

LA COMPÉTENCE LEXICALE DES ÉTUDIANTS FINNOPHONES EN FRANÇAIS

Étude sur la production écrite des apprenants

Maarit Mutta
Thèse de Doctorat
de 3ème cycle

Université de Turku
Institut de langues
classiques et romanes
Département d'études
françaises
Octobre 1999

TURUN YLIOPISTO

Klassillisten ja romaanisten kielten laitos / Humanistinen tiedekunta

MUTTA, MAARIT: La compétence lexicale des étudiants finnophones en français.

Etude sur la production écrite des apprenants

Lisensiaatintutkimus, 172 s., 12 liitesivua

Ranskan kieli ja kulttuuri

Lokakuu 1999

Tutkimuksen aiheena on sanaston käyttö yliopisto-opiskelijoiden kirjoittamissa ranskankielisissä aineissa. Korpus koostuu 106:sta (2x53) kahdessa eri vaiheessa kootusta aineesta, ja 50:stä (2x25) kontrolliryhmän aineesta. Ensimmäisen vaiheen materiaali on koottu ennen koehenkilöiden osallistumista opiskelija-vaihtoon, joka ajoittuu vuosiin 1987-1993 ja oli lukukauden tai -vuoden mittainen. Kontrolliryhmän jäsenet eivät ole osallistuneet vaihtoon, mutta ovat suorittaneet muulla tavoin pakollisen ranskankielisessä maassa suoritettavan kieliharjoittelun. Toisen vaiheen materiaali on koottu samoin perustein oleskelun jälkeen. Tutkimus kuuluu soveltavan kielitieteen, tarkemmin interlingvatutkimuksen alaan. Tutkimuksen tarkoituksena on selvittää, miten ulkomaisessa yliopistossa oleskelu vaikuttaa sanaston oppimiseen ja miten tätä tietoa voidaan käyttää pedagogisesti hyväksi. Työ on luonteeltaan pääasiallisesti kvantitatiivinen, mutta kvalitatiivinen analyysi (mm. virheanalyysi) täydentää käytettyä menetelmää. Työn perushypoteeseinä ovat seuraavat oletukset: 1) opiskelijavaihtoon perustuva oleskelu ulkomailla parantaa leksikaalisia kykyjä paremmin kuin muu oleskelu ulkomailla, 2) kirjallinen esitys tulee luonnollisemmaksi mm. siksi, että ns. pragmaattisten organisaattoreiden käyttö lisääntyy ja 3) puhekielille ominaiset piirteet ilmestyvät kirjoitettuun tekstiin.

Tutkimuksen teoriaosassa kartoitettiin interlingvaan eli välikieleen liittyviä olennaisia piirteitä sekä niiden suhdetta vieraan kielen opettamisen problematiikkaan. Tutkimuksessa keskityttiin virheiden, kommunikatiivisen ja metakommunikatiivisen kompetenssin sekä kehitystutkimuksen merkitykseen vieraan kielen oppimisessa ja opettamisessa. Lisäksi pohdittiin kirjoitetun ja suullisen kielen eroja, sillä viestikanavan valinta vaikuttaa olennaisesti viestin tuottamisprosessiin. Kvantitatiivista analyysiä varten määriteltiin keskeinen käsite **sana** ortografisen määrittelyn mukaisesti, vaikka pedagogisesti ottaen olennaista on puhua leksikaalisesta yksiköstä. Materiaalianalyyssissä apuna käytetty tietokoneohjelma **WordCruncher** vaikutti osaltaan sanan määrittelyyn. Lisäksi määriteltiin pragmaattiset organisaattorit, jotka vaikuttavat tekstin koherenssiin: tekstin segmentointi kappalejaon ja välimerkkien avulla sekä kielellisten ja tekstuaalisten merkkien käyttö.

Materiaalin analyysi perustui pääasiallisesti Moira Linnarudin käyttämään metodiin teoksessa *Lexis in composition* (1986), mutta sitä sovellettiin työn tarpeiden mukaisesti. Tutkitut kielellis-kvantitatiiviset tekijät olivat: sanojen määrä ainetta kohti, leksikaalinen variaatio, tiheys, yksilöllisyys ja vaikeusaste. Sanojen frekvenssejä tarkasteltiin verrattuna *Dictionnaire des fréquences* ja *Français fondamental* teoksiin. Tilastollisesti tutkimustuloksia tarkkailtiin **Pearsonin tulomomenttikertoimen, Spearmanin ja Kendallin järjestykskorrelaatioiden** avulla, sekä **t-testin** ja **ANOVA**-testin avulla; laskelmissa käytettiin **SPSS**-tilasto-ohjelmaa.

Tutkimustulokset osoittavat, että ryhmien välillä ei ollut tilastollisesti merkittävää eroa sanastoon liittyvissä kvantitatiivisissa laskelmissa, mutta kvalitatiivisissa analyyseissä löytyy eroa koeryhmän hyväksi, mm. siten, että näiden opiskelijoiden käyttämien pragmaattisten organisaattoreiden valikoima on laajempi. Puhekielisiä piirteitä ilmestyy tekstiin, varsinkin niillä opiskelijoilla, jotka opiskelivat pidempään ulkomaisessa yliopistossa tai oleskelivat muuten pidempään ulkomailla. Erityisesti oleskelu ulkomailla paransi näiden opiskelijoiden sanaston tuntemusta, joiden lähtökielitaito oli alhaisempi. Yksilölliset erot ovat kylläkin merkittäviä. Yleisesti ottaen tekstuaaliset kyvyt paranevat ulkomailla oleskelun jälkeen. Koeryhmän opiskelijat osaavat käyttää myös kommunikatiivisia strategioita sanaston ja pragmaattisten organisaattoreiden valinnoissa saadakseen viestinsä perille. Sanaston ja tekstuaalisten kykyjen parantamiseksi opiskelijoita voidaan rohkaista lukemaan erilaisia tekstejä mielekkäällä tavalla, rohkaista käyttämään kommunikatiivisia strategioita ongelmatilanteissa sekä rohkaista opiskelijoita lähtemään opiskelijavaihtoon, jotta he saisivat hyvän kielipohjan tuleville opinnoille.

Asiasanat: interlingva; vieraan kielen oppiminen; ranskan kieli; sanasto; kirjoitettu kieli

La conscience se reflète dans le mot comme le soleil dans une petite goutte d'eau. Le mot est à la conscience ce qu'est un petit monde à un grand, ce qu'est une cellule vivante à l'organisme, un atome au cosmos. C'est bien un petit monde de conscience. Le mot doué de sens est un microcosme de la conscience humaine.

(Vygotski 1896-1934)

TABLE DES MATIERES

1. Introduction	8
1.1. Cadre théorique	9
1.2. Choix du corpus	12
1.3. Remarques terminologiques et hypothèses de travail	14
2. L'Interlangue en tant qu'hypothèse de recherche	17
2.1. Histoire du concept d'interlangue	17
2.2. Caractéristiques – hypothèses sur la nature de l'interlangue	20
2.3. Etat actuel de la recherche interlinguale	27
3. L'interlangue et la didactique des langues étrangères	31
3.1. Importance des erreurs dans l'apprentissage d'une langue seconde	32
3.2. Compétence communicative	35
3.3. Compétence métacommunicative	40
3.4. Etude longitudinale	42
4. Production écrite	46
4.1. Langue parlée vs. langue écrite	46
4.2. Norme linguistique	49
4.3. Processus cognitif dans l'activité d'écriture	51
5. Mot, vocabulaire et lexique en tant qu'objet d'étude	55
5.1. Relation entre les recherches sur le lexique et la syntaxe	55
5.2. Rôle actuel du vocabulaire dans l'apprentissage et l'enseignement d'une langue	58
5.3. Définitions du concept 'mot'	64
6. Analyse textuelle	71
6.1. Analyse textuelle du point de vue lexical	71
6.2. Activité métatextuelle	73
6.3. Organismes pragmatiques du texte	75
7. Méthodes utilisées pour l'analyse du corpus	80
7.1. Le logiciel WordCruncher : de son utilité dans la recherche	80
7.2. Facteurs linguistiques et quantitatifs	81

7.2.1. Nombre de mots	82
7.2.2. Variation lexicale	83
7.2.3. Densité lexicale	84
7.2.4. Individualité et sophistication lexicales	85
7.3. Statistiques	87
7.4. Mots en contexte	87
8. Analyse détaillée du corpus	89
8.1. Analyse quantitative des deux corpus	89
8.1.1. Réflexions sur le nombre de mots	89
8.1.2. Indice lexical – vocable / occurrence	92
8.1.3. Pourcentage des mots lexicaux	98
8.1.4. Originalité et difficulté lexicales	101
8.1.4.1. Individualité	101
8.1.4.2. Sophistication	105
8.1.4.2.1. Mots les plus fréquents	105
8.1.4.2.2. Mots réellement uniques	106
8.1.4.2.3. Cas particuliers	111
8.2. Résultats statistiques	114
8.3. Analyse qualitative : la langue des apprenants	117
8.3.1. Erreurs dans l'utilisation du vocabulaire	118
8.3.1.1. Fautes orthographiques	118
8.3.1.2. Autres formes erronées	121
8.3.1.3. Erreurs de sens	122
8.3.2. Organismes pragmatiques du texte	126
8.3.2.1. Remarques sur le découpage en paragraphes et sur la ponctuation	126
8.3.2.2. Cohérence textuelle	129
8.3.2.3. Développement interlingual	135
8.3.2.3.1. Cas particuliers	139
9. Conclusion	145
9.1. Récapitulation	145
9.2. Résultats principaux	150
9.3. Bilan et perspectives	158
Bibliographie	162
Annexe 1-4	173

LISTE DES FIGURES ET TABLEAUX

Figure 1. Types de variabilité dans l'IL	25
Figure 2. Grammaticalité de la langue	35
Figure 3. Composantes de la compétence communicative	37
Figure 4. Processus cognitifs dans l'activité d'écriture	53
Figure 5. Nombre de mots dans le corpus d'Aix	90
Figure 6. Nombre de mots dans le corpus témoin	90
Figure 7. Densité lexicale dans le corpus d'Aix	99
Figure 8. Densité lexicale dans le corpus témoin	99
Figure 9. Individualité lexicale (a) dans les deux corpus	104
Figure 10. Individualité lexicale (b) dans les deux corpus	104
Tableau 1. Corpus	14
Tableau 2. Test pilote : le mot selon la définition mixte	69
Tableau 3. Test pilote : le mot selon la définition orthographique pure	69
Tableau 4. Variation lexicale : extrait du corpus d'Aix	93
Tableau 5. Variation lexicale : extrait du corpus témoin	96
Tableau 6. Mots les plus fréquents	106
Tableau 7. Division des mots lexicaux réellement uniques selon les parties du discours	107
Tableau 8. Coefficients significatifs de Bravais-Pearson aux deux niveaux	116
Tableau 9. Coefficients significatifs de Kendall aux deux niveaux	117
Tableau 10. Catégories des organisateurs pragmatiques	130
Tableau 11. Nombre et répartition de mots : le corpus d'Aix (Annexe 2)	
Tableau 12. Nombre et répartition de mots : le corpus témoin (Annexe 2)	
Tableau 13. Facteurs quantitatifs : le corpus d'Aix (Annexe 3)	
Tableau 14. Facteurs quantitatifs : le corpus témoin (Annexe 3)	

LISTE D'ABREVIATIONS

AC = Analyse contrastive

AE = Analyse des erreurs

Ddl = Dictionnaire de linguistique

DL = Densité lexicale

IL = Interlangue

ILe = Individualité lexicale

L1 = Langue de départ / langue première / langue maternelle

L2 = Langue-cible / langue seconde / langue étrangère

SL = Sophistication lexicale

VL = Variation lexicale

VL/M = Variation lexicale adaptée au nombre de mots

1. INTRODUCTION

Les êtres humains ont la capacité de communiquer entre eux, soit dans leur langue maternelle, soit dans une langue étrangère. Ceci est possible tout d'abord grâce à l'existence de la **Grammaire universelle**, c'est-à-dire d'un mécanisme biogénétique inné leur permettant l'apprentissage d'une première langue, dite maternelle, dès l'enfance (cf. Chomsky 1965). Plus tard, ils utilisent ces connaissances du système de leur langue maternelle et les principes d'acquisition de différents savoir-faire cognitifs pour apprendre d'autres langues (Bley-Vroman 1989 : 41, 51 ; Vogel 1995 : 133-134). Dans les deux cas, les apprenants d'une langue étrangère ont besoin d'une certaine compétence linguistique (formelle et sémantique) et de moyens, c'est-à-dire de mots, pour pouvoir faire passer leur message, autrement dit, communiquer.

Même s'il est très important de maîtriser les règles d'une langue pour pouvoir communiquer, il est encore plus primordial d'avoir une bonne connaissance du vocabulaire. Ce dernier joue en effet un rôle central aussi bien dans la production que dans la compréhension d'une langue (Coady 1997 : 273-290 ; Mäkipää 1994 : 19). Ce sont les adultes apprenant une langue seconde en particulier qui peuvent se sentir frustrés de ne pas pouvoir exprimer leur idées aussi librement et facilement qu'ils le voudraient. C'est une des principales raisons pour lesquelles nous avons voulu nous pencher sur le rôle du vocabulaire dans l'apprentissage des langues étrangères dans un contexte particulier, celui de l'apprentissage au niveau universitaire.

Le but de notre travail est de montrer comment le séjour en France dans un cadre universitaire a influencé les connaissances du français, et plus précisément, la compétence écrite des étudiants universitaires finlandais, et de quelle façon. Dans cette recherche, nous nous concentrerons essentiellement sur l'utilisation du vocabulaire. Bien que nous soyons consciente de l'importance des systèmes hiérarchiques (par exemple la syntaxe de la phrase simple et complexe) de langue dans la production langagière soit orale, soit écrite, nous nous permettons, compte tenu de la nature de cette recherche, d'exclure ces autres aspects de langue. Cependant, nous nous y référerons occasionnellement lorsque cela s'avérera nécessaire pour la progression du travail.

1.1. Cadre théorique

Notre recherche a pour cadre théorique la linguistique appliquée, et plus précisément la question de l'interlangue, c'est-à-dire, la langue des apprenants. Quand on fait des travaux de recherche dans ce domaine, on pourrait aussi bien étudier la production que la réception d'une langue seconde par les apprenants. Ces deux aptitudes peuvent encore être subdivisées en expression et compréhension orales et écrites. Alors pourquoi choisir de se concentrer sur la production écrite des étudiants finlandais ayant le français comme langue cible ? Ces dernières années, la production orale des apprenants de français de toutes origines a été largement étudiée, tandis que leur production écrite l'a été dans de moindres mesures ; auparavant, c'était le cas contraire : l'écrit était le centre d'intérêt dans le domaine de l'apprentissage d'une langue seconde comme ailleurs dans la linguistique (cf. Benveniste & Jeanjean 1987 ; Elo 1993)¹. Nous ne voulons pas dire qu'il n'existe aucune étude importante dans ce domaine (cf. Moirand 1993 ; Pontecorvo 1997), mais que récemment l'accent semble plutôt être sur la production orale (cf. les études sur la variation interlinguale). En tant qu'enseignante universitaire dans le département de français de l'Université de Turku, Finlande, nous avons voulu examiner la partie écrite de la compétence langagière, car la plupart de nos étudiants deviendront un jour professeurs de langue. Cela veut dire que ces derniers auront besoin d'une bonne compétence écrite en plus d'une compétence orale qui sert principalement à la communication de tous les jours (voir chapitre 4). De cette façon, notre recherche a aussi un but pédagogique car les résultats, au moins nous le souhaitons, permettront de réunir des informations pertinentes sur les problèmes d'apprentissage. En outre, cette recherche fait partie d'un projet de recherche mené par le département d'études françaises de notre université ; le corpus recueilli est issu d'un examen global portant sur les compétences langagières des étudiants finlandais. Plusieurs travaux de recherche sur ce corpus ont été achevés (cf. Elo 1993, Mäkipää 1994, Norro

¹ Nous citons ici Elo sur la recherche de l'oral (1993 : 5) : "Cependant, les recherches sur le français parlé, inaugurées par Bauche en 1920 et Frei en 1929, ne cessent de se développer. Citons quelques-uns des travaux et chercheurs les plus importants : le groupe pour l'élaboration du français fondamental (1958), Guiraud (1959), Sauvageot (1962 et 1972), Richer (1964), Martinet (1969), Rigault (1971), D. François (éd. 1983) et le groupe aixois, le G.A.R.S. C. Blanche-Benveniste et C. Jeanjean présentent une bibliographie exhaustive des ouvrages sur le français parlé dans *Le français parlé* (1987)". Néanmoins, ceux-ci excluent les études lexicales dans leur inventaire (1987 : 4). Nous pourrions y ajouter plusieurs ouvrages publiés dans les années 1990, mais, puisque notre objet est la production écrite, nous nous contentons à mentionner quelques chercheurs dans le domaine concernant l'apprentissage du français oral comme Bartning, Bange, Perdue, Porquier, Vaissière et Véronique. Cf. aussi Russier, Stoffel & Véronique (Eds.) 1991. *Interactions en langue étrangère*. Publications de l'Université d'Aix-en-Provence.

1995, Happonen 1999). Ces travaux traitent de questions sur le langage oral, alors que nous examinons les problèmes de l'écrit. De cette façon nous complétons les connaissances disponibles sur l'apprentissage du français par les étudiants finnophones.

Ayant choisi de traiter plus spécifiquement le vocabulaire, nous aurions pu mener notre recherche en nous appuyant sur des approches très différentes. Dugast (1980 : 5-7) nous présente la statistique lexicale, soit la lexicologie quantitative ou lexicométrie, qui est "l'étude de l'organisation du vocabulaire dans le discours" et qui est une branche de la statistique linguistique (voir aussi Muller 1979). Ces études se fondent principalement sur des textes littéraires et essaient de tirer des conclusions sur la fréquence des mots utilisés, sur la variance des mots, etc. D'autre part, si on s'approche du lexique du point de vue sémantique, on a à faire avec la lexicologie, étude théorique du lexique sémantique, qui examine, entre autres, les relations entre le lexique et la syntaxe, et aussi les rapports avec la société environnante (Picoche 1992 : 8-10). La lexicographie, pour sa part, représente un aspect plutôt pragmatique de la lexicologie car les lexicographes utilisent leur savoir lexicologique pour créer des lexiques, autrement dit, des dictionnaires ; néanmoins, il faut noter que la lexicographie est "largement antérieure à la lexicologie, démarche scientifique très récente" (*Dictionnaire de linguistique (= Ddl)* 1989 : 289, 293). Un autre domaine intéressant bien que fort complexe, et qui intrigue depuis longtemps les lexicologues et plus récemment aussi les sociolinguistes, est celui de la néologie (c'est-à-dire, "le processus de formation de nouvelles unités lexicales" ; *Ddl* 1989 : 334) et surtout la dimension argotique de celle-là. Il va de soi que dans la néologie il s'agit de la langue elle-même, des mots nouveaux qui, soit apparaissent d'abord dans la **parole** et dans la **langue**², autrement dit, qui deviennent acceptables et entrent dans les dictionnaires, ou qui tombent dans l'oubli après un certain temps. En somme, tous ces domaines mentionnés (lexicométrie, lexicologie, lexicographie, néologie) étudient l'état de langue, d'une perspective synchronique ou diachronique, et surtout l'état de la langue maternelle.

² **Langue** et **parole** au sens plus large que selon Ferdinand de Saussure (voir *Ddl* 1989 : 276-181, 358-360).

Pourtant, nous sommes intéressée par l'apprentissage des langues secondes, ce qui nous amène à opter pour l'approche interlinguale³ afin d'essayer de trouver des résultats pertinents quant à l'apprentissage des langues secondes, notamment du français. Comme nous l'avons mentionné ci-dessus, nous nous intéressons au développement éventuel des interlangues des apprenants⁴. De ce point de vue, notre travail se classe parmi les études longitudinales (voir chapitre 3.4.), et en conséquence, la production de chaque étudiant est traitée individuellement. Etant donné la taille de notre corpus, il semble évident que nous pourrions tirer aussi des conclusions pertinentes sur les tendances interlinguales des étudiants finnophones.

En ce qui concerne l'étude des phénomènes interlinguaux, on l'approche le plus souvent en utilisant une analyse qualitative. L'utilisation d'une analyse quantitative n'est pas exclue, mais elle est souvent l'objet d'une controverse. L'opposition à l'approche quantitative concerne tout le domaine des sciences humaines, car celles-ci ne sont pas considérées comme des sciences exactes (Suomela-Salmi 1997 : 16-17). Cela avait déjà soulevé de vives discussions entre Chomsky et Herdan⁵ à cause des cadres théoriques trop étroits dus à des raisons épistémologiques (Ménard 1983 : 2). Quoi qu'il en soit, de nombreuses recherches antérieures montrent l'utilité des mesures d'ordre objectif, c'est-à-dire quantitatif, dans diverses études de linguistique appliquée (voir p. ex. Elo 1993 ; Johansson 1997 ; Lavonius 1998 ; Linnarud 1986 ; Mutta 1995 ; Suomela-Salmi 1997). Nous sommes d'accord avec Svartvik (1992 : 9) qui déclare que

“ Verifiability is a normal requirement in scientific research, and it is hard to see why linguistics (which is often claimed to be “the scientific study of language”) should be exempt from this standard mode of research procedure.”

(voir aussi Mialaret 1991). D'autant plus que d'après Ménard (1983 : 2) “ on ne saurait nier toutefois que la statistique a toujours été un outil indispensable à bien des égards au linguiste”. Certes, ni l'un ni l'autre ne parle d'études interlinguales, mais cela n'exclut pas la possibilité d'utiliser la statistique appliquée dans le domaine en question. Cossette va dans le même sens en affirmant que la richesse

³ Nous adoptons dans cette étude le terme *interlingual* utilisé, entre autres, par Vogel comme adjectif qui correspond au concept d'*interlangue*. Les termes *interlinguistique* vs. *intra-linguistique* sont employés, p. ex. dans le cas des interférences (1995 : 16-17) ; cf. chapitre 3.

⁴ Goudaillier (1997 : 6-7) réfère à la langue des cités (ou des banlieues) avec ce même terme *interlangue*, mais nous nous contentons de l'utiliser seulement en parlant des langues des apprenants.

⁵ Gustav Herdan est d'après Dugast “un des grands ouvriers de la statistique lexicale” (1980 : 7) qui a publié de nombreux ouvrages fondamentaux sur ce sujet, par exemple G. Herdan. 1960. *Type-token Mathematics. A Textbook of Mathematical Linguistics*. The Hague : Mouton & Co. / 1964. *Quantitative Linguistics*. London : Butterworths. / 1966. *The Advanced Theory of Language as Choice and Chance*. Berlin-New York : Springer-Verlag.

lexicale⁶, qui est normalement calculée d'après des formules statistiques, est une notion importante en didactique (1994: 1). Preston & Bailey (1996 : xiv), pour leur part, soulignent les avantages obtenus grâce à l'analyse quantitative lorsqu'il s'agit d'étudier la variation interlinguale, en relation avec le point de vue sociolinguistique. Tarone, en outre, fait remarquer que les résultats d'une étude qualitative sur l'interlangue devraient être étayés par les études quantitatives afin de pouvoir déterminer la généralité des résultats obtenus parmi les apprenants d'une langue seconde (1988 : 131).

Nous aurons ainsi recours aussi bien à l'analyse quantitative qu'à l'analyse qualitative pour donner un aperçu global sur notre sujet. Nous adopterons la méthode quantitative utilisée par Moira Linnarud dans son travail en 1986, mais vu son caractère, nous serons obligée de l'adapter en partie à nos propres besoins (voir chapitre 7). Quant à l'approche qualitative, nous devons constater qu'on ne peut pas parler de l'analyse textuelle proprement dite dans cette étude. Pourtant, nous concentrant sur le niveau lexical, nous voulons expliquer quelques phénomènes dans les interlangues des apprenants, par exemple l'utilisation des organisateurs pragmatiques du texte (cf. chapitre 6), puisque nous avons pu remarquer des changements dans leur usage pragmatique lors de l'observation initiale. Bref, le lexique est examiné en contexte entier, ce qui veut dire qu'il ne s'agit pas d'une analyse de phrases isolées, mais du texte global en termes, entre autres, de cohérence.

1.2. Choix du corpus

Notre corpus se compose d'abord de 106 essais rédigés par les étudiants finnophones dans l'examen des connaissances langagières⁷ entre 1985 et 1993. La longueur de chaque essai est environ de deux cents mots. La moitié des essais ont

⁶ Pour la définition de la richesse lexicale dans cette étude, voir chapitre 7.

⁷ Cet examen global vise à mesurer les connaissances du français du point de vue de la compréhension ainsi que de la production. Il est composé de six tests différents : premièrement, de deux tests portant sur le français parlé : 1) compréhension du français parlé et 2) production du français parlé lors d'une interview dont le but est de tester la maîtrise du français dans une situation de communication interactive. Deuxièmement d'une partie écrite se concentrant sur les compétences en français écrit – un test à choix multiples sur les structures grammaticales et le vocabulaire et un troisième test à trous ayant pour but de vérifier la compréhension globale du texte français ainsi que la maîtrise de sa cohérence, et finalement, un dernier test dont l'objectif est de tester la maîtrise de la production libre du français écrit à travers la rédaction d'un essai sur un sujet donné. Les étudiants passent l'examen deux fois durant leurs études ; d'abord au niveau d'études de base (à la fin de leur première année d'études ou pendant le premier semestre de leur seconde année), et ensuite à la fin de leurs études de licence.

été rédigés au niveau d'études de base avant un séjour linguistique à l'Université de Provence – Aix-Marseille 1 (l'université d'Aix-en-Provence par la suite) entre 1987 et 1993 - c'était à l'époque la seule université d'échange en France pour notre département ; l'autre moitié a été rédigée à la fin des études de licence après ce stage linguistique qui dure normalement soit un semestre soit une année universitaire entière. Le nombre total des étudiants ayant fait le stage à l'université d'Aix pendant cette période était de 63, mais nous avons éliminé dix étudiants du corpus parce qu'ils ne répondaient pas à tous les critères requis ; ils avaient, par exemple, effectué leur séjour à Aix au niveau maîtrise. L'ensemble des essais retenus peut être considéré comme étant relativement homogène car les circonstances de la collecte ont été identiques (cf. Tableau 1. infra).

Mais pourquoi ces années-là ? L'échange interuniversitaire (plus tard transformé en échange ERASMUS et SOCRATES) entre l'université de Turku et l'université d'Aix-en-Provence est entré en vigueur en 1987. Il était donc naturel de commencer avec cette année-là. Pour justifier la date de l'année de clôture de la collecte, nous devons avoir recours à des explications moins précises. Nous avons voulu choisir un corpus suffisamment large pour pouvoir tirer quelques conclusions définitives sur les interlangues des étudiants et pour pouvoir montrer si ces phénomènes sont statistiquement significatifs. Le nombre total des étudiants était donc fixé autour d'une cinquantaine ($2 \times 50 = 100$ dissertations), et comme pendant les six années concernées on a abouti à 53 étudiants remplissant les conditions, nous avons alors clos notre corpus.

Deuxièmement, notre corpus contient 50 (2×25) dissertations qui forment un groupe de contrôle, c'est-à-dire un groupe témoin pour notre corpus primaire. Comme ci-dessus, la moitié des essais ont été faits au niveau de base et l'autre moitié à la fin des études de licence, mais les étudiants faisant partie de ce dernier groupe n'ont pas étudié à l'université d'Aix. Pourtant, puisqu'il est obligatoire d'inclure un séjour linguistique, d'une durée minimum de quatre semaines dans les études de licence, ce groupe consiste en partie aussi d'étudiants qui ont effectué un séjour plus long dans un pays francophone⁸. De toute façon, nous supposons que cela n'a pas d'importance sur les résultats. D'autant plus que nous

⁸ Nous avons collecté cette information des rapports du séjour linguistique. Deux tiers des étudiants dans ce groupe ont passé entre un mois et six mois à l'étranger, tandis qu'un tiers d'entre eux y ont séjourné entre 6 mois et 18 mois, travaillant comme jeune fille au pair et/ ou assistant à des cours de langue type "Alliance française". Un étudiant n'a pas du tout effectué de séjour à l'étranger (il a étudié selon l'ancien programme d'études).

utilisons ici un choix aléatoire parmi tous les étudiants n'ayant pas été à l'université d'Aix-en-Provence et répondant aux mêmes critères requis que pour le corpus primaire (au total 114 étudiants durant la période en question) ; les essais d'un étudiant sur quatre environ ont été sélectionnés (soit 25 au total). Ce groupe témoin sert à vérifier nos hypothèses sur le corpus principal.⁹ Le corpus est présenté dans le Tableau 1.

Examen des connaissances langagières Dissertations (environs 150-200 mots)				
Corpus	niveau I	séjour linguistique	niveau II	Total
primaire	53	1 ou 2 semestres à l'Université d'Aix-en-Provence	53	106
secondaire/ groupe de contrôle	25	entre 4 semaines et 18 mois dans un pays francophone ; pas d'études universitaires	25	50

Tableau 1. Corpus

1.3. Remarques terminologiques et hypothèses de travail

Pour le déroulement de cette recherche, nous avons adopté la disposition suivante : dans les chapitres deux et trois nous parlerons des questions relatives à l'apprentissage des langues et à la didactique. Puis, le chapitre quatre contiendra la production écrite comparée à la production orale et en tant que processus cognitif. Avant de passer aux méthodes utilisées dans ce travail, nous examinerons les rapports entre le vocabulaire et la syntaxe et nous définirons le 'mot'. En outre, nous traiterons brièvement les organisateurs pragmatiques du texte du point de vue lexical. Les chapitres principaux, 8 et 9, dévoileront les analyses quantitative et qualitative menées sur le matériel et les perspectives pour l'avenir.

Nous devons encore faire quelques remarques terminologiques avant de passer à la partie théorique. Les termes **lexique** et **vocabulaire** sont définis de façons différentes selon le domaine dans lequel on les utilise ; quelquefois ils peuvent désigner la même chose, mais d'autres fois, ils sont utilisés distinctement l'un de l'autre. Le dictionnaire de linguistique (*Ddl* 1989 : 297-298, 510-511) donne quatre définitions pour le terme **lexique** et seulement une pour **vocabulaire** ; comme terme linguistique général, " le mot *lexique* désigne l'ensemble des unités

⁹ Nous remercions vivement M. Kalle Astala et Mlle Sari Summanen pour l'aide dans la mise au propre du corpus.

formant la langue d'une communauté [...]”, tandis qu’ “un vocabulaire est une liste exhaustive des occurrences figurant dans un corpus” D’après *Ddl*, Hjelmslev est un des chercheurs qui emploie indifféremment ces deux termes, mais d’autres font une nette distinction entre ces deux termes. D’autres encore envisagent le lexique de divers points de vue ; entre autres, on peut parler du lexique (autrement dit, des potentialités lexicales, ou de la compétence) du locuteur (ibid. p. 297, 511)¹⁰. En parlant de la statistique lexicale, le lexique réfère à la langue, et le vocabulaire au discours. Le terme vocabulaire comporte aussi deux définitions distinctes selon qu’il s’agit du **mot** ou du terme **vocable** (*Ddl* 1989 : 297 ; Dugast 1980 : 56) (voir plus en détail chapitre 5.3.). Malgré toute cette discussion, nous utiliserons les deux concepts d’une manière interchangeable. Néanmoins, dans la partie d’analyse le concept **vocabulaire**, si nécessaire, se réfère à l’utilisation des mots de chaque individu ; autrement dit, chacun a son propre vocabulaire, tandis que le terme **lexique** est réservé au sens de stock global des mots d’une langue en question. Dans les statistiques, nous emploierons les termes **vocabulaire(s)**, **mot** et **vocable** (cf. ch. 5.3.).

Krashen est un des chercheurs en anglais qui a adopté les termes l’**acquisition** (un procès plus implicite et inconscient, normalement dans un environnement de la langue-cible) et l’**apprentissage** (un procès plus explicite et conscient, normalement dans une salle de classe) dont il fait une distinction nette dans son modèle du contrôle (*Monitor theory* 1981, 1982) (cf. Stern 1987 : 403-404 ; Vogel 1995 : 99). Nous ne voulons néanmoins pas faire cette distinction et par conséquent, nous utiliserons ces termes indifféremment. En outre, nous emploierons des abréviations pour les concepts suivants :

analyse contrastive (**AC**)
analyse des erreurs (**AE**)
interlangue(s) (**IL**)
langue de départ / langue première / langue maternelle (**L1**)
langue-cible / langue seconde¹¹ / langue étrangère (**L2**)

Quant à l’utilisation des pronoms personnels, nous emploierons, pour des raisons de simplicité et d’anonymat, le pronom **il** pour faire référence à l’apprenant même s’il s’agit d’une femme.

¹⁰ Dans la littérature anglophone on utilise le terme *learners' mental lexicon* (*Oxford English Dictionary* 1989, vol. VIII).

¹¹ Au sens de Véronique (1992 : 5), qui utilise le terme **langue seconde** “pour désigner la ou les langues autres que la langue de première socialisation, apprise(s) par contacts fortuits ou lors d’un enseignement, que celle(s)-ci soi(en)t acquise(s) dans une situation où elle est/ elles sont employée(s) également hors de l’école [...] ou qu’elle(s) soi(en)t véritablement étrangère(s)”.

Il ressort de notre étude préliminaire sur le corpus qu'en somme, il s'agit de vérifier les hypothèses suivantes :

1. Le séjour linguistique dans le cadre d'un échange universitaire améliore les connaissances lexicales plus efficacement qu'un simple séjour linguistique.
2. La production devient plus naturelle, c'est-à-dire qu'entre autres, l'utilisation des organisateurs pragmatiques augmente dans la production.
3. Des traits du langage parlé s'immiscent dans la production écrite.¹²

¹² Nous remercions vivement Mlle Marie-Laure Baudin et Mme Marie-Pierre Lahouati pour leur contribution à la correction de notre thèse. Il va de soi que nous sommes seule responsable des déviations dans le contenu.

2. L'INTERLANGUE EN TANT QU'HYPOTHESE DE RECHERCHE

En présentant le cadre théorique de la présente étude (voir chapitre 1.1.), nous avons indiqué que ce travail fait partie de la linguistique appliquée. Nous voulons en effet nous concentrer surtout sur la question de l'interlangue, la langue des apprenants. Dans ce chapitre, nous allons discuter des caractéristiques de l'interlangue. Mais avant cela, nous étudions l'origine de l'interlangue et sa place au sein de la linguistique appliquée.

2.1. Histoire du concept d'interlangue

Le concept d'**interlangue** (IL) a été introduit par Larry Selinker en 1972 (le terme apparaît dans ces travaux en 1969 pour la première fois ; cf. Vogel 1995 : 34). Il est plus facile de comprendre ce concept si nous regardons de plus près les études linguistiques qui ont mené jusque là.

L'**analyse contrastive** (AC) est une méthode créée dans les années 1940 pour enseigner des langues étrangères ; par exemple, Charles Fries a publié son *Teaching and Learning English as a Foreign Language* en 1945 (Carcedo 1998 : 28). On ne peut pourtant pas donner la date exacte de son origine, car les chercheurs définissent ses débuts un peu différemment ; selon James (1981 : 8), l'AC moderne commence avec la publication de *Linguistics Across Cultures* de Lado (1957).¹³ Elle puise ses origines dans la théorie behavioriste et le structuralisme taxinomique. Elle est donc orientée sur la linguistique (cf. Vogel 1995 : 16, 29). Dans cette analyse, les structures de deux ou plusieurs langues sont comparées afin de trouver des similitudes ou des différences entre ces langues. Quelques structures des langues sont décrites d'une manière détaillée d'abord, puis en général, elles sont comparées entre elles à l'aide des équivalences de traduction (Sajavaara 1980 : 206) ; par exemple, on compare les équivalences de l'article partitif français en finnois. James précise que cette analyse a une base psychologique qui est la théorie du **transfert** élaborée dans la théorie behavioriste (stimulus-réponse) (1981 : 11, 20). En plus, il est parfaitement convaincu que l'AC est toujours prédictive quant à l'influence de la langue

¹³ On peut trouver des travaux sur l'AC appliquée à l'enseignement des langues dans l'école fonctionnelle dite Prague dans les années 1920 et 1930, mais en général, on considère Lado comme un des pionniers principaux dans l'AC moderne.

maternelle (L1) ; ce qui veut dire que le linguiste peut prédire le comportement des apprenants parce qu'il a préalablement analysé les langues en question ; ceci est appelé l'hypothèse forte contrairement à l'hypothèse faible qui a seulement un pouvoir diagnostique (James 1981 : 181-185).

La plupart des chercheurs ont critiqué l'AC à cause de sa conception trop étroite des phénomènes d'apprentissage. Vogel (1995 : 16), entre autres, dit que son inconvénient était de ne prendre en compte que certains aspects spécifiques de l'acquisition des langues secondes. Candlin (1981 : iii-iv) donne l'avis des enseignants en disant que, des points de vue psycholinguistique et pédagogique, l'AC pouvait prédire seulement une partie des problèmes d'apprentissage. Il conclut cependant en disant que l'AC a énormément contribué à faire progresser la théorie de la traduction, la description des langues différentes, la typologie des langues et l'étude des caractéristiques universelles des langues. Pour sa part, Klein (1989 : 40-41) déclare que personne ne soutient plus vraiment l'hypothèse forte de l'AC aujourd'hui. Il voit cependant l'importance de la découverte de l'existence de plusieurs formes de transfert, même si selon lui "les résultats décevants de l'analyse contrastive jettent un certain discrédit sur la notion de "transfert" de la LM [langue maternelle] sur la LE [langue étrangère]".¹⁴ James lui-même affirme que l'AC concerne plutôt la problématique de l'enseignement que celle de l'apprentissage des langues (1981 : 12), ce qui ne semble pas suffire actuellement.¹⁵

Dans les années 1960 et 1970, la linguistique était influencée par d'autres domaines de recherches (par exemple la psychologie et la sociologie¹⁶) ce qui a élargi le domaine linguistique. Les linguistes ont commencé à prendre de plus en plus en considération les phénomènes extralinguistiques ce qui a eu des répercussions sur la recherche de l'acquisition des langues secondes. Du fait de ce changement général, l'AC a cédé sa place, au moins en partie, à l'**analyse des erreurs** (AE). L'AE est une méthode qui se base aussi bien sur l'analyse linguistique que psycholinguistique. James considère que cette approche basée sur la performance et orientée sur le processus d'apprentissage est propre à l'AE qui, d'après lui, n'a qu'un rôle diagnostique contrairement à l'AC (op. cit. : 181, 185).

¹⁴ Cf. aussi Gass & Selinker (Eds.) 1983. *Language Transfer in Language Learning*.

¹⁵ Voir aussi Giacobbe (1992 : 44-49) sur la polémique autour de l'AC.

¹⁶ Quelques réflexions sur l'aspect psychologique, p. ex. dans Carroll (1968 : 46-49) et Stern (1987 : 289-336) et sur l'aspect sociologique et sociolinguistique, p. ex. dans Halliday (1985 : 27-45).

Corder distingue trois phases principales dans l'AE ; premièrement, il faut identifier les erreurs, puis les classer (la description), et finalement, les interpréter (l'explication) pour pouvoir tirer des conclusions pour la planification de l'enseignement (1975 : 166-170). Nous reviendrons sur la question des erreurs dans le chapitre 3.1.

De plus, Corder (1975 : 158, 165) indique que l'AE, est étroitement liée à l'investigation de la langue des apprenants d'une langue seconde (L2). L'AE a créé des techniques sur lesquelles se base l'étude du développement langagier des apprenants. Selon Corder (op. cit. : 169), l'AE a deux buts principaux : primo, clarifier ce que l'apprenant apprend et comment il l'apprend quand il étudie une L2 (le but théorique), et secundo, faire apprendre l'apprenant d'une façon plus efficace en exploitant le savoir sur son dialecte idiosyncrasique (le but pédagogique) (voir infra). On peut donc distinguer aussi un autre changement : celui du point d'intérêt. L'accent n'est plus mis sur l'enseignement, mais sur l'apprentissage. Comme l'AC, l'AE a également été critiquée à cause de ses méthodes de recherche limitées, surtout parce qu'elle s'est concentrée sur les erreurs. Les chercheurs n'ont donc pas traité la situation d'apprentissage dans son ensemble. Larsen-Freeman & Long indiquent que les chercheurs ont étudié où les apprenants commettaient des erreurs, mais pas où ils réussissaient. De plus, il était presque impossible de trouver une seule cause pour les erreurs (1992 : 61-62). Quelques chercheurs suggèrent que l'AE a remplacé l'AC totalement, ou au moins qu'elle la complète. D'autres s'opposent à cette affirmation et disent en revanche que ces deux approches devraient être considérées comme des morceaux d'une image plus grande, en d'autres termes, de la problématique d'apprendre et d'enseigner une langue (Sajavaara 1977 : 70-71). Par exemple, Faerch et al. (1984 : 289-290) disent brièvement que l'AE est implicitement contrastive, même si d'une façon modifiée (cf. l'AC traditionnelle), et que toutes ces analyses contrastives sont essentielles pour l'étude interlinguale. De toute façon, l'AC et l'AE sont considérées aujourd'hui comme inadéquates pour expliquer à elles seules les problèmes de l'acquisition de L2 (cf. Besse & Porquier 1984 : 200-215 ; Larsen-Freeman & Long 1992 ; Vogel 1995).¹⁷

¹⁷ Nous voulons préciser qu'aujourd'hui même on fait des recherches pertinentes en appliquant les méthodes de l'analyse contrastive pour faciliter l'apprentissage d'une L2 ; voir p. ex. Carcedo 1998. Härmä affirme que l'analyse contrastive, ou plutôt l'étude contrastive, est importante aussi bien pour les professeurs de langue que pour les apprenants afin d'augmenter leur conscience linguistique (1998 : 10-11).

2.2. Caractéristiques – hypothèses sur la nature de l’interlangue

Même si l’AC et l’AE sont du point de vue historique de “vieilles méthodes”, elles se fondent bien dans un domaine plus vaste qu’est l’étude de l’interlangue (IL). L’IL est, en termes simples, la langue des apprenants, donc la parole et l’écriture produites par des apprenants d’une L2 essayant de produire en L2. Selinker, qui a introduit ce terme, approche cette question du point de vue psychologique ; en étudiant les structures et processus psycholinguistiques sous la performance sensée (*attempted meaningful performance*), il nous propose l’existence d’une structure psychologique latente activée lors du traitement d’une L2. Cette structure diffère de la structure linguistique latente¹⁸ qui semble présente dans l’acquisition de L1, et qui correspond biologiquement à la grammaire universelle de Chomsky (Selinker 1975 : 31-33). En outre, Selinker détermine le terme “*meaningful performance*” en disant qu’il est utilisé en référant à la situation

“where an ‘adult’ [plus de 12 ans] attempts to express meanings, which he may already have, in a language which he is in the process of learning” (1975 : 32).

Ce phénomène a d’autres descriptions et appellations aussi. Nensser (1971) parle du **système approximatif** (*approximative system*), Klein du **lecte d’apprenant** (all. *Lernervarietät* ; 1989 : 45), Corder parle d’abord de la **compétence transitoire** (*transitional competence*) (1967), et puis du **dialecte transitoire** (*transitional dialect*), soit **idiosyncrasique** (*idiosyncratic dialect*) (1971). Tous ces chercheurs parlent des apprenants qui développent des systèmes intermédiaires successifs et dynamiques pour se rapprocher de plus en plus de la langue-cible (cf. Besse & Porquier 1984 ; Corder 1975 ; Gaonac’h 1991 ; James 1981 ; Vogel 1995). Néanmoins, ces appellations ne sont pas tout à fait synonymes ; par exemple, Corder adopte une approche plutôt linguistique (cf. Corder 1975 : 158-159 ; Giacobbe 1992 : 24-25). Pour sa part, Vogel critique le système de Nensser en disant :

“[...]d’où une réduction injustifiée du concept de l’interlangue qui, au sens de “système approximatif”, est au service de l’analyse contrastive (l’analyse du “contact” entre L1 et L2), et qui de ce fait limite ses investigations aux seules dimensions accessibles et explicables par le contraste (“contact”) entre L1 et L2 telles que les erreurs ou les aides à l’apprentissage” [italique de Vogel] (1995 : 31).¹⁹

¹⁸ Au sens de Lenneberg (1967) ; voir p. ex. Gaonac’h (1991 : 101, 126) et Selinker (1975 : 33).

¹⁹ Vogel critique aussi d’autres modèles de description de l’interlangue fondés sur l’AC (1995 : 32-34).

Le concept de Selinker, non plus, n'est pas exempt de critiques. Selon Porquier, la notion a évolué pendant une vingtaine d'années au cours de son existence, ayant été d'abord liée à l'apprentissage guidé et élargissant ensuite son champ d'investigation à l'acquisition naturelle des langues ce qui a fait apparaître de nouvelles notions décrivant mieux le phénomène (1986 : 101-107). A l'instar de Porquier (1986), Giacobbe signale "à juste titre le manque d'homogénéité des différentes définitions du terme 'interlangue'" (1992 : 24). Porquier lui-même accepte ce terme même si modifié en cours de route, mais il préfère parler de l'apprenant en situation de communication **exolingue**²⁰ où un des interlocuteurs au moins parle dans une langue seconde ; d'après lui, l'apprentissage exolingue articule l'apprentissage d'une langue et l'apprentissage interculturel (1984 : 18-24, 44). Cette discussion sur la notion d'interlangue a déjà été vive entre Gaonac'h et Moirand en 1984. Moirand (1984 : 81, 84) a voulu aborder la question sur la redéfinition de ce concept, et de bien d'autres (p. ex. régression, fossilisation et simplification ; cf. ci-dessous), car elle était "restée un peu sur [sa] faim" en ce qui concerne la notion d'interlangue. Gaonac'h a signalé pour sa part qu'il faudrait "s'assurer que l'on parle bien de la même chose" quand on discute des différentes approches concernant le phénomène en question (1984 : 78). De toute façon, comme Giacobbe, nous utilisons le terme **interlangue**, devenu le plus fréquent dans la bibliographie même francophone, indépendamment de la terminologie utilisée par maints auteurs (cf. 1990 : 115 ; 1992 : 23).

Les recherches sur l'IL ont conduit à plusieurs hypothèses sur ses caractéristiques. Une des suppositions de base est qu'elle est systématique et contient des règles. On peut y trouver des caractéristiques de la L1, mais il y a aussi des règles communes avec la langue-cible (Corder 1975 : 161-162 ; Faerch et al. 1984 : 272-273 ; cf. aussi Besse & Porquier 1984 : 216-227). C'est donc un système linguistique à part. D'autre part, Corder ajoute que "ces règles interlinguales ne sont pas uniques à un individu, mais qu'elles sont partagées par d'autres ayant une situation culturelle, des buts ou une histoire linguistique similaires" (op. cit. : 163). En d'autres termes, les apprenants d'une même L1 ont des structures

²⁰ Terme introduit par Porquier en 1978 (1984 : 17). Cependant, la communication exolingue est plus largement définie qu'une communication interlinguale des apprenants de langue (op. cit. : 29-37 ; cf. aussi la discussion terminologique de Porquier 1994). Ce terme est généralement utilisé parmi les chercheurs français et suisses décrivant le langage parlé. Etroitement lié aux phénomènes interactifs et à l'acquisition est le concept de **séquences potentiellement acquisitionnelles (SPA)**, introduit par De Pietro, Matthey & Py (1988), et qui désigne "les 'séquences que les interlocuteurs eux-mêmes construisent ensemble en leur attribuant des vertus acquisitionnelles' et dans lesquelles ils semblent établir entre eux une sorte de 'contrat didactique'" (Vasseur 1991 : 49-50).

analogues dans leur IL (cf. Elo 1993 : 9 ; James 1981 : 52). Cependant, Giacobbe fait remarquer que les IL présentent plutôt une pluralité de micro-systèmes, un ensemble hétérogène et partiellement contradictoire, qu'un système unifié et cohérent (1990 : 119). Tarone rappelle aussi que, bien évidemment, la performance varie d'une situation à l'autre, mais que ceci n'exclut pas la systématisme (1988 : vii). Larsen-Freeman & Long (1992 : 88-96) expliquent dans leur livre comment plusieurs chercheurs²¹, eux y compris, ont par ailleurs constaté que les IL semblent avoir des ordres et des séquences d'acquisition communs (séquences d'acquisition naturelle) lorsqu'il s'agit de l'apprentissage de différents traits linguistiques d'une L2 (p. ex. l'apprentissage de négation et de formes interrogatives) ; en d'autres termes, il y aurait une succession naturelle de stades d'apprentissage propre à l'IL (cf. Ellis 1992 : 55-63 ; Vogel 1995 : 50, 63).

Une autre caractéristique est la simplification, ou plutôt la simplicité (cf. Elo 1993 : 10), des systèmes comparés avec ceux de la L2 ; cela concerne tous les niveaux hiérarchiques de la langue, aussi bien la grammaire que le lexique. Corder préfère le terme "simple" à celui de "simplification", car même si les IL sont moins complexes que les L2, elles ne sont pas consciemment simplifiées (James 1981 : 159) ; autrement dit, pourquoi les apprenants d'une L2 simplifieraient-ils une grammaire ou un lexique difficilement acquis ? Nous sommes d'accord avec Elo (op. cit. : 10) qu'on ne peut pas simplifier quelque chose qu'on ne possède pas. Faerch et al. (1984 : 273) utilisent le terme *reduced systems*, systèmes réduits, qui décrit bien les IL. On a pu remarquer que ces systèmes des IL ont des traits communs et universels - notamment un vocabulaire limité, une grammaire simple et un champ de fonction moins étendu que celui d'autres langues - avec d'autres systèmes réduits : le langage-pour-enfants (*motherese/ baby talk*, p. ex. *Voilà un petit nounours qui ronfle "rrron pschii, rrron pschiii"*) et le registre-pour-étrangers ou xénolecte (le "petit nègre"/ *foreigner talk*, p. ex. *moi attendre bus*), et même le pidgin qui est un système de communication basé sur des compromis lexicaux et grammaticaux issu du contact entre différentes langues (cf. Faerch et al. 1984 : 273 ; James 1981 : 159 ; Klein 1989 : 65). Pourtant, il existe des différences entre ces systèmes, par exemple quant au langage-pour-enfants, il s'agit de l'acquisition de la L1. Le niveau ultime de ces autres systèmes réduits reste le même, tandis qu'une IL tend vers une compétence maximale d'une L2 lors de l'apprentissage (cf. p. ex. Ellis 1992 : 252-253 ; Klein 1989 : 47-49 ;

²¹ Entre autres, Dulay & Burt (1973, 1974).

Larsen-Freeman & Long 1992 : 251-266 sur la discussion de la pidginisation²²). L'acquisition des IL ressemble aussi à l'acquisition d'une L1 (p. ex. l'ordre acquisitionnel des structures comme la négation et l'interrogation), mais on a pu cependant y trouver plusieurs dissemblances, par exemple la latéralisation des fonctions cérébrales due à l'âge critique, le degré de l'utilisation du dispositif inné d'acquisition et le fait qu'un enfant applique un système linguistique réduit au savoir du monde également réduit, tandis qu'un apprenant, surtout adulte, d'une L2 applique un système linguistique réduit au savoir du monde connu déjà en L1 (cf. p. ex. Elo 1993 : 10-11 ; Gaonac'h 1991 : 129-133 ; Giacobbe 1990 : 115 ; Klein 1989 : 28 sur la discussion d'identité de processus d'acquisition en L1 et L2).

La **fossilisation** est un autre terme, introduit par Selinker, pour désigner une partie des processus psycholinguistiques qui se manifestent dans une situation d'apprentissage. Pour Selinker (1975 : 36),

“Fossilizable linguistic phenomena are linguistic items, rules, and subsystems which speakers of a particular NL [= native language] will tend to keep in their IL relative to a particular TL [= target language], no matter what the age of the learner or amount of explanation and instruction he receives in the TL [...]. It is important to note that fossilizable structures tend to remain as potential performance, re-emerging in the productive performance of an IL even when seemingly eradicated”.

La fossilisation peut émerger dans l'interlangue, par exemple quand l'apprenant pense qu'il connaît suffisamment la L2 pour pouvoir communiquer, c'est-à-dire, quand il croit se débrouiller bien avec ses connaissances dans une situation donnée. Grosjean affirme : “La maîtrise d'une langue donnée pourra se fossiliser (tout au moins chez l'adulte) lorsque les besoins de la communication seront assurés” (1984 : 52). Ou il peut arriver que l'input²³ linguistique (les éléments linguistiques auxquels l'apprenant est exposé) est extrêmement restreint (cf. Selinker & Lakshmanan 1994 : 212). Dans ce cas, on peut aussi parler de “pidginisation” de l'interlangue (cf. Schumann 1978 : 140-150 ; Vogel 1995 : 39). Ces formes fossilisées peuvent être correctes ou incorrectes. Les formes incorrectes devraient être défossilisées, par exemple à l'aide du feed-back²⁴ des locuteurs/ des enseignants et éradiquées de l'IL. Pour que cela soit efficace du point de vue de l'apprentissage, il faut un feed-back négatif dans le champ cognitif et un feed-back positif dans le champ affectif (Vigil & Oller 1976 : 281).

²² Voir aussi Andersen (1983a, b) et Schumann (1976, 1978) sur la pidginisation des IL.

²³ Corder fait la différence entre l'input et l'intake (1975 : 23).

²⁴ Utilisé sous la forme du *Dictionnaire de linguistique* (1989 : 212).

Adjemian (1976 : 308, 317-319) parle de la perte de la perméabilité en référant à la fossilisation. Selon lui, la perméabilité est un phénomène propre à l'IL qui la rend flexible (p. ex. permet le transfert d'une L1 et la surgénéralisation de règles de l'IL) ; cela n'existerait pas dans les langues naturelles.²⁵ Selinker (1975 : 36-37) appelle *backsliding* (régression) la réapparition des erreurs fossilisées qu'on croyait déjà éradiquées de la production; il s'agit d'une régression à un stade antérieur de l'acquisition, mais toujours vers la norme d'une IL (cf. Klein 1989 : 72-73). Cette régression peut apparaître quand l'étudiant essaie de résoudre un sujet difficile, quand il est anxieux, ou même quand il est décontracté (Schwarte 1982 : 26).

Nous pouvons aussi décrire la fossilisation ou la perméabilité des IL d'une façon alternative, notamment en parlant du dynamisme des IL. Les IL sont systématiques, mais en même temps en évolution constante, en d'autres termes, elles sont instables en changeant avec le temps. Elles sont susceptibles de changements qui les mènent de plus en plus vers une L2 (cf. Elo 1993 : 11 ; Faerch et al. 1984 : 274 ; Stern 1987 : 399-400). Pour pouvoir constater ce progrès, il faut recourir à une étude développementale²⁶. Dans notre recherche, nous essayons d'examiner ce développement éventuel sur la base d'un corpus longitudinal, ce qui semble le seul moyen de détecter les formes stabilisées et erronées (ou non) dans la production des apprenants (voir chapitre 3.3.). Ainsi continuons-nous dans la ligne tracée par plusieurs chercheurs. Voir aussi ce qu'en disent, entre autres, Bartning (1994), Eubank et al. (1997) et Regan (1996).

Enfin, cela nous conduit à présenter une caractéristique saillante de l'IL. Comme on l'a dit supra, les IL sont donc systématiques, mais en même temps en évolution constante. De plus, cela n'exclut pas leur variabilité dans la situation de production, c'est-à-dire dans la performance. Cette variabilité est différenciée de multiples façons et à cause de cette variabilité, il est d'autant plus difficile de détecter les formes stabilisées des IL. Comme les IL se transforment au fur et à mesure qu'on s'achemine vers la compétence d'une L2, il faut recourir à une étude développementale, soit longitudinale, afin de pouvoir constater ce progrès. De ce point de vue, les enseignants à l'école ont un avantage, bien évidemment,

²⁵ Voir la discussion critique sur ce modèle d'Adjemian (Vogel 1995 : 46-49).

²⁶ James (1981 : 4) parle même d'une étude interlinguale diachronique au sens du changement individuel d'un être humain (différent donc du sens saussurien).

de pouvoir suivre le développement interlingual de leurs élèves durant des années et ainsi découvrir les réelles étapes de leur progrès.

Selon Ellis (1992 : 75-76)²⁷, il existe deux principaux types de variabilité : systématique et non-systématique, eux-mêmes divisés en sous-catégories. Ils sont résumés dans la Figure 1, d'après Ellis (op. cit. : 76) :

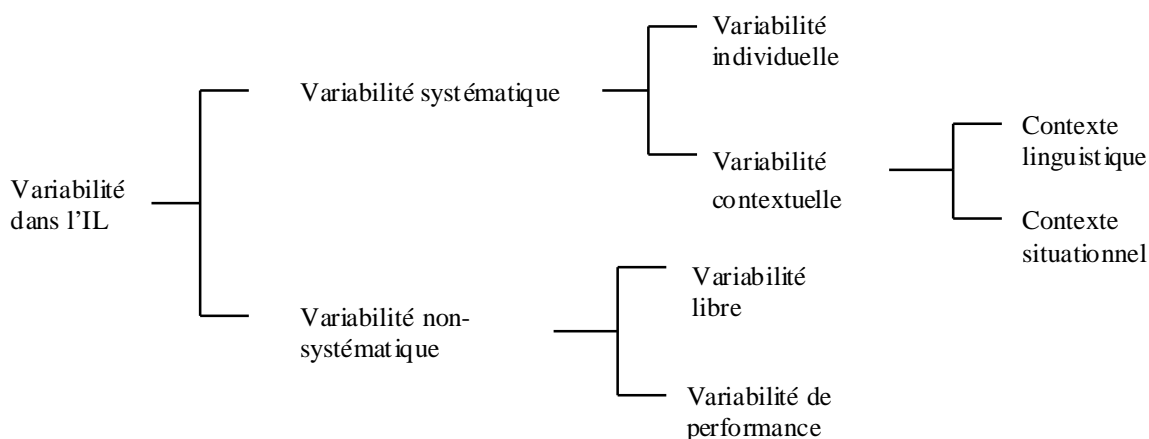


Figure 1. Types de variabilité dans l'IL.

Il va de soi que la variabilité systématique est explicable et prévisible. Elle a deux sous-catégories : individuelle et contextuelle. La variabilité non-systématique, aléatoire et non-prévisible, peut encore être divisée en variabilités libre et de performance. La variabilité individuelle peut être expliquée par des facteurs comme l'âge, la motivation et la personnalité relatifs à l'apprenant (cf. aussi Larsen-Freeman & Long 1992 ; Vogel 1995), tandis que la variabilité contextuelle est due soit au contexte linguistique, soit au contexte situationnel. Dans le contexte linguistique, cela veut dire que l'étudiant utilise différentes variantes, par exemple, pour le phonème français /z/ qui dépend de l'environnement linguistique. Dans le contexte situationnel, la production de l'étudiant varie selon les attentes relatives au style, c'est-à-dire que celui-ci devrait être plus/ moins soigné (cf. Tarone 1988). Vogel (op. cit. : 67-68) appelle cette sorte de variabilité la variabilité par rapport aux situations de production, et il y inclut des facteurs susceptibles d'influencer le comportement de l'étudiant tels que :

²⁷ Ellis ajoute que la variabilité n'est pas seulement une caractéristique des IL, mais qu'elle existe dans tout usage linguistique. Pourtant, il y a une différence de degré, la variabilité des IL est extensive (1992 : 76, 81).

- communication formelle/ informelle (cf. supra)
- communication entre apprenants/ avec des locuteurs natifs
- communication en cours de langue/ en dehors des cours
- langue parlée/ écrite
- tests/ communication libre
- production libre et spontanée/ production guidée ou tests formalisés
- tests ou exercices à trous/ tests exigeant la production de phrases ou de textes
- description/ interview/ résumé/ traduction

Son classement chevauche en partie la variabilité du contexte situationnel et de performance d'Ellis. Pour ce dernier, la variabilité de performance ne concerne pas la compétence en langue étrangère de l'utilisateur, mais elle est plutôt due à sa condition physique et émotionnelle ; cela s'exprime par les lapsus, l'hésitation et la répétition, etc. Quant à la variabilité libre, elle est visible dans l'utilisation fortuite de deux ou plusieurs formes alternatives dans l'IL de l'apprenant (Ellis 1992 : 75-85) ; par exemple il peut utiliser des formes comme *il faut que je suis/sois prêt* dans une même production. Cette variabilité libre, d'après Ellis, est une caractéristique importante dans l'évolution des IL. Corder tient ce phénomène pour un trait essentiel dans l'acquisition en disant que l'apprenant teste ses hypothèses sur la nature d'une L2 à apprendre (1975 : 26-27 ; cf. aussi Selinker 1992 : 150-151). Giacobbe (1990 : 119) parle de l'apprenant qui construit hypothétiquement sa théorie pour développer ses interactions exolingues :

“Mais l'apprenant, exposé à une grande variété de situations d'interaction langagière, peut avancer différentes hypothèses sur un même aspect du système de la langue cible, sans pour autant arriver à en percevoir toujours les contradictions possibles [...]. C'est la possibilité d'arriver à “percevoir” (de façon probablement non consciente) quelques-unes de ces contradictions qui amènent les apprenants à infirmer certaines hypothèses idiosyncrasiques, permettant ainsi l'évolution du système vers des emplois proches de la langue cible.”

Selon Tarone (1988 : 22-58), on peut trouver deux courants théoriques majeurs pour expliquer la variation interlinguale. Le premier s'appelle '*inner processing theories* (théories de procédure innée), et le deuxième, '*sociolinguistic and discourse theories* (les théories sociolinguistiques et du discours). Le premier contient des méthodes, comme la théorie du contrôle (*Monitor*) (Krashen 1981), le modèle de la perméabilité (Adjemian 1976) et les théories basées sur le changement de style dans la L1 (*style-shifting patterns*, cf. *observer's paradox*²⁸) (Labov 1972), tandis que le second comprend des modèles comme le modèle socio-psychologique (Schumann 1978) et le modèle de *form-function* (relations

²⁸ Labov a exposé le paradoxe de l'observateur (*observer's paradox* ; Labov 1972 : 113) ce qui veut dire que le fait de se savoir observé rend la production de l'apprenant moins naturelle (cité dans Preston 1996 : 2). Tous deux, Labov et Preston, parlent de la production orale, mais nous pouvons supposer qu'il en est ainsi aussi à l'écrit.

entre les formes et leurs fonctions sémantiques et discursivo-pragmatiques) (Huebner 1985). Elle n'est satisfaite ni de l'un ni de l'autre de ces courants majeurs puisqu'ils ne peuvent pas, selon elle, expliquer exhaustivement la variation interlinguale. Elle en conclut que les théories les plus adéquates seraient les théories sociolinguistiques et discursives, car elles sont empiriquement mieux vérifiables (Tarone 1988 : 56-57).

Preston regarde partiellement dans la même direction en nous présentant la variation de l'acquisition d'une L2 d'un point de vue sociolinguistique ; en fait, il fait remarquer qu'il existe une relation étroite entre l'étude sur la L2 et la sociolinguistique. Il discute deux modèles, notamment celui de Labov supposant que les formes alternatives d'un élément linguistique n'apparaissent pas d'une façon aléatoire – d'où une approche quantitative quant à la variation linguistique - et deuxièmement, le paradigme dynamique qui explique la variation liée à l'environnement linguistique et à un groupe social changeant avec le temps. Il en conclut que les facteurs les plus importants influençant la variabilité de la performance interlinguale sont linguistiques ; les recherches sur la variabilité pourraient clarifier l'interaction entre la compétence et les autres facteurs qui relèvent de la performance (1996 : 1-2, 18-19, 38-40).²⁹

2.3. Etat actuel de la recherche interlinguale

Comment voit-on la situation aujourd'hui ? Pourquoi s'intéresse-t-on toujours à la recherche interlinguale ? Vu les caractéristiques interlinguales et la place de l'IL dans le domaine de la linguistique appliquée, Vogel (1995 : 276) fait le bilan en concluant que

“La recherche sur l'interlangue apparaît [donc] comme une recherche nécessairement interdisciplinaire et intégrative sur l'enseignement et l'apprentissage des langues, qui étudie aussi bien le développement d'une compétence linguistique que le comportement communicationnel dans la langue-cible, qui conçoit l'apprentissage d'une langue seconde comme un processus variable, non prédéterminé, qualitativement différent de l'acquisition de la langue première, générant des structures linguistiques correctes ou erronées et qui peut s'observer en contexte d'enseignement ou dans tout autre contexte”.

²⁹ Voir par exemple Bogaards (1991 : 33-88) sur les caractéristiques personnelles de l'apprenant qui peuvent soit favoriser soit empêcher l'apprentissage des L2, p. ex. aptitude, intelligence, attitude, motivation, personnalité, âge, sexe, milieu. Il présente aussi un modèle des facteurs déterminant le succès en langue étrangère à l'école (op. cit. : 153) [ici, il fait la distinction entre langue seconde et langue étrangère].

A son avis, les recherches futures devraient être réalisées dans des contextes de communication “réalistes” à toutes les étapes de l’acquisition surtout d’un point de vue diachronique. Il veut en plus quitter le domaine linguistique et psycholinguistique et aller davantage vers la psychologie cognitive et la neurobiologie ; l’étude devrait se placer “dans la perspective idéale de l’élaboration d’une *grammaire universelle des interlangues*” (op. cit. : 286-287). Si on compare notre étude à cette perspective, nous pouvons constater que nous n’allons pas si loin. Nous nous contentons d’une approche linguistique, soit psycholinguistique, mais du point de vue développemental.

Preston (voir supra) souligne également le côté interdisciplinaire de la recherche interlinguale ; son point de vue est plutôt sociolinguistique. Il veut notamment voir une relation étroite entre les recherches sur les IL et sur les aspects sociologiques et socio-psychologiques du langage (1996 : 2). Tout le volume de Bayley & Preston (1996) est consacré à l’approche quantitative de la variation linguistique ; les points de vue varient évidemment selon les articles (voir par exemple la discussion de Labov sur la variation dans la L1 et sa contribution éventuelle à la L2, op. cit. : 245-252). Comme nous l’avons précisé dans l’introduction, notre approche est essentiellement quantitative et se base sur l’évolution interlinguale ; elle suit donc la lignée des études mentionnées ci-dessus.

Giacobbe (1992 : 50-52) voit toujours l’importance de l’influence de la L1 sur l’apprentissage des L2. Pour lui, le recours au système préalable qu’est la L1 fait partie de l’activité langagière de l’apprenant, voire de la démarche hypothétique. Pour pouvoir examiner cette démarche, le chercheur a à sa disposition le moyen de la comparaison contrastive des systèmes des deux langues en question. Il ne pense pas à l’AC au sens traditionnel, mais estime que

“Le recours à la langue première dans la construction de l’interlangue et dans la production du discours doit être ainsi replacé, avec toute sa complexité, dans le cadre cognitif du processus d’acquisition” (op. cit. : 52).

Donc, il accentue le processus cognitif comme Vogel, même si son approche est différente. Pour sa part, Carcedo a pu prouver le transfert continu des phénomènes propres à la L1 dans la L2 au niveau de la prononciation (1998 : 311-312) ; il voit aussi l’analyse contrastive comme un moyen d’améliorer l’enseignement des L2 en utilisant notre savoir linguistique des langues en

question (cf. aussi Härmä 1998). Notre corpus étant composé d'étudiants de français finlandais, nous devons également prendre en considération l'influence de la L1, en particulier dans l'analyse des erreurs (cf. chapitre 8.3.1.)

En 1992, Selinker a fait l'éloge des travaux de Corder en présentant une quarantaine d'hypothèses que celui-ci avait introduites dans la discussion sur les IL qui ne cessent d'influer sur la problématique interlinguale (1992 : 167-168). Il y évoque déjà la réflexion sur la situation actuelle de l'IL ; il souligne, entre autres, le rôle des équivalences de traduction comme stratégie d'apprentissage, et cela dans la perspective du transfert (op. cit. : 259-264). Selinker & Lakshmanan (1994 : 197, 212) se concentrent justement sur cette relation entre le transfert et la fossilisation, un phénomène réel et qui, à leur avis, a besoin d'études longitudinales pour être mieux compris. Dans l'introduction de *The current state of interlanguage* (1997), Eubank, Selinker, et Sharwood Smith font un compte rendu des thèmes centraux dans la recherche interlinguale actuelle [en fait, la situation en 1995]. Tout en restant assez prudents dans leurs réflexions, ils mentionnent néanmoins la possibilité d'arriver à un consensus sur un nombre de thèmes jusqu'à alors controversés. Il s'agit de :

- la grammaire universelle et le transfert existent dans la formation des grammaires interlinguales, mais leur point d'intersection est encore incertain
- l'acquisition de L1 et L2 ne sont pas totalement identiques
- la fossilisation existe ; afin de l'examiner, il faut des études de cas longitudinaux
- les variations de différentes sortes doivent être examinées
- quelques L2 sont plus difficiles que d'autres, mais aussi quelques structures à l'intérieur d'une même L2 sont plus difficiles que d'autres (op. cit. : 8-9).

Pour sa part, Véronique (1992), faisant un état des lieux des recherches sur l'acquisition des langues secondes en 1992, a trouvé deux grandes rubriques concernant les travaux récents qui se concentrent sur le langage oral. Selon lui, les dimensions interactives et les dimensions cognitives de la saisie sont indissociables. Ainsi, l'apprenant collabore-t-il avec un locuteur natif [ou non-natif] dans une situation exolingue pour construire ensemble le sens d'un énoncé. Dans le cas idéal, l'apprenant porte attention aux fonctionnements de la L2 et développe ses activités réflexives sur la langue. En d'autres termes, il y aura une collaboration métalinguistique en interaction (1992 : 10). Véronique voit émerger de ce bilan des zones à explorer, notamment "une analyse plus fine de la prise

d'informations en interaction, un approfondissement de la représentation des opérations cognitives mises en oeuvre lors de la saisie, ou encore un réexamen des 'stratégies' de l'apprenant" (op. cit. : 27). De plus, il ajoute que "le procès d'appropriation devrait permettre d'évaluer quelques grands débats au sein des théories des langues et du langage" (ibid.). Il réfère, en effet, partiellement aux mêmes thèmes qu'Eubank et al. - entre autres - notamment à la prise en compte de l'innéité, à l'hétérogénéité des variétés et leurs dynamiques et à l'existence d'universaux. Ces thèmes, surtout le côté métalinguistique, étaient effectivement présents dans les travaux acquisitionnels³⁰ du colloque FOCAL en 1999³¹. Dans notre étude sur le langage écrit, nous essayons d'apporter notre contribution à cette problématique interlinguale.

³⁰ Ce concept s'utilise principalement pour les travaux sur la communication exolingue.

³¹ Les communications peuvent être consultées dans les actes de la revue électronique *Marges Linguistiques*. In M.L.M.S. (éd.) *Actes du XIe colloque international : Acquisition d'une langue étrangère : perspectives et recherches. Usages pragmatiques et acquisition des langues étrangères*. Paris 19, 20 et 21 avril 1999, France. Disponible à partir du site <http://www.marges-linguistiques.com/>

3. L'INTERLANGUE ET LA DIDACTIQUE DES LANGUES ETRANGERES

On peut postuler de tout ce qui précède que l'interlangue et la pédagogie, voire la didactique, sont étroitement liées l'une à l'autre. D'après Gaonac'h, l'IL est un bon relais entre la linguistique et la psycholinguistique dans le domaine de la didactique des langues étrangères (1985 : 63). Coste examine la situation de la didactique des langues et de la linguistique et en conclut que la didactique des langues s'est éloignée du "noyau dur" de la linguistique et est allée vers les bords de ce domaine, notamment vers l'analyse du discours et la pragmatique (1985 : 10).³² Ceci se voit également dans la collection et la recherche des corpus qui essayent d'éclaircir le processus d'acquisition d'une L2. L'analyse de la performance a aidé à trouver des séquences développementales et des stratégies d'apprenant dans les IL, tandis que l'analyse du discours a éclairé la relation entre l'input et les formes interlinguales, par exemple le recours aux stratégies de communication (voir infra) (cf. Larsen-Freeman & Long 1992 : 62-74).

Porquier déclare que "l'enseignement des langues a pour premier objectif l'acquisition d'une compétence de communication" en L2 (1984 : 42). Pour y arriver, il propose que :

"[...] l'apprentissage d'une langue implique l'apprentissage de la communication exolingue et interculturelle, et des pratiques exolingues d'apprentissage [...]. On n'apprend pas seulement à communiquer dans une langue n [= langue étrangère quelconque], mais aussi à communiquer ailleurs qu'en L1, dans une diversité de situations exolingues dont l'enseignement institutionnel ne peut préjuger. Aussi paraît-il nécessaire d'intégrer à l'enseignement/apprentissage initial des langues, surtout pour les adolescents et les adultes, une sensibilisation au fait exolingue et à la communication exolingue et interculturelle" (op. cit. : 44).

En vue de cet objectif, nous pouvons d'emblée, passer à la discussion sur les erreurs qui jouent un rôle essentiel dans l'apprentissage d'une L2 puisqu'elles contribuent à élaborer la compétence de communication. L'étude longitudinale constitue le meilleur moyen de les détecter.

³² Voir aussi la discussion de Cortès & Cortès sur une taxinomie de la recherche en didactique des langues (1987 : 9-20).

3.1. Importance des erreurs dans l'apprentissage d'une langue seconde

Maints chercheurs ont été intéressés par la véritable situation d'apprentissage et par différents facteurs influant sur celle-ci. Les erreurs forment une partie primordiale dans ce processus et c'est pourquoi leur rôle dans l'acquisition d'une L2 a été abordé par plusieurs chercheurs et professeurs de langues. On affirme très souvent que plus il y a d'éléments en L2 causant des difficultés, plus les apprenants produisent d'erreurs dans leur production écrite ou orale. Par exemple, Lado écrit dans *Linguistics Across Cultures* (1957) à propos de l'apprenant d'une L2 :

"Those elements that are similar to his native language will be simple for him, and those elements that are different will be difficult" (cité dans James 1981 : 188).

Cette hypothèse suggère que le degré de proximité entre la L1 et L2 prédit la quantité d'erreurs que les apprenants d'une certaine L1 commettent et de quelle nature elles seront vraisemblablement. Pourtant, ce n'est pas toujours le cas. Les apprenants semblent éviter les constructions difficiles (la stratégie d'évitement ; voir infra), et ainsi, selon James, ils ne commettent sans doute pas beaucoup d'erreurs dans les constructions présumées difficiles à l'avance (1981 : 183).

Jusqu'à assez récemment, les études portant sur l'acquisition d'une L2 ont été principalement préoccupées par les erreurs des apprenants. L'enseignement consistait à prévenir ou à minimiser les erreurs - en adoptant une attitude négative vis-à-vis de ces dernières - et à mesurer les connaissances langagières en comptant le nombre d'erreurs. Elles étaient le signe d'un apprentissage incomplet et devaient être éradiquées pendant l'apprentissage (cf. Galisson 1980 : 57-64 ; Johansson 1978 : 167). En d'autres termes, les erreurs étaient une preuve du fait que les habitudes correctes et automatisées d'une L2 n'étaient pas encore acquises. Pour les éradiquer, il fallait répéter les formes correctes intensivement (*drilling*).³³ La nature des erreurs n'était pas signifiante. Aujourd'hui, le fait de commettre des erreurs est considéré comme une partie inévitable et même un composant nécessaire dans le processus d'apprentissage. La correction des erreurs fait découvrir à l'apprenant le concept ou la règle corrects (cf. Corder 1971 : 169-170). A l'instar d'un proverbe italien, Porquier résume bien cette idée dans la phrase suivante : "On n'apprend pas sans faire d'erreurs et les erreurs servent à apprendre" (1977 : 28).

³³ Voir p. ex. Gaonac'h sur l'influence du behaviorisme sur la méthode audio-orale (1991 : 11-35).

L'identification des erreurs n'est pourtant pas toujours facile. Avant d'analyser les erreurs des apprenants plus précisément, Corder (1975 : 25) remarque qu'il faut faire la distinction entre les vraies **erreurs** (*errors*) de compétence et les **fautes** et **lapsus** (*mistakes and lapses*) faisant partie de la performance³⁴. Les fautes et les lapsus sont des formes erronées qui peuvent être également produites par des locuteurs autochtones, voire natifs, tandis que les erreurs de compétence sont produites seulement par les apprenants d'une L2. Normalement, nous sommes tout de suite conscients des fautes quand elles apparaissent dans la production et nous pouvons les corriger assez facilement ; elles sont dues aux lapsus de mémoire, à l'état physique, etc. C'est pourquoi elles doivent être séparées des erreurs de compétence des apprenants qui sont systématiques et révèlent leur compétence transitionnelle (Corder op. cit. : 24-25). Corder lui-même critique plus tard les notions **erreur**, **déviant** ou **mal formé** comme mal appropriées à la description des phrases idiosyncrasiques car, selon lui, ces notions correspondent avant tout à des préjugés concernant les performances de l'apprenant (1975 : 162-163)³⁵. Nous nous passons de cette discussion épistémologique. Elo (1993 : 11-12) ne veut pas parler de fautes (ni d'erreurs), mais plutôt de **déviations** de la norme linguistique, car d'après elle, il ne serait pas "logique de qualifier une interlangue fautive, étant donné qu'elle forme un système à part". Hammerly propose de parler de **faute** (*fault*) pour désigner les erreurs que les apprenants commettent quand la forme n'est pas encore apprise et de **déformation** (*distortion*) pour les formes déjà apprises, mais mal utilisées (1991 : 85-87). De plus, il veut encore distinguer les fautes et déformations de gestion maladroite (*mismanagement distortions and faults*) ce qui veut dire, entre autres, que les formes ont été apprises de façon inadéquate lors de l'enseignement. Nous parlerons des erreurs et des fautes dans l'étude présente (cf. infra).

Durant une période relativement longue, les chercheurs suggéraient que la raison la plus évidente des erreurs était l'interférence, c'est-à-dire le transfert négatif de la L1. Plus tard, on a découvert que d'autres langues, apprises auparavant, pouvaient également influencer sur la production. Ces erreurs sont aussi qualifiées

³⁴ Il faut noter que Galisson (1980 : 57) explique autrement la différence entre *faute* et *erreur*. Nous utilisons les termes comme présentés dans l'article de Corder (1980 : 13) traduit par Perdue et Porquier.

³⁵ Corder pense que le dialecte idiosyncrasique est grammatical car il faut l'examiner du point de vue de la langue de l'apprenant, et non en comparaison avec la langue-cible (1975 : 163).

d'erreurs **interlinguistiques**³⁶. Il faut cependant noter que le transfert peut aussi être positif. Cela signifie que les phénomènes en question sont fonctionnellement identiques dans les deux langues d'une part, et que le transfert (et donc le résultat) a réussi d'autre part. Si le transfert est négatif, il indique une sorte d'ignorance sur le système d'une L2 (cf. James 1981 : 22). Pour leur part, les erreurs **intralinguistiques** sont dues à la surgénéralisation ou à la simplification des règles syntaxiques de la langue-cible, ou bien aux stratégies communicatives ou d'apprentissage, ou encore aux procédures d'entraînement spécifique, par exemple quand le professeur utilise seulement le pronom *il* et non *elle* dans les exercices (cf. Gaonac'h 1991 : 126-128 ; Jain 1975 : 189-215 ; Larsen-Freeman & Long 1992 : 58-59). Hammerly (1991 : 63-65) parle d'interférence intruse/importune (*intrusive interference*) et d'interférence inhibitive (*inhibitive interference*). La première aboutit à une conduite inappropriée dans un nouveau contexte, et la deuxième inhibe l'apprentissage. Il ajoute que dans le cas d'interférence inhibitive, les apprenants auraient tendance à ne pas apprendre de structures d'une L2 qui sont inexistantes dans leur L1.

De plus, afin de détecter et de déterminer les erreurs, il faut encore introduire les termes **acceptabilité** et **grammaticalité**. Selon Enkvist (1977 : 1), l'acceptabilité veut dire la relation entre l'énoncé et son évaluation par un groupe d'informateurs sélectionné (dépendant du contexte), tandis que la grammaticalité réfère à la relation entre l'énoncé et la description spécifique de la langue en question, c'est-à-dire la grammaire - là où les autochtones utilisent normalement leur intuition. Il va de soi qu'il est souvent difficile pour l'apprenant d'une L2 de trouver les formes acceptables en question (même si grammaticales), et de plus, qu'il ne peut comprendre la grammaticalité que des phénomènes déjà appris. Schachter et al. (1976 : 70) expriment la grammaticalité entre les natifs et les apprenants d'une L2 de la façon suivante :

³⁶ Ces termes d'**inter-** et **intralinguistique** sont utilisés dans le sens de Vogel (1995 : 17) ; nous utilisons le terme **interlingual**, entre autres, au sens du développement interlingual.

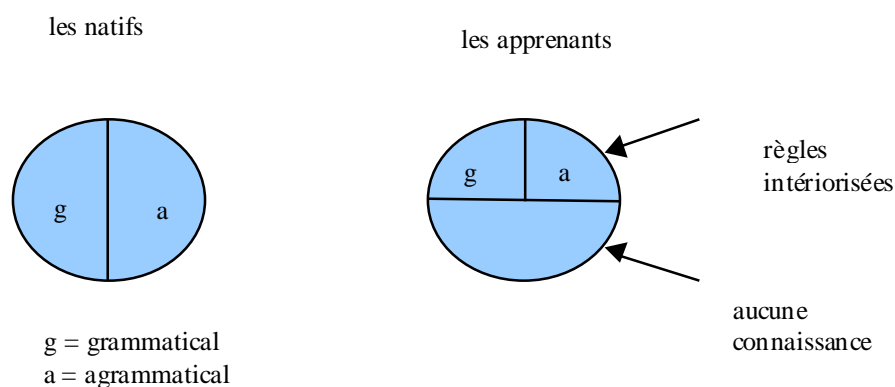


Figure 2. Grammaticalité de la langue

Riegel et al. (1997 : 19-20) déclarent que l’acceptabilité fait fondamentalement partie des phrases énoncées et dépend de tous les facteurs qui conditionnent la performance ; ainsi serait-il possible de trouver un ou plusieurs contextes dans le(s)quel(s) l’interprétation d’une même phrase ne poserait pas de problèmes. La grammaticalité (au sens étroit ou au sens large³⁷) ne recouvre ainsi qu’une partie de l’acceptabilité ; une phrase grammaticale peut être non-acceptable, entre autres, pour des raisons de longueur, de complexité ou d’obscurité. En outre, Vogel voit mal comment l’apprenant pourrait juger – contrairement au locuteur natif – la grammaticalité de sa langue, car

”Dès que l’apprenant s’exprime, ses propos ont une existence et lui semblent, de ce fait, grammaticalement praticables” (1995 : 78).

Selon lui (op. cit. : 79), cela fait partie de l’élaboration des hypothèses que l’apprenant teste et révisé afin de s’approcher des normes de la langue-cible (cf. chapitre 2.2. ; Vogel op. cit. : 109-118).

3.2. Compétence communicative

Etant donné que la nature des erreurs et des fautes est discutable (cf. supra) et qu’elles sont quelquefois difficiles à découvrir sans avoir recours au contexte, nous allons traiter maintenant brièvement ce point de vue. Afin d’analyser la pertinence de la production de l’étudiant dans un certain contexte, il faudrait recourir à des méthodes encore plus élaborées que l’analyse contrastive ou

³⁷ Brièvement, la grammaticalité au sens étroit : règles morphologiques et syntaxiques dans une grammaire traditionnelle ; au sens large : règles morphologiques, syntaxiques, sémantiques et éventuellement pragmatiques (cf. Riegel et al. 1997 : 20).

l'analyse des erreurs, notamment à l'analyse de la performance et du discours³⁸. La première souligne toute la production de l'apprenant, alors que la seconde traite des aspects de la communication dans des situations interactives variées (cf. Larsen-Freeman & Long 1992 : 62-74 ; Sajavaara 1980 : 210). Le domaine de l'analyse de la performance et du discours est un sujet extrêmement intéressant, mais beaucoup trop vaste pour être étudié exhaustivement ici. Néanmoins, dans le chapitre 6, nous allons aborder la question de l'analyse textuelle, mais seulement du point de vue du cadre théorique concernant ce travail. La partie de l'analyse du corpus (ch. 8.3.1.) comprendra une étude sur les vocabulaires des étudiants en contexte ; nous aurons alors recours à l'environnement linguistique pour examiner, entre autres, les formes erronées et tenter d'expliquer leur origine. Ces analyses rendent donc possible l'étude sur la compétence de communication, ou compétence communicative, des apprenants.³⁹ Comme on l'a dit supra, celle-ci est aujourd'hui l'objectif cognitif principal dans l'enseignement des langues (cf. Leino 1980: 172 ; Porquier 1984 : 42). James, référant à Hymes (1972)⁴⁰ définit la compétence de communication comme suit :

“a language user who is able to put the structural resources to use as functionally appropriate is said to have communicative competence” (1981 : 162).

Il existe plusieurs modèles mettant l'accent sur les caractéristiques multicomponentielles de cette compétence.⁴¹ Canale & Swain (1980 : 27-31) distinguent trois composantes au minimum, notamment :

- a. compétence grammaticale – connaissances des structures linguistiques
- b. compétence sociolinguistique – règles socioculturelles et discursives
- c. compétence stratégique – capacité à utiliser des stratégies de communication pour atteindre les buts communicatifs

Moirand (1990 : 20), pour sa part distingue quatre composantes :

- a. compétence linguistique
- b. compétence discursive – connaissance des différents types de discours en liaison avec la situation

³⁸ Ici, nous délimitons ces deux termes du point de vue de l'apprentissage des langues.

³⁹ La compétence et les connaissances de l'apprenant sur la langue sont testées de façons multiples (*proficiency tests* en anglais) ; voir par exemple Carroll (1975 : 47-79) et Stern (1987 : 341-359).

⁴⁰ Hymes discute les différentes acceptions de la compétence de communication dont les origines semblent venir de la grammaire générative transformationnelle et de l'ethnographie de la communication (1984). Il refuse le concept de compétence de Chomsky, celui-ci étant selon lui trop étroitement lié aux traits linguistiques. Il est pour une acception plus large du concept de compétence. Il admet que cette notion est utilisée pour désigner diverses capacités, mais selon lui, c'est un terme bien utile comme terme général. Hymes souligne l'importance du langage dans la vie sociale. ; cf. aussi Besse & Porquier (1984 : 234-239) et Bolton (1991 : 24-25).

⁴¹ Cf. Gaonac'h (1991 : 177-187) sur la même question. Rück discute la compétence de communication du point de vue de la compétence textuelle ; d'après lui, si l'on comprend le concept de texte au sens large, la compétence communicative et la compétence textuelle recouvrent la même chose (1980 : 52).

- c. compétence référentielle – connaissance du monde
- d. compétence socioculturelle – connaissance des règles sociales et des normes d’interaction

Ces modèles se chevauchent en partie ; par exemple la compétence stratégique de Canale & Swain ne correspond que partiellement à la compétence socioculturelle de Moirand, qui, pour sa part, recouvre en partie la compétence sociolinguistique de Canale & Swain. Moirand ajoute que, pour elle, “les ‘stratégies’ [cf. Canale & Swain] ne relèvent pas d’une ‘compétence stratégique’, [...] mais interviendraient lors de l’actualisation de cette compétence dans une situation de communication concrète” (op. cit. : 20). Faerch et al. (1984 : 167-169) décrivent sept sous-catégories qui forment quatre composantes principales. Leur modèle chevauche aussi les deux modèles précédents. On peut présenter la relation entre les différentes composantes de la compétence communicative de la façon suivante (d’après Faerch et al. 1984 : 169) :

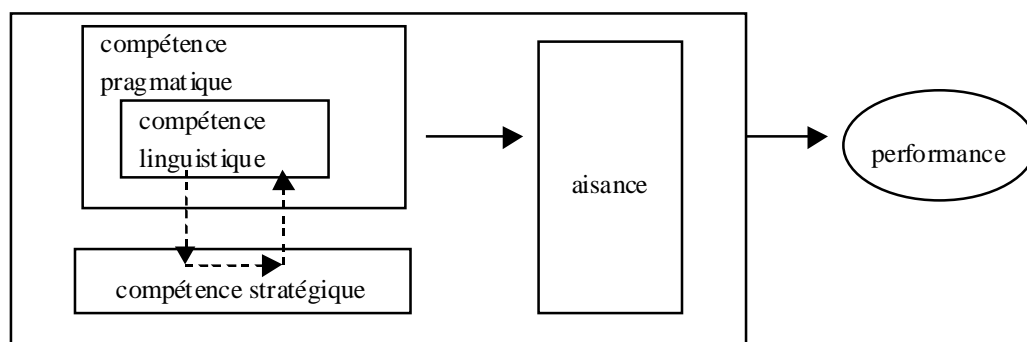


Figure 3. Composantes de la compétence communicative

Dans ce modèle, la compétence linguistique comprend les connaissances orthographiques/phonologiques, grammaticales et lexicales et la compétence pragmatique contient les connaissances pragmatiques et discursives, tandis que la compétence stratégique réfère à la capacité du locuteur à résoudre des problèmes communicatifs à l’aide de stratégies (voir ci-dessous). Finalement, l’aisance, qui selon Sajavaara (1980 : 132) est un continuum naturel et acceptable de parole, réfère à la capacité du locuteur à s’exprimer aisément, de façon naturelle (cf. Freed 1995 : 123-148). Cela ne veut pas dire qu’il n’y ait pas d’hésitation et de pauses dans la production, mais que ces phénomènes sont exprimés de façon naturelle, comme chez un locuteur natif. L’aisance, sous cette forme, réfère souvent au langage parlé, mais à l’écrit il existe sous une autre forme, notamment, à travers des connecteurs et d’autres éléments intra- et interphrastiques qui augmentent la cohérence/la cohésion du texte (cf. Bouchard 1994 ; Péry-Woodley 1993 : 28-29 ; voir chapitre 6).

Communiquer dans une langue étrangère n'est pas toujours facile, et les apprenants doivent quelquefois recourir à des stratégies variées pour transmettre leur message d'une manière réussie et efficace. Tarone distingue trois sortes de stratégies qui font partie des processus utilisés par l'apprenant⁴². Les limites entre ces stratégies sont assez floues, d'autant plus que l'apprenant peut les utiliser aussi bien de façon consciente qu'inconsciente. Selon Tarone (1980 : 419-420), il existe a) les stratégies d'apprentissage dont le but principal est d'apprendre, par exemple en mémorisant une structure, b) les stratégies de production dont le but est d'utiliser le système linguistique de manière efficace et claire, et c) les stratégies de communication, par lesquelles les interlocuteurs essayent de se mettre d'accord mutuellement sur le sens dans des situations ; ces catégories (b+c) peuvent être regroupés sous le nom de "stratégies de l'utilisation de langue". Nous nous contentons de parler ici des stratégies de communication. Corder définit les stratégies de communication de la façon suivante :

"Strategies of communication have essentially to do with the relationship between ends and means. In a native speaker it is ideally assumed that these are in balance, that is, that he always has the linguistic means to express the messages he wishes to communicate. In a learner, however, these are not in balance [...]. He can either tailor his message to the resources he has available [...]. These procedures we can call *message adjustment strategies* [...]. Or he can attempt to increase his resources by one means or another in order to realize his communicative intentions. These strategies we can call *resource expansion strategies*" (1992 : 17).

Selon lui, les stratégies d'**adaptation de message** sont, entre autres, l'évitement du sujet, tandis que les stratégies d'**expansion de source linguistique** comprennent, par exemple, le changement de code (*code-switching*). Tarone, pour sa part, a fait une liste des stratégies que des apprenants ont utilisé (1980 : 429⁴³ ; cf. 1992 : 62-63). Elle trouve trois catégories primaires, qui sont la **paraphrase** (p. ex. *une machine à faire la vaisselle pour lave-vaisselle*), le **transfert** (p. ex. *assez beaucoup - possible en finnois - pour suffisamment*) et l'**évitement** (*avoidance* en anglais, p.ex. la rupture de communication) (cf. Larsen-Freeman & Long 1992 : 126-127 ; Tarone et al. 1992 : 4-14). Faerch et al. (1984 : 154-162) divisent les stratégies en deux groupes principaux, notamment les stratégies

⁴² Selinker a déjà introduit les stratégies d'apprentissage et de communication comme faisant partie des processus psycholinguistiques essentiels dans l'apprentissage d'une L2 (1975 : 35). Voir cependant la discussion de Vogel sur ces processus (1995 : 35-38). Voir aussi les exemples d'Elo sur les stratégies d'apprentissage qui chevauchent les stratégies de communication traitées ici (1993 : 12-13). En outre, Stern présente une liste de dix stratégies d'un bon apprenant de langue (1987 : 406, 414-415).

⁴³ Cette liste date originellement de Tarone (1978).

d'**atteinte**⁴⁴ et de **réduction**. La liste de Tarone correspond à leur catégorie dans la mesure où la paraphrase et le transfert sont considérés comme stratégie d'atteinte et l'évitement comme stratégie de réduction. Les stratégies d'adaptation de message de Corder correspondent à peu près à la stratégie de réduction, et ses stratégies d'expansion de source linguistique à la stratégie d'atteinte de Faerch et al. Pour sa part, Perdue (1993: 14) les appelle stratégies compensatoires et d'évitement.⁴⁵ Quant au profit des stratégies pour l'apprenant, Gaonac'h suggère que les stratégies d'évitement ne satisfont pas la demande d'efficacité de communication car elles visent plutôt l'économie ou la prudence ; la prudence peut néanmoins contribuer, sur le plan cognitif, à la construction du système provisoire (1984 : 76).⁴⁶ Haastrup & Phillipson (1992 : 156-157) sont plutôt d'un avis mitigé : les apprenants semblent utiliser les stratégies d'après un choix personnel et ainsi leurs stratégies de communication ne changeraient pas, même si l'enseignant essayait d'influencer leur choix (cela peut être dû aux stratégies déjà acquises en L1, cf. Herschensohn 1999). D'autre part, ils trouvent positive la tentative d'essayer de rendre les apprenants conscients des stratégies en général y compris leurs propres styles stratégiques (voir ci-dessous).

D'après les définitions, les stratégies de communication semblent, référer plutôt au langage parlé, mais nous postulons qu'on peut les utiliser aussi dans le langage écrit. D'autant plus que Hymes suggère que la compétence de communication s'applique à la compétence dans et pour la communication qui

“est un terme général, qui inclut aussi la réflexion et le dialogue avec soi-même. En écrivant maintenant à la main ces lignes ajoutées à une page tapée à la machine, j'entre en interaction avec moi-même à différents moments, passés, présents, et futurs, dans le mouvement d'une pratique de communication somme toute courante” (1984 : 129).

Donc, la communication peut être intérieure ou extérieure et toutes deux s'appuient sur des stratégies s'il le faut (cf. chapitre 4.3.)⁴⁷. Les stratégies

⁴⁴ Le terme de **stratégie d'atteinte** (*achievement strategy*) est emprunté à Norro (1995). Herschenchon, entre autres, utilise les termes **stratégies d'évitement** et **stratégies d'accomplissement** (1999).

⁴⁵ Par exemple, Noyau & Porquier (Eds.) (1984) et Faerch & Kasper (Eds.) (1992) présentent un ensemble d'études de cas sur le langage oral où les stratégies de communication sont explicites. Voir aussi Norro (1995) à propos du même sujet, mais où la langue source est le finnois et la langue-cible est le français ; dans l'étude de Herschenchon (1999), la langue source est l'anglais et la langue-cible, le français.

⁴⁶ Blum-Kulka & Levenston (1992 : 119-139) font preuve de l'utilisation des stratégies pour compenser surtout le manque de connaissances lexicales.

⁴⁷ L'idée que tout discours, même monologique, est fondamentalement dialogique par structure vient de Bakhtine (cité dans Bessonnat 1988 : 83).

d'évitement sont probablement plus utilisées à l'écrit, mais difficiles à déceler car elles ne sont pas explicites (cf. James 1981 : 183 ; Mutta 1995 : 45).

3.3. Compétence métacommunicative

Faerch et al. postulent de plus que la compétence de communication devrait incorporer dans l'enseignement des langues étrangères aussi bien la capacité linguistique que la conscience métacommunicative, c'est-à-dire "la connaissance consciente des composantes de la compétence de communication, de leur interdépendance, et de leurs fonctions sociales" (1984 : 178). Gaonac'h parle des degrés de métacommunication, "des activités orientées plus ou moins vers un problème de communication" (1991 : 181). Tout apprentissage et enseignement semblent prôner le processus mental actif, car selon Vogel "*il n'y pas d'apprentissage sans compréhension*" [italique de l'auteur] (1995 : 280).

Pour expliciter le concept de **métacommunication**, retournons un peu en arrière. Le terme de **métacognition** vient du domaine de la psychologie comme tant d'autres termes concernant l'apprentissage. La métacognition est un phénomène qui est connu depuis longtemps, mais le concept lui-même n'est devenu largement utilisé qu'à la fin des années 1970 (Vauras & Silvén 1985 : 1). Selon Vauras & Silvén, la métacognition, c'est-à-dire la conscience de ses propres processus cognitifs et de leur régulation, est considérée comme un phénomène central dans le développement et la direction de l'activité intellectuelle (op. cit. : 1). Déjà dans les années 1930, Vygotski, un psychologue russe, a vu que la conscience et le contrôle volontaire étaient les signes d'activités intellectuelles. Il postule (1997 : 309-310) que

"[...] c'est justement à cet âge, comme le montrent les recherches, qu'apparaissent au centre du développement *les fonctions psychiques supérieures, dont les traits distinctifs fondamentaux sont précisément l'intellectualisation et la maîtrise, c'est-à-dire la prise de conscience et l'intervention de la volonté.*

Le fait central à l'âge scolaire est le passage des fonctions inférieures de l'attention et de la mémoire aux fonctions supérieures de l'attention volontaire et de la mémoire logique.[...] Cela tient à ce que l'intellectualisation des fonctions et leur maîtrise sont deux moments d'un seul et même processus : le passage aux fonctions psychiques supérieures. Nous maîtrisons une fonction dans la mesure où elle s'intellectualise."

Dans la métacognition on peut également distinguer divers niveaux. En nous référant à l'apprentissage dans un milieu scolaire, ou guidé, nous pouvons grosso modo faire la distinction entre la prise de conscience et l'automatisation des choses à apprendre. La distinction krashenienne entre l'apprentissage conscient et l'acquisition inconsciente d'une langue n'est pourtant pas la seule façon d'aborder la problématique. Faerch (1986 : 126-127), entre autres, à partir des exemples de règles de grammaire données à apprendre, voit la situation d'une manière plus complexe ; pour lui les connaissances implicites (c'est-à-dire, l'utilisation de la règle sans réflexion) et les connaissances explicites (la description de la règle en termes métalinguistiques) sont placées sur une échelle graduelle (cité dans Bange 1999 : 2-3). Bange postule que cette façon de traiter la question est satisfaisante car, dans ce cas-là, la différence "entre l'implicite et l'explicite n'est pas conçue comme une différence catégorique [...], mais comme un passage graduel" (1999 : 4). Il justifie son point de vue en concluant qu'à chaque niveau de l'échelle le contrôle cognitif est présent. A l'instar de Gombert (1990 : 20-22) auquel il se réfère, Bange fait la distinction entre **activité métalinguistique** et **activité épilinguistique**⁴⁸. La première est, au sens strict, une activité consciente, réfléchie et délibérée qui fonctionne au niveau le plus haut, tandis qu'aux autres niveaux on peut parler d'une activité métalinguistique inconsciente, notamment épilinguistique, qui consiste en une autorégulation et en une auto-observation interne (Bange 1999 : 4-5). Il ajoute que c'est le degré d'attention qui varie sur cette échelle mentionnée par Faerch et qui transforme la nature du contrôle (op. cit. : 5).⁴⁹ En d'autres mots, selon Griggs, "l'automatisation des processus cognitifs ne signifie pourtant pas la disparition du système de contrôle", mais la différence réside "essentiellement dans la présence ou non de l'attention du sujet" (1999 : 7).

Nous pensons que l'exigence d'incorporer dans l'enseignement des langues étrangères aussi bien la capacité linguistique que la conscience métacommunicative est justifiée, car l'ignorance crée une sorte d'incertitude

⁴⁸ Le terme de Culioli signifie d'après Gombert (1990 : 22) "les activités épilinguistiques (...) [qui] représentent l'autoréférenciation implicite automatiquement présente dans toute production linguistique" (cité dans Bange 1999 : 4-5).

⁴⁹ Bange nous rappelle qu'il faut se garder de confondre l'attention et la conscience :

"1. il n'y a de conscience que dans le cadre de l'attention

2. il est faux qu'il n'y ait d'attention que dans le cadre de la conscience ; il existe une attention non-consciente" (1999 : 5). Nous n'entrons pas plus en détail dans cette discussion.

(partie cependant essentielle et non-négligeable de l'apprentissage). Grâce à la connaissance métacognitive, l'apprenant pourrait peut-être diminuer cette incertitude et ainsi aboutir à une communication réussie, par exemple en ayant recours aux stratégies de communication (cf. aussi Faerch & Kasper 1992: 55-56 ; Pendanx 1998 : 46-48). Nous sommes néanmoins consciente du fait que l'utilisation extrême des stratégies de communication peut même inhiber l'acquisition d'une L2 car une fois que les besoins de la communication sont assurés, la fossilisation (c'est-à-dire la stagnation de l'IL à un niveau erroné en dépit d'efforts pédagogiques) peut entrer en jeu et interrompre ainsi le développement interlingual (cf. chapitre 2.2.).

3.4. Etude longitudinale

Etant donné que notre but est de montrer comment le séjour en France dans le cadre d'un échange universitaire (d'abord interuniversitaire et ensuite ERASMUS) a influencé les connaissances lexicales des étudiants universitaires finnophones en français, et vu notre cadre théorique, notre corpus et nos hypothèses, il nous semble pertinent de parler aussi de l'évolution des interlangues. Comme nous l'avons déjà dit supra, les interlangues sont instables et susceptibles aux changements qui les mènent de plus en plus vers une langue-cible (cf. ch. 2.2. ; voir aussi Besse & Porquier 1984 : 227-234). Nos deux corpus (primaire et du groupe de contrôle) ayant été recueillis à deux moments différents (avant et après un séjour linguistique), nous pourrions les examiner du point de vue synchronique, mais notre objet de recherche serait alors plutôt la performance (cf. Elo 1993 ; Mutta 1995, 1997). Or, pour pouvoir examiner la compétence des étudiants, il faut recourir à une étude développementale. De cette manière, nous essayons de constater le progrès dans le temps et éventuellement de détecter des formes stabilisées des IL.⁵⁰

Gaonac'h parle même d'appauvrissement de la notion d'interlangue lorsqu'on a recours à l'analyse synchronique pure car celle-ci ne peut se comprendre dans le cadre évolutif de l'acquisition des systèmes successifs orientés vers la compétence en L2. Du point de vue pédagogique, il propose de plus une hypothèse sous-jacente à ces développements : "l'élève a une certaine connaissance [...] de 'son'

⁵⁰ Il faut pourtant noter que tous les changements ne sont pas dus au séjour linguistique en France, car l'enseignement de français dans notre département consiste également en cours de production écrite.

interlangue à chaque moment de son apprentissage, et [qu'] il faut lui permettre d'en tirer le meilleur parti" (1984 : 66, 77). Ceci nous renvoie encore à la question du profit de la compétence métalinguistique dans l'enseignement des langues étrangères.

La variation dans le temps peut aussi être vue d'une autre manière. Regan mentionne les plans vertical et horizontal (1996 : 178-179). Le progrès sur le plan vertical comprend l'acquisition de structures de langue, tandis que sur le plan horizontal, le progrès comprend l'acquisition d'une compétence sociolinguistique ou dialectale ; les connaissances du premier plan précèdent celles du deuxième. D'après l'exemple de Regan sur la négation française dans la langue parlée, il existe une variation entre **je ne vois* et *je ne vois pas* sur le plan vertical, et *je vois pas* et *je ne vois pas* sur le plan horizontal ; dans le langage écrit, bien évidemment, la forme *je vois pas* serait normalement précédée d'un astérisque. Regan a trouvé une évolution nette de la compétence sociolinguistique d'apprenants anglophones après un séjour linguistique en France (op. cit. : 197 ; cf. aussi Regan 1995 : 245-267). Ce qui nous intéresse ici, c'est l'intervention éventuelle des traits du langage oral dans les productions écrites des étudiants dans notre corpus (voir nos hypothèses, ch. 1.3.). Ces traits peuvent ainsi être considérés soit comme des formes erronées par rapport aux règles normatives, soit comme des preuves d'un apprentissage sur le plan horizontal (voir chapitre 8.3.1.).

De plus, Bartning (1994 : 24) nous rappelle la discussion entre les recherches développementales faites sur l'acquisition en milieu social naturel ou non guidé et les recherches faites en milieu institutionnel scolaire ou guidé.⁵¹ Dans ces recherches, il s'agit d'examiner, entre autres, si les apprenants, dans un milieu scolaire, apprennent également de façon implicite et dans quelle mesure, et si l'intervention pédagogique facilite l'apprentissage. Elle-même a mené un projet à partir d'un corpus (essentiellement) parlé concernant l'acquisition et le développement des étudiants universitaires suédois (cf. op. cit. : 25-27). Dans notre cas, il s'agit de comparer deux corpus différents : celui d'étudiants universitaires ayant passé un semestre ou une année académique dans une

⁵¹ On parle souvent de l'immersion d'une L2, surtout dans un milieu naturel. Dans un milieu guidé, on parle plutôt de bains linguistiques. Pourtant, il faut remarquer qu'il y a aussi de l'apprentissage par immersion dans un milieu guidé, mais il s'agit alors plutôt de sociétés bilingues, par exemple le Canada et la Finlande (en Finlande, il s'agit de l'enseignement du suédois qui est une langue minoritaire).

université française et celui d'étudiants universitaires ayant passé au moins quatre semaines en France, plutôt dans des conditions naturelles ou non guidées (cf. ch. 1.2.). Notre corpus primaire peut être considéré comme étant situé entre les deux extrêmes ; nous pouvons parler de l'acquisition mixte. Même si les étudiants en question participent à l'enseignement universitaire, il ne s'agit cependant pas de cours de langue à part, mais de cours pour les étudiants (francophones) autochtones.

Quant au processus d'apprentissage, Vogel (1995 : 134-144) fait des remarques sur l'input linguistique optimal (par exemple l'effet de la simplification linguistique par les enseignants) et compare l'input avec les situations d'acquisition, soit naturelles soit institutionnelles. Il en résume six points essentiels tout en montrant les différences (par exemple le degré du feed-back correctif), et il présente des opinions pour et contre les deux situations, mais il en conclut que l'apprentissage institutionnel a une supériorité sur l'apprentissage naturel qui est indépendante de l'âge et du niveau linguistique du départ.⁵² Une de nos hypothèses est que le séjour dans un cadre universitaire améliore les connaissances lexicales plus efficacement qu'un simple séjour linguistique. Autrement dit, même si la situation d'apprentissage des étudiants du corpus primaire n'est pas purement institutionnelle, donc guidée, nous supposons que l'apprentissage mi-institutionnel dans un milieu naturel surpasse en qualité et en quantité l'apprentissage dans un milieu naturel. Pourtant, il faut se souvenir du fait que, dans le groupe témoin, la plupart des étudiants ont passé environ un mois à l'étranger, qu'un tiers d'entre eux y ont séjourné entre 6 mois et 18 mois (voir chapitre 1.2.), et qu'une partie de ces derniers étudiants a assisté à des cours de langue du type "Alliance française" qui peuvent être considérés comme de l'enseignement institutionnel. Nous postulons cependant que cela n'a pas eu d'influence sur nos résultats car a) le séjour est relativement court, et que b) même si le séjour était plus long, les cours de langue de ce type n'améliorent pas nécessairement les connaissances requises à un niveau universitaire avancé. En outre, nous nous concentrons sur le vocabulaire utilisé dans les essais en situation d'examen, ce qui exige une certaine réflexion sur la production écrite. Certes, le groupe témoin sert à vérifier nos hypothèses.

⁵² Voir aussi la discussion de Bogaards (1991 : 142-148) sur l'apprentissage du vocabulaire dans le contexte scolaire et social.

En 1988, Tarone a présenté ce qui lui semblait important pour des recherches futures sur la variation interlinguale. Comme d'autres chercheurs, elle considérait que les études longitudinales étaient pertinentes et elle trouvait l'approche quantitative nécessaire pour révéler la fréquence des phénomènes étudiés - elle soulignait l'intérêt, par exemple de la comparaison de deux situations différentes à l'aide de l'analyse corrélacionnelle entre plusieurs variables (1988 : 131, 136-138). Cela est également notre objectif. Ajoutons qu'elle se référait aux études faites sur le langage oral car la plupart des études longitudinales faites sur la variation traitent de la production orale et de son développement.⁵³ Il en est de même dans l'étude de Trévisé (1985) sur les réfugiés politiques latino-américains en France, qui présente une idée intéressante quant à notre propre travail. Elle trouve difficile de mesurer objectivement la compétence d'un locuteur d'interlangue "car c'est une mesure de l'activité de reconstruction du sens par le francophone, donc par quelqu'un qui ne parle pas cette interlangue par définition" (1985: 19, 22). Etant finnophone nous-même, nous n'avons pas cette difficulté et nous pouvons peut-être mieux expliquer quelques formes interlinguales. Néanmoins, il va de soi que, quant au français, nous ne pouvons pas recourir à notre intuition de la même manière qu'un francophone natif. C'est pourquoi nous avons eu recours à un informateur autochtone chaque fois que cela s'est avéré nécessaire.

⁵³ Voir par exemple les études présentées dans *Langages* 84 (1986) et Perdue (1993).

4. PRODUCTION ECRITE

Dans les chapitres précédents, nous avons évoqué plusieurs facteurs qui influencent l'apprentissage d'une L2, ainsi que l'importance de prendre en considération la compétence communicative, incluant également le côté métacommunicatif, dans tout apprentissage et enseignement. Dans la présente étude, comme nous l'avons indiqué supra, nous nous intéressons à la langue écrite, et plus précisément au vocabulaire des apprenants. Avant de passer au lexique, nous allons discuter la spécificité du moyen de communication qu'est le langage écrit, la norme linguistique et le processus cognitif de l'écriture.

4.1. Langue parlée vs. langue écrite

Les connaissances langagières des apprenants se basent sur les quatre "skills", soit les quatre domaines d'aptitude, ou les quatre "zones" d'aptitude⁵⁴ (cf. Widdowson 1981: 69 ; Elo 1993 : 2 ; Boyer & Rivera 1979 : 18, respectivement). Sont distingués : les skills réceptifs, comprendre et lire et les skills productifs, parler et écrire. Ces skills fonctionnent souvent en même temps, aussi les limites entre eux sont-elles assez floues. Dans cette étude, comme nous nous intéressons aux skills productifs et en particulier à l'écriture, cela nous mène à la comparaison des caractéristiques propres à la langue parlée et écrite.

Il va de soi que la langue orale précède la langue écrite, car l'oral existe avant l'écrit dans la communication entre les êtres humains. Toutes les langues ont des formes orales, mais elles n'ont pas toutes de forme écrite ; on peut parler des sociétés de "l'oralité" ou des langues dites "sans écriture" (Blanche-Benveniste & Jeanjean 1987 : 22, 43), par exemple des langues aborigènes d'Afrique. A l'école et à l'université, la plupart des étudiants recourent à la lecture et à l'écriture pour effectuer leurs devoirs, mais après leur formation, ils utilisent plutôt la langue orale, notamment dans leurs contacts en langues étrangères.

L'écrit et l'oral ne sont pas tout à fait identiques ; on peut dire que l'écrit est plus stable, autrement dit, moins flexible et moins perméable aux changements linguistiques. L'oral, pour sa part, change plus ; quelques-uns de ces changements

⁵⁴ En anglais *language skills*, cf. J. B. Carroll (1968, 1975) et Stern (1987 : 347-359). On parle aussi des habiletés ou des capacités pour décrire ce phénomène.

sont temporaires, ce sont des formes ‘à la mode’ qui disparaissent assez vite. Quelques formes ont cependant tendance à résister et elles restent d’abord dans le langage oral avant de passer dans la langue écrite. Bref, l’oral joue un facteur évolutif dans le développement linguistique⁵⁵.

Il semble évident que la langue parlée et la langue écrite ont de nombreuses différences syntaxiques et grammaticales. On peut remarquer, entre autres, la différence suivante : la syntaxe de l’oral est moins structurée (si on la compare à la norme linguistique) ; elle comporte notamment moins de subordonnées et plus de phrases incomplètes, tandis qu’à l’écrit, il est fréquent de trouver beaucoup de déictiques (dans le sens du *Dictionnaire de linguistique* 1989 : 137), ou d’organisateur pragmatiques, p. ex. *premièrement, en somme* (voir chapitre 6.), de marqueurs métalinguistiques entre les énoncés et de phrases nominales lourdes de compléments. En outre, la structure interne du message est organisée différemment ; la langue parlée manifeste, entre autres, plus d’hésitations, de répétitions, d’ellipses, de pauses, et de faux départs, tandis que la langue écrite donne l’impression d’une précision plus poussée. De plus, la ponctuation et le découpage en syllabes doivent être pris en considération dans l’écriture (cf. Blanche-Benveniste & Jeanjean 1987 : 21-25, 138, 150-152 ; Brown & Yule 1989 : 15-17 ; Moirand 1993 : 8 ; Reichler-Béguelin et al. 1988 : 15-22). De plus, Reichler-Béguelin souligne la présence du contexte verbal (ou cotexte) comme caractéristique du texte écrit. Une phrase énoncée à l’oral peut avoir sa signification déictiquement par le destinataire, mais à l’écrit, en revanche, il existe une exigence de “co(n)textualisation de la référence en français écrit” qui découle “de la règle sociolinguistique plus générale de coopération” (1988 : 21-22). Turco & Coltier (1988 : 73) vont dans le même sens en annonçant la valeur d’engagement contractuel du texte écrit : le lecteur doit pouvoir se fier aux instructions de traitement fournies par le sujet écrivant.

L’autre différence principale entre les deux types de communication est le feed-back (soit l’échange) direct de l’interlocuteur qui n’est possible qu’à l’oral⁵⁶. Le feed-back pour la communication écrite est dans une certaine mesure retardé (cf. Luukka 1995 : 53). Pour compenser ce défaut, la communication écrite est

⁵⁵ Voir cependant Blanche-Benveniste & Jeanjean sur la discussion du rôle évolutif de l’oral (1987 : 29-37).

⁵⁶ On laisse ici de côté les nouvelles technologies, notamment le feed-back direct rendu possible par le réseau Internet (groupe de discussion, en anglais *irc and chat*), et aussi l’écriture spéciale utilisée dans les messages écrits par le mél, voir par exemple Minna-Riitta Luukka (1998).

beaucoup plus explicite. De plus, l'écrit ne permet pas non plus de faire référence aux sujets traités à l'aide de moyens extralinguistiques (par exemple les gestes et l'intonation), ce qui est apparent à l'oral. En revanche, le locuteur a le désavantage de ne pas avoir le temps de préparer son discours à l'avance ; il doit le planifier en parlant parce qu'il s'agit en fait de la communication immédiate et en situation (Linnaud 1986 : 12 ; Moirand 1993 : 8). Brown & Yule (1989 : 13) suggèrent que dans la vie quotidienne dans une culture alphabétisée, nous utilisons le moyen de communication oral principalement dans l'interaction avec les autres, tandis que le rôle de l'écrit est de transmettre l'information ; ils ajoutent qu'il est pourtant possible, dans certains cas, d'utiliser le langage oral pour transmettre l'information. Nous n'entrons pas ici dans une discussion plus détaillée sur cette affirmation, même si on peut la considérer comme un peu simpliste. A cause de la planification, le message écrit est, par nature, plus dense, plus révisé et plus permanent que l'oral, qui lui, par contre, contient plus de connecteurs et de remplisseurs de pauses, p. ex. *bon, euh* (cf. Benveniste 1987 : 53-57 ; Brown & Yule 1989 : 15-18 ; Riegel et al. 1997 : 30-31). Il faut noter cependant qu'il y a plusieurs degrés de planification dans le canal écrit aussi ; les lettres privées et les messages du courrier électronique, entre autres, sont moins planifiés (cf. Luukka 1995 : 55 ; Luukka 1998). Charolles (1988 : 12) distingue, dans un texte normal, quatre plans d'organisation textuelle (cf. ch. 6.3.) pour planifier le contenu du texte ; selon lui, aucun sujet écrivain, même bien entraîné, n'écrit d'un seul jet, sans ratures ni corrections. A l'oral, on peut aussi distinguer des variables telles que la vitesse de parole, la longueur des pauses silencieuses et les syllabes allongées qui n'existent pas dans la communication écrite (cf. Léon & Léon 1997).⁵⁷

Koch (1997 : 151) voit la distinction entre le langage oral et écrit un peu différemment. Il souligne le rôle important de paramètres communicatifs qui fonctionnent au niveau de la conception linguistique. Il y distingue quatre paramètres principaux qui ont des sous-catégories et qui peuvent varier par degrés/ par étapes :

1. proximité vs. distance physique, par exemple l'interaction face-à-face vs. à distance temporelle et spatiale

⁵⁷ On peut trouver, néanmoins, des différences aussi à l'intérieur d'un moyen de communication ; voir par exemple Dewaele (1997 : 233-240) sur la discussion de l'oral informel (production libre) et formel (oral "écrit"). Pour sa part, Kerbrat-Orecchioni envisage la situation communicative d'une façon interactionnelle. Elle parle de la pression du destinataire où le *degré d'interactivité* entre en jeu (1998 : 55).

2. proximité vs. distance sociale, par exemple les circonstances privées vs. publiques d'une communication, la familiarité des personnes
3. proximité vs. distance référentielle, par exemple les références 'ego-hic-nunc' existantes ou non-existantes
4. proximité vs. distance d'éloquence, par exemple dialogue vs. monologue, thème libre vs. fixé, spontanéité vs. réflexion

En somme, tout cela influence d'une façon pertinente la problématique et la recherche sur des corpus oraux et écrits.

4.2. Norme linguistique

L'enseignement d'une langue, soit maternelle soit étrangère, a besoin d'un fondement sur lequel baser le reste des connaissances linguistiques. Même si nous disposons d'un système langagier inné pour apprendre des langues, et notamment la langue maternelle, il nous faut aussi un descriptif de ce système qui est la norme linguistique. Dans l'enseignement d'une langue seconde, il est encore plus indispensable d'avoir un recueil de normes, voire de règles, pour transmettre le système aux apprenants. Pour dévier d'une manière acceptable de la norme, il faut quand même connaître cette norme, c'est-à-dire la grammaire (voir aussi chapitre 3.1.). La norme du français est établie depuis le XVIII^e siècle particulièrement sur l'écrit. Cela vient principalement du prestige fondé sur la littérature classique (Riegel et al. 1997 : 30).

Mais comment décrire cette norme ? Et quelle norme enseigner aux étudiants en cours de langue ? Chomsky (1970 : 26 ; d'après Riegel et al. 1997 : 17) déclare que

"La personne qui a acquis la connaissance d'une langue a intériorisé un système de règles qui relie les sons et les significations d'une manière particulière. Le linguiste qui construit la grammaire d'une langue ne fait que proposer un système sur ce langage intériorisé".

Chomsky parle là de l'acquisition d'une langue maternelle, mais il faut noter que la compétence d'une grammaire chez le locuteur natif et non-natif est fondamentalement différente (voir chapitre 3.1.). Cependant, les professeurs utilisent les mêmes manuels de grammaire dans l'enseignement, surtout au niveau universitaire, ou au moins, le cas échéant y recourent. Riegel et al. (1997 : 13-14) constatent brièvement qu'on peut distinguer quatre types de grammaire selon les objets spécifiques des observations et de procédures sur la langue : la grammaire synchronique (ou descriptive), la grammaire diachronique (ou historique), la

grammaire comparée et la grammaire générale. Ils ajoutent qu'une "grammaire descriptive est un **modèle théorique** qui se propose de décrire de façon explicite la grammaire-système, par définition implicite, d'une langue". Vogel (1995 : 86-87) s'oppose à la norme puriste et confinée à un registre, voire à la langue littéraire écrite, qui est souvent présentée dans les manuels scolaires. Il se réfère à Coseriu (1979)⁵⁸ qui déclare que la norme (*Norm*) se situe, en fait, entre les deux extrêmes, c'est-à-dire, entre la parole concrète (*Rede*) et l'abstraction (*System*) qu'est le système linguistique. Vogel en déduit que "l'application du concept linguistique de norme à l'interlangue est problématique en ce qu'elle implique une orientation exclusive sur la langue-cible" (op. cit. : 87). Dans ce cas-là, on n'accorderait pas assez d'importance aux autres facteurs étroitement liés aux productions linguistiques des apprenants (par exemple les facteurs pédagogiques, situationnels, etc. ; voir chapitre 2.2.). Il va même jusqu'à cette conclusion :

"On notera en guise de résumé et de bilan que les hypothèses et les constructions conceptuelles sur l'acquisition des langues et la production langagière qu'on a pu déduire de certaines orientations plus ou moins récentes de la linguistique descriptive sont, du seul point de vue épistémologique de l'antinomie entre empirisme et rationalisme, inadaptées à l'étude de l'interlangue du fait de la perspective indûment restreinte qu'elles impliquent. [...] pour la psychologie de langage en revanche, elles sont tout aussi inefficaces, pour ne pas dire néfastes, que les approches behavioristes" (Vogel 1995 : 91).

Nous pouvons trouver une idée parallèle chez Moirand (1993 : 146-147), qui propose de laisser les apprenants en langue étrangère structurer eux-mêmes leur grammaire⁵⁹, par exemple au travers de documents écrits authentiques. De cette façon, les apprenants pourraient intérioriser les modèles nécessaires à leurs productions. Il semble évident que cette approche rendrait les cours de langue plus motivants, même amusants, mais d'autre part, il ne faut pas oublier que les étudiants ont des besoins différents et le nombre d'étudiants désirant connaître les règles grammaticales exactes n'est pas négligeable. Selon Pendanx (1998 : 127-130), la grammaire a pris une place secondaire dans l'enseignement, puisque les enseignants sont plutôt préoccupés par les activités communicatives, croyant qu'elles seraient accompagnées en même temps d'une compétence linguistique. L'enseignement de la grammaire suscite des opinions contradictoires, d'autant plus que la terminologie concernant ce sujet est polysémique (cf. Besse & Porquier 1984 : 10-12). Pendanx propose quatre acceptions parmi lesquelles les

⁵⁸ E. Coseriu. 1979. "System, Norm und Rede". In Coseriu, E. *Sprache –Strukturen and Funktionen*. Zwölf Aufsätze zur allgemeinen und romanischen Sprachwissenschaft. Erste Auflage 1970. Tübingen : Gunter Narr Verlag, pp. 45-59.

⁵⁹ Elle nous renvoie à R. Porquier. 1975. "Analogie, généralisation et systèmes intermédiaires dans l'apprentissage d'une langue non maternelle". *BULAG* 3, 37-64.

enseignants pourraient choisir celle qui correspond à leur pratique ; elle fait la distinction entre la grammaire descriptive, la grammaire pédagogique (souvent éclectique), la grammaire intériorisée (cf. Corder 1975) - chez l'apprenant la grammaire d'apprentissage - et la grammaire d'enseignement (qui réfère à la pratique en classe).⁶⁰

Quoi qu'il en soit, nous devons, pour parler de l'enseignement, nous référer à des règles précises pour pouvoir effectuer notre analyse et il s'ensuit que ce qui nous intéresse partiellement ici, ce sont la grammaire descriptive et la grammaire comparée. Nous voulons souligner le mot *partiellement*, car nous ne parlons pas, par exemple, de la structure des textes étudiés, mais seulement du vocabulaire. Mais, est-ce qu'une grammaire du lexique existe ? Au sens étroit, probablement non, mais au sens large, oui ; il existe une norme lexicale même si celle-ci change au cours du temps. Pour comparer notre corpus avec cette norme, nous avons recours aux dictionnaires ainsi qu'aux grammaires monolingues et bilingues, car comme nous l'avons indiqué supra, nous examinons les mots tels quels, mais aussi dans leur contexte. En bref, citons Elo sur l'objectif de l'enseignement du français langue étrangère au niveau universitaire, qui

“a pour but de donner à l'étudiant la capacité d'utiliser le français dans différentes situations exigées par sa future profession, (enseignant où [sic] autre), sans faire de fautes de syntaxe, de morphologie ou de vocabulaire qui gênent la communication” (1993 : 18).

Nous recourons principalement aux grammaires *Le bon usage* de Grevisse (1993) et *La grammaire méthodique du français* de Riegel et al. (1997), et au dictionnaire *Le Petit Robert*. Dans notre analyse qualitative, cependant, nous nous servons d'un informateur natif pour les cas délicats.

4.3. Processus cognitif dans l'activité d'écriture

Nous pouvons également introduire ici l'idée de Pontecorvo (1997 : xviii) de rapprocher l'écriture et la lecture : “dans la pratique quotidienne d'une personne alphabétisée, il est presque impossible de séparer les skills de l'écriture et de ceux de la lecture. De la même manière, s'il est possible, en théorie, de lire sans savoir écrire, il est presque impossible d'écrire sans savoir lire”. Ces deux skills sont exprimés à l'aide d'un support visuel, même si l'un est un skill réceptif, et l'autre

⁶⁰ Besse & Porquier (1984 : 195-199) décrivent la relation étroite et complexe entre grammaires descriptives, grammaires pédagogiques et grammaires d'apprentissage qui “est à envisager comme un réseau d'interrelations [...]”.

productif (cf. Widdowson 1981 : 69). Nous avons mentionné au début de ce chapitre que nous nous intéressons surtout au skill productif qu'est l'écriture. Néanmoins, nous ne pouvons pas nous éluder totalement l'étude du côté réceptif. D'autant plus que plusieurs recherches prouvent que l'acquisition lexicale en langue seconde est étroitement liée à la lecture (voir par exemple Coady 1997). Nous reviendrons sur ce sujet dans le chapitre 5.2.

Quant au processus d'écriture, nous pouvons noter qu'il peut être considéré d'après cinq contraintes primaires qui renforcent et influencent différents aspects du processus même, ou bien individuellement ou bien ensemble (Frederiksen & Dominic 1981 : 2-3). Ces contraintes peuvent se superposer, mais toutes donnent quelques informations sur l'activité d'écriture. C'est une activité cognitive qui inclut, entre autres, les connaissances, les stratégies et l'aptitude de réflexion de celui qui écrit ; elle est influencée par ce que le sujet écrivant sait des formes et des principes d'une langue. Troisièmement, quand on écrit, on est obligé de penser à son interlocuteur, en d'autres termes, de penser à la fonction communicative du texte. Finalement, l'écriture est une activité qui a un but spécifique déterminé par le contexte, à savoir, la situation et les consignes. Il faut encore ajouter une cinquième contrainte, celle de la cohérence textuelle aux différents niveaux de langue (cela veut dire qu'il est important de prendre en compte le contenu et la forme du texte antérieur) (op. cit. pp. 18-20). En écrivant, il nous faut donc prendre en considération, d'une façon consciente ou inconsciente, les relations entre ces différentes contraintes. On peut constater explicitement ces relations dans la Figure 4 à la page suivante (d'après Frederiksen & Dominic 1981 : 19).

Pour expliquer cette figure, ajoutons que l'activité d'écriture, voire la composition, se divise en trois processus différents⁶¹ : a) la planification, où celui qui produit le texte planifie, organise et découvre l'information ; b) l'interprétation, où il s'agit de transformer les idées en mots ; et c) l'évaluation, autrement dit, la modification du texte. Enfin, la vérification répétée du texte produit, c'est-à-dire la révision du texte (cf. Flower & Hayes 1981 : 41 ; Nold 1981 : 68-77).

⁶¹ Charolles (1986 : 10-13) parle de la planification, de la mise en texte et de la révision. La planification comprend les activités de mobilisation, activation, sélection, recherche et composition de connaissances, tandis que les processus de mise en texte portent sur les opérations plus locales, c'est-à-dire le contrôle de la verbalisation. La révision, pour sa part, comprend les retouches et les corrections qui affectent la forme superficielle du texte.

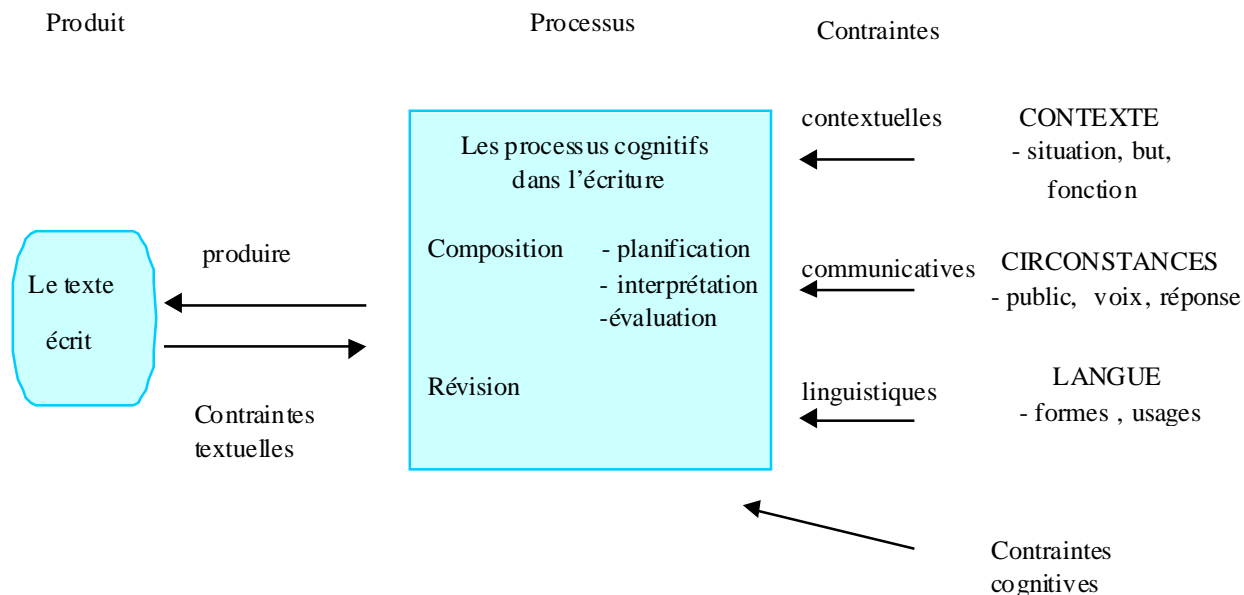


Figure 4. *Processus cognitifs dans l'activité d'écriture*

Puisque le producteur d'un texte a plus de temps pour composer et réfléchir sur son message que le locuteur, il peut aussi le réviser en fonction des contraintes mentionnées ; le but du devoir, son contexte (voire la situation), les consignes, le public imaginaire, et de plus, le temps disponible pour la composition (cf. Frederiksen & Dominic 1981 : 1).

De son côté, Moirand distingue aussi la situation de production de celle de réception qui cependant forment ensemble la situation d'écrit, voire de communication écrite. Nous nous contentons de mentionner brièvement ces composantes de base. Dans la situation de communication, le scripteur et le lecteur eux-mêmes influencent la situation, entre autres, par leur statut social, leur rôle et leurs attitudes. Les relations entre le scripteur et le lecteur jouent un rôle également ; le scripteur se fait une image du lecteur potentiel et vice versa. Le scripteur a aussi une intention dans sa communication ; il veut créer un effet sur le lecteur, mais celui-ci ne réagit pas toujours de la façon souhaitée. Il y a aussi des facteurs extra-linguistiques qui entrent en jeu, par exemple le lieu, le moment et les connaissances antérieures (1993 : 10-12). Ces composantes correspondent en partie, bien sûr, aux contraintes présentes dans le processus cognitif de l'activité d'écriture (voir supra). En outre, Moirand (1993 : 14-15) propose, "bien que ce soit particulièrement délicat", un classement de différents types de modalités qui, à son avis, est opératoire en didactique de l'écrit. Ces modalités expriment les rapports du scripteur avec son énoncé; on peut trouver des modalités

appréciatives, par exemple un jugement du scripteur par rapport à ce qu'il écrit, et des modalités logiques, soit pragmatiques, c'est-à-dire que le scripteur situe son énoncé en termes de probabilité ou d'éventualité, par exemple.⁶² De même, Péry-Woodley postule qu'au niveau discursif ou énonciatif, l'écrit a un double rôle en situation d'apprentissage :

“il est toujours *argumentatif* puisque le but est, quel que soit le sujet traité, d'exhiber connaissances et compétences de manière à convaincre le lecteur-évaluateur ; en même temps, un aspect important de ces connaissances/compétences concerne précisément le respect des conventions liées au type de texte exigé par le sujet (implicitement ou explicitement) : “raconter”, “discuter”, “décrire”, etc.” (1993:100).

Dans notre corpus, il s'agit d'écrire une dissertation le mieux possible sur un thème optionnel. La situation est un examen sur les connaissances langagières des apprenants, et le temps disponible pour tout l'examen est de trois heures, mais les participants doivent effectuer cinq tests différents durant ce temps (voir chapitre 1.2. ; la production orale est un test à part). L'idéal est, bien évidemment, d'écrire dans un français compréhensible qui se rapproche de celui d'un locuteur natif ; les correcteurs sont des professeurs français natifs. Le style requis peut être catégorisé entre formel et informel (cf. Koch 1997 : 150) car les thèmes varient d'un sujet purement objectif à un sujet personnel (voir Annexe 1). En outre, du point de vue de l'apprenant, l'intention est également de convaincre le lecteur-évaluateur de ses capacités en production écrite.

Ayant vu ce que cela signifie d'avoir le langage écrit comme moyen de communication, nous abordons à présent le sujet central de cette étude, le vocabulaire.

⁶² Voir aussi Boyer & Rivera sur la problématique de l'écrit (1979 : 110-144).

5. MOT, VOCABULAIRE ET LEXIQUE EN TANT QU'OBJET D'ETUDE

Dans le chapitre 1.3. sur les remarques terminologiques, nous avons expliqué brièvement que les termes **lexique** et **vocabulaire** sont définis de façons différentes selon le domaine où on les utilise ; quelquefois ils peuvent désigner la même chose, d'autres fois, ils sont utilisés bien distinctement l'un de l'autre. Nous avons choisi d'utiliser ces concepts principalement d'une manière interchangeable sauf dans l'analyse du corpus où, quand c'est nécessaire, le terme **vocabulaire** réfère à l'utilisation des mots de chaque individu. Autrement dit, chacun a son propre vocabulaire, tandis que le terme **lexique** est réservé au sens de stock global des mots d'une langue en question. Dans la partie statistique, nous parlons en termes de **vocabulaire(s)**, **mot** et **vocabulaire** (cf. infra).

Dans l'introduction, nous avons également évoqué le rôle pertinent du lexique aussi bien dans la production que dans la compréhension d'une langue. Dans ce nouveau chapitre, nous allons aborder cette question plus en détail, d'abord du point de vue de la linguistique appliquée et ensuite du point de vue de l'apprenant. Enfin, nous traiterons des définitions du mot en général et puis dans l'étude présente.

5.1. Relation entre les recherches sur le lexique et la syntaxe

La sémantique et le lexique ont souvent été considérés comme des domaines de moindre intérêt dans la recherche linguistique. Les systèmes clos, comme la phonologie, la morphologie et la syntaxe, ont davantage intéressé les linguistes car on pouvait supposer qu'ils étaient plus faciles à traiter (Takala 1989 : 1-2). Auparavant dans l'enseignement, en particulier à l'époque de la méthode audio-orale, les *pattern drills* (ou cadres syntaxiques ou structuraux) et l'apprentissage d'automatismes linguistiques étaient mis en valeur, tandis que le vocabulaire occupait une place secondaire par rapport aux structures syntaxiques (Germain 1993 : 141-151). Or, la situation n'a pas toujours été la même. D'après Kennedy (1992 : 336-341), entre les années 1920 et mi-1950, il y avait une relation étroite

entre les études de cas sur le vocabulaire⁶³ et la pédagogie d'une langue seconde. Il nous renvoie par exemple à l'oeuvre de Thorndike (1921), qui est une base pour la recherche méthodologique sur le vocabulaire (Takala 1989 : 1-2). A partir de la fin des années 1940, le point d'intérêt dans la théorie de l'enseignement des langues s'est principalement fixé sur les structures grammaticales (cf. la méthode audio-orale). Bogaards (1994 : 7-9) fait remarquer que beaucoup d'auteurs célèbres (p. ex. Corder) se sont totalement concentrés sur l'étude de l'acquisition grammaticale. Il faut pourtant noter qu'en France, dans les années cinquante, un groupe de linguistes a entrepris un projet pour préparer une méthode avec laquelle les bases lexicales et grammaticales du français (langue étrangère) seraient apprises d'une manière efficace ; il s'agit du *Français fondamental* dont l'idée se rapproche du *Basic English*, une méthode d'enseignement de l'anglais éditée par l'armée britannique aux Indes (Blanche-Benveniste & Jeanjean 1987 : 78-81 ; Lavonius 1998 : 25-30). Cette méthode a reçu aussi bien des louanges que des critiques⁶⁴, mais c'est surtout la partie concernant le vocabulaire qui est considérée comme étant importante. Ainsi arrive-t-il que, par exemple en Finlande, d'après *Les bases du programme scolaire pour le lycée (Lukion opetussuunnitelman perusteet)*, le vocabulaire à apprendre au lycée se base sur le *Français fondamental* ; cela se voit encore explicitement dans la version de 1985, et implicitement dans la dernière version datant de 1994 (Lavonius op. cit. : 5-9, 70). Ce n'est que vers la fin des années 1970 que les spécialistes de linguistique appliquée se penchent de nouveau sur le vocabulaire et ce n'est que très récemment, à savoir à partir du milieu des années 80, que le vocabulaire est redevenu pertinent dans la pédagogie des langues.

Takala postule qu'il est devenu évident, à travers maintes études, qu'une personne peut comprendre une langue étrangère, même si elle ne peut pas produire des énoncés dans cette langue. Elle peut également être comprise même si elle ne connaît pas toutes les règles morphologiques et syntaxiques de la langue étrangère. C'est possible grâce aux connaissances lexicales de cette personne ; les mots essentiels peuvent être compris à l'aide du contexte et de la situation (1989 : 4). Ainsi les connaissances lexicales sont-elles devenues une partie centrale dans l'enseignement et l'apprentissage d'une langue (cf. Bogaards 1994 : 6) (cf. aussi

⁶³ Un domaine de la linguistique appliquée est celui des études sur les corpus, donc des études de cas. Kennedy (1992 : 336-341) présente brièvement l'histoire de ces études dans le domaine lexical.

⁶⁴ Entre autres, Galisson écrit : "Dans le Français Fondamental, la sélection a tué la motivation en asséchant la matière à enseigner" (1979 : 19).

ch. 3.2.). Un témoignage de l'importance du vocabulaire est aussi le fait que les erreurs concernant la sélection lexicale sont considérées plus gênantes quant à la compréhension que les erreurs grammaticales (Carter 1987 : 145 ; Harley & King 1989 : 415).⁶⁵ Néanmoins, il y a des liens étroits entre le vocabulaire et les structures syntaxiques ; les mots et les constructions ont des significations qui forment ensemble la syntaxe (Clark 1998 : 49). Tréville et Duquette (1996 : 99-100) décrivent cette relation comme un va-et-vient indispensable à tout moment

“parce que la régularité d'une règle de grammaire ne peut se vérifier qu'en faisant varier un grand nombre d'entrées lexicales dans la structure syntaxique considérée et parce que, inversement, les régularités dans le vocabulaire ne peuvent se vérifier qu'en faisant varier les structures syntaxiques susceptibles de l'accueillir.”

Bogaards va dans le même sens en affirmant que récemment les grammairiens et lexicologues ont découvert que les frontières entre les deux domaines sont moins grandes qu'on pouvait l'imaginer ; en d'autres termes, en utilisant un mot quelconque on ne peut pas ignorer les règles d'utilisation qui y sont liés (1994 : 38). Courty (1989 : 146) va jusqu'à dire : “Il est difficile de séparer l'apprentissage lexical de l'apprentissage de la syntaxe, car c'est à partir du lexique que s'organise la syntaxe.” Il faut noter cependant que du point de vue de la communication, les connaissances lexico-sémantiques sont primordiales par rapports aux connaissances grammaticales. Quand la syntaxe manque, le locuteur est obligé de recourir à son intuition pour former une interprétation acceptable, qui, bien sûr, peut être fautive (Bogaards op. cit. : 39-40). Adjemian (1983 : 265) voit également l'importance des connaissances grammaticales du lexique,⁶⁶ mais du point de vue du transfert des caractéristiques lexicales. Il dit :

“Lexical items are complexes of sets of features reflecting diverse phonological, morphological, syntactic, and semantic properties. Furthermore, once the lexicon is viewed as an active component of the grammar, lexical rules reflecting the speaker's knowledge of “lexical relatedness” become of great interest.”

Giacobbe (1992 : 61-64) décrit le même phénomène, mais il parle de règles idiosyncrasiques en déclarant que le lexique, à l'égal de la grammaire, est une construction instable de l'apprenant ; l'apprenant teste donc ses hypothèses sur le système de règles lexicales afin de mémoriser des lexèmes pour pouvoir produire des énoncés dans le discours. Pour sa part, Vogel étudie la question de la

⁶⁵ Il faut remarquer que les locuteurs non-natifs semblent être plus stricts envers les erreurs lexicales que les natifs (cf. Amaud 1984 : 18).

⁶⁶ Vives (1985 : 48-65) présente dans son article le terme de **lexique-grammaire**. Il base son analyse sur les constructions nominales dont la description est présentée sous la forme de matrices binaires. Il veut montrer l'utilité de cette entreprise linguistique pour l'enseignement/apprentissage des langues. Il s'agit de la description profitable des contraintes lexicales, des locutions propres à une langue, etc. (cf. Tréville & Duquette 1996 : 19-21).

mémorisation du savoir linguistique qui se passe à plusieurs niveaux. Entre autres, il pense que d'après la description psycholinguistique les connaissances syntaxiques sont représentées dans la mémoire sous la forme de "structures propositionnelles", tandis que les traits sémantiques, la représentation des connaissances lexicales et sémantiques, correspondent à des "capacités cognitives générales en rapport avec les données psycho-physiques de l'être humain" (1995 : 284-286).

5.2. Rôle actuel du vocabulaire dans l'apprentissage et l'enseignement d'une langue

Les enfants bâtissent les fondations de leur langue maternelle au cours des quatre premières années. L'acquisition du lexique est relativement rapide après l'âge de deux ans où ils disposent de 200 à 600 mots. A six ans, ils en connaissent déjà environ 14.000, ce qui correspond à l'apprentissage de presque un mot par heure (Clark 1998 : 49). Les chiffres pour estimer l'étendue du vocabulaire d'un natif adulte varient d'après les façons de calculer⁶⁷. Pour leur part, Tréville & Duquette estiment qu'un locuteur natif disposerait d'un vocabulaire productif, voire actif, de 2000 mots tandis que le vocabulaire réceptif, soit passif, serait de cinq ou six fois supérieur (1996 : 69). Elles ajoutent que l'apprentissage du vocabulaire, même de la L1, n'est jamais achevé (op. cit. : 120).

Quand on étudie la situation des apprenants d'une L2, on peut remarquer que la différence entre le vocabulaire actif et le vocabulaire passif est moindre, au moins parmi les débutants. Takala (1984 : 202-206) a trouvé qu'à l'école secondaire la différence entre l'étendue du vocabulaire productif et réceptif des élèves apprenant l'anglais n'était pas significative ; en d'autres termes, l'étendue des connaissances lexicales productives et réceptives était presque la même. L'étendue des connaissances lexicales des apprenants finnophones du français est encore mal connue. Lavonius présume pourtant dans sa recherche que, quant au français, la situation est identique à celle de l'anglais et qu'à partir du lycée le vocabulaire passif s'accroît du fait des exigences du contenu du programme scolaire dans lequel la quantité des textes longs étudiés d'une manière extensive⁶⁸

⁶⁷ Voir par exemple Laufer (1997 : 20-34), Nation (1990 : 11-12) et Takala (1984 : 48-55).

⁶⁸ Lire des textes d'une manière extensive veut dire, entre autres, que le texte n'est pas traduit, mais que la compréhension du contenu se fait à l'aide des questions à choix multiples, des résumés, etc.

augmente (1998 : 33). Comme nous l'avons dit supra, le vocabulaire à apprendre au lycée finlandais se base sur le *Français fondamental* qui contient (les deux degrés) au total environ 3500 mots qui sont en partie un peu vieillis aujourd'hui ; il s'ensuit que dans l'épreuve du français de l'examen du baccalauréat, et notamment dans la partie de compréhension écrite (c'est-à-dire celle qui teste les connaissances du vocabulaire passif), à peu près 25 % des mots des textes n'appartiennent pas au *Français fondamental* (Lavonius 1998 : 26-29, 67-68). Cela ne nous donne pas de chiffres exacts quant aux connaissances lexicales des lycéens en français, mais nous aide à cerner la situation de manière approximative.⁶⁹

Blum-Kulka & Levenston (1992 : 119-121) discutent les universaux concernant la simplification lexicale et arrivent à une conclusion selon laquelle à tous les niveaux d'apprentissage les connaissances lexicales des apprenants d'une L2 sont moindres en comparaison avec celles d'un autochtone. Ceci est lié à certains aspects de la compétence sémantique (par exemple la prise de conscience d'hyponymie et d'antonymie et d'autres relations systémiques entre les unités lexicales) qui, selon eux, représentent la base d'utilisation des stratégies communicatives afin d'exprimer "des significations complexes par des moyens indirects avec un vocabulaire minimal" (cf. Nation 1990 : 159-174). Le manque de connaissances sémantiques lexicales se reflète ainsi dans la simplification, consciente ou inconsciente, du vocabulaire. Pour compenser ce déficit, les apprenants ont recours à leurs connaissances sémantiques en langue maternelle. Puisque les langues réalisent ces relations au moins en partie différemment, il s'ensuit que la production de l'apprenant n'est pas toujours appropriée par rapport à celle d'un natif (Blum-Kulka & Levenston op. cit. : 137 ; cf. aussi Bogaards 1994 : 148-151 ; Harley & King 1989 : 431, 434).

Nous voulons aussi aborder la discussion sur la difficulté du vocabulaire à apprendre : en d'autres termes, quels sont les mots les plus difficiles à apprendre ? Les langues qui appartiennent à la même famille linguistique sont souvent très proches syntaxiquement, ou ont une base lexicale commune, ou les deux.

⁶⁹ Jaatinen & Mankkinen (1993) ont effectué une recherche sur l'étendue du vocabulaire passif des apprenants universitaires d'anglais et de son développement. Le test à choix multiples se basait sur des entrées de dictionnaire, donc des mots sans contexte ; elles estimaient l'étendue du vocabulaire à 17.100 mots chez les étudiants de première année, et à 19.500 chez les étudiants avancés. Takala (1984 : 100-117) discute des méthodes de mesure de l'étendue du vocabulaire appris et enseigné d'une L2 ; il s'agit principalement de l'anglais.

Normalement, ceci facilite l'apprentissage entre ces langues, surtout la compréhension d'un texte en langue étrangère (cf. Tréville & Duquette 1996 : 58). Quant au vocabulaire, il s'agit de **cognates**, de mots qui sont rapprochés ou transparents aussi bien graphémiquement que sémantiquement (cf. Elo 1993 : 48 ; Takala 1989 : 59), par exemple *touriste* en français et *tourist* en anglais.⁷⁰ Le finnois n'étant pas une langue indoeuropéenne, le nombre des cognates entre cette langue et le français est minimal, par exemple le mot *turisti* justement qui est un emprunt ; il faut pourtant noter que la connaissance d'autres langues indoeuropéennes apprises à l'école aide les apprenants finlandais. Mais la similarité formelle des langues peut aussi être trompeuse si la signification n'est pas la même. Dans ce cas-là il s'agit de **faux amis**, qui peuvent rendre l'apprentissage plus difficile (Takala op. cit. : 5), par exemple *journée* en français et *journey* en anglais (nos étudiants du niveau élémentaire ont fait ce rapprochement plusieurs fois). Ces difficultés peuvent souvent causer des erreurs dans l'utilisation du vocabulaire ce qui gêne la compréhension. Laufer (1997 : 20-34) discute cette question du point de vue du vocabulaire passif, autrement dit, lié à la compréhension écrite. Laufer observe qu'il existe trois sortes de mots qui peuvent créer une sorte d'anxiété (*lexical plight*) chez l'apprenant et ainsi empêcher l'apprentissage ; il y a des mots qu'on ne connaît pas (donc complètement opaques), des mots qu'on croit connaître (transparence déceptive) et des mots qu'on ne peut pas deviner (il n'y a pas d'indices contextuels suffisants).

En outre, on pourrait penser que les mots qui sont les plus fréquents seraient les plus faciles à apprendre et que les mots rares, porteurs de beaucoup d'informations, seraient les plus difficiles. Ce n'est que partiellement vrai, car les mots dont la fréquence est élevée sont souvent des mots à plusieurs significations, par exemple les verbes *faire* et *avoir* (cf. Harley & King 1989). Arnaud & Savignon (1997 : 158) abordent la question de la fréquence des mots du point de vue du vocabulaire passif et d'après eux, la non-connaissance des mots rares suscite un obstacle dans la lecture car ces mots transmettent une information

⁷⁰ Arnaud (1993 : 14) conclut, d'après une expérience, que quand les apprenants accomplissent une tâche métalinguistique sur le vocabulaire d'une L2, ils semblent avoir recours à leur langue maternelle, ce qui, selon lui, témoigne de "l'omniprésence du vocabulaire de la L1 derrière celui de la L2 dans les situations d'apprentissage scolaire".

dense.⁷¹ Comme on l'a dit supra, les erreurs lexicales sont significatives car elles empêchent plus facilement la compréhension, soit de l'apprenant, soit du locuteur, que les erreurs grammaticales. Faerch et al. (1984 : 174-175, 292-294) ont tiré la même conclusion de leur "test de tolérance", qui est une variante de l'analyse de performance des apprenants, portant sur le langage écrit et ayant comme but de provoquer des réactions chez les lecteurs natifs afin de découvrir si les lecteurs ont compris le message comme prévu.

Il semble évident que quand on écrit dans une langue étrangère, il ne suffit pas d'utiliser correctement, c'est-à-dire sans erreurs, la grammaire et le vocabulaire pour que le texte ressemble à la production d'un scripteur natif. Linnarud dit :

"A poor knowledge of the vocabulary of a foreign language leads not only to the wrong choice of a word but also to a less imaginative, dull and uninteresting composition with repetition of highly frequent lexical items and a simple and unelaborated theme" (1986 : 3).

Ceci implique que l'on peut utiliser le vocabulaire d'une manière certes correcte mais inadéquate en comparaison avec les normes d'un locuteur natif (cf. ch. 3.1., 3.2.). Linnarud (1986 : 6) ajoute qu'il ne suffit pas de connaître les mots isolés, mais qu'il faut également les traiter en contexte puisqu'il existe des collocations et associations qui forment, pour ainsi dire, une unité de sens (cf. Tréville et Duquette 1996 : 34-36).⁷² Le but de l'apprentissage des langues n'est pas toujours nécessairement d'arriver à une production comparable à celle d'un natif ; c'est un but rarement approprié, et même impossible à atteindre par la plupart des apprenants. James (1981 : 164) a postulé qu'à l'école secondaire en Angleterre à cette époque, autour de 90 % d'élèves en cours de langues étaient des *under-achievers*, c'est-à-dire, des élèves qui obtenaient des résultats qui ne répondaient pas aux attentes. Normalement à l'école, il s'agit plutôt de garantir une communication adéquate, autrement dit, de garantir que le message soit compris.

⁷¹ Il faut se souvenir que la lecture et l'écriture sont étroitement liées l'une à l'autre (cf. ch. 4.3. ; Reichler-Béguelin et al. 1988 : 11). Arnaud, dans un article antérieur (1993 : 7), explicite également le lien entre la fréquence des mots d'un texte et sa lisibilité ; la fréquence semble être attachée aux représentations mentales des mots. Voir aussi la discussion de Nation sur les difficultés d'apprendre un mot ; il parle, entre autres, des effets comiques de l'utilisation excessive des mots rares dans la production (1990 : 33-49).

⁷² Les concordanceurs, c'est-à-dire les logiciels permettant de chercher une chaîne de caractères dans un texte et d'imprimer toutes les occurrences avec les contextes dans lesquels ils apparaissent, peuvent être utiles dans des applications pédagogiques ; normalement, ils sont conçus pour une analyse linguistique ou littéraire (Péry-Woodley 1993 : 185-186).

Mais que veut dire apprendre ou enseigner un mot ? Ou que veut dire connaître un mot ? Selon Tréville & Duquette (1996 : 135), à l'instar de Dale (1965), la connaissance d'un mot se passe par étapes : la reconnaissance, la compréhension et le ré-emploi des unités. Il faut noter que l'apprentissage du vocabulaire se passe d'une manière différente s'il s'agit du vocabulaire productif, voire actif, ou du vocabulaire réceptif, ou passif. Le premier requiert des techniques de mémorisation et donc l'apprentissage direct de vocabulaire, tandis que le second nécessite l'apprentissage indirect par le contexte (cf. Coady & Huckin 1997 ; Nation 1990 : 31-32 ; Tréville & Duquette 1996 : 65-69). En outre, selon Nation (op. cit. : 147-148), la capacité d'utiliser les mots dans l'écriture implique l'apprentissage de l'orthographe, de l'utilisation des mots dans les phrases, de l'organisation et de la structuration du texte écrit. Des trois objectifs concernant l'élargissement des connaissances lexicales en L1 présentés par Graves (1987), Bogaards concidère qu'ils peuvent être transposés aussi dans l'apprentissage des L2 (1994 : 162-163)⁷³ ; ces objectifs sont :

- a. apprendre les mots – par exemple l'apprentissage de nouvelles significations pour des concepts déjà connus
- b. apprendre à apprendre les mots – par exemple l'utilisation de diverses sources d'informations : contexte, parties analysables de mots
- c. apprendre des faits sur les mots – par exemple se rendre compte du fait que les sens varient et comment ils varient

Il en résulte que l'apprentissage du vocabulaire ne devrait jamais se réduire à une simple mémorisation de mots nouveaux (cf. Nation & Newton 1997). Par ailleurs, il est aussi primordial de souligner l'importance des métaconnaissances linguistiques des apprenants comme nous l'avons déjà indiqué supra (cf. ch. 3.3. ; Tréville & Duquette 1996 : 101-102). Quant à l'enseignement du vocabulaire, Galisson (1981 : 41) postule que : "C'est gageure de vouloir enseigner les vocabulaires, il faut enseigner à les apprendre" (cité dans Bogaards 1994 : 162). Pour sa part, Sautermeister (1989 : 123-124, 132) juge que la signification de la métaconnaissance lexicale est importante surtout pour les étudiants au niveau

⁷³ Discutant la nature des connaissances lexicales d'une L2, Bogaards (1980) en distingue quatre types : la connaissance très vague d'un mot et les connaissances des aspects formels, sémantiques et syntaxiques d'un mot ; par exemple connaître les aspects syntaxiques et sémantiques du verbe *tenir* signifie que la personne connaît son utilisation dans des phrases comme *il tient la caisse – il la tient, il tient à sa réputation – il y tient, il tient de son père – il tient de lui, il tient bon – (-)* (cité dans Takala 1984 : 56-57). Ringbom, de son côté, parle des connaissances lexicales qui varient selon plusieurs dimensions en continuum allant du non-savoir au savoir complet, certes théorique, du lexique ; selon lui, ces dimensions sont au moins les suivantes : l'accessibilité lexicale (c'est-à-dire le mot est accessible, au niveau le plus bas, seulement dans un contexte spécifique, et il est accessible même sans contexte au niveau le plus haut), la morpho-phonologie lexicale, la syntaxe, la sémantique, les collocations et les associations lexicales (1987 : 36-37).

universitaire afin que ces derniers s'orientent vers "un auto-apprentissage plus réfléchi". En outre, Bogaards signale quatre thèses mentionnées dans maintes recherches actuelles sur l'enseignement du vocabulaire :

1. Les mots doivent être présentés en contexte.
2. Les mots doivent être appris avec la totalité de leur contenu sémantique.
3. La présentation du vocabulaire est exclusivement affaire de sémantique.
4. Le vocabulaire doit être proposé en ensembles organisés.

(1994 : 180-186)

Combien ces thèses sont pratiquées ou même praticables dans les cours de langues est une autre question.

Cela nous mène à la question : comment choisir les mots à apprendre ? Dugast (1980 : 29) présente une méthode longtemps utilisée par les enseignants ; il s'agit de listes de **fréquence** qui ont plus tard été complétées par les critères de **répartition** et de **disponibilité** (par exemple des mots concrets comme *dent* qui n'est pas fréquent mais familier –les mots du *Français fondamental* étaient également choisis selon ces critères (cf. Bogaards 1994 : 124-128 ; Lavonius 1998 : 20-21). Pour sa part, Bogaards nomme l'**utilité** le critère le plus pertinent. Il le définit comme suit : "[...] est utile ce qui sert à atteindre un but qu'on s'est fixé" [en gras dans le texte original] (op. cit. : 111-112). De plus, il présente les critères de **familiarité**, d'**items concrets** et de **contrastivité**. Il semble que les mots familiers et concrets soient plus faciles à mémoriser, mais qu'avec les mots en contraste (p. ex. *rapide – lent*), il faille être plus prudent car il y a un risque de confusion (Bogaards 1994 : 151-153). Picoche (1989 : 91) conclut que, d'après les recherches, la fréquence est un bon indicateur des mots à apprendre – c'est-à-dire que les mots de haute fréquence (> 7.000) représentent plus de 90 % de l'ensemble des occurrences de n'importe quel texte écrit en français. Elle ajoute que celle-ci doit être complétée par une étude sur la polysémie des mots et sur les structures actanciennes aussi bien quant aux mots fondamentaux qu'aux mots sémantiquement apparentés et moins fréquents. Nation (1990 : 15-21) trouve aussi que le critère de fréquence est très utile car il représente, d'après ses recherches sur l'anglais, 87 % d'un texte ; il propose ainsi de choisir les mots à enseigner principalement d'après leur fréquence et leur ordre, mais il ajoute que ces critères doivent être complétés par d'autres, à savoir ceux mentionnés plus haut. En général, on peut dire qu'on apprend oralement les mots les plus fréquents, mais par la lecture, on apprend également des mots difficiles et avec une fréquence moins

élevée (Paribakht & Wesche 1997 : 174-175). Ainsi, les mots rares sont également à enseigner, mais de façon différente (cf. supra la remarque de Nation sur les mots rares).

5.3. Définitions du concept ‘mot’

A cause de nos méthodes utilisées dans la présente étude (cf. chapitre 7), nous devons déterminer l’unité de base utilisée pour effectuer nos calculs, à savoir le mot. Le concept **mot** peut sembler à première vue facile à définir, mais en fait, il n’en est rien. Ménard affirme qu’“aucun linguiste, à notre connaissance, n’y est parvenu” (1983 : 20). Est-ce qu’un mot consiste en un seul mot ou en plusieurs items qui ensemble forment la signification de ce mot ? Qu’est-ce que cela veut dire, un mot ? Comment appeler ce phénomène, cette unité pratique pour de multiples analyses ? Au sens courant, le mot signifie un son ou un groupe de sons correspondant à un sens (*Le Petit Robert* 1989). Mais cela ne satisfait pas les linguistes car il est facile de remarquer qu’un mot n’est pas une unité minimale de la langue (cf. Elo 1993 : 37 ; Lavonius 1998 : 11).

Pergnier (1986 : 13-15) postule que cette unité significative minimale est un point commun chez les linguistes contemporains, même si elle est nommée de diverses façons, par exemple **monème**⁷⁴, **sémantème** et **morphème** (*Ddl* 1989), et que ce point commun concerne son caractère unissant les deux constituants inséparables de la langue, notamment son et signification (ou signifiant et signifié au sens saussurien). Selon lui, il n’existe cependant pas d’unanimité quant aux autres unités de langue, et il en est ainsi également quant au concept de mot. Il déclare que

“le seul point commun aux différentes théories et pratiques descriptives évoquées ici est négatif : c’est l’absence de référence implicite ou explicite à la notion de *mot*” (op. cit. : 15).

Greimas (1989 : 58) affirme que “le mot reste le noyau mythique de la linguistique”. Un des problèmes concernant la définition du mot est la diversité d’approches. Nous pouvons trouver, entre autres, des approches linguistiques,

⁷⁴ Martinet distingue les monèmes grammaticaux (= morphèmes) et les lexèmes (cité dans *Ddl* 1989). De plus, Martinet fait la distinction entre un monème (= la plus petite unité significative, par exemple dans *chant/er/ons*, il y a trois monèmes = 1 mot), et un syntème (= 1 forme + 1 forme = 1 fonction), et un syntagme (= 1 forme + 1 forme = 2 fonctions) (cf. 1985 : 33-42, 83). Pour donner un exemple, un monème *chemin*, un syntème *chemin de fer*, un syntagme *par ce chemin-là*.

pédagogiques et statistiques qui retiennent différentes définitions pour l'unité traitée ; ces définitions peuvent également se chevaucher.⁷⁵

Picoche⁷⁶ (1992 : 13-25) présente quelques critères de délimitation pour le signe lexical du point de vue phonique et syntactico-sémantique afin d'essayer d'éclairer la situation terminologique dans une perspective pédagogique. Nous nous passons du premier puisque nous examinons seulement le langage écrit. Le deuxième est divisé en deux parties principales : les unités graphiquement simples mais morphologiquement complexes et les unités graphiquement complexes. Les unités graphiquement simples mais morphologiquement complexes, par exemple *anti/constitution/nelle/ment*, posent des problèmes quant à leur décomposition en **lexèmes** (= un mot dépouillé de ses affixes) et en morphèmes (entre autres chez Martinet), car jusqu'où peut-on aller dans la décomposition pour trouver le véritable radical ? Picoche présente encore les **mots simples** (p. ex. *couteau*), les **mots construits**, ou **dérivés** (l'exemple ci-dessus) et les **mots fléchis** (p.ex. *donnerons*) (1992 : 17-18, 21). De plus, elle considère le verbe, global et abstrait, comme un **mot**, et parle de formes de ce mot en référant à sa diversité morphologique (1992 : 23). Pour leur part, selon Picoche, les unités graphiquement complexes, par exemple *pomme de terre*, remplissent deux critères primordiaux, notamment celui d'inséparabilité et celui de commutation, voire de substitution. Sur le plan syntaxique, il n'est donc pas possible, entre autres, d'ajouter un adjectif dans l'exemple mentionné et de dire **pomme jaune de terre*, ni remplacer une phrase *il a un joli pied à terre à la campagne* par * *un pied luxueux à terre* (1992 : 15-16). Cette unité nommée fonctionnelle a d'autres appellations également : **lexie** (Pottier), **synapsie** (Benveniste), **synthème** (Martinet), **unité syntagmatique** (Guilbert), **unité phraséologique** (Dubois) ; Picoche ajoute qu'en France, le terme le plus couramment utilisé dans ce sens-là est **lexie**, mais qu'elle préfère utiliser le terme **mot** et préciser si nécessaire s'il s'agit de mots composés ou de locutions (op. cit. : 23).

Bogaards aborde ce sujet du point de vue pédagogique et étudie l'élément qui pourrait "assumer la fonction d'unité de base dans l'apprentissage du

⁷⁵ Nous laissons ici de côté la discussion sur l'unité minimale dans le langage parlé ; voir par exemple Blanche-Benveniste (1997 : 21-45) et Elo (1993 : 37-41).

⁷⁶ Picoche donne une bibliographie de base exhaustive concernant les revues linguistiques et ouvrages généraux et spécialisés sur le lexique (1992 : 11-12).

vocabulaire”. Il l’appelle, à l’instar de Cruse (1986: 77)⁷⁷, une **unité lexicale** décrite comme “l’union d’une forme lexicale et d’un sens unique” (1994 : 19) ; ceci correspond à peu près à la lexie. Bogaards (op. cit. : 180-189) et Coady (1997 : 225-237) montrent qu’il semble évident que les mots devraient être appris plutôt comme unités lexicales dans un contexte plus large, et non comme éléments séparés. De cette façon, les significations et utilisations différentes d’un mot deviendraient plus claires, et l’utilisation des collocations spécifiques à chaque langue deviendrait plus adéquate⁷⁸ (voir aussi la discussion dans le chapitre 5.2. supra).

En statistique lexicale, le **lexique** réfère à la langue (au sens saussurien), et le **vocabulaire** au discours. Le **vocabulaire** est encore divisé en **vocab**le et en **mot** (*Ddl* 1989 : 297 ; Dugast 1980 : 56)⁷⁹. Les vocables sont des entrées du dictionnaire, tandis que les mots sont appelés également **occurrences**, c’est-à-dire que par exemple le vocable *avoir* peut se trouver dans un texte sous différentes formes *as, a, avons*, etc., qui seront comptées comme un mot/ une occurrence chaque fois qu’elles apparaissent dans un texte. Ces occurrences peuvent être lemmatisées de façon que toutes les formes d’un mot sont groupées sous une forme lexicale, le **lemme** (cf. Cossette 1994 : 4, 47 ; Dugast op. cit. : 57 ; Gustafsson 1990 : 46, 48). Etant donné que nous étudions la richesse lexicale des apprenants (cf. ch. 7.), citons encore Cossette (op.cit. : 4-5) :

“Le nombre de formules ou indices qui ont été proposés depuis 40 ans pour mesurer la richesse lexicale laisse soupçonner un problème complexe. Du point de vue statistique cependant, c’est toujours le problème du rapport entre le nombre de vocables et le nombre de mots d’un texte qui est envisagé [...]. Il s’agit là d’un postulat qui s’applique aussi, au moins dans ce qu’il a de plus fondamental, quand les deux textes ne sont pas de la même longueur et que l’on doit recourir à une analyse beaucoup plus complexe du rapport de V [= nombre de vocable] à N [= nombre de mots].”

Les analyses plus complexes auxquelles il se réfère sont la “caractéristique” K de Yule, l’indice V_m , de Yule-Herdan, l’indice R de Guiraud, $\log V/\log N$ de Herdan, la loi binomiale de Muller, l’indice W de Brunet et l’indice U(ber) de Dugast (op. cit. : 67-160). Cossette a trouvé qu’un problème essentiel des mesures est la longueur du texte qui semble perturber les formules diverses ; seuls K et la formule binomiale résistent aux effets perturbateurs, mais il a trouvé de plus que,

⁷⁷ D. A. Cruse. 1986. *Lexical semantics*. Cambridge : CUP.

⁷⁸ Il s’agit de groupes de mots comme *un ton péremptoire, une preuve irréfutable et grièvement blessé* (cf. Bogaards 1994 : 65). Faerch et al. donnent un autre exemple en anglais: “*Tea can be strong, but not heavy, whereas a person can be a heavy drinker, but not a strong drinker*” (1984 : 95). Tréville & Duquette (1996: 34-35) préfèrent le terme général de **co-occurent** et tiennent la collocation comme un cas particulier de co-occurrence.

⁷⁹ En anglais, on parle de *type* et *token*.

premièrement, K n'était pas un véritable indice de richesse lexicale, et que deuxièmement, la loi binomiale est lourde à appliquer (1994 : 167, 182). Certes, il parle de textes littéraires plus vastes (> 20.000 mots) que nos dissertations. Néanmoins, on pourrait appliquer ces formules également pour mesurer la richesse lexicale des apprenants. Nous nous référons à Grönholm (1992 : 239-252) qui a étudié l'utilité de différents indices pour mesurer la richesse lexicale ; elle a traité, entre autre, l'indice de vocable (V) / occurrence (N) (*type/token ratio*) et l'indice de Guiraud ($R = V/\sqrt{N}$). Il ressort de cette étude qu'il est difficile de montrer un indice qui serait beaucoup plus objectif ou exact que les nombres absolus pour décrire la richesse et la variation lexicales. Pour mesurer la richesse lexicale, on utilise souvent également les listes de hapax legomena ou mots qui n'apparaissent qu'une fois dans le texte (cf. Dugast 1980 : 8, 31 ; Grönholm op. cit. : 239-240).

Après la discussion supra, comment définir le mot dans l'étude présente ? Nous en sommes arrivées à la conclusion que nous devons recourir à plusieurs définitions. Nous considérons comme pertinent que du point de vue pédagogique, l'unité de base à enseigner et à apprendre est l'**unité lexicale** apprise plutôt dans le contexte approprié, mais que du point de vue statistique et linguistique, afin de pouvoir effectuer nos calculs, nous préférons parler respectivement de **vocable**, d'**occurrence** et de **mot**, (cf. le chapitre 7.). Voyons donc plus en détail ce que nous entendons par mot dans les calculs. On pourrait définir le mot comme une unité lexicale, mais de cette façon, nous devrions aussi déterminer quels éléments forment une unité étroite considérée comme un ensemble, par exemple mots composés, prépositions complexes, et expressions figées. D'autre part, nous pourrions le définir simplement comme une unité graphique, mais Ménard postule que "l'orthographe française est tout sauf un système cohérent" avec ses apostrophes et autres phénomènes spécifiques du français (1983 : 21 ; Muller 1979 : 128-130). D'autre part, selon Carter, la définition orthographique, selon laquelle le mot est une unité d'écriture comprise entre deux blancs, est "une définition pratique du bon sens" qui se réfère au moyen du langage écrit (1987 : 4).

A l'instar d'autres recherches sur le même domaine, nous avons trouvé cette dernière définition bien pratique. Ainsi en était-il dans notre étude antérieure concernant le vocabulaire utilisé dans les essais d'un examen d'entrée du

département d'études anglaises (Mutta 1995, 1997 ; cf. aussi Elo 1993 ; Mäkipää 1994). En partie aussi en raison de nos méthodes quantitatives réalisées à l'aide du logiciel **WordCruncher**⁸⁰ (cf. chapitre 7.1.), qui calcule chaque élément séparé comme un mot. Pourtant, nous avons trouvé difficile de considérer, entre autres, mots composés (p. ex. *swimming pool*), prépositions, conjonctions et pronoms complexes (p. ex. *because of, even though, each other*), comme éléments séparés, et ainsi nous les avons comptés comme un mot ; d'autres unités fixes comme des noms propres et des prédicats de phrase⁸¹ (p. ex. *of course*) ont également été considérées. Dans une certaine mesure, nous avons eu recours à une définition mixte. Nous n'étions cependant pas trop satisfaite de cette solution, car durant l'analyse nous sommes tombée sur un autre problème, celui de la délimitation des unités. En d'autres termes, où tracer la ligne de démarcation entre les unités étroitement liées et les unités moins étroitement liées ?

Dans l'étude présente, nous avons d'abord voulu procéder de la même manière que dans notre étude précédente sur l'anglais, mais ensuite, nous avons pensé pouvoir éviter la question de la délimitation en recourant à la définition orthographique pure.⁸² Afin de découvrir si les définitions différentes influent sur les résultats, nous avons effectué un test pilote sur un facteur quantitatif, notamment la variation lexicale (VL), c'est-à-dire l'indice de vocable (V) / occurrence (N) de mots lexicaux (voir aussi le chapitre 7.2.2.). Une VL élevée indique qu'il y a beaucoup de mots lexicaux différents. Une VL élevée n'est cependant pas une fin en soi car une VL modérée ne révèle que le caractère limité du vocabulaire en question, mais pas sa qualité. Nous avons choisi dix dissertations (rédigées par cinq étudiants) représentant les deux niveaux différents (voir ch. 1.2.). Les tableaux suivants montrent les pourcentages obtenus par cette analyse. Dans le premier tableau, le mot est défini selon la définition mixte (cf. supra ; cf. Mutta 1998), et dans le deuxième selon la définition orthographique pure, c'est-à-dire que chaque unité entourée par deux blancs est un mot ; ainsi les formes avec une apostrophe sont-elles des éléments séparés (p. ex. *l', d'*).

⁸⁰ WordCruncher est un logiciel simple qui permet à l'utilisateur d'obtenir et de manipuler des données sur les fichiers de texte de façon diverse. Par exemple, on peut examiner un mot dans son contexte ou imprimer des rapports statistiques incluant le nombre des mots et la distribution des fréquences lexicales.

⁸¹ Riegel et al. appellent aussi les prédicats de phrases "adverbes de commentaires phrastiques" (1997: 379).

⁸² Nous avons eu des discussions sur le choix des définitions, entre autres, avec Lavonius (1998) qui a analysé son étude lexicale à l'aide du même logiciel.

étudiant	NIVEAU I				NIVEAU II				différence de %
	nombre des mots	vocab le V	occurrence N	VL %	nombre des mots	V	N	VL %	
1	165	72	99	72,7 %	187	75	98	76,5	+3,8
2	175	81	96	84,4 %	192	87	106	82,1	-2,3
3	175	70	94	74,5 %	181	84	105	80	+5,5
4	196	97	107	90,7 %	231	112	129	86,8	-3,9
5	272	114	142	80,3 %	315	137	173	79,2	-1,1

Tableau 2. Test pilote : le mot selon la définition mixte

étudiant	NIVEAU I				NIVEAU II				différence de %
	nombre de mots	vocab le V	occurrence N	VL %	nombre de mots	V	N	VL %	
1	197	67	100	67 %	205	74	102	72,6	+5,6
2	197	85	104	81,7 %	211	86	108	79,6	-2,1
3	203	74	104	71,2 %	203	86	115	74,8	+3,6
4	220	95	114	83,3 %	246	114	135	84,4	+1,1
5	312	109	155	70,3 %	357	136	191	71,2	+0,9

Tableau 3. Test pilote : le mot selon la définition orthographique pure

Il ressort de ces deux tableaux que les étudiants ayant des VL moins élevées au niveau I (ex. 1 et 3), semblent avoir profité le plus de leur séjour en France puisque leur pourcentage de VL s'améliore au niveau II. Les autres étudiants ont des pourcentages moins élevés ou légèrement plus élevés au niveau II qu'au niveau I, mais cela résulte plutôt du fait que la longueur de l'essai (c'est-à-dire le nombre total des mots) a un effet sur le pourcentage puisqu'il est plus difficile d'éviter la répétition dans un essai plus long (cf. Grönholm 1992 ; Linnarud 1986). En résumé, nous pouvons constater d'après le test pilote que les résultats ont les mêmes tendances en dépit des définitions différentes. Par conséquent, le plus pertinent est d'avoir les mêmes critères d'un bout à l'autre du corpus. D'autant plus que d'après Dugast (1980 : 57) : "L'important est d'abord de rester fidèle à sa propre norme [...]".

Il s'ensuit de tout ce qui précède que le mot est déterminé d'une façon orthographique pure dans l'analyse quantitative. Néanmoins, nous incluons dans cette catégorie "orthographique" une quantité limitée d'unités qui n'ont pas de signification propre séparément (p. ex. *parce que*). D'ailleurs, à l'instar de Gustafsson (1990 : 48-49), nous laissons les formes non-lemmatisées dans le

dépouillement des textes⁸³, sauf quand nous examinons la sophistication lexicale (voir ch. 7.2.4.). En outre, nous sommes consciente d'un autre problème qui se manifeste lorsqu'on utilise la définition orthographique : les homonymes ou plutôt ici les homographes, c'est-à-dire les mots écrits de la même façon, mais ayant une signification différente, par exemple *chasse* et *cocher* (tantôt verbes, tantôt substantifs) (*Dictionnaire des fréquences* 1971). Comme nous utilisons le logiciel WordCruncher pour effectuer nos calculs, les homographes sont groupés sous les mêmes formes, car le logiciel ne peut pas les traiter séparément ; cela est cependant fait manuellement dans l'analyse qualitative quand cela s'avère nécessaire. Pour conclure, nous voulons encore souligner l'importance de traiter les "mots" comme unités lexicales en parlant des unités d'enseignement et d'apprentissage ; c'est ce que nous essayons de faire dans la partie de l'analyse qualitative aussi afin de prendre en compte les unités dans leur contexte.

⁸³Cossette (1994 : 47) fait la distinction entre le travail à partir d'index de formes et à partir d'index de vocables ; celui-ci réfère aux lemmes, celui-là aux occurrences. Voir aussi par exemple Lavonius (1998) pour la définition orthographique plus large et pour les formes lemmatisées.

6. ANALYSE TEXTUELLE

Etant donné la nature de notre recherche, nous n'entrons pas dans une étude exhaustive sur les phénomènes discursifs dans le corpus. Ayant le lexique comme centre d'intérêt, nous voulons éclaircir, entre autres, l'utilisation des organisateurs pragmatiques dans les textes étudiés. Dans ce chapitre, nous traitons brièvement ce domaine avant de délimiter le sujet de l'étude présente.

6.1. Analyse textuelle du point de vue lexical

Comme maints autres domaines, l'analyse textuelle fait naître des problèmes de définition. Pour définir l'analyse textuelle, on peut recourir à l'approche basée sur des théories d'origine francophone, anglosaxonne ou du groupe de Bakhtine. Ainsi entre-t-on dans la discussion terminologique portant sur l'analyse du discours et sur l'analyse textuelle. Quant à l'étude présente, la distinction entre les deux concepts semble peu pertinente, et ainsi, à l'instar de plusieurs chercheurs, nous préférons parler de l'analyse textuelle.⁸⁴

Lundquist dit, quant à l'analyse textuelle, que les concepts de base dans celle-ci sont *les actes de langage* et *la cohérence textuelle* ; les deux permettent l'articulation du textuel qui comprend les phénomènes linguistiques de la surface du texte et l'extra-textuel "qui s'étend de la situation de communication vers la sphère de production de textes de l'organisation sociale" (1983 : 9-10). Etant donné la délimitation du travail, nous nous passons des actes de langage. Il nous reste donc les termes de **cohésion** et de **cohérence**. Widdowson (1981 : 37-39) indique clairement une dichotomie : la cohésion est un rapport explicite entre les

⁸⁴ Lundquist (1980 : 1), entre autres, parle de la linguistique textuelle ; selon elle le texte est l'unité linguistique primaire, non la phrase, car on "s'exprime en texte et non pas en phrases isolées". Néanmoins, elle postule que ceci est dû à un flou conceptuel de différentes écoles linguistiques ; en France le texte, de tradition allemande, se mêle avec le discours et l'énoncé. Elle nous renvoie à Maingueneau qui parle de la polysémie du concept 'discours' (1976 : 11). Lundquist dit que le texte peut être pris dans des sens divers qui chevauchent les définitions de l'approche française (op. cit. : 12). En outre Suomela-Salmi (1997 : 25-28) nous rappelle que texte et discours sont "souvent utilisés sans distinction. Or ils ont des acceptions qui varient selon les traditions scientifiques" ; la distinction entre l'écrit et l'oral, différences théoriques et méthodologiques, le discours étant un concept plus large que le texte (discours = texte + conditions de production) (op. cit. : 33-36). Ce dernier nous renvoie à la définition d'Adam, résultant de la discussion sur divers points de vue, entre autres, de Bakhtine, de Foucault et de Maingueneau (1990 : 23) : "En d'autres termes, un discours est un énoncé caractérisable certes par des propriétés textuelles, mais surtout comme un acte de discours accompli dans une situation (participants, institutions, lieu, temps) ; ce dont rend bien compte le concept de "conduite langagière" comme mise en oeuvre d'un type de discours dans une situation donnée. Le texte, en revanche, est un objet abstrait résultant de la soustraction du contexte opérée sur l'objet concret (discours)".

propositions, tandis que la cohérence peut être implicite.⁸⁵ Pour leur part, Riegel et al. font remarquer que si on abandonne la distinction entre le discours et le texte, on peut ainsi arriver à parler communément de cohérence qui caractérise “la bonne formation aussi bien du texte que du discours” (1997 : 603)⁸⁶. Vu la discussion supra, nous nous contentons de parler de la cohérence en général.

En effet, la cohérence textuelle consiste en différents éléments linguistiques. Lundquist (1980 : 21, 28, 49) parle de la **cohérence par coréférence** (par exemple l’anaphore et la cataphore), de la cohérence par **contiguïté sémantique** (p. ex. *la victoire – la défaite*) et des **connecteurs** (p.ex. *et, mais, entre autres*) qui sont essentiels pour la cohérence textuelle. Pour leur part, Riegel et al. (1997 : 604) postulent qu’il y a deux règles principales de cohérence du texte : règle de répétition et de progression. Ils y ajoutent aussi une règle de non-contradiction implicite ou explicite entre les différents éléments du texte. Le degré de l’application de ces règles indique la cohérence du texte en question. Pour la continuité cohérente des séquences linguistiques, la répétition semble être une condition nécessaire, mais pas suffisante telle quelle ; il faut aussi garantir la progression (Adam 1990 : 52, 104). Adam dit dans une de ses quatre hypothèses principales pour une pragmatique textuelle :

“Le texte est un produit connexe, cohésif, cohérent (et non pas une juxtaposition aléatoire de mots, phrases, propositions ou actes d’énonciation)”. (hypothèse 2 ; op. cit. : 109)

“Tout texte est soumis à une tension constitutive entre continuité-répétition et/vs progression”. (hypothèse dérivée 2.3. ; op. cit. : 113)

Il réfère à des moyens linguistiques comme nominalisation, co-référence, mémoire textuelle, structure thème-rhème, etc. (op. cit. : 107-120).⁸⁷

Dans le chapitre 5.2., nous avons souligné l’importance du contexte quant à l’apprentissage et l’enseignement du vocabulaire. Le contexte est également

⁸⁵ L’exemple bien connu de Widdowson où il y a cohérence, mais pas de cohésion est le suivant : “A: That’s the telephone. B : I’m in the bath. A : O. K.” (1981 : 39). Voir aussi la définition et les remarques de Suomela-Salmi (1987 : 115, 131-138).

⁸⁶ Ils expliquent leur raisonnement : “En s’appuyant sur l’opposition entre le discours et le texte, on distingue la *cohérence* et la *cohésion*. La cohérence est une propriété du discours, qui est mis en relation avec les conditions de l’énonciation, alors que la cohésion est une propriété du texte, qui est envisagé fermé sur lui-même. [...] Dans la pratique, il est difficile d’opérer une stricte répartition entre des règles de cohérence à portée externe et des règles de cohésion à portée interne [...]” (Riegel et al. 1997 : 603).

⁸⁷ Certes, il y a également d’autres modèles d’approche de cette problématique chevauchant les classifications présentées ; par exemple Charolles (1978 : 12) a discuté la question brièvement en termes de quatre méta-règles de cohérence, notamment celles de répétition, de progression, de non-contradiction et de relation.

primordial dans la recherche lexicale, en particulier dans l'analyse qualitative. C'est ici qu'entre en jeu la cohérence. Linnarud (1986: 8) postule que :

“A composition is not just a collection of separate sentences which happen to have been written out one after the other. It is, or should be, a text which is both coherent and cohesive. Lexis plays an important part in this respect”.

Portine (1983 : 135-135) voit également le rapport étroit entre l'écriture, le lexique et la dimension textuelle. Selon lui, l'ensemble cohérent du texte est composé, entre autres, de cohérence anaphorique, de résolution des contradictions internes, de cohérence lexicale et de cohérence sociolinguistique (c'est-à-dire des changements de registre volontaires). Reichler-Béguelin et al. (1988 : 7) sont arrivées à la même conclusion quant à l'importance de la cohésion textuelle comme partie de l'apprentissage et de l'enseignement de l'écriture ; elles ont déduit cette conclusion de leurs pratiques pédagogiques. En d'autres termes, leurs apprenants non francophones commettaient beaucoup d'erreurs liées aux problèmes de mise en texte, par exemple il y avait un manque de cohérence logique et argumentative. Ce qui, comme nous l'avons déjà vu dans le chapitre 5.2., peut gêner la compréhension. En outre, Péry-Woodley suppose que la cohérence d'un texte est liée à nos connaissances préalables et à nos préjugés : “la cohérence n'est pas dans le texte, elle est le résultat d'une interaction (réussie ?) entre texte et lecteur.” (1993 : 60).

6.2. Activité métatextuelle

Quand on fait de la recherche sur le langage écrit, il faut prendre en considération le fait que le processus de l'écriture est déjà en partie une activité métalinguistique à cause de la nature même du processus de l'écriture (Bouchard 1994 : 122). Comme nous l'avons déjà évoqué dans le chapitre 4.3., à la différence du langage oral, le scripteur a plus de temps pour réfléchir à sa production et pour la réécrire, mais en même temps, il doit être plus explicite dans sa production pour que le message soit compris comme voulu ; cela est dû au fait que le lecteur ne peut pas recourir au feed-back direct pour avoir des renseignements supplémentaires sur le contenu. Une partie essentielle dans l'amélioration des capacités d'écriture est l'amélioration des capacités textuelles de l'apprenant car celles-ci favorisent l'augmentation de la cohérence textuelle et ainsi également l'augmentation de la compréhension textuelle. Au début de l'apprentissage, l'apprenant mobilise plutôt ses capacités métaphonologiques, métasyntaxiques et métasémantiques,

tandis qu'au fur et à mesure de l'apprentissage l'apprenant laisse la priorité aux capacités métapragmatiques et métatextuelles (Gombert 1991 : 152, cité dans Bouchard 1994 : 122). Quant à notre étude, on peut remarquer que les métaconnaissances lexicales sont mises en évidence surtout au niveau universitaire où les étudiants s'orientent vers un auto-apprentissage plus réfléchi (Sautermeister 1989 : 123-124, 132). Cet auto-apprentissage plus réfléchi s'exprime par une aisance dans l'écriture qui se manifeste dans l'amélioration des capacités textuelles de l'apprenant de langue ; en d'autres termes, la cohérence interne du texte augmente.

Pour effectuer nos analyses, nous pourrions faire comme l'ont fait Tréville & Duquette (1996 : 39-44) dans leur ouvrage (certes sur l'enseignement) et étudier l'usage systématique des indices contextuels, par exemple la répétition des mots, l'emploi des synonymes, etc. Au lieu de cette approche, nous voulons éclaircir l'utilisation des organisateurs pragmatiques des textes étudiés (voir ci-dessous pour la définition des termes). Pourquoi cela ? Primo, pour une question de délimitation, et secundo, parce que l'étude préliminaire du corpus indique que les textes deviennent plus "naturels" au niveau II (c'est-à-dire après le séjour linguistique ; cf. ch. 1.2.), partiellement à cause de l'augmentation de l'utilisation des organisateurs pragmatiques. Phénomène déjà constaté par d'autres recherches ce qui prouve que ceci est un développement naturel chez l'apprenant d'une L2. Chez les apprenants enfants, la différence développementale est plus nette (cf. Verhoeven 1997), tandis que chez les apprenants adultes, cela se passe un peu différemment, car ces derniers peuvent avoir recours à leurs connaissances préalables, donc à celle de leur L1 (cf. Noyau & Vasseur 1986 : 117 ; Perdue 1993 : 9-10). D'ailleurs, il y a, en effet, une similitude également entre l'ordre acquisitionnel des connecteurs par les enfants d'une L1 et celui des apprenants adultes d'une L2 (Kerr-Barnes 1998 : 206). Ce qui diffère dans la production des enfants et des adultes, c'est que les adultes disposent déjà de techniques pour rendre leur production plus naturelle, mais ils ont cependant quelquefois des difficultés à se débarrasser des habitudes acquises dans la L1, surtout à l'oral (cf. Kerr-Barnes 1998 : 207). Certes, les problèmes d'apprentissage des adultes sont multiples par ailleurs.

6.3. Organismes pragmatiques du texte

On peut également aborder autrement la question de la cohérence textuelle. Pour sa part, Charolles (1988 : 4) postule :

“La simple *cohérence thématique* (entendue au sens vague de convergence des contenus) ne suffit donc pas à garantir la bonne formation textuelle. [...], les unités qui le [= texte] constituent doivent satisfaire à certains critères d'organisation. [...] il [= le sujet écrivain] doit *aménager* ce qu'il dit de façon à ce que ceux à qui il s'adresse puissent interpréter sans difficultés insurmontables les rapports qu'il établit entre les énoncés qu'il produit successivement”.

Ainsi, Charolles distingue-t-il quatre plans différents organisant le texte, à savoir **période**, **chaîne**, **portée** et **séquence** (1988 : 6-10). Selon lui, la période consiste en éléments phrastiques qui entretiennent des rapports de dépendance (par exemple la coordination et la juxtaposition), tandis que les chaînes réfèrent à des suites d'expressions coréférentielles (par exemple la pronominalisation). Les portées, quant à elles, sont définies comme toute portion de texte dont l'interprétation s'effectue dans un certain cadre (p. ex. *selon X, pour X*). Finalement, les séquences font références au découpage du texte par le sujet parlant/écrivain (p.ex. *quant à, donc*). Selon Charolles, tous ces plans d'organisation cohabitent dans le texte.

Il existe, en effet, plusieurs façons de classifier l'organisation du texte⁸⁸. Entre autres, Adam, reformulant l'idée de Charolles, distingue six plans différents pour l'organisation textuelle : 1. les chaînes et phénomènes locaux de liage (par exemple l'anaphore), 2. les espaces sémantiques (par exemple la prise en charge par des marques d'attribution des propositions), 3. la segmentation textuelle (par exemple la ponctuation, les paragraphes), 4. la période (les aspects rythmiques du texte), 5. la structuration séquentielle ou hiérarchique, et 6. la dimension pragmatique-configurationnelle (par exemple l'intentionnalité de texte) (1990 : 51-104). Pour sa part, Lundquist (1983 : 10-11) distingue six niveaux différents dans une analyse textuelle:

1. le niveau pragmatique – par exemple le rapport entre le texte et ses usagers dans un environnement particulier
2. le niveau thématique - par exemple la progression thématique du texte (thème et rhème)
3. le niveau sémantique - par exemple les relations interphrastiques établies par les connecteurs

⁸⁸ Par exemple, Péry-Woodley discute les écrits du point de vue didactique, par exemple elle parle des moyens à la disposition du scripteur pour réaliser la structure thématique dans l'écrit (1993 : 66-83).

4. le niveau syntaxique – par exemple la syntaxe et la cohérence
5. le niveau rhétorique - par exemple la répétition et l'antithèse
6. le niveau idéologique - le rapport avec l'environnement social

En bref, nous pouvons constater que ce qui nous intéresse dans cette étude, c'est le développement métadiscursif de l'apprenant pour organiser son texte en unités sémantiquement et logiquement plus naturelles qu'avant. Charolles dit (1988 : 9) :

“Les marques de séquentialité sont l'indice d'une activité *métadiscursive* chez celui qui les utilise. Elles dénotent un travail explicite d'organisation de l'énonciation visant en particulier à faciliter la tâche de l'interprétation.”

Ce travail explicite est d'autant plus nécessaire dans la production écrite à cause de ses caractéristiques inhérentes, en d'autres termes, des contraintes spécifiques du processus d'écriture (voir ch. 4.3.). De plus, Péry-Woodley (1993 : 61-62) postule que dans les milieux de la didactique de l'écrit, à savoir dans les milieux scolaire ou universitaire, les exigences explicites guidant le processus mental du lecteur vers celui du scripteur, sont explicites dans des textes produits pour être évalués “où le scripteur a tout à gagner à être directif”.

Il s'ensuit que le niveau d'analyse qui nous intéresse dans la présente étude est principalement le niveau sémantique et plus précisément les relations interphrastiques, en termes de Lundquist ; en termes d'Adam, la segmentation textuelle et la structuration séquentielle ; et en termes de Charolles, la période, la portée et les séquences (en partie). Nous aboutissons ainsi dans notre étude à une analyse bipartite d'organisation textuelle : primo, la segmentation, soit le découpage en paragraphes et l'utilisation de la ponctuation, et secundo, les marqueurs visibles dans le texte qui jouent un rôle de passeurs d'une phrase à l'autre. Les deux parties forment ce que nous appelons les organisateurs pragmatiques du texte.

A première vue, la ponctuation semble un phénomène peu pertinent quant à la cohérence textuelle. Mais cela n'est qu'illusoire car la ponctuation dispose, en effet, d'un fonctionnement textuel. Fayol, à maintes reprises, a pu prouver que la ponctuation interpropositionnelle marque un degré de liaison ou de distance entre les propositions adjacentes (1989 : 37). Cela témoigne des relations interphrastiques qui sont signalées par des regroupements d'unités significatives d'une manière hiérarchique. La ponctuation représente une sorte de modèle

mental sous-jacent du texte que le sujet écrivain émet en organisant le texte (cf. aussi Coirier et al. 1996 : 133-134). Dans le même sens va Bessonnat (1988 : 94) en discutant du découpage d'un texte en paragraphes, ce qui est un outil métatextuel d'ordre métalinguistique, selon lui. Il estime que le découpage en paragraphes a un rôle triparti, à savoir un rôle de facilitation de la lecture, un rôle de programmation hiérarchique du texte et finalement, un rôle dialogique (op. cit. : 85-87). Nous avons également pu remarquer dans une étude antérieure sur un corpus anglais (il s'agissait de dissertations d'apprenants d'anglais dans un examen d'entrée) que les apprenants organisant leur texte en paragraphes réussissent mieux à l'examen (Mutta 1995, 1997). Dans la présente étude, nous n'entrerons pas dans une analyse exhaustive de la segmentation, mais nous ferons des réflexions générales sur l'utilisation des paragraphes et de la ponctuation.

Les marqueurs visibles du texte auxquels nous référons sont ceux qu'on appelle souvent les **connecteurs** qui d'après Lundquist sont "des particules de liaison, telles que "ainsi", "là", "donc", "pour cette raison", qui jouent le rôle de "passeurs" d'une phrase à une autre" (Lundquist 1980 : 47). Riegel et al. appellent **connecteurs** les éléments de liaison entre des propositions ou des ensembles de propositions dans l'enchaînement linéaire du texte. D'après eux, on peut définir les connecteurs soit au sens restreint, c'est-à-dire en tant que "termes qui assurent la liaison à l'intérieur d'une phrase complexe : les conjonctions de coordination et de subordination" ; soit les connecteurs considérés au sens plus large, c'est-à-dire tous les termes assurant l'organisation d'un texte. Dans ce dernier sens, les connecteurs se chevauchent avec certains procédés anaphoriques comme des expressions spatiales (1997 : 616-617). Cependant, Riegel et al. veulent limiter ce concept aux unités "qui ne font pas partie intégrante des propositions, mais qui assurent leur liaison et organisent leurs relations, sans être des expressions anaphoriques" (op. cit : 617). Ils suivent ici l'exemple de Schneuwly et al. (1989). Schneuwly et al. (1989 : 40-44) introduisent le terme **organiseurs textuels** pour désigner ce qui dépasse l'analyse au plan local, c'est-à-dire tout ce qui permet d'établir une liaison étroite entre deux propositions. Selon eux, les organisateurs textuels, qui peuvent avoir plusieurs sous-catégories (par exemple organisateurs temporels), comprennent aussi bien des opérations de connexion que de segmentation qui varieront en fonction du contexte. Il s'agit donc d'unités comme les connecteurs, les diverses locutions et les expressions temporelles ou argumentatives (p. ex. *et, mais, parce que, soudain, en effet, d'une part*).

Suomela-Salmi (1993 : 241-242) adopte également ce terme d'organisateur textuels, mais elle compte parmi eux aussi certains symboles numériques et certains faits relevant plutôt de la ponctuation.

Selon Adam, la mise en texte est orientée par des contraintes de linéarisation qui sont explicites dans les propositions. Ces structures transphrastiques peuvent être marquées grammaticalement par des connecteurs ou des organisateurs, ou uniquement par la ponctuation et/ ou les phénomènes de reprise. Il fait la distinction entre la catégorie générale de **connecteurs** consistant en conjonctions, locutions conjonctives et adverbales qui sont "argumentativement marqués" et celle des **organisateur**s, "non marqués argumentativement" (1990 : 41). Adam donne des exemples concrets dans un texte descriptif sur l'utilisation de certains connecteurs, notamment *mais*⁸⁹ et *certes*, ainsi que de certains organisateurs énumératifs qu'il appelle, à l'instar de Turco & Coltier (1988)⁹⁰, **marqueurs d'intégration linéaire** (p. ex. *d'abord, puis*). Selon lui, il est difficile de classer des unités linguistiques dans cette catégorie d'intégration linéaire car il n'existe pas de morphèmes spécifiques exclusivement réservés à cette fonction. Selon lui, il faut recourir à d'autres sous-systèmes comme les organisateurs temporels, spatiaux et additifs (p. ex. *la veille, au nord, en plus, par ailleurs*). Les marqueurs d'intégration peuvent avoir une fonction de marquage d'ouverture, de relais ou de clôture d'une séquence (op. cit. : 142-167). Adam compte les organisateurs *il y a* et *c'est* comme un groupe à part soulignant l'enchaînement et l'empaquetage, c'est-à-dire le traitement hiérarchique des unités linguistiques. Lundquist, de son côté, parle de *c'est* présentatif (1980 : 170). L'organisateur *c'est* a également une fonction reformulative qui représente un autre moyen essentiel dans un texte descriptif de garantir non seulement la cohésion, mais aussi la progression du texte. Dans cette catégorie, il s'agit de marqueurs comme *bref, c'est-à-dire, autrement dit* (op. cit. : 168-174).

⁸⁹ Maingueneau (1987 : 120-123, 135) indique que le connecteur *mais* a été le plus étudié et qu'il présente un intérêt certain pour l'analyse du discours dû, entre autres, à sa fréquence. Adam (1990 : 192-211) s'oppose à la distinction traditionnelle (par exemple de Ducrot et Anscombre) qui postule qu'il y aurait seulement deux *mais* en français : un *mais* de réfutation et un *mais* d'argumentation. Il distingue donc cinq fonctionnements de *mais* (de renforcement-renchérissement, réfutatif, phatique, concessif et argumentatif) qui certes sont quelquefois proches en apparence. A son avis, il serait profitable de distinguer nettement les fonctionnements procéduraux différents de *mais*. Pourtant, nous ne le ferons pas dans cette étude.

⁹⁰ En effet, Turco & Coltier (1988 : 57-58) appellent les marqueurs d'intégration linéaire (MIL), à la suite de A. Auchlin [1981. "Réflexions sur les marqueurs de structuration de la conversation". *Etudes de Linguistique appliquée* 44: 97], les morphèmes qui "accompagnent l'énumération sans fournir de précision autre que le fait que le segment discursif qu'ils introduisent est à intégrer de façon linéaire dans la série".

Il s'ensuit de ce qui précède que nous comptons parmi les **organiseurs pragmatiques** les liens transphrastiques qui assurent la cohérence entre les différentes propositions organisant le texte. Cependant, nous allons limiter notre analyse à la macro-structure ou au niveau macro-textuel (cf. Adam 1990 : 115), ce qui veut dire, entre autres, que nous n'examinerons pas les connecteurs de subordination (cf. Riegel et al. 1997 : 617)⁹¹. Ces organisateurs pragmatiques correspondent ainsi aux connecteurs et aux organisateurs textuels, mais d'une façon sélective (cf. le chapitre 8.3.2.2.) ; nous allons recourir principalement aux classifications d'Adam (1990), Lundquist (1980, 1983), et de Riegel et al. (1997) en les modifiant et en les regroupant.

De plus, nous voulons ajouter dans ce classement une partie des "modalisateurs" qui indiquent un certain type d'attitude du locuteur par rapport à son énoncé. Il s'agit des adverbes ou des locutions adverbiales qui sont souvent en position de complément de phrase comme *évidemment* et *à mon avis* (cf. Riegel et al. 1997 : 581-582) ; nous les examinerons seulement quand ils représentent clairement un lien sémantique entre deux propositions organisant ou faisant progresser le texte. De même, nous y incluons, à l'instar d'Elo, des expressions qu'on peut appeler connecteurs métatextuels, par exemple *je crois que* et *je pense que* quand les étudiants les "utilisent plutôt comme des introducteurs sans vraiment y attacher de contenu sémantique" (1993 : 93). Elo parle des connecteurs dans le langage oral, mais ce phénomène est également présent dans le langage écrit. D'autant plus que l'une de nos hypothèses pose que des traits du langage parlé s'immiscent dans la production écrite après le séjour à l'étranger. Un bon exemple de ce phénomène est le marqueur de clôture *voilà* qui est propre au discours oral (cf. Adam 1990 : 157 ; Schneuwly et al. 1989 : 45).

⁹¹ Vargas discute explicitement la différence entre les deux niveaux d'organisation textuelle, notamment la micro- et macro-organisation textuelle. Selon lui, les connecteurs, d'après une définition minimale, font partie de la micro-organisation textuelle (c'est-à-dire du niveau local référant aux phrases), mais qu'il existe également des connecteurs textuels de la macro-organisation (autrement dit, au niveau global référant à de grandes unités de contenu) (1999 : 55-58). Notre approche d'analyse se situe plutôt entre ces deux extrêmes.

7. METHODES UTILISEES POUR L'ANALYSE DU CORPUS

Nous avons fait référence aux principes qui dirigent cette recherche au fur et à mesure de l'encadrement théorique de notre étude. Avant l'analyse du corpus, explicitons encore les méthodes quantitatives et qualitatives.

7.1. Le logiciel WordCruncher : de son utilité dans la recherche

En parlant de la définition du mot dans la présente étude, nous avons brièvement mentionné dans le chapitre 5.3. que pour effectuer nos calculs nous avons recours à un logiciel, WordCruncher, qui calcule chaque élément entre deux blanc comme un mot isolé. C'est un logiciel qui a été développé à l'université de Brigham Young, dans l'Utah aux Etats-Unis, pour la recherche lexicale. Une fois que le texte à examiner est introduit dans le programme, on peut imprimer des listes de mots par ordre alphabétique accompagnées d'indication concernant la distribution des fréquences lexicales. De plus, le mot peut être examiné dans son entourage linguistique, donc dans son contexte, et le logiciel permet de faire des listes de concordance. Même si ce logiciel est simple, il facilite notre recherche quantitative du corpus.

Quant à la définition du mot, nous sommes arrivée à une définition orthographique à laquelle nous avons inclu quelques unités étroitement liées l'une à l'autre que nous avons estimées difficiles à séparer. Nous sommes consciente du fait qu'il est extrêmement difficile de rester dans des limites raisonnables et ainsi nous avons seulement une toute petite quantité de mots traités de cette manière. Lavonius, par exemple, a choisi un groupe plus large dans cette perspective (1998 : 9-10, 37), comme nous-même d'ailleurs dans une recherche antérieure portant sur l'anglais (Mutta 1995). En voici quelques exemples de notre corpus : *n'est-ce pas, parce que, par exemple, peut-être, d'après*. Selon Nation (1990 : 29-30), les mots devraient être définis d'après des critères clairs pour que leur apprentissage devienne moins contraignant. Un de ses critères étant la fréquence des mots, il faut bien entendu voir de plus près comment la fréquence est calculée ; par exemple, on peut compter les mots séparément d'après les formes et ainsi le singulier et le pluriel d'un mot sont des mots différents. Dans ce cas-là, on parle de formes non lemmatisées, donc des occurrences dans le texte étudié. Nous avons fait nos calculs d'après les formes non lemmatisées (cf.

Gustafsson 1990 : 48-49), pour des raisons de simplicité (cf. néanmoins infra).

7.2. Facteurs linguistiques et quantitatifs

L'analyse quantitative du corpus a été donc faite partiellement à l'aide du logiciel WordCruncher qui rend possible le traitement objectif du vocabulaire, au moins plus objectif qu'une simple analyse d'erreurs. On pourrait dire que ce qui nous intéresse, c'est la richesse lexicale⁹² des apprenants aussi bien en tant que groupes d'étudiants qu'en tant qu'apprenants individuels et le développement, soit l'augmentation de cette richesse lexicale. Dans le chapitre 5.3. nous avons déjà cité Cossette (1994 : 4-5) qui affirme que la façon de mesurer la richesse lexicale soulève un problème complexe depuis 40 ans, mais que du point de vue statistique, il s'agit cependant presque toujours des rapports entre le nombre de vocables et d'occurrences. Il en va de même avec le postulat de Ménard affirmant que les critères d'ordre qualitatif sont inopérants quant à la richesse lexicale (1983: 16-17). Néanmoins, il ajoute que ces critères aident à préciser les résultats de l'analyse quantitative. Il présente un tableau récapitulatif (certes sur la littérature comme la plupart des recherches statistiques sur le lexique) des aspects de ce problème, des solutions proposées (par exemple l'indice de Guiraud) et des observations faites sur des modèles statistiques différents. Harley & King⁹³ (cité dans Tréville & Duquette 1996 : 159-160) basent la quantification de la richesse lexicale, à l'instar de plusieurs chercheurs, sur l'utilisation des verbes. Elles élaborent des calculs comme la justesse lexicale, la variété lexicale, la spécificité lexicale et la sophistication lexicale (voir infra).

Nous avons adopté la méthode quantitative utilisée par Moira Linnarud dans sa recherche *Lexis in composition* (1986). Elle a étudié divers aspects lexicaux dans les dissertations écrites par les apprenants suédois et les natifs à partir de facteurs simples comme le nombre de mots par phrases, et en progressant vers des facteurs

⁹² Rappelons que la richesse lexicale est déterminée au sens large dans cette étude, incluant notamment aussi bien des phénomènes quantitatifs que qualitatifs. Comme le dit Arnaud, la richesse lexicale comprend normalement le nombre de vocables et d'occurrences dans la recherche statistique lexicale (1984 : 14).

⁹³ B. Harley & M. King. 1989. "Verb lexis in the written compositions of young L2 learners". *Studies in Second Language Acquisition* 11(4) : 415-439.

plus complexes comme la sophistication lexicale⁹⁴. Cependant, étant donné le caractère de son travail, nous avons été obligée de l'adapter en partie à nos propres besoins. Nous examinerons des facteurs linguistiques comme la variation lexicale, la densité lexicale, l'individualité et la sophistication lexicale (voir ci-dessous). L'ensemble de ces phénomènes donnera probablement une image plus complète de la richesse lexicale de chaque étudiant que l'analyse d'un phénomène isolé. On a calculé également les corrélations entre ces phénomènes afin de voir si les changements prévus sont statistiquement significatifs. L'analyse des erreurs, l'étude des organisateurs pragmatiques et les vocabulaires des apprenants seront examinés dans leur contexte entier dans la partie de l'analyse qualitative.

7.2.1. Nombre de mots

Après avoir décidé comment définir le mot pour nos calculs, le logiciel WordCruncher nous a calculé le nombre total des mots dans chaque dissertation ainsi qu'au niveau des corpus différents. Le nombre des mots des dissertations nous a intéressé car nous avons voulu voir si la longueur du texte était en corrélation avec les autres facteurs quantitatifs. Selon Arnaud (1984 : 16), on a pu prouver dans maintes recherches qu'il existe une corrélation significative entre la longueur de la dissertation et sa qualité générale. Pourtant, il faut noter qu'un texte long devient facilement répétitif et incohérent (Arnaud 1984 : 25). Avant de pouvoir calculer les autres facteurs quantitatifs, nous devons encore faire la distinction entre les mots pleins et les mots outils. Cette distinction peut s'avérer quelquefois ambiguë (cf. Arnaud 1984 : 15-16), mais nous avons adopté une acception généralement admise sans entrer dans cette discussion.⁹⁵ Les mots outils, c'est-à-dire les termes grammaticaux, comprennent, entre autres, les articles, les pronoms, et les prépositions, tandis que les mots pleins, c'est-à-dire les termes lexicaux, contiennent entre autres les noms, les adjectifs, les adverbes suffixés en *-ment* et les verbes. Les mots grammaticaux servent à montrer les rapports syntaxiques entre les mots lexicaux ou entre le discours et le monde

⁹⁴ Les facteurs linguistiques que Linnarud (1986) a étudiés sont : le nombre de mots par composition, le nombre d'erreurs par composition et leurs pourcentages, le nombre de mots par phrases, l'individualité lexicale, la sophistication lexicale, la variation lexicale et la densité lexicale. De plus, elle a étudié les corrélations entre les différents facteurs linguistiques et le choix des mots utilisés et leur distribution en parties du discours. De plus, elle discute ces résultats du point de vue des évaluations données sur les dissertations.

⁹⁵ Pour les catégories des mots pleins et des mots outils, voir par exemple Bogaards (1994 : 50), Carter (1987 : 8), Linnarud (1979 : 221 ; 1986 : 4) Riegel et al. (1997 : 381) et Tréville & Duquette (1996 : 12).

extralinguistique (Bogaards⁹⁶ 1994 : 50-51; Carter 1987 : 8).

7.2.2. Variation lexicale

Nous avons utilisé la formule de Linnarud (1986 : 46-47) pour calculer la **variation lexicale (VL)** dans ces dissertations. Dans la présente étude, la variation lexicale comprend l'indice des mots lexicaux (*type/token ratio*) de chaque dissertation⁹⁷ :

$$VL = \frac{\text{vocables lexicaux}}{\text{occurrences lexicales}} \times \frac{100}{1}$$

vocable = mot lexical différent
occurrence = tout mot lexical⁹⁸

Selon Linnarud (op. cit. : 46), une variation lexicale élevée indique qu'il y a beaucoup de mots lexicaux différents, autrement dit, qu'il y a peu de répétition lexicale. D'ailleurs, elle n'a pas adapté la VL à la longueur de la dissertation comme l'ont fait Tiilikainen & Ylimentalo (1989 : 133-134). Celles-ci avaient d'abord calculé la VL et l'avaient ensuite proportionnée à la longueur de la dissertation (**VL/M**, où M = nombre total des mots). Tiilikainen & Ylimentalo nous font cependant remarquer qu'un indice très élevé de VL/M n'est pas une fin en soi car un indice modéré ne révèle que le caractère limité du vocabulaire en question, mais pas sa qualité, c'est-à-dire sa sophistication. De toute façon, une variation lexicale élevée serait souhaitable dans les dissertations des étudiants finlandais du point de vue de la richesse lexicale et ainsi également de la cohérence du texte écrit en français (cf. Arnaud 1984 : 17; Charolles 1988 : 8), même si ce n'est pas le cas pour le finnois. Suomela-Salmi en témoigne :

“Comme les répétitions sont mal tolérées dans la langue française où elles sont plutôt interprétées comme des signes de mauvais style, on peut supposer que leur part n'est pas

⁹⁶ Bogaards élabore néanmoins cette idée plus loin en se référant à A. Wierzbicka (1988 : 8) et en disant : “(Wierzbicka) s'inscrit en faux contre cette opposition en déclarant que “tous les mots sont des ‘mots outils’”. Je suis de son avis, et cela dans un double sens. En premier lieu, il n'est pas superflu de rappeler que tous les mots remplissent nécessairement une fonction grammaticale, le contenu sémantique étant toujours exprimé dans un cadre syntaxique quelconque. En second lieu, le poids sémantique ou grammatical d'un mot, ou plutôt d'une UL [= unité lexicale], n'est pas une question de tout ou de rien, mais se présente comme un alliage et doit être compris en termes de proportions, d'apports relatifs”. Pour les calculs de la présente étude, nous nous contentons de la distinction traditionnelle.

⁹⁷ Normalement (par exemple dans les études de cas), le pourcentage des vocables et des occurrences concerne tous les mots, soit pleins ou outils, mais dans cette recherche, nous nous intéressons seulement à la variation des mots lexicaux, c'est-à-dire pleins, à l'instar de Linnarud.

⁹⁸ Inversement à la façon de calculer qu'on trouve par exemple dans Linnarud (1979, 1986) et Gustafsson (1990), Cossette (1994 : 16-17) et Ménard (1983 : 51), entre autres, parlent de la relation N/V (où N = nombre de mots ou occurrence et V = nombre de vocables) dans ce même sens.

aussi importante que dans les textes finnois ou anglais par exemple” (1997 : 207-208 ; voir aussi Suomela-Salmi 1987).

7.2.3. Densité lexicale

La **densité lexicale (DL)** indique le pourcentage des mots lexicaux par rapport au nombre total des mots dans chaque dissertation. La formule de Linnarud est la suivante (1986 : 47) :

$$DL = \frac{\text{mots lexicaux}}{\text{nombre total des mots}} \times \frac{100}{1}$$

Linnarud ajoute que quant aux calculs portant sur la richesse lexicale au sens large, la densité lexicale est un bon indice car il n’est pas affecté par la longueur du texte. Dans un texte écrit par un autochtone, la quantité de mots grammaticaux dépasse normalement celle des mots lexicaux. Tréville & Duquette (1996 : 12) précisent que les mots grammaticaux représentent à peu près 0,5 % de l’inventaire dans le lexique, soit dans le dictionnaire, tandis qu’au niveau du vocabulaire, donc de la parole, ils peuvent représenter environ 60 % des énoncés. En outre, Ménard postule qu’une fréquence très élevée de mots grammaticaux appauvrit le vocabulaire d’un texte. Néanmoins, même si la densité lexicale est un indice de la compétence lexicale, une DL élevée n’est pas non plus une fin en soi; il faut recourir à une analyse qualitative pour y voir de plus près. D’autant plus que cette proportion varie pour des raisons stylistiques (Ménard op. cit. : 52)⁹⁹. Linnarud va dans le même sens aussi. Certes, les résultats de Linnarud sur les dissertations écrites en anglais ne sont pas directement comparables avec un corpus français, mais ils peuvent donner une idée générale de la situation. Elle a pu montrer qu’arriver à une DL au-dessus de 40 % correspond au niveau d’un scripteur natif dans une dissertation (= un texte non-interactif). Elle ajoute cependant qu’une DL élevée n’est pas toujours un facteur positif et qu’elle ne s’accompagne pas non plus automatiquement d’une progression dans le taux de réussite des étudiants, car quand le texte est trop dense, elle devient facilement moins lisible (1986 : 82, 118).

⁹⁹ Gustafsson présente également la densité lexicale comme un marqueur de style, mais elle a recours à des formules différentes des nôtres (1990).

7.2.4. Individualité et sophistication lexicales

L'**individualité lexicale (ILe)** correspond ici à l'originalité lexicale qui est calculée d'après les mots à occurrence unique chez un étudiant dans l'ensemble du corpus ; ces "mots uniques" ont été comparés au nombre des mots lexicaux dans la dissertation de l'étudiant en question. En voici la formule (d'après Linnarud 1986 : 44) :

$$\text{ILe} = \frac{\text{mots lexicaux à occurrence unique chez un étudiant}}{\text{nombre total des mots lexicaux dans la dissertation}} \times \frac{100}{1}$$

Cet indice correspond à peu près au concept de **hapax legomena**, c'est-à-dire aux mots dont on ne relève qu'un exemple dans un texte (cf. Dugast 1980 : 8 ; Gustafsson 1990 : 48, 53 ; Grönholm 1992 : 239-240). C'est une mesure de la richesse lexicale (cf. ch. 5.3.). Une ILe élevée indique donc une grande originalité individuelle dans le texte. Cet indice semble pertinent car c'est une preuve quantitative aussi bien que qualitative du corpus. Il faut remarquer que nous avons calculé deux sortes de "mots uniques", à savoir, primo, les mots ont été calculés tels quels, c'est-à-dire, non-lemmatisés et/ ou avec des formes erronées – ceci a été fait pour avoir de l'information exacte sur les dissertations - et secundo, les "mots réellement uniques" d'après les consignes suivantes : les mots sont uniques a) s'ils ne font pas partie des formes lemmatisées - ceci a été fait pour pouvoir analyser le niveau de la sophistication lexicale des dissertations (voir au-dessous), b) s'ils sont uniques aussi après la correction d'une forme erronée (p. ex. *mirole* → *mioule* – c'est un mot unique ; *général* → *général* – ce mot n'est plus unique, car il y en a d'autres occurrences sous cette forme), et c) s'ils ne sont pas donnés aux consignes de la dissertation en question.

La **sophistication lexicale (SL)**, pour sa part, comprend ici le degré de difficulté¹⁰⁰. Elle est étroitement liée à l'individualité lexicale car selon des calculs, plus le texte est original, plus le niveau de sa sophistication est élevée (cf. Linnarud 1986 ; Mutta 1995, 1997). Ménard (1983 : 53) démontre qu'un texte de pauvre qualité est généralement caractérisé, entre autres, par "la prédominance des

¹⁰⁰ De toute façon, nous sommes consciente du fait que les mots rares ne sont pas toujours les mots les plus difficiles à apprendre, par exemple étant souvent plus concrets, et que les mots les plus fréquents ne sont pas toujours les plus faciles à apprendre, par exemple ayant souvent plusieurs sens. Voir aussi Bogaards (1994 : 151-152) pour les mots "faciles ou difficiles" à apprendre. Voir aussi chapitre 5.2.

vocables fondamentaux courants ou polysémiques”. De son côté, l’expérience d’Arnaud témoigne d’une corrélation étroite entre la richesse lexicale et le pourcentage des mots rares ; pour lui, les mots rares sont ceux qui ne figurent pas sur la liste faite pour la classe de troisième du Ministère de l’Education Nationale de 1976 (1984 : 15, 18, 23). Pour cet indice, Linnarud a eu recours à la méthode de Thorén¹⁰¹ où les items du vocabulaire étaient également listés d’après l’année scolaire où ces items devraient être introduits à l’école (Linnarud op. cit. : 45-46). La SL est égale, pour Harvey & King (cf. supra), au “quotient du nombre de verbes employés [...] divisé par le nombre de verbes non familiers (ou absent du *Français fondamental 1 ou 2*, suivant le niveau des candidats)” (cité dans Tréville & Duquette 1996 : 160). Il en va de même dans la présente étude où nous avons décidé d’utiliser la méthode employée dans une recherche antérieure portant sur l’anglais (Mutta 1995), avec quelques modifications¹⁰². La sophistication lexicale a été ainsi calculée manuellement en comparant le vocabulaire des apprenants, notamment les “mots réellement uniques”, avec le *Dictionnaire des fréquences* et le *Français fondamental* (qui est la base de l’enseignement du vocabulaire au lycée, cf. chapitre 5.1.). De plus, nous avons étudié les 20 mots les plus fréquents utilisés respectivement dans les deux corpus et les avons comparés avec ces ouvrages-là. Cossette (1994 : 47) nous rappelle encore qu’il faut faire des choix méthodologiques avant les calculs ; on peut donc travailler à partir d’index de vocables, c’est-à-dire de lemmes, ou d’index d’occurrences (cf. ch. 5.3.). Il lui-même a opté pour l’index de vocables. Ainsi ont également procédé les auteurs du *Français fondamental* et du *Dictionnaire des fréquences*, ce qui veut dire que nous avons utilisé ici également des formes lemmatisées, contrairement au reste de notre travail (cf. ch. 5.3.), pour pouvoir comparer notre corpus avec ces oeuvres. Dans la lemmatisation, nous avons adapté principalement la même pratique que dans le *Dictionnaire des fréquences* (1971) et que chez Lavonius (1998 : 14, 38)¹⁰³.

¹⁰¹ B. Thorén. 1976. *10000 ord för tio års engelska*. Lund : Liber.

¹⁰² Dans cette étude antérieure, les mots uniques étaient comparés, à la différence de Linnarud, au dictionnaire de fréquence de Holland & Johansson (1982).

¹⁰³ Entre autres, toutes les formes des verbes sont lemmatisées sous la forme de l’infinitif, les substantifs sous la forme du singulier, les adjectifs sous la forme masculine singulière. En revanche, par exemple, les articles ne sont pas lemmatisés. De plus, les homonymes sont divisés, quand c’est possible, dans différentes parties du discours.

7.3. Statistiques

Comme nous l'avons déjà mentionné supra, nous avons également calculé des corrélations entre les facteurs linguistiques divers afin de voir si les changements sont statistiquement significatifs. Pour calculer les corrélations entre les différents facteurs à chaque niveau, nous avons eu recours à la corrélation des valeurs, ou **coefficient de Bravais-Pearson (r)**¹⁰⁴ et à la corrélation des rangs, ou **coefficient de Spearman (ρ)** (Dugast 1980 ; Mialaret 1991). Nous avons utilisé un logiciel (SPSS)¹⁰⁵ pour effectuer les calculs. En outre, un t-test et un test ANOVA ont été effectués pour indiquer d'abord s'il y a une différence quant aux changements des facteurs linguistiques entre les deux niveaux dans tout le corpus et ensuite, s'il y a une différence dans les changements entre les deux corpus. Également, les corrélations entre les facteurs linguistiques dans tout le corpus ont été calculées séparément à ces deux niveaux ; on a eu recours au coefficient de Bravais-Pearson et au **coefficient de Kendall (*Kendall's tau-b*)**.¹⁰⁶

7.4. Mots en contexte

Cette partie de l'analyse consiste en une analyse qualitative. Cela veut dire que les vocabulaires des apprenants ont été examinés dans leur contexte linguistique et non comme items séparés. L'analyse des erreurs ne contient pas d'étude minutieuse des erreurs (cf. Corder 1975 : 166-171 ; 1980 : 24- 27), mais d'une façon générale la reconnaissance, la description et l'explication des formes idiosyncrasiques. Nous sommes allée plutôt dans le même sens que, entre autres, Péry-Woodley qui souligne l'importance d'une nouvelle tendance dans les analyses des erreurs, notamment l'évolution vers leur interprétation en "termes de pertinence et d'adéquation au contexte textuel et extra-textuel" (1993 : 20). Elle parle d'une évaluation globale des textes écrits où les erreurs en *langue* (par exemple l'orthographe) pourraient être abordées dans une perspective de *parole*,

¹⁰⁴ En anglais, *Pearson's product-moment correlation coefficient* (cf. Pitkänen & Kohonen 1984).

¹⁰⁵ SPSS (8.0 for Windows) est un logiciel qui peut calculer diverses formules statistiques, entre autres des corrélations.

¹⁰⁶ Nous remercions vivement Kalle Lertola, lecteur du département des statistiques, qui a effectué les calculs du t-test et du test ANOVA et a calculé les corrélations entre les différents facteurs dans le corpus total ; nous avons nous-même effectué les autres calculs.

notamment en jugeant leur impact sur la lisibilité.¹⁰⁷ Des recherches sur les phénomènes “discursifs” témoignent de cette approche (op. cit. : 30-31). Nous avons essayé, entre autres, à l’aide d’une étude sur l’utilisation des organisateurs pragmatiques (cf. ch. 6.3.), d’évaluer les dissertations des apprenants, notamment du point de vue de la lisibilité et de la clarté car une bonne cohérence textuelle contribue à la compréhension du texte écrit. Ce qui nous a intéressé également, comme nous l’avons mentionné, c’était la ponctuation qui joue aussi un rôle essentiel dans l’écriture en tant qu’organisateur structurel et cognitif. Selon Suomela-Salmi (1993 : 255) les marqueurs de structuration (par exemple typographiques) peuvent influencer beaucoup sur l’interprétation du message par le locuteur. Nous donnerons cependant seulement un aperçu général sur l’utilisation de la ponctuation des apprenants et cela également du point de vue du développement interlingual (cf. ch. 8.3.2.1.).

¹⁰⁷ Toutes les dissertations ont été évaluées par des professeurs français autochtones après l’examen. Leurs indications nous ont aidées dans l’interprétation des formes erronées, mais puisque les professeurs n’ont pas toujours été les mêmes, il y a des différences dans la façon de noter les corrections ; certains sont plus méticuleux que d’autres. Dans les cas délicats, nous avons recouru à l’aide d’un informateur natif. En outre, à cause de ces différences, nous n’avons pas pu directement tenir compte des notes mises dans les dissertations, comme nous l’avons fait dans une étude antérieure (cf. Mutta 1995).

8. ANALYSE DETAILLEE DU CORPUS

Nous voulons insister sur le fait que tous les résultats obtenus dans cette analyse sont étroitement liés à la méthode utilisée, notamment à l'utilisation du logiciel WordCruncher. Le recours à une autre méthode aurait probablement donné des résultats différents des nôtres à certains égards, mais dans l'ensemble ils se sont avérés identiques à ceux des recherches antérieures.

8.1. Analyse quantitative des deux corpus

Dans ce chapitre, nous traiterons les résultats des corpus dans leur ensemble, car c'est à partir de ces résultats que nous pouvons discerner des tendances générales. Dans la deuxième partie, notamment dans l'analyse qualitative, nous analyserons les dissertations des apprenants de manière plus détaillée.

8.1.1 Réflexions sur le nombre de mots

Le corpus principal consiste en 22.401 mots (ou occurrences) dont 11.069 pour le corpus des étudiants de première année (I) et 11.332 pour celui des étudiants de deuxième année (II) ; tandis que le groupe de contrôle consiste en 11.356 mots, distribués en 5.368 et 5.988 mots respectivement pour les différents niveaux. Ainsi, la totalité des mots de nos deux corpus est de 33.757 mots. Ces chiffres montrent que le nombre de mots augmente dans les dissertations au niveau II et cela dans les deux corpus. Les consignes données pour la rédaction des dissertations précisait que le nombre de mots devait être de 150 à 200, mais en examinant les moyennes des mots par dissertation, on peut remarquer que cette limite est quelque peu dépassée dans le corpus d'Aix (= **A**) (c'est-à-dire le corpus principal) - 209 et 214 mots respectivement – tandis que dans le corpus témoin (= **B**) (c'est-à-dire le groupe de contrôle) la limite est largement dépassée au niveau II - 215 et 240 mots respectivement. Le nombre de mots et leur répartition en mots pleins et en mots outils figurent dans les Tableaux 11 et 12 (Annexe 2). Bien entendu, le nombre de mots employés varie d'une copie à l'autre ; par exemple un des étudiant dépasse les limites imposées aux deux niveaux, 311 et 354 mots, tandis qu'un autre verbalise surtout au niveau II, 246 contre 477 mots. Dans le corpus A, c'est-à-dire le corpus des étudiants ayant séjourné en France dans le cadre d'un échange universitaire, le nombre de mots varie entre 139 et 311 mots

au niveau I et entre 151 et 354 au niveau II. Dans le corpus B, c'est-à-dire dans le corpus témoin, les chiffres vont de 159 à 300, et de 147 à 477 respectivement aux deux niveaux. Les Figures 5 & 6 permettent de visualiser les résultats obtenus pour les deux corpus.

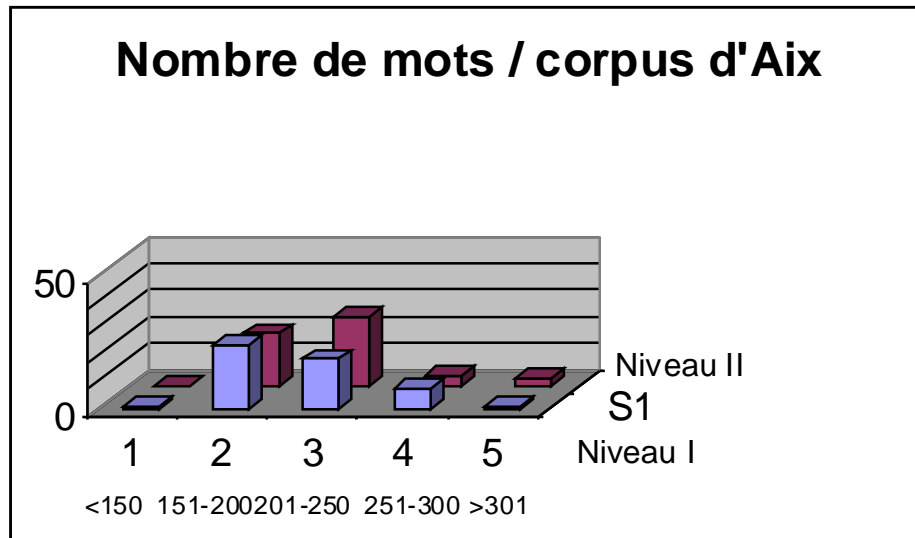


Figure 5. Nombre de mots dans le corpus d'Aix

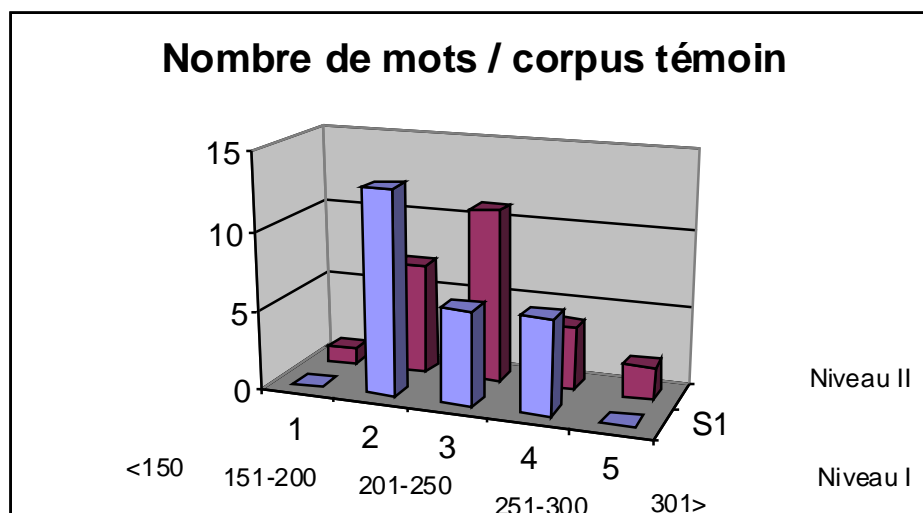


Figure 6. Nombre de mots dans le corpus témoin

Globalement, nous pourrions penser qu'au niveau I, le respect des consignes données est due aux habitudes d'écriture gravées dans l'esprit des apprenants depuis l'école. De plus, on pourrait supposer qu'au niveau II, les étudiants ayant passé un certain temps dans une université française ont été confrontés à des consignes assez strictes concernant certains travaux écrits et ainsi, qu'ils ont gardé cette habitude après leur séjour, contrairement aux apprenants du corpus B. Nous croyons que cette hypothèse n'est pas valable telle quelle, car tous les étudiants participent à l'enseignement dispensé au sein de notre département par les

professeurs autochtones qui imposent les mêmes consignes d'écriture. Selon nous, il s'agit donc plutôt d'une variabilité individuelle. En effet, dans le corpus d'Aix, 28 étudiants sur 53 (soit 53 %) ont moins de mots au niveau II qu'au niveau I. Le corpus témoin diffère du groupe d'Aix, car le nombre de mots augmente chez 16 étudiants sur 25 (64 %), résultats semblables à ceux constatés pour l'ensemble des corpus. De plus, il semble que la durée du séjour n'influence pas la longueur des dissertations ; la moitié des étudiants du corpus d'Aix dont le nombre de mots diminue après le séjour, ont passé un semestre à l'Université d'Aix, l'autre moitié deux semestres. Il en va de même pour les apprenants du corpus témoin ; ceux dont le nombre de mots diminue ont fait un séjour linguistique d'une durée d'un à onze mois dans un pays francophone. En outre, le fait d'étudier le français comme matière principale ou secondaire ne semble pas avoir eu d'effets sur les deux corpus.

Cette explication sur le nombre de mots dans les dissertations peut paraître inutile, mais en réalité, il a été prouvé que la longueur de la dissertation a une corrélation significative avec sa qualité générale (cf. Arnaud 1984 : 16). La recherche de Linnarud va également dans le même sens. Elle a trouvé que la longueur extrême d'une dissertation ne garantissait pas nécessairement l'obtention de la meilleure note à un examen, même si, à l'inverse, une dissertation brève tend à recevoir une mauvaise note (1986 : 116).¹⁰⁸ Regardons les statistiques. Ici, le coefficient de Bravais-Pearson (cf. ch. 7.3.) concernant la corrélation entre le nombre de mots au niveau I et II est d'une signifiante faible, à savoir $r = 0,258$ ($p \leq 0,05$) dans le corpus d'Aix, tandis que le coefficient de Spearman n'est statistiquement pas significatif. Cela indique qu'il existe bien une corrélation entre les deux niveaux, mais que la corrélation par rang, c'est-à-dire le rang que la variable (= le résultat de l'étudiant) occupe, n'a pas d'importance ici. Autrement dit, le nombre de mots entre les différents étudiants varie d'un niveau à l'autre, mais de manière aléatoire. Dans le corpus témoin, le résultat est inversé : le coefficient de Bravais-Pearson n'est statistiquement pas significatif, tandis que le coefficient de Spearman est d'une signifiante faible ($\rho = 0,369$, $p \leq 0,05$). Cela indique que dans le corpus témoin, ceux qui ont plus de mots ou moins de mots à l'intérieur de leur groupe au niveau I, ont une tendance, mais faible, de l'avoir aussi au niveau II.

¹⁰⁸ Nous sommes arrivés à la même conclusion dans notre étude antérieure portant sur l'anglais (1995 : 30).

Récapitulons brièvement les résultats : les taux moyens des nombres de mots augmentent au niveau II dans les deux corpus. Ce changement n'est statistiquement pas d'une grande signifiante, mais nous pouvons discerner cette tendance dans les deux corpus, en particulier dans le corpus témoin. Dans le corpus d'Aix, il s'agit plutôt d'une variabilité individuelle, car chez plus de la moitié des étudiants, le nombre de mots diminue respectant néanmoins les consignes données pour la rédaction. Cette diminution peut-être due à l'obéissance aux consignes données ce qui vient des habitudes adoptées à l'école ou des habitudes françaises adoptées en France ou dans notre département. Il faut remarquer que la longueur extrême d'une dissertation n'est nécessairement pas une fin en soi, car un texte long devient facilement répétitif (cf. infra).

Contrairement à notre étude antérieure (1995), nous n'avons pas comparé, dans la présente étude, la longueur d'un essai à la note donnée, car notre objectif ici est différent. Ce qui nous intéresse, c'est l'influence du nombre de mots sur les différents facteurs quantitatifs, notamment la variation, la densité, l'individualité et la sophistication lexicales ainsi que la corrélation entre ces phénomènes. Les facteurs quantitatifs calculés dans le corpus d'Aix et dans le corpus témoin sont présentés dans les Tableaux 13 et 14 (Annexe 3). Nous n'allons pas commenter ce tableau exhaustivement, mais à partir de quelques hypothèses portant sur les changements, nous présenterons les tendances et les écarts de ces tendances.

8.1.2. Indice lexical - vocable / occurrence

L'indice lexical d'une dissertation montre sa variation lexicale. Autrement dit, il indique le niveau de répétition des mots pleins véhiculant de l'information ; les mots outils, c'est-à-dire les mots grammaticaux, structurent quant à eux le message. Une variation lexicale (VL) élevée indique qu'il y a beaucoup de mots lexicaux différents. Elle n'est cependant pas une fin en soi car une VL modérée ne révèle que le caractère limité du vocabulaire, mais pas sa qualité, c'est-à-dire sa sophistication (cf. Tiilikainen & Ylinentalo 1989 : 133-134). De plus, il faut remarquer que la longueur de l'essai (c'est-à-dire le nombre total de mots) a un effet sur le pourcentage de la VL parce qu'il est plus difficile d'éviter la répétition dans un essai plus long et vice versa. Examinons d'abord le corpus d'Aix (= A).

La moyenne de la VL dans le corpus A s'améliore au niveau II, notamment de 79 % à 80% (cf. Annexe 3). L'augmentation ne semble pas être importante, mais si nous regardons de plus près les résultats des apprenants, nous pouvons remarquer que l'amélioration de la VL est un phénomène réel. Nous nous expliquons. Dans 21 essais sur 53, la VL diminue entre le niveau I et le niveau II. Comme nous l'avons mentionné supra, il est plus difficile d'éviter l'utilisation des mêmes mots dans un essai plus long. Nous avons montré qu'effectivement le nombre de mots augmente au niveau II. Dans le Tableau 4 nous voyons les écarts, c'est-à-dire les cas dans lesquels il y a une diminution de la VL dans le corpus. Le numéro de l'étudiant correspond à son numéro de référence dans l'ensemble du corpus A, voir Annexe 3.

corpus d'Aix				
étu- diant	I		II	
	mots	VL %	mots	VL %
2.	151	80,0	193	69,7
3.	189	81,7	225	79,4
10.	241	77,3	197	73,8
11.	195	84,3	211	82,0
16.	194	79,8	285	73,0
19.	219	88,5	244	86,8
20.	182	87,6	181	76,2
22.	203	92,9	172	78,4
26.	180	92,4	179	86,5
30.	170	72,0	228	71,7
31.	236	75,5	194	72,0
36.	212	88,5	209	64,4
37.	193	88,2	228	71,9
39.	181	80,8	219	80,2
41.	206	65,9	331	64,8
42.	200	89,3	244	71,6
43.	157	76,0	166	75,0
45.	200	76,8	368	70,8
47.	231	79,2	178	67,1
48.	228	87,1	208	82,8
49.	254	85,9	269	69,1

Tableau 4. Variation lexicale : extrait du corpus d'Aix

Il ressort du Tableau 4 concernant le corpus d'Aix, que dans 13 cas sur 21, le nombre des mots augmente, entraînant ainsi une diminution de la valeur de la VL et un prétendu appauvrissement du vocabulaire. Ceci est en partie illusoire, car en examinant les résultats de plus près, nous remarquons que chez les étudiants 3, 30, 39 et 41, par exemple, le pourcentage de la VL est presque de la même importance aux deux niveaux, même s'il y a beaucoup plus de mots au niveau II, par exemple chez l'étudiant 41 (+ 125 mots). Ainsi, chez ces étudiants, l'augmentation de la variation lexicale est véritable. Ce qui est intéressant ici, c'est que tous ces

étudiants énumérés supra n'ont passé qu'un semestre dans une université française et que trois d'entre eux sont des étudiants ayant une autre matière principale que le français. Comment expliquer ce phénomène chez ces apprenants ? Nous pourrions supposer que ces étudiants ont bien profité de leur séjour linguistique dans une université française et qu'ils ont bénéficié d'un input varié. Il semble que les étudiants dont le niveau de base était moins élevé au début profite davantage de leur séjour dans une université française (cf. Mutta 1999a). Le fait qu'ils ont participé à l'examen du niveau II peu après leur retour et que leur apprentissage a été étayé entre-temps par un enseignement dans notre département (autrement dit, il n'y a pas eu d'interruption de l'input favorable) a probablement eu un effet positif sur leur résultat. Cependant, l'étudiant 49 qui n'a que + 15 mots au niveau II, mais dont la VL diminue considérablement constitue un contre exemple. Cet étudiant semble régresser de manière significative dans son utilisation du vocabulaire, mais nous ne pouvons pas en tirer de véritables conclusions avant d'avoir examiné de plus près l'utilisation des mots uniques, c'est-à-dire la sophistication lexicale, et avant d'avoir réalisé l'analyse qualitative (cf. infra). C'est le cas également chez les huit autres étudiants (10, 20, 22, 26, 31, 36, 47 et 48) dont la variation lexicale diminue un peu ou radicalement lorsqu'ils ont moins ou autant de mots au niveau II qu'au niveau I. Dans ces huit cas, la durée du séjour en France ne semble pas jouer de rôle déterminant.

A l'instar de Tiilikainen & Ylimentalo (1989 : 133-134), on doit noter que la variation lexicale est étroitement lié au facteur de la VL/M qui indique la variation lexicale (calculée uniquement à partir des mots lexicaux) adaptée à la longueur de la dissertation (le nombre total de mots). Après l'arrondissement, la moyenne de ce facteur (voir Annexe 3) est la même aux deux niveaux : 0,39¹⁰⁹. Ce résultat va de pair avec l'augmentation du nombre de mots et de la VL. Nous supposons donc que lorsque la valeur de la VL augmente, celle de la VL/M augmente également, et vice versa. Cela est effectivement le cas dans notre corpus. Des écarts à cette tendance existent cependant : dans 14 cas sur 53 (voir Annexe 3). On peut diviser ces cas à part en trois groupes. **1)** Dans le premier groupe (les étudiants 5, 21, 27 et 40), la valeur de la VL/M est la seule qui s'écarte des tendances générales des autres facteurs et ceci dans le sens où la VL/M diminue alors que la VL s'accroît. Ceci est probablement aussi dû au fait que le nombre de mots s'accroît beaucoup en même temps ; par exemple chez l'apprenant 5, la VL

¹⁰⁹ Avant arrondissement, 0,3894 (I) et 0,3900 (II), donc il y a une très légère augmentation.

change de 75,7 % à 86,2 % même si le nombre de mots augmente de + 82. **2)** Dans le deuxième groupe (les étudiants 7, 14, 24, 28 et 29), la différence entre la VL/M au niveau I et II est moindre que dans le premier groupe ; cette valeur diminue quand la VL augmente. Cette fois-ci, la diminution est liée également à l'augmentation du nombre de mots. Mais le phénomène est modéré par l'augmentation de la densité lexicale (DL), c'est-à-dire le pourcentage des mots lexicaux de la dissertation. **3)** Dans le troisième groupe (les étudiants 10, 22, 31, 47 et 48), tous les résultats, notamment la VL, la VL/M et la DL, s'écartent des tendances générales. Cela veut dire que leur VL diminue au niveau II (cf. supra), et que par contre, la VL/M augmente au niveau II à cause du nombre de mots qui est inférieur à celui du niveau I. La DL s'écarte de la tendance, car elle augmente, peu certes, chez ces apprenants au niveau II (sauf chez l'étudiant 31 davantage) (cf. infra). Ni la matière principale, ni la durée du séjour n'ont eu tendance à influencer les résultats de la VL/M.

Quant au corpus témoin (= B), nous remarquons une différence dans les moyennes de la VL et de la VL/M par rapport à celles du corpus d'Aix. Dans les deux cas, les moyennes du niveau II sont inférieures à celles du niveau I, respectivement 80 % et 77 %, et 0,39 et 0,34. Ceci est clairement dû au fait que la moyenne du nombre de mots dans ce corpus est également plus élevée au niveau II. Voyons brièvement les écarts par rapport aux tendances obtenues pour le corpus A ; le corpus B est là pour vérifier les hypothèses du corpus A. Le Tableau 5 à la page suivante montre les cas à part ; les numéros des étudiants correspondent à leur numéro de référence dans l'ensemble du corpus B, voir Annexe 3.

Il ressort de ce tableau que 17 cas sur 25 s'écartent de la tendance du corpus A – à l'intérieur de ce corpus B cette tendance est néanmoins dominante – puisque la valeur de la VL diminue au niveau II. Dans douze cas, la diminution est due à l'augmentation du nombre de mots. Comme dans le corpus A, la détérioration de la VL n'est donc pas toujours la preuve d'une régression totale de la variété lexicale. Prenons comme exemple

corpus témoin				
étu- diant	I		II	
	mots	VL %	mots	VL %
1.	180	87,5	263	82,3
3.	159	87,3	180	75,3
4.	202	80,0	291	65,0
5.	228	79,6	246	74,8
7.	300	83,5	199	65,5
9.	290	75,4	274	74,0
10.	258	76,0	340	74,1
13.	268	91,6	298	79,4
14.	246	71,2	477	69,6
15.	187	86,3	179	76,0
16.	195	79,0	234	74,8
17.	179	81,8	214	80,5
18.	176	80,6	298	79,4
19.	163	78,5	242	71,3
21.	165	85,9	209	78,5
24.	200	86,2	147	77,6
25.	256	83,3	206	70,4

Tableau 5. Variation lexicale : extrait du corpus témoin

les étudiants 14, 17 et 18 dont la VL reste presque identique malgré l'augmentation du nombre de mots. Dans les cinq cas qui restent (les étudiants 7, 9, 15, 24 et 25), la VL s'affaiblit, même si le nombre de mots va en diminuant. Pour voir s'il s'agit d'une vraie régression, nous devons recourir à l'analyse qualitative (cf. infra) comme nous l'avons déjà mentionné en parlant du corpus A. Nous pouvons peut-être aussi mentionner que parmi les cinq derniers étudiants cités quatre sont des étudiants ayant une autre matière principale que le français et n'ayant fait qu'un séjour bref dans un pays francophone.¹¹⁰ Quant à la VL/M, il y a six cas sur 25 qui s'écartent de la tendance générale du corpus A (cf. Annexe 3). Dans deux de ces cas, cela est dû à l'augmentation du nombre de mots au niveau II (les étudiants 6 et 22), et dans deux autres cas le nombre des mots diminue et la VL aussi (les étudiants 9 et 24). Contrairement au corpus A, pour les cas parallèles, la DL diminue. Les deux derniers cas (les étudiants 7 et 25) correspondent au troisième groupe du corpus A, à savoir que tous les résultats, c'est-à-dire la VL, la VL/M et la DL, s'écartent des tendances générales. Cela veut dire que leur VL diminue au niveau II (cf. supra), mais que la VL/M augmente au niveau II à cause du nombre de mots qui est inférieur à celui du niveau I. La DL s'écarte de la tendance, car elle augmente chez ces apprenants au niveau II (cf. infra). La durée du séjour ne semble pas influencer les résultats.

¹¹⁰ Il faut remarquer que les connaissances dont nous disposons quant à leur séjour linguistique dans un pays francophone proviennent des rapports qu'ils ont écrits sur ce séjour. Nous ne disposons donc pas d'information sur leurs autres visites ou séjours possibles dans un pays francophone ou même sur leurs contacts personnels avec des locuteurs natifs. Cela est également le cas pour les étudiants du corpus d'Aix.

Le coefficient de Bravais-Pearson concernant la corrélation entre la VL au niveau I et II est significatif, mais faible, notamment $r = 0,252$ ($p \leq 0,05$) dans le corpus d'Aix, et c'est ainsi aussi quant au coefficient de Spearman, notamment $\rho = 0,239$ ($p \leq 0,05$). Ceci signifie qu'il existe une corrélation entre les deux niveaux, et que le rang des étudiants est faiblement le même à l'intérieur du corpus. Autrement dit, ceux dont la VL est plus ou moins élevée (à l'intérieur du corpus A) au niveau I, ont une faible chance d'être aussi au même rang au niveau II. Quant à la valeur VL/M, seulement le coefficient de Bravais-Pearson est significatif, mais faible, à savoir $r = 0,255$ ($p \leq 0,05$). Dans le corpus témoin, ni le coefficient de Bravais-Pearson, ni le coefficient de Spearman ne sont significatifs quant à la valeur de la VL. Par contre, la corrélation entre la VL/M au niveau I et II, est modérément significative dans le corpus témoin calculée par les deux coefficients, respectivement $r = 0,419$ ($p \leq 0,05$), et $\rho = 0,420$ ($p \leq 0,05$). En d'autres termes, nous pouvons constater une légère relation entre les valeurs de la VL et VL/M aux deux niveaux dans les deux corpus, mais nous ne pouvons pas, en nous basant sur ces chiffres, tirer de conclusions exactes sur le développement éventuel de ces valeurs. La seule valeur remarquable est la valeur VL/M du corpus témoin ce qui montre qu'il y a une relation étroite entre les deux niveaux. Ceci est probablement lié au nombre de mots qui augmente davantage dans le corpus B.

En somme, d'après les résultats obtenus, nous pouvons conclure que la variation lexicale s'améliore, c'est-à-dire que l'éventail de mots lexicaux utilisés dans une dissertation par un étudiant se diversifie après le séjour linguistique dans une université française ; dans le corpus témoin, l'amélioration est moins évidente, car la longueur des dissertations influe sur les résultats puisqu'il est plus difficile d'éviter la répétition dans une dissertation plus longue que courte. Récapitulons : dans le corpus d'Aix, il y avait 21 cas sur 53 contre 17 cas sur 25 dans le corpus témoin où la valeur de la VL diminuait, mais cette régression était en partie illusoire à cause de l'augmentation du nombre de mots. Chez certains étudiants, nous pouvons parler d'une véritable amélioration car leur VL augmente, bien que leur nombre de mots augmente nettement en même temps. La valeur de la VL/M (= la variation lexicale calculée à partir des mots lexicaux et divisée par le nombre total des mots) était étroitement liée à la valeur de la VL, à savoir lorsque la VL augmentait, la VL/M augmentait également et vice versa. Dans le corpus d'Aix, il y avait 14 cas sur 53 qui s'écartaient de cette tendance, dans le corpus témoin, il y

en avait 6 sur 25. Dans la plupart de ces cas, les écarts étaient dus à l'augmentation des nombres de mots ; dans le corpus d'Aix, chez certains étudiants, ce phénomène était modéré par l'augmentation de la densité lexicale ce qui veut dire que le nombre de mots lexicaux (par exemple des substantifs et des adjectifs) augmentait également. Dans un tiers de ces cas, les résultats s'écartaient des tendances générales. Nous reviendrons encore sur ces cas particuliers dans l'analyse qualitative afin de tirer des conclusions. Bref, l'amélioration chez les apprenants semble logique, surtout dans le corpus d'Aix, car les étudiants ont eu accès à un input linguistique varié aussi bien oralement, par exemple dans les contacts de tous les jours, que par écrit, par exemple lorsqu'ils ont lu des oeuvres littéraires ou scientifiques pour les cours universitaires qu'ils ont suivis. Ce résultat est positif, car une variation lexicale élevée est souhaitable chez les apprenants du point de vue de la cohérence du texte écrit (cf. Arnaud 1984 : 17 ; Charolles 1988 : 8). Néanmoins, il faut se souvenir que la VL montre si le vocabulaire d'un essai est plus ou moins varié, mais qu'elle ne révèle pas sa qualité, c'est-à-dire sa sophistication (voir infra ; cf. Tiilikainen & Ylimentalo 1989 : 133-134).

8.1.3. Pourcentage des mots lexicaux

La densité lexicale permet de savoir quelle est la quantité des mots lexicaux dans l'ensemble d'un texte. Nous avons remarqué (cf. ch. 7.2.3.) que dans un texte écrit par un autochtone, la quantité des mots grammaticaux dépasse normalement celle des mots lexicaux, bien que leur grande quantité soit supposée appauvrir le texte (cf. Ménard 1983 : 52). La limite de 40 % des mots lexicaux est censée correspondre au niveau des scripteurs natifs dans une dissertation, c'est-à-dire dans un texte non-interactif ; cependant, une fois que cette limite est dépassée, la croissance de la DL n'est pas nécessairement accompagnée de la réussite à l'examen (cf. Linnarud 1986 : 82). Linnarud postule par ailleurs que cette valeur n'est pas affectée par la longueur du texte (1986 : 47). Nous avons donc retenu cette limite de 40 % comme référence par rapport à laquelle nous avons comparé les résultats des apprenants. Dans le corpus d'Aix, la moyenne de la DL décroît légèrement, notamment de 42,8 % à 42,5, et il en va de même dans le groupe témoin, dans lequel la DL décroît de 41,1 % à 41,0 %. De ces moyennes, nous ne pouvons pas tirer de conclusions directes, même si le pourcentage est proche, pour

ainsi dire, de la limite de référence. Dans les Figures 7 et 8, les pourcentages de la DL sont regroupés en cinq classes.

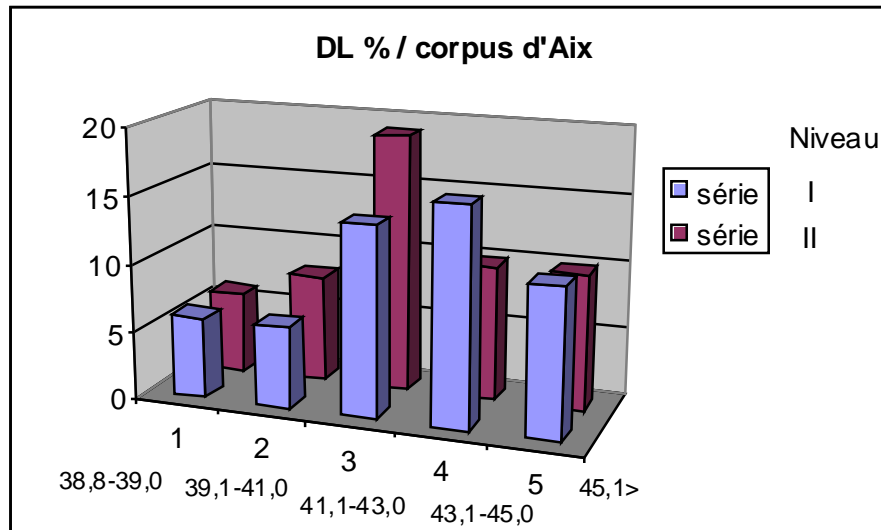


Figure 7. Densité lexicale dans le corpus d'Aix

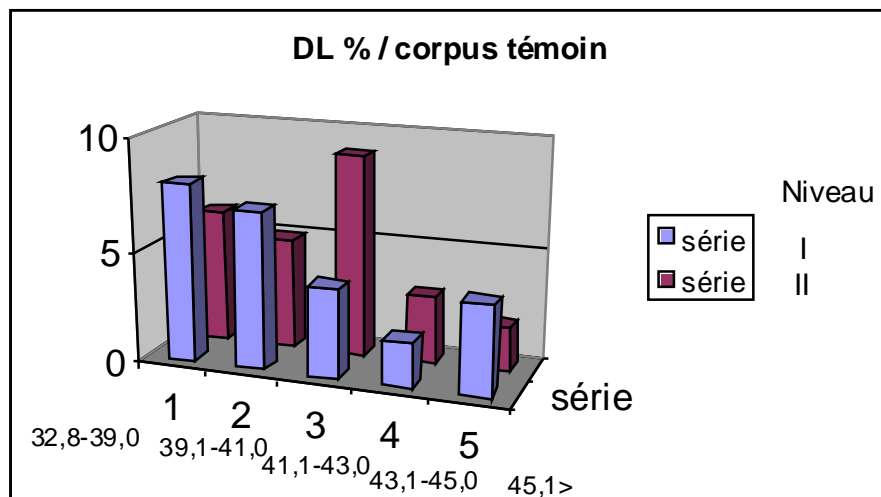


Figure 8. Densité lexicale dans le corpus témoin

Pour aller plus loin dans notre analyse de résultats, concentrons-nous sur les étudiants individuellement (voir Annexe 3). Comme le montrent les figures, et même si les pourcentages se rapprochent de la limite de 40 %, des écarts individuels sont visibles : les pourcentages minimums sont 33,8 au niveau I et 35,3 au niveau II dans le corpus d'Aix, et 36,0 et 32,8 dans le corpus témoin. Les pourcentages maximums sont respectivement 48,9 et 50,6 dans le corpus d'Aix, et 47,9 et 47,0 dans le corpus témoin.

Dans le corpus d'Aix, il y a onze étudiants dont le pourcentage des mots lexicaux est inférieur à la limite de 40 %, notamment les étudiants 4, 5, 6, 12, 13, 16, 27,

34, 40, 50 et 51. Ce sont les étudiants 13, 27 et 40 dont le pourcentage est largement au-dessous de la limite de 40 %, surtout l'étudiant 27 dont le pourcentage est de 35,3. Pour expliquer la raison de ce changement qui semble appauvrir le contenu du texte écrit par les étudiants (leur texte contient une trop grande quantité de mots grammaticaux), nous devons regarder leur texte de plus près. Ce que nous ferons dans la partie de l'analyse qualitative (cf. infra). C'est aussi le cas des étudiants dont le pourcentage est très élevé, et qui dévient donc de la tendance générale observée chez les natifs. Tendance qui rend leur texte moins naturel et lourd du point de vue de la lisibilité. Les étudiants 2, 14, 20, 23, 24, 31, 39, 43, 49 et 52 ont des pourcentages supérieurs à 45 % ; le cas de l'étudiant 49 est intéressant à examiner de plus près, car le pourcentage des mots lexicaux présents dans son texte est le seul à dépasser le pourcentage des mots grammaticaux au niveau II (50,6 %). De plus, nous pouvons ajouter que ni la durée du séjour ni la matière principale n'influent sur les résultats.

Dans le corpus témoin par contre, il y a huit étudiants dont le pourcentage de la DL est inférieur à la limite de 40 %, notamment 6, 9, 10, 11, 13, 19, 23 et 25, tandis que seulement trois étudiants, à savoir les étudiants 3, 20 et 24 ont dépassé les 45 %. Les cas les plus intéressants, comme dans le corpus d'Aix, sont les extrêmes, en particulier, les étudiants 13, 19, 20 et 23. Nous verrons leur essai d'une manière plus détaillée dans la partie de l'analyse qualitative avant de tirer des conclusions sur la richesse lexicale de leurs travaux. Ce qui diffère le plus du corpus d'Aix, c'est que sur les 11 étudiants, 6 ont eu une DL inférieure à 40 % aux deux niveaux, et un étudiant a eu une DL supérieure à 45 % aux niveaux I et II. Dans le corpus d'Aix, il n'y a aucun cas semblable pour les pourcentages inférieurs à 40 %, et seulement 4 étudiants sur 10 ont les pourcentages supérieurs à 45 % au niveaux I et II. Les corrélations quant à la DL ne sont statistiquement pas significatives, sauf dans le corpus témoin, la corrélation entre la DL au niveau I et II, est peu significative du point de vue du coefficient de Bravais-Pearson, notamment $r = 0,374$ ($p \leq 0,05$).

Bref, plus de la moitié des dissertations dans les deux corpus ont eu des valeurs de la DL qui se rapprochaient de la limite de 40 % observée chez les scripteurs natifs dans un texte non-interactif. Au niveau II, les taux moyens décroissaient légèrement dans les deux corpus. Dans le corpus d'Aix, il y avait 22 étudiants sur 53 contre 11 étudiants sur 25 dans le corpus témoin qui se sont écartés de cette

limite (soit inférieur à 40 % soit supérieur à 45 %). Les cas les plus intéressants sont les extrêmes, que nous analyserons plus en détail dans la partie qualitative. Si la valeur de la DL est moins élevée, cela signifie qu'il y a plus de mots grammaticaux dans le texte ce qui peut appauvrir le contenu du texte, car les mots lexicaux sont ceux qui portent principalement l'information. En revanche, si la valeur de la DL dépasse 50 %, ce n'est pas non plus une fin en soi, car de cette manière, le texte devient plus lourd (avec beaucoup de substantifs par exemple) ce qui gêne la lisibilité.

8.1.4. Originalité et difficulté lexicales

Dans la partie théorique (cf. ch. 7.2.4.), nous avons indiqué que d'après les résultats obtenus dans des recherches antérieures (cf. Linnarud 1986 ; Mutta 1995), les facteurs de l'individualité lexicale et de la sophistication lexicale sont étroitement liés l'un à l'autre, car plus le texte est original, plus le niveau de sa sophistication est élevé. Nous verrons maintenant si cette affirmation est également valable dans le corpus étudié dans la présente étude.

8.1.4.1. Individualité

L'originalité, c'est-à-dire, l'individualité du texte de l'apprenant a été mesurée en calculant le nombre de mots à occurrence unique chez un étudiant en comparaison avec tous les mots lexicaux de l'ensemble du corpus, à savoir les 106 dissertations du corpus d'Aix et les 50 dissertations du corpus témoin ; les corpus ont été traités à part. Pour obtenir le pourcentage de ces "mots uniques" (nous utilisons désormais ce terme pour des raisons de facilité), ceux-ci ont été comparés au nombre de mots lexicaux dans la dissertation de chaque apprenant. L'individualité lexicale est une mesure objective, mais laborieuse (cf. Linnarud 1986 ; Mutta 1995), d'où l'importance donnée à l'utilité du logiciel WordCruncher qui a facilité le dépouillement des mots. On pourrait critiquer le fait que les essais rédigés par les étudiants ont été collectés dans plusieurs examens de connaissances langagières entre 1985 et 1993 (cf. ch. 1.2.), ainsi que la diversité des sujets. En effet, comment peut-on comparer les vocabulaires de devoirs écrits alors que les examens sont différents et les sujets aussi ? En comptant sur nos connaissances préalables de l'examen et sur les sujets de dissertation, et en comptant également sur le fait qu'en règle générale, les

étudiants ont participé aux mêmes examens, et ont respecté les mêmes consignes au cours de ces années, nous en avons conclu que ce n'était pas un problème majeur, surtout en ce qui concernait le corpus d'Aix. En effet, les sujets de dissertation ont des thèmes assez généraux, comme par exemple *Les parents d'aujourd'hui sont-ils trop libéraux ?* ; *Quelle est l'originalité (...) de la Finlande par rapport aux autres pays ?* ; *Faites le compte rendu d'un livre que vous avez lu* ; *Les étrangers en Finlande* ; *Que faire pour sauver la Terre ?* ; *Le sida* (voir Annexe 1 aussi). Pourtant, lorsque nous parlerons de la sophistication lexicale, nous verrons combien le sujet influe en réalité sur l'utilisation des mots uniques (voir infra).

Etant intéressée par la langue des apprenants, nous avons voulu calculer deux sortes de mots uniques : 1) tous les mots uniques incluant, entre autres, les formes erronées et toutes les formes des verbes – ces mots-là sont comparables avec les autres facteurs calculés à partir d'index d'occurrences ; 2) les “mots réellement uniques” (nous les appelons ainsi pour des raisons de facilité) qui sont des mots excluant les formes lemmatisées et les mots donnés dans les consignes de rédaction des dissertations en question, ainsi que les mots qui ne sont plus uniques après la correction de la forme (cf. ch. 7.2.4.). Ces mots réellement uniques sont ainsi des hapax legomena (cf. Dugast 1980; Gustafsson 1990 ; Grönholm 1992) dont on ne relève qu'un exemple dans le corpus ; les deux corpus ont été traités séparément. En outre, ces mots réellement uniques sont étudiés de manière plus approfondie dans le chapitre suivant (8.1.4.2.).

Quand nous examinons les pourcentages de l'individualité lexicale (ILe), il ne faut pas oublier que la plupart du temps, il s'agit de quantités de mots peu importantes, même s'il y a, bien sûr, des différences individuelles. Autrement dit, l'ajout ou la réduction d'un mot se voit clairement dans les pourcentages, et ceci plus particulièrement lorsqu'il s'agit de mots réellement uniques. Il ne faut donc pas survaloriser les changements dans les pourcentages ; par exemple chez l'étudiant 28, un mot réellement unique correspond à 1,7 % (au niveau I), tandis que chez l'étudiant 43, trois mots réellement uniques correspondent à 4,0 % (I). Les moyennes de la valeur ILe concernant les mots uniques (a) sont en hausse dans le corpus d'Aix, notamment, 19,0 % contre 21,8 %, tandis que dans le corpus témoin, il y a une légère baisse, à savoir, 25,4 % contre 25,3 %.

Quant aux mots réellement uniques (b), il y a une hausse dans les deux corpus : respectivement 8,8 % et 11,2 % (corpus d'Aix) et 12,9 % et 13,3 % (corpus témoin). En ne regardant que les pourcentages, on pourrait supposer que les étudiants du corpus témoin utilisent un vocabulaire plus original, mais, à notre avis, cela est illusoire. Premièrement, la quantité d'apprenants, qui est plus modeste, et la diversité des sujets entrent en jeu ici. Les 25 étudiants ont été choisis d'une façon aléatoire, mais couvrent la même période d'examens que celle à laquelle ont participé les apprenants du corpus d'Aix; il y avait donc moins d'étudiants par examen (cf. ch. 1.2.). Deuxièmement, le nombre de mots chez les étudiants du corpus témoin est moyennement plus élevé, et on peut supposer que quand le nombre de mots augmente, le nombre de mots uniques augmente également, même si, d'un autre côté, la variation lexicale semble diminuer dans un essai plus long (cf. aussi ch. 8.1.4.2.). Troisièmement, comme nous l'avons déjà indiqué supra, il faut se souvenir qu'il s'agit principalement de petites quantités de mots, surtout en ce qui concerne les mots réellement uniques et ainsi il faut interpréter les chiffres avec prudence. Finalement, nous pouvons constater qu'en fait, dans le corpus d'Aix, les pourcentages augmentent plus que dans le corpus témoin. Les changements sont indiqués dans les Figures 9 et 10 (à la page suivante); les chiffres exacts sont présentés dans l'Annexe 3.

Nous ne pouvons pas dire grande chose sur l'utilisation des mots uniques ou réellement uniques dans les dissertations si l'on ne se base que sur les pourcentages. C'est pourquoi nous les examinons de plus près dans le chapitre 8.1.4.2; par exemple, combien les erreurs ou les fautes d'orthographe influent-elles sur les chiffres? Nous pouvons quand même postuler, comme nous l'avons déjà évoqué supra, que quand le nombre de mots augmente, le nombre de mots uniques augmente également, même si, d'un autre côté, la variation lexicale semble diminuer dans un essai plus long.

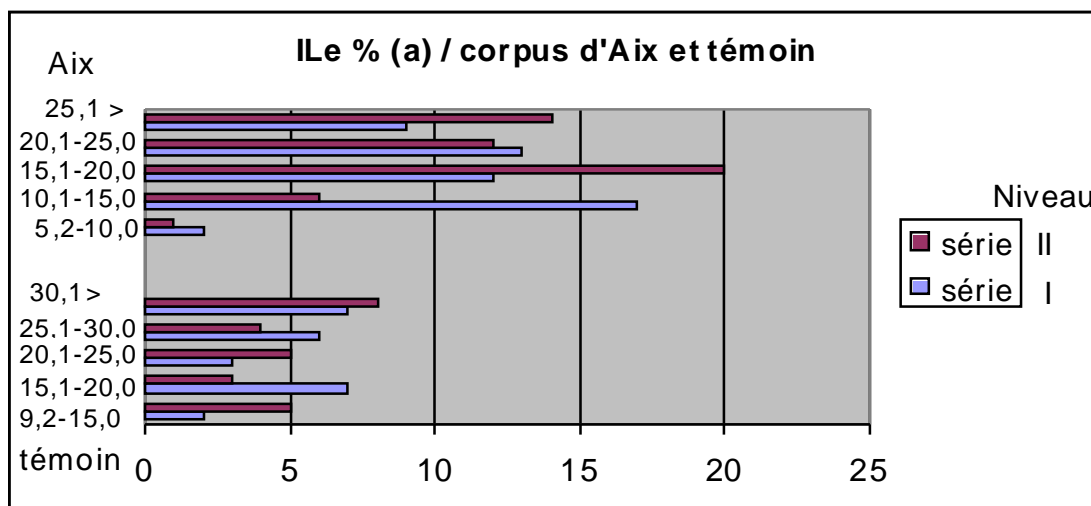


Figure 9. Individualité lexicale (a) dans les deux corpus

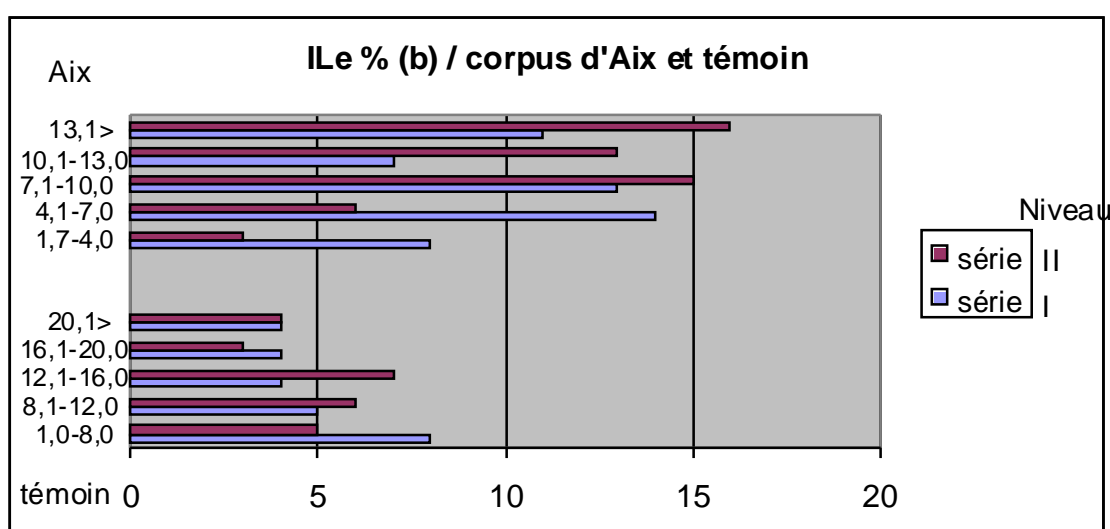


Figure 10. Individualité lexicale (b) dans les deux corpus

Quant aux corrélations, nous pouvons constater qu'il existe une corrélation modérée entre la ILe (a) au niveau I et II dans le corpus d'Aix, à savoir, les deux coefficients sont respectivement $r = 0,441$ ($p \leq 0,01$), et $\rho = 0,345$ ($p \leq 0,01$). Dans le corpus témoin, le coefficient de Bravais-Pearson montre une corrélation faible, $r = 0,401$ ($p \leq 0,05$), et le coefficient de Spearman le montre également, à savoir $\rho = 0,386$ ($p \leq 0,05$). Quant à la valeur ILe (b), le coefficient de Bravais-Pearson montre une corrélation modérée entre les niveaux I et II dans le corpus d'Aix, $r = 0,352$ ($p \leq 0,01$), tandis que les autres corrélations ne sont pas significatives. En somme, il faut interpréter les chiffres concernant les pourcentages des mots uniques et réellement uniques avec prudence. De toute façon, nous pouvons constater une hausse dans les pourcentages des mots réellement uniques dans les deux corpus ; ce sont les mots qui n'ont qu'une occurrence dans l'ensemble du corpus (les deux corpus ont été traités séparément) et dont on a enlevé les formes lemmatisées, entre autres. Au niveau I et II, en

particulier dans le corpus d'Aix, les mots uniques semblent avoir un lien étroit qui s'explique probablement en partie par le nombre de mots ; lorsque le nombre de mots augmente, le nombre de mots uniques augmente également, bien que la variation lexicale semble diminuer dans un essai plus long. Pour la qualité des mots uniques et des mots réellement uniques, nous allons à présent voir leur degré de difficulté.

8.1.4.2. Sophistication

La sophistication lexicale (SL) indique ici le degré de difficulté des dissertations. Dans la partie méthodologique, nous avons expliqué pourquoi la difficulté correspond dans cette étude à la sophistication (cf. ch. 7.2.4.), et comment nous avons effectué l'analyse. La SL n'est pas chiffrable en pourcentages comme les autres facteurs quantitatifs, même si elle est étroitement liée à la valeur de l'ILE. Aussi avons-nous effectué l'analyse manuellement. Nous avons fait deux sortes d'analyse. D'abord, nous avons comparé les 20 mots les plus fréquents avec le *Dictionnaire des fréquences* et le *Français fondamental*, et puis nous avons examiné les mots réellement uniques dans l'ensemble du corpus toujours par rapport à ces mêmes ouvrages.

8.1.4.2.1. Mots les plus fréquents

Commençons par les mots les plus fréquents dans nos deux corpus et dans le *Dictionnaire des fréquences* ; nous pouvons laisser de côté le *Français fondamental*, car il comprend automatiquement les mots les plus fréquents – la fréquence était un des critères principaux dans la collecte des mots. Les mots et leur fréquence dans les deux corpus sont présentés dans le Tableau 6 à la page suivante.

Il ressort du tableau que les onze premiers mots sont les mêmes dans les deux corpus et dans le *Dictionnaire des fréquences*, mais que l'ordre des mots n'est pas tout à fait identique. Dans les deux corpus, seul le premier mot occupe la même place. Dans le corpus d'Aix, les trois premiers mots correspondent aux trois premiers mots du *Dictionnaire*, même si leur ordre varie. A partir du rang 12, il y a deux mots (*on* et *pas*), dans le corpus d'Aix qui ne se trouvent pas parmi les 20 premiers mots du *Dictionnaire des fréquences*, et vice versa (*ne* et *se*). Dans le

corpus témoin, la variation est déjà plus importante ; quatre mots notamment (*pas*, *on*, *plus* et *mais*) sont exclus de la liste des 20 mots les plus fréquents du *Dictionnaire des fréquences*. Il y a trois mots du *Dictionnaire des fréquences* (*se*, *une* et *dans*) qui ne se trouvent pas parmi les 20 mots les plus fréquents du corpus témoin – il faut remarquer que plusieurs mots occupent le même rang dans le corpus témoin. Il s’ensuit de cette présentation que les deux corpus respectent assez bien les tendances générales concernant l’utilisation des mots fréquents, en particulier le corpus primaire. Nous pouvons donc conclure que plus il y a de mots

corpus d’Aix mot / fréquence		dictionnaire des fréquences	corpus témoin mot / fréquence	
1. de	1028	1. de (de + d’)	1. de	513
2. être	903	2. la	2. être	455
3. la	648	3. être	3. avoir	347
4. avoir	637	4. et	4. les	314
5. les	506	5. que (que + qu’)	5. la	279
6. le	448	6. le	6. et	254
et	448	7. à	7. que	249
8. que	430	8. l’	8. le	211
9. à	409	9. avoir	9. l’	202
10. il	356	10. les	10. à	173
11. l’	354	11. il	11. il	171
12. des	327	12. ne	12. des	160
13. en	306	13. je	13. pas	155
14. on	303	14. un	on	155
15. pas	285	15. se	15. je	144
16. un	249	16. des	en	144
17. qui	217	17. en	17. un	128
18. une	208	18. qui	18. plus	113
19. je	207	19. une	19. mais	105
20. dans	198	20. dans	20. qui	102
			j’	102
		23. pas	ne	102
		28. plus		
		30. on		
		34. mais		

Tableau 6. Mots les plus fréquents

dans le corpus, mieux les mots sont distribués de façon comparable à la tendance générale des textes français. Autrement dit, les dissertations des étudiants se rapprochent globalement des textes des scripteurs natifs quant à l’utilisation des mots les plus fréquents qui sont, comme nous pouvons le voir dans le Tableau 6, des mots grammaticaux.

8.1.4.2. Mots réellement uniques

Comme nous l’avons déjà indiqué auparavant, le corpus d’Aix consiste en 22.401 mots dont 3.514 mots différents (c’est-à-dire avec une fréquence = / > 1). Parmi

ces mots, 2.032 ont une occurrence unique (fréquence = 1) répartis comme suit : 62 mots grammaticaux et 1.970 mots lexicaux. De ces 1.970 mots lexicaux uniques, il nous reste, après avoir éliminé les formes lemmatisées et autres (cf. ch. 8.4.1.1.), **971** mots réellement uniques que nous examinerons de plus près dans le paragraphe suivant. Dans le corpus témoin, les nombres correspondants sont : 11.356 mots au total, 2.117 mots différents, 1.267 mots à occurrence unique (fréquence = 1), dont 79 mots grammaticaux et 1.188 mots lexicaux. Après élimination, il ne reste plus que **612** mots réellement uniques.

Les 971 mots du corpus d'Aix se divisent en deux niveaux, de la façon suivante : 432 mots au niveau I et 539 mots au niveau II, ce qui représente une moyenne de 8 mots et 10 mots, respectivement, par dissertation. Dans le corpus témoin, le nombre concernant les 612 mots réellement uniques est de 292 mots (au niveau I) et 320 (II), c'est-à-dire une moyenne de 12 et 13 mots respectivement par dissertation. Avant de comparer les mots réellement uniques avec le *Dictionnaire des fréquences* et le *Français fondamental*, regardons encore la distribution des mots lexicaux réellement uniques selon les différentes parties du discours (Tableau 7).

partie du discours	corpus d'Aix				corpus témoin			
	niveau I		niveau II		niveau I		niveau II	
substantifs	203	47 %	224	42 %	138	47 %	150	47 %
adjectifs	96	22 %	145	27 %	69	24 %	89	28 %
verbes	83	19 %	102	19 %	60	20 %	57	18 %
adverbes (-ment)	21	5 %	23	4 %	13	5 %	13	4 %
numéraux	8	2 %	10	2 %	3	1 %	5	1 %
noms propres	21	5 %	35	6 %	9	3 %	6	2 %
Total	432	100 %	539	100 %	292	100 %	320	100 %

Tableau 7. Division des mots lexicaux réellement uniques selon les parties du discours

Il ressort de ce tableau que la plupart des mots réellement uniques sont des substantifs dans les deux corpus et aux deux niveaux. Ce qui augmente le plus dans les deux corpus, c'est le nombre des adjectifs. Ce résultat va dans le même sens avec le postulat de Ménard (1983 : 52) qui dit que dans un texte riche, la fréquence relative des substantifs par rapports aux verbes, et celle des adjectifs par rapport aux substantifs sont en général plus élevées que dans un texte pauvre. Ainsi, on peut tout d'abord supposer que les textes des apprenants deviennent plus riches au niveau II, et secundo, qu'il est plus facile d'apprendre un substantif ou

un adjectif nouveau et de le mémoriser, que, par exemple, d'apprendre un nouveau verbe qui semble avoir de nombreux sens différents. Ceci n'est que partiellement vrai, car nous avons également compté comme mots réellement uniques les formes erronées, si elles demeuraient toujours uniques après correction ; il s'agissait principalement de fautes d'orthographe, ce qui indique que ces mots ne sont pas complètement acquis. Mais comme nous l'avons évoqué dans les méthodes, nous avons approché les erreurs et les fautes, à l'instar de Péry-Woodley (1993 : 20), d'une manière globale, notamment en jugeant leur impact sur la lisibilité et sur la compréhension. Ajoutons encore que les pourcentages des noms propres s'expliquent par la nature des sujets de dissertation ; par exemple *Liège, K. Rosberg, Salman Rushdie, Rudolf Steiner, Venice* (cf. Annexe 1). Tous les exemples sont transcrits dans la forme trouvée dans les dissertations.

Quand nous parlons de la sophistication lexicale, il faut remarquer que dans cette étude, la sophistication correspond au degré de difficulté ; la comparaison a été faite à partir de mots réellement uniques. Néanmoins, la quantité ou l'utilisation de ces mots ne suffisent pas pour indiquer toute la sophistication du texte. Pour examiner les mots réellement uniques et les comparer au *Dictionnaire des fréquences*, nous avons divisé les mots du dictionnaire en quatre catégories de fréquence d'occurrences : a) la fréquence entre $1 \leq 50$, b) $51 - 500$, c) $501 - 5000$, et d) > 5000 . Les mots des catégories a) et b) représentent donc un vocabulaire plus sophistiqué. En comparant les mots réellement uniques aux mots du *Français fondamental*, nous avons voulu voir si les mots que les étudiants utilisent sont appris dès l'école, au moins théoriquement (le *Français fondamental* étant la base de l'enseignement du vocabulaire au lycée, cf. ch. 5.1.). Nous avons examiné chaque mot réellement unique dans toutes les dissertations pour pouvoir tirer des conclusions sur les tendances générales. Mais pour donner des exemples du corpus, nous le ferons à partir des dissertations contenant un grand nombre de mots réellement uniques, ainsi qu'à partir des dissertations ayant nombre moindre de ces mots.

Nous avons postulé supra que l'individualité et la difficulté lexicales sont étroitement liées l'une à l'autre, c'est-à-dire que plus le texte est original, plus le niveau de sa sophistication est élevé. Dans la présente étude, cela se manifeste de la manière suivante : plus il y a de mots uniques dans une dissertation, moins il est

probable de les trouver tous dans le *Français fondamental* dont le contenu est principalement fondé sur la fréquence des mots. Il faut cependant se souvenir du fait que la première partie du *Français fondamental* contient uniquement des mots recueillis grâce à des enregistrements et que les listes de mots ainsi recueillis ont été modifiées en fonction de l'utilité de certains mots. Cela veut dire que le *Français fondamental* contient des mots dont la fréquence n'est pas toujours si élevée. Notre corpus contient en partie ces mots, par exemple *émission*, *géographique*, *réglement*. Par ailleurs, les étudiants utilisent quelquefois des mots moins fréquents qui ne figurent pas dans le *Français fondamental*, mais qui sont soit des mots de tous les jours, soit des mots "faciles"¹¹¹ pour nos étudiants, par exemple *chouette*, *pessimiste*, *prostituée*, *shopping* et *tartine*. Nous pouvons également déduire de notre analyse que plus il y a de mots réellement uniques appartenant aux catégories a) et b) dans le *Dictionnaire des fréquences*, moins il y a de mots qui se trouvent dans le *Français fondamental* ; il s'agit de mots qui en sont exclus comme *aléatoire*, *fracturée*, *hermétique*, *paralysé* et *refoulés*. Inversement, les mots apparaissant dans le *Français fondamental I & II* se situent le plus souvent dans les catégories c) et d), donc parmi les mots les plus fréquents ; par exemple *attention*, *dents*, *directeur*, *élever*, *neuve* et *soin*. De plus, nous pouvons noter que lorsque l'étudiant s'exprime avec un nombre élevé de mots réellement uniques, il utilise des mots réellement uniques même de la catégorie c) qui ne se trouvent pas dans le *Français fondamental I & II* (p. ex. *caresse*, *contact*, *continent*, *exclusif*, *fidélité*, *monotone* et *naïve*). Les mots utilisés par les étudiants qui ne figurent ni dans le *Français fondamental*, ni dans le *Dictionnaire des fréquences*, sont principalement des noms propres (p. ex. *Ajaccio*, *Alvar Aalto*, *Peugeot*, *San Diego* et *Salieri*), ou parfois des mots spécifiques, culturels comme *sauna*, ou encore des mots récents, comme *caméraman*, *postmoderne*, *recycler* et *supermarché*. Nous pouvons également ajouter qu'au niveau I, il y a plus de formes erronées exclues de l'analyse qu'au niveau II (autrement dit, ces mots-là n'étaient plus uniques après correction). Au niveau II, les formes exclues sont plutôt des formes lemmatisées et les formes erronées, étant encore uniques après correction, sont moins nombreuses.

En conclusion, il semble que de manière générale, la correction orthographique des mots réellement uniques s'améliore au niveau II. Il reste cependant des fautes

¹¹¹ C'est-à-dire des mots proches d'un mot qui existe en finnois ou dans une autre langue que les apprenants connaissent.

d'orthographe, mais ce sont principalement des fautes dans l'utilisation des accents, par exemple *enormes*, *ététernité*, *Génèse*, *negatif*, *prémier* et *représentent* (cf. ch. 8.3.1. sur les erreurs). Dans l'Annexe 4, il y a des exemples de dissertations avec peu ou beaucoup de mots réellement uniques et des exemples de dissertations où la quantité des mots réellement uniques augmente du niveau I au niveau II.

Nous pouvons en déduire que les apprenants connaissant plus de mots réellement uniques les ont appris en dehors du programme scolaire (*Les bases du programme scolaire pour le lycée* ; cf. ch. 5.2.) qui se base, au niveau du vocabulaire français, sur le *Français fondamental* ; cependant il faut remarquer que même aux épreuves de baccalauréat, les textes de compréhension écrite comprennent environ 25 % des mots des textes n'appartenant pas au *Français fondamental* (Lavonius 1998 : 26-29, 67-68). Comment expliquer les différences dans l'utilisation de mots réellement uniques ? Certes, il y a des variabilités individuelles : ceux qui mémorisent mieux les nouveaux mots, et ceux qui ne le peuvent pas. Mais, nous pouvons également supposer qu'il s'agit de différentes stratégies de communication. Certains étudiants prennent plus facilement des risques ; s'ils ne connaissent pas le mot exact, ils essayent néanmoins d'utiliser le mot qu'ils pensent connaître d'où le grand nombre de mots uniques sous forme erronée. En revanche, il y a des étudiants dont l'utilisation des mots réellement uniques est pour ainsi dire "sans surprise" ; ils utilisent seulement les mots qu'ils connaissent ou les mots plus simples avec un minimum de risques. Les exemples de l'Annexe 4 donnent un aperçu de ces stratégies.

En ce qui concerne le corpus témoin, nous pouvons constater que les phénomènes présents dans le corpus d'Aix se manifestent aussi dans le corpus témoin, mais que la moyenne du nombre de mots réellement uniques par dissertation étant plus élevée, il y a plus de mots qui se situent dans la catégorie c) de fréquence d'occurrences du *Dictionnaire des fréquences*, autrement dit, ils ont également des mots réellement uniques qui sont assez fréquents. Ceci est dû au fait que la quantité des apprenants à l'intérieur du corpus témoin est plus modeste que dans le corpus d'Aix mais ils couvrent la même période d'examens que celle à laquelle ont participé les apprenants du corpus d'Aix. Autrement dit, dans certains examens, il n'y a eu qu'un apprenant par examen et ainsi les différents sujets de

dissertation influent ici l'utilisation des mots qui peuvent devenir plus facilement uniques dans le corpus.

8.1.4.2.3. Cas particuliers

Dans la partie sur la variation lexicale (VL), nous avons évoqué la question de la régression apparente de l'utilisation du vocabulaire lexical chez quelques étudiants (cf. ch. 8.1.2.), en particulier chez les apprenants dont la VL diminue un peu ou radicalement au niveau II. Voyons maintenant leur dissertation du point de vue de l'utilisation des mots réellement uniques afin d'essayer de tirer quelques conclusions sur leur dissertation. Il s'agit des étudiants 10, 20, 22, 26, 31, 36, 47, 48 et 49 dans le corpus d'Aix, et des étudiants 7, 9, 15, 24 et 25 dans le corpus témoin (voir Annexe 3). Dans le corpus d'Aix, tous les étudiants possèdent moins de mots réellement uniques au niveau II qu'au niveau I. Les étudiants 10, 20, 22, 26, 31, 36 et 47 ont très peu ou assez peu de mots réellement uniques dans les deux niveaux et ces mots-là font principalement partie des catégories c) et d), donc des mots plus fréquents. Les étudiants 10, 22, 31 et 47 ont beaucoup moins de mots au niveau II et ainsi on pourrait supposer que la quantité de mots réellement uniques diminue automatiquement (cf. supra) ; les étudiants 20, 26 et 36 ont presque la même quantité de mots au niveau II, mais ils ont aussi moins de mots réellement uniques au niveau II. Nous voudrions supposer que la VL diminuée de ces étudiants va de pair avec un vocabulaire moins sophistiqué. Autrement dit, la régression semble encore plus évidente, mais il faut encore voir la partie qualitative proprement dite avant de tirer des conclusions définitives. Les étudiants 48 et 49 pour leur part, ont beaucoup de mots réellement uniques, même s'ils en ont moins au niveau II qu'au niveau I. Si on regarde de plus près les mots uniques de l'étudiant 48 – qui a, au total, 20 mots de moins au niveau II – nous pouvons remarquer qu'il y a une amélioration dans l'utilisation des mots réellement uniques qui n'est pas manifeste à travers le comptage des mots (21 et 14 respectivement) : au niveau I, il y a quatre noms propres et quatre formes erronées (comptées après correction) parmi les mots uniques, tandis qu'au niveau II, il y a deux noms propres et aucune forme erronée. L'étudiant 49 a au total 15 mots de plus au niveau II, tandis que ses mots réellement uniques diminuent (20 et 14 respectivement), même si l'on attendait le contraire. Chez cet étudiant, le niveau de sophistication ne change pas d'une manière significative (c'est-à-dire la distribution que nous avons faite dans les catégories du *Dictionnaire des*

fréquences). Ses résultats concernant les mots réellement uniques sont plutôt liés au fait que sa variation lexicale diminue radicalement, passant de 85,9 % à 69,1 %. En somme, chez lui, un vocabulaire plus répétitif est accompagné d'un nombre moins important de mots uniques qui restent néanmoins assez sophistiqués. Il utilise des mots comme *s'abstenir*, *flexible*, *mutisme* et *rigoureuse* qui ne figurent pas dans le *Français fondamental*.

Dans le corpus témoin, le total de mots des étudiants 7, 9, 15, 24 et 25 (voir Annexe 2) est moins important au niveau II qu'au niveau I. Leur VL devrait donc augmenter, mais comme nous l'avons vu supra (cf. ch. 8.1.2.), ce n'est pas le cas. Contre toute attente, le nombre de leurs mots réellement uniques diminue également entre le niveau I et le niveau II, sauf pour l'étudiant 24 qui a autant de mots réellement uniques aux deux niveaux malgré un nombre total de mots inférieur au niveau II (-53 mots). Les étudiants 15 et 25 possèdent une bonne quantité de mots uniques aux deux niveaux et la sophistication de leur dissertation reste presque identique. Pour les étudiants 15, 24 et 25, nous pouvons néanmoins dire que la VL en baisse est quand même équilibrée grâce à une sophistication lexicale ; en d'autres termes, bien que leur vocabulaire soit moins varié, il y a une "pincée" de sophistication, par exemple ils utilisent des mots comme *dépourvu*, *commodités*, *recession* et *urbanisation*. Pour sa part, l'étudiant 9 a légèrement moins de mots au total au niveau II, et sa VL et le nombre de ses mots réellement uniques décroissent un peu également. Nous pouvons remarquer une légère baisse dans la sophistication de son vocabulaire, c'est-à-dire qu'au niveau I, il a utilisé plus de mots correspondant aux catégories b) et c) (donc plus fréquents) non présents dans le *Français fondamental* qu'au niveau II. S'il s'agit d'une régression proprement dite, nous devons encore examiner sa dissertation de plus près (cf. ch. 8.3.2.3.1.). Finalement, chez l'étudiant 7, la régression semble encore plus évidente, car bien qu'il ait 101 mots de moins au niveau II qu'au I, il a une VL nettement détériorée, 83,5 % et 65,5 % respectivement. De plus, cette VL est également accompagnée d'un vocabulaire moins sophistiqué, à savoir 16 mots réellement uniques au niveau I et seulement 6 au niveau II. Au niveau II, ses mots réellement uniques correspondent principalement à la catégorie d), donc les mots les plus fréquents. Cependant, il faut encore examiner sa dissertation de plus près avant de conclure définitivement.

Récapitulons les résultats : les valeurs de l'individualité et de la sophistication lexicales sont étroitement liées l'une à l'autre. Dans cette étude, l'individualité correspond à l'originalité du texte qui est calculée à partir des mots à occurrence unique chez l'étudiant dans l'ensemble du corpus, tandis que la sophistication est le degré de difficulté calculé en le comparant au *Dictionnaire des fréquences* et au *Français fondamental*, référence de base de l'enseignement du vocabulaire au lycée. On pourrait également calculer la sophistication d'une autre manière, mais puisque la fréquence semble un bon indicateur des mots à apprendre (cf. Picoche 1989 : 91 ; voir aussi ch. 5.2. et 7.2.4.), nous avons fait ce choix méthodologique. Les mots les plus fréquents dans les deux corpus, en particulier dans le corpus d'Aix, sont distribués de façon comparable à la tendance générale des mots les plus fréquents utilisés par des scripteurs natifs ; ces mots-là sont des mots grammaticaux, par exemple des prépositions et des articles. Les mots réellement uniques des deux corpus sont principalement des substantifs et la part des adjectifs augmente le plus au niveau II. Ce résultat va de pair avec l'hypothèse de Ménard qu'un texte plus riche contient proportionnellement plus de substantifs et d'adjectifs que de verbes (1983 : 52) qu'un texte pauvre. Pour examiner les mots réellement uniques et les comparer au *Dictionnaire des fréquences*, nous avons divisé les mots du dictionnaire en quatre catégories de fréquence d'occurrences : a) la fréquence entre $1 \leq 50$, b) $51 - 500$, c) $501 - 5000$, et d) > 5000 . Les mots des catégories a) et b) représentent donc un vocabulaire plus sophistiqué. Nous pouvons conclure que plus il y a de mots réellement uniques dans une dissertation, moins il est probable de trouver tous ces mots dans le *Français fondamental*. En outre, il semble que les apprenants ayant plus de mots réellement uniques les ont appris en dehors du programme scolaire (*Les bases du programme scolaire pour le lycée* ; cf. ch. 5.2.) qui se base, quant au vocabulaire français, sur le *Français fondamental*. A notre avis, les différences entre les apprenants et entre les deux corpus dans l'utilisation des mots réellement uniques s'expliquent par des variabilités individuelles d'une part, et par l'utilisation de différentes stratégies de communication d'autre part ; certains étudiants prennent plus facilement des risques et il y en a d'autres pour qui l'utilisation de mots réellement uniques est pour ainsi dire "sans surprise".

8.2. Résultats statistiques

Dans la partie de l'analyse quantitative 8.1., nous avons déjà présenté quelques corrélations entre les différents facteurs linguistiques dans les deux corpus décrivant le changement entre les deux niveaux. Dans ce chapitre, nous présentons les résultats quant à l'ensemble du corpus, c'est-à-dire, le corpus d'Aix et le corpus témoin réunis.¹¹²

Avec le t-test, nous avons voulu montrer qu'il existe une différence concernant les facteurs linguistiques, notamment le nombre de mots, la variation lexicale (VL), la densité lexicale (DL) et l'individualité lexicale (ILe) a (= les mots uniques) & b (= les mots réellement uniques), entre les deux niveaux dans l'ensemble du corpus. Autrement dit, nous avons voulu voir si les changements du niveau I au niveau II sont significatifs. Les résultats du t-test montrent qu'il y a effectivement une différence statistiquement significative entre les deux niveaux concernant les valeurs du nombre de mots, de l'ILe (a) et de ILe (b). Les valeurs *t* sont pour l'ILe (a) 2,008 (au niveau $< 0,05$), pour l'ILe (b) 2,416 (au niveau $< 0,01$), et pour le nombre de mots 1,766 (au niveau $< 0,05$).¹¹³ Ces résultats nous permettent donc peut-être de conclure ce que nous avons déjà indiqué supra, à savoir que le nombre de mots augmente au niveau II - ce qui influence au moins partiellement les autres résultats, par exemple, la quantité de mots uniques qui augmente également.

Le test ANOVA, pour sa part, nous a permis de mettre en évidence les taux moyens pour les changements des différents facteurs quantitatifs au niveau I et II, ainsi que les écarts dans les deux corpus séparément. Les calculs entre les deux corpus ont également été effectués et les résultats obtenus nous permettent de conclure s'il y a ou non des différences significatives entre les deux corpus quant aux changements dans l'un et l'autre niveau.¹¹⁴ Le test ANOVA montre qu'il n'y a pas de différence significative entre les deux corpus concernant les changements de ces facteurs quantitatifs. La seule valeur où nous pouvons discerner une tendance statistique qui donne une idée de différence entre les corpus, même si

¹¹² Rappelons que ces statistiques ont été effectuées dans le département de statistiques par Kalle Lertola.

¹¹³ Ces différences réfèrent à une analyse "dans un sens", *1-tailed analysis* ; les résultats sont aussi significatifs dans une analyse "dans deux sens", *2-tailed analysis*, mais ici nous sommes intéressée par les changements du niveau I au niveau II.

¹¹⁴ Pour le test ANOVA, voir par exemple, Yli-Jokipii (1994 : 101-103).

non significative statistiquement, est la valeur VL/M, c'est-à-dire la variation lexicale en fonction du nombre de mots ($F = 3,370$, *sig.* 0,070). Nous avons déjà pu remarquer cette tendance quand nous avons examiné la variation lexicale ; cette différence est due au fait que la moyenne du nombre de mots est beaucoup plus élevée au niveau II dans le corpus témoin (215 mots et 240 mots respectivement aux deux niveaux) que dans le corpus d'Aix (209 et 214 respectivement). En bref, nous pouvons conclure que les changements dans les deux corpus sont presque parallèles. De plus, le test ANOVA a calculé s'il y a une différence dans les changements entre le niveau I et le niveau II qui puisse être liée au fait que le français est une matière principale ou non (dans le corpus d'Aix, 40 sur 53 sont des étudiants de la matière principale, tandis que dans le corpus témoin seulement 12 sur 25 le sont) ; ici, les différences ne sont pas statistiquement significatives. Nous avons également effectué le test ANOVA sur le corpus d'Aix afin de voir s'il y avait une différence dans les changements entre le niveau I et le niveau II parmi les étudiants qui ont passé un ou deux semestres à l'université d'Aix-en-Provence ; dans le corpus, 26 étudiants y ont passé un semestre et 27 y ont passé deux semestres. Le test ANOVA montre qu'il n'y a pas de différence statistiquement significative qui seraient liées à la durée du séjour.¹¹⁵

Examinons maintenant les corrélations entre les différents facteurs quantitatifs. Les corrélations sont calculées pour l'ensemble du corpus et puis séparément pour les deux niveaux. Les corrélations indiquent seulement s'il existe une corrélation entre les phénomènes en question, mais ne sont pas explicites concernant la cause et l'effet de ces phénomènes. Le coefficient de Bravais-Pearson indique une corrélation linéaire, tandis que le coefficient de Kendall indique une corrélation par rang – ce dernier coefficient est moins sensible aux écarts que le coefficient de Spearman. Les Tableaux 8 et 9 (à la page suivante) montrent les coefficients significatifs de Bravais-Pearson et de Kendall.

¹¹⁵ Nous n'avons pas pu calculer l'effet de la durée du séjour dans le corpus témoin, puisque les séjours étaient de durée variable.

	mots I/II	VL I/II	VL/M I/II	DL I/II	ILe a I/II	ILe b I/II
mots I/II						
VL I/II	/ -.322		.478 / .617		.563 / .528	.511 / .559
VL/M I/II	-.900 / -.882					/ .228
DL I/II						
ILe a I/II						.876 / .837
ILe b I/II						

Tableau 8. Coefficients significatifs de Bravais-Pearson aux deux niveaux

Il ressort de ce tableau que, tout d'abord, la densité lexicale n'a de corrélation significative avec aucun autre facteur aux deux niveaux. Au niveau I, la variation lexicale (VL) a une corrélation modérément significative avec les valeurs VL/M, l'ILe (a+b) ($p < 0,01$), tandis qu'au niveau II, la VL corrèle avec les valeurs VL/M (une corrélation considérable, $p < 0,01$) et l'ILe (a+b) (une corrélation modérée, $p < 0,01$), ainsi qu'avec le nombre de mots où la corrélation est modérée, mais négative ($p < 0,01$) – ce qui semble normal, car l'augmentation du nombre de mots est souvent accompagnée d'un vocabulaire moins varié. En effet, les corrélations significatives avec le nombre de mots sont toutes négatives ; en témoigne également la VL/M : une corrélation forte, mais négative, aux deux niveaux ($p < 0,01$). De plus, la VL/M corrèle faiblement avec l'ILe (b) au niveau II ($p < 0,05$). Les valeurs ILe (a) et l'ILe (b) corrélaient fortement aux deux niveaux ($p < 0,01$). En bref, la VL semble bien corrélérer avec les autres facteurs, et notamment avec les mots uniques et réellement uniques, comme nous l'avons déjà évoqué supra. Les mots réellement uniques (ILe (b)) corrélaient également bien avec les autres facteurs, sauf la densité lexicale et le nombre de mots. Ainsi, contrairement à notre hypothèse supra, et d'après ces statistiques, nous pouvons conclure que la quantité de mots réellement uniques est plutôt liée à la valeur VL qu'au nombre de mots.

	mots I/II	VL I/II	VL/M I/II	DL I/II	ILe a I/II	ILe b I/II
mots I/II						
VL I/II	/.182		.335 / .462		.405 / .336	.349 / .380
VL/M I/II	-.746 / -.740					/.165
DL I/II						
Ile a I/II						.695 / .584
Ile b I/II	.160 /					

Tableau 9. Coefficients significatifs de Kendall aux deux niveaux

Comme nous pouvons le constater, les résultats de la corrélation par rang vont dans le même sens que les corrélations linéaires. C'est encore la VL qui corrèle d'une manière significative avec les autres facteurs. Par contre, la DL n'a pas ici de corrélation significative avec les autres facteurs quantitatifs. La valeur ILe (b) a une corrélation plus ou moins forte avec les autres facteurs. La seule différence est que la valeur ILe (b) corrèle également, même si faiblement ($p < 0,05$), avec le nombre de mots au niveau I. Autrement dit, comme la corrélation n'est que faible, on peut en déduire que ces deux phénomènes, nombre de mots et nombre de mots réellement uniques, sont entrelacés mais de manière lâchée au niveau I.

Nous avons ainsi examiné la partie quantitative de notre analyse. Pourtant, même si cette partie est centrale dans notre étude, cette analyse seule ne suffit pas à révéler la richesse lexicale des dissertations. Entre autres, même si le test ANOVA a montré qu'il n'y a pas de différence significative entre les deux corpus quant aux changements de ces facteurs quantitatifs, nous ne pouvons pas en tirer de conclusions définitives avant la présentation de l'analyse qualitative, puisque ces résultats quantitatifs ne révèlent qu'un aspect des dissertations. Les résultats peuvent donc encore s'avérer différents.

8.3. Analyse qualitative : la langue des apprenants

Tous les apprenants du corpus étudient le français comme langue étrangère, et ainsi nous pouvons examiner leur utilisation de cette langue comme des exemples d'interlangues différentes, c'est-à-dire des langues d'apprenants. Dans une étude comme la nôtre, il est hors de question d'examiner chaque dissertation individuellement de manière exhaustive. C'est pourquoi, dans un premier temps, nous donnerons seulement un aperçu global des deux corpus, tandis que dans un

deuxième temps, nous présenterons quelques résultats de l'analyse qualitative dans les dissertations des apprenants auxquelles nous avons référé auparavant afin de tirer des conclusions. Tous les exemples que nous donnons sont retranscrits tels que nous les avons trouvés dans les dissertations.

8.3.1. Erreurs dans l'utilisation du vocabulaire

Dans la partie théorique de cette étude (cf. ch. 3.1.), nous avons expliqué ce que le concept d'erreur signifie. Les fautes et les lapsus sont des formes erronées qui peuvent être produites aussi bien par des apprenants d'une langue étrangère que par des locuteurs autochtones, voire natifs, tandis que les erreurs de compétence ne sont produites que par les apprenants d'une L2 (Corder 1975 : 24-25). Nous avons également souligné, à l'instar de Péry-Woodley (1993 : 30-31), l'importance d'une nouvelle tendance dans les analyses d'erreurs concernant les textes écrits où les erreurs en *langue* (par exemple les fautes d'orthographe) pourraient être abordées dans une perspective de *parole*, notamment en jugeant leur impact sur la lisibilité. Comme nous avons effectué une étude développementale, nous pouvons probablement faire quelques remarques sur la fossilisation, c'est-à-dire la stagnation, des formes lexicales que les apprenants utilisent. Dans une étude synchronique, cet aspect ne peut normalement pas être étudié.

8.3.1.1. Fautes orthographiques

Dans les deux corpus, les erreurs les plus évidentes, ou plutôt les formes erronées, sont les fautes d'orthographe, et surtout les fautes dans l'utilisation des accents. Les fautes d'orthographe se trouvent fréquemment dans des mots comme *americaine*, *amériquin*, *chômeurs*, *chommeurs*, *developpement*, *développement*, *néccéssaire*, *nécessaire*, *nécessaires*. Dans quelques cas, il s'agit de la fossilisation (dans le sens de Selinker) d'une forme erronée, c'est-à-dire que l'étudiant utilise tout le temps la forme erronée dans la dissertation, au niveau I comme au niveau II (p. ex. *obligé à mentir* – *obligés à penser* ; *verité* ; *semble d'être* – *semble d'avoir*) ; il peut être question d'une surgénéralisation d'une structure, p. ex. *égallement* – *finallement* – *spécialement* ; *égaals* – *idéals*. Il s'agit clairement de fossilisation quand l'étudiant utilise une forme incorrecte aux deux niveaux (*bien sur* I, *bien-sûr* II ; *on essaye aider* I, *essayer vivre* II ; *vraiment*). Il

peut cependant arriver que l'étudiant n'utilise la forme incorrecte qu'au niveau I (*par exemple I, par exemple II ; system I, système II ; reussir I, nous avons réussi II*) – dans ce cas-là, on peut parler de la défossilisation de la forme erronée au niveau II, donc du progrès dans l'utilisation du français. Ou bien, l'étudiant utilise la forme erronée seulement au niveau II (*presque I, presque II*) – dans ce cas-là, on parle de régression (*backsliding*), c'est-à-dire que la forme erronée apparaît même si elle semblait avoir été éradiquée de la production. Comment expliquer cette variation ? Il semble quelquefois qu'il s'agit de la variabilité libre, comme Ellis la définit (1992 : 75-76) ; l'apprenant utilise probablement la forme d'un mot sans avoir aucune idée de la manière dont il doit être écrit (p. ex. *problem I, problèms – problèmes II ; differents, différents II ; enfance, enfant, enfants II*). Cela peut être pour tester une hypothèse sur la langue, en particulier sur le vocabulaire (cf. Giacobbe 1990 : 119 ; 1992 : 61-64) afin de construire son système vers des emplois proches de la L2. Ou cela peut être aussi une sorte de stratégie communicative quand l'étudiant n'est pas sûr de la forme correcte ; pour citer l'un de nos étudiants : "J'écris le mot de différentes façons pour faire croire au professeur que je connais le mot, mais que par mégarde, je l'ai mal écrit."

Une catégorie des fautes d'orthographe s'explique par l'interférence, c'est-à-dire un transfert négatif. Il s'agit de l'influence de la L1 ou d'autres langues étrangères apprises auparavant (ou en même temps que le français comme matière principale ou secondaire) ; par exemple *Koran* (< finnois) ; *neutrale* (< finnois, anglais) ; *stil* (< suédois) ; *archipelago, s'apofiter* (< italien) ; *incroyable, vivir* (< espagnol). Pour la même raison, probablement, les apprenants hésitent souvent entre les majuscules et les minuscules concernant les adjectifs et les substantifs de nationalité ; entre autres, en finnois, le nom du pays s'écrit avec une majuscule, mais la nationalité avec une minuscule ce qui cause des problèmes dans l'utilisation des substantifs et des adjectifs en français (p. ex. *un peu fou à la Scandinave ; un auteur Anglaise ; beaucoup de finlandais*). L'interférence explique également l'utilisation de quelques mots qui sont généralement connus, mais qui s'écrivent différemment en différentes langues (p. ex. *Hannover* → *Hanovre* ; *Mouhamed* → *Mohammed, Muhammed*). Dans ce cas-là, l'erreur interlinguistique est quelquefois liée à une erreur intralinguistique (cf. Larsen-Freeman & Long 1992 : 58-59), c'est-à-dire que l'apprenant ne connaissant pas la forme exacte emprunte le mot d'une autre langue et en même temps surgénéralise ses connaissances de l'emploi du vocabulaire français et ainsi aboutit à une erreur

(p. ex. *Genesis* < finnois → *Génèse* ; *trauma* < finnois → *traumes*). Les erreurs de ce genre sont présentes aux deux niveaux et dans les deux corpus, et il est plutôt question de cas individuels.

Autre explication primordiale permettant d'expliquer l'origine de certaines fautes d'orthographe : ce qu'Ellis (1992 : 75-76) appelle la variabilité de performance et Vogel, la variabilité par rapport aux situations de production (1995 : 67-68). Cela veut dire que la situation d'examen elle-même rend les étudiants nerveux et donne lieu à des lapsus dans l'écriture. Dans le cas de certains étudiants, il s'agit pourtant plutôt d'étourderies, en particulier quand il y a beaucoup de fautes de ce genre aux deux niveaux. Quelques-unes de ces fautes peuvent aussi être considérées comme faisant partie de la correction grammaticale.

Ex. 1. niveau I :

[...] Elle aura un autre nom, une autre identité si elle reusit. Elle l'accepte parce qu'elle ne peut pas faire autrement. Puis après un certain temps de "l'éducation" en prison elle en sort. Elle comence sa vie en zero, except qu'elle est obligé de suivre des ordres qu'elle va revoir par l'Etat. [...]

Ex. 2. niveau II :

J'ai toujours été assez contante ici en Finlande. On dit souvent de mal à propos de la Finlande, mais si on reflechi vraiment on est obligé d'accepter le fait que la Finlande est un pays où on vit très bien. [...] Ils sourient et sont plus contents de leur vie qu'en hiver quand il fait froi. [...] Moi-même, je suis fier d'être finlandaise, mais cela ne m'epeche pas d'aimer les autres pays et j'ai toujours eu envie de voyager et de voir le monde.

Là, il s'agit d'une variabilité individuelle. Néanmoins, il est difficile de tracer une ligne absolue entre ces deux phénomènes à partir de deux dissertations écrites en situation d'examen seulement. Il faudrait suivre les productions de ces étudiants dans différentes sortes de tâches, par exemple une production libre ou guidée, un résumé fait à la maison avec des dictionnaires, etc.

Même si nous avons effectué une analyse longitudinale, il est difficile de déterminer s'il est question d'erreurs de compétence ou s'il s'agit seulement de fautes de performance. Pour cela, nous aurions dû faire une recherche métacommunicative pure dans laquelle il y aurait des enregistrements sur la réflexion du processus de l'écriture ou des commentaires des apprenants sous forme d'interview (cf. ch. 3.3. sur la compétence métacommunicative). A cause de la délimitation de notre travail, nous ne l'avons pas fait. De plus, tous les problèmes dans l'utilisation du vocabulaire ne sont pas évidents, car les étudiants utilisent des stratégies communicatives différentes afin de faire passer leur message. Comme nous l'avons déjà mentionné supra (ch. 3.2.), les apprenants

utilisent des paraphrases, des transferts et de l'évitement – pour reprendre les termes de Tarone (1980 : 429) - quand ils ont un problème de communication. Et l'évitement est particulièrement difficile à étudier car, n'étant pas explicite, il passe inaperçu.

8.3.1.2. Autres formes erronées

Pour présenter globalement les erreurs dans l'ensemble du corpus, et faire des remarques sur les différences entre les deux corpus à la fin, nous avons réparti les erreurs en deux catégories principales : les formes erronées et les erreurs de sens. Les formes erronées se chevauchent en partie avec les fautes d'orthographe. Voilà quelques exemples supplémentaires. Il arrive que l'étudiant conjugue mal les verbes (p. ex. *les écoles primaires et les collègues recevront [au présent] ; la vérité peux se présenter ; des fois j'oublis de vivre ; j'ai compris qu'il s'agitait) ; plus souvent, il est question de ne pas savoir si le verbe est pronominal ou pas, et cela arrive aux deux niveaux (p. ex. *qui se tue en essayant d'enfuire ; chacun doit se veiller aussi*) ou encore le régime du verbe est incorrecte (p. ex. *on a commencé à parler espagnol pour moi ; je voudrais me spécialiser par exemple aux langues*) ; ou encore, l'étudiant utilise l'auxiliaire incorrect au passé composé, ce qui est assez rare dans le corpus (p. ex. *j'ai allé la voir*). Il arrive même que l'étudiant écrive d'une manière incorrecte des formes du passé composé aux deux niveaux (p. ex. *il a reveiller au niveau I, on a donner au niveau II ; on les a "oublie", j'ai remarque au niveau I*) ; ceci est plus général dans les copies du corpus témoin. Nous pouvons également noter l'utilisation incorrecte du genre des mots ce qui se voit dans les formes liées aux substantifs, comme les adjectifs (p. ex. *les pays occidentales ; quelques événements plus intéressantes ; le façon le plus habituel*). Dans deux dissertations du corpus d'Aix, niveau I, nous trouvons des erreurs de forme assez rares dont l'origine reste un mystère : *A mon avis l'éducatons des garçons et des filles devraient être identiques ; C'est pourquoi je ne pense pas que l'éducatons des filles et celle des garçons devraient être différentes [...]*. Il y a également quelques fautes qui sont probablement dues à l'inattention, à savoir que l'étudiant écrit le mot sous une forme incorrecte, même s'il figure dans les consignes de la dissertation (p. ex. *les banlieux → banlieues ; exlure → exclure ; une façon différante → différente ; les verset sataniques → versets*) ; dans ce cas-là, il peut même arriver que l'étudiant utilise le mot tantôt d'une manière correcte, tantôt incorrecte dans la même dissertation. Néanmoins, il*

y a un contre-exemple : le mot figurant dans les consignes est incorrect, mais l'étudiant l'utilise sous la forme correcte : *l'anecdote*, dans les consignes *l'anecdote* ; on pourrait supposer que dans ce cas-là, l'étudiant a mémorisé le mot et n'est pas influencé par les déviations.

Nous trouvons en outre des formes erronées dues à la prononciation du français, c'est-à-dire que l'étudiant adapte ses connaissances sur le système phonétique français à l'écriture (p. ex. *invisible* → *invisibile*; *l'embience* → *l'ambiance* ; *on miole* → *miaule* ; *que la Finlande soie* → *soit* ; *dizaines de kilomètres* → *kilomètres* ; *efficase* → *efficace* ; *par hazard* → *hasard* ; *le Brésil* → *Brésil*) ; cela peut-être dû également à l'influence d'une autre langue (p. ex. *notre civilization* < anglais → *civilisation*). On peut dire que, dans le premier cas, il s'agit de la surgénéralisation des structures françaises, c'est-à-dire des erreurs intralinguistiques, et dans le deuxième, des erreurs interlinguistiques – parfois, c'est un mélange de ces deux influences. Il en va de même aussi pour quelques cas où l'étudiant, ne connaissant pas la forme exacte, invente un mot qui ressemble au système français (p. ex. *conservatif* → *conservateur* ; *decisseurs* → *décideurs* ; *les maladies venères* → *vénériennes* ; *beaucoup de choses, trop théorique* → *théoriques*). On peut également trouver quelques exemples de la surgénéralisation d'une forme qui devient une forme mixte dans les deux corpus, par exemple *chocquant* ← *le choc* et *choquer* ; *rechercher* ← *la recherche* et *chercher*.

8.3.13. Erreurs de sens

Les erreurs de sens sont plus difficiles à expliquer, mais normalement, il s'agit de problèmes de sélection lexicale (p. ex. *j'ai allé la voir beaucoup de temps* → *de fois*) liés à une connaissance imparfaite du mot en question – à moins que cela ne vienne d'un problème lié à l'acceptabilité du terme (cf. Enqvist 1977 : 1). Cela veut dire que l'apprenant connaît probablement la signification du terme, mais l'utilise dans une situation inappropriée (cf. dans le chapitre 5.3., l'exemple de **heavy tea*) ; voici un exemple tiré du corpus : *La promenade faite nous sommes rentrées à la voiture [...] – le verbe voulu serait plutôt retourner*. Voilà quelques exemples de plus (nous ne corrigeons pas toutes les formes erronées) :

à situation actuelle → à l'heure actuelle ; c'est d'immigrer dans un pays étrange → étranger ; qui peut par son attitude créer des catastrophes → déclencher ; il y a des problèmes sévères → sérieux ; l'école est plutôt visée à ceux qui [...] → destinée à ; c'est une peste de notre ère → notre époque ; la division entre les leçons de beaux-arts et de sciences → la division entre les cours artistiques et obligatoires ; pour battre contre le sida → lutter ; où le risque de transfusion du virus est plus grand → de transmission ; les jeunes ne sont plus du tout mûres → mûrs ; la plupart des garçons et des filles sont privés de leur meilleure façon d'apprendre → des meilleures conditions d'apprentissage ; il n'y a pas un tel écrivain qui n'aimerait pas de la publicité → il n'y a pas un seul écrivain qui n'aimerait pas la publicité ; tout d'un coup il est devenu une des auteurs les plus parlés dans le monde entier → célèbres ; à mon opinion → avis ; la production des armes plus efficaces → efficaces, puissants ; de suivre le ton de son cœur → la voix ; je vais rompre les obstacles → vaincre, affronter ; qui ont une plus grande risque d'avoir cette maladie → qui courent

Quelquefois, il arrive que la sélection lexicale empêche la compréhension du message écrit, ou qu'elle rende le sens du message laborieux à déduire à partir du texte. Cela peut être dû au fait que l'apprenant a traduit la forme directement du finnois ou d'une autre langue, souvent de l'anglais, par exemple *j'ai entendu dire que le rire cure mieux que [...] → guérit*. Autres exemples :

Ex. 3. :

L'absurdité du monde nous a attrapé, on s'y attend seulement. [...] Les conséquences dont l'imam Khomeini est coupable sont les menaces de mort. L'imam Khomeini a dit que chaque Musulman a le droit de le tuer en plein aire.

Ex. 4. :

En fait, la Finlande voudrait être une partie internationale d'Europe mais bien à part et entre les Finlandais, pour qu'elle puisse prendre ses décisions à elle et commettre ses déractions à elle tout tranquillement.

Ex. 5. :

Bruce Willis est l'acteur que je voudrais rencontrer définitivement si j'avais la possibilité. Certainement il est différent de la caractère dans la série mais comment il est chez soi ; cela est la devi nette que j'aimerais résoudre.

Ex. 6. :

Car, bien qu'on sache qu'il ne faudrait pas pas conduire la voiture après avoir eu du vin, de la bière ou de n'importe quel genre d'alcool, on le fait trop souvent. A mon avis, personnellement, les chauffeur et les chauffeuses, qui ont causé des morts [...]. Il y a une règle simple, mais toujours aussi juste : si on a de l'alcool, on ne conduit pas et si on conduit, on n'en a pas.

Mais il y a aussi des exemples où, même si la sélection lexicale n'est pas correcte, le sens de la phrase est compréhensible grâce au contexte - dans l'exemple 7, l'influence de l'italien est visible, tandis que dans l'exemple 8, c'est celle de l'anglais qui intervient.

Ex. 7. :

Ils ont peur, les pauvres finlandaises et c'est ça que les impedit de comprendre que la richesse d'une culture se manifeste dans sa variété, dans ses différents couleurs. [...] A mon avis la Finlande peut bien augmenter le numéro des réfugiés, les problèmes qu'ils rencontreront ici ne seront jamais [...].

Ex. 8 :

A mon avis, ce serais vraiment pitoyable à si grossièrement sous-estimer les possibilités de notre civilization nordique. En effet, il s'agirait d'un suicide, ou plus précisément notre influence et existence dans le monde seraient nullifiées complètement.

L'exemple suivant a semblé incompréhensible au correcteur français, mais étant de la même langue maternelle et de la même culture que l'apprenant, nous avons pu déchiffrer la phrase dans laquelle l'apprenant utilise une paraphrase pour parler d'une poubelle de recyclage du papier :

Ex. 9. :

Pourtant, j'admets qu'en approchant des grandes villes le paysage entouré est bien décevant et triste à regarder. Des énormes poubelles de papiers d'occasion, des voitures cassées et une odeur puante qui donne l'envie à vomir et à s'enfuir plutôt possible.

C'est le cas également avec l'extrait suivant où l'apprenant utilise une image facilement compréhensible en finnois qui s'exprime différemment en français, par exemple *on doit rester à sa place même si on est prêt à éclater de rire* :

Ex. 10. :

Il y a aussi des situations où on ne peut pas rire en haute voix sans faire un scandale. Et c'est vraiment pénible, car on doit trembler sur sa place en souffrant de l'attaque de fou rire.

Pour expliciter ces différences entre les deux langues et les deux cultures, on pourrait recourir à une sensibilisation à la communication exolingue et interculturelle, en termes de Porquier (1984 : 44).

Les extraits suivants illustrent deux erreurs typiques de l'utilisation des adverbes dues à un transfert de la langue maternelle ; ce sont des dissertations du niveau I :

Ex. 11. :

Je trouve que dans notre pays les créateurs de la mode des grandes entreprises ont assez beaucoup de pouvoir dans un certain domaine.

Ex. 12. :

N'importe qui peut s'inscrire à l'hôpital quand il se sent malade et on va lui soigner très bien. Et ce qui est important, cela ne coûte pas assez cher.

Nous donnons deux derniers exemples sur la sélection lexicale. Dans le premier, l'étudiant a voulu utiliser une expression française, mais il ne s'est pas souvenu de la forme correcte ; dans le deuxième, l'étudiant a réussi à utiliser l'expression toute faite d'une manière métaphorique dans un contexte donné (avis de notre informateur natif) :

Ex. 13. :

Mon ami, je l'aime parce qu'elle est la personne la plus proche de moi. Elle m'aide, elle me donne des conseils mais aussi elle me dit mes quatre cents vérités quand c'est le temps. [au lieu de mes quatre vérités]

Ex. 14. :

La vie bat son plein quand on est jeune. Mais peut-être n'arrive-t-on pas vivre sa jeunesse comme prévu. La société nous demande de réussir, de pouvoir montrer notre capacité, d'être actif...

Dans aucun des deux corpus, nous n'avons trouvé d'exemples où l'étudiant aurait inventé un mot qui ne soit pas reconnaissable, c'est-à-dire dont le sens présumé puisse être déduit du contexte. Par exemple quand l'étudiant écrit : *pour dissoudre le problème*, nous savons que le terme voulu est *résoudre* ; en effet, sachant que l'étudiant a étudié le latin, nous pouvons penser qu'il a surgénéralisé l'emploi des préfixes français ; ou par exemple quand l'étudiant parle des *traiteurs du pays* au lieu des *traîtres* (< anglais *traitor*). Ce qui est plus problématique, c'est l'utilisation incorrecte du vocabulaire associée à l'emploi d'une structure erronée car cela rend quelquefois la production incompréhensible. Dans les dissertations, il y aurait beaucoup de choses intéressantes à examiner au niveau de la structure des phrases, mais nous ne pouvons pas les traiter dans le cadre de cette étude. Comme nous l'avons déjà mentionné supra, nous avons voulu analyser des erreurs, à l'instar de Péry-Woodley (1993 : 30-31), dans une perspective de *parole*, notamment en évaluant leur effet sur la lisibilité. Dans la partie théorique (cf. ch. 5.1.), nous avons présumé que les erreurs concernant la sélection lexicale sont considérées plus gênantes quant à la compréhension que les erreurs grammaticales (Carter 1987: 145). Nous pensons qu'il en est sûrement ainsi à l'oral mais qu'à l'écrit, même les fautes d'orthographe peuvent avoir un effet négatif sur la compréhension et sur les réactions du lecteur (nous avons parlé d'une grammaire, soit de la norme lexicale au chapitre 4.2). Nous nous expliquons. S'il s'agit seulement de lire la production écrite des apprenants, et d'évaluer la lisibilité et la compréhension générale du texte, les fautes et les erreurs de formes semblent moins importantes que la sélection lexicale. Mais s'il s'agit d'examiner et d'analyser la production en détail, par exemple afin de faire des commentaires précis aux étudiants quant à leur production, alors même les petites fautes peuvent gêner la lecture et avoir un effet négatif sur l'évaluation. C'est ce que Linnarud (1986 : 83) conclut également dans sa recherche. Elle postule qu'un grand nombre d'erreurs dans une dissertation courte a un effet néfaste quant à l'évaluation des essais, ce qui semble tout à fait logique ; nous sommes arrivés à la même conclusion dans un travail précédent (Mutta 1995). Néanmoins, notons que le plus pertinent semble de toute façon être le contenu du texte écrit ; en effet, si l'étudiant a de bonnes idées et qu'il argumente bien ses points de vue, les erreurs ont alors une influence moindre sur l'évaluation.

En ce qui concerne les erreurs, nous n'avons pas remarqué de grandes différences entre les deux corpus. Nous pensons qu'il s'agit plutôt de variabilités individuelles, c'est-à-dire que si l'étudiant est étourdi au niveau I, il l'est probablement au niveau II également. Cependant, nous avons pu remarquer un progrès entre le niveau I et le niveau II en général, à savoir que la correction lexicale est meilleure, malgré les fautes d'orthographe (en particulier les accents), et que les notes données par les professeurs natifs sont normalement meilleures aussi après le séjour à l'étranger. A notre avis, ceci est étroitement lié au progrès dans l'utilisation des organisateurs pragmatiques (cf. ch. 6.3.), car une bonne cohérence textuelle augmente la compréhension et la lisibilité du texte écrit.

8.3.2. Organismes pragmatiques du texte

Nous avons indiqué dans la partie théorique (cf. ch. 6.3. et 7.4.) que ce qui nous intéresse dans cette étude du point de vue de l'organisation textuelle, ce sont les organisateurs pragmatiques du texte qui comprennent, primo, la segmentation : le découpage en paragraphes et l'utilisation de la ponctuation - qui jouent un rôle essentiel à l'écrit en tant qu'organismes structurels et cognitifs ; et secundo, les marqueurs visibles interphrastiques dans le texte. Comme nous l'avons déjà mentionné supra, nous donnerons seulement un aperçu général de la segmentation et cela également du point de vue du développement interlingual.

8.3.2.1. Remarques sur le découpage en paragraphes et sur la ponctuation

Nous aurions pu calculer le nombre de phrases et de paragraphes (cf. Mutta 1995) ou le nombre de mots par une phrase (cf. Linnarud 1986), mais nous ne l'avons pas fait. A notre avis, dans cette étude, cela n'aurait probablement pas apporté de renseignements pertinents sur l'évolution de la production écrite.¹¹⁶ Ce que nous avons pu constater d'emblée dans les deux corpus, c'est que chaque étudiant organise son texte en plusieurs paragraphes. En d'autres termes, il est extrêmement rare que l'étudiant n'écrive son texte sans paragraphes. Ainsi, nous pourrions dire que les apprenants savent utiliser le découpage du texte en

¹¹⁶ Le nombre de mots par phrase peut procurer des données intéressantes sur la qualité des dissertations des apprenants. Nous avons pu prouver dans une étude antérieure, certes sur l'anglais, que l'on a plus de chance d'être bien noté si l'on organise son texte en plusieurs paragraphes et que l'on a un certain nombre de phrases (cf. Mutta 1995). Néanmoins, il faut se souvenir du fait que le nombre de phrases n'est pas une valeur en soi, car une grande quantité de phrases peut rendre le texte trop fragmenté.

paragraphes ce qui, d'après Bessonnat (1988: 94), est une activité métalinguistique qui facilite la lecture et organise le texte hiérarchiquement. En outre, nous pouvons ajouter après avoir observé notre corpus que, plus le texte est long, plus il y a de paragraphes, ce qui semble logique. Néanmoins, il y a parfois des cas où l'étudiant semble avoir plus de paragraphes que nécessaire si l'on considère le nombre total des mots de son texte. Dans ces cas-là, l'étudiant a écrit une ou deux phrases isolées qui servent à introduire le sujet ou conclure le texte, ou bien les deux ; donc au début ou à la fin du texte.

Ex. 15. :

Si on compare le nombre d'étrangers en Finlande avec ce des autres pays nordiques par. ex la Suède et le Danemark, on peut bien voir que le nombre est petit. A mon avis, il n'y en a pas trop.

[...]

Je ne connais pas personnellement des étrangers en Finlande mais au contraire j'ai des amis en étranger.

Chez certains étudiants, cette façon d'organiser le texte correspond soit à une stratégie soit à un style personnel, car aux deux niveaux, l'étudiant termine sa dissertation par une phrase isolée.

Ex. 16. niveau I :

Donnez la possibilité aux étrangers !

Ex. 17. niveau II :

Le rire fait vivre plus longtemps.

Dans notre analyse de l'emploi de la ponctuation, nous avons laissé de côté les virgules que nous considérons moins importantes que les autres signes de ponctuation quant à l'organisation macro-structurale du texte (cf. ch. 6.3.). Pourtant, nous pouvons peut-être mentionner qu'il y a des erreurs dans l'utilisation des virgules qui sont dues à l'influence de la langue maternelle, le finnois ; par exemple quand l'étudiant met une virgule devant les mots "que" et "qui" – ce type d'emploi des virgules est plus fréquent au niveau I :

Ex. 18. :

Bien sûr la Finlande ne peut pas accueillir tous les réfugiés et il faut choisir ceux, qui ont le meilleur besoin.

Ex. 19. :

On constate souvent, que la Finlande est le plus cher pays du Monde.

Pour les autres signes de ponctuation, nous pouvons conclure que dans les deux corpus, la plupart des étudiants utilisent un éventail de signes de ponctuation dès le niveau I (dans le corpus d'Aix 41 sur 53, et dans le corpus témoin, 18 sur 25), à savoir, des point-virgules, des tirets et des points d'exclamation, etc. (cf. infra). Dans le corpus d'Aix, il y a 7 étudiants sur 53 et 4 étudiants sur 25 dans le corpus

témoin qui continuent à ponctuer leur texte avec des points uniquement. Cela ne veut pas dire que leurs essais sont moins organisés que ceux des autres étudiants, mais qu'ils utilisent d'autres moyens pour organiser leur texte (cf. infra). Nous pouvons donc constater que l'utilisation des signes de ponctuation est déjà stabilisée au niveau I, ce qui semble normal, puisqu'il s'agit d'apprenants adultes - nous ne commentons pas la correction de l'utilisation des signes de ponctuation. Nous pouvons néanmoins remarquer un léger changement chez certains étudiants. Au niveau II, dans le corpus d'Aix, 15 étudiants sur 53, contre 11 étudiants sur 25 dans le corpus témoin, ponctuent leur texte par d'autres signes de ponctuation que le point (unique ponctuation présente au niveau I), ou qu'ils utilisent un éventail plus large de signes qu'au niveau I. Nous ne pouvons cependant pas savoir si ce changement est dû à une réflexion métalinguistique sur l'organisation du texte ou à une autre cause.

Pourtant, dans trois cas (deux dans le corpus d'Aix et un dans le corpus témoin), nous savons qu'il s'agit d'un acte délibéré dû à une réflexion consciente pour ainsi dire. Ce n'est pas une question de règle d'utilisation de la ponctuation, mais plutôt, une question de style ou d'influence culturelle. Les deux premiers étudiants ont passé un ou deux semestres dans une université française, le troisième a passé une année à Lausanne et a participé à des cours de langues pour les étrangers. Après leur séjour à l'étranger, ces apprenants utilisent un stylo à encre pour écrire leur dissertation. Cette observation peut sembler banale, mais en Finlande, on écrit au crayon pour les travaux réalisés à la maison, en classe ou même aux examens, tandis qu'en France, entre autres, on utilise un stylo à encre ou à bille. Certes, cet exemple n'est pas pertinent tel quel, mais il donne néanmoins une idée de la manière dont l'immersion culturelle peut influencer les étudiants. D'autant plus que l'étudiant du corpus témoin a séjourné en Suisse avant l'examen du niveau I; en d'autres termes, il écrit avec un stylo à encre au niveau I, mais avec un crayon au niveau II, après avoir commencé ses études en Finlande.

Notre objectif principal n'est pas de nous pencher sur l'analyse des fautes dans l'utilisation des signes de ponctuation, mais plutôt de les envisager en tant qu'élément structurant du texte au même titre que les autres organisateurs pragmatiques qui assurent la cohérence textuelle. Nous avons pu remarquer que les étudiants utilisent le plus souvent les autres signes de ponctuation que le point

d'une façon métalinguistique ou métacommunicative (cf. Fayol 1989). Reprenons encore une phrase déjà présentée :

Ex. 16. niveau I :

Donnez la possibilité aux étrangers !

Nous avons dit que l'étudiant avait mis cette phrase à la fin de sa dissertation pour conclure. Le point d'exclamation final renforce sa conclusion, et donne de poids à son dernier argument.

8.3.2.2. Cohérence textuelle

Comme nous l'avons vu dans la partie théorique (cf. ch. 6.3.), les autres organisateurs pragmatiques, que le découpage en paragraphes et la ponctuation, correspondent aux connecteurs et aux organisateurs textuels classifiés principalement par Adam (1990), par Lundquist (1980, 1983), et par Riegel et al. (1997). Dans le Tableau 10 (à la page suivante), sont présentés les termes utilisés regroupés.

Les catégories d'Adam, de Lundquist et de Riegel et al. se chevauchent, et c'est pourquoi nous avons fait quelques regroupements n'ayant pas besoin d'une catégorisation méticuleuse dans la présente étude. Les explicatifs et les résumatifs de Lundquist et les reformulatifs d'Adam sont utilisés côte à côte. De plus, nous appelons parfois le connecteur *mais* un connecteur argumentatif s'il l'est clairement - sinon, nous n'entrons pas dans la discussion sur les différentes valeurs de *mais* - de même que les autres termes de ce genre, par exemple *même si*. La dernière catégorie de marqueurs métatextuels est compilée à partir des catégories d'Elo (1993), de Lundquist et de Riegel et al. Comme nous l'avons dit supra, les signes de ponctuation autres que le point, sont souvent utilisés plutôt d'une façon métalinguistique ou métacommunicative et sont étroitement liés à la façon de structurer et de commenter le texte des apprenants (cf. Fayol 1989).

organisateur		exemples	catégorisation originale
additif		<i>et, encore, de plus</i>	Lundquist
énumératif		<i>d'abord – ensuite – enfin</i>	Lundquist
transitif		<i>d'ailleurs, d'autre part</i>	Lundquist
explicatif / reformulatif		<i>car, à savoir, en d'autres termes</i>	Adam, Lundquist
illustratif		<i>par exemple, notamment</i>	Lundquist
comparatif		<i>plutôt</i>	Lundquist
adversatif		<i>or, mais, en revanche</i>	Lundquist
concessif		<i>néanmoins</i>	Lundquist
causatif / consécutif / conclusif		<i>donc, ainsi, en effet</i>	Lundquist
résumé / reformulatif		<i>bref, en somme</i>	Adam, Lundquist
présentatif		<i>c'est, il y a</i>	Lundquist
organisateur temporel		<i>puis, en même temps</i>	Adam, Lundquist, Riegel et al.
organisateur spatial		<i>ici, là</i>	Adam, Riegel et al.
connecteur argumentatif		<i>mais, même si</i>	Adam, Riegel et al.
marqueur de clôture / d'ouverture		<i>voilà</i>	Adam
marqueur métatextuel	modalisateur argumentatif	<i>évidemment, justement, même</i>	Riegel et al.
	commentaire métatextuel	<i>voir p., cf.</i>	Lundquist
	introduceur	<i>je crois / pense / trouve que</i>	Elo

Tableau 10. Catégories des organisateurs pragmatiques

En voilà quelques exemples qui donnent un aperçu de l'utilisation des signes de ponctuation par les étudiants. Dans l'exemple 20, l'étudiant utilise le tiret comme un explicatif et dans l'exemple 21 comme un conclusif :

Ex. 20. :

Je suis une fille de la campagne – j'aime les forêts et les lacs, et il y en beaucoup en Finlande.

Ex. 21. :

Leur rire était fou. J'avais peur – je me suis réveillée.

Dans l'exemple 22, le premier paragraphe se termine par une interrogation, une sorte de marqueur d'ouverture (cf. Adam 1990 : 157) :

Ex. 22. :

Il y a quelques années les Finlandais pouvaient dire en souriant : "Naître en Finlande, c'est gagner le gros lot !" Mais qu'est-ce qu'ils disent aujourd'hui ?

Le point d'exclamation est souvent utilisé en même temps qu'un modalisateur argumentatif comme dans l'exemple 23 :

Ex. 23. :

A mon avis : jamais. Dans les autobus les gens n'osent même pas parler !

Trois points à la fin du texte sont des marqueurs de clôture, mais en même temps, ils cachent aussi un sous-entendu, comme dans l'exemple 24 :

Ex. 24. :

*Poli, il se présentait : “Je m’appelle Mark Finlandais.” “Enchanté”, ai-je m-entendu répondre.
Des amis pour toujours...*

Les guillemets, pour leur part, peuvent indiquer soit un illustratif, soit un explicatif, soit un commentaire métatextuel, comme dans les extraits présentés dans l’exemple 25 :

Ex. 25. :

Ceux qui voudraient avoir une formation professionnelle (technicien, etc.), pourraient se concentrer sur des matières plus importantes pour eux et même avoir une période de travail d’un mois.

C’est une maladie de virus HIV, transporté par le sang, la salive, le sperme (~ les liquides du corps).

Dans la vie quotidienne on a besoin de connaissances pratiques, comme comment préparer une tartine. (je pense maintenant aux petits enfant...)

Comme différence entre les deux corpus, nous pouvons remarquer que dans le corpus d’Aix, les étudiants emploient une variété plus large de signes de ponctuation que les étudiants du corpus témoin, même s’il y a, certes, des différences individuelles. Il en va de même également en ce qui concerne les autres organisateurs pragmatiques. Les étudiants étant des apprenants adultes, disposent généralement déjà de techniques pour rendre leur production plus naturelle en ayant recours à leurs connaissances préalables, donc à celles de leur L1 (cf. Noyau & Vasseur 1986 : 117 ; Perdue 1993 : 9-10). Mais on peut constater ici une nette différence entre les deux corpus, à savoir que la plupart des étudiants du corpus d’Aix essayent d’employer divers moyens pour organiser leur texte dès le niveau I, tandis que dans le corpus témoin le recours aux organisateurs pragmatiques est moindre au même niveau ; certes, ils utilisent d’autres moyens, comme l’anaphore pronominale pour relier leurs idées. Il existe par ailleurs des différences individuelles d’une production à l’autre. Nous pourrions peut-être postuler que ceci est lié au fait que dans le corpus d’Aix, il y a 40 étudiants sur 53 qui ont le français comme matière principale, tandis que dans le corpus témoin, il n’y en a que 12 sur 25 ; ainsi, les étudiants de la matière principale accèdent plus facilement aux organisateurs pragmatiques français.

Il ressort donc de l’analyse qu’il y a certaines tendances de changement entre les deux niveaux. Entre autres, les apprenants des deux corpus utilisent des organisateurs pragmatiques aux deux niveaux étudiés, mais l’éventail des

organisateur est plus diversifié au niveau II (cf. aussi Mutta 1999).¹¹⁷ Au lieu d'utiliser seulement des connecteurs additifs (p. ex. *et*) et adversatifs (p. ex. *mais*), les apprenants ont davantage recours aux conclusifs (p. ex. *donc*) et aux combinaisons de divers connecteurs (p. ex. *et ainsi, mais pourtant*). De plus, l'utilisation "correcte" de certains organisateurs semble augmenter (ex. 26 et 27):

Ex.26. niveau I:

Ils voyagent beaucoup, mais ils savent peu des autres cultures. A l'autre côté, la culture de la Finlande n'est pas connue à l'étranger non plus.

Ex. 27. niveau II :

J'ai une amie qui va au cinéma presque tous les jours. D'un autre côté, elle ne regarde jamais la télévision, même s'il y a des bons films.

Il y a d'autres cas où l'organisateur choisi n'est pas correct ; dans l'exemple 28, il s'agit d'un adversatif, en termes de Lundquist, mais l'étudiant aurait plutôt dû mettre *d'autre part*. Pour mieux illustrer la fonction de cet organisateur, nous devrions l'appeler marqueur d'intégration linéaire (Adam 1990 : 157), ou marqueur de relais.

Ex. 28. niveau II :

Peut-être que ai-je un point de vue trop pessimiste. Bien sûr les gens "savent" et osent rire, mais pour quelque'un c'est] une chose privée et, paradoxalement, sérieuse. Dans un tel cas on rit quand on a vraiment envie de rire. Au contraire tout le monde connaît les gens qui rient tout le temps, sans arrêt presque 24 heures sur 24.

Le conclusif *donc* est un cas spécial, pour ainsi dire. Son utilisation augmente beaucoup au niveau II ce qui est probablement dû à l'influence du langage parlé. Ce qui est à remarquer, c'est qu'il est tantôt utilisé de manière incorrecte en tant qu'enchaîneur, une sorte de transitif (ex. 28), tantôt de manière correcte en tant que conclusif (ex. 29).

Ex. 28. :

[...] J'ai pris le train de Turku à Aix avec une bonne copine, et le voyage s'est bien passé. A l'arrivée nous avons été surpris par la chaleur après le temps froid en Finlande !

Donc, nous nous sommes installées à nos chambres à la cité universitaire, où la vie sociale était très vivante. [...]

Ex. 29. :

[...] En plus il y a aussi des usines qui ne s'occupent pas de la nature et qui laissent sortir des gaz qui contiennent même des poisons. Donc, petit à petit la nature mourra à cause de ces deux menaces, si on n'agit pas vite et radicalement. [...]

Bien qu'il y ait moins de *mais* et de *et* utilisés seuls au niveau II, il peut arriver que l'étudiant ait recours à ces mots, mais d'une façon sensiblement déviante. En

¹¹⁷ Ce résultat est semblable aux résultats obtenus par Kerr-Barnes (1998 : 206-208) dans une étude sur l'utilisation des connecteurs par des apprenants adultes de français langue étrangère.

d'autres termes, l'étudiant utilise ces mots plutôt comme des transitifs, ou marqueurs de relais (cf. Adam 1990 : 157) entre deux énoncés. Ces marqueurs, s'ils étaient absents, ne changeraient pas énormément le sens du texte et auxquels on pourrait substituer un autre organisateur. Ceci est probablement dû aussi à l'influence du langage oral français, ou même du finnois. Voici deux exemples des ces cas de figure :

Ex. 30. :

[...]

Et quelle est notre réaction ? Je pense que la plupart de nous accepte un certain nombre d'étrangers chez nous. Ils pensent qu'il est bien de montrer aux autres pays, que nous en accueillons aussi, mais qu'il ne faut pas être trop gentil... Et qu'ils ne causent pas de problèmes !

[...]

A mon avis il est bien d'avoir des étrangers, ça change. En ce moment il n'y en a pas trop, mais je trouve qu'il y a la limite quelque part. Et il faut les accueillir peu à peu pour que tout le monde puisse s'habituer à la nouvelle situation.

Ex. 31. :

[...]

Mais la Suisse n'est pas la seule possibilité... Il y a six mois, j'ai fait un grand voyage en Chine qui m'a beaucoup impressionnée. C'est une culture tellement différente de la nôtre : parfois on avait l'impression d'être non de l'autre côté du monde, mais sur une autre planète ! C'est pourquoi je ne pourrais pas décider d'aller vivre en Chine pour toute ma vie, mais pour quelques années, pourquoi pas ?! Je serais d'accord de prendre des risques, de vivre plus modestement que maintenant, et d'apprendre une langue difficile mais fascinante. Mais il faut faire de ce genre d'expérience quand on est jeune ; plus tard cela peut être difficile et peut-être que je n'aurai plus envie de le faire !

[...]Pour l'instant, j'ai envie de voyager, de découvrir d'autres cultures, d'apprendre d'autres langues. Mais qui sait ? Peut-être que dans vingt ans, après avoir vu les quatre coins du monde, [...].

En outre, ce qui semble augmenter dans l'utilisation des organisateurs, ce sont les marqueurs métatextuels, c'est-à-dire, les modalisateurs argumentatifs, les commentaires métatextuels et les introducteurs. Nous avons supposé dans la partie théorique qu'au fur et à mesure de l'apprentissage l'apprenant laisse la priorité à ses compétences métapragmatiques et métatextuelles (Gombert 1991 : 152, cité dans Bouchard 1994 : 122) ce qui se manifeste par une aisance dans l'écriture. Ceci constitue une part essentielle de l'amélioration des capacités d'écriture, car l'amélioration des capacités textuelles de l'apprenant favorisent l'augmentation de la cohérence textuelle et ainsi, l'augmentation de la compréhensibilité ou de la lisibilité textuelle en même temps. Dans l'exemple 32, il y a également plusieurs organisateurs temporels au niveau I car il s'agit d'un texte narratif.

Ex. 32. niveau I :

J'ai vu plusieurs films dont le sujet a été la guerre de Vietnam mais le dernier de ceux-ci était peut-être le meilleur. C'était "Né le 4 juillet". Le personnage principal était un jeune homme. Sa famille était très patriotique et très "américaine". Les parents voulaient faire un hero de leur fils.

Puis quand la guerre de Vietnam avait éclaté, cet jeune homme était plein d'envie d'y aller.

Finale­ment il pour­rait être un hé­ro, dé­fendre le monde libre contre la dan­ger hor­rible, le com­mu­nisme. Mais il n'é­tait pas si fa­cile d'être un hé­ro. Par ac­ci­dent il a tué un Amé­ricain et plus tard il est blessé gra­ve­ment. Il est re­venu aux Etats Unis, mais dans un hô­pital des vé­té­rans de la guerre. L'hô­pital était hor­rible. On n'avait pas d'ar­gent pour cet hô­pital, parce que la guerre était si cher. Quel paradox ! Il est resté pa­ralysé.

Puis il est re­venu à la maison, mais il ne pou­vait plus s'ha­bituer à la vie nor­male. Il s'est aussi aper­çu qu'il y avait beau­coup de pro­tes­ta­tions contre la guerre entre les jeunes gens. Il avait beau­coup de dif­fi­cultés, mais à la fin il trouve sa place dans la vie. Il est de­venu un ac­tiviste et en­tré dans la poli­tique.

La fin du film était trop heu­reuse et idéaliste, mais au­tre­ment il était un bon film. Les pro­blèmes de cet homme, qui est de­venu in­valide et qui est perdu tous ses rêves, étaient plus im­por­tantes que la dé­scrip­tion de la guerre.

Ex. 33. niveau II :

Il est évident qu'il existe une culture jeune qui est sé­parée de la culture "géné­rale" de la société. Le commerce en profit et on fait beau­coup de publi­cité spé­cia­le­ment visée à la jeunesse. Pour­tant on peut se de­mander si cette culture est unifiée et homogène. Comme on peut di­viser la société en­tière dans les classes (sociales p. ex.) sé­parées, on peut aussi trouver dans la jeunesse des groupes très dif­fé­rents.

Au dé­but (dans les années 1960 et 1970) la culture jeune vou­lait faire une ré­volte. Les jeunes s'op­po­saient à la mode de vie moyenne de leurs parents. Ils vou­laient chan­ger le monde. A mon avis, ce "but" commun fa­isait un lien unifiant entre les jeunes, bien qu'il y ait des groupes dif­fé­rents.

En ce moment je pense que la jeunesse est plus frac­tu­rée et hété­ro­gène. Bien sûr les jeunes veulent toujours se sé­parer des gens plus agés et de ses parents, un senti­ment de ré­volte ap­partient à la jeunesse, mais, a mon avis, il n'est plus si évident qu'avant.

Enfin je suis d'accord avec M. Djian quand il dit qu'il s'agit d'une culture atomisée, mais avec des lan­gages communs. Au­tre­ment dit qu'il existe une culture propre aux jeunes, mais qui n'est pas unifiée et simple.

En ce qui concerne les stratégies métacommunicatives, nous avons trouvé des méthodes utilisées dès le niveau I chez certains étudiants. Il s'agit de stratégies d'ouverture du texte ou de phrases fixes pour organiser le texte, entre autres, par des marqueurs métatextuels (cf. supra).

Ex. 34. niveau I :

Je pense absolument que la culture générale est meilleure alternative que la spécialisation intense. Après le lycée les jeunes ne sont plus du tout mûres et raisonnables bien qu'ils aient dix-huit ans. Il ne sont pas encore capables de choisir leur métier.
[...]

Ex. 35. niveau II :

Je suis absolument d'avis que la culture humaine n'avance guère s'il n'y a pas d'exploration profonde de tous les éléments de la vie humaine. D'une part les sciences avancent en vitesse mais d'autre part on a toujours besoin d'artistes. Ce sont justement les écrivains et les peintres qui peuvent décrire la situation dans une société ou dans le monde telle qu'elle est. [...]

Certes, comme nous l'avons déjà mentionné plusieurs fois supra, les variabilités individuelles jouent également un rôle dans les productions ; certains utilisent, par exemple, beaucoup d'organismes pragmatiques très divers dès le premier niveau dans les deux corpus. En somme, les résultats du corpus témoin semblent aller

dans le même sens que ceux du corpus d'Aix - sauf que la variété des organisateurs semble moindre - infirmant ainsi notre hypothèse sur les différences entre les corpus.¹¹⁸ Nous pouvons donc conclure ici, d'après les résultats étudiés dans l'analyse, que les capacités métatextuelles des apprenants s'améliorent effectivement pendant le séjour linguistique à l'étranger.

8.3.2.3. Développement interlingual

Dans l'ensemble, on peut discerner une cohérence plus élevée au niveau II qu'au niveau I. Cela se voit clairement dans les dissertations des apprenants étudiant le français comme matière secondaire. Au niveau I, les apprenants utilisent beaucoup de phrases isolées. Ou alors les moyens linguistiques utilisés pour contribuer à la cohérence du texte ne sont présents qu'à l'intérieur des paragraphes qui eux, restent isolés. Au niveau II, par contre, ce n'est plus uniquement au niveau phrastique que l'enchaînement est cohérent ; il l'est également entre les paragraphes.

Ex 36. niveau I :

Si on compare le nombre d'étrangers en Finlande avec ce des autres pays nordiques par. ex la Suède et le Danemark, on peut bien voir que le nombre est petit. A mon avis, il n'y en a pas trop.

Si la Finlande ne rejoindra pas la C.E.E., notre pays sera trop extérieur d'Europe. Les étrangers ne me font pas peur même s'il y en avait beaucoup. Les étrangers peuvent nous apprendre de ce qu'ils savent bien, des difficultés sont été créés pour les résoudre.

Je pense que les réfugiés ont seulement deux possibilités : souvent c'est d'être tortu ou même mourir en son propre pays et la deuxième possibilité c'est d'immigrer dans un pays étrange. Je ne connais pas bi en le nombre exact de réfugiés en Finlande mais il y en a pas trop.

C'est souvent les préjugés des gens qui constituent des obstacles à l'intégration des étrangers.

Je ne connais pas personnellement des étrangers en Finlande mais au contraire j'ai des amis en étranger.

Ex. 37. niveau II :

Après avoir passé quelques mois à Aix-en-Provence, j'ai commencé à réfléchir si je veux habiter en Finlande pour l'éternité. En Finlande je suis attaché à mes amis. Mon loisir et mes études me plaisent et la variation des saisons. Mais en ce qui concerne la vie culturelle, il y a quelque chose qui me manque. A Aix, j'ai mais bi en m'asseoir dans les cafés avec mes amis. C'était magnifique de passer le temps en regardant les passants, en regardant la vie animé et en parlant avec des chaleureux méridionaux.

C'est triste qu'en Finlande il nous manque de bons cafés et que dans les plusieurs cafés on doit se servir soi-même ! En plus, on ne peut pas voir des films français ici autant qu'en France. Cela me manque aussi. Je ne sais pas, peut-être que je me trouverai à l'étranger après les études !

¹¹⁸ Il faut noter pourtant que tous les changements ne sont pas dus au séjour linguistique en France, car l'enseignement du français dans notre département consiste également en cours de production écrite.

De plus, les idées semblent évoluer au fur et à mesure que le texte progresse au lieu de se répéter d'un paragraphe à l'autre.

Ex. 38. niveau I :

Il y a plusieurs personnes qui sont d'avis que l'éducation des filles et celle des garçons doivent être différentes. Ils pensent aussi très souvent que les métiers doivent être divisés en deux groupes : l'une devrait se consister seulement des métiers des hommes et l'autre des métiers des femmes. A mon avis, l'éducatons des garçons et des filles devraient être identiques.

Si on pense qu'aujourd'hui il n'existe pas des métiers qui sont traditionnellement et seulement des métiers des hommes ou des métiers des femmes, il est bien sûr naturel que les éducations des deux sexes doivent être identiques. Il est très important que les filles et les garçons apprennent les mêmes choses à l'école maternelle et aussi au lycée. De telle façon, ils obtiennent la même base élémentaire pour leur métier et pour leur vie. En parlant des universités et des écoles qui préparent les élèves pour les métiers, on peut dire la même chose.

Quant à la situation de l'éducation en Finlande, est-elle tolérable ? On peut répondre qu'elle est peut-être idéale. L'état donne – ou surtout essaye de donner – une éducation identique à et les garçons à et les filles, et cela c'est important.

Ex. 39. niveau II :

De nos jours, le monde et la société finlandaise changent très rapidement. Le développement est quelquefois très profond. Tout cela influence ma vie, mais pourtant à mon opinion, la raison la plus évidente des changements dans ma vie, c'est moi-même.

Le changement très important concernant la Finlande, c'est l'entrée dans le CEE. On dit que cela améliore la situation économique de la Finlande, mais d'autre part on n'est pas sûr si le nombre des chômeurs diminue. Un changement très sûr est que la Finlande s'internationalisera. Dans quinze ans, les changements les plus évidents dans le monde concerneront le développement des sciences : de la technologie et de la médecine. Par exemple dans les années 70 et 80, on pensait souvent que le développement mène à la production des armes plus effectives et à cause de cela, à la troisième guerre mondiale et à la destruction du monde. Aujourd'hui, la situation politique mondiale n'est pas si grave. Ce qui va changer dans ma propre vie, c'est que je terminerai mes études. Si tout va bien, j'aurai du travail aussi. Les études universitaires durent longtemps, plusieurs années. A cause de cela, on s'habitue à la manière de vivre d'un étudiant et il n'est pas si facile de changer de manière de vivre.

A mon avis, il y a plus de raisons de pessimisme que de raisons d'optimisme. Une des raisons de pessimisme est le menace de rester au chômage après les études. Les personnes avec une formation académique n'obtient pas toujours du travail à cause de la situation économique mauvaise de la Finlande. Le développement de cette situation est aussi une raison de pessimisme. Une raison d'optimisme est qu'après avoir terminé mes études, je peux penser que j'ai terminé le travail dur de plusieurs années.

En général, il est très difficile d'imaginer comment la vie sera à l'avenir. On ne peut jamais être absolument sûr.

D'une certaine manière, on peut également dire que les dissertations sont plus correctes du point de vue lexical au niveau II (cf. supra), même s'il reste encore des fautes d'orthographe. Dans le corpus d'Aix, nous pouvons par ailleurs remarquer, une petite liberté concernant le sujet de dissertation. Autrement dit, dans le corpus témoin, les apprenants écrivent toujours sur le sujet donné, tandis que dans le corpus d'Aix, certains étudiants, tout en restant dans le cadre du sujet donné, laissent leurs pensées vagabonder vers des considérations qui s'en écartent. Dans l'exemple 1 qui est dans l'Annexe 4, le sujet proposait aux étudiants de décrire un ou plusieurs de leurs amis, et l'étudiant a choisi de parler de l'argent

(cf. ex. 24). Dans l'exemple suivant, le sujet était : *Dans quel pays aimeriez-vous vivre ?*

Ex. 40. niveau II :

Je voudrais vivre dans le pays des merveilles, comme Alice. Vivre dans le monde réel est trop compliqué. Il faut penser sans cesse. Des fois j'oublis de vivre – je suis dans la prison de mes pensées. Pour en sortir il faut apprendre à voler. Je ne sais pas voler encore et tous ces clichés que j'ai écrits viennent de "la prison de mes pensées", si vous permettez une citation.

J'ai visité le pays des merveilles plusieurs fois. En fait, je voudrais y aller maintenant, mais je suis obligée de rester dans cette salle et d'écrire aux gens que je ne connais pas. Quelle misère. "Tu vis, donc tu souffres" a dit quelqu'un qui avait tout-à-fait raison. Il me reste encore une vingtaine de minutes de souffrance. A une heure je serai libérée. Le pays, où il n'y a pas de productions écrites, m'attend. Ça y est. J'ai fini.

Il est difficile de savoir d'où vient cette liberté. Nous avons vu supra que les étudiants du corpus d'Aix suivent plus strictement les consignes données quant à la longueur des dissertations. Dans nos derniers exemples, il s'agit d'une liberté individuelle. Nous pourrions peut-être penser qu'une des raisons de cette liberté est due au fait que ces étudiants ont effectué leur séjour à l'université d'Aix-en-Provence la même année, et qu'ils ont participé en partie aux mêmes cours (cf. aussi ex. 41 infra). Néanmoins, nous ne pouvons pas être sûre de l'influence de ces cours sur leurs productions.

Les exemples supra, en particulier les exemples 37 et 40, témoignent également d'un phénomène que nous avons postulé dans une de nos hypothèses après la première lecture du corpus. Nous avons supposé que des traits du langage parlé s'immiscent dans la production écrite. En effet, ceci peut être observé dans les dissertations du niveau II, par exemple, la particule de négation *ne* a tendance à tomber, mais d'une façon irrégulière ; dans les dissertations, cela arrive le plus souvent dans les phrases où le sujet est le pronom neutre *on*, par exemple, *on peut pas savoir*. En termes de Regan (1995 : 178-179), nous pouvons ainsi parler de l'apprentissage au plan horizontal, soit sociolinguistique, même si cette forme est incorrecte selon les règles normatives (cf. ch. 3.4.). De plus, nous pouvons ajouter que du point de vue du vocabulaire appartenant à un registre du langage familier ou parlé, ce sont surtout les étudiants qui ont passé deux semestres à l'université d'Aix-en-Provence qui y ont recours. Il en est de même également dans le corpus témoin : les étudiants ayant séjourné assez longtemps dans un pays francophone, utilisent plus facilement le vocabulaire du français parlé. Nous trouvons des mots ou des expressions comme *ben ; bof ; ça y est ; voilà ; ils le ratent ; chouette ; pas mal de ; je ne sais pas ; je me trouverai à l'étranger ; peut-être que ; tout*

simplement ; et moi, je le connaissait, et aussi des abréviations comme télé, sympa et prof. Un des étudiants est allé encore plus loin (cf. supra), puisqu'il a écrit partiellement son anecdote de voyage sous la forme d'un dialogue :

Ex. 41. :

Il y a un an, quand j'étudiais en France, j'avais l'habitude de remplacer le déjeuner par un sandwich et par un chocolat chaud dans le café de la faculté.

Il est arrivé une fois, je crois que c'était un de ces lundis où j'ai raté le premier cours, que je suis allé dans le café du coin pour y avoir mon sandwich, au lieu de me promener jusqu'à l'Université. C'était le café que je fréquentais, mais normalement je n'ai pris là qu'un café ou une bière. Cette fois-là, j'avais faim.

- Bonjour, pour moi un sandwich au jambon, s'il vous plaît.

- Bon, et avec le sandwich ?

- Un café crème.

Mes mots ont provoqué un effet incroyable. Il faut constater que on se connaissait déjà bien, le garçon et moi, parce qu'il s'agit de mon café préféré, mais quand même la réponse du garçon était surprenante :

- Non, ça ne va pas.

- Mais pourquoi, la cafetière est cassée ?

- Non, mais en France on ne prend jamais du café avec un sandwich.

Ah bon, maintenant j'ai compris : il ne suffit pas, qu'on connaît la langue, il faut connaître les mœurs aussi. Ce sont les mœurs qui sont les plus importantes. J'ai essayé de nouveau :

- Alors, qu'est-ce que je pourrais boire avec mon sandwich ?

- Du vin, par exemple.

- Non, ça ne va pas, je dois aller à l'Université.

- Ben, dans ce cas-là, une bière.

- Mais non, il faut quelque chose d'autre, une boisson sans alcool.

Enfin, après cette longue négociation, le garçon a servi mon sandwich avec l'eau minérale. Et après avoir mangé le sandwich, j'ai eu mon café.

En revanche, au niveau II, nous trouvons dans le groupe témoin deux phénomènes plutôt structurels et stylistiques inexistants dans le corpus d'Aix. Dans le premier cas, il s'agit de l'utilisation du passé simple et dans le second, d'une syllepse facultative concernant le pronom *on*. Dans l'exemple 42, le passé simple est acceptable à cause du contexte, mais dans l'exemple 43, il est mal justifié.

Ex. 42. :

[...] On croyait au américanisme et quand la guerre de Vietnam éclata, Ron était prêt pour y aller et pour défendre son pays contre la menace du communisme.

[...]

Ex. 43. :

[...] Donc, un soir à la fin du dîner, j'ai réussi à donner un souris amical à la mère (c'était plutôt une grimace) et je lui ai dit plaine de bonne volonté : "Merçi. C'était délicieux ! Je suis plaine !" Claire me lanca un regard douteux et puis elle eclata de rire.

L'exemple 44 montre la syllepse facultative :

Ex. 44. :

[...]

Il pleuvait à Helsinki lorsqu'on est parties pour l'Italie, mais à Rome, le soleil nous a souhaitées bienvenues.

Notre court séjour à Rome était merveilleux. On a visité tant de monuments historiques, on s'est

bien reposées, on a bien mangé et bu. En somme, on a passé les vacances idéales.
[...]

Pourquoi les étudiants ont-ils utilisé ces formes ? Nous ne le savons pas, mais peut-être s'agit-il d'une forme d'hypercorrection, en particulier dans l'exemple 43. Cela veut dire que l'étudiant a peut-être eu envie de donner à son texte une tournure plus soutenue.

8.3.2.3.1 Cas particuliers

Il nous reste encore à examiner les dissertations des étudiants dont les résultats quant à la variation lexicale et à la densité lexicale déviaient des tendances générales (cf. ch. 8.1.2. et 8.1.3.) pour voir s'il y a une régression dans leur production d'une part et tenter d'expliquer pourquoi il y a une déviation d'autre part.

Commençons par les étudiants dont la densité lexicale dépasse 45 %, c'est-à-dire qu'ils utilisent plus de mots lexicaux que de mots grammaticaux dans leur essais. Nous avons indiqué supra que la limite de 40 % de mots lexicaux est supposée correspondre à un niveau de scripteurs natifs dans une dissertation qui est un texte non-interactif. Ici, cette limite est largement dépassée, ce qui pourrait indiquer qu'il y a une amélioration dans les dissertations, mais paradoxalement celle-ci n'est généralement pas accompagnée d'une réussite à l'examen (cf. Linnarud 1986 : 82). Il s'agit des étudiants 2, 14, 20, 23, 24, 31, 43, 49 et 52 dans le corpus d'Aix et des étudiants 3, 20 et 24 dans le corpus témoin (cf. Annexe 3). Les étudiants 20, 31 et 49 (Aix) et l'étudiant 24 (témoin) ont aussi une VL moins élevée au niveau II. Nous pouvons déduire de l'analyse de ces dissertations du corpus d'Aix que dans deux cas seulement, à savoir chez les étudiants 24 et 43, l'évolution entre les deux niveaux est moindre ou même régressive si l'on considère leur production du point de vue lexical. Il est difficile de savoir si ce phénomène est dû à la situation d'examen, à une démotivation de l'étudiant, ou à une stagnation des compétences d'écriture. Il faudrait suivre leurs productions à long terme et d'une manière systématique pour pouvoir se prononcer. Ajoutons cependant que ces étudiants de matière principale n'ont passé qu'un semestre à l'université d'Aix-en-Provence. Dans les autres cas, on ne peut pas parler de régression. Il semble que les résultats sont affectés par la méthode que nous avons utilisée, à savoir le recours au logiciel WordCruncher. Celui-ci a influencé les

résultats car, bien que les étudiants utilisent plus d'organismes pragmatiques qui sont souvent des mots grammaticaux, ils emploient également des organismes qui sont formés de plusieurs mots et qui, par conséquent, sont calculés séparément, par exemple, *je pense que*. Par ailleurs, certains adverbes en formes -*ment* sont considérés comme des mots lexicaux, par exemple, *personnellement, premièrement, deuxièmement*. Ceci est à noter également dans le corpus témoin, sauf que chez l'étudiant 24, il y a une légère régression quant à l'utilisation des organismes pragmatiques. Néanmoins, l'étudiant a fait moins de fautes d'orthographe au niveau II et sa dissertation comporte moins de mots dus à l'interférence de l'italien qui est sa matière principale.

L'exemple 45 présente les dissertations de l'étudiant 24 (corpus d'Aix) dont le développement lexical est moindre ; toutefois, nous pouvons remarquer un changement quant au degré de la cohérence du texte entre les deux niveaux.

Ex. 45. niveau I :

Bien sûr, on a toujours des principes magnifiques qui concerne l'éducation des enfants quand on n'a pas encore d'enfants. En pratique c'est difficile à garder les principes.

Moi, j'ai un frère qui a quatre ans et je pense que mes parents ont une attitude trop libérale pour lui. Quand il ira à l'école je sais déjà qu'il va avoir des difficultés avec ses professeurs.

Actuellement, on devrait avoir l'attitude plus stricte à l'école. Les enfants devraient vouvoyer les professeurs ; de telle sorte on apprend de respecter des gens plus âgés. Cela serait aussi obligatoire de monter quand on parle et quand on réponde aux questions.

A la maison on ne doit pas être aussi stricte qu'à l'école, mais en tout cas, je pense que les parents doivent être des exemples pour les enfants. De dire merci et pardon est très important, aussi.

Niveau II :

Quand le baccalauréat, à l'âge de dix-huit ans, quitte le lycée, il est très savant. Mais s'il a vraiment appris tout ce qu'on lui a enseigné, il n'a guère eu le temps à faire autre chose.

Par exemple en Finlande les pauvres lycéens passent en viron sept ou huit heures par jour à l'école et en plus ils ont des devoirs à faire.

On remarque que l'école joue le rôle principal dans leur vie. Pour cette raison l'école doit aussi être responsable du développement psychique de ses élèves.

La situation actuelle n'est pas très bonne, pour améliorer la situation, il faudrait supprimer quelques cours théorétiques et ajouter des cours créatifs et des cours qui aideraient les jeunes à former sa personnalité.

Bien qu'il soit difficile à préparer les jeunes à la vie, l'école devrait, malgré tout, donner aux jeunes le temps à trouver leur propre caractère. Quand on est sûr de soi-même, on est prêt à vaincre les difficultés dans la vie.

L'exemple 46, pour sa part, présente les dissertations d'un étudiant de matière principale ayant passé deux semestres à Aix-en-Provence ; au niveau II, il y a encore beaucoup de fautes d'orthographe, mais l'utilisation des organismes est plus développée par rapport au niveau I (l'étudiant 49).

Ex. 46. niveau I :

Naturellement les pays capitalistes ont suivi avec satisfaction la chute des structures sociales des pays communistes. Les anciens ennemis ne sont plus des lions dangereux mais on les voit comme chats sans dents. Bon, la risque de la guerre en Europe a diminué mais avec quel genre de problèmes les pays ex-communistes doivent se poser vis-à-vis maintenant ? Ce sont les difficultés économiques et les temps très difficiles. Il s'agit d'un proces d'apprendre à penser d'une manière toute neuve – à la manière capitaliste.

Il est probable que la majorité des gens en Est est heureuse de cette changement totale parce que la situation avait été catastrophale très longtemps – rien ne marchait plus – mais certainement il y a aussi des gens qui ne sont pas tout à fait satisfait.

On peut demander pourquoi le système communiste ne fonctionnait pas. Les idées et les buts en étaient grands et bons mais... D'où venait le totalitarisme, la corruption et tous les maux possibles ?

Maintenant nous vivons dans le monde, ou plutôt dans l'Europe, unissant et on supprime les frontières mais on peut demander si c'est vraiment une bonne changement. Dans le monde uni il n'y aura pas de choix ! On ne peut pas changer de "l'atmosphère" si l'on veut. Où pourront aller les dissidents futurs ? Avec qui ou quel système pourrons-nous comparer notre société dans l'avenir ? Qui sera là pour nous faire remarquer notre fautes dans le monde de demain ? Personne à nous blamer ? Personne ? Quelle horreur !

Niveau II :

Il y a toujours des situations où la vérité peut se présenter bien difficiles à dire. Surtout, si on doit dire quelque chose désagréable à quelqu'un qu'on aime. Bien sûr il y a des moments où l'on veut mentir pour sa propre cause aussi, pour se justifier. On peut toujours "retoucher" de ses actions un peu par les explications.

Personnellement, je ne voudrais jamais mentir aux autres, mais trop souvent j'ai dû rendre cette principe rigoureuse plus flexible. Peut-être suis-je un grand égoïste, mais je crois que j'ai menti plus souvent pour justifier mes propres actions aux yeux d'autrui que ne pas peiner quelqu'un d'autre par la vérité.

Naturellement j'essaie de m'abstenir de pieux mensonges et la meilleure façon de le faire, à mon avis, est la suivante : ne pas dire la chose difficile du tout si la nature de cette chose permet un tel mutisme !

Enfin, je ne voit pas que les mensonges et les pieux mensonges soient la même chose. Si on est sûr qu'on peut éviter la peine inutile chez quelqu'un, on peut se recourir aux pieux mensonges ou bien garder la silence.

Il est d'un grand naïveté de s'imaginer que nous autres hommes puissions toujours être absolument francs ou honnêtes ! C'est pour les anges et les saints !

Bien sûr, chacun doit se veiller aussi qu'on ne dépasse la limite du bon sens. Les mensonges ne doivent pas devenir de moyens d'échappement de toute situation difficile.

On doit choisir prudemment sa propre conduite et ses paroles selon le cas.

Voyons à présent les dissertations des étudiants dont la DL est inférieur à la limite de 40%. Nous avons déjà remarqué (cf. ch. 7.2.3. et 8.1.3.) que dans un texte écrit par un autochtone, la quantité de mots grammaticaux dépasse normalement celle des mots lexicaux, bien que l'on prétende que leur grand nombre appauvrit le texte (cf. Ménard 1983 : 52). Regardons les dissertations des étudiants 13, 27, 40, 4, 5, 6, 12, 16, 34, 50 et 51 du corpus d'Aix, et des étudiants 6, 9, 10, 11, 13, 19,

23 et 25 du corpus témoin. Les étudiants 9 et 25 du corpus témoin ont aussi une VL moins élevée au niveau II. Nous pouvons en conclure que la diminution de la densité lexicale est clairement due à l'augmentation de l'utilisation des organisateurs pragmatiques dont une partie est des adverbes appartenant aux mots grammaticaux (p. ex. *plutôt*), ou à l'augmentation de l'utilisation des anaphores pronominales, par exemple dans les textes narratifs. Cependant, dans le corpus témoin, les étudiants 10, 11, 23 et 25 ont une DL légèrement plus élevée au niveau II, même si elle reste inférieure à 40 %. Ainsi, la distribution entre les mots grammaticaux et lexicaux devient plus naturelle dans leur dissertation ; une raison pour cela peut être que l'étudiant a moins d'adverbes et plus d'organisateur métatextuels au niveau II. Nous ne donnons qu'un exemple de ce phénomène ; nous nous référons cependant à d'autres exemples présentés supra, dans notre partie concernant la cohérence textuelle par exemple. L'exemple 47 vient du corpus d'Aix (l'étudiant 27):

Ex. 47. niveau I:

Aujourd'hui le sida est peut-être la maladie dont on a le plus peur dans le monde entier. La maladie s'est étendue partout, surtout dans les pays peuplés.

Combien des victimes du sida on aurait pu sauver, s'ils avaient su se protéger dans leurs relations sexuelles, s'ils avaient su qu'il existe une maladie pareille ? Il y en aurait eu beaucoup. Mais jusqu'à ces dernières années on n'a pas eu beaucoup d'information sur cette maladie et maintenant le problème est de trouver un moyen pour guérir les gens qui l'ont eue. C'est aussi important de donner à tout le monde d'information correcte sur cette maladie. Dans plusieurs pays on a déjà commencé les campagnes de prévention, mais par exemple au Tiers Monde la plupart des gens ne sait rien du tout du sida.

Chez nous en Finlande on en a bien donné d'information, mais si on a eu assez, j'en doute. Il faudrait surtout penser des jeunes, garantir leur future sans sida. Il faudrait faire les gens penser raisonnablement: "On va ensemble à l'examen de dépistage du sida avant qu'on commence une relation sexuelle" ou bien "on utilise la protection en faisant l'amour". Pourquoi prendre des risques ? Il faut battre ce sida, maladie horrible.

Niveau II :

Ces dernières années le nombre d'étrangers a beaucoup augmenté aussi chez nous ; La Finlande s'internationalise ! Mais va-t-on s'habituer les uns aux autres ?

Quand les étrangers arrivent en Finlande, ils trouvent très souvent devant eux la vie qui diffère totalement de celle dans leur propre pays. Ils sont dans une situation où ils doivent tout recommencer. Mais en décidant de quitter leur pays, ils ont dû accepter le fait que leur vie va changer. Ils doivent apprendre les coutumes et la langue du pays et sûrement ils font de leur mieux en s'adaptant à nous, bien que ça ne soit pas toujours facile.

Et quelle est notre réaction ? Je pense que la plupart de nous accepte un certain nombre d'étrangers chez nous. Ils pensent qu'il est bien de montrer aux autres pays, que nous en accueillons aussi, mais qu'il ne faut pas être trop gentil... Et qu'ils ne causent pas de problèmes !

Mais d'autre part il y a entre nous des gens qui ne veulent pas accueillir des étrangers. A leur opinion la situation sociale n'est pas suffisamment bonne pour les accueillir ; il faut d'abord prendre soin de nous-mêmes.

A mon avis, il est bien d'avoir des étrangers, ça change. En ce moment il n'y en a pas trop, mais je trouve qu'il y a la limite quelque part. Et il faut les accueillir peu à peu pour que tout le monde puisse s'habituer à la nouvelle situation.

Finalement, examinons les cas où la VL a diminué au niveau II, même si le nombre de mots a également diminué ce qui devrait normalement augmenter la VL. Il s'agit des étudiants 10, 22, 26, 36, 47, et 48 dans le corpus d'Aix et des étudiants 7 et 15 dans le corpus témoin (en plus de ceux que nous avons mentionnés supra). Nous pouvons déduire de l'analyse de ces dissertations que la valeur de la variation lexicale est effectivement étroitement liée à la quantité des mots réellement uniques, comme nous l'avons déjà indiqué grâce aux calculs statistiques concernant les corrélations entre les différents facteurs quantitatifs dans le chapitre 8.2. Ainsi, quand la VL diminue, le nombre de mots réellement uniques diminue également. Dans ces dissertations, nous pouvons voir une augmentation de la cohérence textuelle, certes, mais en même temps, nous constatons une plus grande répétition du lexique. Cela peut-être dû au sujet de la dissertation, par exemple *Le rire est le propre de l'homme [...]* comme nous présente l'exemple 48 (l'étudiant 10). On peut supposer que l'étudiant a eu des difficultés à trouver des synonymes pour le terme "rire". L'exemple 46 nous a également bien montré comment le sujet peut influencer sur le contenu et sur l'utilisation des mots.

Ex. 48. niveau I :

Aujourd'hui il y a très peu d'étrangers et de réfugiés en Finlande, quand on compare cette situation de Finlande à celle de Suède. Et peut-être que cela est une cause pourquoi les Finlandais ne savent pas, comment ils devraient se comporter vers les étrangers. L'homme noir est presque une attraction touristique aux petits villages. S'il est médecin et s'il s'intéresse à la politique en même temps, c'est quelque chose d'extraordinaire, un journaliste va chez lui pour faire un interview. Nous les Finlandais, nous ne laissent pas les étrangers vivre normalement.

La Finlande accueille quelques centaines de réfugiés par ans. Ces personnes ont dû laisser leurs parents, leurs maisons, leurs pays et leurs cultures derrière elles et ils n'ont pas fait cela avec volentier. Mais pourtant nous les envient, parce que l'Etat va leur donner de l'argent, des appartements et des vêtements. Il y a des pauvres, des sans-abris et des chômeurs aussi ici en Finlande, c'est vrai mais je pense que nous sommes aussi avare quant à eux que quant aux réfugiés.

Bien sûr la Finlande ne peut pas accueillir tous les réfugiés et il faut choisir ceux, qui ont le meilleur besoin. Les étrangers rend notre culture plus riche, mais notre culture n'est pas meilleure que les autres cultures. Par exemple nous devons donner aux enfants étrangers la possibilité d'apprendre leurs propres langues ici.

Donnez la possibilité aux étrangers !

Niveau II :

Une bonne plaisanterie me fait toute de suite rire et c'est toujours la première fois qu'on entend une plaisanterie qui fait rire au fond du coeur. S'il nous arrive quelque chose de stupide ou fou, nous ne rions pas à ce moment mais après avoir réfléchi sur ce qui s'est passé. Et cela peut être du fou rire, on rit jusqu'aux larmes. Lorsqu'on est vraiment fatigué on rit de tout possible et après avoir se comporté de cette manière on a plutôt honte, car ce sont surtout les petites filles qui rient et rient sans pouvoir cesser le rire. Il y a aussi des situation où on ne peut pas rire en haute voix sans faire un scandale. Et c'est vraiment pénible, car on doit trembler sur sa place en souffrant de l'attaque de fou rire.

Les gens rient bien sûr de différentes choses, mais il y a ceux qui ont pris l'habitude de rire des échecs d'une personne. Il s'agit d'un rire méchant. Mais dans une situation tendue le rire "convenable" peut rendre l'ambiance plus agréable à tous.

Le rire fait vivre plus longtemps.

Une autre raison peut être également une stratégie de communication, comme nous l'avons déjà mentionné supra. L'étudiant a peut-être peur de prendre des risques quant au vocabulaire et a recours à l'utilisation du vocabulaire qu'il connaît bien ; donc une stratégie "sans surprise" (l'étudiant 36) :

Ex. 49. niveau I :

Salman Rushdie, un écrivain anglais, est soudain devenu connu par tout le monde, surtout par les Musulmans. Ils l'accusent d'avoir insulté leur religion dans le livre "Versets sataniques".

Supposons qu'un écrivain musulman aurait écrit un livre insultant par exemple l'église catholique. Il n'aurait pas été bien respecté parmi les prêtres et les cures, mais le pape n'aurait sûrement pas donné l'ordre de le tuer. Au contraire, l'imam Khomeiny l'a fait et tel est son pouvoir que même les petits enfants seraient prêts à tuer Salman Rushdie.

En Finlande, comme toujours, le gouvernement et le président refusent de dire leurs vraies opinions. On dit des phrases qui serviraient aussi bien à décrire le temps qu'il faisait hier. Heureusement les écrivains ne se contentent pas d'écouter ces "déclarations", ils se sont réunis pour garantir la liberté des écrivains d'écrire ce qu'ils veulent. Cela s'est passé dans autres pays aussi.

Quand même Salman Rushdie a déjà causé une centaines de morts avec son livre dans les manifestations pour ou contre cet ouvrage. Vaut-il mieux mourir pour ses opinions ou ne pas en avoir du tout, ce qui est le cas en Finlande ? Ou bien a-t-on besoin de rushdies pour le savoir ?

Niveau II :

Pourquoi devient-on écrivain ? Qu'est-ce qui pousse une personne à s'exprimer par la littérature ?

Dans la vie chacun cherche ses limites. L'écrivain les cherche avec les moyens de la littérature en cherchant les limites du monde en même temps. Heureusement il arrivera à ses propres limites d'abord et le monde restera toujours ouvert.

La littérature est un moyen de sortir de notre univers personnel, de l'élargir vers les niveaux imaginaires et inconscients si on est l'écrivain et d'y ajouter des éléments nouveaux des autres univers si on est le lecteur. Le monde n'est pas donc trop petit, il est même trop grand pour qu'une personne puisse en découvrir une petite partie et l'ajouter dans son univers pendant sa vie trop courte.

Antonine Maillet dit : "J'écris pour achever le monde". Pour lui la tâche de la littérature, c'est de compléter le travail de la création. L'écrivain cherche à compléter son bonheur et le bonheur des autres et il pourra certainement apporter des grains de bonheur ou des indices pour le trouver à ceux qui le cherchent en vain. Mais il ne réussira jamais à achever le monde qui restera toujours ouvert et inachevé pour que de nouveaux écrivains veuillent l'achever.

9. CONCLUSION

9.1. Récapitulation

Le but de notre travail a été d'étudier et de décrire la compétence lexicale des étudiants finlandais en français. Nous avons effectué une étude longitudinale visant à observer le développement éventuel de la langue des apprenants. La recherche a eu aussi un but pédagogique ; nous avons examiné la production écrite des apprenants de manière à ce que les résultats permettent de réunir des informations importantes sur les problèmes d'apprentissage du vocabulaire, et de trouver des moyens pour faciliter l'apprentissage. Nous avons voulu donner une image plus complète de la production des étudiants en ayant recours aussi bien à une analyse quantitative que qualitative. A partir de ces analyses, nous avons examiné comment et à quel point un séjour d'échange universitaire en France a pu influencer la compétence lexicale de nos étudiants.

Notre corpus est composé de deux parties: la partie primaire est constituée de 106 essais rédigés par des étudiants finlandais lors de l'examen des connaissances langagières. La moitié de ces essais (niveau I) ont été écrits avant un séjour linguistique effectué à l'Université de Provence Aix-Marseille I (l'université d'Aix-en-Provence par la suite) dans le cadre d'un échange avec notre université, l'autre moitié (niveau II) a été faite à la fin des études de licence après ce stage linguistique d'une durée d'un semestre à une année universitaire entière. La partie secondaire regroupe 50 essais (2x25) écrits par des étudiants n'ayant pas étudié à Aix ; les critères de collecte ont été les mêmes que pour le premier corpus. Cette seconde partie a formé le groupe témoin, c'est-à-dire le groupe de contrôle de notre corpus primaire.

Les hypothèses de départ dans cette étude ont été les suivantes :

1. Le séjour linguistique dans un cadre universitaire améliore les connaissances lexicales plus efficacement qu'un simple séjour linguistique.
2. La production devient plus naturelle : entre autres, l'utilisation des organisateurs pragmatiques augmente dans la production.
3. Des traits du langage parlé s'immiscent dans la production écrite.

Dans la partie théorique qui se divise en cinq chapitres (2. – 6.) - le sixième étant le chapitre explicitant nos méthodes - nous avons présenté le cadre théorique de

notre étude. La recherche a pour cadre la linguistique appliquée, et plus précisément la question de l'interlangue, c'est-à-dire la langue des apprenants. Nous avons d'abord traité le concept de l'interlangue (IL) en tant qu'hypothèse de recherche et étudié ses caractéristiques. A l'instar de Giacobbe (1992), nous avons adopté le terme d'interlangue (de Selinker). Les IL sont systématiques, mais en même temps en évolution constante ; en d'autres termes, elles sont instables, changeantes avec le temps. Elles se transforment au fur et à mesure qu'on s'achemine vers la compétence de L2 (cf. Faerch et al. 1984). Selon Ellis (1992), il existe deux types de variabilité principaux : systématique et non-systématique, eux-mêmes divisés en sous-catégories. Il va de soi que la variabilité systématique est explicable et prévisible. Elle a deux sous-catégories : individuelle et contextuelle. La variabilité non-systématique, aléatoire et non-prévisible, peut être divisée encore en variabilités libre et de performance. Cette variabilité libre, d'après Ellis (1992) et Corder (1980), est une caractéristique importante dans l'évolution des IL. Giacobbe (1990) parle de l'apprenant qui construit hypothétiquement sa théorie pour développer ses interactions exolingues. En outre, dans ce chapitre, nous avons évoqué la discussion sur l'état actuel de la recherche interlinguale qui se dirige plutôt vers les travaux acquisitionnels du langage oral et la collaboration métalinguistique en interaction (cf. Véronique 1992).

Le chapitre 3 présente l'interlangue du point de vue pédagogique. Nous avons examiné la problématique autour des erreurs des apprenants, car elles sont jugées primordiales pour l'apprentissage d'une langue étrangère. L'apprenant, pour ainsi dire, teste ses hypothèses sur la langue et les révisé afin de s'approcher des normes de la langue-cible. L'étude des erreurs ne suffit néanmoins pas à révéler tous les aspects de la langue des apprenants. Il faut s'appuyer sur une autre étude sur la compétence communicative pour y arriver, c'est-à-dire la compétence qui inclue aussi bien les connaissances linguistiques d'une langue que la compétence stratégique, socioculturelle, référentielle et discursive (cf. Canale & Swain 1980 ; Moirand 1990). Il est vrai que l'enseignement des langues a pour objectif l'acquisition de cette compétence (Porquier 1984). Faerch et al. (1984) postulent également que l'incorporation de la conscience métacommunicative dans l'enseignement des langues serait pertinente. Nous avons trouvé cette exigence bien justifiée, car ainsi l'apprenant pourrait peut-être diminuer l'incertitude envers l'apprentissage et ainsi aboutir à une communication réussie. Dans le cas des

problèmes de communication, l'apprenant peut recourir, de façon consciente ou inconsciente, aux différentes stratégies de communication pour transmettre son message soit oralement, soit par écrit. L'enseignant peut encourager les étudiants à utiliser ces stratégies, mais d'une façon modérée, car l'utilisation extrême des stratégies de communication peut en réalité inhiber l'acquisition d'une langue étrangère. Une fois que les besoins de la communication sont assurés, la fossilisation, c'est-à-dire la stagnation de l'IL à un niveau erroné, peut entrer en jeu et interrompre le développement interlingual. Ce développement interlingual a été détecté ici par une étude longitudinale en examinant deux examens différents avant et après un séjour linguistique des apprenants à l'étranger. A l'étranger, l'apprentissage s'est passé dans un milieu naturel, mais en partie aussi dans un milieu institutionnel ; néanmoins, il faut se souvenir du fait qu'entre les deux examens, la plupart des étudiants ont également participé à l'enseignement de la production écrite qui est dispensé par notre département d'études françaises.

Dans le chapitre 4, nous avons traité des caractéristiques du langage oral versus le langage écrit. Quand on fait de la recherche sur le langage écrit, il faut prendre en considération le fait que le processus de l'écriture est déjà en partie une activité métalinguistique à cause de la nature même du processus de l'écriture (Bouchard 1994). A la différence du langage oral, le scripteur a plus de temps pour réfléchir sur sa production et pour la réécrire, mais en même temps, il doit être plus explicite dans sa production pour que le message soit compris comme voulu ; cela est dû en partie au fait que le lecteur ne peut pas recourir au feed-back direct pour avoir des renseignements supplémentaires sur le contenu (cf. Frederiksen & Dominic 1981). Le sujet écrivant doit également prendre en considération la fonction communicative du texte, c'est-à-dire à qui adresser le message voulu (cf. Moirand 1993). Chez nos étudiants, il s'agissait d'écrire une dissertation le mieux possible sur un thème optionnel dans une situation d'examen. Le but idéal était d'écrire en un français compréhensible s'approchant de la production d'un Français natif. Et également essayer de convaincre l'évaluateur de ses capacités communicatives à l'écrit. Nous avons également justifié dans ce chapitre notre point de vue sur la grammaire du lexique qui, à notre avis, existe bel et bien au sens large.

Le chapitre 5 est consacré au sujet central de notre étude, à savoir le lexique. Nous avons évoqué la relation entre les recherches faites sur le lexique et la syntaxe, car

ces deux pôles se mêlent constamment. Il existe une sorte de va-et-vient indispensable entre la syntaxe et le lexique (Tréville & Duquette 1996). Il faut cependant noter que du point de vue de la communication, les connaissances lexico-sémantiques sont primordiales par rapport aux connaissances grammaticales ; quand la syntaxe manque, le lecteur ou l'interlocuteur a recours à son intuition pour former une interprétation acceptable, ce qui peut s'avérer être faux à la fin. Les erreurs concernant la sélection lexicale sont considérées plus gênantes quant à la compréhension que les erreurs grammaticales (Carter 1987). Ceci témoigne de l'importance du vocabulaire. Ainsi, à partir du milieu des années 80, le vocabulaire est redevenu pertinent dans l'enseignement des langues. Nous avons également abordé la discussion sur la difficulté du choix du vocabulaire à apprendre et à enseigner. Du point de vue de l'apprentissage, le plus important est que l'apprentissage du vocabulaire ne devrait jamais se réduire à une simple mémorisation de mots nouveaux. Quant à l'enseignement du vocabulaire, nous postulons, à l'instar de Galisson (1981), qu'il faut enseigner aux étudiants à les apprendre. Autrement dit, les étudiants devraient s'orienter vers un auto-apprentissage plus réfléchi, surtout au niveau universitaire (cf. Sautermeister 1989). Notre choix pour le concept **mot** dans la présente étude a été déterminé à partir de notre méthode utilisée, c'est-à-dire que nous avons eu recours au logiciel **WordCruncher** qui compte chaque item séparé comme un mot isolé. Il s'ensuit de notre discussion sur la définition du mot qu'il est déterminé d'une façon orthographique pure dans l'analyse quantitative. Du point de vue pédagogique, nous avons souligné l'importance de traiter les mots, à chaque fois, comme des unités lexicales, c'est-à-dire des unités ayant une forme lexicale et un sens (cf. Bogaards 1994).

En ce qui concerne l'organisation générale des textes étudiés, nous avons présenté brièvement l'analyse textuelle du point de vue lexical dans le chapitre 6. Ce qui nous a intéressé dans cette étude, c'est le rapport étroit entre l'écriture, le lexique et la dimension textuelle. La cohérence textuelle (nous n'avons pas fait de distinction entre les termes de **cohésion** et de **cohérence**) est une partie problématique dans la production écrite, car les apprenants ne connaissent pas toujours suffisamment de moyens linguistiques pour pouvoir lier les phrases ensemble d'une manière logique et argumentative (cf. Reichler-Béguelin 1988). Une partie essentielle dans l'amélioration des capacités d'écriture est justement l'amélioration des capacités textuelles qui favorisent l'augmentation de la

cohérence textuelle et en même temps celle de la compréhension textuelle. L'écriture étant par nature une activité métalinguistique, nous pouvons remarquer que, au fur et à mesure de l'apprentissage, l'apprenant laisse la priorité aux capacités métapragmatiques et métatextuelles (Gombert 1991). Dans notre étude, nous parlons du développement métadiscursif, c'est-à-dire des marques de séquentialité qui facilitent la tâche de l'interprétation (cf. Charolles 1988) ; ceci se manifeste par l'augmentation de l'utilisation des organisateurs pragmatiques. Nous avons défini les organisateurs pragmatiques, primo, comme des moyens pour découper le texte en paragraphes et la ponctuation, secundo, comme des marqueurs explicites dans le texte jouant le rôle de marqueurs de relais d'une phrase à l'autre. Nous avons compilé notre catégorisation principalement à partir des classifications d'Adam (1990), de Lundquist (1980, 1983) et de Riegel et al. (1997) sur les connecteurs et sur les organisateurs textuels.

Le chapitre 7 présente nos méthodes dans cette recherche. Nous avons adopté la méthode quantitative utilisée par Moira Linnarud dans sa recherche (1986), mais nous avons de toute façon dévié de sa méthode, car nos objectifs n'étaient pas les mêmes. Nous avons ainsi examiné les facteurs linguistiques quantitatifs suivants : le nombre de mots, la variation lexicale, la densité lexicale, l'individualité et la sophistication lexicales (cf. infra). L'analyse quantitative du corpus a été effectuée partiellement à l'aide du logiciel WordCruncher qui a facilité le dépouillement des textes. L'individualité lexicale et la sophistication lexicale ont été examinées en partie manuellement ; contrairement aux autres facteurs quantitatifs, les "mots réellement uniques" ont été traités sous forme lemmatisée afin d'examiner la sophistication lexicale. Pour ces facteurs, nous avons calculé des corrélations afin de voir si les changements entre les niveaux sont statistiquement significatifs ; nous avons eu recours aux coefficients de **Bravais-Pearson** et de **Spearman** (cf. Dugast 1980) que nous avons calculés à l'aide du logiciel **SPSS**. En outre, un **t-test** et un **test ANOVA** ont été effectués pour indiquer s'il existe des différences dans les changements du niveau I au niveau II entre les deux corpus. De plus, les corrélations entre les facteurs linguistiques quantitatifs ont été calculées séparément aux deux niveaux en recourant aux coefficients de Bravais-Pearson et de **Kendall**. Ces derniers calculs ont été effectués dans le département des statistiques. L'analyse quantitative a été complétée par une analyse qualitative qui consistait en analyse des erreurs dans une perspective de *parole*, en jugeant leur impact sur la lisibilité (cf. Péry-Woodley 1993). L'étude sur des organisateurs

pragmatiques fait partie de cette analyse qualitative.

9.2. Résultats principaux

A partir de notre analyse, nous avons pu conclure que le taux moyen du nombre de mots augmentait dans les dissertations au niveau II et cela dans les deux corpus, à savoir 209 et 214 respectivement dans le corpus d'Aix, et 215 et 240 respectivement dans le corpus témoin. L'augmentation est plus grande dans le corpus témoin, mais tout ce changement semble, en effet, être plutôt lié à la variabilité individuelle. Le nombre de mots a, néanmoins, une valeur importante, car elle influe, entre autre, la valeur de la variation lexicale (VL). Le pourcentage de la variation lexicale élevée indique qu'il y a beaucoup de mots lexicaux différents dans le texte. La longueur de l'essai a un effet sur le pourcentage de la VL parce qu'il est plus difficile d'éviter la répétition dans un essai plus long et vice versa. D'après l'analyse des pourcentages de la VL, nous avons pu constater que la variation lexicale s'améliore après le séjour linguistique, alors que dans le corpus témoin, l'amélioration est moins évidente, justement à cause de la longueur des dissertations qui a influé sur les résultats. Etroitement lié à la VL est le facteur de la VL/M qui indique la variation lexicale adaptée à la longueur de la dissertation. Nous avons supposé que quand la VL augmente, il en va de même quant à la VL/M, et vice versa. Ceci était effectivement à remarquer dans le corpus d'Aix et dans le corpus témoin. Statistiquement les corrélations de la VL au niveau I et au niveau II sont faiblement significatives dans le corpus d'Aix, tandis que dans le corpus témoin, ce n'est pas le cas. Par contre, les corrélations de la VL/M sont modérément significatives entre les deux niveaux dans le corpus témoin, mais seul le coefficient de Bravais-Pearson est peu significatif dans le corpus d'Aix. Bien que ces résultats ne soient pas statistiquement très significatifs, nous avons pu remarquer que dans le corpus d'Aix en particulier, même un séjour plus court semble améliorer les résultats. Ceci semble logique car les étudiants ont eu accès à un input linguistique varié aussi bien oralement que par écrit. Ce résultat est positif car une variation lexicale élevée est souhaitable chez les apprenants du point de vue de la cohérence du texte écrit (cf. Amaud 1984 ; Charolles 1988). Néanmoins, la VL montre uniquement si le vocabulaire d'un essai est plus au moins varié, mais elle ne révèle pas sa qualité, c'est-à-dire sa sophistication (cf. Tiilikainen & Ylinentalo 1989).

La densité lexicale indique la quantité de mots lexicaux dans les dissertations. Dans un texte écrit par un autochtone, la limite de 40 % de mots lexicaux est normale dans un texte non-interactif. Nous avons ainsi comparé les résultats des étudiants à cette limite. Si la valeur passe largement au-dessous de cette limite, le texte risque de s'appauvrir à cause d'un grand nombre de mots grammaticaux, mais une fois que la limite est dépassée, la croissance de la DL n'est pas nécessairement accompagnée d'une réussite : une grande quantité de mots lexicaux, entre autres de substantifs et d'adjectifs, peut rendre le texte moins lisible. Dans les deux corpus, les taux de moyen décroissent légèrement au niveau II, restant toujours au-dessous de la limite de 40 %. Les écarts individuels sont plus grands ; le pourcentage minimum est de 33,8 % au niveau I et de 35,3 % au niveau II (le corpus d'Aix), et respectivement de 36,0 % et de 32,8 % (le corpus témoin). Les pourcentages maximums étaient respectivement de 48,9 % et de 50,6 % dans le corpus d'Aix, et de 47,9 % et de 47 % dans le corpus témoin. Seul le coefficient de Bravais-Pearson montre une faible corrélation entre la DL au niveau I et II dans le corpus d'Aix. Grâce à ces données, nous n'avons pas pu tirer de conclusions directes sur l'appauvrissement du vocabulaire des dissertations ou sur le foisonnement des mots lexicaux – normalement dans un texte écrit par un autochtone, la quantité des mots grammaticaux dépasse celle des mots lexicaux. Nous avons donc repris ces dissertations déviantes par rapport aux moyennes et les avons étudiés de plus près dans l'analyse qualitative (cf. infra).

Les facteurs quantitatifs de l'individualité lexicale et de la sophistication lexicale sont étroitement liés l'un à l'autre. L'individualité, c'est-à-dire, l'originalité du texte a été mesurée en calculant le nombre de mots à occurrence unique chez un étudiant en comparaison avec tous les mots lexicaux dans l'ensemble du corpus, et en calculant le pourcentage de ces mots-là dans une dissertation en question. Nous les avons appelés des **mots uniques**. Nous avons encore exclu une partie de ces mots de l'analyse, pour avoir ainsi des hapax legomena que nous avons appelés des **mots réellement uniques** ; afin de pouvoir examiner la sophistication lexicale des textes. La sophistication lexicale a été effectuée manuellement en comparant les mots réellement uniques des apprenants avec le *Dictionnaire des fréquences* et le *Français fondamental* pour voir le degré de difficulté de leurs dissertations. Les résultats des pourcentages de l'individualité lexicale doivent être interprétés avec prudence, car le nombre de mots concernés est restreint ; de cette manière, un ajout ou une réduction d'un mot se voit clairement dans les

pourcentages. De toute façon, ces résultats montrent des tendances dans le corpus ; il y a une hausse quant aux mots réellement uniques dans les deux corpus. Ceci est statistiquement étayé par une corrélation modérée (le coefficient de Bravais-Pearson) entre les deux niveaux dans le corpus d'Aix – les autres corrélations ne sont pas statistiquement significatives. L'analyse de la sophistication lexicale est constituée de deux parties : primo, nous avons comparé les 20 mots les plus fréquents dans les deux corpus avec le *Dictionnaire des fréquences*, et secundo, nous avons comparé les mots réellement uniques au *Dictionnaire des fréquences* et au *Français fondamental*. Il ressort de cette analyse que les onze premiers mots étaient les mêmes dans les deux corpus et dans le *Dictionnaire de fréquences*, mais que leur rang n'était pas tout à fait le même, sauf pour le premier (la préposition *de*). Nous avons pu conclure que les deux corpus respectent assez bien les tendances générales concernant l'utilisation des mots fréquents, surtout le corpus d'Aix. Deuxièmement, nous avons pu déduire de cette analyse que plus il y a de mots dans le corpus, mieux les mots sont distribués comme ils le sont généralement dans les textes français. Autrement dit, les dissertations des étudiants s'approchent globalement des textes natifs quant à l'utilisation des mots les plus fréquents, c'est-à-dire les mots grammaticaux.

Dans le corpus d'Aix, il y a 971 mots lexicaux réellement uniques dont 432 au niveau I et 539 au niveau II, tandis que dans le corpus témoin, les nombres sont respectivement 292 et 320, au total 612. La plupart des mots réellement uniques sont des substantifs dans les deux corpus aux deux niveaux. Le nombre des adjectifs augmente le plus dans les deux corpus en passant du niveau I au niveau II. Ce résultat rejoint le postulat de Ménard (1983) qui dit que dans un texte riche, la fréquence relative des substantifs par rapport aux verbes, et celle des adjectifs par rapport aux substantifs sont en général plus élevées que dans un texte pauvre. Bref, nous pouvons conclure que le vocabulaire dans les dissertations des apprenants est devenu plus riche au niveau II. Nous avons supposé supra que l'individualité et la difficulté lexicales sont étroitement liées l'une à l'autre, c'est-à-dire que plus le texte est original, plus le niveau de sa sophistication est élevé. Dans la présente étude, cela se manifeste de la manière suivante dans les deux corpus : plus il y a de mots uniques, moins il est probable de trouver tous ces mots dans le *Français fondamental* dont le contenu est principalement fondé sur la fréquence des mots. En outre, si l'étudiant s'exprime avec beaucoup de mots réellement uniques, même les mots d'une fréquence élevée qu'il utilise ne se

trouvent pas dans le *Français fondamental* contre toute attente. Les mots utilisés par les étudiants qui ne figurent ni dans le *Français fondamental*, ni dans le *Dictionnaire des fréquences*, sont principalement des noms propres, des mots liés à une culture spécifique ou encore des mots récents. En outre, la correction orthographique des mots réellement uniques s'améliore en général au niveau II. Il semble également que les apprenants ayant plus de mots réellement uniques dans leurs dissertations les ont appris en dehors du programme scolaire qui se base, quant au vocabulaire français, sur le *Français fondamental*. Nous pouvons conclure que l'utilisation des mots uniques est premièrement dû à des variabilités individuelles. Mais il se peut qu'il s'agisse de différentes stratégies de communication ; certains étudiants prennent plus facilement des risques : bien qu'ils ne connaissent pas le mot exact, ils essaient néanmoins d'utiliser le mot qu'ils croient connaître. D'autre part, certains étudiants utilisent des mots réellement uniques pour ainsi dire "sans surprise" : ils n'utilisent que des mots qu'ils connaissent bien ou des mots avec le moins de risques possibles.

Les résultats statistiques montrent que, dans l'ensemble du corpus, c'est-à-dire le corpus d'Aix et le corpus témoin, il y a une différence statistiquement significative dans l'évolution du niveau I au niveau II, quant à certains facteurs quantitatifs, notamment la valeur du nombre de mots, de l'IIe (a) et de l'IIe (b) – ces derniers sont des mots uniques et des mots réellement uniques respectivement. Ce résultat soutient les résultats que nous avons présentés supra. Pour sa part, le test ANOVA a effectué les calculs entre les deux corpus et montre qu'il n'y a pas de différence significative entre les deux corpus quant aux changements de ces facteurs quantitatifs. Bref, nous pouvons conclure que les changements entre les deux corpus sont presque identiques. A l'aide du test ANOVA, on a aussi testé l'existence d'une différence dans les changements du niveau I au niveau II qui serait liée au fait que le français est une matière principale ou non. Les différences ne sont pas non plus statistiquement significatives. En outre, dans le corpus d'Aix, on a effectué le test ANOVA pour voir s'il y avait une différence dans les changements du niveau I au niveau II parmi les étudiants qui ont passé soit un ou deux semestres à l'université d'Aix-en-Provence. Le test ANOVA montre qu'il n'y a pas de différence statistiquement significative dans les changements qui seraient liée à la durée du séjour. Nous avons également calculé des corrélations entre les différents facteurs quantitatifs séparément aux deux niveaux. Bref, il ressort de l'analyse que la variation lexicale semble bien corrélérer avec les autres

facteurs, et notamment avec les mots uniques et réellement uniques, comme nous l'avons déjà évoqué supra. Les mots réellement uniques (IIe (b)) corrélaient également bien avec les autres facteurs, sauf la densité lexicale et le nombre de mots. Ainsi, nous pouvons en conclure que les valeurs de la variation lexicale et des mots réellement uniques sont des facteurs quantitatifs centraux dans les calculs, et que la quantité des mots réellement uniques est plus liée à la valeur VL qu'à la valeur des nombres de mots. Ces données statistiques infirment en partie notre première hypothèse qui postule que le séjour linguistique dans un cadre universitaire améliore les connaissances lexicales plus efficacement qu'un simple séjour linguistique. Pourtant, ce résultat ne suffit pas à révéler la richesse lexicale des dissertations (cf. infra).

Les erreurs les plus évidentes, ou plutôt les formes erronées, dans les deux corpus sont les fautes d'orthographe, et surtout les fautes dans l'utilisation des accents. Dans quelques cas, il s'agit de la fossilisation (dans le sens de Selinker) d'une forme erronée, c'est-à-dire que l'étudiant utilise tout le temps la forme erronée dans la dissertation, soit au niveau I ou II, ou il l'utilise seulement à un des niveaux. Il semble parfois qu'il s'agit de la variabilité libre (Ellis 1992) : l'apprenant utilise une forme du mot probablement sans savoir comment l'utiliser correctement. La forme erronée peut être un signe pour tester une hypothèse sur la langue afin d'élaborer son propre système s'approchant de la L2 (Giacobbe 1990, 1992). Ou cela peut être aussi une sorte de stratégie communicative quand l'étudiant n'est pas sûr de la forme correcte. Une catégorie des fautes d'orthographe s'explique par l'interférence, le transfert négatif (cf. Larsen-Freeman & Long 1992). Il s'agit d'influence de la L1 ou d'autres langues étrangères apprises auparavant (ou en même temps que le français comme matière principale ou secondaire). Il y a également une raison primordiale pour les fautes d'orthographe : la variabilité de performance (Ellis 1992). Cela signifie que la situation d'examen peut rendre les étudiants nerveux ce qui les amène à écrire de nombreux lapsus ; ceci est probablement un signe de formes non automatisées de la L2. Quant à certains étudiants, nous pouvons plutôt parler d'étourderies, surtout quand il y a beaucoup de fautes aux deux niveaux. De plus, nous avons remarqué deux catégories principales dans les erreurs de vocabulaire: les formes erronées et les erreurs de sens. Les formes erronées se chevauchent en partie avec les fautes d'orthographe. Outre les fautes d'accents, les apprenants conjuguent mal les verbes, entre autres, ou bien les formes erronées peuvent être liées à la

surgénéralisation du système phonétique français. En ce qui concerne les erreurs de sens, elles proviennent de la sélection lexicale qui est probablement due au manque de connaissance exacte du sens du mot en question ou de l'acceptabilité du terme (cf. Enqvist 1977), ce qui signifie que l'apprenant connaît probablement la signification du terme, mais l'utilise dans une situation inappropriée. Nous avons étudié les erreurs, à l'instar de Péry-Woodley (1993), dans une perspective de *parole*, notamment en examinant leur effet sur la lisibilité. En général, les erreurs concernant la sélection lexicale sont considérées plus gênantes quant à la compréhension que les erreurs grammaticales (Carter 1987). Nous postulons qu'il en est sûrement de même quant à la production orale, mais qu'à l'écrit, même les fautes peuvent avoir un effet négatif sur la compréhension et sur les réactions du lecteur. S'il faut examiner et analyser la production en détail afin de donner aux étudiants de l'information exacte sur leur production, même les fautes mineures peuvent gêner la lecture et avoir un effet négatif sur l'évaluation. C'est ce que Linnarud (1986) a également démontré dans sa recherche. Néanmoins, il faut noter que le plus pertinent semble être de toute façon le contenu du texte écrit, c'est-à-dire que si l'étudiant s'exprime et argumente bien ses points de vue, les erreurs ont une moindre influence sur l'évaluation. Nous ne pouvons pas mettre en évidence de grandes différences entre les deux corpus quant aux erreurs lexicales dans la production. Nous voudrions plutôt dire qu'il s'agit de variabilités individuelles. Cependant, nous pouvons remarquer un développement du niveau I au niveau II en général. Autrement dit la correction lexicale est meilleure, malgré les fautes d'orthographe (surtout les accents), et aussi les notes données par les professeurs natifs sont normalement meilleures après le séjour à l'étranger. Ces notes nous ont aidés dans l'analyse du corpus bien qu'elles n'aient pas été un objet d'étude. A notre avis, ce développement est étroitement lié à l'utilisation des organisateurs pragmatiques car une bonne cohérence textuelle augmente la compréhension et la lisibilité du texte écrit.

En ce qui concerne l'utilisation du découpage en paragraphes et des signes de ponctuation, nous pouvons constater que, dans les deux corpus, la plupart des étudiants découpent leur texte en plusieurs paragraphes. Ainsi, nous concluons que les apprenants savent utiliser le découpage du texte en paragraphes, qui est une activité métalinguistique facilitant la lecture et organisant le texte hiérarchiquement (Bessonnat 1988). En outre, nous pouvons conclure à partir du corpus que, plus le texte est long, plus il y a de paragraphes, ce qui semble

logique. Nous pouvons également constater que l'utilisation des signes de ponctuation est déjà stabilisée au niveau I, ce qui semble normal, car ce sont des apprenants adultes – nous n'avons pas commenté si l'utilisation des signes de ponctuation était correcte. On peut quand même remarquer une légère évolution chez certains étudiants, c'est-à-dire qu'au niveau II, ils ponctuent leur texte par d'autres signes de ponctuation que le point (seule marque de ponctuation au niveau I), ou qu'ils utilisent un éventail plus large de signes de ponctuation qu'au niveau I. De plus, les signes de ponctuation autres que le point sont souvent utilisés plutôt d'une façon métacommunicative et sont étroitement liés à la façon des apprenants de structurer et commenter le texte (cf. Fayol 1989). La différence majeure entre les deux corpus est que dans le corpus d'Aix, les étudiants emploient une variété plus large de signes de ponctuation que ne le font les étudiants du corpus témoin. Mais, il y a aussi des différences individuelles. Il en va de même des autres organisateurs pragmatiques. Principalement, les étudiants, étant des apprenants adultes, disposent déjà de techniques pour rendre la production plus naturelle en ayant recours à leurs connaissances préalables (cf. Noyau & Vasseur 1986 ; Perdue 1993). En général, il ressort de l'analyse qu'il y a certaines tendances d'évolution à noter entre les deux niveaux. Entre autres, les apprenants des deux corpus utilisent des organisateurs pragmatiques aux deux niveaux étudiés, mais l'éventail des organisateurs est plus diversifié au niveau II (cf. Kerr-Barnes 1998). De plus, l'utilisation "correcte" de certains organisateurs semble augmenter. En outre, ce qui semble augmenter dans l'utilisation des organisateurs, ce sont les marqueurs métatextuels, c'est-à-dire les modalisateurs argumentatifs, les commentaires métatextuels et les introducteurs ce qui se manifeste par une aisance plus grande dans l'écriture (cf. Gombert 1991). Nous pouvons néanmoins constater une petite différence entre les deux corpus en faveur du corpus d'Aix. Nous pourrions supposer que ceci est lié au fait que dans le corpus d'Aix, il y a 40 apprenants sur 53 qui étudient le français comme matière principale, tandis que dans le corpus témoin, il n'y en a que 12 sur 25 ; ainsi, les étudiants de la matière principale auraient à leur disposition un accès plus facile aux organisateurs pragmatiques français. En somme, les résultats du corpus témoin semblent aller dans le même sens que ceux du corpus d'Aix - sauf que la variété des organisateurs semble moindre - infirmant ainsi en partie notre hypothèse sur les différences entre les corpus. De toute façon, nous pouvons conclure que les capacités métatextuelles des apprenants s'améliorent comme prévu pendant le séjour linguistique à l'étranger. Il faut remarquer néanmoins que

tous les changements ne sont pas dus au séjour linguistique en France, car l'enseignement de français dans notre département consiste également en cours de production écrite. L'hypothèse que la production devient plus naturelle, entre autres, grâce à l'augmentation des organisateurs pragmatiques est ainsi vérifiée.

Quant au développement interlingual de la cohérence textuelle, dans l'ensemble, nous discemons une cohérence plus élevée au niveau II qu'au niveau I. Nous pouvons le constater clairement dans les dissertations des étudiants de matière secondaire. Au niveau II, la cohérence ne se trouve pas uniquement au niveau phrastique, mais par contre, il y a un enchaînement cohérent également entre les paragraphes. De plus, dans le corpus d'Aix, au niveau II, nous pouvons remarquer, une petite liberté concernant le sujet de dissertation qui ne se voit pas dans le corpus témoin ; certains étudiants écrivent, certes, sur le sujet donné, mais leurs pensées s'éloignent du sujet original. Nous supposons que cette liberté pourrait être liée aux cours dispensés à l'université d'Aix-en-Provence, même si nous ne pouvons pas affirmer cette influence. Notre troisième hypothèse de travail postule que des traits du langage parlé s'immiscent dans la production écrite. En effet, ceci était à noter dans les dissertations du niveau II. Nous remarquons, entre autres, que du point de vue du vocabulaire appartenant à un registre du langage familier ou parlé, ce sont essentiellement les étudiants qui ont passé deux semestres à l'université d'Aix-en-Provence ou ont séjourné plus longtemps dans un pays francophone qui ont recours à ce vocabulaire.

En outre, nous pensons que le choix entre les différentes façons d'organiser le texte ainsi que la sélection lexicale sont parfois les preuves des stratégies de communication, conscientes ou inconscientes. En effet, quand l'étudiant prend des risques avec le vocabulaire, il s'agit plutôt d'une stratégie d'atteinte, alors que quand il a recours au vocabulaire plus simple, on pourrait parler d'une stratégie de réduction (Faerch et al. 1984). Cependant, nous ne pouvons pas être sûre de cette dernière hypothèse car l'évitement passe inaperçu ; il serait plus prudent de parler uniquement d'une stratégie de communication que l'étudiant utilise afin de faire passer le message. Nous pouvons parler également de stratégies métacommunicatives quand l'apprenant a recours à un moyen quelconque d'une manière consciente ; par exemple, quand l'étudiant écrit un mot de différentes façons pour faire croire au professeur qu'il connaît le mot, mais par mégarde, a commis une erreur.

9.3. Bilan et perspectives

De tout ce qui précède, nous pouvons finalement conclure qu'il est vrai que la production écrite des apprenants devient plus naturelle dans l'examen des connaissances langagières auquel les apprenants participent à la fin de leur année de licence. Cela est dû partiellement au fait que leurs compétences métalinguistiques et métatextuelles s'améliorent. Le séjour à l'étranger, soit dans un milieu naturel à proprement parlé, soit dans un milieu guidé ou semi-guidé, a un effet positif sur ce développement, même si nous ne pouvons pas oublier l'importance de l'enseignement dispensé au sein de notre département. Il est également vrai que des traits du langage parlé s'immiscent dans la production écrite, surtout chez les étudiants qui ont passé deux semestres dans une université française ou qui ont passé un séjour plus long dans un pays francophone sans étudier dans une université. Ce que nous ne pouvons pas directement conclure à partir de nos résultats, c'est que le séjour linguistique dans un cadre universitaire améliorerait les connaissances lexicales plus efficacement qu'un simple séjour linguistique. Nous avons supposé que les résultats du corpus d'Aix, surpasseraient en quantité et en qualité les résultats du corpus témoin. Cela n'est que partiellement vrai, car les analyses statistiques ont infirmé notre hypothèse de la supériorité quantitative. Néanmoins, les statistiques nous ont aidée à prouver les tendances générales dans les deux corpus concernant les changements du niveau I au niveau II. Sur le plan qualitatif, nous pensons que les différences entre les deux groupes sont quand même légèrement en faveur du corpus d'Aix, même s'il y a, certes, des variabilités individuelles. De cet état de chose témoigne en partie l'utilisation de l'éventail plus large des organisateurs pragmatiques.

En outre, quand nous parlons des différences entre les deux corpus, il faut se souvenir du fait que nous avons seulement étudié l'utilisation du vocabulaire. Si nous avions également examiné la structure des dissertations, les résultats se seraient probablement avérés fort différents. Nous avons pu remarquer, entre autres, l'interférence du finnois ou d'autres langues étudiées dans l'ordre des mots et dans l'utilisation des prépositions ; ceci est à remarquer surtout quand le niveau des capacités langagières de l'étudiant est moins élevé. Il en va de même avec les différences dans les deux corpus quant au langage parlé. Dans la présente étude, nous n'avons discuté que de la production écrite où le développement des capacités langagières semble un apprentissage à long terme, mais si nous avions

étudié la langue orale, les résultats seraient largement en faveur du corpus d'Aix. Il en est également de même à l'intérieur du corpus d'Aix. Bien que les résultats statistiques n'aient pas montré de différences significatives dépendant de la durée du séjour, nous savons par expérience que les résultats globaux dans l'examen des connaissances langagières indiquent une amélioration plus nette aussi bien dans les capacités de compréhension que dans celle de production chez les étudiants ayant passé toute une année universitaire en France. Surtout, les connaissances de la compréhension et de la production du français parlé s'améliorent. De plus, nous pouvons penser que les capacités d'écriture s'amélioreraient encore plus si les étudiants suivaient l'enseignement du niveau de maîtrise en France ; dans la présente étude, nous avons exclu ces étudiants du corpus.

Nous pouvons aussi déduire de cette analyse que bien que les résultats statistiques ne montrent pas de différence significative entre les étudiants ayant le français comme matière principale et ceux de matière secondaire, au moins, la cohérence textuelle des étudiants de matière secondaire semble nettement s'améliorer au niveau II. Il semble que le séjour en France, indépendamment de sa durée, ait influencé d'une façon plus positive les étudiants dont le niveau de base était moins élevé, ce qui est souvent le cas chez les étudiants de matière secondaire. Les variabilités individuelles sont grandes, mais normalement il y a rarement des polyglottes qui sont aussi forts dans plusieurs langues en même temps ; en général, la langue de matière principale prime sur les autres langues. Cependant, il faut se souvenir qu'une partie des étudiants a une matière principale autre que les langues et on ne peut donc pas trop généraliser cette hypothèse. Cette amélioration des compétences textuelles est probablement due au fait que les apprenants ont accès à l'input plus varié dans un milieu naturel qui leur manque en Finlande. Étroitement liée à l'accès à cet input varié est également la motivation des apprenants partant à l'étranger. Sans motivation à apprendre, on n'apprend pas, même si l'input linguistique est optimal. On pourrait supposer que la motivation d'apprendre des étudiants qui participent à l'échange ERASMUS/SOCRATES ou partent pour un séjour plus long est plus grande que celle des étudiants qui effectuent un séjour linguistique d'un mois. Cette supposition est partiellement vraie ; pour preuve, nous pouvons voir comment les étudiants du corpus sont répartis : 40 étudiants de matière principale et 13 étudiants de matière secondaire dans le corpus d'Aix, 12 étudiants de matière principale et 13 étudiants de matière secondaire dans le corpus témoin. Donc, ce sont principalement des

étudiants de français de matière principale qui partent étudier dans une université française et qui sont, pour ainsi dire, plus motivés. D'un autre côté, le choix de partir seulement pour un mois n'est pas toujours une question de motivation ; il y a aussi des raisons personnelles, même familiales ou bien financières. Les étudiants n'ont accès à la bourse ERASMUS qu'une fois pendant les études et ils participent donc plus probablement au programme d'échange ERASMUS de leur matière principale.

Ainsi notre étude soulève-t-elle peut-être plus de questions qu'elle ne donne de réponses ou plutôt elle s'ouvre davantage sur de nouvelles perspectives. Même si les résultats statistiques n'ont pas pu clairement prouver l'hypothèse des différences entre les deux corpus, nous n'allons pas rejeter l'approche de l'analyse quantitative. A l'instar de Tarone (1988), nous postulons que les résultats d'une étude qualitative sur l'interlangue devraient être étayés par les études quantitatives afin de pouvoir déterminer la généralité des résultats obtenus parmi les apprenants d'une langue seconde. Pour améliorer notre approche, nous pouvons définir les mots différemment, à savoir selon la définition orthographique, certes, mais en regroupant les organisateurs pragmatiques comme une catégorie à part. De cette manière, nous croyons pouvoir mieux expliciter les différences entre les niveaux et entre les corpus. En outre, pour compléter l'analyse des dissertations, nous pourrions examiner encore plus en détail leur structure et leur cohérence textuelle en ayant recours aux autres moyens de cohérence tels que les mécanismes référentiels, par exemple la pronominalisation. Comme nous l'avons évoqué supra, les connaissances du français parlé de ces étudiants restent à être également examinées ce qui perfectionnerait l'image de leurs capacités langagières du point de vue lexical (cf. Elo 1993).

De plus, ce qui semble également pertinent quant à l'étude du vocabulaire, c'est le côté métacommunicatif. Entre autres, Blum-Kulka & Levenston (1992) postulent qu'à tous les niveaux d'apprentissage les connaissances lexicales des apprenants d'une L2 sont moindres en comparaison avec celles d'un autochtone. Le manque de connaissances sémantico-lexicales se reflète ainsi dans la simplification, consciente ou inconsciente, du vocabulaire. Pour compenser ce déficit, les apprenants ont recours aux connaissances sémantiques de leur langue maternelle. Il s'ensuit que la production de l'apprenant n'est pas toujours appropriée dans une situation de communication comme l'est la production d'un natif (cf. ch. 5.2). Le

postulat de Kellerman (cité dans Ellis 1994) va dans le même sens quand il parle des structures que l'apprenant sent comme transférables d'une langue à une autre à partir des connaissances préalables dans sa langue maternelle. Afin d'expliquer ce phénomène, il serait intéressant d'examiner les processus qui déterminent, entre autres, la sélection lexicale dans la production écrite ou orale.

Pour conclure, il semble que la meilleure façon d'apprendre le vocabulaire reste la méthode traditionnelle, connue depuis déjà longtemps, c'est-à-dire l'acquisition d'une langue dans un milieu naturel étayée par un apprentissage guidé ; on parle alors d'immersion naturelle. Pourtant, nous pouvons aussi améliorer les capacités lexicales et les capacités d'écriture des apprenants par l'enseignement dispensé dans notre département. La lecture est un atout primordial dans l'apprentissage de nouveaux mots et de nouveaux concepts, surtout la lecture consciente, qui est une partie de la compétence métacommunicative des apprenants. La lecture de différents types de textes améliore les connaissances lexicales passives ou réceptives des apprenants. Comme toutes les aptitudes se chevauchent et que les limites entre elles sont floues, il s'ensuit que l'apprenant ayant de meilleures connaissances lexicales passives peut aussi probablement écrire plus aisément dans une langue étrangère. En d'autres termes, les connaissances lexicales passives facilitent également la production active qu'est l'écriture. Il en est également ainsi quant aux organisateurs pragmatiques. Nous pouvons aussi encourager les apprenants à utiliser des stratégies de communication diverses pour faire passer leur message, même si les moyens exacts manquent. Et finalement, nous pouvons les encourager à partir étudier dans une université française afin d'acquérir une bonne base pour leurs futures études en Finlande ou ailleurs.

BIBLIOGRAPHIE

- Adam**, J-M. 1990. *Eléments de linguistique textuelle. Théorie et pratique de l'analyse textuelle*. Deuxième éd. Liège : Pierre Mardaga.
- Adjemian**, C. 1976. "On the nature of interlanguage systems". *Language Learning* 26/2 : 297-320.
- Adjemian**, C. 1983. "The transferability of lexical properties". In S. Gass & L. Selinker (Eds.) *Language Transfer in Language Learning*. pp. 250-268.
- Andersen**, R. W. (Ed.) 1983a. *Pidginization and creolization as language acquisition*. Rowley, Mass. : Newbury House.
- Andersen**, R. W. 1983b. "Transfer to somewhere". In S. Gass & L. Selinker (Eds.) *Language Transfer in Language Learning*. pp. 177-201.
- Arnaud**, P. J. L. 1984. "The lexical richness of L2 written productions and the validity of vocabulary tests". In T. Culhane, C. Klein Braley & D. K. Stevenson (Eds.) *Practice and Problems in Language Testing*. Occasional Papers n. 29. Department of Language and Linguistics. Essex : University of Essex. pp. 14-28
- Arnaud**, P. J. L. 1993. "Estimations subjectives des fréquences des mots en L1 et en L2". In P. J. L. Arnaud & P. Thoiron (Eds.) *Aspects du vocabulaire*. Lyon : Presses universitaires de Lyon. pp. 7-16.
- Arnaud**, P. J. L. & S. J. Savignon. 1997. "Rare words, complex lexical units and the advanced learner". In J. Coady & T. Huckin (Eds.) *Second Language Vocabulary Acquisition*. pp. 157-173.
- Auchlin**, A. 1981. "Réflexions sur les marqueurs de structuration de la conversation". *Etudes de Linguistique appliquée* 44 : ?
- Bakhtine**, M. 1984. *Esthétique de la création verbale*. Traduit du russe par A. Aucouturier. Paris : Gallimard.
- Bange**, P. 1999. "L'usage de la règle dans l'enseignement et l'apprentissage de langue étrangère". In M.L.M.S. (éd.) *Actes du XIe colloque international : Acquisition d'une langue étrangère : perspectives et recherches. Usages pragmatiques et acquisition des langues étrangères*. Paris 19, 20 et 21 avril 1999, France. Disponible à partir du site <http://www.marges-linguistiques.com/>
- Bartning**, I. 1994. "Recherches sur l'acquisition des langues secondes/étrangères et du français langue étrangère – aspects de développement d'interaction et de variation". In G. Boysen (Ed.). 1994. *Actes du XIIIe Congrès des Romanistes Scandinaves*. pp. 23-32.
- Bayley**, R. & D. R. Preston (Eds.) 1996. *Second Language Acquisition and Linguistic Variation*. SiBil 10. Amsterdam/Philadelphia : John Benjamins Publishing Company.
- Besse**, H. & R. Porquier. 1984. *Grammaires et didactique des langues*. Paris : Hatier-CREDIF.
- Bessonnat**, D. 1988. "Le découpage en paragraphes et ses fonctions". *Pratiques* 57 : 81- 105.
- Blanche-Benveniste**, C. 1997. "The unit in written and oral language". In C. Pontecorvo (Ed.) *Writing Development : An interdisciplinary view*. pp. 21-45.
- Blanche-Benveniste**, C. & C. Jeanjean. 1987. *Le français parlé. Transcription et édition*. Institution National de la Langue Française (INALF) (CNRS). Paris : Didier.
- Bley-Vroman**, R. 1989. "What is the logical problem of foreign language learning". In S. Gass & J. Schachter (Eds.) *Linguistic Perspectives on Second Language Acquisition*. pp. 41-68. Cambridge : Cambridge University Press (CUP).
- Blum-Kulka**, S. & E. A. Levenston. 1992. "Universals of lexical simplification". In C. Faerch & G. Kasper (Eds.) *Strategies in Interlanguage Communication*. pp. 119-139. Originally published in *Language Learning* 28 (1978) : 399-415.

- Bogaards, P.** 1980. "Geheugen en woordverwerving in een vreemde taal [Memory and acquisition of the vocabulary in a foreign language]. *Levende Talen* 351 : 272-283.
- Bogaards, P.** 1991. *Aptitude et affectivité dans l'apprentissage des langues étrangères*. Paris : Hatier-CREDIF.
- Bogaards, P.** 1994. *Le vocabulaire dans l'apprentissage des langues étrangères*. Paris : Hatier-CREDIF.
- Bolton, S.** 1991. *Evaluation de la compétence communicative en langue étrangère*. Traduit de l'allemand par Y. Bertrand. Paris : Hatier-CREDIF.
- Bouchard, R.** 1994. "De l'oral à l'écrit en français langue étrangère : les procédés d'intégration discursive". *Bulletin suisse de linguistique appliquée* 59 : 103-125.
- Boyer, H. & M. Rivera.** 1979. *Introduction à la didactique du français langue étrangère*. Paris : CLE International.
- Boysen, G. (Ed.)**. 1994. *Actes du XIIIe Congrès des Romanistes Scandinaves*. Aalborg, 11 – 15 août 1993. Vol. I. Publications of the Department of Languages and Intercultural Studies, Aalborg University. Vol 10. Aalborg: Aalborg University Press.
- Brown, G. & G. Yule.** 1989. *Discourse Analysis*. First published in 1983. Cambridge-New York-Melbourne : CUP.
- Canale, M. & M. Swain.** 1980. "Theoretical bases of communicative approaches to second-language teaching and testing". *Applied Linguistics* 1 : 1-47.
- Candlin, C. N.** 1981. "Preface". In C. James. *Contrastive Analysis*. First published 1980. London : Longman. pp. iii-v.
- Carcedo González, Alberto.** 1998. *La pronunciación del español por hablantes nativos de finés : particularidades de un acento extranjero*. Acta ibero-americana fennica. series Hispano-Americana 2. Helsinki : Instituto Iberoamericano de Finlandia.
- Carroll, J. B.** 1968. "The psychology of language testing". In A. Davies (Ed.) *Language Testing Symposium : A Psycholinguistic Approach*. London : OUP.
- Carroll, J. B.** 1975. *The Teaching of French as a Foreign Language in Eight Countries*. New York-London-Sydney-Toronto: A Halstead Press Book, John Wiley & Sons. Stockholm: Almqvist & Wiksell International.
- Carter, R.** 1987. *Vocabulary. Applied Linguistic Perspectives*. London : Allen & Unwin Publishers Ltd.
- Charolles, M.** 1978. "Introduction aux problèmes de la cohérence des textes". *Langue française* 38 : 7-41.
- Charolles, M.** 1986. "L'analyse des processus rédactionnels : aspects linguistiques, psycholinguistiques et didactiques". *Pratiques* 49 : 3-21.
- Charolles, M.** 1988. "Les plans d'organisation textuelle : période, chaînes, portées et séquences". *Pratiques* 57 : 3-13.
- Chomsky, N.** 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass. : MIT Press.
- Clark, E. V.** 1998. "Lexique et syntaxe dans l'acquisition du français". *Langue française* 118 : 49-60.
- Coady, J.** 1997. "L2 vocabulary acquisition through extensive reading". In J. Coady & T. Huckin (Eds.) *Second Language Vocabulary Acquisition*. pp. 225-237.
- Coady, J. & T. Huckin (Eds.)** 1997. *Second Language Vocabulary Acquisition*. A Rationale for Pedagogy. Cambridge : CUP.
- Coirier, P. & D. Gaonac'h & J-M. Passerault.** 1996. *Psycholinguistique textuelle. Une approche cognitive de la compréhension et de la production des textes*. Paris : Armand Colin.

- Corder**, S. P. 1975. "The significance of learners' errors". In J. C. Richards (Ed.) *Error Analysis – Perspectives on Second Language Acquisition*. Reprinted from *IRAL*, vol. V/4. 1967. pp. 19-27.
- Corder**, S. P. 1975. "Idiosyncratic dialects and error analysis". In J. C. Richards (Ed.) *Error Analysis – Perspectives on Second Language Acquisition*. Reprinted from *IRAL*, vol. IX/2. 1971. pp. 158-171.
- Corder**, S. P. 1980. "Dialectes idiosyncrasiques et analyse d'erreurs". *Langages* 57 : 17-28. Traduit de l'anglais par C. Perdue et R. Porquier. Originellement publié dans *IRAL*, vol. IX/2. 1971.
- Corder**, S. P. 1992. "Strategies of communication". In C. Faerch & G. Kasper (Eds.) *Strategies in Interlanguage Communication*. pp. 15-19.
- Cortès**, J. (Ed.) 1987. *Une introduction à la recherche scientifique en didactique des langues*. Ecole normale supérieure de Saint-Cloud. Paris : Didier/ Crédif.
- Cortès**, J. & J. Cortès. 1987. "Une taxonomie de la recherche en didactique des langues". In J. Cortès (Ed.) *Une introduction à la recherche scientifique en didactique des langues*. pp. 9- 20.
- Coseriu**, E. 1979. "System, Norm und Rede". In Coseriu, E. *Sprache – Strukturen and Funktionen. Zwölf Aufsätze zur allgemeinen und romanischen Sprachwissenschaft*. Erste Auflage 1970. Tübingen : Gunter Narr Verlag. pp. 45-59. Aus dem Italienischen von H. Bertsch.
- Cossette**, A. 1994. *La richesse lexicale et sa mesure*. Travaux de linguistique quantitative 53. Paris : Honoré Champion Éditeur.
- Coste**, D. 1985. "Sur quelques aspects des relations récentes entre grammaire et didactique du français langue étrangère". *Langue française* 68 : 5-17.
- Courtilon**, J. 1989. "Lexique et apprentissage de la langue". *Le français dans le monde*. Août-septembre 1989 : 146-153. Paris : EDICEF.
- Cruse**, D. A. 1986. *Lexical semantics*. Cambridge : CUP.
- Dale**, E. 1965. "Vocabulary measurement : Techniques and major findings". *Elementary English* 42 (5) : 895-901.
- Dewaele**, J.-M. 1997. "Style-shifting in oral interlanguage : Quantification and definition". In L. Eubank et al. (Eds.) *The Current State of Interlanguage*. pp. 233-240.
- Dictionnaire de linguistique (Ddl)*. 1989. Voir **Dubois et al.** *Dictionnaire des fréquences*. 1971. Etudes statistiques sur le vocabulaire français. Vocabulaire littéraire des XIX et XX siècles. Quatre tomes. Nancy : C. N. R. S. – T. L. F.
- De Pietro**, J. F., M. **Matthey** & B. **Py**. 1988. "Acquisition et contrat didactique : séquences potentiellement acquisitionnelles dans la conversation exolingue". In *Actes du 3e Colloque Régional de Linguistique*. Strasbourg.
- Dubois**, J., M. **Giacomo**, L. **Guespin**, Chr. et J.-B., **Marcellesi** & J.-P. **Mével**. 1989. *Dictionnaire de linguistique*. 1ère éd. 1973. Paris : Larousse.
- Dugast**, D. 1980. *La statistique lexicale*. Genève : Editions Slatkine.
- Dulay**, H. & M. **Burt**. 1973. "Should we teach children syntax ?". *Language Learning* 23 : 245-58.
- Dulay**, H. & M. **Burt**. 1974. "Natural sequences in child second language acquisition". *Language Learning* 24 : 37-53.
- Ellis**, R. 1992. *Understanding Second Language Acquisition*. First published 1985. Oxford: Oxford University Press (OUP).
- Ellis**, R. 1994. *The Study of Second Language Acquisition*. Oxford : OUP.
- Elo**, A. 1993. *Le français parlé par les étudiants finnophones et suédophones*. Turun yliopiston julkaisu. Annales universitatis turkuensis. Sarja B, osa 198. Turku : Turun yliopisto.
- Enkvist**, N. E. 1977. "Contextual acceptability and error evaluation". In R. Palmberg & H. Ringbom (Eds.) *Föredrag vid konferensen om kontrastiv lingvistik*

och felanalys. Stockholm & Åbo, 7-8 Februari 1977. Meddelanden från stiftelsen för Åbo Akademi forskningsinstitut n. 19. Åbo : Åbo Akademi. pp. 1-16.

Eubank, L. & L. Selinker & M. Sharwood Smith (Eds.). 1997. *The Current State of Interlanguage*. Amsterdam/Philadelphia : John Benjamins Publishing Company.

Faerch, C. 1986. "Rules of thumb and other teacher-formulated rules in the foreign language classroom". In G. Kasper (Ed.) *Learning, teaching, and communication in the foreign language classroom*. Aarhus : Aarhus U. P., 125-143.

Faerch, C. & K. Haastруп & R. Phillipson. 1984. *Learner Language and Language Learning*. København : Gyldendalske Boghandel, Nordisk Forlag A.S.

Faerch, C. & G. Kasper. 1992. "Plans and strategies in foreign language communication". In C. Faerch & G. Kasper (Eds.) *Strategies in Interlanguage Communication*. pp. 20-60.

Faerch, C. & G. Kasper (Eds.) 1992. *Strategies in Interlanguage Communication*. First published in 1983. New York : Longman.

Fayol, M. 1989. "Une approche psycholinguistique de la ponctuation : étude en production et en compréhension". *Langue française* 81 : 21-39.

Flower, L. & J. R. Hayes. 1981. "Plans that guide the composing process". In C. H. Frederiksen. & J. F. Dominic. (Eds.) *Writing : The Nature, Development, and Teaching of Written Communication*. pp. 39-58.

Français fondamental (1er degré). 1972. Première édition 1959. Paris : Publication de l'institut National de Recherche et de Documentation Pédagogiques.

Français fondamental (2e degré). (s.a.) 2e édition. Paris : Publication de l'Institut pédagogique national.

Freed, B. F. (Ed.) 1995. *Second Language Acquisition in a Study Abroad Context*. Studies in Bilingualism (SiBil) 9. Amsterdam/Philadelphia : John Benjamins Publishing Company.

Freed, B. F. 1995. "What makes us think that students who study abroad become fluent ?" In B. F. Freed (Ed.) *Second Language Acquisition in a Study Abroad Context*. pp. 123-148.

Frederiksen, C. H. & J. F. Dominic. 1981. "Introduction : perspectives on the activity of writing". In C. H. Frederiksen. & J. F. Dominic (Eds.) *Writing : The Nature, Development, and Teaching of Written Communication*. pp. 1-20.

Frederiksen, C. H. & J. F. Dominic (Eds.) 1981. *Writing : The Nature, Development, and Teaching of Written Communication*. Volume 2. Writing : Process, Development and Communication. Hillsdale, New Jersey : Lawrence Erlbaum Associates, Inc., Publishers.

Fries, Ch. 1945. *Teaching and Learning English as a Foreign Language*. Ann Arbor, Michigan : University of Michigan.

Galisson, R. 1979. *Lexicologie et enseignement des langues*. Paris : Hachette

Galisson, R. 1980. *D'hier à aujourd'hui la didactique des langues étrangères*. Paris :CLE International.

Galisson, R. 1981. "Approches communicatives et acquisitions des vocabulaires (du concordancier à l'auto-dictionnaire personnalisé)". *Bulletin CILA* 34 : 13-49.

Gaonac'h, D. 1984. "La notion d'interlangue et la psychologie cognitive de langage". In B. Py (Ed.) *Acquisition d'une langue étrangère III*. pp. 63-79.

Gaonac'h, D. 1991. *Théories d'apprentissage et acquisition d'une langue étrangère*. Paris : Hatier-Crédif.

Gass, S. & L. Selinker (Eds.) 1983. *Language Transfer in Language Learning*. Series on Issues in Second Language Research. Rowley, Massachusetts : Newbury House Publishers, Inc.

- Gass, S. & L. Selinker** (Eds.) 1994. *Language Transfer in Language Learning*. LALD 5. Reprinted with corrections. Amsterdam/ Philadelphia : John Benjamins Publishing Company.
- Germain, C.** 1993. *Evolution de l'enseignement des langues : 5000 ans d'histoire*. Paris: CLE International.
- Giacobbe, J.** 1990. "Le recours à la langue première – une approche cognitive". *Le français dans le monde*. Février 1990 : 115-123. Paris : EDICEF.
- Giacobbe, J.** 1992. *Acquisition d'une langue étrangère: Cognition et interaction*. Paris : CNRS Editions.
- Gombert, J. E.** 1990. *Le développement métalinguistique*. Paris : Presses Universitaires de France (PUF).
- Gombert, J. E.** 1991. "Le rôle des capacités métalinguistiques dans l'acquisition de la langue écrite". *Repères* no3 : ?
- Goudaillier, J.-P.** 1997. *Comment tu t'achètes!* Dictionnaire du français contemporain des cités. Paris : Maisonneuve et Larose.
- Graves, M. F.** 1987. "The roles of instruction in fostering vocabulary development". In M. G. McKeown & M. E. Curtis (Eds.) *The nature of vocabulary acquisition*. Hillsdale NJ : Erlbaum. pp. 165-184.
- Greimas, A. J.** 1989. "Qu'est-ce qu'un mot ? Le statut incertain du mot". *Le français dans le monde*. Août-septembre 1989 : 58. Paris : EDICEF.
- Grevisse, M.** 1993. *Le bon usage*. Grammaire française. Treizième édition refondue et revue par André Goosse. Paris-Louvain-la-Neuve : Duculot.
- Griggs, P.** 1999. "A propos de l'effet de l'activité métalinguistique sur les processus de production en L2". In M.L.M.S. (éd.) *Actes du XIe colloque international : Acquisition d'une langue étrangère : perspectives et recherches. Usages pragmatiques et acquisition des langues étrangères*. Paris 19, 20 et 21 avril 1999, France. Disponible à partir du site <http://www.marges-linguistiques.com/>
- Grosjean, F.** "Réponse à Rémy Porquier – Communication exolingue et communication bilingue". In B. Py (Ed.) *Acquisition d'une langue étrangère III*. pp. 49-61.
- Grönholm, M.** 1992. "Eri mittarien soveltuvuus toisen kielen oppijan sanaston rikkouden mittaamiseen". *Virittäjä* 1992 (2-3) : 239-252.
- Gustafsson, M.** 1990. "Lexical density as a style marker". In K. Battarbee & R. Hiltunen (Eds.) *Alarums & Excursions : Working Papers in English*. University of Turku : Publications of the Department of English. pp. 45-60.
- Haastруп, K. & R. Phillipson.** 1992. "Achievement strategies in learner/ native speaker interaction". In C. Faerch & G. Kasper (Eds.) *Strategies in Interlanguage Communication*. pp. 140-158.
- Halliday, M. A. K.** 1985. "Towards a sociological semantics - extracts". In C. J. Brumfit & K. Johnson (Eds.) *The Communicative Approach to Language Teaching*. First published in 1979. Oxford : OUP. pp. 27-45.
- Hammerly, H.** 1991. *Fluency and Accuracy. Towards Balance in Language Teaching and Learning*. Clevedon, Avon : Multilingual Matters Ltd.
- Happonen, J.-M.** 1999. *Dynamisme de l'interlangue – étude de cas sur la production orale des étudiants de français*. Mémoire de maîtrise. Département d'études françaises. Université de Turku.
- Harley, B. & M. L. King.** 1989. "Verb lexis in the written compositions of young L2 learners". *Studies in Second Language Acquisition* 11 (4) : 415-439.
- Herdan, G.** 1960. *Type-token Mathematics*. A textbook of mathematical linguistics. The Hague : Mouton & Co.
- Herdan, G.** 1964. *Quantitative Linguistics*. London : Butterworths.
- Herdan, G.** 1966. *The Advanced Theory of Language as Choice and Chance*. Berlin-New York : Springer-Verlag.

- Herschensohn, J.** 1999. "Deux profils de stratégies de communication". In M.L.M.S. (éd.) *Actes du XIe colloque international : Acquisition d'une langue étrangère : perspectives et recherches. Usages pragmatiques et acquisition des langues étrangères*. Paris 19, 20 et 21 avril 1999, France. Disponible à partir du site <http://www.marges-linguistiques.com/>
- Holland, K. & S. Johansson.** 1982. *Word frequencies in British and American English*. Bergen : The Norwegian Computing Centre for the Humanities.
- Huebner, T.** 1985. "System and variability in interlanguage syntax". *Language Learning* 35/2 : 141-163.
- Hymes, D. H.** 1972. "On communicative competence". In J. B. Pride & J. Holmes (Eds.) *Sociolinguistics*. London : Penguin. pp. 269-293.
- Hymes, D. H.** 1984. *Vers la compétence de communication*. Paris : Hatier-CREDIF. Traduit par F. Mugler.
- Härmä, J.** 1998. "Mitä hyötyä on kontrastiivisesta kielitieteestä ?" *Tempus* 1/98 : 10-11.
- Jaatinen, S. & T. Mankkinen.** 1993. "The size of vocabulary of university students of English". In K. Sajavaara & S. Takala (Eds.) *Finnish as Learners of English : Three Studies*. Jyväskylä Cross-Language Studies no 16. Jyväskylä : Department of English. pp. 147-211.
- Jain, M. P.** 1975. "Error analysis : source, cause and significance". In J. C. Richards (Ed.) *Error Analysis – Perspectives on Second Language Acquisition*. pp. 189-215.
- James, C.** 1981. *Contrastive Analysis*. First published 1980. London : Longman.
- Johansson, M.** 1997. *Les interviews médiatico-politiques françaises. Analyse pragmatique interactionniste*. Thèse de doctorat de 3e cycle. Département d'études françaises. Université de Turku.
- Johansson, S.** 1978. "Learners' errors and communicative efficiency". In V. Kohonen & N. E. Enkvist (Eds.) *Text Linguistics, Cognitive Learning and Language Teaching*. AFinLA 22. Turku. pp. 167-180.
- Kalmbach, J.-M.** 1994. *Suomi-ranska opiskelusanakirja*. Porvoo-Helsinki-Juva : WSOY.
- Kennedy, G.** 1992. "Preferred ways of putting things with implications for language teaching". In J. Svartvik (Ed.) *Directions in Corpus Linguistics*. pp. 335-373.
- Kerbrat-Orecchioni, C.** 1998. "La notion d'interaction en linguistique : origine, apports, bilan". *Langue française* 117 (février) : 51-67.
- Kerr-Barnes, B.** 1998. "The acquisition of connectors in French L2 narrative discourse". *Journal of French Language Studies* 8 (2) : 189-208.
- Klein, W.** 1989. *L'acquisition de langue étrangère*. Paris: Armand Colin. Traduit de W. Klein. 1987. *Zweitspracherwerb: Eine Einführung*. 2. Auflage. Frankfurt am Main : Athenäum Verlag GmbH.
- Koch, P.** 1997. "Orality in literate cultures". In C. Pontecorvo (Ed.) *Writing Development : An interdisciplinary view*. pp. 149-171.
- Krashen, S. D.** 1981. *Second Language Acquisition and Second Language Learning*. Oxford : Pergamon Press.
- Krashen, S. D.** 1982. *Principles and Practice in Second Language Acquisition*. Oxford : Pergamon Press.
- Labov, W.** 1972. *Sociolinguistic patterns*. Philadelphia : University of Pennsylvania Press.
- Labov, W.** 1996. "Some notes on the role of misperception in language learning". In R. Bayley & D. R. Preston (Eds.) *Second Language Acquisition and Linguistic Variation*. pp. 245-252.
- Lado, R.** 1957. *Linguistics Across Cultures*. Applied Linguistics for Language Teachers. Ann Arbor, Michigan : University of Michigan.

- Langages* 84. 1986. L'acquisition du français par des adultes immigrés : aspects psycholinguistiques. Composé par C. Perdue.
- Larsen-Freeman, D. & M. H. Long.** 1992. *An Introduction to Second Language Acquisition Research*. First published 1991. New York: Longman.
- Lauffer, B.** 1997. "The lexical plight in second language reading : Words you don't know, words you think you know, and words you can't guess". In J. Coady & T. Huckin (Eds.) *Second Language Vocabulary Acquisition*. pp. 20-34.
- Lavonius, E.** 1998. *Sanasto ranskan kielen ylioppilaskokeessa*. Mémoire de maîtrise. Département d'études françaises. Université de Turku.
- Leino, A-L.** 1980. "Vieraan kielen opettamisesta". In K. Sajavaara (Ed.) *Soveltava kielitiede*. pp. 170-180.
- Lenneberg, E. H.** 1967. *Biological foundations of language*. New York : Wiley.
- Léon, M. & P. Léon.** 1997. *La prononciation du français*. Paris : Editions Nathan.
- Linnarud, M.** 1979. "Lexikal täthet och betydelsesamband". In K. Hyltenstam (Ed.) *Svenska i invandrarperspektiv*. Lund : Liber Läromedel. pp. 219-230.
- Linnarud, M.** 1986. *Lexis in Composition. A Performance Analysis of Swedish Learners' Written English*. Doctoral dissertation. Lund Studies in English 74. Malmö : GWK Gleerup, Liber Förlag.
- Lundquist, L.** 1980. *La cohérence textuelle : syntaxe, sémantique, pragmatique*. København : Nyt Nordisk Forlag Arnold Busck.
- Lundquist, L.** 1983. *L'analyse textuelle : méthode, exercices*. Paris : CEDIC.
- Luukka, M.-R.** 1995. *Puhuttua ja kirjoitettua tiedettä : Funktionaalinen ja yhteisöllinen näkökulma tieteen kielen interpersonaalisiin piirteisiin*. Jyväskylä studies in communication 4. Jyväskylä : Jyväskylän yliopisto.
- Luukka, M.-R.** 1998. "Keskustelua verkossa. Sähköpostikeskustelu rekisterinä". *AFinLA Yearbook* 56 : 199-218.
- Mainueneau, D.** 1976. *Initiation aux méthodes de l'analyse du discours*. Paris : Hachette.
- Mainueneau, D.** 1987. *Nouvelles tendances en analyse du discours*. Paris : Hachette.
- Mainueneau, D.** 1994. *L'Enonciation en linguistique française*. Paris : Hachette.
- Martinet, A.** 1985. *Syntaxe générale*. Paris : Armand Colin.
- Ménard, N.** 1983. *Mesure de la richesse lexicale*. Genève – Paris : Slatkine-Champion.
- Mialaret, G.** 1991. *Statistiques appliquées aux sciences humaines*. Paris : PUF.
- Moirand, S.** 1984. "Réponse à Daniel Gaonac'h". In B. Py (Ed.) *Acquisition d'une langue étrangère III*. pp. 81-86.
- Moirand, S.** 1990. *Enseigner à communiquer en langue étrangère*. 1ère éd. 1982. Paris : Hachette.
- Moirand, S.** 1993. *Situations d'écrit. Compréhension / production en français langue étrangère*. 1ère éd. 1979. Paris : CLE International.
- Muller, Ch.** 1979. *Langue française et linguistique quantitative*. Recueil d'articles. Genève: Editions Slatkine.
- Mutta, M.** 1995. *A study of the vocabulary used in entrance examination compositions*. Mémoire de maîtrise. Département d'études anglaises. Université de Turku.
- Mutta, M.** 1997. "Lexical factors in entrance examination compositions". In H. Dufva (Ed.) *FINLANCÉ. A Finnish Journal of Applied Linguistics*. Vol. XVII. Jyväskylä : Centre for Applied Language Studies. pp. 80-91.
- Mutta, M.** 1998. "La compétence lexicale des étudiants finnophones en français". In *Actes du premier congrès des romanistes scandinaves pour étudiants en doctorat*. PERLES no 6. Lund : Lunds Universitets romanska institution. pp. 11-18.

- Mutta**, M. 1999a. "Le développement lexical des apprenants finlandais en français". In M.L.M.S. (éd.) *Actes du XIe colloque international : Acquisition d'une langue étrangère : perspectives et recherches. Usages pragmatiques et acquisition des langues étrangères*. Paris 19, 20 et 21 avril 1999, France. Disponible à partir du site <http://www.marges-linguistiques.com/>
- Mutta**, M. 1999b. "La conscience métacommunicative ou savoir apprendre à communiquer". In Y. Gambier & E. Suomela-Salmi (Eds.) *Jalons. 75 ans de l'enseignement du français à l'Université de Turku*. Publications du département d'études françaises de Turku 2. Turku. (à paraître)
- Mäkipää**, M. 1994. *Äidinkielen vaikutus vieraan kielen sanaston oppimisessa*. Mémoire de maîtrise. Département d'études françaises. Université de Turku.
- Nation**, I. S. P. 1990. *Teaching and learning vocabulary*. New York : Newbury House.
- Nation**, P. & J. **Newton**. 1997. "Teaching vocabulary". In J. Coady & T. Huckin (Eds.) *Second Language Vocabulary Acquisition*. pp. 238-254.
- Nemser**, W. 1971. "Approximative systems of foreign language learners". *IRAL* 9 : 115-123.
- Nold**, E. W. 1981. "Revising". In C. H. Frederiksen & J. F. Dominic. (Eds.) *Writing : The Nature, Development, and Teaching of Written Communication*. pp. 67-79.
- Norro**, S. 1995. *Les stratégies de communication dans la performance des étudiants de français*. Mémoire de maîtrise. Département d'études françaises. Université de Turku.
- Noyau**, C. & R. **Porquier** (Eds.) 1984. *Communiquer dans la langue de l'autre*. Paris : Presses Universitaires de Vincennes.
- Noyau**, C. & M.-Th. **Vasseur**. 1986. "L'acquisition des moyens de la référence temporelle en français langue étrangère chez des adultes hispanophones". *Langages* 84 : 105-117.
- The Oxford English Dictionary*. 1989. Second Edition. Volume VIII. Oxford : Clarendon Press.
- Paribakht**, T. S. & M. **Wesche**. 1997. "Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition". In J. Coady & T. Huckin (Eds.) *Second Language Vocabulary Acquisition*. pp. 174-200.
- Pendanx**, M. 1998. *Les activités d'apprentissage en classe de langue*. Paris : Hachette.
- Perdue**, C. 1993. "Comment rendre compte de la "logique" de l'acquisition d'une langue étrangère par l'adulte ?". *Etudes de linguistique appliquée* 92 : 8-22.
- Pergnier**, M. 1986. *Le mot*. Paris : PUF.
- Péry-Woodley**, M.-P. 1993. *Les écrits dans l'apprentissage*. Paris : Hachette.
- Le Petit Robert 1*. 1989. Dictionnaire alphabétique et analogique de la langue française. Rédaction dirigée par Rey, A. & J. Rey-Debove. Paris : Dictionnaires Le Robert.
- Picoche**, J. 1989. "Orientations en lexicologie". *Le français dans le monde*. Août-septembre 1989 : 86-91. Paris : EDICEF.
- Picoche**, J. 1992. *Précis de lexicologie française. L'étude et l'enseignement du vocabulaire*. 1ère édition 1977. Paris : Nathan.
- Pitkänen**, A. J. & V. **Kohonen**. 1984. *Johdatus kvantitatiiviseen kielentutkimukseen ja alan atk-sovelluksiin*. Helsinki : Gaudeamus.
- Pontecorvo**, C. 1997. "Introduction : Studying writing and writing acquisition today. A multidisciplinary view". In C. Pontecorvo (Ed.) *Writing Development : An interdisciplinary view*. pp. xv-xxxi.
- Pontecorvo**, C. (Ed.) 1997. *Writing Development : An interdisciplinary view*. Studies in written language and literacy 6. Amsterdam/Philadelphia : John Benjamins Publishing Company.

- Porquier, R.** 1975. "Analogie, généralisation et systèmes intermédiaires dans l'apprentissage d'une langue non maternelle". *BULAG* 3 : 37-64.
- Porquier, R.** 1977. "L'analyse des erreurs – Problèmes et perspectives". *Etudes de linguistique appliquée* 25 : 23-43.
- Porquier, R.** 1984. "Communication exolingue et apprentissage des langues". In B. Py (Ed.) *Acquisition d'une langue étrangère III*. pp. 17-47.
- Porquier, R.** 1986. "Remarques sur les interlangues et leurs descriptions". *Etudes de Linguistique Appliquée* 63 : 101-107.
- Porquier, R.** 1994. "Communication exolingue et contextes d'appropriation : Le continuum acquisition /apprentissage". *Bulletin suisse de linguistique appliquée* 59 : 159-169.
- Portine, H.** 1983. *L'argumentation écrite : expression et communication*. Paris : Hachette.
- Preston, D. R.** 1996. "Variationist perspectives on second language acquisition". In R. Bayley & D. R. Preston (Eds.) *Second Language Acquisition and Linguistic Variation*. pp. 1-45.
- Preston, D. R. & R. Bayley.** 1996. "Preface". In R. Bayley & D. R. Preston (Eds.) *Second Language Acquisition and Linguistic Variation*. pp. xiii-xviii.
- Py, B. (Ed.)** 1984. *Acquisition d'une langue étrangère III*. Paris-Neuchâtel : Presses Universitaires de Vincennes et Centre de Linguistique Appliquée de Neuchâtel.
- Regan, V.** 1995. "The acquisition of sociolinguistic native speech norms : effects of a year abroad on second language learners of French". In B. F. Freed (Ed.) *Second Language Acquisition in a Study Abroad Context*. pp. 245-267.
- Regan, V.** 1996. "Variation in French interlanguage : a longitudinal study of sociolinguistic competence". In R. Bayley & D. R. Preston (Eds.) *Second Language Acquisition and Linguistic Variation*. pp. 177-201.
- Reichler-Béguelin, M.-J.** 1988. "Anaphore, cataphore et mémoire discursive". *Pratiques* 57 : 15- 43.
- Reichler-Béguelin, M.-J., M. Denervaud & J. Jespersen.** 1988. *Ecrire en français. Cohésion textuelle et apprentissage de l'expression écrite*. Neuchâtel-Paris : Delachaux & Niestlé S. A.
- Richards, J. C. (Ed.)** 1975. *Error Analysis – Perspectives on Second Language Acquisition*. First ed. 1974. London : Longman.
- Riegel, M., J.-Ch. Pellat & R. Rioul.** 1997. *Grammaire méthodique du français*. 1ère éd. 1994. Paris : PUF.
- Ringbom, H.** 1987. *The Role of First Language in Second Language Learning*. Clevedon, Avon : Multilingual Matters Ltd.
- Rück, H.** 1980. *Linguistique textuelle et enseignement du français*. Paris : Hatier-CREDIF. Traduit de H. Rück. 1978. *Textlinguistik und Französischunterricht*. Dormund : Lambert Lensing.
- Russier, C., H. Stoffel & D. Véronique (Eds.)** 1991. *Interactions en langue étrangère*. Aix-en-Provence : Publications de l'Université de Provence.
- Sajavaara, K.** 1977. "Kontrastiivinen analyysi ja kommunikatiivinen kompetenssi". In K. Sajavaara (Ed.) *Näkökulmia kieleen*. AFinLA 18. Helsinki. pp. 59-82.
- Sajavaara, K.** 1980. "Kontrastiivinen kielentutkimus ja virheanalyysi". In K. Sajavaara (Ed.) *Soveltava kielitiede*. pp. 202-221.
- Sajavaara, K. (Ed.)** 1980. *Soveltava kielitiede*. Helsinki : Gaudeamus.
- Sajavaara, K.** 1980. "Toisen kielen oppiminen ja omaksuminen". In K. Sajavaara (Ed.) *Soveltava kielitiede*. pp. 115-135.
- Sautermeister, C.** 1989. "Pour une meilleure compétence lexicale". *Le français dans le monde*. Août-septembre 1989 : 122-133. Paris : EDICEF.

- Schachter, J., A. F. Tyson & F. J. Diffley.** 1976. "Learner intuitions of grammaticality". *Language Learning* 26/1 : 67-76.
- Schneuwly, B., M.-C. Rosat & J. Dolz.** 1989. "Les organisateurs textuels dans quatre types de textes écrits : étude chez les élèves de dix, douze et quatorze ans". *Langue française* 81 : 40-58.
- Schumann, J.** 1976. "Second language acquisition : the pidginization hypothesis". *Language Learning* 26/2 : 391-408.
- Schumann, J.** 1978. "Implications of pidginization and creolization for the study of adult second language acquisition". In J. Schumann & N. Stenson (Eds.) *New Frontiers in Second Language Learning*. First ed. 1975. Rowley, Mass. : Newbury House Publishers, Inc. pp. 137-152.
- Schwarte, B. S.** 1982. *The Acquisition of English Sentential Complementation by Adult Speakers of Finnish*. Jyväskylä Cross-Language Studies 8. Jyväskylä : Department of English, University of Jyväskylä.
- Selinker, L.** 1975. "Interlanguage". In J. C. Richards (Ed.) *Error Analysis – Perspectives on Second Language Acquisition*. Reprinted from *IRAL*, vol. X/3. 1972. pp. 31-54.
- Selinker, L.** 1992. *Rediscovering Interlanguage*. New York : Longman.
- Selinker, L. & U. Lakshmanan.** 1994. "Language transfer and fossilization : the multiple effects principle". In S. Gass & L. Selinker (Eds.) *Language Transfer in Language Learning*. pp. 197-216.
- Stern, H. H.** 1987. *Fundamental Concepts of Language Teaching*. First published in 1983. Oxford : OUP.
- Sundelin, S.** 1993. *Ranska-suomi opiskelusanakirja*. Porvoo-Helsinki-Juva : WSOY.
- Suomela-Salmi, E.** 1987. "Quelques observations sur la cohésion et la cohérence des textes traduits". In J. Härmä & I. Mäkinen-Schwanck (Eds.) *Communications*. (8^e rencontre des professeurs de français de l'enseignement supérieur. Helsinki : Université de Helsinki. pp. 115-140.
- Suomela-Salmi, E.** 1993. "Remarques sur l'organisation pragmatique du discours – une étude de cas". In *Romanistica Turkuensis. Mélanges d'études romanes*. Offerts à Lauri Lindgren à l'occasion de son 60^e anniversaire. Annales universitatis turkuensis. Sarja B, osa 202. Turku : Turun yliopisto. pp. 241-257.
- Suomela-Salmi, E.** 1997. *Les syntagmes nominaux (SN) dans les discours économiques français : repères textuels*. Turun yliopiston julkaisu. Annales universitatis turkuensis. Sarja B, osa 218. Turku : Turun yliopisto.
- Svartvik, J. (Ed.)** 1992. *Directions in Corpus Linguistics*. Proceedings of Nobel Symposium 82. Stockholm, 4-8 August 1991. Trends in Linguistics. S. a. M. 65. Berlin-New York : Mouton de Gruyter.
- Svartvik, J.** 1992. "Corpus linguistics comes of age". In J. Svartvik (Ed.) *Directions in Corpus Linguistics*. pp. 8-13.
- Takala, S.** 1984. *Evaluation of Students' Knowledge of English Vocabulary in the Finnish Comprehensive School*. Kasvatustieteiden tutkimuslaitoksen julkaisu 350/1984. Jyväskylä : University of Jyväskylä.
- Takala, S. (Ed.)** 1989. *Sanaston opettaminen ja oppiminen*. Kasvatustieteiden tutkimuslaitoksen julkaisusarja B. Teoriaa ja käytäntöä. Jyväskylä : Kasvatustieteiden tutkimuslaitos.
- Takala, S.** 1989. "Sanaston oppimisen uudet haasteet". In S. Takala (Ed.) *Sanaston opettaminen ja oppiminen*. pp. 1-11.
- Tarone, E.** 1978. "Conscious communication strategies in interlanguage : a progress report". In H. Brown, C. Yorio & R. Crymes (Eds.) *On TESOL '77 : Teaching and Learning ESL*. Washington, D. C. : TESOL. pp. 194-203.
- Tarone, E.** 1980. "Communication strategies, foreigner talk and repair in interlanguage". *Language Learning* 30/2 : 417-431.

- Tarone**, E. 1988. *Variation in Interlanguage*. London : Edward Arnold.
- Tarone**, E. 1992. "Some thoughts on the notion of 'communicative strategy'". In C. Faerch & G. Kasper (Eds.) *Strategies in Interlanguage Communication*. pp. 61-74.
- Tarone**, E., A. D. **Cohen** & G. **Dumas**. 1992. "A closer look at some interlanguage terminology : a framework for communication strategies". In C. Faerch & G. Kasper (Eds.) *Strategies in Interlanguage Communication*. pp. 4-14.
- Thorén**, B. 1976. *10000 ord för tio års engelska*. Lund : Liber.
- Thorndike**, E. L. 1921. *Teacher's word book*. New York : Columbia Teachers College.
- Tiilikainen**, R. & K. **Ylimentalo**. 1989. "Sanaston käyttö lukion I-luokkalaisten englannin aineissa". In S. Takala (Ed.) *Sanaston opettaminen ja oppiminen*. pp. 131-139.
- Tréville**, M-C. & L. **Duquette**. 1996. *Enseigner le vocabulaire en classe de langue*. Paris : Hachette.
- Trévisse**, A. 1985. "Acquisition d'une langue 2 en milieu naturel : quelles méthodologies de description ?". *Langue française* 68 : 18-31.
- Trévisse**, A. 1994. "Représentations métalinguistiques des apprenants, des enseignants et des linguistes : un défi pour la didactique". *Bulletin suisse de linguistique appliquée* 59 : 171-190.
- Turco**, G. & D. **Coltier**. 1988. "Des agents doubles de l'organisation textuelle : les marqueurs d'intégration linéaire". *Pratiques* 57 : 57-79.
- Vargas**, Cl. 1999. *Grammaire pour enseigner / 1 : L'énoncé, le texte, la phrase*. 2 éd. Paris : Armand Colin.
- Vasseur**, M.-Th. 1991. "Solliciter n'est pas apprendre (initiative, sollicitation et acquisition d'une langue étrangère)". In C. Russier, H. Stoffel & D. Véronique (Eds.) *Interactions en langue étrangère*. pp. 49-59.
- Vauras**, M. & M. **Silvén**. 1985. *Metakognition kehittyminen kouluikässä*. Psykologian tutkimuksia 75. Turku : Turun yliopiston psykologian laitos.
- Verhoeven**, L. 1997. "Acquisition of literacy by immigrant children". In C. Pontecorvo (Ed.) *Writing Development : An interdisciplinary view*. pp. 219-240.
- Véronique**, D. 1992. "Recherches sur l'acquisition des langues secondes : un état des lieux et quelques perspectives". *AILE* 1 : 5-35.
- Vigil**, N. A. & J. W. **Oller**. 1976. "Rule fossilization : a tentative model". *Language Learning* 26/2 : 281-295.
- Vives**, R. 1985. "Lexique-grammaire et didactique du français langue étrangère". *Langue française* 68 : 48-65.
- Vogel**, K. 1995. *L'interlangue - la langue de l'apprenant*. Toulouse : Presses Universitaires du Mirail. Titre original (1990) *Lernersprache*. Tübingen : Gunter Narr Verlag.
- Vygotski**, L. 1997. *Pensée et langage*. 3e éd. Traduction revue. Paris : La dispute. (originellement paru en 1934).
- Widdowson**, H. G. 1981. *Une approche communicative de l'enseignement des langues*. Paris : Hatier-CREDIF. Titre original (1978) *Teaching language as communication*. Oxford : OUP.
- Wierzbicka**, A. 1988. *The semantics of grammar*. Amsterdam/Philadelphia : John Benjamins Publishing Company.
- Yli-Jokipii**, H. 1994. *Requests in Professional Discourse : A cross-cultural Study of british, American and Finnish Business Writing*. Annales academiae scientiarum fennicae dissertationes humanarum litterarum 71. Helsinki : Suomalainen tiedeakatemia.

Exemples des sujets dans les dissertations :

1.
Vos meilleur(e)s ami(e)s. Comment les avez-vous connu(e) ? Décrivez-les (physique, personnalité, occupations...). Qu'est-ce que vous aimez en eux/elles, et qu'est-ce que vous aimez moins ? Racontez un bon souvenir que vous avez en commun avec l'un(e) d'eux.
2.
Faites le compte rendu d'un livre que vous avez lu.
3. Le sida
Que savez-vous du sida ? Quels sont les pays les plus touchés par ce virus ? Les campagnes de prévention menées par les gouvernements sont-elles suffisantes pour endiguer ce fléau ? Quelles sont les personnes qu'il faudrait soumettre à l'examen de dépistage du sida ? A votre avis, cet examen doit-il être obligatoire ? Une entreprise ou une administration a-t-elle le droit d'exclure (de licencier) une personne séropositive ? Juridiquement, qu'en est-il en Finlande ? Quelle est l'attitude de votre pays face à ce grave problème ? Peut-on aider les gens touchés par cette maladie ? Et comment ? Les séronégatifs sont-ils à l'abri du sida ? A votre avis, pourra-t-on guérir très rapidement cette maladie ?
4.
Que pensez-vous de cette affirmation : « Naître en Finlande, c'est gagner le gros lot » ? Justifiez votre opinion.
5.
L'école prépare-t-elle à la vie ? Sinon, que faudrait-il y changer ?
6.
Existe-t-il de pieux mensonges (des mensonges que l'on peut faire pour éviter de blesser des gens que l'on aime) ou tous les mensonges sont-ils condamnables ? Vous répondrez à cette question en vous aidant, si vous le souhaitez, d'exemples personnels et concrets.

ANNEXE 2

étudiant	matière principale	séjour à Aix 1-2 semestres	niveau I				niveau II			
			nombre des mots	mots lexicaux		mots grammaticaux	nombre des mots	mots lexicaux		mots grammaticaux
				vocabulaire	occurrence			vocabulaire	occurrence	
1.	art	2	197	68	87	110	155	52	64	91
2.	fr.	2	151	52	65	86	193	62	89	104
3.	suéd.	1	189	76	93	96	225	77	97	128
4.	fr.	2	251	92	117	134	201	67	79	122
5.	angl.	1	157	53	70	87	239	81	94	145
6.	suéd.	2	175	54	73	102	151	49	59	92
7.	fr.	1	189	50	69	120	204	61	83	121
8.	angl.	1	253	107	119	134	215	86	89	126
9.	fr.	1	253	76	100	153	182	67	76	106
10.	fr.	2	241	75	97	144	197	59	80	117
11.	fr.	1	195	70	83	112	211	73	89	122
12.	fr.	1	231	84	107	124	209	66	82	127
13.	fr.	1	190	73	85	105	183	61	68	115
14.	fr.	1	201	81	91	110	236	104	109	127
15.	fr.	2	203	61	87	116	203	67	88	115
16.	art	2	194	63	79	115	285	81	111	174
17.	fr.	1	198	64	77	121	182	66	76	106
18.	angl.	2	159	51	73	86	154	52	68	86
19.	fr.	2	219	85	96	123	244	92	106	138
20.	suéd.	1	182	78	89	93	181	64	84	97
21.	fr.	2	180	65	78	102	252	91	108	144
22.	fr.	2	203	78	84	119	172	58	74	98
23.	fr.	1	201	47	68	133	212	81	97	115
24.	fr.	1	153	48	64	89	170	63	78	92
25.	fr.	1	193	60	72	121	182	73	77	105
26.	fr.	1	180	73	79	101	179	64	74	105
27.	fr.	2	217	64	87	130	252	66	89	163
28.	fr.	2	139	47	58	81	188	72	84	104
29.	fr.	2	311	99	135	176	354	121	156	198
30.	angl.	1	170	54	75	95	228	66	92	136
31.	fr.	1	236	74	98	138	194	67	93	101
32.	fr.	1	288	97	119	169	202	75	85	117
33.	fr.	2	251	82	122	129	209	65	87	122
34.	fr.	2	207	67	87	120	174	62	68	106
35.	fr.	2	257	83	113	144	214	68	92	122
36.	fr.	2	212	77	87	125	209	58	90	119
37.	fr.	2	193	60	68	125	228	69	96	132
38.	fr.	2	268	85	120	148	205	69	88	117
39.	fr.	1	181	59	73	108	219	81	101	118
40.	fr.	2	173	54	76	97	252	79	94	158
41.	all.	1	206	60	91	115	331	94	145	186
42.	fr.	2	200	67	75	125	244	73	102	142
43.	fr.	1	157	57	75	82	166	57	76	90
44.	angl.	1	165	55	74	91	188	68	79	109
45.	suéd.	2	200	63	82	118	368	114	161	207
46.	fr.	1	254	82	107	147	174	61	72	102
47.	fr.	1	231	76	96	135	178	51	76	102
48.	fr.	1	228	88	101	127	208	77	93	115
49.	fr.	2	254	91	106	148	269	94	136	133
50.	fr.	2	213	64	93	120	206	65	82	124
51.	fr.	1	238	77	100	138	209	67	81	128
52.	fr.	2	270	97	128	142	250	91	113	137
53.	it.	2	212	80	96	116	196	70	84	112
total			11069	3743	4744	6325	11332	3817	4814	6518
moyen			208,9	70,6	89,5	119,3	213,8	72,0	90,8	123,0
écart			37,43	14,84	18,05	21,78	45,08	14,97	20,87	26,04

Tableau 11. Nombre et répartition de mots : le corpus d'Aix

ANNEXE 2 (continue)

étudiant	matière principale	séjour linguistique	niveau I				niveau II			
			nombre des mots	mots lexicaux		mots grammaticaux	nombre des mots	mots lexicaux		mots grammaticaux
				vocabulaire	occurrence			vocabulaire	occurrence	
1.	fr.	> 1 an	180	63	72	108	263	93	113	150
2.	angl.	3 mois	178	54	71	107	186	63	77	109
3.	fr.	2 mois	159	55	63	96	180	61	81	99
4.	fr.	10 mois	202	64	80	122	291	76	117	174
5.	angl.	> 1 an	228	70	88	140	246	77	103	143
6.	fr.	10 mois	192	57	77	115	207	62	79	128
7.	all.	1 mois	300	96	115	185	199	55	84	115
8.	angl.	6 mois	236	90	113	122	192	66	80	112
9.	fr.	2 mois	290	95	126	164	274	77	104	170
10.	fr.	1 mois	258	73	96	162	340	97	131	209
11.	fr.	1 mois	258	57	98	160	250	72	98	152
12.	angl.	6 mois	238	90	112	126	240	87	102	138
13.	fr.	5 mois	268	98	107	161	298	85	107	191
14.	fr.	> 1 an	246	79	111	135	477	135	194	283
15.	angl.	-	187	69	80	107	179	57	75	104
16.	fr.	8 mois	195	64	81	114	234	74	99	135
17.	all.	1 mois	179	63	77	102	214	70	87	127
18.	angl.	1 mois	176	54	67	109	298	100	126	172
19.	fr.	1 mois	163	51	65	98	242	67	94	148
20.	angl.	2,5 mois	167	52	75	92	166	64	78	88
21.	angl.	1 mois	165	55	64	101	209	73	93	116
22.	all.	1 mois	200	69	86	114	231	89	100	131
23.	fr.	11 mois	247	68	89	158	219	69	82	137
24.	it.	1 mois	200	81	94	106	147	52	67	80
25.	all.	1 mois	256	80	96	160	206	57	81	125
total			5368	1747	2203	3164	5988	1878	2452	3536
moyen			214,7	69,9	88,1	126,6	239,5	75,1	98,1	141,4
écart			42,11	14,94	18,07	26,82	67,73	18,15	25,86	42,84

Tableau 12. Nombre et répartition de mots : le corpus témoin

ANNEXE 3

étudiant	niveau I						niveau II					
	VL %	VL/ M	DL %	mots uniques a / b	ILe %		VL %	VL/ M	DL %	mots uniques a / b	ILe %	
					a	b					a	b
1.	78,2	0,40	44,2	13 / 8	14,9	9,2	81,3	0,52	41,3	16 / 6	25,0	9,4
2.	80,0	0,53	43,1	6 / 2	9,2	3,1	69,7	0,36	46,1	9 / 3	10,1	3,4
3.	81,7	0,43	49,2	24 / 14	25,8	15,1	79,4	0,35	43,1	18 / 10	18,6	10,3
4.	78,6	0,31	46,6	23 / 9	19,7	7,7	84,8	0,42	39,3	23 / 15	29,1	19,0
5.	75,7	0,48	44,6	17 / 5	24,3	7,1	86,2	0,36	39,3	20 / 11	21,3	11,7
6.	74,0	0,42	41,7	8 / 2	11,0	2,7	83,1	0,55	39,1	10 / 6	17,0	10,2
7.	72,5	0,38	36,5	9 / 3	13,0	4,4	73,5	0,36	40,7	23 / 12	27,7	14,5
8.	89,9	0,36	47,0	44 / 25	37,0	21,0	96,6	0,45	41,4	41 / 26	46,1	29,2
9.	76,0	0,30	39,5	14 / 4	14,0	4,0	88,2	0,49	41,8	11 / 5	14,5	6,6
10.	77,3	0,32	40,3	11 / 6	11,3	6,2	73,8	0,38	40,6	12 / 6	15,0	7,5
11.	84,3	0,43	42,6	21 / 11	25,3	13,3	82,0	0,39	42,2	15 / 10	16,9	11,2
12.	78,5	0,34	46,3	18 / 8	16,8	7,5	80,5	0,39	39,2	15 / 12	18,3	14,6
13.	85,9	0,45	44,7	28 / 15	32,9	17,7	89,7	0,49	37,2	13 / 10	19,1	14,7
14.	89,0	0,44	45,3	25 / 12	27,5	13,2	95,4	0,40	46,2	35 / 18	32,1	16,5
15.	70,1	0,35	42,9	13 / 8	14,9	9,2	76,1	0,38	43,4	17 / 7	19,3	8,0
16.	79,8	0,41	40,7	17 / 8	21,5	10,1	73,0	0,26	39,0	21 / 12	18,9	10,8
17.	83,1	0,42	38,9	16 / 9	20,8	11,7	86,8	0,48	41,8	15 / 6	19,7	7,9
18.	69,9	0,44	45,9	9 / 2	12,3	2,7	76,5	0,50	44,2	13 / 3	19,1	4,4
19.	88,5	0,40	43,8	30 / 17	31,3	17,7	86,8	0,36	43,4	41 / 29	38,7	27,4
20.	87,6	0,48	48,9	22 / 11	24,7	12,4	76,2	0,42	46,4	21 / 9	25,0	10,7
21.	83,3	0,46	43,3	11 / 3	14,1	3,9	84,3	0,34	42,9	30 / 18	27,8	16,7
22.	92,9	0,46	41,4	18 / 11	21,4	13,1	78,4	0,46	43,0	14 / 8	18,9	10,8
23.	69,1	0,34	33,8	7 / 4	10,3	5,9	83,5	0,39	45,8	21 / 9	21,7	9,3
24.	75,0	0,49	41,8	8 / 3	12,5	4,7	80,8	0,48	45,9	10 / 7	12,8	9,0
25.	83,3	0,43	37,3	10 / 4	13,9	5,6	94,8	0,52	42,3	22 / 11	28,6	14,3
26.	92,4	0,51	43,9	15 / 8	19,0	10,1	86,5	0,48	41,3	16 / 6	21,6	8,1
27.	73,6	0,34	40,1	6 / 4	6,9	4,6	74,2	0,29	35,3	13 / 3	14,6	3,4
28.	81,0	0,58	41,7	10 / 1	17,2	1,7	85,7	0,46	44,7	17 / 8	20,2	9,5
29.	73,3	0,24	43,4	31 / 9	23,0	6,7	77,6	0,22	44,1	38 / 18	24,4	11,5
30.	72,0	0,42	44,1	16 / 8	21,3	10,7	71,7	0,31	40,4	18 / 5	19,6	5,4
31.	75,5	0,32	41,5	19 / 7	19,4	7,1	72,0	0,37	47,9	14 / 7	15,1	7,5
32.	81,5	0,28	41,3	26 / 11	21,9	9,2	88,2	0,44	42,1	15 / 8	17,7	9,4
33.	67,2	0,27	48,6	17 / 7	13,9	5,7	74,7	0,36	41,6	13 / 7	14,9	8,1
34.	77,0	0,37	42,0	13 / 4	14,9	4,6	91,2	0,52	39,1	22 / 12	32,4	17,7
35.	73,5	0,29	44,0	23 / 14	20,4	12,4	73,9	0,35	43,0	18 / 8	19,6	8,7
36.	88,5	0,42	41,0	18 / 7	20,7	8,1	64,4	0,31	43,1	16 / 5	17,8	5,6
37.	88,2	0,46	35,2	12 / 5	17,7	7,4	71,9	0,32	42,1	19 / 7	19,8	7,3
38.	70,8	0,26	44,8	27 / 9	22,5	7,5	78,4	0,38	42,9	22 / 9	24,7	10,1
39.	80,8	0,45	40,3	11 / 4	15,1	5,5	80,2	0,37	46,1	25 / 19	24,8	18,8
40.	71,1	0,41	43,9	9 / 4	11,8	5,3	84,0	0,33	37,3	20 / 11	21,3	11,7
41.	65,9	0,32	44,2	13 / 4	14,3	4,4	64,8	0,20	43,8	25 / 7	17,2	4,8
42.	89,3	0,45	37,5	25 / 11	33,3	14,7	71,6	0,29	41,8	23 / 7	22,6	6,9
43.	76,0	0,48	47,8	12 / 3	16,0	4,0	75,0	0,45	45,8	23 / 16	30,3	21,1
44.	74,3	0,45	44,9	17 / 10	23,0	13,5	86,1	0,46	42,0	18 / 9	22,8	11,4
45.	76,8	0,38	41,0	9 / 3	11,0	3,7	70,8	0,19	43,8	25 / 14	15,5	8,7
46.	76,6	0,30	42,1	18 / 7	16,8	6,5	84,7	0,49	41,4	21 / 13	29,2	18,1
47.	79,2	0,34	41,6	17 / 6	17,7	6,3	67,1	0,38	42,7	4 / 2	5,2	2,6
48.	87,1	0,38	44,3	25 / 21	24,8	20,8	82,8	0,40	44,7	25 / 14	26,9	15,1
49.	85,9	0,34	41,7	30 / 20	28,3	18,9	69,1	0,26	50,6	24 / 14	17,7	10,3
50.	68,8	0,32	43,7	12 / 9	12,9	9,7	79,3	0,39	39,8	13 / 6	15,9	7,3
51.	77,0	0,32	42,0	17 / 9	17,0	9,0	82,7	0,40	38,8	23 / 12	28,4	14,8
52.	75,8	0,28	47,4	22 / 11	17,2	8,6	80,5	0,32	45,2	29 / 12	25,7	10,6
53.	83,3	0,39	45,3	29 / 12	30,2	12,5	83,3	0,43	42,9	24 / 11	28,6	13,1
total				921/432						1049/539		
moyen	78,99	0,39	42,82	17 / 8	19,0	8,8	79,88	0,39	42,45	20 / 10	21,8	11,2
écart	6,86	0,08	3,29	7,9 / 5,1	6,73	4,80	7,58	0,08	2,86	7,6 / 5,4	7,04	5,46

Tableau 13. Facteurs quantitatifs : le corpus d'Aix

ANNEXE 3 (continue)

étudiant	niveau I						niveau II					
	VL %	VL/M	DL %	mots uniques a / b	ILe %		VL %	VL/M	DL %	mots uniques a / b	ILe %	
					a	b					a	b
1.	87,5	0,49	40,0	26/ 16	36,1	22,2	82,3	0,31	43,0	34/ 16	30,1	14,2
2.	76,1	0,43	39,9	24/ 12	33,8	16,9	81,8	0,44	41,4	24/ 15	31,2	19,5
3.	87,3	0,55	39,6	19/ 8	30,2	12,7	75,3	0,42	45,0	10/ 5	12,4	6,2
4.	80,0	0,40	39,6	21/ 7	26,3	8,8	65,0	0,22	40,2	22/ 10	18,8	8,6
5.	79,6	0,35	38,6	22/ 6	25,0	6,8	74,8	0,30	41,9	19/ 9	18,5	8,7
6.	74,0	0,39	40,1	10/ 3	13,0	3,9	78,5	0,38	38,2	16/ 10	20,3	12,7
7.	83,5	0,28	38,3	34/ 16	29,6	13,9	65,5	0,33	42,2	18/ 6	21,4	7,1
8.	79,7	0,34	47,9	40/ 21	35,4	18,6	82,5	0,43	41,7	25/ 17	31,3	21,3
9.	75,4	0,26	43,5	33/ 20	26,2	15,9	74,0	0,27	38,0	32/ 16	30,8	15,4
10.	76,0	0,30	37,2	25/ 9	26,0	9,4	74,1	0,22	38,5	37/ 16	28,2	12,2
11.	58,2	0,23	38,0	9/ 1	9,2	1,0	73,5	0,29	39,2	14/ 5	14,3	5,1
12.	80,4	0,34	47,1	40/ 32	35,7	28,6	85,3	0,36	42,5	46/ 23	45,1	22,6
13.	91,6	0,34	39,9	48/ 27	44,9	25,2	79,4	0,27	35,9	40/ 12	37,4	11,2
14.	71,2	0,29	45,1	17/ 11	15,3	9,9	69,6	0,15	40,7	29/ 14	15,0	7,2
15.	86,3	0,46	42,8	18/ 14	22,5	17,5	76,0	0,43	41,9	11/ 9	14,7	12,0
16.	79,0	0,41	41,5	30/ 20	37,0	24,7	74,8	0,32	42,3	23/ 13	23,2	13,1
17.	81,8	0,46	43,0	14/ 5	18,2	6,5	80,5	0,38	40,7	26/ 18	29,9	20,7
18.	80,6	0,46	38,1	19/ 11	28,4	16,4	79,4	0,27	42,3	44/ 22	34,9	17,5
19.	78,5	0,48	39,9	11/ 4	16,9	6,2	71,3	0,30	32,8	19/ 11	20,2	11,7
20.	69,3	0,42	44,9	15/ 6	20,0	8,0	82,1	0,50	47,0	15/ 11	19,2	14,1
21.	85,9	0,52	38,8	12/ 4	18,8	6,3	78,5	0,38	44,5	26/ 13	28,0	14,0
22.	80,2	0,40	43,0	14/ 5	16,3	5,8	89,0	0,39	43,3	44/ 24	44,0	24,0
23.	76,4	0,31	36,0	17/ 10	19,1	11,2	84,2	0,38	37,4	19/ 5	23,2	6,1
24.	86,2	0,43	47,0	22/ 11	23,4	11,7	77,6	0,53	45,6	17/ 11	25,4	16,4
25.	83,3	0,33	37,5	26/ 13	27,1	13,5	70,4	0,34	39,3	12/ 9	14,8	11,1
total				566/292						622/320		
moyen	79,52	0,39	41,09	23/12	25,4	12,9	77,02	0,34	41,02	25/13	25,3	13,3
écart	6,92	0,09	3,33	10,2/7,7	8,69	7,14	5,95	0,09	3,16	10,8/5,4	9,03	5,38

Tableau 14. Facteurs quantitatifs : le corpus témoin

EXEMPLE 1 : beaucoup de mots réellement uniques dans les dissertations d'un étudiant

Niveau I (25):

Naître en Finlande, c'est gagner le gros lot

Les Finlandais ni parlent ni *s'embrassent*. C'est assez rare de voir une sourire chaleureuse mais quand cela se passe, on en peut être sûr que c'est *authentique*. Un vrai Finlandais !

Nos politiciens, bien payés et *chauvins*, sont fiers de notre *neutralité*. Cette attitude nous a sauvés de plusieurs de *calamités*, des *brutalités* d'une guerre, de la nourriture *bonmarchée*, et des étrangers dangereux dont le peau n'est pas aussi pâle que le nôtre. Pourvu que nous puissions garder notre identité de haute classe, et ne pas être directement en *contact* avec les nouvelles idées et coutumes d'autres peuples européens, les *decisseurs* hésitent d'ouvrir les portes vers le *continent*. L'isolation totale est notre *sort* inévitable. La culture finlandaise ne disparaîtra pas, mais elle ne va pas *s'enrichir*, et sans la concurrence des alternatives, elle restera sans *nuances* et deviendra *monotone*, pas de *dynamique*. A mon avis, ce serait vraiment *pitoyable* si *grossement sousestimer* les possibilités de notre civilisation nordique. En effet, il s'agirait d'une *suicide*, ou plus précisément notre influence et *existence* dans le monde seraient *nullifiées complètement*. Toutes les personnes qui ont du talent de quelque sorte, partiraient pour jamais retourner, et aideraient d'autres pays d'autres peuples, se développer et réussir.

Si "naître en Finlande, c'est gagner le gros lot" est une *proposition valable*, c'est seulement ainsi à une condition : La Finlande commencera à participer plus *activement* dans la discussion mondiale, et n'oubliera pas de la confiance en soi.

Niveau II (26):

Mon "ami"

Il est la dernière *tentation* sur la terre avant des événements inévitables, *apocalyptiques* après lesquels l'histoire humaine deviendra transformée, dans un peu de temps, tout fait *anéantie*.

Il s'habille en blanc pour nous *tromper*. La couleur d'*innocence* l'*enveloppe* dans sa *caresse*, mais elle ne touche pas son âme, bien qu'on puisse *témoigner* la façon dont le Bien, le *souverain honorable* essaie de gagner du terrain, de *pénétrer* dans son cœur *malsain*. Il reste calme et, *nonchalemment*, il *serre* les mains en souriant. Par son regard il nous *manipule* et nous *excite*. Cette fois-ci on ne peut pas apercevoir le danger d'avance. Mon ami, il a l'air si *tendre*, si *magnanime* qu'on est prêt de le suivre partout! Où nous sommes-nous rencontrés pour la première fois ?

C'était dans un *supermarché* près d'où j'habite ; moi, j'étais en train de *peser* mes *artichauts* quand, soudainement, je l'ai remarqué. Je ne souviens maintenant qu'en même temps que cela s'est passé, j'ai éprouvé une émotion bien étrange, celle de la *reconnaissance ardente*. Il y avait une odeur d'argent qui *chatouillait* mes *narines*. Poli, il se présentait : "Je m'appelle *Mark* Finlandais." "Enchanté", ai-je m'entendu répondre. Des amis pour toujours...

EXEMPLE 2 : beaucoup de mots réellement uniques dans les dissertations d'un étudiant

Niveau I (17) :

Le sida

Lorsqu'on a lancé une nouvelle de la maladie fatale au début des années 80, le monde entier était bouleversé par la *terreur*. On a accusé les homosexuels de leurs relations *anormales* et on les a jeté dans un coin de la société. Pourtant, aujourd'hui il faut constater que le sida n'est pas *uniquement* 'le droit *exclusif* des homosexuels ou des *toxicomans*.

En Finlande, puisque le nombre des *séropositifs* n'est pas grand, les jeunes sont trop libéraux à ses rapports sexuelles. Même si le sida ne leur fait pas la peine, les autres maladies *venères* ont gagné du terrain de se *fleurir* tranquillement. Le gouvernement Finlandais a donné sa *contribution* en faisant *distribuer* des *préservatifs*. Mais pour *dissoudre* le problème il faudra partir du fond - en changeant des attitudes on pourra atteindre des résultats. Les vieux principes, comme la *fidélité*, entre homme et femme devront être *remis* en valeur pour pouvoir maîtriser le sida.

Il ne faut pas qu'on condamne les gens qui sont attirés par les représentants du même sexe. Il ne faut non plus *abandonner* les *malheureux* qui ont été touchés par la maladie - ce sont quand-même des humains qui ont le droit de vivre. Le sida, c'est une peste de notre *ère*, on n'a qu'à faire face à la réalité.

Niveau II (29):

Réflexions sur le rôle de la littérature à partir de l'avis d'Antoine Maillet

En considérant l'*extrait* sur la tâche de la littérature d'Antoine Maillet, il est *requis* de prendre comme point de départ *aléatoire* le '*mythe*' de la création du monde de la *Bible*. Dans la *Génèse* on ne parle que de l'*essence* physique de l'être humain ; on ne parle ni de raison ni de conséquent. Ainsi pour l'homme rest-t-il cette éternelle réflexion *métaphysique* et *psychologique* qui est *transmise* en paroles par les écrivains. Ceux-ci *possèdent* la capacité de *sublimation* de leurs pensées, ou bien de *traumas refoulés*. Les grandes lignes des sentiments propres à tout homme prennent forme par leur *plume*. Tout écrivain considéré comme *classique* révèle, peut-être à son insu, un *message* universel ou parfois même bouleversant.

Néanmoins ce *processus* de création n'est pas complété sans le lecteur. Bien que la *critique* littéraire vise à *refuter* la justification de la lecture *naïve*, elle est pourtant essentielle pour faire vivre l'écrivain. Aujourd'hui, à l'époque *postmoderne*, où il semble que tout ait déjà été dit et fait, la littérature doit aussi être *confrontée* aux nouvelles méthodes. Par conséquent le lecteur trouve souvent *hermétique* et *répulsive* la lecture, lorsque la forme *surpasse* le sens. En dépit de ce dernier la littérature ne cesse jamais à attirer le public. Elle est un *substitut* de la vie et représente une *fuite romantique* même jusqu'au point qu'elle devient plus réelle que la réalité - et ainsi le comble de la création.

EXEMPLE 3 : peu de mots réellement uniques dans les dissertations d'un étudiant

Niveau I (2):

Sujet 1

"Naître en Finlande, c'est gagner le gros lot", dit-on. En Finlande nous avons toutes les choses qui sont vraiment nécessaires pour vivre. Si on compare la Finlande et les pays du Tiers Monde, la Finlande est un pays riche.

Les Finlandais pensent qu'ils ont de la chance ; ils peuvent acheter toutes sortes de choses. Mais cela coûte. La Finlande est un pays riche mais aussi très chère. Pour vivre en Finlande, on doit avoir beaucoup d'argent. Les maisons et les appartements sont chers. Il faut que nous travaillions beaucoup parce que nous devons aussi payer plus de la nourriture que les autres Européens en paient.

Malgré les problèmes déjà *mentionnés*, la vie en Finlande n'est pas mal. Nous avons les lacs purs et les grands forêts, qui sont moins *pollués* que ceux en Allemagne, par exemple. J'aime vivre en Finlande. Je ne quitterai jamais mon pays si j'ai assez d'argent pour habiter ici.

Niveau II (3):

Sujet I. Existe-t-il de pieux mensonges ?

Les pieux mensonges existent, certainement. S'ils sont utilisés beaucoup ou pas autant, cela dépend de la culture et de la relation entre les personnes.

Si les gens se connaissent bien, il vaudrait mieux d'être honnête car tard ou tôt la vérité *apparaîtra*. La vérité est très *déstructive* pour la relation si elle est construite aux mensonges. Il y a des façons pour dire la vérité sans qu'on blesse des gens.

Les pieux mensonges peuvent bien servir des gens quand il s'agit des voisins ou des connaissances du travail. Avec eux, il est surtout important d'être poli et garder les relations amicales.

Il y a des cultures où les vérités inagréables sont strictement *tabou* dans la conversation. Dans le monde arabe la politesse est "obligatoire" et les plus pieux mensonges font une grande partie dans la communication. Dans la société finlandaise la politesse n'a pas un rôle aussi grand.

EXEMPLE 4 : peu de mots réellement uniques dans les dissertations d'un étudiant

Niveau I (4) :

Le Sida

Aujourd'hui le sida est peut-être la maladie dont on a le plus peur dans le monde entier. La maladie s'est *étendue* partout, surtout dans les pays *peuplés*.

Combien des victimes du sida on aurait pu sauver, s'ils avaient su se protéger dans leurs relations sexuelles, s'ils avaient su qu'il existe une maladie pareille ? Il y en aurait eu beaucoup. Mais jusqu'à ces dernières années on n'a pas eu beaucoup

d'information sur cette maladie et maintenant le problème est de trouver un moyen pour guérir les gens qui l'ont eue. C'est aussi important de donner tout le monde d'information correcte sur cette maladie. Dans plusieurs pays on a déjà commencé les campagnes de prévention, mais par exemple au Tiers Monde la plupart des gens ne sait rien du tout du sida.

Chez nous en Finlande on en a bien donn d'information, mais si on a eu assez, j'en doute. Il faudrait surtout penser des jeunes, garantir leur *future* sans sida. Il faudrait faire les gens penser *raisonnablement* : "On va ensemble à l'examen de dépistage du sida avant qu'on commence une relation sexuelle" ou bien "on utilise la protection en faisant l'amour". Pourquoi prendre des risques ? Il faut battre ce sida, maladie horrible.

Niveau II (3):

Les étrangers en Finlande

Ces dernières années le nombre d'étrangers a beaucoup augmenté aussi chez nous ; La Finlande s'internationalise ! Mais va-t-on s'habituer les uns aux autres ?

Quand les étrangers arrivent en Finlande, ils trouvent très souvent devant eux la vie qui diffère totalement de celle dans leur propre pays. Ils sont dans une situation où ils doivent tout recommencer. Mais en décidant de quitter leur pays, ils ont dû accepter le fait que leur vie va changer. Ils doivent apprendre les coutumes et la langue du pays et sûrement ils font de leur mieux en s'*adaptant* nous, bien que ça ne soit pas toujours facile.

Et quelle est notre réaction ? Je pense que la plupart de nous accepte un certain nombre d'étrangers chez nous. Ils pensent qu'il est bien de montrer aux autres pays, que nous en accueillons aussi, mais qu'il ne faut pas être trop gentil... Et qu'ils ne causent pas de problèmes !

Mais d'autre part il y a entre nous des gens qui ne veulent pas accueillir des étrangers. A leur opinion la situation sociale n'est pas *suffisamment* bonne pour les accueillir ; il faut d'abord prendre *soin* de nous-mêmes.

A mon avis il est bien d'avoir des étrangers, ça change. En ce moment il n'y en a pas trop, mais je trouve qu'il y a la limite quelque part. Et il faut les accueillir peu à peu pour que tout le monde puisse s'habituer à la nouvelle situation.

EXEMPLE 5 : l'amélioration dans la quantité des mots réellement uniques chez un étudiant

Niveau I (3):

L'affaire des "Versets sataniques"

Tout le monde occidental a été *stupéfait* par la menace de mort que l'imam Khomeiny a proférée à l'encontre de Salma Rushdie. On a dit que cette menace est une attaque contre toute l'humanité et contre le droit à s'exprimer librement et il ne le faut pas accepter. Mais à dire quelque chose n'est pas la même chose qu'à faire quelque chose. On l'a marqué bien ici en Finlande.

Même le gouvernement finlandais ne peut pas dire que la menace de l'imam Khomeiny est justifiée, mais si on devrait aussi *réagir* contre cette menace, on veut mieux l'oublier. Bien sûr Finlande a des relations d'affaires avec l'Iran et elle peut en profiter, particulièrement si les autres pays occidentaux sont rompus ses relations avec l'Iran, mais c'est incroyable que quand il s'agit de la justification d'exécuter quelqu'un à cause de ses opinions, le gouvernement finlandais veut éviter le problème. On a l'impression que la vie humaine ne coûte pas cher quand les affaires peuvent être *dérangées*. Si la position officielle finlandaise est telle sur cette affaire, on peut seulement demander quelle est la qualité de la vie en Finlande.

Niveau II (14) :

Les étrangers en Finlande

En ce moment le monde semble très différent et un peu effrayant aussi à beaucoup de finlandais. Pendant des centaines d'années les finlandais ont été les seuls à habiter ce pays *isolé* sauf quelques *suédois* qui ressemblent quand-même beaucoup aux finlandais. Tout de suite la vie tranquille et normale a l'air d'être bouleversé. A la télévision on ne parle que de l'intégration et des réfugiés politiques qui demandent de l'*asile* en Finlande. Il y a beaucoup de finlandais qui *craignent* que tout cela va détruire notre culture nationale et encore pire, cela va *détériorer* notre niveau de vie.

Il y a toujours des problèmes si de grandes masses de l'extérieur du pays entrent à un autre pays qui a une culture différente de celle des *immigrés*. Mais jusqu'à aujourd'hui il y a tellement peu d'étrangers ici que il est *exagéré* d'avoir peur pour notre culture nationale. Au contraire, les finlandais pourraient pour une fois sortir de leur isolation et rechercher ce qu'ils pourraient apprendre des autres cultures et des gens qui présentent une différente façon de vivre. Une des fautes des finlandais semble être que souvent nous pensons que ce qui est différent est *impérativement redoutant*, donc *negatif*.

De point de vue économique la question d'étrangers n'est pas facile. Tant que les étrangers sont *éduqués* et ils travaillent il n'y a pas de problèmes. Quand il s'agit des réfugiés politiques, le problème est plus compliqué. C'est vraie que chaque réfugié politique coûte énormément à l'état. A mon avis la Finlande doit accueillir sa part des réfugiés politiques, parce que c'est quand-même un pays riche. Mais bien-sûr il faut être modéré et tenir compte de la capacité du pays. Cela n'aide personne, ni les réfugiés ni les finlandais, si les quantités *croissantes* causent tant de problèmes *sociologiques* et économiques qu'on ne peut plus les maîtriser.

Je pense qu'il n'y a pas de retour au passé et aussi pour la Finlande c'est *indispensable* de ouvrir ses frontières aux étrangers et aux marchés étrangers. Cette chose va causer des problèmes, mais à l'autre cas les problèmes seraient plus grands.

EXEMPLE 6 : l'amélioration dans la quantité des mots réellement uniques chez un étudiant

Niveau I (3) :

Naître en Finlande, c'est gagner le gros lot"

La Finlande - un petit pays neutre, calme et formidable avec les milliers de *dizaines* de lacs, la nature pure et beaucoup de forêts, bien sûr ! Puis n'oublions pas les nuits d'été de *Saint-Jean*, la politesse des gens et notre système sociale très bien. Oui, tout ça est bien vrai, mais ce n'est pas toute la vérité...

Le niveau de la vie, en générale, est assez haut en Finlande, mais les prix sont hauts aussi. Il me semble que vraiment tout ici coûte cher. Aujourd'hui il n'y a pas assez de logements pour les jeunes, il y a des chômeurs et on devrait penser plus à la *protection* de la nature - si non, peut-être dans quelques ans la nature verte n'existera plus...

Alors, moi, je suis fière d'être Finlandaise, mais il faut qu'on comprenne les problèmes lesquels existent aussi dans notre pays. A cause de cela je dois dire que cette phrase "Naître en Finlande, c'est gagner le gros lot" a deux cotés différentes - c'est vrai et pas vrai en même temps !

Niveau II (18) :

Sujet 2

Après avoir passé *onze* mois en Provence au Sud de la France, j'ai réalisé que je pourrais vraiment y aller vivre. Pourquoi pas s'installer sous le ciel bleu de Provence ?

Si un jour je quitte la Finlande pour toujours, cela va sans dire que je déménagerai en France et surtout dans le Midi. Le mode de vie des Méridionaux me plaît énormément. Il me semble qu'ils vivent *pleinement* chaque *instant* que la journée leur porte mais pourtant ils vivent d'une façon calme, sans être angoissés par les soucis du monde. Les gens de Provence sont tout simplement *aimables* : ils sont amicaux, souriants et très ouverts. Là, il m'est arrivée même de parler avec des *inconnus* - une chose qui serait impossible en Finlande ! Ce que je trouve extraordinaire, c'est *l'humour* des Méridionaux ; il est vrai qu'ils rigolent *constamment* ! En général, je dirais que ces gens-là ne sont pas *superficiels*, ils ont pu garder encore la "vraie nature humaine".

Quant à la nature de cette région, je l'aime bien aussi. Qui n'aimerait pas voir le *printemps* avec toutes les *fleurs fleurissantes* déjà en *février* ou bien les *champs de lavande* en été ? De plus, la Provence a ses propres odeurs des herbes, comme le *thym*, le *romarin* et la *sauge*. Que c'est beau la vie en Provence !

Où est-ce que je me trouve dans vingt ans, seulement Dieu le sait ! Mais si j'ai un peu de chance, il m'*envoyera* en Provence...