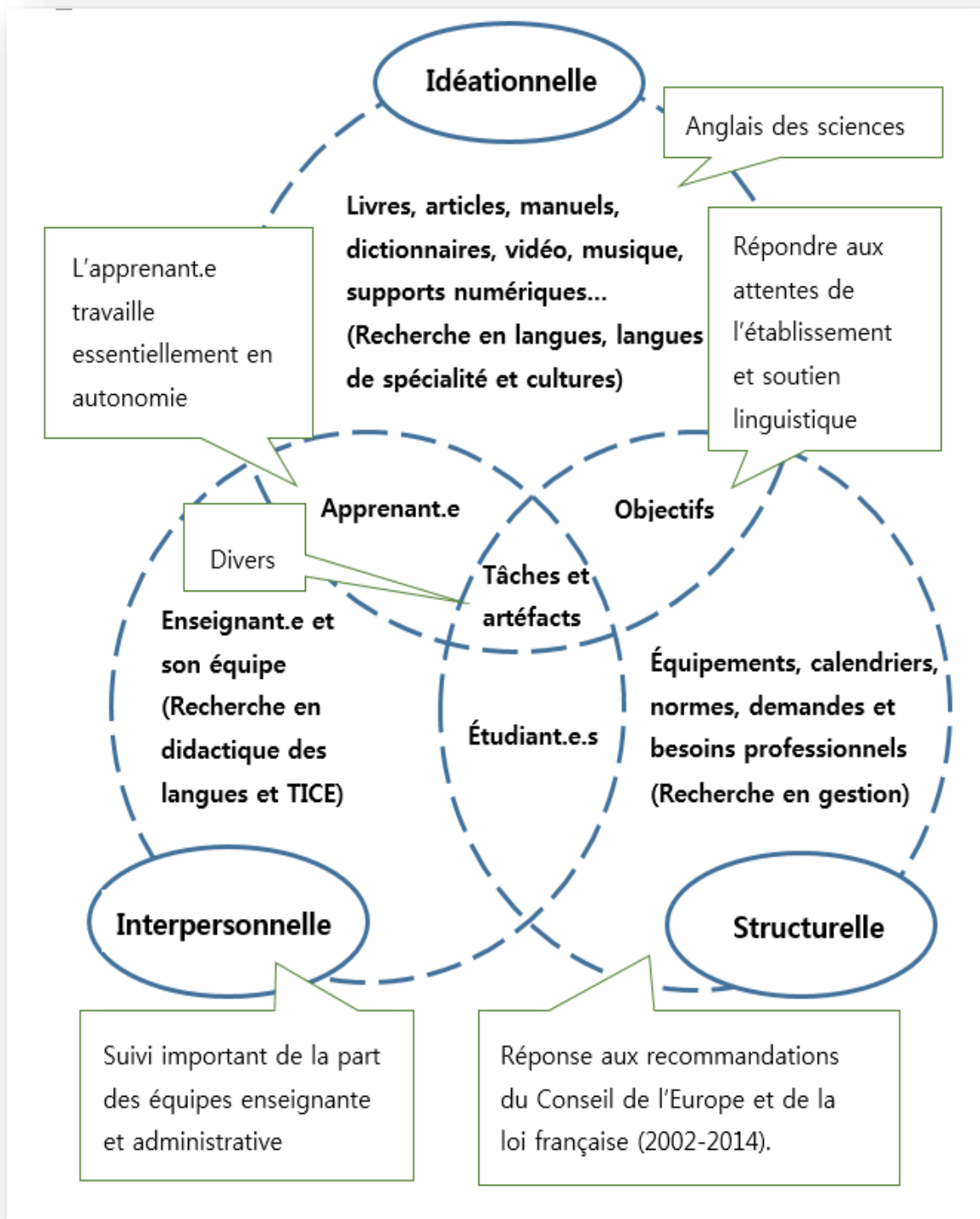


Figure 11 : Modèle d'un dispositif d'aide en ligne, l'EBCM

Le résultat de la réorientation à la fois des dates du déroulement et des objectifs a eu un impact direct sur la production sous forme de tâches et celle des artéfacts. En dehors des exercices effectués par les étudiant.e.s, nous pouvons concevoir une amélioration

pour les travaux en classe. Cette hypothèse demande une réponse qualitative et quantitative, qui mérite d'être traitée par la suite. Malheureusement, les données des pratiques sur l'EBCM de l'année 2014-2015 ont été effacées lors d'un nettoyage du système. Il faut attendre les prochains résultats.

2.1.3.3. DISPOSITIF BASÉ À PARTIR DE CORPUS : CORPUS4ALL

Le troisième et dernier cas que nous explorons, le projet Corpus4All, est un travail collaboratif en cours de création. Le financement de l'Université numérique en région Rhône-Alpes (UNR-RA) obtenu par Joséphine Rémon (ICAR) et moi-même est accompagné par des contraintes importantes en ce qui concerne le calendrier.

Il ne s'agit pas d'un parcours d'apprentissage proprement dit, mais d'une ressource pour les apprenant.e.s de niveau avancé (niveau B2 du CECRL ou plus) qui rédigent un article de recherche ou préparent une présentation de conférence. Pour la sphère idéationnelle située au sommet du modèle, les données sont issues des corpus. Le principe de s'appuyer sur les données de corpus n'a rien d'original, car à l'heure actuelle beaucoup d'éditeurs de dictionnaires adoptent cette méthode : les dictionnaires *Pearson Longman* sont compilés à partir du *Longman Corpus Network*, les dictionnaires Macmillan à partir du *World English Corpus*, les dictionnaires d'Oxford sont basés sur l'*Oxford English Corpus* (cf. 1.2.1.2.)

La différence entre ces dictionnaires et cette ressource est qu'elle cible les aspects linguistiques pouvant poser des problèmes à un public de niveau avancé et ceci pour des tâches linguistiques complexes à haute valeur professionnelle. Elle est divisée, pour l'instant, en trois grandes sections : *Structuring a text or talk*, *Giving opinions* et *Difficult Expressions*. Au sein de chaque section, les exercices seront introduits par la fonction (*Guiding the audience, introducing a conflicting fact...*) avec le point lexico-grammatical correspondant : « *Focus on 'see' and 'look'* », « *Referring to time: 'within' or 'by'* ».

Dans la Figure 12 du modèle du projet Corpus4All en cours de réalisation, nous voyons que les aspects structurels sont influencés par les objectifs du financeur du projet, l'Université Numérique en Région Rhône-Alpes. Il s'agit là d'une ressource, et non d'un environnement d'apprentissage au sens traditionnel du terme. Elle comporte cependant

des exercices en relation avec les notions explorées. Comme les données et les tâches ne sont pas encore définies, elles sont représentées par des nuages. Dans ce sens, un modèle permet aussi d'identifier une lacune.

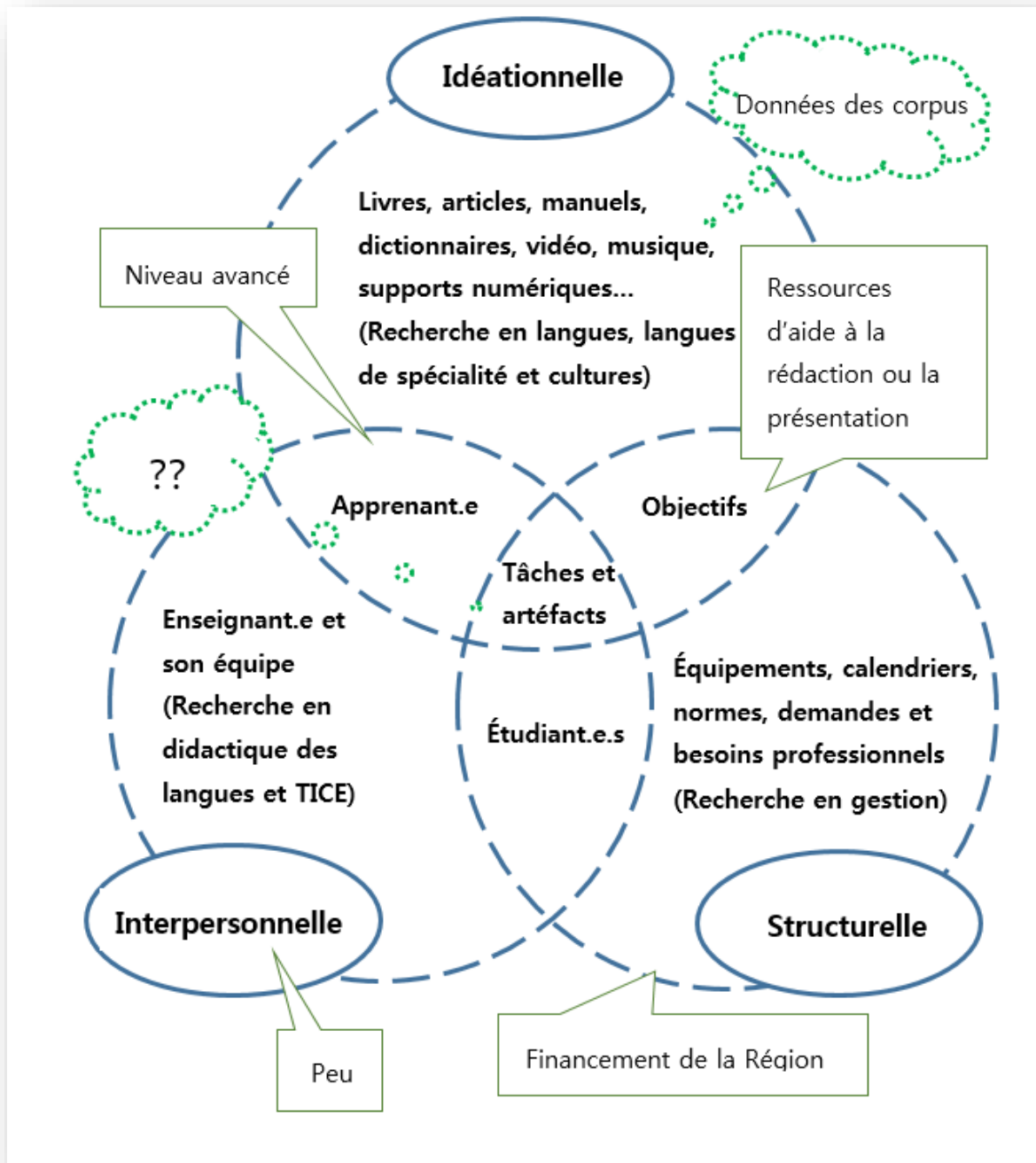


Figure 12 : Modèle du dispositif Corpus4All

Un autre point critique est le peu d'interactivité interpersonnelle (cf. 2.1.2.1). Pour cela, j'ai cherché une compensation sur le plan esthétique afin de motiver les utilisateurs/trices en l'absence d'interaction sociale (cf. Figure 13). La proposition de Cavalla (2012, 2014) sous forme de carte heuristique à visée sémantique illustrant visuellement les collocatifs par catégorie grammaticale me semble particulièrement robuste (cf. Annexe IV). La plateforme Chamilo offre la possibilité de consulter les temps et horaires de connexion, les activités tentées et les résultats. Nous pourrions ainsi examiner les attirances, les démarches, les écueils et les décrochements.

The screenshot displays the Chamilo Learning & Collaboration Software interface. At the top, there is a navigation bar with 'Homepage', 'My courses', and 'Blog enseignant'. Below this, a search bar contains 'Corpus for all' and a 'Teacher View' button. The main content area is titled 'Advanced activities for preparing oral presentations and written documents in English'. It features three distinct sections, each with a list of activities and a small image:

- Structuring a sentence:** Includes tasks like 'Further meanings of further', 'Prepositions of time: within and by', 'Introducing conditions: given that and provided that', 'that as a substitute for the one', 'Reporting that as a subordinating conjunction', 'This or That?', and 'Frequent confusion: actually or currency'. Accompanied by an image of a modern building.
- Describing data:** Includes tasks like 'Specifying facts: given', 'Resulting in / from and as an adjective', 'Reality or comparison: as or like?', 'Oral presentation: let's focus on see and look', and 'Conflicting fact: that while'. Accompanied by an image of a red flower.
- Offering opinions:** Includes tasks like 'Significantly emphasised', 'Introducing cause: due to', 'Reporting results or stating a hypothesis: if', 'Levels of possibility: may or might', 'Possibility or past ability: could', and 'Citing a source: par and as par'. Accompanied by an image of a microscope.

At the bottom of the page, there are logos for various institutions: ALP-STI, Université Grenoble Alpes, Rhône-Alpes, UNR-RA, and SIM SU. A footer note states: 'Réalisé dans le cadre de l'UNRRA avec le soutien financier de la Région Rhône-Alpes.' Below the logos, there are icons for 'Course description', 'Offering opinions', 'Corpus and writing resources', 'Structuring a sentence', and 'Describing data'.

Figure 13 : Saisie d'écran du projet UNR-RA Corpus4All

Dans cette section, nous avons vu l'intérêt de travailler à partir de modèles visuels afin de mettre en œuvre une analyse holistique des environnements d'apprentissage à contexte, objectif et organisation variables. Cette démarche heuristique est essentielle, car, selon Demaizière et Narcy-Combes, J.-P. (2005) :

La théorie conduit au choix de la technologie. Les possibilités offertes par les TIC peuvent être précieuses. Encore conviendra-t-il de relier l'introduction à des hypothèses et des références théoriques et méthodologiques mûrement réfléchies. On pourra peut-être alors rêver à un passage harmonieux de la théorie à la pratique recherche-action et TIC offrant à l'apprenant un environnement d'apprentissage riche et cohérent (Demaizière & Narcy-Combes, J.-P., 2005 : en ligne).

Le modèle que je propose permet d'identifier les influences premières et les points de tension dans l'objectif d'y remédier. Au centre, la focalisation étant mise sur la production des apprenant.e.s, les tâches et les artefacts représentent l'œil du cyclone idéationnel, interpersonnel et structurel. Dans le champ disciplinaire de la didactique des langues, le terme *retrofitting* évoque l'identification de points de faiblesse détectables lors de l'utilisation et de la rénovation mise en place dans un objectif d'amélioration (Fulcher & Davidson, 2009). Il se distribue dans trois grandes sphères d'influence : l'idéationnelle, où la recherche en phraséologie devrait trouver sa place ; l'interpersonnelle, qui est fortement impactée par l'utilisation (ou non) des TICE et la sphère structurelle, qui influence considérablement l'enseignement des langues dans le cadre universitaire. Dans la prochaine section, nous examinons plus finement la sphère idéationnelle à travers un rappel historique de la place des listes lexicales dans l'enseignement de l'anglais académique.

2.2. RESSOURCES ET PROPOSITIONS DIDACTIQUES

2.2.1. LE LEXIQUE ACADÉMIQUE ET SCIENTIFIQUE

2.2.1.1. L'INTRODUCTION DE LISTES LEXICALES ACADÉMIQUES

L'acquisition d'un vocabulaire ciblé est essentielle aux compétences linguistiques et a fait l'objet de nombreuses recherches dont la liste emblématique en anglais général est le GSL (*General Service List*), établie par Michael West en 1953. Le GSL reprend les travaux précédents, dont *The Teacher's Word Book* de Thorndike (1921), *A Basic Writing*

Vocabulary de Horn (1926) et la liste de Palmer (1931), qui a introduit la notion de *headword*. Ces ouvrages ont nourri l'*Interim Report on Vocabulary Selection*, publié par l'équipe de Faucett en 1936 (Gilner, 2011). Chacune de ces listes comporte environ 2000 lexèmes qui couvrent entre 70% et 95% des mots-formes des textes.

La notion d'une liste spécifique au cadre académique est défendue par Coxhead qui établit une première AWL (*Academic Word List*) en 1998, renouvelée grâce à l'évolution technologique en 2012. Le corpus sur lequel travaille Coxhead est constitué de 414 textes de longueur variable tirés de 28 sujets de quatre grandes disciplines, dont les arts, le commerce, le droit et les sciences. Exclus de l'AWL sont tous les lexèmes du GSL, ainsi que tout lexème qui n'est pas présent au moins 10 fois dans chacune des 4 grandes disciplines, au moins dans 15 sujets et dont la fréquence dans le corpus est supérieure à 100, pour un total de 570 « familles de mots » ou lexèmes (Coxhead, 2000). Dans le Tableau 5, nous voyons que le GSL et l'AWL couvrent moins bien l'anglais des sciences que les trois autres disciplines, le droit est mieux couvert par le GSL et le commerce par l'AWL.

Tableau 5 : Taux de couverture des 4 sous-corpus par le GSL et l'AWL en pourcentages (Coxhead, 2000 : 224)

Sous-corpus par discipline	AWL	Lemmes 1-1000 du GSL	Lemmes 1001-2000 du GSL	Total
Arts	9,3	73,0	4,4	86,7
Commerce	12,0	71,6	5,2	88,8
Droit	9,4	75,0	4,1	88,5
Sciences	9,1	66,7	5,0	80,8

Ceci veut dire que les lexèmes des textes scientifiques sont moins représentés par le GSL (71,7%, comparé à 76,8%-79,1% pour les autres disciplines), mais aussi, à un moindre degré, par le complémentaire AWL (9,1%, comparé à 9,3%-12% pour les autres disciplines). L'AWL présente les lexèmes par ordre alphabétique, avec un codage par fréquence. En réponse à ce manque de représentation en sciences, Coxhead et Hirsch

(2007) ont composé un corpus de textes issus de quatorze domaines scientifiques (1 761 380 mots) leur permettant d'identifier 318 lexèmes, absents du GSL et de l'AWL et couvrant 3,79% du corpus. Il est intéressant de noter que cette nouvelle liste pilote ne couvre que 0,61% du sous-corpus « arts », 0,54% du sous corpus « commerce » et 0,34% du sous-corpus « droit » (cf. Tableau 5). Nous voyons que les études de corpus permettent de cibler le vocabulaire de base nécessaire à des compétences linguistiques dans un cadre académique général ou spécialisé à condition que le corpus soit d'une taille conséquente et composé de documents récents et en rapport avec le cadre ciblé.

2.2.1.2. LA RÉPRÉSENTATION DES SCIENCES

Le manuel *Minimum Competence in Scientific English* (Blattes et al., 2003) a été conçu dans cette approche par le lexique de haute fréquence ciblé. Le lexique du GSL et le lexique de l'anglais général ont été d'abord retirés de la liste CEL (*Cambridge English Lexicon*), créée par Hindmarsh en 1980. Ensuite, un corpus papier d'une cinquantaine d'articles scientifiques a été analysé dans l'objectif de cibler le lexique scientifique de base nécessaire à des étudiant.e.s se préparant à une poursuite des études en anglais après le DEUG ou à une carrière professionnelle en sciences (Upjohn, 1999). Le manuel comporte 782 lexèmes, qui sont organisés en douze chapitres : *measurement, frequency, comparison, modification, link words, time (present and past), cause and consequence, hypothesis, modality, purpose and process*, puis repris dans le glossaire, chaque entrée étant composée d'une transcription phonologique, d'un exemple d'utilisation par catégorie grammaticale, de synonymes et parfois d'antonymes. Ces activités permettent à la fois de revenir au sens, élément essentiel à l'apprentissage, mais aussi de développer les potentialités linguistiques d'un lemme.

2.2.1.3. RÉPONDRE AUX BESOINS LEXICAUX DES ÉTUDIANT-ES

Travailler avec Jonathan Upjohn autour de la mise en place de la version sonore d'*Enigma Plus* et explorer les pratiques des utilisateurs/trices à partir des données récoltées de manière informatique, m'a permis de saisir l'importance des besoins lexicaux, notamment des étudiants en Licence scientifique (cf. Annexe IV). Par la suite, j'ai éprouvé le besoin de connaître les ressources disponibles, notamment en ce qui

concerne l'utilisation de ces lexèmes en contexte [HART11b, HART13a], piste que j'ai suivie au sein du Laboratoire LIDILEM.

L'équipe du laboratoire LIDILEM travaille sur le discours académique et universitaire depuis de nombreuses années (Boch & Rinck, 2010 ; Tutin, 2007 ; Tutin & Grossmann, 2014). Dans ce prolongement, Marie-Paule Jacques et moi-même avons bénéficié en 2012 d'un financement institutionnel autour d'un projet « Structure discursive de l'article de recherche comparative français/anglais sciences humaines et sociales/sciences exactes. L'objectif fut d'analyser les qualités discursives, énonciatives et textuelles de l'article de recherche à des fins didactiques et dans la continuation du projet ANR « Scientext » (2007-2010) [HART12a, HART13b, HART14a]. Cette recherche a contribué à l'élaboration de l'activité « Scientext Enigma », dont le contenu, basé sur les difficultés linguistiques des apprenant.e.s, cible un niveau d'anglais plus avancé que celui visé par la plateforme *Enigma Plus* initiale. Dans ce cadre, j'ai aussi comparé les ressources numériques proposées pour l'anglais, dont quatre dictionnaires et trois outils de corpus particulièrement intéressants pour les apprenant.e.s de l'anglais des sciences, ceci afin d'avoir une meilleure vision des ressources lexicales existantes.

2.2.1.4. L'APPORT DES DICTIONNAIRES

Nous avons vu que les corpus ont modifié la rédaction des dictionnaires (*cf.* 1.2.1.2.). Diverses analyses de dictionnaires ont été effectuées dont celle de Moon (2008) autour des collocations de trois lemmes et celle d'Arnaud, Lewis et Maniez (2008) autour de la lexicalisation des séquences [Adj + N] (avec ou sans trait d'union) servant à modifier des noms. J'ai voulu savoir comment les éditeurs se sont adaptés aux corpus et plus largement à la transformation informatique. Une analyse de quatre dictionnaires en ligne, conçus pour un public universitaire, dont deux pour les apprenant.e.s non anglophones : le *Merriam-Webster Learner's Dictionary* (MWLD), l'*Oxford Advanced American Dictionary for Learners of English* (OADL), le *Merriam-Webster Dictionary* (MW) et l'*Oxford English Dictionary* (OED) présentent une gamme de ressources et d'activités d'apprentissage, parfois ludiques, comme indiqué dans le Tableau 6.

Tableau 6 : Comparaison de ressources disponibles sur quatre sites de dictionnaires

	MWLD	OADL	MW	OED
Définitions avec exemples	X	X	X	X
Dictionnaire de synonymes	X		X	
Prononciation audio	X	X	X	
Mot du jour	X		X	X
Quizz et jeux	X		X	X
Tendances et néologies		X	X	
Liste des 3000 lexèmes académiques	X	X		
Espagnol, anglais			X	X
Français, allemand, italien				X
Anglais médical			X	
Aide à la rédaction				X

Ces outils multimodaux mettent l'accent sur des aspects ludiques (jeux, tendances, « mot du jour ») et pratiques, et pour les deux dictionnaires à destination des apprenant.e.s, le MWLD et l'OADL, une liste de 3 000 lexèmes à connaître. Le MW propose un dispositif autour de l'anglais médical, domaine important en anglais de spécialité. Tous les dictionnaires proposent en plus de la définition, des exemples du mot ciblé en contexte, mais l'espace dédié reste limité. Nous pouvons y constater que les dictionnaires n'ont plus le même format traditionnel, bien qu'une définition et des exemples soient présentés dans chacun. Ce qui m'intéressait dans cette exploration était les différentes façons d'aborder l'acquisition lexicale. Le « mot du jour » est sans contexte, mais reste une activité ludique et peu chronophage, une section sur l'anglais médical, au contraire est fortement contextualisée. L'aide à la prononciation ainsi que l'aide à la rédaction répondent à des besoins réels, par contraste avec la description des néologismes, qui relève plutôt de la curiosité. Cette évolution correspond à une évolution technologique, mais aussi une prise en compte des envies, besoins et pratiques en anglais de la société.

2.2.1.5. L'APPORT DES CORPUS À DESTINATION DES NON LINGUISTES

Par contraste, les trois outils basés sur corpus, SKELL (*Sketch Engine for Language Learning*) (Baisa, 2014), StringNet (Wible & Tsao, 2010) et Springer Exemplar (www.springerexemplar.com) proposent essentiellement des séquences ou des motifs lexico-grammaticaux à partir d'un mot précis. SKELL propose 40 occurrences d'un lexème en contexte, les collocations et les mots similaires sous forme de nuage. La technologie de Sketch Engine a été initialement développée pour identifier des exemples authentiques pour une version électronique du dictionnaire anglais de Macmillan. Ses exemples d'anglais général sont choisis pour illustrer des patrons fréquents et ayant une bonne « lisibilité », à savoir sans recours à une terminologie compliquée dans un contexte limité (Kilgarriff, Husák, McAdam, Rundell & Rychlý, 2008). Faisant appel au BNC, StringNet permet de saisir un ou plusieurs mots ou catégories grammaticales. Les résultats sont présentés sous forme de listes d'expressions ou de « n-grams hybrides » pouvant être composée de mots-formes, de lexèmes, de catégorie grammaticales ou de catégories grammaticales ayant une flexion précise (telle que : *there [BE] no point in [V-ing]*) ou des mots « similaires ». Les résultats sont nettoyés afin de mettre en avant des réponses accessibles et non redondantes.

Enfin, le Springer Exemplar est basé sur 9 424 904 articles ou livres en libre accès dans les sciences, publiés par Springer. Il est possible de rechercher des chaînes de plusieurs mots, mais le résultat ne comporte aucune autre information lexico-grammaticale. Le site propose quelques informations concernant l'origine des textes. Il est possible de faire une recherche par sujet ou par publication. Les résultats comprennent essentiellement l'unité lexicale ou la séquence dans la phrase. Les résultats pour SKELL et StringNet sont plus intéressants au niveau de la technologie et de la capacité d'identifier des motifs, même pour un public de non-linguistes. En revanche, seul Springer Exemplar permet d'accéder à l'anglais scientifique. Ces ressources, dictionnaire, outil ou corpus, peuvent apporter une quantité importante d'information lexicale ou grammaticale, mais, à la différence d'un manuel, sans démarche pédagogique précise.

2.2.2. ASPECTS DIDACTIQUES

2.2.2.1. USAGE-BASED GRAMMAR

Il me semble important d'évoquer la proposition d'*Usage-based language learning* dans le cadre de l'exploration des pratiques des apprenant.e.s. Cette perspective de l'acquisition de langues (Ellis, O'Donnell & Römer, 2013 ; Ellis, O'Donnell & Römer, 2015) se base sur la théorie psycholinguistique selon laquelle l'être humain apprend des patrons grâce à un contact avec la langue et la construction de catégories et ceci selon la distribution d'usage. Cette distribution peut être identifiée par une analyse de très grand corpus. Ellis *et al.* (2013, 2015) prennent comme point de départ de leur démonstration vingt-trois constructions [Verbe + across + Nom] parmi les 700 *patterns* cités dans l'ouvrage *Verb Grammar Patterns* (Hunston & Francis, 1996). L'intérêt de cette approche, qui met l'accent sur les constructions syntaxiques, reste indéniable, mais je pense, néanmoins, que la complexité des grammaires pose des problèmes dans un cadre d'apprentissage auprès d'un public LANSAD.

L'approche basée sur l'utilisation (*usage-based grammar*) comporte des éléments intéressants, dont des analyses statistiques de l'usage de verbes dans une construction donnée. Nous pouvons néanmoins mettre en cause la conclusion concernant l'apport didactique. Les images graphiques sont particulièrement esthétiques, mais méritent d'être testées dans un cadre d'apprentissage avant de tirer des conclusions sur leur pertinence didactique. Dans mes recherches, les aspects de lisibilité et de pertinence pour un public d'apprenant.e.s sont des critères essentiels et j'adopte, dans ce sens, une approche lexicale.

Legallois (2014) propose aussi un apprentissage similaire par « la Grammaire de Construction », c'est-à-dire « les structures syntaxiques [qui] sont des unités linguistiques possédant une pertinence cognitive, à l'image du mot ». Pour la séquence [N1 V N2 de Inf.], il propose deux exemples tirés de l'Internet :

- 1- Par des conditions de reprise intenable, Mittal empêche le gouvernement de trouver des repreneurs sérieux et voue à l'échec les tentatives de sauvetage de Florange et de ses emplois.

- 2- Le lobby pharmaceutique a convaincu le gouvernement de remplacer tous les psys par des pilules (Legallois, 2014 : 42).

Mais, contrairement à Ellis (2013, 2015), il émet clairement l'idée qu'il faut évaluer, dans un contexte d'apprentissage, les capacités des apprenant.e.s à appréhender des aspects lexico-grammaticaux généraux. Ma position se rapproche de cette vision de Legallois, mais à cela je pense qu'il faut ajouter que, dans le cas de l'anglais de spécialité, il est essentiel de cerner les aspects lexico-grammaticaux pertinents et accessibles selon le contexte discursif donné. Pour reprendre l'expression de Deyrich (2001), il faut une « transposition didactique » des ressources documentaires de spécialité (avec par extension des corpus) en opérant des articulations stratégiques afin que des données linguistiques simplifiées puissent être appropriables et réinvesties par l'apprenant.e.

2.2.2.2. ÉQUILIBRE ENTRE CONTENU ET ACCESSIBILITÉ AU CORPUS

L'accès au corpus dans un contexte didactique peut s'opérer de trois façons : par l'analyse de corpus d'apprenant.e.s (cf. 1.1.3.), par un mode indirect via l'utilisation de dictionnaires ou autres ressources conçues à partir de corpus et enfin, par un accès direct ou ASC (Apprentissage sur corpus) (Boulton & Tyne, 2014). Dans leur méta-analyse de 116 études empiriques au sujet de l'ASC, Boulton et Cobb (2015) concluent, que, de manière générale, l'utilisation de corpus peut être efficace (*efficient* et *efficace*) en classe notamment pour les aspects lexico-grammaticaux, mais il resterait certaines lacunes dans la recherche sur l'enseignement des aspects discursifs, dont le positionnement et le genre. Dans ce sens, il serait intéressant, dans un deuxième temps, d'évaluer l'utilisation de ressources médiatisées du dispositif Corpus4All (cf. 2.1.3.6.), qui offrent des activités de niveau linguistique avancé en anglais des sciences.

La consultation de dictionnaires, souvent à base de données de corpus, reste accessible, bien que parfois sous-utilisée. Nous voyons apparaître des ressources avec un accès direct au corpus. Cependant, ces ressources sont moins adoptées dans les contextes d'apprentissage que la consultation de dictionnaires, un frein important étant un manque de connaissance de la part de beaucoup d'équipes enseignantes autour de l'utilisation de corpus, mais aussi la difficulté d'exposer une grande quantité de données complexes sous des formats réutilisables. Il faut aussi construire ou avoir accès à un


corpus dont le sujet et le genre correspondent à l'objectif d'apprentissage. Dans la création de ressources numériques, il faut trouver justement un équilibre entre ces trois éléments : accessibilité, complexité et pertinence.

2.2.3. LE CONTEXTE INITIAL DU PROJET ENIGMA PLUS


2.2.3.1. L'INTÉRÊT DE LA RECONSTRUCTION DE TEXTE

En 1991 fut publiée la première version du manuel *Minimum Competence in Scientific English* (cf. Blattes *et al.*, 2011) conçu à destination des étudiant.e.s universitaires en sciences ou en technologies ayant déjà des bases en anglais général et bâti sur un recensement de mots fréquents dans les domaines scientifiques (Upjohn, 1999). Le format pap-ebook, format fusionnant le manuel de base avec un accompagnement numérique important comprenant des tests lexicaux d'entrée et de sortie de chapitre, la lecture à voix haute de certains extraits du manuel et des activités de reconstruction de texte, est sorti vingt ans plus tard, en 2011. Cette démarche visant à intégrer l'oralité permet de consolider les connaissances lexico-grammaticales, mais surtout de mettre en pratique la compréhension de l'oral, précurseur de l'expression orale, compétences souvent problématiques pour une grande majorité des apprenant.e.s.

Dans les années quatre-vingt, Jonathan Upjohn du Pool des langues de l'UJF et l'ingénieur Kim Upjohn travaillaient déjà sur l'activité de reconstruction de texte, nommée *Enigma Plus*. La reconstruction de texte est une activité élaborée prenant une forme entre la dictée, le texte lacunaire et les jeux de mémoire, dans la construction d'un texte porteur de sens et à portée scientifique (cf. Figure 14). Construite à partir d'un squelette informatisé, elle est similaire à *Storyboard*, mis au point par John Higgins (1982), à partir de son programme *Rebuild*, inspiré du programme *Textbag* de Tim Johns. *Storyboard* et ses versions proches ont été publiés par Wida Software, puis sous une forme plus accessible par ESM Cambridge deux ans plus tard. De nombreuses autres versions, dans des langues différentes, ont vu le jour dans les années 1980 (Davies, 2007).



ENIGMA PLUS ...



Hypothesizing

The act of hypothesizing is a ce... no... of sc... re..., as can be seen in the Scientext corpus. Running a search query of this corpus shows that authors cl... si... th... **hypothesis** within a de... ar... Hence, the act of hypothesizing is li... to one or mo... pr... st... in the te..., using ex... such as th... or th... A that-clause introduces the authors' pr... As the **hypothesis** is not yet pr..., the that-clause often contains a mo... verb such as co... or ma...

One way of proposing a re... qu... is re... to te... the **hypothesis**. Al..., te... is the mo... frequent verb used in this context, other verbs include in..., ex..., ad..., ex..., or ve... Note that be... wj... an in... verb, for example, to te... or to in..., is another ac... way of presenting ho... the **hypothesis** is ac... ex...

In many cases, the re... va... the **hypothesis** or hy... The mo... frequent means of expressing this po... ou... is the verb su... Again, the subject of the sentence ma... va..., bu... frequent ones are fi..., da..., re..., or st... In other words, while the authors ma... te... the **hypothesis**, it is the ou... of the ex... that ma... su..., st... or fa... the **hypothesis**. Re... a nu... **hypothesis** is fr... found in English as in other languages.

Finally, re... ma... di... a **hypothesis**. In this ca..., the subject of the sentence can be the authors or the fi..., re... or st... Once again, the verb su... is found: the re... of this st... do not su... the **hypothesis**. Notice al... that a that-clause is not always used in this context. We st... su... consulting other re... ar... or an on-line corpus to fi... other ex... of au... language in use. Th... yo...

Type in your word :

Return to menu

1. Proposing a hypothesis

2. Data on findings in a corpus

3. Examples of hypotheses

4. Examples of hypotheses

5. Examples of hypotheses

6. Examples of hypotheses

Figure 14 : Saisie d'écran de Scientext Enigma Plus

Sa spécificité est une réponse instantanée à l'utilisateur/trice après chaque entrée de mot, une interactivité pour ainsi dire impossible à l'échelle de la classe ou sur un support papier. Entre 1994 et 2002 une version sans son du logiciel *Enigma Plus* a été vendue à une quinzaine d'instituts d'enseignement supérieur en France, Allemagne et Suisse dont le Pôle Universitaire Européen à Strasbourg, l'École Suisse d'ingénieurs à Lausanne et les Écoles de Mines de Nantes et de St Étienne.

2.2.3.2. PROJET DE L'ÉTABLISSEMENT

Dans le cadre de la politique des langues du contrat quadriennal 2006-2010 de l'UJF, le Service des langues m'a confié la responsabilité de piloter la mise en place de supports audio, dont la création d'enregistrements d'extraits du manuel et plus spécifiquement le renouvellement et la mise en ligne d'*Enigma Plus* en coordination avec J. Upjohn et K. Upjohn en complément du manuel *Minimum Competence in Scientific English* (Blattes et al., 2003).

2.2.3.3. ANALYSE DES PRATIQUES

Il n'est pas évident que l'apprenant.e exécute un programme informatique de la manière imaginée par les concepteurs. La question de la façon dont sont utilisés ces dispositifs, en l'absence d'enseignant.e qui motive l'apprenant.e ou l'oblige à y travailler, fut la source de questions que j'ai traitées dans plusieurs publications. J'ai observé, à partir des résultats sur *Enigma Plus*, qu'il y a un travail initialement linéaire des utilisateurs/trices dans le choix des textes et le remplissage du squelette, que les supports visuels impactent la rétention du lexique. J'ai aussi constaté qu'il y a des différences très claires de choix de texte, d'horaire et de durée de travail selon le genre, que l'interaction de plusieurs facteurs influence la poursuite ou l'arrêt du travail et enfin, que les utilisateurs/trices travaillent de manière appliquée, même en dehors de la présence d'enseignant.e [HART10c, HART10d].

2.2.3.4. AJUSTER LES PARAMÈTRES EN FONCTION DES PRATIQUES

Cazade (2012a, 2012b, pas de pagination) suggère qu'*Enigma Plus*, un accompagnement numérique de *Minimum Competence in Scientific English* (Blattes et al., 2003), pourrait bénéficier « d'aide au moment du traitement de l'erreur commise par l'utilisateur ». À

mon sens, cette proposition ne prend pas suffisamment en compte le corrigé qui est proposé en fin d'exercice, ni les véritables pratiques d'utilisation, dont l'analyse des résultats montre de multiples tentatives de trouver un mot difficile. La différence entre un contexte en classe et ce contexte en ligne est qu'avec *Enigma Plus*, l'utilisateur/trice sait instantanément si sa réponse est correcte ou incorrecte. Il est envisageable d'offrir un commentaire plus spécifique selon le type de mot, par exemple de vérifier s'il y a interférence de la langue française sur l'orthographe (*gaz/gas*) ou bien d'offrir des rappels simples de la conjugaison anglaise, mais cela risque d'avoir un impact négatif sur le facteur temps, qui n'est pas extensible.

Le nombre relativement faible d'utilisateurs/trices qui terminent complètement chaque exercice est un signe de difficulté. En effet, indépendamment de la quantité d'erreurs, la durée semble être le facteur déclenchant la fin de l'activité. Les mots en début de texte, les mots les plus faciles et les mots répétés sur les supports visuels sont privilégiés. Lors de la création de l'activité « Scientext Enigma », je me suis appuyée sur ces données afin de rationaliser la quantité de réponses à trouver et de rationaliser la présence de lexique pertinent intégré aux supports visuels. J'ai conçu « Scientext Enigma » à partir des données linguistiques tirées du corpus Scientext, ainsi que de ma recherche en didactique des langues, notamment autour de la pédagogie innovante, dans l'objectif de répondre aux lacunes des apprenant.e.s constatées en classe. À l'avenir, une étude sur corpus d'apprenant.e.s permettrait de vérifier ces choix et l'acquisition des éléments.

2.2.4. ANALYSE DES PRATIQUES EN ENVIRONNEMENT MÉDIATISÉ

2.2.4.1. PRATIQUES INDIVIDUALISÉES

J'ai pu percevoir, lors de l'analyse des réponses sur Enigma, que les pratiques des apprenant.e.s sont fortement individuelles. Je l'ai encore remarqué lors de la conception d'un nouveau parcours en ligne pour les étudiants en deuxième année de master, l'EBCM, que j'ai piloté en collaboration avec Emmanuelle Esperança-Rodier et Richard Kennedy (cf. 2.1.3.5.). En regardant les données de 224 étudiant.e.s ayant consulté le

site pendant au moins 30 minutes lors du premier semestre⁶, nous découvrons un temps de connexion moyen de 6 heures, mais 42 étudiant.e.s (18,8%) ont passé plus de 10 heures. Il y a donc une forte diversité des pratiques au niveau de l'assiduité. Une dizaine d'étudiant.e.s a commencé dès le 24 octobre, alors que deux autres attendront le 23 novembre, mais tous les jours de la semaine sont fréquentés. La constante est que les supports sur lesquels j'ai travaillé sont réellement consultés dès le matin et très tard le soir, et ceci tous les jours de la semaine, car les pratiques des apprenant.e.s sont loin d'être homogènes sur ce point de planification.

L'autre constante est que bon nombre d'utilisateurs/trices dépassent le strict minimum d'activités proposées, bien que cela soit optionnel. L'observation selon laquelle pratiques en autoformation sont individualisées sur *Enigma Plus* est confirmée par les résultats des pratiques sur l'EBCM. Le Tableau 7 montre les ratios de personnes, selon leur niveau, ayant complété trois exercices de l'EBCM : *Enigma Plus* (environ 20 minutes de travail en moyenne plus le temps de connexion), un exercice de grammaire autour du temps présent spécifiquement conçu pour la préparation à une présentation orale (20 questions à choix multiple) et enfin, une série de questions qui nécessite la consultation d'une fiche SUP (Service Universitaire de Pédagogie) au sujet de la prise de notes (cf. 2.3.3.).

Nous pouvons noter la présence de deux groupes établis par niveau : un groupe faible ayant obtenu un niveau A1-A2 du CECRL ($n = 159$) et au groupe ayant été évalué au niveau B1 du CECRL ($n = 190$). Il est indiqué au groupe des étudiant.e.s ayant obtenu un niveau A1-A2 que l'EBCM est obligatoire dans les programmes du master, mais au groupe ayant été évalué au niveau B1 il est indiqué que l'EBCM est seulement recommandé et devient obligatoire uniquement en cas de non-validation des

⁶ L'utilisateur ou l'utilisatrice est automatiquement déconnecté.e de la plateforme en l'absence d'activité.

compétences en fin de semestre. Les chiffres ayant été recueillis avant que les étudiant.e.s passent les épreuves de validation, les exercices étaient donc encore optionnels pour ce groupe.

Tableau 7 : Assiduité des tâches de l'EBCM

	EBCM obligatoire : Pourcentage des étudiant.e.s de niveau A1-A2 ayant complété l'exercice	EBCM optionnel : Pourcentage des étudiant.e.s de niveau B1 ayant complété l'exercice
Enigma	86%	63%
Grammaire	100%	40%
SUP Prise de notes	77%	18%

Il est possible que des étudiant.e.s aient téléchargé le document SUP au sujet de la prise de notes sans répondre aux questions, ce qui pourrait expliquer le chiffre assez bas de 18% pour le groupe B1. Ce que nous voyons, c'est que les étudiant.e.s n'ayant pas d'obligation directe vont sélectionner les activités qu'ils ou elles souhaitent en fonction de leurs intérêts, et pas uniquement en fonction du temps, car l'activité la plus chronophage est *Enigma Plus*.

2.2.4.2. PRENDRE EN COMPTE L'ASPECT SOCIAL

L'approche socio-constructiviste du langage met l'accent sur la notion que le langage est acquis dans un contexte social, ce qui semble être, au premier abord, en contradiction avec les dispositifs d'apprentissage numériques. Les deux activités, *Enigma Plus* et l'EBCM, sont prévues en accompagnement de la formation en présentiel et donc l'aspect social n'est pas complètement absent. La contrainte de temps est forte, et le peu de

temps consacré à l'apprentissage doit être efficace et complémentaire sur le plan lexical, et notamment apporter des informations phonologiques.

En effet, des dispositifs de compréhension de l'oral permettent à l'apprenant de ralentir la tâche par un recours aux répétitions, ce qui n'est pas toujours possible en classe. Nous voyons que les apprenant.e.s abordent ces deux activités de leur propre manière, choisissant le moment, la durée et parfois les thèmes. Essentiellement tournées vers la compréhension de l'oral ou pour l'EBCM vers un rappel des notions prérequisées, elles offrent un travail en autonomie, sans remplacer l'interaction et la créativité qui se produisent ou devraient se produire en classe.

2.3. ÉVALUATION DES COMPÉTENCES

2.3.1. LA PRESSION DE LA CERTIFICATION

2.3.1.1. ADAPTION AUX BESOINS PROFESSIONNELS

Dans les écoles d'ingénieur en France, il y a recours au certificat TOEIC (Test of English for International Communication), (cf. 2.3.4.1.) pour valider le niveau B2 du CECRL en fin de diplôme, suite aux recommandations de l'Europe. Cette décision politique a un impact important (*washback*) sur les formations d'anglais pour un public LANSAD, puisque l'apprenant.e doit maîtriser l'anglais des affaires et le format très structuré autour des questions à choix multiples énoncées dans un rythme soutenu.

L'équipe du Service des Langues de l'UJF a fortement rejeté le recours aux certificats externes, comme le TOEIC, qui ne correspondent pas aux futurs besoins professionnels des étudiant.e.s et dont l'adoption impliquerait nécessairement un éloignement des objectifs linguistiques en anglais des sciences. Seule une aide à l'obtention d'une certification en langues dans des cas précis, dont la poursuite d'études ou de stage en pays anglophone, est soutenue.

2.3.1.2. PROPOSITION D'ÉVALUATION SUR MESURE

Le CEVU (Conseil des études et de la vie universitaire) de l'UJF a voté en 2009 la mise en place d'un dispositif de validation du niveau B2/Niveau indépendant du CECRL en réponse aux exigences européennes en matière de délivrance de diplômes de master. Ce dispositif de validation en anglais de spécialité se divise en trois axes : compréhension de l'écrit, expression lors d'une présentation orale et prise de notes à partir de sources orales (cf. 2.1.1.5.). L'expression écrite se manifeste en classe, à travers les supports visuels des présentations et des activités de prise de notes, mais sans faire l'objet d'une évaluation spécifique.

Mes expériences comme enseignante certifiée affectée dans l'enseignement supérieur assurant une charge importante de formation, à côté de la responsabilité de coordonner l'enseignement des langues à UJF Valence et enfin, comme Directrice du Service des Langues à l'UJF m'ont conduite à une forte responsabilité dans l'évaluation des compétences linguistiques des apprenant.e.s. En parallèle, mes étudiants en licence à l'UJF Valence devaient obtenir un certificat en langues dans le cadre de l'entrée à l'IUFM/ESPE (Institut Universitaire de la Formation des Maîtres/École Supérieur du Professorat et de l'Éducation) ou à la titularisation en tant que professeur des écoles. L'accompagnement de ces deux populations m'a amenée à examiner les propositions d'évaluation des compétences linguistiques et notamment de la compréhension de l'oral.

2.3.2. LA COMPRÉHENSION DE DOCUMENTS ORAUX

2.3.2.1. ADÉQUATION DU CONTENU ET DE LA POPULATION

Mes recherches et travaux autour d'*Enigma Plus*, proposé aux étudiant.e.s en licence, m'ont apporté une première vision concernant la compréhension de l'oral. Le niveau scientifique du contenu d'*Enigma Plus* est trop faible pour motiver une population en master et les sujets traités ne correspondent pas à la préparation d'un certificat CLES2 (Certificat en Langues de l'Enseignement Supérieur), démarche entreprise par un grand nombre d'étudiant.e.s souhaitant intégrer l'ESPE. Dans ces deux populations, en master

comme pour la préparation au CLES, *Enigma Plus* n'est pas adapté et je souhaitais explorer plus largement la compréhension de l'oral.

Ici, j'exposerai d'abord mes recherches en compréhension de l'oral, puis j'évoquerai la prise de notes, thèmes qui ont nourri la mise en place du dispositif de validation du niveau B2/Indépendant au Service des Langues voté par le CEVU à l'UJF en 2009 [HART11c, HART13a]. L'équipe de l'UJF n'avait pas les mêmes lacunes en ce qui concerne la préparation et la mise en place d'épreuves de compréhension de l'écrit ou de la présentation orale, car ces épreuves se situent dans le cadre d'une réflexion internationale beaucoup plus avancée.

2.3.2.2. *UNE APPROCHE PAR LES TÂCHES*

L'évaluation de la compréhension de l'oral nécessite l'accomplissement d'une tâche dans un contexte précis sur un support donné, au moyen d'une autre compétence, le plus souvent l'écrit (*cf.* 1.2.4.1.). Bachman (2002) propose que la conceptualisation d'évaluation basée à partir de tâches doit établir une distinction entre les caractéristiques de la tâche, les attributs de la personne et l'interaction entre ces deux facteurs.

Vandergrift (2007) a effectué une synthèse d'études autour de la compréhension de l'oral et a conclu que la recherche actuelle n'a pas encore déterminé comment les apprenants traitent simultanément des informations à la fois visuelles et sonores dans les environnements multimédias. De même, Wagner et Meros (2010) constatent les résultats contradictoires des différentes études concernant l'influence des supports audio ou audio-vidéo lors des épreuves de compréhension de l'oral. Son étude sur 122 apprenant.e.s adultes a montré une augmentation significative de la performance lorsque des supports vidéo ont été utilisés. Jensen et Hansen (1995) postulent que les caractéristiques du discours et le style de présentation peuvent affecter la compréhension détaillée (micro) et globale (macro). De plus, ils ont trouvé qu'une connaissance préalable du sujet abordé a un impact important sur les résultats d'une épreuve de compréhension, mais uniquement lorsqu'il s'agit de sujets techniques, ce qui est souvent le cas dans les documents oraux que nous utilisons lors des épreuves à

l'UJF. Par conséquent, il semble y avoir une série de facteurs qui influent sur la compréhension de l'oral dans des contextes multimédias, et donc toute conception d'évaluation de la compréhension doit prendre ces facteurs en considération [HART07a, HART07b, HART09a, HART10a, HART10b, HART10c, HART10d, HART10e, HART11b].

2.3.2.3. L'AUTENTICITÉ DES TÂCHES

Dans le cadre du dispositif de validation du niveau indépendant/B2 voté par le CEVU à l'UJF en 2009, l'équipe enseignante souhaitait maintenir l'utilisation de documents correspondant à chaque programme disciplinaire, à la fois pour encourager la motivation des étudiant.e.s et pour éliminer un biais disciplinaire. Cela suppose des supports différents par discipline, mais un mode d'évaluation commun, l'objectif étant de trouver un dispositif qui étaye l'apprentissage de pratiques proches des futures expériences professionnelles, comme une réunion de travail ou une conférence en langue anglaise. Le format habituel d'une évaluation demande de répondre à une série de questions écrites pendant l'écoute. Or, cela n'est que rarement une pratique de la vie courante. La prise de notes lors de l'écoute préalable à la restitution du contenu se rapproche davantage d'une situation « authentique », c'est-à-dire qui imite la performance lors de contextes de la vie réelle (concernant la notion d'authenticité d'une évaluation, voir McNamara, 2000 ou Bachman, 2002). Ceci va dans le même sens que l'étude de Song (2011) selon laquelle la qualité des notes, évaluée à partir de leur contenu et de leur organisation, est un indicateur positif de la compétence en compréhension de l'oral chez les étudiant.e.s en master.

La compréhension de l'oral est sans doute la compétence la moins étudiée, car en fait, elle reste invisible. La compréhension et l'expression écrite, à forte valeur académique et sociale, sont aussi étudiées dans la langue maternelle. Pourtant, pour l'apprenant.e non natif, elle pose des problèmes particuliers, notamment parce qu'elle se trouve souvent combinée à d'autres moyens de communication, comme les supports visuels. L'évaluation de la compréhension passe presque toujours par des moyens écrits, dont les modalités doivent être prises en compte lors de la conception des épreuves.

2.3.3. LA PRISE DE NOTES

2.3.3.1. DEMANDES COGNITIVES

La prise de notes en langue maternelle fait appel à plusieurs compétences, y compris une attention prolongée, la capacité d'identifier des concepts complexes, des arguments secondaires et les détails, la capacité de prêter attention simultanément aux visuels, à la voix et la capacité d'organiser et de trier les informations, ainsi que de les réutiliser à une date ultérieure (Boch & Piolat, 2004 ; Piolat & Boch 2005 ; Salehzadeh, 2006). La prise de notes est peut-être encore plus difficile pour les apprenant.e.s d'une langue étrangère qui peuvent avoir des difficultés avec une ou plusieurs des compétences linguistiques requises (Buck, 2001 ; Chaudron, 1995, Flowerdew, 1994 ; Piolat, 2003) [HART09a, HART10a].

2.3.3.2. GRILLE D'ÉVALUATION DANS UN CONTEXTE HÉTÉROGÈNE

À partir de l'observation des pratiques de l'équipe enseignante et des fruits de la recherche, j'ai proposé une grille d'évaluation de la performance en prise de notes pendant l'unité d'enseignement d'anglais des sciences, qui se déroule sur un semestre (cf. Figure 15). Au sommet de la grille est évoquée succinctement la complexité croissante des besoins cognitifs nécessaires pour la prise de notes, à commencer par la compréhension de l'oral, puis la sélection, l'organisation et la réaction aux informations et aux concepts présentés. Les colonnes verticales de la grille représentent les niveaux successifs de compétences telles que désignées par les descripteurs du CECRL ou de l'ALTE (*Association of Language Testers in Europe*).

Dans cette grille, la première rangée horizontale est destinée à une tâche en conditions d'examen, dont l'écoute d'un document audio ou le visionnement d'un document audiovisuel trois fois avec prise de notes suivi d'une série de questions de compréhension et de commentaires sur le contenu. La deuxième rangée destinée au contrôle continu étalé sur le semestre comporte des traces écrites des activités du cours et au moins six heures de travail en autonomie à partir de documents, notamment audio ou audiovisuels (par exemple, *Enigma Plus* pour les plus faibles ou *TED Talks* pour les plus avancés) et se termine par les notes prises lors des présentations orales des autres

étudiant.e.s du groupe. Ce contrôle continu est rendu sous forme de cahier de bord rédigé tout au long du semestre. Il inclut une introduction, parfois une conclusion et le glossaire.

Octobre 2015

Marking grid for Master's degree note-taking (See CEF on verso for note-taking skills)

Student number _____ Last name _____ First name _____
 UFR _____

	“survie” A1/A2 <i>Comprehension</i>	“seuil” B1 <i>Selection</i>	“independent” B2 <i>Organization</i>	“autonome” C1 <i>Reaction</i>	“maîtrise” C2 <i>In depth understanding</i>	Total Points
	Basic information. Relies heavily on visual aids. Very simple summary from an outline or direct questions. Errors may hinder comprehension.	Records key or important points when more or less dictated. Can write a basic summary of main ideas. Some errors in the association or importance of the ideas. Structural and lexical errors do not hinder comprehension.	Misses some facts. Rare structural and lexical errors. Relates work to personal objectives. Rich vocabulary. Gives basic opinion. Notices points of personal interest.	Detailed notes useful for other people. Notes and summarizes clearly and in detail. Gives opinions. Main ideas clearly given priority. Can reformulate with own structural and vocabulary elements.	Complex summaries relating different sources and objectives. Nuances, allusions and implicit facts included. Elegant use of vocabulary, idioms and complex structures. No errors.	
In-class note-taking from oral document(s).	1 - 2	4	5	7	8	In-class note-taking. /8
Class notes, follow up & from student presentations.	1 - 2	4	5	7	8	Class notes, follow up, and from student presentations. /8
Introduction/Conclusion 0 - ½ - 1	Glossary 1 - 2 - 3		Organisation and Objectives /4		Total	/20

Remarks:

Validation of B2-level note-taking skills : Can understand a clearly structured lecture on a familiar subject, and can take notes on points which strike him/her as important, even though he/she tends to concentrate on the words themselves and therefore to miss some information. (CEFRL)

Tick one:

Pas de validation

Passage en commission (9.5-9.99)

Validation (10+) or Validation by commission

Creative Commons : Attribution - Pas d'utilisation commerciale
 Citez : UGA S4L - ALP Sciences et Technologies.

Figure 15 : Grille d'évaluation pour la prise de notes en Master2 en anglais de spécialité

Le point unique accordé pour l'introduction/conclusion visait à encourager la réflexion personnelle sur ses propres compétences en prise de notes. L'équipe enseignante a convenu que l'acquisition d'un vocabulaire spécifique au domaine était indispensable à la maîtrise de la langue et devrait être incluse dans la grille. Après délibération sur l'importance de l'acquisition de la terminologie et les pratiques de l'équipe enseignante en matière de notation, trois points ont été accordés à cet aspect. Nous retrouvons dans cette décision l'importance de l'acquisition de lexiques académiques et disciplinaires (cf. 2.2.1.). La grille a été modifiée suite à des réunions successives avec l'équipe permanente sur une période de deux ans. L'équilibre délicat du volume de texte et de chiffres à cocher, l'exhaustivité des descripteurs, et un espace pour les commentaires furent discutés. L'affectation de points pour chaque niveau a été rééquilibrée afin que le résultat global reflète les pratiques de notation déjà ancrées au sein de l'équipe. En d'autres termes, j'ai adapté la feuille de notation aux pratiques de l'équipe et aux critères du CECRL, sans les imposer. En parallèle, j'ai orchestré l'établissement de la grille d'évaluation de l'épreuve de présentation orale en collaboration avec l'équipe enseignante et grâce à une revue de la littérature (cf. Annexe VI).

2.3.3.3. *CORRÉLATION DES RÉSULTATS*

Afin de vérifier la validité d'un dispositif d'évaluation, nous pouvons comparer les scores obtenus à partir des grilles d'évaluation avec les résultats du test de positionnement OOPT (Oxford Online Placement Test) passé en début de semestre en laboratoire de langues sous surveillance. La population était en deuxième année de master dans les disciplines d'activités physiques et sportives, astrophysique, biologie, chimie, informatique, génie mécanique, géographie, mathématiques, physique ou pharmacie. Sont exclues des résultats les personnes ayant manqué l'OOPT pour des raisons de calendrier ou qui n'ont pas assisté au cours d'anglais. Nous voyons dans le Tableau 8, une corrélation étroite entre le niveau du CECRL ($n = 1\ 007$) selon l'OOPT et la note obtenue lors des épreuves de prise de notes et évaluée à travers le filtre de la grille d'évaluation. Une analyse ANOVA montre une différence significative pour les 6 niveaux $F(5, 1000) = 41.67, p < 0.0001$. Ces résultats chiffrés suggèrent une excellente validité

du dispositif, car il y a une corrélation des résultats des deux types d'évaluation. Par ailleurs, l'évaluation de prise de notes tente de répliquer un contexte authentique.

Tableau 8 : Scores d'un test normalisé (OOPT) et la note en prise de notes

Niveau du CECRL selon l'OOPT	Quantité d'étudiant.e.s	Moyenne de la note	Écart
A1 (niveau débutant)	46	11.10	8.09
A2	277	11.51	5.70
B1	405	12.69	3.77
B2	213	13.89	4.89
C1	47	13.93	5.91
C2 (niveau avancé)	19	15.36	6.59

2.3.3.4. ÉTAYAGE MÉDIATISÉ

En parallèle, j'ai conçu une plateforme d'aide à la prise de notes, Notable Science, publiée sur le site du Service des Langues (*cf.* Figure 16). Une brève auto-évaluation comportant 20 questions permet à l'étudiant.e de réfléchir à ses stratégies de prise de notes. Lors de la rédaction du présent dossier de synthèse, j'ai consulté les réponses récoltées en 2014-2015. Quarante-cinq étudiant.e.s de l'UJF (sur une population d'environ 750) ont complété l'auto-évaluation en dehors du temps de classe entre le 9 septembre et le 26 mai, dont 53 (62,3%) en décembre 2014, période d'examen pour la plupart. Chaque question à choix multiple vaut 10 points, avec des réponses allant de 0 (jamais/non) à 10 (oui/toujours/souvent).



Figure 16 : Saisie d'écran du dispositif numérique d'aide à la prise de notes

Les résultats indiquent que la vaste majorité des étudiant.e.s qui ont pris l'initiative de s'autoévaluer estime savoir produire une carte de concepts (schéma conceptuel ou *mind maps*) (moyenne = 9,5/10), mais seule la moitié sait se servir du système Cornell de prise de notes (moyenne = 5,7/10). Cela peut indiquer une préférence culturelle des étudiant.e.s ou de l'équipe enseignante, car les deux méthodes sont explicitées sur le

site Notable Science (cf. Figure 17). Selon leurs réponses, la plupart des participant.e.s de cette population disent prêter attention aux mots clés (moyenne = 9,1/10) et aux marqueurs linguistiques (*to begin...*) (moyenne = 8,4/10) dans leur recherche du sens.

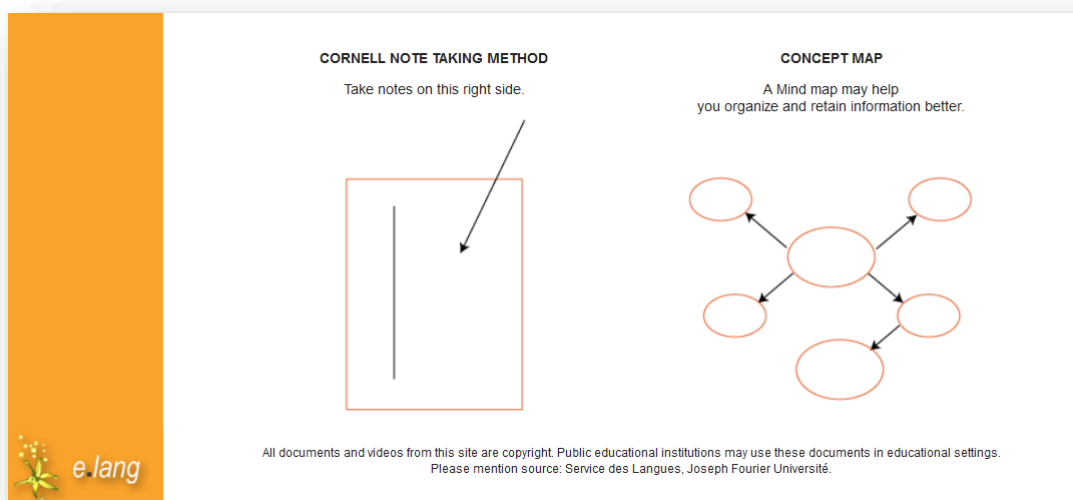


Figure 17 : Extrait de *Notable Science* : Carte de concepts

Moins présent est le recours aux stratégies transférables de notation rapide, telle que l'utilisation d'abréviations (e.g., i.e., dvpt, tech, csq, pb, t°) (moyenne = 7,7/10) ou de symboles (=, +, !) (moyenne = 8,1/10). Les pratiques souvent utiles, comme la relecture des notes dans les 24 heures (moyenne = 6,5/10) ou la comparaison occasionnelle des notes avec celles d'une autre personne (moyenne = 6,7/10) laissent penser qu'il s'agit d'un exercice personnel et souvent réservé au moment de la prise de notes et aux révisions juste avant l'examen. Ces résultats sont encourageants, car l'identification des mots clés et même des marqueurs linguistiques semblent être une stratégie adoptée par la population concernée. Cependant, l'utilisation d'une gamme d'abréviations et de symboles est corrélée avec une meilleure rétention des informations [HART07a, HART09a] et mériterait plus d'attention.

Afin de connaître les pratiques de la population en deuxième année de master, il serait intéressant de reproduire une analyse de la prise de notes en situation d'épreuve, afin de comparer les résultats collectés en 2008, c'est-à-dire avant la publication de *Notable*

Science et une moindre connaissance de la part de l'équipe enseignante concernant des méthodes de prise de notes. Par exemple, l'analyse de 2008 a dévoilé l'utilisation d'une troisième méthode de prise de notes basée sur la séparation graphique de la page en plusieurs colonnes, une colonne pour chaque question « qui ?, quoi ?, quand ?, comment ? », méthode qui semble peu adaptée au contexte scientifique où les méthodologies, les concepts et les théories sont omniprésents.

2.3.3.5. *PORTÉE TRANSDISCIPLINAIRE*

J'ai pu mettre ma recherche sur la prise de notes au service de deux autres projets universitaires. D'abord, j'ai participé au groupe de travail « S'appropriier les cours » (en langue française) mis en place par le Service interuniversitaire de pédagogie (SUP) à l'UJF en 2012, suite à un constat par les enseignants disciplinaires de lacunes chez les étudiant.e.s en matière de prise de notes. Puis, j'accompagne deux collègues anglicistes, Nathalie Grumel et Carole Liatard, dans leur conception d'une série d'exercices numériques d'aide à la prise de notes pour les étudiant.e.s de niveau avancé.

Les stratégies de prise de notes sont transdisciplinaires et translinguistiques. Une approche de l'évaluation par les compétences en prise de notes permet d'aborder cet acte dans des contextes académiques variés. Seule une recherche approfondie permet de vérifier la validité d'une évaluation.

2.3.4. LE DÉCALAGE ENTRE BESOINS ET CERTIFICATION

2.3.4.1. *EFFET DE L'OBLIGATION DE CERTIFICATION*

En douze ans d'enseignement dans l'enseignement supérieur, j'ai accompagné à de multiples reprises les étudiant.e.s confronté.e.s aux demandes institutionnelles ou professionnelles en matière de certification. Le marché juteux de la certification en langue s'installe au niveau international au sein de démarches de plus en plus sélectives. Deux questions fondamentales alors s'imposent : qui est exclu ou gratifié par cette démarche (l'exemple le plus critique étant sans doute la certification en langues des migrants) et est-ce que les épreuves de certification, qui doivent aussi obéir aux contraintes du marché, évaluent réellement les besoins langagiers pour le contexte en question ?

D'après mon expérience, la situation la plus compliquée sur le plan administratif fut les tergiversations de l'Education Nationale au sujet de la certification en langues et l'introduction du CLES pour le professorat [HART11c]. Heureusement, l'obligation d'obtenir un CLES3 (niveau C1 du CECRL) a été abandonnée, car elle aurait exclu d'excellent.e.s enseignant.e.s de la titularisation. Le plus ennuyeux, pour les étudiant.e.s, comme pour l'équipe enseignante, est la préparation au certificat TOEIC certificat demandé par les écoles d'ingénieurs et dont l'aspect principal de réussite semble être la vitesse et les automatismes autour de certaines structures. Si à l'heure actuelle, les ingénieurs et les scientifiques en général ont besoin de pratiquer l'anglais, leurs besoins portent sur la communication à objectifs spécifiques, et non sur la capacité de répondre à des questions à choix multiples autour de l'anglais général (évaluation moins coûteuse, car informatisée), comme c'est le cas pour le TOEIC.

2.3.4.2. MANQUE DE RECHERCHE CRITIQUE

L'étude du TOEFL 2000 Spoken and Written Academic Language (T2K-SWAL) Project, financé par l'Education Testing Service a établi une vision globale du langage universitaire dans l'objectif de permettre l'évaluation des ressources en anglais pour non natifs, et particulièrement les documents incorporés dans l'examen de certification TOEFL (Test of English as a foreign language), une certification à visée académique, (Biber, 2006 : 19-20). Cette étude d'envergure, dont un corpus de l'anglais oral et écrit des cours, des manuels et d'autres contextes académiques ou administratifs (2 737 200 mots) a mis en avant la différence importante entre l'anglais académique écrit et l'anglais académique oral (*ibid* : 24). Plus que des différences disciplinaires ou contextuelles, cette différence entre oralité et écrit est la variante la plus déterminante (*ibid* : 213-218). Cette étude montre l'importance de la recherche linguistique sur corpus dans le cadre de la certification.

Ce qui m'interpelle est le manque de recherche pertinente et indépendante au sujet de la validité et de l'authenticité de certains dispositifs de certification aux enjeux élevés. Lawson (2008 : 16) défend également cette position concernant le TOEIC, une certification à visée professionnelle, en soulignant des lacunes au niveau de la validité du contenu. Les constats faits par Chapman et Newfields (2008) concernant les

problèmes de validité, en particulier dans la conception des épreuves de compréhension de l'oral (format des questions à choix multiples, quantité importante de texte, impossibilité de prendre des notes) sont des faiblesses que j'ai déjà décrites [HART13a].

Membre de jury du CLES2 tous les ans depuis 2008, j'ai administré un CLES3 (niveau C1 du CECRL) en 2010 et rédigé un sujet CLES3 en 2011. En 2010, j'ai publié dans *ALSIC*, la première publication, à ma connaissance, traitant du CLES et portant notamment sur les difficultés de compréhension d'un document oral, difficultés dues à l'interférence entre un support visuel ne comportant aucune information et de nombreuses questions écrites [HART10c]. En d'autres termes, les supports visuels sont trop facilement associés à une aide à la compréhension. Or, dans certains cas, ils provoquent une surcharge cognitive ou une distraction sans apporter au sens transmis par voie orale. La mise en place d'un comité scientifique du CLES début 2016 devrait affermir l'assise scientifique de cette certification à but non lucratif.

2.3.4.3. *EFFET DE RETOUR*

Ce qui reste problématique est l'effet de retour (*washback*), car le contenu lexical et pragmatique du certificat CLES2 ne correspond guère au contexte de la classe de primaire. Or il s'agissait du certificat le plus mis en avant lors de la titularisation dans l'enseignement primaire. Au lieu de peaufiner les capacités de débattre autour des questions sociétales trouvées dans le CLES, les enseignant.e.s du premier degré ont besoin d'un bagage linguistique et littéraire dans les thèmes ludiques traditionnellement liés à l'enfance, sous forme d'histoires, de chansons ou de jeux, ainsi que des connaissances sur les théories de l'acquisition du langage.

Les avantages du CLES par rapport au TOEIC pour un contexte scientifique sont les sujets abordés et les types de tâches qui peuvent engendrer de la discussion, car le TOEIC comporte des exercices en anglais général ou en anglais des affaires peu compatibles avec le discours intellectuel ou la vie professionnelle scientifique [HART10e, HART11c, HART13e].

2.3.4.4. REPRÉSENTATIONS SOCIÉTALES MISES À MAL

Autre faiblesse, les aspects esthétiques et sociaux sont peu développés dans le TOEIC. En fait, les illustrations sont très souvent des photographies de qualité moyenne de personnes en situation professionnelle, avec une surreprésentation des hommes blancs en situation de pouvoir. Nous trouvons, par ailleurs dans les supports visuels des examens ou dans les préparations aux examens environ deux hommes pour chaque femme, un déséquilibre qui est insupportable sur le plan didactique [HART10b, HART11d]. Dans ce sens, le recours à la certification massive sans adopter une attention à la langue de spécialité n'est pas toujours dans l'intérêt des étudiant.e.s, ni dans celle de la société.

RÉSUMÉ DE LA DEUXIÈME PARTIE

Dans cette deuxième partie, j'ai traité d'une modélisation du processus d'enseignement, de la place du lexique, des projets et considérations didactiques, ainsi que des modes d'évaluation. J'ajouterais que j'ai bénéficié d'une grande marge de manœuvre me permettant d'agir de manière créative, ceci dans une atmosphère collaborative et collective lors de la mise en place du dispositif de validation de niveau B2/indépendant en anglais de spécialité et des supports pédagogiques, dont *Enigma Plus* et *English Booster Course for Master's*. J'ai pu transmettre les résultats de ma recherche en didactique et en anglais de spécialité directement au service d'une équipe enseignante, en plus des publications et des communications.

Depuis, le dispositif d'évaluation a été reçu avec un intérêt certain lors de rencontres avec des collègues d'autres universités, lors des rencontres du groupe de travail « Sciences et Academia » du GERAS et lors de la journée « Les certifications en langues dans le secteur LANSAD » organisée par le laboratoire « Image Langage » à l'Université de Bourgogne le 13 novembre 2015. La présentation à Dijon, en collaboration avec Marie-Hélène Fries et Emmanuelle Esperança-Rodier sera aussi l'objet d'une publication. Le dispositif mériterait une exploration plus fine, par exemple, autour du contenu des supports audiovisuels et des capacités des étudiant.e.s en situation de handicap.

3. TROISIÈME PARTIE : ANALYSE DU DISCOURS DES EXPERT.E.S

Nous avons vu (*cf.* 1.1.4.) qu'en didactique des langues de spécialité, effectuer des tâches pédagogiques permet d'acquérir des connaissances et des savoir-faire linguistiques appartenant à une communauté de discours. Le fil conducteur de cette troisième partie est l'identification de quelques notions linguistiques constitutives de l'anglais académique. Pour cela, je fais de nouveau appel à l'analyse du discours pour explorer les notions de causalité (*cf.* 1.3.1.) et de présence auctoriale (*cf.* 1.3.3.) dans les textes des expert.e.s en sciences exactes. J'expose ici mes contributions à la recherche dans la tradition théorique de l'analyse de discours et dans l'analyse du discours spécialisé abordés dans la première partie et tout particulièrement les travaux selon le genre de Bakhtine (*cf.* 1.1.) et de Swales (*cf.* 1.2.3.), la métaphore grammaticale de Halliday (*cf.* 1.3.2.) et le positionnement d'Hyland (*cf.* 1.3.3.2.) et du LIDILEM (*cf.* 1.3.3.).

Carter-Thomas (2009) décrit la linguistique de corpus, comme étant « la pierre angulaire » du travail d'investigation en langue spécialisée (2009 : 124). La linguistique de corpus est relativement nouvelle et ses capacités puissantes d'exploration linguistique méritent une brève introduction. Afin d'illustrer l'évolution des corpus, nous prenons comme exemple quelques corpus et outils avec lesquels je travaille depuis plusieurs années. Puis en deuxième partie, j'expose ma contribution à la recherche en langues académiques élaborée à partir du corpus Scientext et qui s'intègre pleinement dans les objectifs du laboratoire LIDILEM.

3.1. MÉTHODES ET MOYENS D'ANALYSE DU DISCOURS DES EXPERT.E.S

3.1.1. UNE APPROCHE FONDÉE SUR DES CORPUS SPÉCIALISÉS

Au cours de mes recherches, j'ai travaillé à partir de différents outils et ressources, dont chacun comportait des intérêts particuliers, mais aussi des inconvénients qu'il fallait contourner, voire aménager. Au laboratoire LIDILEM, j'ai bénéficié d'une intégration rapide dans l'axe 1 « Descriptions linguistiques, TAL, corpus » (TAL : Traitement Automatique des Langues), où les connaissances en linguistique de corpus et de production de corpus de langue française sont reconnues au niveau international. Dès 2010, je me plonge dans les possibilités d'analyse du discours de spécialité offertes par ces techniques et ressources.

Dans la première section de cette partie, je développe brièvement les intersections et comparaisons entre quatre corpus constitués de documents académiques ou scientifiques. Les genres académiques sont intégrés dans les quatre corpus que nous développons ici : le John Swales Conference Corpus (JSCC) comportant 100 000 mots, l'Identité Culturelle dans la Prose Académique (KIAP) comportant 3 152 022 mots, le corpus constitué lors du projet Étude EIIDA (en cours) et enfin, Scientext (35 244 378 mots pour le corpus d'articles de recherche en anglais). Ensuite, je développerai trois notions, examinées à travers le filtre d'une recherche linguistique sur corpus. Halliday a introduit la notion de métaphore grammaticale (*cf.* 1.3.2.), que nous développons ici en examinant l'articulation entre métaphores grammaticales possibles par une séquence Adverbe + *V-ing*. La subjectivité du discours scientifique constatée par Hyland (2005, 2012) et Grossmann et Tutin (2014) (*cf.* 1.3.3.) est abordée au travers du filtre des lemmes HOPE et WISH [HART16b]. Enfin, suite aux travaux notamment de Fløttum et d'autres (Fløttum, 2010 ; Fløttum *et al.*, 2006), je me focalise sur les aspects interpersonnels en discours oral des sciences.

3.1.1.1. JOHN SWALES CONFERENCE CORPUS

L'English Language Institut à l'University of Michigan a organisé une conférence autour de l'analyse de discours lors du départ en retraite officielle de John Swales en juin 2006. Sous le pilotage d'Ute Römer, un corpus de cette conférence a été constitué à partir des

23 transcriptions des présentations et des séances des questions et réponses, pour un total d'environ 100 000 mots, dont 77% issus des présentations et 23% des questions et réponses. Le corpus JSCC (John Swales Conference Corpus) présente un grand intérêt, car nous pouvons y découvrir des discussions autour de thèmes centraux et les détails de la langue des expert.e.s en linguistique. Le corpus est accessible en ligne à des fins non commerciales.

3.1.1.2. *CORPUS DE L'ÉTUDE INTERDISCIPLINAIRE ET INTERLINGUISTIQUE DU DISCOURS ACADÉMIQUE*

Le corpus JSCC est intégré au corpus du projet transdisciplinaire et trilingue, EIIDA porté par Shirley-Carter Thomas et Jeanne-Marie Debaisieux dans le cadre du labex TransferS, auquel je participe depuis 2012. L'objectif principal de ce projet est généralement l'identification des marqueurs d'organisation textuelle ; les analyses portent sur :

des faits de syntaxe et sur les constructions traditionnellement considérées soit comme caractéristiques de l'écrit (tels les passifs, les nominalisations, les structures impersonnelles, et certains connecteurs argumentatifs), ou de l'oral (structures détachées, phrases pseudo clivées ou encore certaines locutions adverbiales énonciatives) (Poibeau, 2012).

La constitution du corpus EIIDA a demandé un travail collectif, car il s'agit d'un corpus anglais-espagnol-français de transcriptions de quinze présentations orales par langue et de quinze articles de recherche par discipline des mêmes orateurs et oratrices sur le même thème, et en linguistique et en sciences exactes pour un total d'environ 180 documents.

Le sous-corpus du projet EIIDA d'anglais scientifique autour du thème de l'eau comporte les transcriptions de 15 présentations orales (37 000 mots) par 15 orateurs natifs ou oratrices natives et 15 articles de recherche publiés (68 400 mots) par les mêmes auteur.e.s sur le même sujet que leur présentation orale. L'ensemble des documents forme un total de 105 400 unités lexicales et un ratio de différence de volume entre les deux genres autour de 1 : 1,85, le sous-corpus écrit étant plus volumineux (cf. 3.2.3).

3.1.1.3. LE CORPUS D'IDENTITÉ CULTURELLE DANS LA PROSE ACADÉMIQUE

L'objectif principal de la constitution du corpus KIAP fut lié à « l'identification d'identités culturelles telles que manifestées par des traces linguistiques de voix scientifiques dans le genre de l'article de recherche » en anglais, en français et en norvégien dans les domaines de l'économie, la politique, la médecine et la linguistique (Fløttum, 2012 : 4). Il s'agit de 450 articles (3 152 022 mots) tirés de revues connues et publiés entre 1992 et 2003 permettant d'examiner l'hypothèse qu'il y a « plus de ressemblances entre les articles provenant de la même discipline qu'entre les articles rédigés dans la même langue » (*ibid* : 8 et 14). De nombreuses publications sont le fruit de cette ressource, mais nous pouvons noter ici la conclusion qu'en anglais et en norvégien, la voix des auteur.e.s est plus présente qu'en français (*ibid* : 9).

Fløttum et d'autres (Fløttum 2010, Fløttum *et al.*, 2006) mettent l'accent sur les qualités polyphoniques du discours scientifique, cette polyphonie au sein d'une communauté de recherche contribuant aux similitudes disciplinaires, malgré les différences inhérentes à une langue ou à une culture donnée. L'existence de cette polyphonie interdisciplinaire est confirmée par l'analyse du corpus KIAP (Fløttum *et al.*, 2007). Les manifestations autoriales permettent d'identifier quatre rôles : *scripteur*, *chercheur*, *argumentateur* et *évaluateur*. Pour le rôle de *scripteur*, nous pouvons trouver, en anglais, des verbes de description tels que DESCRIBE, ILLUSTRATE, PRESENT, BEGIN BY, FOCUS ON, pour celui de *chercheur*, des verbes portant sur l'engagement cognitif, ANALYSE, ASSUME, CONSIDER, EXAMINE, FIND et STUDY ; la prise de position par l'*argumentateur* peut être relevée dans les verbes ARGUE, CLAIM, DISPUTE, REJECT, enfin, plus rare, le rôle d'*évaluateur* est incarné par des verbes de jugements de valeur ou d'évaluation, tels que FEEL ou BE SCEPTICAL (voir aussi pour le français : Fløttum & Vold, 2010) [HART14a] (*cf.* 3.2.2.).

3.1.1.4. CORPUS SCIENTEXT

Les sections françaises en linguistique, en économie et en médecine du corpus KIAP ont été intégrées à la version non publique du corpus Scientext, projet financé par l'ANR « Corpus » entre 2007-2010. Le projet Scientext autour des écrits scientifiques, a réuni trois équipes de linguistes ayant le double objectif de mener une étude théorique du

positionnement et du raisonnement du discours scientifique et d'offrir aux linguistes un corpus librement disponible en ligne. L'équipe coordinatrice du projet, le laboratoire LIDILEM, a élaboré l'interface informatique, ScienQuest, permettant d'interroger le corpus annoté selon trois modes : des requêtes ou « grammaires » sémantiques préétablies, des requêtes pouvant intégrer formes, lemmes, catégories morphosyntaxiques et relations syntaxiques et enfin, des requêtes sous forme de grammaires locales (Falaise, Tutin & Kraif, 2011a ; Falaise, Tutin & Kraif, 2011b ; Falaise, Tutin, Kraif & Rouquet, 2012).

Scientext comporte quatre corpus spécifiques :

- un corpus d'écrits scientifiques en français (4,8 millions de mots),
- un corpus d'évaluation de propositions de communications (520 commentaires) constitués par le laboratoire LIDILEM,
- un corpus d'écrits universitaires en anglais langue étrangère autour de la sociologie (1,1 million de mots) élaboré par le laboratoire LLS (Langage, Littérature, Société) de l'Université de Chambéry (maintenant l'Université de Savoie Mont Blanc),
- le corpus qui nous intéresse dans le cadre de ce dossier de synthèse, un corpus de publications scientifiques (biologie et médecine) en anglais mis en place par le laboratoire LiCoRN (Linguistique de corpus) à l'Université de Lorient.

Ce projet a donné lieu à une soixantaine de publications fondées sur plusieurs pistes de recherche. Un travail important s'est fait autour de l'énonciation et de la rhétorique (cf. Boch & Rinck (Eds.), 2010), dont les marques d'évidentialité (ce sur quoi se fonde une assertion) (Grossmann, 2010 ; Grossmann & Tutin, 2010), les collocations (Grossmann & Tutin (Eds.), 2003) ainsi qu'en TAL en didactique ou en création de dictionnaire (Grossmann & Tutin (Eds.), 2014). Ayant intégré le LIDILEM lorsque le projet ANR se terminait, mes travaux ont été influencés par les avancées importantes de l'équipe en linguistique du discours académique en français. Sans se limiter à un travail comparatif, j'ai repris les mêmes axes de travail, dont les verbes, les collocations lexicales (*j'é mets l'hypothèse / I hypothesize*) et le positionnement à partir d'un corpus Scientext en anglais. Je participe à l'évolution de ces travaux par l'animation scientifique,

dont des demandes de financement, l'organisation de conférences et des projets en didactique et de recherche menés en partenariat avec des membres internes et externes au laboratoire LIDILEM. J'encadre l'étudiante en master Melanie Wilkos, qui traduit l'interface en anglais.

Le corpus d'articles de recherche scientifique en anglais est constitué de 7 564 publications, essentiellement des articles de recherche publiés entre 1997 et 2005 dans 137 revues en accès libre par BioMed Central dans les domaines de la biologie et de la médecine. Les textes sont formatés selon le modèle TEI (*Text Encoding Initiative*) et balisés pour les catégories grammaticales (*parts of speech*) et les lemmes avec Treetagger. Les annotations morphologiques et syntaxiques sont obtenues grâce à Syntex (Tutin & Falaise, 2014). Nous verrons que le balisage et l'annotation ont un effet très important sur les possibilités d'analyse. Bien que le contenu du corpus privilégie les domaines biomédicaux, sa taille importante (35 millions de mots) lui permet d'être une source valide pour des études plus générales, par exemple une étude des verbes scientifiques (Williams, 2012).

3.1.2. LA PUISSANCE DES OUTILS : TXM, PLATEFORME D'ANALYSE TEXTUELLE

3.1.2.1. L'IMPORTANCE DES OUTILS

En plus du contenu, un autre facteur à considérer lors d'une recherche sur corpus est l'ensemble des outils mis à disposition. Lors de certaines analyses d'auxiliaires modaux, je me suis appuyée sur la plateforme TXM. TXM est une plateforme logicielle ouverte d'analyse textométrique développée par le laboratoire ICAR (Interactions, Corpus, Apprentissages, Représentations) à l'Université de Lyon depuis 2007. Cela veut dire qu'elle ne comporte pas de corpus propre comme Scientext, le BNC ou le COCA, mais permet d'importer et d'analyser son propre corpus. Une interface API (Application Public Interface) donne accès aux diverses composantes raccordées entre elles. Cette interface comporte quatre parties principales, dont l'accès et la manipulation des données, les apports de certains calculs statistiques, des fonctionnalités spécifiques à la textométrie (calcul des spécificités, concordance...) qui dépendent des deux premières

parties et enfin, les capacités d'importation (Heiden, 2010 ; Heiden, Magué & Pincemin, 2010 ; Poudat & Follette, 2012).

3.1.2.2. *UNE APPROCHE QUANTITATIVE*

Des formations TXM proposées par l'équipe ICAR m'ont permis d'explorer cet outil aux multiples avantages. L'expérience de collaboration avec Marie-Paule Jacques dans ce cadre fut enrichissante, car nous avons traité, de manière quantitative, les énoncés relatifs à un objectif de recherche dans un corpus de résumés d'articles scientifiques en anglais, puis effectué une comparaison de l'emploi de pronoms en français et en anglais dans les diverses sections d'un article de recherche en sciences [HART14a]. Pour ce dernier travail, le financement UJF « Structure discursive de l'article de recherche – approche comparative français / anglais, sciences humaines et sociales / sciences exactes » (cf. 2.2.1.3) a permis à une doctorante, Tran Thi Thu Hoai, de travailler sur un élargissement du corpus en français, au mastérisant Julien Corman, de faire une analyse linguistique et au développeur Achille Falaise d'avancer sur plusieurs points de Scientext.

3.1.2.3. *LES AUXILIAIRES MODAUX EN LINGUISTIQUE DE L'ORAL*

À l'aide de TXM, j'ai aussi exploré le corpus JSCC dans l'objectif d'examiner l'emploi de verbes modaux, qui, d'après mon expérience, est souvent une source de difficultés pour les apprenant.e.s francophones. La Figure 18 montre une partie des résultats d'une requête des verbes modaux *can/could/may/might/will/would/shall/should* dans un corpus de 13 transcriptions de présentations orales (environ 45 000 mots) du JSCC. La colonne du centre comporte le pivot, ici le mot *would*. De chaque côté, on trouve le contexte de gauche et de droite. Ces contextes peuvent être triés par ordre alphabétique, le contexte de gauche étant trié par le mot le plus près du pivot. Enfin, la colonne complètement à gauche indique le texte source. Nous voyons donc de grandes similitudes avec Scientext. Il s'agit d'une présentation classique KWIC, comme pour beaucoup d'interfaces de corpus.

Query: [word="can|could|may|might|will|would|shall|should"]

sort keys: #1 Keyword #2 None #3 None #4 None #5 None

301-397 / 397

text_id	Left context	Keyword	Right context
JSC02_script	i direct and when i was asked who	would	be your first choice of external reviewer i said John Swales
JSC02_script	Swales but i didn't really think he	would	agree to come < EVENT DESC = " laugh " / >
JSC14_script	one hand these R-P-T reports seem like they	would	be highly individual, as they are centrally evaluative and hence filled
JSC14_script	texts with their own voices. today i	would	like to report on a study of all th- these writers'
JSC14_script	university. these texts fit into what Swales	would	call a genre chain, in which one genre triggers the creation
JSC14_script	agree that the that the English department probably	would	be a better place if we were all replaced < EVENT DESC
JSC15_script	to gain from a graduate degree, that	would	build on previous knowledge. however the rhetorical forces of these
JSC15_script	roles i've always felt that an M-B-A	would	complement my experience < / SEG > and so on... in
JSC15_script	note however, the discussion of extracurricular activities	would	be present in the statement for undergraduate admission, and is eve
JSC15_script	shed more light on this question. it	would	also be interesting to compare statements written for entry to PhD p
JSC15_script	i had not received those invitations, i	would	never have come to the University of Michigan. - < U
JSC16_script	removed from any sense of audience, who	would	be interested in reading about how a bicycle pump works ? <
JSC16_script	. Madison Avenue and other promoters of advertising	would	have us believe that ads provide information that enables consumer
JSC16_script	, and we all have negative information that	would	be of interest < EVENT DESC = " laugh " / >
JSC16_script	when he predicted that the prince of Wales	would	remarry quote, because he is the kind of person who almost
JSC16_script	, to pontificate on constitutional issues. he	would	offer the simplest explanation of how the queen could give up her
JSC16_script	was quick to assure the queen that she	would	be safe from Islamic fundamentalism, because he had discovered, st
JSC16_script	reflects the patrilineal nature of Egyptian society and	would	have to be part of the prototypical Egyptian obituary, whether comp

Figure 18 : Résultats d'une requête sur TXM : auxiliaires de modalité (JSCC)

L'auxiliaire modal le plus fréquent dans ce corpus de présentations orales dans le domaine de la linguistique est *can* ($n = 120$; 30,3%), suivi par *would* ($n = 73$; 20,2%) et *will* ($n = 58$; 14,6%). Des 396 occurrences, les modaux *may* ($n = 49$), *might* ($n = 36$), *could*

(*n* = 30), *should* (*n* = 26) constituent un gros tiers (35,6%) de cette catégorie. Enfin, *must* (*n* = 16) et *shall* (*n* = 3) ne sont que très peu employés. Cette distribution montre que dans l'expression orale dans le domaine de la linguistique, les anglophones s'appuient sur une large gamme de possibilités de modalité. Or, quatorze ans d'enseignement auprès de divers publics me laissent penser que les modaux *can* et *will* sont assez bien maîtrisés, mais que les modaux *would*, *may* et *might* le sont beaucoup moins et demanderaient souvent un travail ciblé d'acquisition (pour l'écrit, voir : Hyland & Milton, 1997). Il serait intéressant d'évaluer ce phénomène à travers des corpus de textes oraux, notamment dans le cas de cours dispensés en France en anglais.

3.1.2.4. LIEN ENTRE L'ANGLAIS GÉNÉRAL ET L'ANGLAIS DE SPÉCIALITÉ

Une requête à partir de TXM sur le corpus JSCC a permis de confirmer de nouveau la présence de verbes modaux dans le discours des expert.e.s. J'ai choisi de travailler ce thème, car en classe, je constate des difficultés autour de l'anglais général et en anglais des sciences. Une erreur fréquente se produit lorsque l'apprenant.e confond *would* avec *voudrais* : **I would a new car* au lieu de *I would like a new car*. Une deuxième source de mauvaise maîtrise concerne les traductions autour d'une marque de respect pouvant être exprimée par le vouvoiement en français ou le modal *would* en anglais. Par exemple, la question classique *Would you like some coffee* peut être traduite *Aimeriez-vous du café ?*, qui est une formulation moins employée que *Voulez-vous un café ?* L'apprenant.e utilisera pour cette dernière *Do you want a coffee ?*, expression qui perd la marque de politesse du vouvoiement ou de la conjugaison avec *would*. Ces exemples en anglais général se glissent dans le discours spécialisé de l'apprenant.e qui aura tendance à dire *today, I report on a study*, alors qu'il serait plus approprié de dire *today I will report on a study* ou dans un registre plus soutenu : *today I would like to report on a study* (cf. Figure 19).

La Figure 19 illustre les nombreuses occurrences autour de la probabilité ou de l'hypothèse :

- the English department would probably be a better place if we were all replaced
- I was asked who would be your first choice of an external reviewer

Cette structure comportant l'auxiliaire *would* est difficile pour beaucoup d'apprenant.e.s, mais la consultation d'exemples authentiques peut aider à explorer la richesse linguistique.

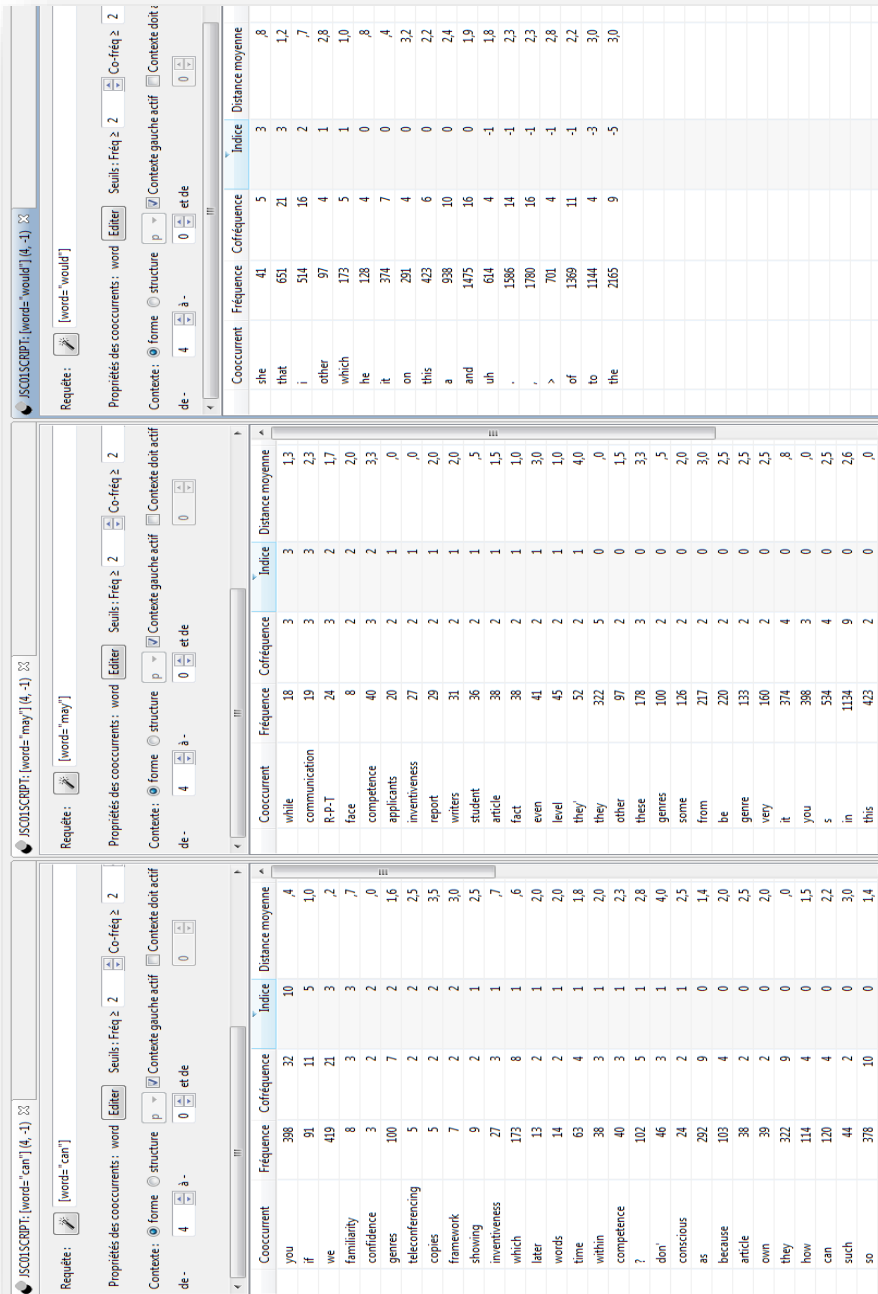


Figure 19 : Saisie d'écran des résultats d'une requête de cooccurrences de *can/may/would* (JSCC)

3.1.2.5. FRÉQUENCE DE COOCCURRENTS

Nous avons vu l'importance des collocations dans l'analyse du corpus (cf. 1.2.2.). TXM permet aussi de détecter les cooccurents, en d'autres termes, les éléments qui se trouvent à proximité. Par exemple, dans la Figure 19, nous voyons les cooccurents jusqu'à quatre places (occurrences) à gauche de *can*, *may* et *would*. Pour chacune des trois listes de réponses à cette requête, sont indiqués le cooccurrent, la fréquence totale du cooccurrent dans le corpus, la fréquence de proximité avec le pivot, un indice reflétant l'importance de cette cofréquence et enfin, la distance moyenne entre le pivot et le cooccurrent. Ces résultats montrent une grande variation dans l'utilisation de ces trois modaux. Alors que *can* s'emploie dans ce corpus avec les pronoms *you* ($n = 32$, avec un fort indice de cofréquence) et *we* ($n = 21$), *may* se trouve davantage en corrélation avec des notions ou des objets (communication, rapport), des groupes (candidat.e.s, scripteurs ou scriptrices), le premier pronom de collocation étant *they* ($n = 4$). Enfin, nous pouvons noter que l'indice de *she would* ($n = 5$) est de trois et que l'indice de l'expression *I would* ($n = 16$) n'est que de deux, car le pronom *I* ($n = 514$) est beaucoup plus fréquent dans ce corpus que *she* ($n = 41$). En d'autres termes, l'emploi de pronoms personnels et de verbes modaux (*you can*, *we can*) participent à l'aspect interpersonnel d'un contexte de conférence.

3.1.2.6. CALCULER LA SIGNIFICATION STATISTIQUE

TXM ne permet pas encore de calculer, de manière statistique, l'importance des collocations. Olivier Kraif a eu la générosité de me fournir une fiche de calculs statistiques, qui permet de calculer le PMI, le résultat d'un t-test et le log-likelihood [Hart13b]. Plusieurs modes de calcul sont disponibles, dont le calcul PMI (*Pointwise Mutual Information*) et le t-test (cf. Hunston et Francis, 1996), le calcul de log-likelihood (cf. Dunning, 1993 ; Ellis et Simpson-Vlach, 2009) et le calcul de MI (*Mutual Information*). Nous voyons dans le Tableau 9, que ces calculs ne donnent pas les mêmes résultats, car le PMI produit des résultats plus importants lorsqu'il s'agit de petits effectifs. Dans le présent cas, *i would*, se trouve avec un PMI plus fort que *we can*, ce qui, à mon sens représente de manière plus juste « l'attirance » entre *I* et *would*.

Tableau 9 : Calculs statistiques de trois modaux et prénoms du JSCC

Pivot	Collocatif	PMI	LogLike	MI
can	you	4,91	241,6	0,0027
can	we	4,23	126,9	0,0014
would	i	4,26	98,2	0,0010

3.1.3. L'INTÉRÊT DE L'ANNOTATION SYNTAXIQUE DE SCIENTEXT

Contrairement à la plupart des autres corpus intégrant des outils, l'annotation morphologique de Scientext est accompagnée d'une annotation syntaxique. Il s'agit donc d'un corpus dit *arboré*. ScienQuest a aussi des capacités de concordances à partir des formes, lemmes et catégories syntaxiques (*parts of speech*), la potentialité de préciser des relations syntaxiques, sans contraintes par rapport à l'ordre linéaire des phrases. Ainsi, une grammaire de requête de Scientext peut capter les séquences ayant un sujet [WE]. Afin d'illustrer l'intérêt de l'annotation syntaxique, je propose une requête à partir de la grammaire locale [*we* + Verbe]. Scientext identifie 59 534 occurrences, sous format KWIC (Figure 20), dont :

We will touch on a couple of these approaches below
 We first identified 50 clones
 We thus may be able to ultimately derive a sequential pattern of development

Grace à l'analyseur syntaxique conçu par Bourigault, nommé Syntex (Bourigault, 2007), intégré à ScienQuest, les dépendances sont repérées. Ici, grâce à l'expression (SUJ|SUJL), la requête ne détecte que les verbes syntaxiquement liés au pronom *we*. Cette capacité de prise en compte de la relation syntaxique élimine des cooccurrences telles que *using this method we aimed* où *using* est identifié comme étant un verbe en cooccurrence avec *we*, alors que le verbe lié syntaxiquement au *we* est le verbe fléchi *aimed*. Sans imposer un lien syntaxique, il y a 65 262 occurrences comportant *we* et un verbe ou environ 5 728 occurrences de bruit inutile.

Scientext		
Recherche	Résultats	
Sémantique Libre Avancée		
pro- inflammatory molecules , we designed a study in which	we	compared isolated monocytes from COPD patients and controls . In addition
acute exacerbations , cigarette smoking , and corticosteroid treatment .	We	were careful to ensure that groups of individuals well were matched
they were in a stable clinical state . Therefore ,	we	infer that the differences observed in basal monocyte behaviour and
monocytes and induces neutrophil and basophil chemotaxis , 29 , 31 .	We	found that monocytes isolated from COPD patients release significantly less
from COPD patient blood monocytes might be related to corticosteroid treatment .	We	also find that the LPS activated COPD monocytes release more
different mechanisms dependent on the basal status of monocytes .	We	compared levels of pro- inflammatory molecules released by monocytes obt
promoter of target genes , 34 . In this study	we	evaluated whether NF & # 954 ; B expression is modified
: molecular abnormalities and therapeutic implications In this review ,	we	will summarize the current state of knowledge regarding the cascade of
potential use in early detection strategies . In this paper	we	will review the genetic basis of lung cancer progression using a
important component underlying genomic instability and eventually tumor growth .	We	will illustrate this point by commenting on 3 classical examples :
than look at the individual impact of isolated changes ,	we	have begun efforts to " cluster " changes into groups of
lung cancer histological subtypes based on array CGH profiling ,	we	first identified 50 clones (most of which were on chr
instability , 6 , 7 and DNA methylation patterns , 8 .	We	thus may be able to ultimately derive a sequential pattern of
cancers of various histological subtypes have been analyzed by genomic approaches .	We	recently published the results of such analysis in a series

Figure 20 : Saisie d'écran de résultats "KWIC" de Scientext

Scientext permet d'identifier les deux éléments alors même qu'ils sont discontinus, séparés par exemple, par un adverbe. Sans cette possibilité, la même requête n'identifie que 52 304 occurrences. Cette différence quantitative peut se traduire par le manque de 7 230 occurrences dans lesquelles le pronom *we* n'est pas adjacent au verbe, principalement à cause d'un adverbe. Nous pouvons remarquer que Scientext détecte le verbe conjugué et la présence de verbes auxiliaires. À partir de l'interface Scientext, il est possible de formuler des requêtes en formulant une grammaire ou via la page de l'interface « Libre », comme illustré dans la Figure 21.

3.1.3.1. UNE VISION GLOBALE GRÂCE AUX STATISTIQUES

L'option « statistiques » offre une synthèse des résultats, sous forme numérique et graphique, selon la forme, le lemme, la catégorie grammaticale, les traits morphosyntaxiques ou la partie textuelle. Cette vision globale est très utile pour le repérage des hautes fréquences ou pour mettre en relief d'autres éléments productifs en contexte d'enseignement. La synthèse des résultats par partie textuelle permet de recenser les occurrences dans l'introduction, le développement, la conclusion ou le résumé [HART14a]. Dans le cas de notre exemple, les statistiques autour de « lemme »

montrent que USE ($n = 7\,170$, 7,4539 % des verbes liés au pronom *we* et 2,0344% du corpus) est le plus fréquent, suivi de FIND ($n = 6\,709$, 6,9746 % des verbes liés au pronom *we*, et 1,9036% du corpus) et SHOW ($n = 3\,229$, 3,3568 % des verbes liés au pronom *we*, et 0,9162% du corpus). En revanche, en consultant les statistiques autour des formes, nous remarquons que SHOW se trouve en neuvième place et qu'il y a une forte présence du temps passé (cf. Figure 22).

La plupart des corpus n'offrent pas la recherche par relation syntaxique, se limitant à une simple recherche par proximité. De même, la capacité d'identifier le lien syntaxique en dépit d'un éloignement, parfois important, permet des résultats de qualité. La recherche sur Scientext peut offrir des possibilités de recherche très fines, et ceci à partir d'une base de données de grande taille.

ScienQuest – Corpus «Écrits scientifiques en anglais»

Corpus | Textes | Recherche | Résultats

Sémantique | Libre | Avancée

Mot 1

Forme: we
Lemme:
Catégorie:

Mot 2

Forme:
Lemme:
Catégorie: Verbe
Traits:

Relations syntaxiques

Mot 1: sujet de (SUJ) | Mot 2:

Chercher 59534 occurrences

#	Contexte gauche	Occurrences	
1	of activities , 23 . In the present study ,	we sought	to directly test the relationship
2	to the group 1 allergens , 20 . Therefore ,	we assayed	the 3 pollen diffusates for prote
3	the protein responsible for the proteolytic activity . Therefore ,	we undertook	additional studies using two-di spectrometry as
4	cytokines and chemokines involved , 17 . In this study	we investigated	pro- inflammatory molecular re
5	after stimulation with LPS . By using this experimental approach	we aimed	to determine the differences in

Figure 21 : Saisie d'écran d'une requête sur Scientext, mode libre

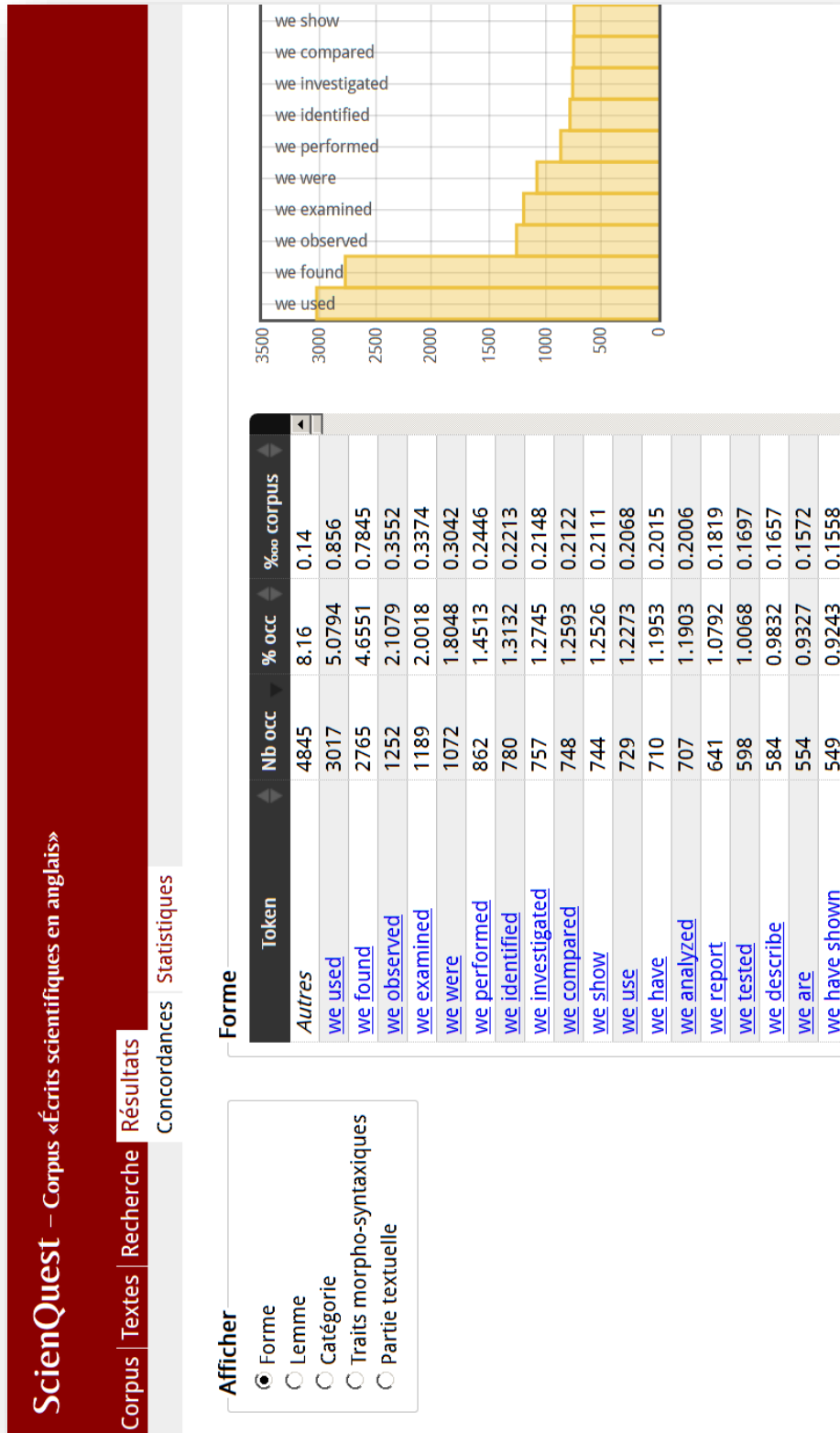


Figure 22 : Saisie d'écran des résultats statistiques d'une requête sur Scientext de [we + Verbe]

3.1.3.2. LA QUALITÉ LINGUISTIQUE DU CORPUS

L'analyse d'un corpus sur Scientext identifie les véritables pratiques langagières existantes dans le contexte de la recherche internationale en biomédecine. Cette recherche permet ainsi de confronter les règles ou consignes morpho-grammaticales administrées dans un cadre didactique avec l'anglais des publications scientifiques. Il est concevable qu'un texte rédigé par des scientifiques n'offre pas les mêmes qualités linguistiques qu'un texte rédigé par des anglicistes. Dans cette éventualité, un corpus de textes rédigés par des scientifiques pourrait comporter des éléments de moins bonne qualité sur le plan linguistique qu'un corpus de textes rédigés par des anglicistes. En réponse à cette possibilité, j'ai examiné une erreur très fréquente des francophones, l'utilisation du pluriel du terme *informations*. Par exemple, il est généralement accepté que *information* est un substantif indénombrable au singulier, c'est-à-dire n'ayant que la forme au singulier.

En français, nous trouvons la forme plurielle *informations* et ce contraste amène les apprenant.e.s francophones à se tromper par interférence avec la langue maternelle, produisant une *non-target form* (Osborne 2008 : 68). Nous trouvons cependant de rares exceptions du pluriel *informations* en anglais dans le domaine de l'informatique. Une requête à partir du terme *informations*, en anglais sur Scientext, montre 26 occurrences contre 19 613 occurrences du terme standard *information*, ce qui correspond à un taux de 0,13% de la forme plurielle.

Les 26 occurrences d'*informations* viennent de 20 articles, notamment en génétique, dont la plupart rédigés dans des pays non anglophones (Allemagne, Autriche, Belgique, Brazil, Grèce, Italie, Iran, Norvège, Thaïlande), ce qui pourrait être une source d'interférence de la langue maternelle. Il y a aussi un article rédigé aux États-Unis, un autre au Pays de Galles, et un troisième avec un des auteur.e.s aux États-Unis. Nous pouvons conclure qu'un si faible taux (0,13%) ne représente pas une tendance vers l'utilisation du pluriel *informations* dans le domaine de la génétique. Aussi, ces chiffres laissent penser que même si une utilisation non orthodoxe de ce type existe dans un corpus, elle reste noyée dans une masse importante de données.

3.1.3.3. ÉVALUER L'EXISTENCE D'ADVERBE EMPLOYÉ ENTRE TO ET UN VERBE (SPLIT INFINITIVES)

Une analyse à partir de Scientext permet aussi d'examiner l'usage autour des questions stylistiques. Cette démarche est particulièrement intéressante lorsque la règle grammaticale est contestée par des linguistes, comme cela est le cas lorsque le *to* d'un verbe à l'infinitif est coupé du verbe lexical *split infinitives*. Traditionnellement, cette pratique est inacceptable, mais reste source de débat :

The prescriptive view of grammar condemns the use of split infinitives [...] We would recommend that you find out your reader's preference for splitting (or not splitting) infinitives and decide what to do. (Both Chris and John use split infinitives in their writing) (Swales & Feak, 2004 : 23).

Parmi les 8 000 occurrences d'un verbe conjugué à l'infinitif comportant un adverbe après le *to* et avant le verbe [*to* + Adv + Verbe] (*split infinitive*) dans le corpus Scientext, nous trouvons 937 motifs lexico-grammaticaux différents dont 46 motifs de 10 occurrences ou plus, notamment *to better understand* ($n = 141$), *to significantly increase / reduce / decrease* ($n = 73$) et *to further investigate* ($n = 40$). Les intrus lexicaux sont essentiellement composés d'adverbes dont les plus fréquents sont *further* (EXPLORE, ELUCIDATE, EXAMINE...) ($n = 288$) et *significantly* ($n = 118$). Il semble donc que certaines formulations de ce type, bien que souvent mises en cause par les grammairiens, soient rentrées dans les mœurs. Contrairement à l'unité lexicale *information*, il n'y a pas, à mon sens, d'erreur, mais une expression rendue plus fluide et compréhensible par l'emploi d'un *split infinitive*, et dont la haute fréquence marque une existence acceptée par la communauté de discours. Ce type de requête permet d'évaluer l'évolution d'une langue et de confirmer ou non le rapport entre sa description et son usage.

3.1.3.4. INVESTIR DANS UN CORPUS

Le corpus Scientext est en constante évolution. Même après la fin du projet ANR « SCIENTEXT : un corpus et des outils pour étudier le positionnement et le raisonnement de l'auteur dans les écrits scientifiques » (2007-2010), Achille Falaise et d'autres chercheur-ses du LIDILEM ont continué à l'améliorer en ajoutant de nouveaux corpus et en identifiant et réparant les bugs ou en modifiant les aspects et les options de l'interface. Il y a aussi un travail de collaboration avec Achille Falaise pour l'amélioration

de l'ergonomie de la plateforme. J'ai aussi contribué à ce travail, notamment en détectant certains bugs dans le corpus d'articles de recherches en anglais, en contribuant au travail de l'incorporation de métadonnées et en portant le projet Corpus4All (Financement UNR-RA) (cf. 2.1.3.6.). Samia Ounoughi (LIDILEM) a récemment rejoint ce projet en initiant une analyse sur Scientext de la citation dans une perspective didactique dans le cadre du Corpus4All. L'ensemble de ces travaux ont pour perspective d'ouvrir cette ressource à d'autres chercheurs, mais aussi aux équipes enseignantes en langues de spécialité ou en traduction.

3.1.3.5. *RICHESSSE ET PORTÉE DIDACTIQUE DE LA PHRASÉOLOGIE TRANSLINGUISTIQUE*

Les thématiques de la phraséologie transdisciplinaire et translinguistique approfondies lors des travaux autour de Scientext, notamment la structuration de l'écrit, l'emploi de verbes de causalité ou de pronoms et la didactique nourrissent mes propres recherches. En participant avec Francis Grossmann à l'obtention du financement ARC6 en 2013 « Perspectives nouvelles en phraséologie : Quels outils pour la description et l'acquisition des langues ? » et en organisant avec Cristelle Cavalla les journées d'étude « Valoriser et développer les outils autour des corpus dans une perspective didactique » depuis 2013, j'ai contribué à une plus grande ouverture vers les applications de corpus en langues étrangères. L'accessibilité aux données linguistiques générées par une requête sur corpus est problématique pour les non linguistes. Dans cette optique, lors de mes travaux en pédagogie numérique sur Scientext Enigma, j'ai trié les données et soigné les aspects visuels (cf. Figure 23). Cela s'est fait en collaboration avec l'ingénieur Hubert Borderiou et le graphiste Tim Catinat au GRENET de l'UGA. Dans cet exemple, l'environnement numérique médiatise le contenu, sans qu'il n'y ait interactivité directe entre l'apprenant.e et le corpus. Cela me semble une étape nécessaire dans l'étayage de l'apprenant.e découvrant de nouveaux outils.

Une approche par les patrons et les motifs me semble productive dans un cadre didactique à cause de sa pertinence en termes de mémoire (cf. Ellis, 2001 ; Fenouillet, 2003 ; Guichon, 2004 et Piolat, 2004). L'être humain lit et retient par séquence sémantique et semble aussi voir « des rapports plutôt que des éléments isolés »

(Gombrich, 1987 : 74). Dans les contextes multimodaux, comme évoqués dans ce dossier, ces deux capacités cognitives sont mises en œuvres simultanément. Selon MacWhinney, les représentations conceptuelles jouent un rôle important dans la mémoire et dans l'interprétation de phrases (MacWhinney, 2001 : 89). À mon sens, le constat de l'existence de la notion de séquence (textuelle) ou de segment (visuel) dans les recherches en compréhension de l'écrit et de l'oral, dans les études sur la réception de l'image, ainsi que dans l'analyse linguistique ont tendance à plaider en faveur d'approches interdisciplinaires, et ceci notamment dans des contextes d'enseignement / apprentissage.

❖ The term *hypothesize* in context
 Samples from a Scientext corpus query: we hypothesize/hypothesized

Link	We + Verb	that-clause
Thus,	we hypothesize	that ATPase YscN could be a suitable candidate
Therefore,	we hypothesized	that alternatively spliced transcripts of CAR may also exist.
Based upon this data,	we hypothesized	that efferent ductules were occluded
Instead,	we hypothesize	that ASW carries a signal
Here,	we hypothesize	that such a knowledge-based function may be used

ENIGMA

Figure 23 : Saisie d'écran de Scientext Enigma [we + HYPOTHESIZE]

3.2. QUELQUES NOTIONS LINGUISTIQUES CONSTITUTIVES DE L'ANGLAIS DES SCIENCES

Dans le présent chapitre, je m'intéresse tout particulièrement à trois phénomènes linguistiques qui sont au cœur du discours scientifique : la causalité, la subjectivité et

l'interpersonnalité. Je pense que le discours scientifique comporte des structures grammaticales permettant d'articuler une terminologie complexe et des qualités subjectives ou interpersonnelles, dont nous avons noté l'importance tout au long de ce dossier. Ces structures peuvent être identifiées grâce à une analyse sur corpus.

La première section est consacrée à la causalité, concept essentiel dans les sciences exactes. J'ai pris comme point de départ pour cette étude des requêtes autour des adverbes, dont *thus*, *thereby*, *therefore*, suivi d'un verbe. Cette construction simple permet souvent de relier des groupes nominaux, qui sont abondants en anglais scientifique. Chaque adverbe s'inscrit dans un contexte spécifique et nous verrons que *therefore* est, en fait, davantage employé dans un contexte de réflexion cognitive que comme description causative.

La deuxième section est consacrée à la subjectivité en anglais scientifique écrit exprimée à travers les lemmes HOPE et WISH. Ces lemmes sont sémantiquement proches, mais leur utilisation semble fortement déterminée selon l'objet du désir.

Enfin, la troisième section explore certaines différences entre le discours scientifique oral et le discours scientifique écrit, car il serait erroné de réduire l'anglais des sciences aux tendances lexico-grammaticales de l'anglais écrit. Une analyse à partir des unités lexicales à haute fréquence du corpus EIIDA (anglais scientifique) montre en effet des éléments linguistiques à visée communicative, qui mettent en relief le caractère interpersonnel de la recherche scientifique.

3.2.1. CAUSALITÉ OU SUBJECTIVITÉ : TROIS ADVERBES D'ARTICULATION

3.2.1.1. CONTRASTES D'EMPLOI

Nous avons vu que l'expression de la causalité est centrale dans le discours scientifique (cf. 1.3.1.). Lors de la conférence *Rencontres Phraséologiques* organisée par le laboratoire LIDILEM en novembre 2013, j'ai présenté un travail autour des collocations *thus/thereby/therefore* + Verbe [HART16a]. Cette structure permet, par exemple, de lier deux groupes nominaux parfois complexes et dans ce sens de tendre vers une stabilité

des notions scientifiques (cf. 1.3.2.). Des requêtes sur Scientext nous dévoilent que les verbes à haute fréquence dans cette structure sont, à part SUGGEST, liés à la causalité : PROVIDE, ALLOW, SUGGEST, REDUCE, INCREASE, MAKE et PREVENT (cf. Tableau 10).

3.2.1.2. CONTRASTES DE FRÉQUENCES

Le deuxième constat que nous pouvons établir est une différence importante de fréquence des trois adverbes puisque *therefore* ($n = 9\,795$) et *thus* ($n = 8\,254$) sont trois fois plus fréquents que *thereby* ($n = 2\,618$), comme indiqué dans le Tableau 10. De plus, environ une fois sur deux, *therefore* (45,3%) et *thus* (58,8%) sont directement suivis d'un verbe, alors que *thereby* (86,8%) est suivi d'un verbe environ quatre fois sur cinq.

Tableau 10 : Fréquence de chaque adverbe et caractéristiques des verbes associés

	Therefore	Thus	Thereby
Occurrences	9795	8254	2618
Occurrences + Verbe	4437 45,3%	4854 58,8%	2272 86,8%
Verbes différents (moyenne d'occurrences par type)	626 (7,1)	694 (7)	380 (5,9)
Occurrences + <i>-ing</i> (Pourcentage en <i>-ing</i>)	89 (0,9%)	2033 (24,6%)	1349 (67,7%)
Lemmes de verbes différents (moyenne d'occurrences par type)	59 (1,5)	377 (5,4)	331 (4)

Nous observons aussi de fortes différences au niveau des collocations verbales. Proportionnellement, la variété des lemmes-verbes est moins conséquente après *thereby*, car la quantité de verbes lexicaux différents est de 380 (moyenne = 5,9), contre 626 et 694 respectivement pour *therefore* (moyenne = 7,1), et *thus* (moyenne = 7) (cf. Tableau 10). Un tri par verbe en *-ing* montre de nouveaux contrastes, car la collocation *therefore* + clause en *-ing* (0,9%) est pour ainsi dire inexistante, bien que la fréquence

globale de *therefore* soit la plus importante. Enfin, les patrons et motifs comportant *thus* ou *thereby* et suivis d'une clause en *-ing* montrent une assez grande variété lexicale.

3.2.1.3. PATRONS LEXICO-GRAMMATICAUX

Ensuite, j'ai analysé de manière plus fine les propriétés syntaxiques des verbes de notre étude en analysant les occurrences les plus fréquentes, mais aussi l'ensemble des verbes selon l'adverbe. Le Tableau 11 montre les cinq verbes les plus fréquents selon l'adverbe ainsi que la forme fléchie la plus fréquente par verbe identifié.

Tableau 11 : Verbes les plus fréquents selon l'adverbe et leur conjugaison

Adverbe	Verbes les plus fréquents	Formes fléchies les plus fréquentes
therefore	suggest (115) use (165) consider (106) represent (80) expect (67)	therefore suggest/s (78) therefore used (53) therefore considered (13) therefore represent/s (44) therefore might be expected (9)
thus	allow (124) make (108) provide (108) reduce (101) suggest (100)	thus allowing (90) thus making (85) thus providing (99) thus reducing (78) thus suggesting (70)
thereby	increase (105) reduce (84) allow (61) prevent (59) provide (46)	thereby increasing (80) thereby reducing (69) thereby allowing (57) thereby preventing (45) thereby providing (45)

Comme nous pouvions l'anticiper suite à l'analyse des fréquences globales, les patrons avec *therefore* contrastent avec ceux des deux autres adverbes par les propriétés sémantiques des verbes et leurs conjugaisons les plus fréquentes. Seul le verbe *suggest* ($n = 115$) se trouve aussi parmi la liste de fréquence (cinquième place) pour l'adverbe *thus* ($n = 100$). SUGGEST, CONSIDER, REPRESENT, EXPECT sont tous des verbes fortement liés à l'activité intellectuelle et les auteurs les emploient souvent avec un

auxiliaire modal. Représentatif de ce phénomène est le verbe CONSIDER, pour lequel nous trouvons la présence de la voix passive et d'une grande variété de modaux :

- therefore should be considered (12)
- therefore can be considered (9)
- therefore may be considered (6)
- therefore must be considered (5)
- therefore could be considered (5)
- therefore might be considered (4)
- therefore has to be considered (1)
- therefore must consider (1)
- therefore may need to consider (1)
- therefore might consider (1)
- therefore should consider (1)

3.2.1.4. CONTRASTES SÉMANTIQUES

Une analyse des autres verbes selon l'adverbe confirme des différences sémantiques. En association avec *therefore*, nous trouvons une série de verbes liés à une décision intellectuelle souvent prise en fonction des données, dont : EXAMINE ($n = 60$), CONCLUDE ($n = 55$), DEVELOP ($n = 35$), et FOCUS ON ($n = 24$). *Therefore* est souvent employé sans recours à un verbe en *-ing* lors de la discussion de la méthodologie, dont :

- we therefore developed a quantitative test for detection
- we therefore focused on the independent tests
- ...therefore no identifications could be performed
- it is therefore important to determine whether

En français scientifique, l'agent est souvent effacé sur le plan linguistique lorsqu'il s'agit de la méthodologie. Par contraste, en anglais scientifique, les démarches méthodologiques sont plus fréquemment introduites par le pronom *we* [HART2014], comme dans les deux premiers exemples ci-dessus (*cf.* Tarone *et al.*, 1981). De manière générale, nous constatons ici des phrases avec *therefore* sans V *-ing* et ayant un agent

humain. *Therefore* sert aussi lors des conclusions pour faire le lien entre les données et le postulat :

- these data **therefore suggest** the loss of candidate tumor suppressor genes.
- which **therefore may be expected to differ** between mice.

Il y a plus de similitudes entre *thus* et *thereby*. Par exemple, parmi les verbes les plus fréquents après les deux adverbes, nous trouvons ALLOW, REDUCE et PROVIDE. Pour chaque verbe, la forme fléchie la plus fréquente est en *-ing* et ceci, à un pourcentage élevé. En effet, après *thereby*, ALLOW ($n = 21$) et PROVIDE ($n = 18$) se trouvent uniquement conjugués en *V-ing*. Une analyse plus fine révèle de nouveau des différences sémantiques.

Au niveau sémantique, *thus*, comme *therefore*, sert à faire un lien entre les données et l'activité scientifique et nous retrouvons SUGGEST ($n = 100$) parmi les verbes le plus fréquents, mais aussi INDICATE ($n = 79$), CONTRIBUTE ($n = 73$), CONFIRM ($n = 69$), RESULT ($n = 57$) and EXPLAIN ($n = 30$). *Thus* introduit aussi, souvent au milieu d'une phrase, des conséquences physiques directes, dont MAKE ($n = 108$), INCREASE ($n = 110$), REDUCE ($n = 101$), RESULT ($n = 57$), LEAD ($n = 61$), PRODUCE ($n = 36$) et CREATE ($n = 27$). La conséquence physique introduite par *thus* peut aussi être une modification des conditions physiques, dont ALLOW ($n = 90$), PREVENT ($n = 79$), FACILITATE ($n = 50$), AVOID ($n = 44$) ou ENABLE ($n = 24$), par exemple dans :

- ...**thus allowing** the nuclear translocation of the subunit
- ...**thus avoiding** information loss.

3.2.1.5. CONTRASTES DE VARIATIONS LEXICALES

Enfin, la similitude entre les huit verbes les plus fréquents dans le patron *thus* et *thereby* + *V-ing* est remarquable, car sept verbes sont identiques (cf. Tableau 12). Seul *make* se trouve en troisième position après *thus* ($n = 85$), mais en 29^e position pour *thereby* ($n = 3$). En miroir, *thereby facilitating* ($n = 30$) se place sixième dans la liste.

S'il existe une forte similitude des constructions avec les deux adverbes en ce qui concerne les verbes les plus fréquents, elle disparaît par la suite. Nous avons vu que *thus*

se trouve avec 626 verbes différents ou une moyenne de 7 occurrences par verbe, alors que la moyenne avec *thereby* + Verbe ($n = 233$) tombe à 2,5 occurrences par verbe. Il est intéressant de noter que le patron *thereby* + Verbe comporte 44 verbes qui ne se trouvent pas en collocation avec *thus*. Ces verbes peu fréquents, dont ANTAGONIZE, DAMPEN, DEPHOSPHORYLATING, EMBED, RECONSTITUTE, UNWIRING ne se distinguent pas par une catégorie sémantique commune, mais par leur rareté. Il semble que les auteurs employant *thereby* cherchent à utiliser un lexique recherché et précis dans un style soutenu, par exemple :

- **thereby facilitating** data exchange and comparison
- **thereby activating** DNA repair

Tableau 12 : Les huit verbes en *-ing* les plus fréquents après *thus* et *thereby*

<i>Thus</i> (ainsi) + V- <i>ing</i> (quantité)	<i>Thereby</i> (de ce fait) + V- <i>ing</i> (quantité)
1. providing (99)	1. increasing (80)
2. allowing (90)	2. reducing (69)
3. making (85)	3. allowing (57)
4. reducing (78)	4. providing (45)
5. suggesting (70)	5. preventing (16)
6. increasing (67)	6. facilitating (30)
7. preventing (63)	7. suggesting (29)

3.1.2.7. PORTÉE DIDACTIQUE

Cette étude confirme la validité de la notion de colligation [*thus* / *thereby* + V-*ing*] et l'importance de la notion de collocation, car nous voyons que certains adverbes s'associent plus fréquemment avec certains verbes. Cette association avec certains verbes semble indiquer une tendance sémantique. Par exemple, *thus* est davantage associé avec des conséquences physiques directes. Au plan linguistique, une comparaison de cette structure avec l'anglais général ou le français scientifique ferait

l'objet d'une étude intéressante. Sur le plan didactique, cette structure, pourtant simple, pose de nombreux défis, dont la maîtrise du verbe en *-ing* et la particularité que ces trois adverbes ne sont pas interchangeables.

3.2.2. SUBJECTIVITÉ : HOPE ET WISH

3.2.2.1. CONTRIBUTION AU DÉBAT AUTOUR DU MODALISATEUR HOPEFULLY

En octobre 2013, j'ai présenté « Niveau d'acceptabilité dans les articles de recherche scientifique » à la conférence organisée par l'ILCEA (*Institut des langues et des cultures d'Europe et d'Amérique*) à l'Université Stendhal, l'objectif étant de contribuer au débat autour d'un seuil d'acceptabilité entre « la tolérance aux imperfections et la condamnation des outrances ». Initialement, j'ai voulu traiter de la notion de langage « non standard » présent dans les corpus d'anglais, car plusieurs personnes de mon entourage suggéraient que l'utilisation de corpus auprès de publics d'apprenant.e.s pourrait avoir pour résultat la transmission d'erreurs lexicales ou grammaticales. Les données linguistiques de corpus permettent cependant d'évaluer l'évolution d'une langue et de cerner les transformations linguistiques. Par exemple, l'acceptation du modalisateur *hopefully* est objet de débat depuis son regain de popularité dans les années 1960.

L'*American Press Stylebook* a accepté son usage moderne le 17 avril 2012 : « *Hopefully, you will appreciate this style update, announced at #aces2012. We now support the modern usage of hopefully: it's hoped, we hope.* » Selon le *Merriam-Webster* dictionnaire en ligne, cet adverbe, dans son emploi de modalisateur, permet d'apporter une appréciation sur le contenu d'une phrase, au même titre que *frankly, interestingly, clearly, luckily, unfortunately*. Une requête sur Scientext, nous montre que cette expression est effectivement employée par des scientifiques, par exemple :

- Hopefully this article will spur new studies...
- Hopefully ovarian tumorigenesis will provide...

Cette première démarche m'a amenée vers une piste plus fascinante, celle de la verbalisation du souhait, car à travers ces requêtes de Scientext, nous découvrons que

l'espoir auctorial (*hope*) ou le souhait auctorial (*wish*) sont des notions acceptables par des éditeurs en sciences exactes.

3.2.2.2. DIFFÉRENCES SÉMANTIQUES DE HOPE ET WISH

Selon le dictionnaire en ligne *Merriam-Webster* (2014), les deux lemmes HOPE and WISH sont similaires, mais présentent différentes notions de désir. WISH est issu du vieil anglais *wyscan*, apparenté à l'ancien haut allemand *wunsken*, qui trouve ses racines, comme le verbe *win*, dans le sanscrit *vanoti* (*il aspire à*). HOPE vient du vieil anglais *hopian* apparenté à l'ancien moyen haut-allemand *hoffen*, ce qui signifie *to hope*. Son utilisation archaïque est proche de *trust*, mais HOPE est actuellement utilisé dans le sens de *espoir, espérance, chance* ou du verbe *espérer*, englobant la notion d'anticipation. Le lemme WISH renvoie aux qualités actives du sanskrit *vanoti* ; nous retrouvons dans sa définition l'idée de *souhait, vœu, désir* ou des verbes *vouloir* et *souhaiter* (Larousse, 2015).

3.2.2.3. RÉSULTATS DES REQUÊTES DE HOPE ET WISH

Des requêtes sur Scientext ont produit des données assez importantes pour une analyse de verbes peu associés à l'anglais des sciences, car constituées de 247 occurrences du verbe HOPE, 312 occurrences du nom HOPE, 118 occurrences de l'adverbe *hopefully* et enfin, 616 occurrences du verbe WISH. Le lemme WISH n'a pas été trouvé dans des formes autres que la forme verbale. Ces données ont ensuite été examinées individuellement pour identifier leur contexte et la signification, les données non pertinentes ont été ensuite supprimées [HART16b].

Seuls les énoncés dans lesquels les auteur.e.s décrivaient leur propres *hopes* ou *wishes* ont été inclus dans la présente étude. Ainsi, les éléments du corpus qui comprennent les commentaires de plusieurs éditeurs ont été exclus. En outre, tous les textes trouvés dans les sections de remerciements (*we wish to thank...*) sont exclus, ainsi que des contextes liés aux sentiments des participant.e.s à une étude :

- people hope for a means of understanding the issue

- GPs said patients often wished radiography because of advice from health care providers.

J'ai retenu 150 occurrences du verbe HOPE, plus 165 du substantif HOPE, et 165 occurrences du verbe WISH.

Les deux verbes sont principalement en collocation avec *we*, et donc très liés au positionnement des auteur.e.s. Nous trouvons 105 occurrences de *we hope* ou *we hoped* et un total de 139 occurrences de *we WISH*, WISH se trouvant aussi avec *one* ($n = 27$) et *researchers* ($n = 11$). Le substantif *hope* se trouve à 19 reprises dans la séquence : *it is our hope that* ($n = 19$).

3.1.3.6. CONTRASTES CONTEXTUELS

Ensuite, j'ai évalué les cotextes des lemmes HOPE et WISH afin de cerner l'objectif du souhait. J'ai trouvé quatre grands axes : être utile pour la société ou la recherche, encourager la recherche sur le sujet, se projeter dans l'avenir et traiter des questions de méthodologie. Comme nous pouvons le voir dans le Tableau 13, une analyse sémantique des données montre que WISH et HOPE sont employés dans des contextes contrastés. L'objet principal du désir est de servir la société en général :

- *hopefully we can offer a solution to alleviate or delay the occurrence of vein graft disease*
- *The results in this paper give hopes for finding a cell-based autologous biological treatment* » ou « *in hopes of improving health outcomes.*

Sans surprise, beaucoup d'énoncés renvoient à l'avenir, ou servent à encourager la recherche en général :

- *We hope this report will stimulate renewed interest in the field*

Ou en référence à leur propre recherche :

- *We hope to pursue this line of investigation in the future*

Enfin, la forte présence de références à la méthodologie ($n = 134$), et ceci pour les quatre collocations (HOPE en tant que substantif, verbe ou adverbe, ainsi que le verbe WISH),

mérite une catégorie à part, bien que plusieurs occurrences peuvent aussi être catégorisées parmi le souhait d’être utile à la société ou la recherche :

- Our hope is to facilitate the compilation of data...
- We hope to extend this approach...
- With this new model, we hope to predict....

Tableau 13 : Nombre d’occurrences par catégorie grammaticale et par contexte

	Hope_N	Hope_V	Hopefully	Wish_V	Total
Utile pour la société ou la recherche	96	72	27	14	209
Encourager la recherche	24	6	23	122	175
Projection pour l’avenir	9	33	28	5	75
Méthodologie	36	39	35	24	134
Total	165	150	113	165	593

3.1.3.7. QUALITÉ SUBJECTIVE

L’article de recherche scientifique est parfois perçu comme étant objectif. Pourtant, les requêtes sur Scientext montrent la présence de la subjectivité exprimée par ces deux lemmes. L’énoncé « while lower than what could be wished for, these results are more encouraging... », montre un engagement presque personnel des auteur.e.s. Il semble que les auteur.e.s emploient le lemme *HOPE*, sous forme de verbe, substantif ou adverbe, notamment pour évoquer l’espoir de servir le public et le verbe *WISH* en particulier pour encourager d’autres chercheurs à approfondir leurs conclusions en entreprenant d’autres études similaires.

3.2.3. NEUTRALITÉ OU QUALITÉS INTERPERSONNELLES

3.2.2.1. DIFFÉRENCES ENTRE LE GENRE ÉCRIT ET ORAL EN ANGLAIS DES SCIENCES

L'anglais oral n'est pas une simple reproduction de l'anglais écrit oralisé. Influencé par une contrainte spatio-temporelle et la présence d'interlocuteurs/trices, l'anglais oral comporte des différences linguistiques importantes. À partir d'un corpus de transcriptions de présentations orales et de leurs actes, Carter-Thomas et Rowley-Jolivet (2001) identifient cinq différences syntaxiques entre les deux genres, dont une plus grande fréquence dans le discours écrit de la voix passive (*the study was conducted*) et de l'extraposition (*it may be possible*). Dans le discours oral, nous trouvons davantage d'énoncés existentiels (*there is*), d'inversions du verbe et du sujet syntaxique (*here is a picture of one of the devices*) et de pseudoclivées (*what we did here was*). Ces formulations permettent à l'auteur.e, à l'orateur ou à l'oratrice de mieux gérer l'introduction des informations et de souligner leur importance dans des contextes contrastés.

3.2.2.2. PORTÉE DIDACTIQUE

Motivées par leur travail, Emmanuelle Esperança-Rodier et moi-même travaillons dans le cadre du projet EIIDA (cf. 3.1.1.2.), sur les différences entre discours oral et écrit en anglais des sciences dans un objectif didactique. Ces recherches nourrissent aussi le projet *Corpus4All*, financé par l'UNR-RA en 2015 (cf. 2.1.3.6.), car l'apprenant.e semble adopter trop souvent des structures typiques de l'anglais scientifique écrit dans un contexte d'oralité :

Non-native speakers seem to differentiate less between written and oral modes of scientific communication, tending to overuse, in oral presentations, structures that are typical of writing such as the passive, and not to use structures that are typical of speech such as pseudo-clefts and inversion. In other words, they lack pragmatic competence. As a result, this makes their discourse much more difficult for the listener to comprehend and process (Carter-Thomas & Rowley-Jolivet, 2001 : 13).

3.2.2.3. RÉSULTATS D'UNE ANALYSE DE FRÉQUENCE LEXICALE

Une première analyse par simple fréquence montre quelques différences, dont la présence importante de *that*, *so* et *can* dans le corpus oral d'anglais des sciences (cf. Tableau 14). Dans cette perspective, nous avons aussi identifié une quarantaine d'unités lexicales dont la fréquence est fortement contrastée à l'oral et à l'écrit et pour lesquelles nous avons constaté des difficultés d'appropriation chez les apprenant.e.s (cf. Annexe VII), dont *actually*, *any*, *as*, *due to*, *further*, *given*, *if*, *important*, *per*, *while*, *within*, ainsi que les verbes modaux.

Tableau 14 : Dix mots les plus fréquents à l'oral et à l'écrit du corpus EIIDA, anglais scientifique

Corpus oral : Ordre de fréquence	Unité lexicale	Corpus écrit : ordre de fréquence	Fréquences orales/écrites
1	of	1	1078/2362
2	to	2	829/1478
3	that	5	792/597
4	in	3	685/1267
5	this	6	465/387
6	so	46	363/32
7	can	12	269/147
8	on	8	250/318
9	for	4	243/637
10	at	10	239/296
12	as	7	192/366
18	by	9	72/306

3.2.2.4. CONCLUSIONS AUTOUR DE THAT SELON LE GENRE

Une première analyse de *that* montre une forte présence de pronoms personnels après *that* dans le corpus oral et, par contraste, de verbes lexicaux avant *that* à l'écrit, parmi

lesquels AFFECT, ALLOW, ASSUME, CAUSE, DISPLAY, ENABLE, OCCUR, POSSESS, RESULT, SHOW, USE. À l'oral, nous trouvons les pronoms *we* ($n = 124$), *I* ($n = 29$), *you* ($n = 24$) ou 177 occurrences de motif [pronom personnel + *that*] pour 792 occurrences de *that* (22,3%). Les séquences lexicales *that we can* ($n = 13$) et *that we need to* ($n = 13$) se retrouvent dans des phrases comportant souvent une qualité subjective :

- I have said that we need to move and that we can move from descriptive to predictive.
- I want to just bring to everyone's forefront attention that we can't sit at our lab benches anymore
- And we think that to do that we need to manage stormwater
- The volume reduction approach I think is the way that we need to go.

Dans ces exemples tirés du corpus oral, il apparaît que l'orateur ou l'oratrice est fortement impliqué.e à travers l'utilisation de *we*, mais aussi du pronom personnel *I*. Par contraste, à l'écrit, nous retrouvons plus fréquemment des verbes lexicaux neutres en cooccurrence avec *that*, dont SHOW ($n = 65$), INDICATE ($n = 27$), FOUND ($n = 16$), le sujet étant typiquement le nom d'un autre chercheur, une table ou figure, ou un sujet neutre, tel que *results, analysis, data, the study*, voire la voix passive :

- This project shows that mixing the land uses within a regional area enables highly beneficial and more sustainable relationships to be put in place, in other words they allow a permaculture approach to be applied.
- However, it has been shown that the Bluebelt approach provides considerable cost savings.
- Perhaps more importantly, it was found that there were inherent difficulties in the spinning process.

3.2.2.5. FORTE PRÉSENCE DE SO À L'ORAL

L'adverbe *so* se trouve dans une grande variété de séquences, signifiant alternativement : une certaine manière (*does so with a constant concentration*), un certain degré (*be so dense, especially so for Australia*) ou une conséquence (*was optimised so it did not*). *So* est surtout remarquable, car 11,3 fois plus présent à l'oral qu'à l'écrit, avec une distribution intéressante pour les quatre séquences indiquées dans le Tableau 15.

Tableau 15 : So dans le sous-corpus EIIDA d'anglais scientifique

	Écrit	Oral	Ratio
<i>so</i>	32	363	1 : 11,3
<i>so that</i>	11	17	1 : 1,5
<i>so what</i>	0	24	N/A
<i>so we</i>	0	46	N/A

Les trois séquences indiquées avec *so* dans le Tableau 15 montrent que *so that* est proportionnellement plus fréquent à l'écrit qu'à l'oral, mais que *so what* et *so we* prédominent à l'oral. *So what* ($n = 24$) est suivi par un pronom dans 19 des occurrences, dont *I* ($n = 5$), *we* ($n = 12$) et *you* ($n = 2$) dans des phrases qui expliquent les étapes de l'étude (*so what we did*), de la présentation (*so what I'd like to talk about*) ou des visuels (*so what I'm showing you here*). En d'autres termes, pour *so what*, comme pour *so we*, les différences d'utilisation relèvent de nouveau de la qualité interpersonnelle et immédiate de la présentation orale, par contraste avec le discours objectif de l'écrit. À l'écrit, l'extraposition (*it is important to*) permet d'éviter une référence directe à l'auteur.e dans les exemples ci-dessous et d'orienter les lecteurs et lectrices vers des concepts :

- It is important to size filters so that they are large relative to their impervious catchments
- It was proposed to amend the Masterplan so that a bioregional stormwater harvesting.

Par contraste, la séquence *so that* à l'oral est suivie presque une fois sur deux ($n = 8$) par un pronom personnel, dont *we* ($n = 6$) ou *you* ($n = 2$) :

- We modeled it very infinitely so that we could really even look at inches of elevation differences
- That tool must be GIS based so that you've got uh access immediately to indi- uh uh individual real land use
- We started off with a bacterial inhibition test [...] and found no bacterial inhibition so that got us thinking it was at least a possible start.

Nous avons déjà constaté que *so we* ($n = 46$) constitue 12,7% des séquences avec *so* à l'oral. Une analyse plus fine montre que *we* réfère presque toujours à l'équipe de recherche ($n = 42$), les quatre exceptions étant inclusives du public en évoquant un support visuel :

- This gives you a better view of the overall system so we have the transferred-arc cathode here
- And so we see here two standards plotted

Cette construction à fonction interpersonnelle autour de *we* et dont l'objectif est d'attirer l'attention vers les supports visuels est plus fréquente à l'oral du fait de la présence du public, mais aussi grâce à la présence importante de supports visuels. Par contraste, à l'écrit, nous retrouvons des expressions dépersonnalisées (*as illustrated/seen in figure 6, the figure shows, it is clear from this figure, electron micrographs presented in figure 5 show examples...*). Je constate qu'en anglais des sciences, une approche pragmatique, par sa contextualisation des énoncés, semble plus pertinente lors de l'analyse des genres oraux, que pour le genre écrit de l'article de recherche.

3.2.2.6. RECOURS AUX PRONOMS PERSONNELS

Le phénomène de recours à un pronom personnel à l'oral se répète avec *so what*, où nous ne trouvons que trois questions rhétoriques :

- So what are the forcing events that then lead to an image
- So what type of biochemistry are bacteria exerting on on [sic] the organic matter

Les autres séquences *so what* ($n = 21$) étant suivies par le pronom *we* ($n = 12$), *I* ($n = 5$), *you* ($n = 2$), ou bien *this paper* ($n = 1$) ou *my student* ($n = 1$).

L'auxiliaire modal *can* n'est employé avec aucun pronom personnel à l'écrit ($n = 147$), mais à l'oral *can* ($n = 269$) est combiné avec un pronom personnel presque trois fois sur quatre ($n = 198$, 73,6%). Nous retrouvons une interaction directe entre l'orateur ou l'oratrice et le public, notamment avec *you can* ($n = 121$). L'attention est attirée vers un aspect des supports visuels qui soutient le développement par la séquence

interactionnelle *you can see* ($n = 68$), dont *you can see that* ($n = 13$) ou *you can see here* ($n = 5$) :

- You can see that the magnitude of floods increases quite a fair bit
- You can see that you get the greatest benefit from the lower amounts
- You can see here for example how the LIDAR detail is actually identified

La séquence la plus fréquente comportant *we* est aussi employée dans l'objectif d'attirer l'attention du public vers un aspect du support visuel : *we can (actually) see* ($n = 8$). Par contraste avec *you*, *we can* ($n = 69$) est employé avec une plus grande variété de verbes lexicaux, essentiellement pour évoquer des démarches de l'équipe de recherche, dont *PUT* ($n = 4$), *DESIGN* ($n = 2$), *GET* ($n = 2$), *RUN* ($n = 2$), *MAKE* ($n = 2$), *MOVE from* ($n = 2$), *MEASURE* ($n = 2$). Comme pour *can*, les adverbes *so* ($n = 9$) et *actually* ($n = 6$) soulignent l'intérêt des capacités techniques ou méthodologiques de l'équipe :

- We can actually test instrumentation
- Under certain circumstances we can design for performance and other objectives

La grande variété d'emploi des pronoms apporte une finesse dans l'interaction avec le public. Par exemple, dans l'exemple ci-dessous, l'auteur emploie d'abord un *we* inclusif suivi de *here* pour évoquer le graphique du support visuel, puis, *we can keep track* (nous pouvons suivre) qui se rapproche davantage des travaux de l'équipe :

- We've got a rainfall graphic here so we can keep track and see what the rainfall was...

Enfin, nous trouvons quelques occurrences avec *I* ($n = 6$) et *one* ($n = 2$). En anglais, il y a moins de recours au pronom personnel *one*, l'étymologie et l'usage dans les deux langues étant contrastés [HART14a]. Ici, nous trouvons deux occurrences avec *can*, dont : *one can think in terms of potentiality*. Par cette utilisation, il y a proposition d'une approche possible, sans l'imposer de force à son public.

3.2.2.7. ASPECTS PRAGMATIQUES

La moitié des occurrences de *I* ($n = 3$) sont employées par l'orateur ou l'oratrice lors de problèmes techniques (*if I can get down here and rotate it around*), mais aussi en incipit pour créer un lien personnel avec le public, qui digère un déjeuner sans doute pris

ensemble dans le cadre de la conférence, avant de situer l'importance de sa recherche et celle de son équipe dans un contexte international :

I've only had one glass of wine at lunch so I'm hoping I can get through okay uh my co-authors of this paper and indeed a lot of other colleagues of us are involved in a really large research project which is uh trying to think of how we can restore the health of urban streams + uh + and + as part of that study we're hoping to address uh some really big questions in urban hydrology and one of those big questions is what's the impact of source control measures or best maintenance practices?

Cet extrait comporte *uh* et des pauses, symbolisées par +, qui sont des éléments propres aux transcriptions des présentations orales. Sur le plan linguistique, ces éléments permettent de mieux évaluer la production, voire la démarche cognitive de l'acte de présenter oralement. Sur le plan didactique, ces éléments et des éléments comme *okay* ou des phrases non standard peuvent être retirés ou évités afin de proposer des exemples accessibles aux apprenant.e.s sans surcharge. Une analyse autour de ces trois unités lexicales (*that, so, can*) laisse à voir un discours oral où l'interpersonnel et la subjectivité jouent un rôle important. La présentation orale devant les pairs, dans un contexte presque intime en comparaison avec la publication d'un article de recherche, semble induire un discours plus direct. Il semblerait que les qualités d'objectivité ou d'impersonnalité du discours écrit ne sont pas, en fait, aussi tranchées dans un contexte d'oralité. En d'autres termes, nous en venons au constat que l'anglais oral n'est pas une réplique de l'anglais écrit dans cette langue fortement codée qu'est l'anglais de spécialité.

RÉSUMÉ DE LA TROISIÈME PARTIE

Cette partie retrace les étapes de la création de quelques corpus à caractère spécialisé, souvent établis dans un cadre de travaux d'équipe et souvent dans un esprit collégial et dans l'intérêt de faire avancer les connaissances à la fois quantitatives et qualitatives. J'y explore les outils de traitement sur lesquels je m'appuie pour approfondir ma recherche, et dans ce sens, il s'agit d'un travail interdisciplinaire, car bien que la recherche en TAL (Traitement Automatique des Langues) soit un domaine de recherche spécifique, il existe des intérêts communs avec les linguistique de corpus.

Mes propres recherches autour de l'anglais des sciences, et notamment sur la subjectivité, s'inscrivent dans les travaux en linguistique du discours académique de l'anglais et du français. Mon approche par la phraséologie, bien plus que par la lexicologie, permet de cerner des spécificités du discours scientifique en anglais contextualisé, car le contexte, qu'il s'agisse des articles de recherche ou des conférences scientifiques, impacte les besoins linguistiques et les configurations lexico-grammaticales. Les analyses linguistiques à partir de corpus à visée spécifique permettent de dévoiler la présence auctoriale notamment à travers des constructions comportant des pronoms personnels, ainsi que par une sélection de lemmes, directement liés à la subjectivité, dont HOPE et WISH. L'analyse du discours oral met en évidence l'importance du facteur humain sur le discours et ouvre de nouvelles pistes autour des codes discursifs de la subjectivité.

PERSPECTIVES

Dans ces dernières pages, j'aborde les projets et les thèmes de ma recherche à venir. Engagée dans des démarches collectives d'animation de la recherche, je participe à l'organisation de journées d'études en recherche pure et pour leur rayonnement auprès des acteurs et des actrices sur le terrain. Membre de jury de thèse et de comités de sélection, je participe aux travaux collectifs d'évaluation. Mes activités de relecture avant publication, de cotutelle de thèse avec Françoise Raby depuis trois ans et de la direction récente de thèse (avec dérogation d'HDR) sont autant d'opportunités pour me préparer à cette responsabilité à diriger des recherches (*cf.* Curriculum Vitae).

Dans le cadre de travaux autour de la linguistique de corpus, Emmanuelle Esperança-Rodier (LIG) et moi-même réalisons une étude sur corpus dans le cadre du projet EIIDA (*cf.* 3.1.1.2.) qui cerne les aspects discursifs interpersonnels de l'anglais oral des sciences en contexte de conférence, par comparaison avec la forme écrite. Nous prêtons une attention particulière aux expressions qui sont source de difficulté pour les apprenant.e.s. Un ouvrage collectif est en projet sous la direction de Shirley Carter-Thomas, Grégory Furmaniak et Marie-Paule Jacques. J'émetts l'hypothèse que l'aspect discursif interpersonnel est fortement influencé par la culture. Cet aspect est d'autant plus difficile qu'il ne peut pas être simplement traduit d'une langue à une autre, les codes étant très dépendants du genre discursif. Cette recherche aura une portée didactique, car le passage de l'écrit vers l'oralité est souvent problématique pour un public LANSAD.

Le thème de l'exploitation de corpus dans une démarche didactique est aussi au cœur de la conception de la plateforme Corpus4All (*cf.* 2.1.3.6.) et de la journée d'étude « Valoriser et développer les outils autour des corpus dans une perspective didactique », organisée annuellement sur le Campus Sud-Rhône-Alpes à Valence (UGA) depuis 2014. Cristelle Cavalla et moi-même travaillons sur un projet de publication collective autour

de ce thème et un appel à contribution sera lancé pour la revue LIDIL en 2016. Il reste des interrogations autour des possibilités de rendre de grands concepts de l'anglais académique ou des sciences accessibles à un public LANSAD. Mes perspectives de recherche sur ces thèmes ont des applications didactiques dans ce sens. Je souhaite contribuer aux connaissances au niveau textuel, mais aussi métalinguistique.

L'approfondissement de ma recherche autour du lien entre la formation en langues étrangères et les compétences langagières (cf. 2.1.1.) s'effectuera par une nouvelle étude sur le terrain. J'émetts l'hypothèse selon laquelle certains facteurs impactent plus ou moins la réussite de l'étudiant.e en contexte anglophone authentique et que ces facteurs négatifs pourraient être minorés par une information ou par une préparation préalable. Pour cela, je mets en place une étude basée sur des entretiens avec les équipes enseignantes et administratives en contact avec les étudiant.e.s francophones en mobilité aux États-Unis, ainsi qu'auprès des étudiants *in situ* dans au moins deux établissements, dont Louisiana State University et l'University of San Diego. Un financement Erasmus+ est en cours dans ce sens en collaboration avec la Vice-Présidence des Relations Internationales à l'UJF.

Enfin, dans la continuité des travaux sur les corpus linguistiques et les compétences langagières, je souhaite contribuer aux recherches dans le processus de l'écriture de textes, dont les textes en anglais de spécialité par les non anglophones apprenant.e.s et expert.e.s dans leur discipline [HART11a]. Quelles sont les démarches pour entrer dans le texte, quels en sont les étapes, les obstacles, les stratégies, les différences textuelles avec les productions d'expert.e.s anglophones, les codes saillants ? Comme pour mes travaux sur la compréhension de l'oral, cela passe par les théories de l'acquisition du langage et par une analyse de discours. Un travail à partir de corpus serait adapté à cette recherche et les applications didactiques pourraient être développées à partir d'une approche multidisciplinaire, dont linguistique, épistémologique, didactique et informatique. Mes futures recherches s'inscrivent clairement dans les travaux d'équipe menés au laboratoire LIDILEM depuis plusieurs années autour de l'écrit, mais aussi du laboratoire LAIRDIL et de l'University of San Diego avec lesquels je garde des liens de collaboration.

CONCLUSION GÉNÉRALE

Ce dossier de synthèse retrace les grandes étapes en linguistique de corpus et en didactique depuis les travaux de Palmer (1933) autour des collocations. Ces travaux mettent en relief l'importance du contexte, du genre et de la communauté de discours sur les codes linguistiques. En anglais des sciences, la causalité et le positionnement font appel à des patrons à portée des apprenant.e.s. Certaines constructions relevant de la métaphore grammaticale sont souvent plus complexes à la fois dans leur identification et dans la possible mise en œuvre d'applications didactiques.

L'approche socio-constructiviste communicationnelle et actionnelle à l'anglais des sciences qui sous-tend ce dossier met en relief et contextualise les spécificités linguistiques des communautés discursives ciblées. La phraséologie permet un ancrage linguistique fort à la hauteur des demandes et besoins identifiés. La didactique des langues étrangères ne peut pas être séparée du contenu disciplinaire, surtout lorsque la vaste majorité du public concerné étudie l'anglais depuis au moins dix ans et se trouve confronté à des besoins académiques ou professionnels importants. Le dispositif de validation détaillé dans ce dossier (*cf.* 2.1.1.) s'inscrit dans un contexte international d'évaluation et de certification en langues où les intérêts financiers des entreprises privées ne doivent pas se substituer à une réponse pertinente aux multiples besoins professionnels en matière de langues étrangères.

À mon sens, une convergence entre les domaines de la didactique, de la linguistique de corpus et de l'anglais de spécialité offre une réponse à ces besoins académiques et professionnels majeurs. Comme nous l'avons vu au fil des pages, les recherches sur corpus informent les ressources pédagogiques en langues, dont les dictionnaires et manuels, mais aussi nos connaissances des langues académiques et des langues de spécialité. Les aspects discursifs de l'anglais académique et l'anglais des sciences, dont la causalité (*cf.* 1.3.1.), le positionnement (*cf.* 1.3.3.), la métaphore grammaticale (*cf.*

1.3.2.), les différences importantes entre les genres oral et écrit (cf. 3.2.3.3.) sont abordés dans ce dossier. Seule une recherche approfondie en langues de spécialité, nous permettra d’offrir des formations adaptées aux besoins des domaines spécifiques de niveau avancé.

Comme souligné par le Conseil de l’Europe (cf. 2.1.), les enjeux de l’enseignement des langues étrangères concernent le monde professionnel, mais aussi la cohésion sociale. Les langues, les littératures et les cultures anglophones dominent l’enseignement des langues étrangères dans les établissements du secondaire, au détriment d’autres langues et cultures, créant ainsi un déséquilibre. La place de l’anglais comme *lingua franca* ne devrait pas avoir pour effet la carence d’autres contextes socio-culturels et dans ce sens, l’enseignement des langues mérite d’être repensé. Il faudrait à la fois ouvrir vers une plus grande variété socio-culturelle et soutenir la recherche et la formation en langues de spécialité dès le secondaire, afin d’équiper la jeunesse avec les outils linguistiques, une ouverture culturelle et des bonnes pratiques d’apprentissage afin qu’elles et ils puissent s’accomplir et contribuer à un monde complexe.

Laura M. Hartwell, Valence, décembre 2015

INDEX DES TRAVAUX

- [HART03a] : 20
- [HART05a] : 65
- [HART07a] : 65, 102, 108
- [HART07b] : 65, 102
- [HART09a] : 102, 103, 108
- [HART10a] : 102, 103
- [HART10b] : 102, 112
- [HART10c] : 95, 102, 111
- [HART10d] : 95, 102
- [HART10e] : 102, 111
- [HART11a] : 150
- [HART11b] : 88, 102
- [HART11c] : 20, 61, 65, 101, 111
- [HART11d] : 28, 76, 112
- [HART11e] : 62
- [HART12a] : 21, 88
- [HART13a] : 20, 72, 88, 101, 111
- [HART13b] : 27, 50, 54, 88, 123
- [HART13c] : 21
- [HART13d] : 28, 74, 76
- [HART13e] : 111
- [HART14a] : 88, 116, 119, 125, 147
- [HART14b] : 21, 74, 76
- [HART14c] : 28
- [HART16a] : 132
- [HART16b] : 114, 139

INDEX DES AUTEURS

Arnaud.....	88	Coxhead.....	86, 87
Bachman.....	64, 101, 102	Dahl	28, 58, 60, 114, 116
Bak Sienkiewicz	49	Daley.....	36
Bakhtine	32, 113	Davidson.....	64, 85
Banks	29, 58	Davies	94
Barber.....	44	Décuré	78, 190
Bertin.....	26, 70, 71	Demaizière	85
Biber	42, 48, 57	Deyrich	26, 31, 62, 70, 72, 92, 161
Blattes	50, 87, 93, 95	Dörnyei	67
Blin.....	68, 69, 72	Douglas.....	64
Boch.....	60, 88, 103, 117	Dunning.....	123
Boulton.....	28, 31, 92	Dwyer	59, 135
Bourigault.....	124	Ellis.....	67, 91, 123, 130
Braud	24	Engeström	68
Bruner.....	62	Erickson	30
Buck.....	103	Falaise, voir aussi HART13c..	21, 88, 117, 118
Carter-Thomas	45, 52, 113, 141, 142	Feak	46, 58, 129
Cavalla	84, 162	Fenouillet.....	130
Cazade	63, 95	Firth	34, 35, 60
Cazden.....	61	Fløttum	28, 48, 58, 60, 114, 116
Chamot.....	67	Follette	119
Chapelle.....	63	Francis	42, 48, 91, 123
Chaudron.....	103	François	48
Cobb	92	Fries-Verdeil	20
Conrad.....	42, 48	Fulcher.....	64
Conseil de l'Europe	64	Gass	63
Cortes	42		

Gee.....	62	Jans.....	50, 87, 95
Gillette	59, 135	Jones	36, 64
Gilner	86	Kennedy	96
Gilquin.....	51	Kilgarriff.....	90
Gjesdal	58, 60, 116	Kinn	28, 58, 60, 114, 116
Gledhill.....	41	Kraif.....	117
Gombrich	130	Krashen	30
Granger	31, 41	Leech.....	35, 48
Gravé.....	70, 71	Legallois.....	41, 42, 48
Gray.....	57	Lewis	88
Greenbaum	35, 48	Loiseau	48
Grosbois.....	190	Mackey.....	63
Gross	51	MacWhinney.....	131
Grossmann.....	28, 58, 60, 88, 117	Magué	119
Guichon.....	70, 130, 160	Maingueneau.....	45
Halliday	36, 52, 54, 55, 56	Maniez	40, 41, 88
Hatier	44	Matthey	67
Heiden.....	119	McAdam.....	90
Henderson	59	McNamara	64, 102
Hilton	67	Meros.....	101
Hirsch	87	Million	60
Hoey.....	40, 43	Millot.....	24
Horn	86	Milton.....	121
Hunston	42, 48, 91, 123	Moon.....	88
Husák	90	Narcy-Combes, J.-P.	65, 67, 70, 85
Hyland.....	48, 57, 58, 60, 121	Narcy-Combes, M.-F.	61
Icke.....	59, 135	Novakova	49
Isani.....	62	O’Malley.....	67
Jacques, voir aussi Hart12a, 13c, 14a		O'Donnell	91
	21, 44, 88, 116, 119, 125, 147	Osborne	31, 128
Jalkanen	68, 69, 72	Palmer	34, 60, 64, 86

Paquot.....	31, 41	Simpson-Vlach.....	123
Payne.....	51	Sinclair.....	36, 37, 41
Pecman.....	43, 47	Skinner.....	30
Perrin.....	25, 26	Song.....	102
Petit.....	26	Stubbs.....	37, 39, 60
Pincemin.....	119	Svartvik.....	35, 48
Piolat.....	103, 130	Swales... 44, 45, 46, 58, 59, 60, 114, 129	
Poibeau.....	115	Tarone.....	59, 135
Poudat.....	119	Tran.....	44
Quirk.....	35, 48	Tutin.....	28, 39, 40, 44, 58, 59, 60, 88, 117, 118
Raby.....	67	Tyne.....	31, 92
Reppen.....	42, 48	Upjohn.....	50, 87, 95
Rinck.....	60, 88, 117	Vandergrift.....	101
Römer.....	91	Véronique.....	67
Rouquet.....	117	Vold.....	60, 116
Rowley-Jolivet.....	141, 142	Vygotski.....	30
Rundell.....	90	Wagner.....	101
Rychlý.....	90	Williams.....	60, 118
Salehzadeh.....	103	Wozniak.....	24
Sancho-Guinda.....	57	Yan.....	44
Sarré.....	24	Zou, voir HART13e et 14b.....	74, 76
Saville.....	64		

LISTE DES ANNEXES

Annexe I : Document explicatif du dispositif de validation du niveau « indépendant / B2 » en anglais de spécialité à l'UJF

Annexe II : Modèle des compétences tutorales pour le tutorat en ligne synchrone (Guichon, 2011 : 100)

Annexe III : Structure expérientielle de la mise en place des tâches (Deyrich, 2011 : 130)

Annexe IV : Carte heuristique des collocatifs (Cavalla, 2012)

Annexe V : Extrait de données informatiques « *Enigma Plus* »

Annexe VI : Grille d'évaluation pour la présentation orale en M2

Annexe VII : Unités lexicales du corpus EIIDA identifiées comme étant fréquentes et problématiques par les apprenant.e.s francophones

Annexe I : Document explicatif du dispositif de validation du niveau « indépendant / B2 » en anglais de spécialité à l'UJF



Épreuves de niveau « indépendant » en anglais de spécialité (B2 du CECRL)

Ce document vise à expliciter les compétences langagières développées dans les programmes de master à l'Université Joseph Fourier de Grenoble, telles que définies par le **Cadre Européen Commun de Référence en Langues (CECRL)**. Elles correspondent à des recommandations *études* de l'ALTE (Association of Language Testers in Europe). Les compétences ciblées comportent un intérêt professionnel pour les scientifiques et préparent à l'utilisation de l'anglais dans le cadre de son futur emploi pour la compréhension de rapports ou d'autres documents écrits, la présentation de ses travaux en réunion et la synthèse d'informations orales.

Vous trouverez dans le tableau ci-dessous les savoirs de l'étudiant(e) associés à la validation de chaque compétence, comme mis en place par le **Service des Langues** de l'UJF et indiqué sur le **relevé de notes de Master2**, sous le titre « Compétences en langue étrangère ».

Compréhension de documents écrits à caractère scientifique	Présentation orale en lien avec sa discipline scientifique	Prise de notes à partir de documents oraux à caractère scientifique
<p>EST CAPABLE de :</p> <ul style="list-style-type: none"> - parcourir un texte pour retrouver l'information pertinente et d'en saisir l'essentiel dans un temps relativement court. - saisir le point de vue, repérer la structure d'un article et réorganiser une structure déconstruite. 	<p>EST CAPABLE de :</p> <ul style="list-style-type: none"> - faire un exposé clair sur un sujet connu en insistant sur les temps forts et répondre à des questions factuelles prévisibles. - s'exprimer avec aisance sans fautes gênant la compréhension. - s'autocorriger et parfois développer le propos à partir de notes succinctes. 	<p>EST CAPABLE de :</p> <ul style="list-style-type: none"> - prendre des notes simples lors de conférences ou de sujets assez longs dont il/elle pourra faire un usage raisonnable pour écrire une dissertation ou faire une révision. - relier sa prise de notes à ses objectifs personnels et souligner ses points d'intérêt. - donner son point de vue et d'argumenter.

La compétence **Compréhension de l'Écrit** consiste en une série de questions sur une épreuve de 1 heure, visant à dégager sur trois textes longs des informations générales et détaillées, de repérer la structure d'un texte mais aussi d'évaluer une compréhension en profondeur du document.

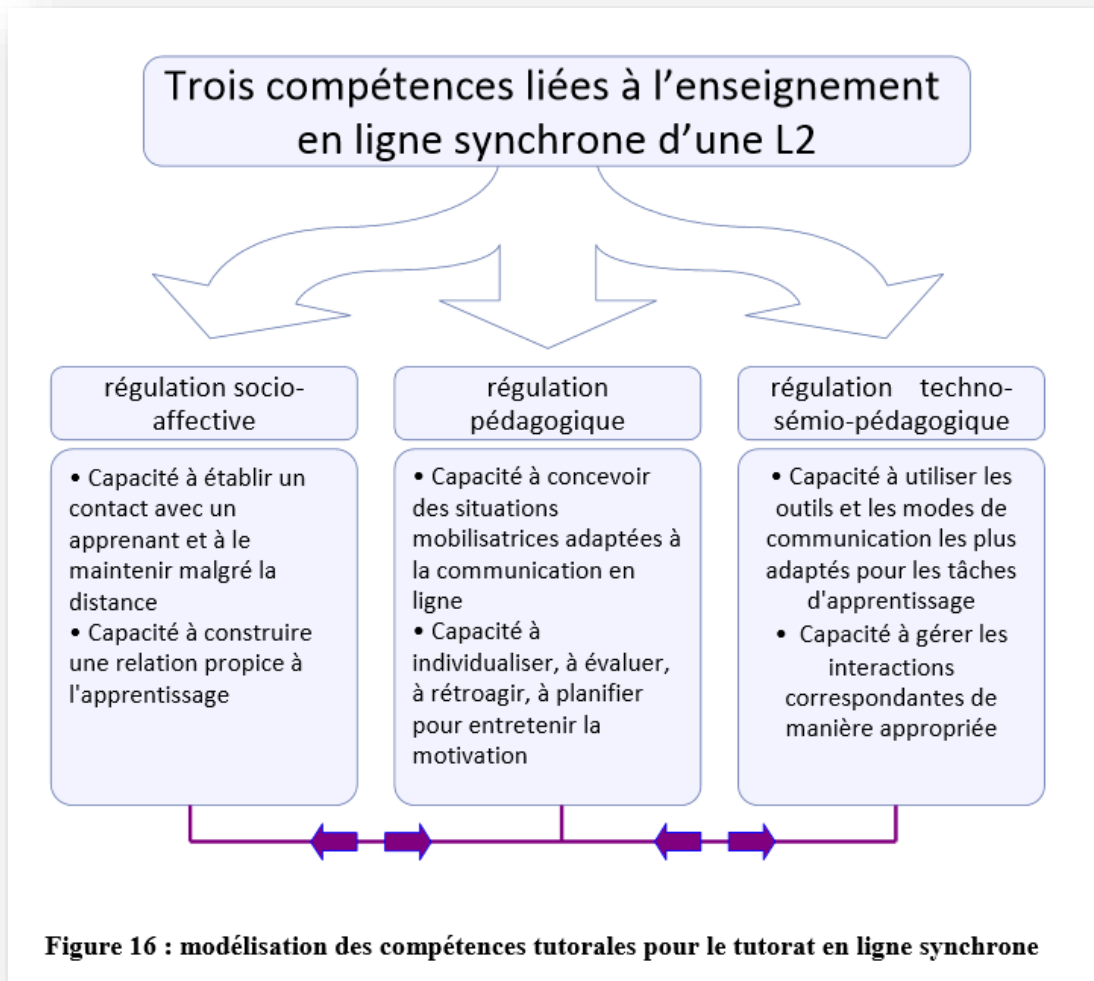
La compétence **Prise de Notes** vise à évaluer la capacité du candidat à comprendre ce qu'il se dit lors d'un cours, d'une conférence d'une durée de 10 à 15 minutes dans sa discipline, et de prendre des notes structurées, signifiantes et efficaces. Il s'agit là d'une compétence qui juge à la fois de la compréhension de l'oral de l'étudiant(e) mais aussi son expression écrite, car l'étudiant(e) est amené(e) à rédiger et à donner un commentaire personnel critique dudit document.

La compétence **Présentation Orale** est évaluée à partir d'un exposé d'une dizaine de minutes fait par l'étudiant(e) et qui porte sur sa spécialité scientifique. Elle est faite devant la classe (qui prend des notes) et devant deux examinateurs. L'un d'eux est leur professeur d'anglais et l'autre est un professeur scientifique de leur spécialité. Le fond et la forme de l'exposé sont ainsi évalués simultanément.

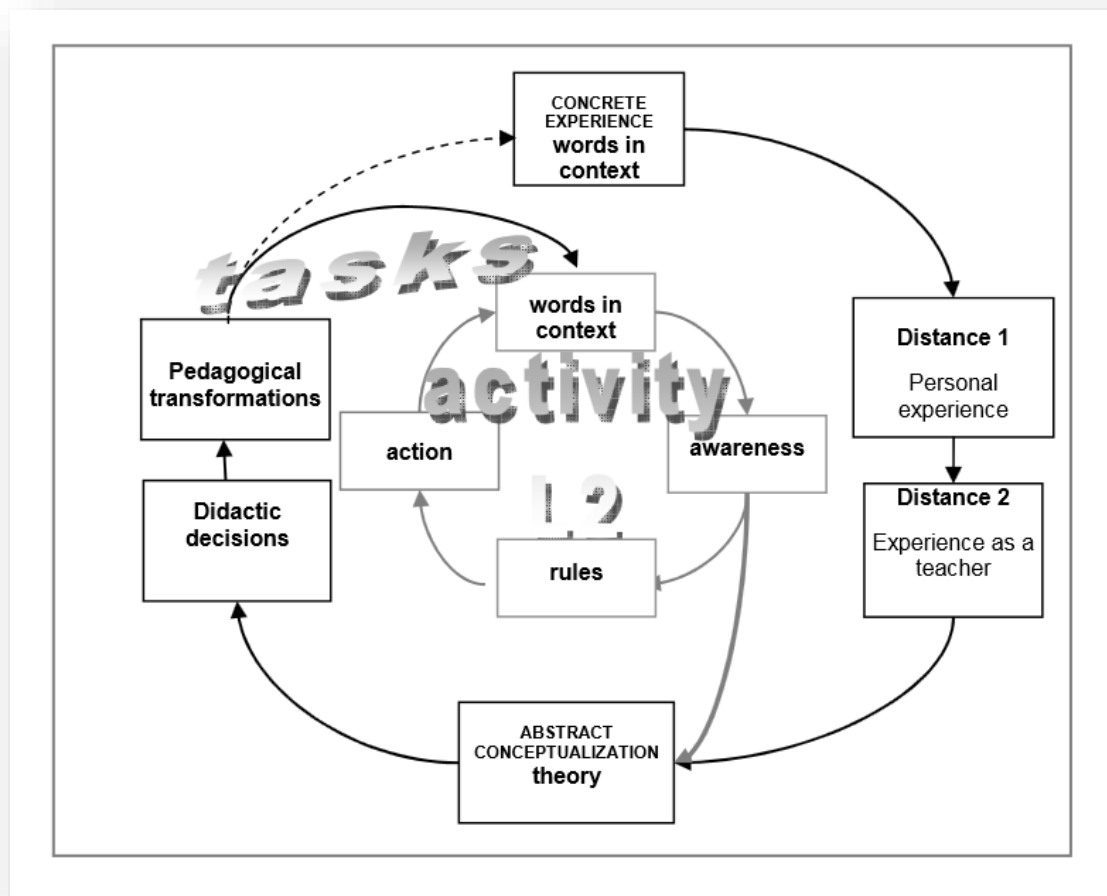
Les étudiant(e)s ayant déjà un certificat en langue étranger de niveau B2 du CECRL (CLES2, TOEIC +780, etc), les étudiant(e)s non francophone ainsi que certaines étudiant(e)s à statut particulier sont dispensés de ce dispositif.

Pour plus d'informations vous pouvez consulter le document du CECRL en suivant le lien suivant : http://www.coe.int/t/dg4/linguistic/source/framework_fr.pdf, pages 173 à 184.

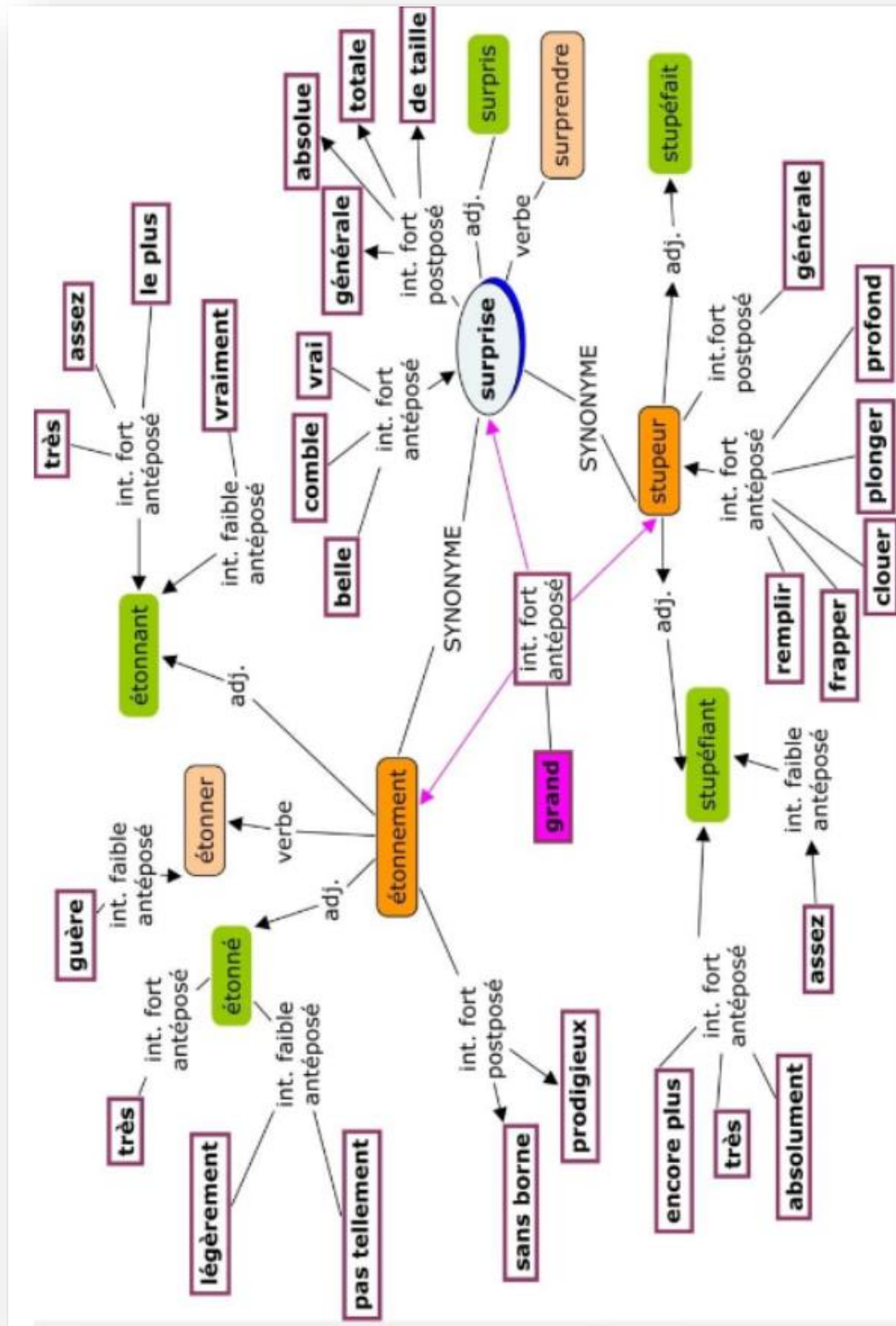
Annexe II : Modèle des compétences tutorales pour le tutorat en ligne synchrone (Guichon, 2011 : 100)



Annexe III : Structure expérientielle de la mise en place des tâches (Deyrich, 2011 : 130)



Annexe IV : Carte heuristique des collocatifs (Cavalla, 2012)



Annexe V : Extrait de données informatiques *Enigma*

[Le nom, le prénom et l'adresse courriel de l'étudiant/e ont été enlevés pour des raisons d'anonymat.]

Début de la session	Fin de la session	Statut de la session	Activité	Mot saisi	Chrono	Mot trouvé
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	osmosis	320	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	has	345	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	figure	458	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	normal	464	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	play	472	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	play	475	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	seen	480	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	even	489	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	other	497	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	subtract	502	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	through	511	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	volume	516	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	liquid	520	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	however	526	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	improve	561	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	volumes	584	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	what	590	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	level	637	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	experience	648	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	illustrate	656	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	attain	661	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	almost	667	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	minute	672	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	middle	718	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	temperatur	735	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	equal	743	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	this	843	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	means	846	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	that	851	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	watter	856	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	town	869	not found

15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a higher	875	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a highest	879	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a unter	884	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a given	898	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a when	907	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a place	916	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a approximate	927	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a number	950	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a large	954	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a level	999	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a levels	1002	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a separate	1014	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a separated	1022	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a illustrated	1051	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a illustrate	1055	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a attain	1071	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a attained	1076	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a minus	1084	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a w with	1114	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a with	1120	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a high	1144	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a widest	1160	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a then	1179	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a difference	1198	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a pression	1210	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a temperature	1249	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a place	1259	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a also	1269	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a used	1273	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a wing	1290	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a wases	1304	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a improve	1311	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a life	1333	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a molair	1361	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a sale	1372	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a through	1395	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a different	1416	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a fundamental	1570	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a place	1593	not found

15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	span	1599	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	wins	1608	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	approximate	1757	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	illustrate	1891	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	illustrates	1895	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	pressure	2048	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	unter	2078	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	seme	2158	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	watter	2176	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	equally	2205	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	equaly	2210	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	systems	2326	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	biologic	2336	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	body	2347	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	ions	2401	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	product	2417	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	somebody	2482	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	somebody	2486	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	sombody	2490	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	water	2502	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	flows	2512	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	from	2517	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	sale	2528	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	sals	2546	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	saly	2552	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	places	2580	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	place	2583	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	role	2587	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	biologic	2595	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	biologique	2600	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	sale	2609	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	explain	2646	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	explained	2653	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	someone	2661	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	drinked	2673	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	drink	2676	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	will	2683	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	flow	2689	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	rest	2694	found

15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	causes	2701	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	these	2821	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	these	2824	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	these	2830	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	substances	2860	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	seme	2875	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	generated	2908	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	hence	2960	found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	sals	2976	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	sale	2978	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	sali	2981	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	saly	2983	not found
15/12/2007 12:55	15/12/2007 13:45	not finished	texte01a	salu	2998	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	osmosis	188	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	has	190	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	fundamental	197	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	for	214	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	any	227	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	life	242	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	can	245	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	seen	249	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	figure	253	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	normal	260	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	these	265	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	even	269	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	molecules	275	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	salt	283	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	other	288	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	substance	294	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	throughout	304	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	volume	308	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	liquid	313	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	however	318	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	two	322	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	volumes	327	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	water	330	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	different	341	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	are	351	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	space	369	found

Annexes

15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	pressure	378	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	generated	392	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	then	398	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	that	403	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	means	406	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	this	408	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	the	411	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	flow	416	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	flows	426	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	from	430	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	low	440	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	high	454	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	under	459	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	until	468	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	hence	478	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	given	483	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	when	493	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	temperature	524	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	plant	530	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	any	558	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	which	573	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	gaz	589	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	and	598	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	places	617	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	place	620	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	role	625	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	all	630	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	biological	634	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	systems	637	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	one	641	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	body	650	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	ensuring	664	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	leave	669	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	level	676	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	levels	680	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	blood	692	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	why	716	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	someone	725	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	someone	728	found

15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	wuth	742	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	with	747	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	result	768	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	chart	781	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	allows	784	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	almost	788	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	applied	792	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	arrow	795	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	compel	808	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	dissolved	815	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	equals	821	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	even	829	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	figure	834	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	flights	840	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	flow	845	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	hence	849	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	levels	854	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	locate	856	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	located	859	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	means	864	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	obtain	869	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	obtained	873	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	plant	884	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	pressure	889	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	produced	896	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	remain	903	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	remove	907	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	removed*	910	not found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	towards	920	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	used	925	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	virtually	930	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	waste	934	found
15/12/2007 13:50	15/12/2007 14:06	not finished	texte01a	widely	939	found
15/12/2007 14:07	15/12/2007 14:31	not finished	texte04a	circulation	241	found
15/12/2007 14:07	15/12/2007 14:31	not finished	texte04a	system	245	found
15/12/2007 14:07	15/12/2007 14:31	not finished	texte04a	cells	251	found
15/12/2007 14:07	15/12/2007 14:31	not finished	texte04a	the	256	found
15/12/2007 14:07	15/12/2007 14:31	not finished	texte04a	egs	266	not found
15/12/2007 14:07	15/12/2007 14:31	not finished	texte04a	egg	268	found

Annexe VI : Grille d'évaluation pour la présentation orale en Master2 en anglais de spécialité

Octobre 2015

Service des Langues de l'Université Joseph Fourier (Grenoble I)

 Student number _____ Last name _____ Subject _____
 UFR _____ First name _____

Marking grid for Master's degree oral presentations (See CEF on verso for oral presentation skills)						
	"survie" A2	"seuil" B1	"independent" B2	"autonome" C1	"maîtrise" C2	Total Points
Structure One mark on each row. Includes half points, ex. 3.	n.a. Can give a short rehearsed presentation. (Introduction / basic transitions / signing off).	2 Presents the main points in a reasonably clear and precise manner.	2.5 Establishes rapport. Clear, systematically developed presentation, highlighting significant points.	3.5 Explains complex subjects clearly and in detail. Well-structured. Use of cohesive devices. "Rounds off" the conclusion.	4 Confident and articulate. Adapts to audience needs. Explains with elaborate and often memorable descriptions.	4
English	n.a. No reading. Numerous grammar and pronunciation mistakes.	2 Noticeable mother tongue influence – but usually understandable.	2.5 Does not make mistakes, which lead to breakdown in understanding. Self-corrects. Pronunciation clear.	3.5 Rare errors difficult to spot. Varies intonation to express finer shades of meaning. Easily uses occasional circumlocutions.	4 Very articulate. Uses idiomatic expressions. Varies intonation to express finer shades of meaning.	English 4
Fluency	0.5 n.a. Pauses, false starts, reformulation very evident.	2 Keeps going comprehensively, despite pauses for planning and repair.	2.5 Good fluency and intonation, despite occasional hesitations. Some chunking.	3.5 Almost effortless fluency. Smooth flow of language. Chunking.	4 Native user fluency. Chunking.	Fluency 4
Questions and answers	0.5 n.a. Frequently asks for repetition. Needs help with formulation of reply.	2 Can manage to answer – may ask for repetition if speech was rapid.	2.5 Ease of expression. Fluency and spontaneity poses no strain (on either self or audience).	3.5 Responds spontaneously almost effortlessly.	4 Can handle difficult and even hostile questions.	Questions and answers 4
Communication Points per criteria are added.	0 - ½ - 1 Clear and aesthetic visuals	0 - ½ - 1 Visuals are integrated into talk.	0 - ½ - 1 Good speed and volume.	0 - ½ - 1 Interactive body language, eye contact.	Communication skills 4	Total 20

Tick one:

- Pas de validation (less than 11.49)
 11.5-12.99 / Passage en commission
 13+/ Validation or Validation by the Commission

 Creative Commons : Attribution – Pas d'utilisation commerciale.
 Citéx : UGA SLL - ALP Sciences et Technologies.

Annexe VII : Unités lexicales du corpus EIIDA identifiées comme étant fréquentes et problématiques par les apprenant.e.s francophones

Forme	Quantité d'occurrences à l'oral	à	Quantité d'occurrences à l'écrit	à	Prédominance
that	792		597		oral
this	465		387		oral
so	363		32		oral
can	269		147		oral
what	236		7		oral
if	131		33		oral
like	69		6		oral
actually	63		4		oral
might	48		11		oral
per	47		38		oral
those	43		24		oral
little	28		11		oral
basically	22		0		oral
me	20		0		oral
something	18		1		oral
who	16		5		oral
fairly	14		2		oral
let's	13		0		oral
ago	8		0		oral
else	7		0		oral
somewhat	7		2		oral
there	99		46		oral
for	243		637		écrit
as	192		366		écrit
by	72		306		écrit
would	68		109		écrit
will	58		119		écrit
than	39		108		écrit

could	34	50	écrit
between	32	91	écrit
within	31	114	écrit
important	30	48	écrit
any	23	36	écrit
each	20	78	écrit
low	19	49	écrit
should	19	35	écrit
such	17	75	écrit
may	15	66	écrit
however	13	83	écrit
further	8	38	écrit
when	62	68	écrit
given	6	29	écrit
currently	8	15	écrit
while	3	35	écrit
likely	2	27	écrit

BIBLIOGRAPHIES

RÉFÉRENCES AUX TRAVAUX PERSONNELS

[HART03a] Hartwell, L. (2003). The Effects of Formal Education on Mother/Daughter Relationships in French Colonies. *Phoebe : Journal of Gender and Cultural Critiques*, 15 (1 & 2), 116-130.

[HART05a] Hartwell, L. (2005). Note pédagogique : Anglais quotidien et scientifique au cœur d'une simulation globale. *Les Cahiers de l'APLIUT*, 24(3), 29-33.

[HART07a] Hartwell, L. (2007). Se documenter sur l'Internet anglophone? Enquête sur les pratiques. 2007. *Les Cahiers de l'APLIUT*, 26 (1), 72-81.

[HART07b] Hartwell, L. (2007). Review: Step Up: Listening, Speaking, and Critical Thinking (1 and 2). *Teaching English as a Second or Foreign Language*, décembre 2007.

[HART09a] Hartwell, L. (2009). On-line support of note-taking strategies for English language learners. *Proceedings of the International EuroCALL Conference. New Trends in CALL: Working together*, Universidad Politécnica de Valencia, Gandia, Espagne, 9-12 septembre 2009 (pp. 129-132).

[HART10a] Hartwell, L. (2010). Visionnement de documents audiovisuels authentiques — Problématiques de contiguïté et d'interprétation culturelle. *ALSIC* (Apprentissage des langues et systèmes d'information et de communication) 13.

[HART10b] Hartwell, L. (2010). Performances selon le genre en anglais de spécialité. *Les Après-midi de LAIRDIL*, 16, 45-62.

[HART10c] Hartwell, L. (2010). Pratiques de reconstruction de texte en autoformation. *Les Cahiers de l'APLIUT*, 29(2), 81-96.

[HART10d] Hartwell, L. (2010). Impact of software design on on-line text reconstruction. *SYSTEM: An International Journal of Educational Technology and Applied Linguistics*, 38(3), 370-378.

[HART10e] Hartwell, L. (2010). Une analyse Bakhtinienne du genre en didactique de l'anglais de spécialité. *Actes du 3èmes RJC en EIAH*, Université Lyon I, Lyon, 6-7 mai 2010, 20-26.

[HART11a] Hartwell, L. (2011). Between FASP and fiction: An interview with author Abigail Padgett. *Les Cahiers de l'APLIUT*, 30, 138-148.

[HART11b] Hartwell, L. (2011). Learning On-Line about Modality in Written and Oral English. *Proceedings of ICT for Language Learning*, Pixel, Firenze, Italie, 20-21 octobre 2011.

[HART11c] Hartwell, L. (2011). Notions of utility: Construction of the B2 foreign language level of competence in Europe and France. *Du mot au concept : Utilité*, (pp. 138-155). Grenoble : Presses Universitaires de Grenoble.

[HART11d] Hartwell, L. (2011). Portraying the language-culture link through spatial representation in three US language textbooks. *Les Après-midis de LAIRDIL* (Laboratoire Inter-universitaire de Recherche en Didactique des Langues), 20, 95-109.

[HART11e] Hartwell, L. (2011). Exploring English for the nuclear industry in biographical films: Oppenheimer and Silkwood. *Les Cahiers de l'APLIUT* (Association des Professeurs de Langues des Instituts Universitaires de Technologie) 30, 96-111.

[HART12a] Hartwell, L. & Jacques, M. P. (2012). A corpus-informed text reconstruction resource for learning about the language of scientific abstracts. *Proceedings of the International EuroCALL Conference. CALL: using, learning, knowing*. University of Gothenburg (Suede), 22-25 août 2012.

[HART13a] Hartwell, L. (2013). Devenir membre d'une communauté de discours : Enquête sur les besoins et les pratiques des docteur-es en sciences. *Les Cahiers d'Acedle*, 10, 217-236.

[HART13b] Hartwell, L. (2013). Corpus-informed descriptions: English verbs and their collocates in science abstracts. *Études en didactique des langues*, 20, 79-94.

[HART13c] Jacques, M.-P., Hartwell, L. & Falaise, A. (2013). TAL et linguistique de corpus pour aider la rédaction scientifique en anglais. *Actes de TALN 2013*. Les Sables d'Olonne, 17-22 juin 2013.

[HART13d] Hartwell, L. & Zou, B. (2013). A Chinese-French case study of English language learning via Wikispaces, Animoto and Skype. *Proceedings of the International EuroCALL Conference: 20 years of Eurocall: Learning from the Past, Looking to the Future*. Université d'Évora (Portugal), 11-14 septembre 2013.

[HART13f] Hartwell, L. (2013). Recension : Natacha Chetcuti et Luca Greco (dir.), 2012, La face cachée du genre. *Bulletin ANEF*, 63, 64-70.

[HART14a] Hartwell, L. & Jacques, M.-P. (2014). Authorial presence in French and English: 'Pronoun + Verb' patterns in Biology and Medicine research articles. *Discours*, 15. <http://discours.revues.org/8941>.

[HART14b] Hartwell, L. & Zou, B. (2014). From anxiety to pleasure: A case study of online foreign language learning. *Études en didactique des langues*, 23-24, 119-131.

[HART14c] Hartwell, L. (2014). Florimond Rakotonoelina (dir.), Perméabilité des frontières entre l'ordinaire et le spécialisé dans les genres et les discours, *Les Carnets du Cediscor*, 12. *LIDIL*, 50, 215-217.

[HART16a] Hartwell, L. (2016) Au-delà des mots : la phraséologie appliquée à l'anglais scientifique. *Cahiers de lexicologie* 1 (108) : 181-195.

[HART16b] Hartwell, L. (2016) Wishful thinking: when scientists hope. *Études en didactique des langues*, 26 : 41-51.

PILOTAGE DE PROJETS DE PÉDAGOGIE NUMÉRIQUE

Enigma Plus (FLODI)

<http://elang.ujf-grenoble.fr/enigma/>

Notable Science (FLODI)

<http://elang.ujf-grenoble.fr/valence/pages/NotableScience/index.php>

Outil de sélection de ressources en ligne – Anglais des sciences (PedagoTICE)

http://elang.ujf-grenoble.fr/anglais/master/accueil/tutorat_guide/index.php

English for Sciences and APS podcasts (MIPE 3)

<http://podcast.grenet.fr/podcast-compact/physical-education-english-version-compacte/>

English Booster Course (PedagoTICE)

Accès réservé

Corpus4All (UNR – RA)

<http://chamilo1.grenet.fr/ujf/courses/CORPUSFORALL/>

CORPUS ET OUTILS RÉFÉRENCÉS

British Academic Written English : <http://www2.warwick.ac.uk/fac/soc/al/research/collect/bawe/>

British National Corpus : www.natcorp.ox.ac.uk/

COBUILD Dictionary : <http://www.collinsdictionary.com/dictionary/english>

Corpus of Contemporary American English : corpus.byu.edu/coca/

John Swales Conference Corpus : <http://jscc.elicorpora.info/>

Learner corpora around the world d'UCL : <http://www.uclouvain.be/en-cecl-lcworld.html>

Louvain Corpus of Native English Essays : <http://www.learnercorpusassociation.org/resources/tools/locness-corpus/>

Pilot Science-specific word list Coxhead & Hirsh (2007): http://www.victoria.ac.nz/lals/about/staff/publications/Sci_EAP_sub_lists_Coxhead_and_Hirsh.pdf

Scientext : <http://scientext.msh-alpes.fr/scientext-site-en/spip.php?article9>

Scientext corpus d'apprenants universitaires français écrivant en anglais : <http://corpora.aiakide.net/scientext18/?do=SQ.setView&view=texts>

SKELL : <https://skell.sketchengine.co.uk/run.cgi/skell#>

StringNet : <http://www.lexchecker.org/>

TermITH : <http://www.atilf.fr/ressources/termith/index.php>

TXM platform : <http://textometrie.ens-lyon.fr/?lang=en>

UCREL - CLAWS : <http://ucrel.lancs.ac.uk/claws/>

BIBLIOGRAPHIE

Arnaud, P. J. L., Lewis, D. M. & Maniez, F. (2008) Disentangling the phraseological web: Observations on lexicalization. In Granger S. & Meunier, F. (Eds.), *Phraseology: An interdisciplinary perspective*. (pp. 111-125). Amsterdam: John Benjamins Publishing.

Association of Language Testers in Europe (Alte). ALTE language for specific purposes (formerly WOLE). <http://www.alte.org/projects/langsp.php>.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 1(1), 1-34.

Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bak Sienkiewicz, M. & Novakova, I. (2013). Le raisonnement dans les textes scientifiques : Le cas des verbes causatifs. In Grossmann, F. & Tutin, A. (Eds.), *L'écrit scientifique : du lexique au discours* (pp. 101-121). Rennes: Presses Universitaires de Rennes.

Bakhtin(e), M. M. (1986/2001). The problem of speech genres. In Jaworski, A. & Coupland, N. (Eds.) *The discourse reader* (pp 121-140). London: Routledge.

Banks, D. (2010). The place of diachronic studies in LSP. *Journée d'étude de LAIRDIL : L'espace en didactique des langues*, le 5 novembre 2010 à l'Université de Toulouse.

Banks, D. (1994). Hedges and how to trim them. In Brekke M., Andersen, Ø. Dahl T. & Myking J. (Eds.) *Applications and Implications of Current LSP Research. Proceedings of the 9th European Symposium on LSP*. Bergen: Fagbokforlaget, 2, (pp. 587–592).

Barber, C. L. (1962). Some measurable characteristics of modern English scientific prose, *Gothenburg Studies in English*, 14, 21-43.

Barthes, R. (1977). Léçon inaugurale de la chaire de sémiologie littéraire du Collège de France.

Bertin, J.-C. (2009). Qu'est-ce que la langue de spécialité ? Quelle en est la problématique spécifique ? Journée d'étude *La langue de spécialité, qu'est-ce que c'est ? Comment l'enseigner ?*, le 23 janvier 2009 à l'UPMF, Grenoble. <http://podcast.grenet.fr/podcast/journees-thematiques-la-langue-de-specialite-quest-ce-que-cest-comment-lenseigner/>.

Bertin, J.-C. (2015). Modélisation en apprentissage des langues médiatisé : quelle utilité ?, *Alsic*, 18. <http://alsic.revues.org/2781>.

Bertin, J.-C., Gravé, P. & Narcy-Combes J.-P. (2000). *Second language distance learning and teaching: Theoretical perspectives and didactic ergonomics*. Hershey, NY: Information Science Reference.

- Biber, D. (2006). *University Language: A corpus-based study of spoken and written registers*. Amsterdam / Philadelphia : John Benjamins Publishing Company.
- Biber, D., Conrad, S. & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks, *Applied Linguistics* 25, 371-405.
- Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Blattes, S., Jans, V. & Upjohn, J. (2003). *Minimum competence in scientific English*. Les Ulis : EDP Sciences. Pap-ebook: http://grenoble-sciences.ujf-grenoble.fr/pap-ebooks/upjohn/unit9_1.
- Blin, F. & Jalkanen, J. (2014). Designing for language learning: agency and languaging in hybrid environments. *Apples – Journal of Applied Language Studies*, 8(1).
- Boch, F. & Rinck, F. (Eds.). (2010). Énonciation et rhétorique dans l'écrit scientifique. *LIDIL*, 41.
- Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. In Boulton, A., Carter-Thomas, S. & Rowley-Jolivet, E. (Eds.) *Corpus-informed research and learning in ESP: Issues and applications* (pp. 261-292). Amsterdam: John Benjamins Publishing.
- Boulton, Alex & Tyne, H. (2014). Des documents authentiques aux corpus : Démarches pour l'apprentissage des langues. Paris: Les Éditions Didier.
- Bourigault, D. (2007). Un analyseur syntaxique opérationnel : SYNTAX. Mémoire de synthèse en vue de l'obtention de l'Habilitation à diriger des recherches. Toulouse: Université de Toulouse.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Braud, V., Millot, P., Sarré, C. & Wozniak, S. (2015). Pour une formation de tous les anglicistes à la langue de spécialité. *Les langues modernes*, 3, 67-76.
- Bruner, J. (1985/2012). The role of interaction formats in language acquisition. In Forgas, J. (Ed.) *Language and Social Situations* (pp. 31-46). New York, NY: Springer Science and Business Media.
- Carter-Thomas, S. (2009). *Texte et contexte : pour une approche fonctionnelle et empirique*. Mémoire de synthèse en vue de l'obtention de l'Habilitation à diriger des recherches. Paris: Université Sorbonne Nouvelle.
- Carter-Thomas, S. (1994). Langue de spécialité : cohésion, culture et cohérence, *ASp*, 5-6, 61-67.

Carter-Thomas, S. & Rowley-Jolivet, E. (2001). Syntactic Differences in Oral and Written Scientific Discourse: The Role of Information Structure. *ASp*, 31-33, 19-37.

Cavalla, C. (2010). Les écrits universitaires des étudiants étrangers : quelles normes présenter ? In Bertrand, O. & Schaffner, I. (Eds.), *Quel Français Enseigner ? La question de la norme dans L'enseignement / apprentissage* (pp. 203-214). Paris: Editions Ecole Polytechnique.

Cavalla, C. & Loiseau, M. (2014). Scientext comme corpus pour l'enseignement. In Grossmann, F. & Tutin, A. (Eds.), *L'écrit scientifique : Du Lexique au discours. Autour de scientext* (pp. 163-180). Rennes: Presse universitaire de Rennes.

Cazade, A. (1995). Multimédia, monometa, pour faire le portrait d'un hypertexte, *ASp*, 7-10. <http://asp.revues.org/3977>.

Cazade, A. (2012). Recension : Sue Blattes, Véronique Jans et Jonathan Upjohn, *Minimum competence in scientific English*, nouvelle édition. *Recherche et pratiques pédagogiques en langues de spécialité : Cahiers de l'Apliut*, 31(1) 130-133.

Cazade, A. (2012). Recension : Sue Blattes, Véronique Jans et Jonathan Upjohn, *Minimum competence in scientific English*, nouvelle édition, *ASp* 61(1) 87-89.

Cazden, C. B. (2001). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.

Chapelle, C. (1997). CALL in the year 2000 - still in search of research paradigms? *Language Learning and Technology Journal*, (1)1.

Chapelle, C. (1998). Multimedia CALL: Lessons to be Learned from Research on Instructed SLA. *Language Learning & Technology*, 2(1), 22-34.

Chapman, M. & Newfields, T. (2008). The "New" TOEIC®: Opinion piece. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 12 : 2, 32-37.

Chaudron, C. (1995). Academic listening. In Mendelsohn, D. & Rubin, J. (Eds.), *A guide for the teaching of second language listening* (pp. 74-96). San Diego, CA: Dominic Press.

Cobb, T. & Boulton, A. (2015). Classroom applications of corpus analysis. In Biber, D. & Reppen, R. (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 478-497). Cambridge: Cambridge University Press.

Collins Learning Catalogue (2007-2013). *English for specific purposes*. <http://www.collins.co.uk/category/English+Language+Teaching/English+for+Specific+Purposes/>

Council of Europe. (2001). *Common European framework of reference for language: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

- Council of Europe. (2003). Relating language examination to the common European framework of references for languages: Learning, teaching, assessment (CEF). Manual, Preliminary Pilot Version.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Coxhead A. & Hirsch D. (2007). A pilot science-specific word list. *Revue française de linguistique appliquée*, (2)7, 65-78.
- Daley, R., Jones, S. & Sinclair, J. (2004). *English Collocation Studies: The OSTI Report*. In Krishnamurthy, R. (Ed.), *Studies in Corpus and Discourse*. London: Continuum.
- Davidson, F. & Fulcher, G. (2007). The common European framework of reference (CEFR) and the design of language tests: A matter of effect. *Language Teaching*, 40(3), 231-241.
- Davies, G. (2007). Total cloze text reconstruction programs: A brief history. www.ict4lt.org/en/FWTHistory.doc.
- Décuré, N. (2000). Poésie et photographie : Deux armes pour un même combat. *Les Cahiers de l'APLIUT*, 19(4), 35-47.
- Décuré, N. (1996). Images et imagination. *Les Après-midi de LAIRDIL*, 7, 29-40.
- Décuré, N. (1994). To see or not to see. *Les Après-midi de LAIRDIL*, 3, 27-31.
- Décuré, N. (1990). Utiliser les images en classe de langue. *Les Langues modernes*, 4, 82-87.
- Demaizière, F. & Narcy-Combes, J.-P. (2005). Méthodologie de la recherche didactique : nativisation, tâches et TIC, *Alsic*. (8)1. <http://alsic.revues.org/326>.
- Deyrich, M.-C. (2007). Médiations et positionnements : deux concepts-clés dans la formation des enseignants en anglais Mémoire de synthèse en vue de l'obtention de l'Habilitation à diriger des recherches. Paris : Université Paris 3 Sorbonne Nouvelle.
- Deyrich, M.-C. (2004). Exploration didactique de la langue du milieu professionnel à l'université : quell apport pour la definition de tâches d'enseignement-apprentissage ? *ASp*, 43-44, 125-134.
- Deyrich, M.-C. (2001). Quelles médiations pour une gestion efficace de la transposition didactique en anglais de spécialité ? *ASp*, 31-33, 143-152.
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Dörnyei, Z. (2006). *Motivational strategies in the language classroom*. Cambridge: Cambridge University Press.

Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics, Morristown, NJ*, (19)1, 61-74.

Ellis, N. C. & Simpson-Vlach, R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory*, 5, 61-78.

Ellis, N. C., O'Donnell, M. B. & Römer, U. (2015). Usage-Based Language Learning. In MacWhinney, B. & O'Grady, W., *Handbook of Language Emergence* (pp. 163-180). Hoboken, NJ: John Wiley & Sons, Inc.

Ellis, N. C., O'Donnell, M. B. & Römer, U. (2013). Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning: A Journal of Research in Language Studies*, 63, Supplement 1, 25-51.

Ellis, N. C. (2001). Memory for language. In Robinson, P. (Ed.) *Cognition and Second Language Instruction* (pp. 33-68). Cambridge: Cambridge University Press.

Engeström, Y. (2008). *From teams to knots: Activity-theoretical studies of collaboration and learning at work*. Cambridge: Cambridge University Press.

Erickson, F. (1996). Going for the zone: The social and cognitive ecology of teacher-student interaction in classroom conversations. In Hicks, D. (Ed.) *Discourse, learning and schooling* (pp. 29-62). New York, NY: Cambridge University Press.

Erickson, F. (1984). School, literacy, and civility: An anthropologist's perspective. *Review of educational research*, 54(4), 525-546.

Falaise, A., Tutin, A. & Kraif, O. (2011a). Exploitation d'un corpus arboré pour non spécialistes par des requêtes guidées et des requêtes sémantiques. *Proceedings from TALN, Montpellier 2011*.

Falaise, A., Tutin, A. & Kraif, O. (2011b). Définition et conception d'une interface pour l'exploitation de corpus arborés pour non-informaticiens : La plateforme ScienQuest du projet Scientext. *TAL* 52(3), 103-128.

Falaise, A., Tutin, A., Kraif, O. & Rouquet, D. (2012). ScienQuest: A treebank exploitation tool for non NLP-specialists. *Actes de COLING 2012, Mumbai, Inde*.

Fenouillet, F. (2003). *Motivation, mémoire et pédagogie*. Paris : L'Harmattan/Éditions Savoir et Formation.

Firth, J. R. (1957). Modes of meaning. *Papers in linguistics 1934-1951*. Oxford: Oxford University Press, 190-215.

Fløttum, K. (2012). Présentation du projet et du corpus KIAP. *Journée d'études EIIDA, Paris, le 20 avril 2012*.

- Fløttum, K. (2010). A Linguistic and discursive view on climate change. *ASp*, 58, 19-37.
- Fløttum, K. (2007). *Language and discipline perspectives on academic discourse*. Cambridge: Cambridge Scholars Publishing.
- Fløttum, K., Dahl, T. & Kinn, T. (2006a). *Academic voices*. Amsterdam: John Benjamins Publishing Company.
- Fløttum, K., Kinn, T. & Dahl, T. (2006b). “We now report on ...” versus “Let us now see how ...”: Author roles and interaction with readers in research articles. In Hyland, K. & Bondi, M. (Eds.), *Academic discourse across disciplines* (pp. 203-224). Bern: Peter Lang.
- Fløttum, K., Dahl, T., Kinn, T., Gjesdal, A. M. & Vold, E. T. (2007). Cultural identities and academic voices. In Fløttum, K., (Ed.), *Language and discipline perspectives on academic discourse* (pp. 14-39). Cambridge: Cambridge Scholars Publishing.
- Fløttum, K. & Vold E. T. (2010). L'éthos auto-attribué d'auteurs-doctorants dans le discours scientifique. *LIDIL*, 41 : 41-58.
- Fries-Verdeil, M.-H. (2009). Mise en cohérence de l'anglais de spécialité et du CECRL en France : difficultés et enjeux, *ASp*, 56 : 105-125.
- Fulcher, G. & Davidson, F. (2009). Test architecture, test retrofit. *Language testing*, 26(1), 123-144.
- Fulcher, G. (2009). Test use and political philosophy, *Annual Review of Applied Linguistics*, 29, 3-20.
- Gee, J. P. (2001). *An introduction to discourse analysis: Theory and method*. New York, NY: Routledge.
- GERAS (Groupe d'Etude et de Recherche en Anglais de Spécialité). (2007). Welcome Page. <http://www.geras.fr/welcome/index.php>.
- GERAS/SAES. (2012). Livre blanc 2012 de la SAES. http://www.geras.fr/bibliotheque/File/LivreblancSAES_contribution%20GERAS_sept%202012-1.pdf
- Gilner, L. (2011). A primer on the general service list. *Reading in a foreign language*, 23(1), 65-83.
- Gilquin, G. (2010). *Corpus, cognition and causative constructions*. Amsterdam: John Benjamins Publishing
- Gjesdal, A. (2013). The Influence of genre constraints on author representation in medical research articles. The French indefinite pronoun on in IMRAD research articles. *Discours*, 12. <http://discours.revues.org/8770>.

Gledhill, C. J. (1997). Les collocations et la construction du savoir scientifique. *ASp*, 15-18, 85-104.

Gledhill, C. J. (2000). *Collocations in science writing*. Tübingen: Gunter Narr Verlag.

Granger, S. (1996). From CA to CIA and back: An integrated contrastive approach to computerized bilingual and learner corpora. In Aijmer, K., Altenberg B. & Johansson, M. (Eds.), *Languages in Contrast. Text-based cross-linguistic studies* (pp. 37-51). Lund: Lund University Press.

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, (1)1 : 7-24.

Granger, S. & Paquot (2008) Disentangling the phraseological web. In Granger S. & Meunier, F. (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 27-50). Amsterdam: John Benjamins Publishing.

Gross, G. (2009). *Sémantique de la cause avec la collaboration de Ramona Pauna et Freiderikos Valetopoulos*. Leuven: Peeters.

Grossmann, F. (2013). Les verbes de constat dans l'écrit scientifique. In Grossmann, F. & Tutin, A. (Eds.), *L'écrit scientifique : du lexique au discours* (pp. 85-100). Rennes : Presses Universitaires de Rennes.

Grossmann F. (2010). L'Auteur scientifique. Des rhétoriques aux épistémologies, *Revue d'anthropologie des connaissances* (4)3, 410-426.

Grossmann F. & Tutin A. (2010). Evidential markers in French scientific writing: the case of the French verb "voir". In Smirnovan E. & Diewald G. (Eds). *Evidentiality in European Languages, Empirical Approaches to Language Typology (EALT)* (pp. 279-307). Berlin / New York: Mouton de Gruyter.

Grossmann, F. & Tutin, A. (Eds). (2014). *L'écrit scientifique : du lexique au discours*. Rennes: Presses Universitaires de Rennes.

Guichon, N. (2011). Apprentissage des langues médiatisé par les technologies : contribution à l'épistémologie de la didactique des langues. Mémoire de synthèse en vue de l'obtention de l'Habilitation à diriger des recherches. Le Havre: Université du Havre.

Guichon, N. (2004). *Compréhension de l'anglais oral et TICE : Les conditions d'un apprentissage signifiant*. Thèse de doctorat. Nantes: Université de Nantes.

Halliday, M. A. K. (2004). *Lexicology*. In Halliday, M. A. K. Teubert, W., Yallop, C. & Čermáková, A. (Eds.) *Lexicology and Corpus Linguistics* (pp. 1-22). London: Continuum.

Halliday, M. A. K. (1997/2004). On the grammar of scientific English (pp. 181-198). *The language of science*. New York, NY: Continuum.

Halliday, M. A. K. (2003). *On language and linguistics: Collected works of M.A.K. Halliday*. London: Bloomsbury Publishing.

Hatier, S., Yan R., Tran, T. T. H., Tutin, A. & Jacques, M.-P. (soumis). French cross-disciplinary scientific lexicon: extraction and linguistic analysis. Euralex Congress Proceedings (Full paper). Ivané Javakhishvili Tbilisi State University, Tbilisi, Georgia, 6 - 10 September 2016.

Heiden, S. (2010). The TXM platform: Building open-source textual analysis software compatible with the TEI encoding scheme. In K. I. Ryo Otaguro, (Ed.), *24th Pacific Asia Conference on Language, Information and Computation - PACLIC24*, (pp. 389-398). Institute for Digital Enhancement of Cognitive Development, Waseda University : Sendai, Japan.

Heiden, Serge., Magué, J-P., Pincemin, B. (2010). TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement. In Sergio Bolasco, I. C. (Ed.), *Proceedings of the 10th International Conference on the Statistical Analysis of Textual Data*, 2, (pp. 1021-1032). Edizioni Universitarie di Lettere Economia Diritto, Roma, Italie.

Henderson, A. (2013). Le positionnement à travers la mention de l'autuer dans les écrits universitaires anglais d'étudiants français de premier cycle : implications pédagogiques. In Grossmann F. & Tutin A. (Eds.), *L'écrit scientifique : du lexique au discours* (pp. 145-162). Rennes: Presses Universitaires de Rennes.

Hilton, H. (2005). Théories de l'apprentissage et didactique des langues. *Les langues modernes*, 3: 12-21.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. New York et London: Routledge, Taylor & Francis Group.

Horn, E. (1926). *A Basic Writing Vocabulary: 10,000 words most commonly used in writing*. Iowa City, IA: College of Education, University of Iowa.

Hunston, S. (2011). *Corpus approaches to evaluation: Phraseology and evaluative language*. New York et London: Routledge, Taylor & Francis Group.

Hunston, S. & Francis, G. (2000). *Pattern grammar: A Corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins Publishing Company.

Hyland, K. & Sancho Guina, C. (2012). *Stance and voice in written academic genres*. London: Palgrave Macmillan.

Hyland, K. (2006). Disciplinary differences: Language variation in academic discourses. In Hyland K. & Bondi, M. (Eds.) *Academic discourse across disciplines* (pp. 17-45). Bern: Peter Lang.

Hyland, K. (2005). *Metadiscourse*. London: Continuum Guides to Discourse.

Hyland, K. (2004). Patterns of engagement: Dialogic features and L2 undergraduate writing. In Ravelli J. & Ellis R. A. (Eds), *Analysing academic writing: Contextualized frameworks* (pp. 5-23). London – New York: Continuum International Publishing Group.

Hyland, K. (2002). Directives: Argument and engagement in academic writing. *Applied Linguistics*, 23(2), 215-239.

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for specific purposes*, 20, 207-226.

Hyland, K. (1996). Writing without conviction?: Hedging in science research articles. *Applied linguistics*, 17(4), 433-454.

Hyland, K. (1995). The author in the text: Hedging in scientific writing. *Hong Kong Papers in Linguistics and Language Teaching*, 18, 33-42.

Hyland, K. & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, 6(2), 183-205.

Isani, S. (2009). Journalism FASP & fictional representations of journalists in popular contemporary literature, *ILCEA : Revue de l'Institut des langues et cultures d'Europe et l'Amérique* 11. <https://ilcea.revues.org/251>.

Jones, N. & Saville N. (2009). European Language Policy: Assessment, Learning, and the CEFR, *Annual Review of Applied Linguistics*, 29, 51-63.

Journal of Academic English. (2015). Description. <http://www.elsevier.com/journals/journal-of-english-for-academic-purposes/1475-1585?generatepdf=true>

Kennedy, G. (2002). Variation in the distribution of modal verbs in the British national corpus. In Reppen, R., Fitzmaurice, S. M. & Biber, D. (Eds.) *Using corpora to explore Linguistic Variation* (pp. 73-90). Amsterdam: John Benjamins Publishing Company.

Kennedy, G. (2008). Phraseology and language pedagogy: Semantic preference associated with English verbs in the British national corpus. In Meunier, F. & Granger, S. (Eds) *Phraseology in Foreign Language Learning and Teaching* (pp. 21-41). Amsterdam: John Benjamins Publishing.

Kilgarriff, A., Husák, M., McAdam, K., Rundell, M. & Pavel R. (2008). GDEX: Automatically finding good dictionary examples in a corpus, *Proceedings of EURALEX*, Barcelona, Espagne, 2008.

Krashen, S. D. & Tracy T. (1983). *The natural approach: Language acquisition in the classroom*. Oxford: Pergamon Press.

Larousse (Société éditions) (2015). *Larousse.fr*. <http://www.larousse.fr/infos/credits>

Lawson, A. (2008). Testing the TOEIC: Practicality, Reliability and Validity in the Test of English for International Communication : MA TEFL/TESL Assignment. Centre for English Language Studies, University of Birmingham.

Legallois, D. (2014). La grammaire d'une langue peut-elle être enseignée à partir de ses unités ? In González Rey, I. *Outils et méthodes d'apprentissage en phraséodidactique*, (pp. 39-58). Fernelmont, Belgique : Editions Modulaires Européennes.

Legallois, D. (2012). La colligation : autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique ?, *Corpus*, 11. <http://corpus.revues.org/2202>.

Legallois, D. (2006). La phraséologie dans la linguistique contextualiste. In Legallois, D. & François, J. (Eds.) *Autour des grammaires de constructions et de patterns* (pp. 28-48). *Cahier du CRISCO*, 21.

Legallois, D. & François, J. (Eds.) (2006). *Autour des grammaires de constructions et de patterns*. *Cahier du CRISCO*, 21.

Mackey, A. & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Mackwhinney, B. (2001). The competition model: The input, the context, and the brain. In Robinson, P. (Ed.) *Cognition and second language instruction* (pp. 69-90). Cambridge: Cambridge University Press.

Maniez, F. (2012). A corpus-based study of adjectival vs. nominal modification in medical English. In Boulton, A. Carter-Thomas, S. & Rowley-Jolivet, E. (Eds.) *Corpus-informed research and learning in ESP: Issues and applications* (pp. 83-102). Amsterdam: John Benjamins Publishing.

Matthey, M. & Véronique, D. (2004). Trois approches de l'acquisition des langues étrangères : enjeux et perspectives, *Acquisition et interaction en langue étrangère*, 21, 203-223.

McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.

Merriam-Webster, Inc. 2014. *Merriam-Webster Beta*. <http://beta.Merriam-webster.com/about-us>.

Moon, R. (2008) Dictionaries and collocation. In Granger S. & Meunier, F. (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 313-336). Amsterdam: John Benjamins Publishing.

Narcy-Combes, J.-P. (2009). La correction dans l'enseignement/apprentissage des langues : un problème malaisé à construire. *Cahiers de l'APLIUT*, 28(3), 26-38.

Narcy-Combes, M.-F. (2005). *Précis de didactique : Devenir professeur de langues*. Paris: Ellipses.

O'Malley, J. M. & Chamot, A. U. (1995). *Learning strategies in second language acquisition*. Cambridge: Cambridge Applied Linguistics/Cambridge University Press.

Osborne, J. (2011). Fluency, complexity and informativeness in native and non-native speech. *International Journal of Corpus Linguistics*, 16(2), 276-298.

Osborne, J. (2008). Phraseology effects as a trigger for errors in L2 English. In Meunier, F. & Granger, S. (Eds) *Phraseology in Foreign Language Learning and Teaching* (pp 67-83). Amsterdam: John Benjamins Publishing.

Paquot, M. (2008). Exemplification in learner writing. In Meunier, F. & Granger, S. (Eds) *Phraseology in Foreign Language Learning and Teaching* (pp 101-119). Amsterdam: John Benjamins Publishing.

Palmer, F. R. (1986/2001). *Mood and modality*. Cambridge: Cambridge University Press.

Payne, T. E. (2008). *Describing morphosyntax: A guide for field linguists*. Cambridge: Cambridge University Press.

Pecman, M. (2005). Les apports possibles de la phraséologie à la didactique des langues étrangères. *Alsic*, 8(2), 109-122.

Pecman, M. (2012). Etude lexicographique et discursive des collocations en vue de leur intégration dans une base de données terminologiques. *The Journal of specialised translation (JoSTrans)* 18, 113-138.

Perrin, M. P. (1995). Les langues de spécialité, facteur de progrès pédagogique. Bordeaux: *Actes du dixième symposium européen des langues de spécialité*, 14-15.

Petit, M. (2006). Les descripteurs du cadre : quelle conception de la langue de spécialité ? <http://www.languesvivantes.ubordeaux2.fr/frsa/pdf/06MPetitCadreEuropean.pdf>.

Piolat, A. (2003). Introduction : Pourquoi étudier la prise de notes en langue seconde ? *Arob@se*, 1-2 : 1-5. <http://www.arobse.to>.

Piolat, A. (2004). La prise de notes : écriture de l'urgence. In Piolat, A. (Ed.) *Ecriture : Approches en sciences cognitives* (pp. 206-229). Aix-en-Provence: Presses Universitaires de Provence.

Poibeau, T. (2012). *Opération étude interdisciplinaire et interlinguistique du discours académique (EIIIDA)*, Site Laboratoire Langues, Textes, Traitements Informatiques, Cognition (UMR 8094). <http://www.lattice.cnrs.fr/Projet-Etude-interdisciplinaire-et>.

- Polguère, A. (2003). *Lexicologie et sémantique lexicale*. Montréal: Les Presses de l'Université de Montréal.
- Poudat, C. & Follette, P. (2012). Corpora and academic writing: A contrastive analysis of research articles in biology and linguistics. In Boulton, A., Carter-Thomas, S. & Rowley-Jolivet, E. (Eds), *Corpus-informed research and learning in ESP: Issues and applications* (pp 167-192). Amsterdam: John Benjamins Publishing.
- Quirk, R. (1985). *A comprehensive grammar of the English language*. London: Longman Publishing Company.
- Raby, F. & Narcy-Combes, J.-P. (2009). Prolégomènes : où en est la recherche sur la motivation en LVE et en L2. *LIDIL*, 40, 5-16.
- Rowley-Jolivet, E. (2012). Oralising text slides in scientific conference presentations: A multimodal corpus analysis. In Boulton, A., Carter-Thomas, S. & Rowley-Jolivet, E. (Eds), *Corpus-informed research and learning in ESP: Issues and applications* (pp. 45-82). Amsterdam: John Benjamins Publishing.
- Salehzadeh, J. (2006). *Academic listening strategies*. Ann Arbor, MI: Series in English for Academic & Professional Purposes.
- Sancho-Guinda, C. & Hyland, K. (2012). Introduction: A Context-Sensitive Approach to Stance and Voice. In Hyland, K. & Sancho-Guinda, C. (Eds.). *Stance and Voice in Written Academic Genres* (pp. 15-33). Cambridge: Palgrave Macmillan.
- Sinclair, J. (2003). Corpora for lexicography. In van Sterkenburg, P. (Ed.). *A Practical Guide to Lexicology, Volume 6 de Terminology and lexicography research and practice* (pp. 167–178). Amsterdam: John Benjamins Publishing.
- Skinner, B. F. (1954). The science of teaching and learning and the art of teaching. *Harvard Educational Review*, 24(2), 86-97.
- Skinner, B. F. (1958). *Teaching machines*. *Science*, 128, 969-977.
- Song, M.-Y. (2012). Note-taking quality and performance on an L2 academic listening test, *Language Testing*, 29(1), 67-89.
- Stubbs, M. (2009). John Sinclair (1933-2007): The search for units of meaning – Sinclair on empirical semantics. *Applied Linguistics*, 30(1), 115-137.
- Stubbs, M. (2009). Technology and phraseology: With notes on the history of corpus linguistics. In Römer, U. & Schulze, R. (Eds). *Exploring the Lexis Grammar Interface* (pp. 15-47). Amsterdam: John Benjamins Publishing.
- Swales, J. M. (1990/2004). *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Swales, J. M. (1984). ESP comes of age? – 21 Years after ‘Some Measurable Characteristics of Modern Scientific Prose.’ *Unesco Alsed-LSP Newsletter*, 7(2), 9-20.

Swales, J. (1971). *Writing Scientific English*. London: Nelson.

Swales, J. & Feak, C. (2004). *Academic Writing for Graduate Students: Essential Tasks and Skills*. Ann Arbor, MI: Michigan Series in English for Academic and Professional Purposes.

Tarone, E., Dwyer, S., Gillette, S., Icke, V. (1981). On the use of the passive in two astrophysics journal papers. *The ESP Journal*, 1(2), 123-140.

Tutin, A. (Éd.). (2007). Lexique des écrits scientifiques. *Revue française de linguistique appliquée*, 12(2).

Tutin, A. (2010). *Sens et combinatoire lexicale : de la langue au discours*. Mémoire de synthèse en vue de l’obtention de l’Habilitation à diriger des recherches. Grenoble: Université de Stendhal.

Tutin, A. (2013). La phraséologie transdisciplinaire des écrits scientifiques : des collocations aux routines sémantico-rhétorique. In Grossmann, F. & Tutin, A. (Eds.), *L’écrit scientifique : du lexique au discours* (pp. 27-44). Rennes: Presses Universitaires de Rennes.

Tutin, A. & Falaise, A. (2014). Expressions polylexicales dans le discours scientifique : une base de données lexicales basée sur corpus. *Actes d’Europhras*, Paris 2014.

Tutin, A. & Grossmann, F. (2014). *L’écrit scientifique : du lexique au discours*. Rennes: Presses Universitaires de Rennes.

Tutin, A., Grossmann, F., Falaise, A. & Kraif, O. (2009). Autour du projet Scientext: étude des marques linguistiques du positionnement de l’auteur dans les écrits scientifiques. http://w3.u-grenoble3.fr/lidilem/labo/file/Lorient_vfinale.pdf.

Tutin, A., Grossmann, F., Falaise, A. & Kraif, O. (2009). Autour du projet Scientext : Étude des marques linguistiques du positionnement de l’auteur dans les écrits scientifiques. Communication aux *Journées Linguistique de Corpus*. Lorient, 10-12 septembre 2009.

Upjohn, J. (1999). *Exit proficiency: The proof of the pudding*. *ASp*, 23-26 : 305-322.

Vandergrift, L. (2008). *Listening: Theory and practice in modern foreign language competence*. Subject center for languages, linguistics and area studies. <http://www.llas.ac.uk/resources/gpg/67>.

Vygotski, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wagner, R. K. & Meros, D. (2010). Vocabulary and reading comprehension: Direct, indirect, and reciprocal influences, *Focus on exceptional children*, 43 : 1. <http://pubmedcentralcanada.ca/pmcc/articles/PMC4031673/>.

Williams, G. (2012). Bringing data and dictionary together. In Boulton, A., Carter-Thomas, S. & Rowley-Jolivet, E. (Eds), *Corpus-informed research and learning in ESP: Issues and applications* (pp. 217-238). Amsterdam: John Benjamins Publishing.

Williams, G. & Millon, C. (2012). Palmer, Firth and Internet: Drawing together collocational threads. *Corpus Linguistics*, 1-3.

