

Le développement de la complexité syntaxique en français langue seconde : complexité structurelle et diversité

Bastien De Clercq

Vrije Universiteit Brussel
bastien.de.clercq@vub.ac.be

Résumé. Dans cette contribution, nous examinerons le rôle de la complexité syntaxique dans le développement du français langue seconde (FLS). Parmi la multitude d'approches envers la complexité linguistique, dans l'étude de l'apprentissage de langues secondes celle-ci a surtout été étudiée dans une optique appliquée et quantitative, qui la considère comme une composante essentielle du développement linguistique. Si la complexité syntaxique est typiquement vue comme une notion multidimensionnelle représentant la complexité structurelle et la diversité des structures maîtrisées, la pratique actuelle l'a toutefois généralement analysée de façon réductrice, en utilisant seulement une ou deux mesures visant uniquement la complexité structurelle. Cette contribution présente les résultats d'une étude empirique sur le développement du français langue seconde en contexte d'apprentissage guidé. Elle propose une démarche systématique qui décompose la notion de complexité en complexité structurelle et diversité à différents niveaux syntaxiques afin de mieux rendre compte du lien entre la complexité syntaxique et la compétence linguistique. À l'aide d'un corpus oral composé de productions de 100 apprenants néerlandophones de FLS et de 25 locuteurs natifs, l'étude indique que la complexité structurelle et la diversité augmentent surtout au niveau multipropositionnel, même si les apprenants n'atteignent généralement pas le degré de complexité observé dans les productions de locuteurs natifs. Un examen détaillé révèle également des différences distributionnelles entre les natifs et les apprenants, surtout au niveau du syntagme nominal.

Abstract. In this paper, we will examine the role of syntactic complexity in the development of L2 French. In second language research, syntactic complexity has frequently been studied through a quantitative approach that considers it one of the main dimensions of linguistic development. While syntactic complexity is typically seen as a multidimensional construct, representing both the internal complexity of structures and the diversity of acquired structures, in practice, however, it has generally been studied in a reductionist way, restricted to one or two measures tapping into structural complexity only. This paper presents the results of an empirical study on the development of L2 French in an instructed language acquisition context. It proposes a systematic approach to syntactic complexity, that represents structural complexity and syntactic diversity in equal measure, at various syntactic levels, in order to better describe the link between syntactic complexity and linguistic proficiency. Using a corpus of 100 oral narratives by L2 learners and 25 narratives by native speakers of French, the study indicates that syntactic complexity mostly increases at the multiclausal level, even if the learners generally don't use the same degree of complexity as the native speakers. A closer look equally reveals distributional differences of different types of subclauses between the L1 and L2 groups.

1 Introduction

La complexité est sans doute un des concepts les plus polyvalents en linguistique, qui a fait l'objet de nombreuses approches théoriques et appliquées. Récemment, cet intérêt inter-disciplinaire a engendré un nombre d'ouvrages et de numéros thématiques de revues dédiées au sujet et qui illustrent l'hétérogénéité de ces approches (e.g. Paprocka-Piotrowska et al., 2013; Kortmann & Szmrecsanyi, 2012; Housen et al., 2012; Housen & Kuiken, 2009; Miestamo et al., 2008). Généralement, la notion de complexité est définie en fonction, d'une part, du nombre d'éléments dans un système, et d'autre part, des connexions ou relations entre ces éléments (Miestamo, 2009). Ici, la complexité linguistique et, plus spécifiquement, la complexité syntaxique seront étudiées dans une optique appliquée et quantitative. Cette démarche, surtout adoptée dans les recherches sur l'apprentissage de langues secondes (ALS), considère la complexité syntaxique comme une composante essentielle du développement linguistique, qui est souvent étudiée au niveau de la production orale ou écrite comme indice de compétence linguistique (cf. Wolfe-Quintero et al., 1998; Housen & Kuiken, 2009; Ortega, 2012).

Or, le rôle de la complexité linguistique dans l'ALS ainsi que sa mesure ont récemment été mis en cause. D'une part, de nombreuses études ont démontré que le degré de complexité au niveau de la parole ne varie pas uniquement en fonction de la compétence linguistique du locuteur, mais aussi en fonction de restrictions cognitives, contextuelles, stylistiques et typologiques (Skehan, 1998, 2009; Pallotti, 2009; Kormos & Trebits, 2012; Lambert & Kormos, 2014; De Clercq, en préparation). En même temps, l'opérationnalisation de la complexité syntaxique a souvent débouché sur une vision réductrice de celle-ci, même si la complexité est typiquement vue comme une notion multidimensionnelle dont les différentes composantes se développent à différents stades d'ALS : le nombre de mesures syntaxiques se limite souvent à une seule mesure de subordination et/ou une seule mesure de longueur d'unité multipropositionnelle (Lambert & Kormos, 2014; Norris & Ortega, 2009). Les recherches antérieures dans ce domaine ont de plus majoritairement été consacrées au développement de la complexité à l'écrit et en anglais (cf. Wolfe-Quintero et al., 1998 pour un aperçu; Bulté, 2013 et Verspoor et al., 2012 pour quelques exemples récents), de sorte que le lien entre complexité et compétence et sa validité interlangagière restent comparativement sous-étudiés à l'oral (Lambert & Kormos, 2014).

Cette contribution présente les résultats d'une étude empirique et quantitative du développement du français langue seconde (FLS) en contexte d'apprentissage guidé. Afin de traiter l'aspect multidimensionnel de la complexité syntaxique avec précision, cette étude propose, à l'instar des études sur la complexité lexicale (Malvern et al., 2004; Jarvis & Daller, 2013; De Clercq, 2015) et, plus récemment, la complexité morphologique (Pallotti, 2015 ; De Clercq 2016), une approche qui prend en compte non seulement la complexité structurelle à différents niveaux syntaxiques, mais aussi la diversité des structures. Elle a ainsi comme objectif principal de rendre compte du développement de la complexité syntaxique dans la production orale à travers différents niveaux de compétence et confrontera, pour ce faire, des données orales de 100 apprenants néerlandophones de FLS, situés à 4 niveaux de compétence, à des données de 25 locuteurs natifs.

Dans la première partie, la notion de complexité syntaxique et son rôle dans le procès de développement linguistique seront discutés. Dans la deuxième partie, la méthodologie de cette étude sera présentée. Enfin, cette section sera suivie par la présentation des résultats et des analyses statistiques.

2 Complexité syntaxique en ALS

La notion de complexité syntaxique (et linguistique) correspond à un nombre d'acceptions touchant différents aspects parfois hétérogènes, tels que l'enchaînement, le degré d'ellipse, le marquage, la redondance, la difficulté cognitive, le rôle dans l'acquisition de la langue, ou encore la variation paradigmatique ou syntagmatique (Kirchmeyer, 2002 ; Bartning & Kirchmeyer, 2003 ; Szmrecsanyi, 2004 ; Bulté & Housen, 2012 ; Szmrecsanyi & Kortmann, 2012 ; Pierrard & Housen, 2013). Généralement, cette multitude de définitions se décompose en deux approches soulignant d'une part la complexité absolue et d'autre part la complexité relative (Dahl, 2009 ; Miestamo, 2008, 2009). La complexité absolue (ou structurelle), qui fera le sujet de cette contribution, réfère typiquement à des caractéristiques linguistiques inhérentes, auxquelles nous avons fait allusion ci-dessus : le nombre d'éléments dans un système donné et le nombre et type de liens entre ces éléments. Par contre, la complexité relative (ou difficulté) est définie par rapport au locuteur en fonction du coût cognitif, ou de la difficulté des structures ou du procès d'apprentissage.

Récemment, un intérêt renouvelé dans la matière a également révélé une préférence à réserver la notion de complexité à l'approche absolue. Pallotti (2015), par exemple, avance que si la complexité absolue et la complexité relative ne sont pas démêlées, nous risquons d'arriver à des constats tels que « les structures complexes [structurellement] sont souvent plus complexes [difficiles] et complexes [acquis tardivement] » (3 ; traduction de l'auteur). Une telle observation s'avère d'autant plus importante, que la pratique de la recherche et, plus spécifiquement, l'opérationnalisation de la complexité syntaxique ne sont que rarement systématiques en ce qui concerne cette distinction. Ainsi, maintes études utilisent des indices représentant les occurrences de certaines structures estimées difficiles, mais pas nécessairement structurellement complexes (pour un aperçu, voir Norris & Ortega, 2009 et Bulté & Housen, 2012).

Bien que nous adhérons en premier lieu à l'approche absolue, nous ferons toutefois également allusion aux aspects cognitifs, car les deux perspectives sont inextricablement liées, surtout dans les recherches en ALS. Ainsi, les objectifs principaux dans la recherche sur la complexité en ALS sont selon Ortega (2012 : 128) de quantifier la compétence linguistique, de décrire la performance et de tracer le développement linguistique, tandis que Bulté & Housen (2012) avancent également l'étude de la complexité en soi comme objectif additionnel. Selon cette dernière approche, l'utilisation de mesures de complexité en tant qu'indices de développement a comme conditions préalables que la complexité soit opérationnalisée indépendamment de facteurs relatifs, que sa mesure représente différentes dimensions de complexité, et que le lien entre complexité et développement soit ensuite démontré de façon empirique.

Même si la dimension syntaxique est, parmi les différentes dimensions de complexité linguistique, la plus fréquemment étudiée dans les recherches sur la complexité en ALS, celles-ci s'y prennent souvent de façon réductrice en utilisant typiquement une seule mesure de complexité représentant la longueur moyenne d'une unité syntaxique ou le nombre de propositions subordonnées dans un texte (cf. Lambert & Kormos, 2014 ; Ortega, 2012). Premièrement, la longueur d'unité (en mots ou en morphèmes) est employée comme approximation de la structure interne de ces unités et suppose que, plus une structure est longue, plus elle sera structurellement complexe. Bien qu'il s'agisse clairement d'une approximation et simplification de la notion de complexité, Szmrecsanyi (2004) note toutefois que, pour un corpus de données écrites en anglais, la corrélation entre des mesures de longueur et le nombre de nœuds syntaxiques est très forte ($r = 0,976-0,989$). De plus, Norris & Ortega (2009) et Pallotti (2015) soulignent l'importance de mesurer l'élaboration syntaxique à différents niveaux, représentant des développements au sein des syntagmes, de la proposition, mais aussi des unités multipropositionnelles, car différents développements de complexité sont attendus à différents niveaux syntaxiques à différents stades acquisitionnels (cf. ci-dessous). Dans certains cas, ces développements, par exemple l'enchaînement de propositions dans les SN, se produiraient aussi plutôt à l'écrit qu'à l'oral (Ortega, 2012 ; Biber et al.,

2011). Deuxièmement, Norris & Ortega (2009) avancent les mesures de subordination (par exemple, le nombre de propositions dépendantes par unité multipropositionnelle) comme catégorie de complexification séparée, car elle représente uniquement les développements du nombre de liens interpropositionnels et non pas les développements au sein des propositions¹.

Or, les recherches sur la complexité syntaxique se limitent typiquement à la complexité structurelle et n'intègrent que rarement la diversité des structures. La notion de diversité a surtout été étudiée pour la complexité lexicale (par exemple, Malvern et al., 2004; Jarvis & Daller, 2013; De Clercq, 2015) et, plus récemment, pour la complexité morphologique (Pallotti 2015, à paraître; De Clercq, 2016). Généralement, la diversité linguistique est perçue comme une dimension au niveau de la production qui représente le répertoire ou la gamme d'éléments (lexicaux, morphologiques, syntaxiques, phonologiques). Cet aspect a déjà été analysé pour la complexité syntaxique par des mesures de distribution, représentant la proportion de différentes structures syntaxiques (par exemple, Welcomme, 2013 pour les jonctions interpropositionnelles). Or, il s'agit ici de mesures de diversité indirectes : ce type de représentation distributionnelle offre un aperçu des structures dans un texte, dont il peut être inféré que plus différentes catégories sont présentes, plus le texte sera syntaxiquement divers. Bien qu'informatrice, cette approche n'est typiquement pas intégrée dans l'étude de la complexité syntaxique telle que décrite ci-dessus (pour une exception, voir Verspoor et al., 2012), et elle n'offre, contrairement aux mesures de diversité lexicale ou morphologique, pas d'indice direct de complexité comparable aux mesures de complexité structurelle.

Les recherches antérieures sur le rôle de complexité syntaxique dans l'ALS ont révélé quelques tendances développementales générales, qui sont, dans le cas du français L2, le mieux documentées pour des apprenants suédophones (cf. Bartning & Kirchmeyer, 2003; Bartning & Schlyter, 2004), mais qui s'appliquent également à d'autres groupes d'apprenants (par exemple, Welcomme, 2013 pour des apprenants néerlandophones). De plus, ces tendances ont également été observées pour d'autres langues cibles, notamment l'anglais (par exemple, Bulté, 2013; Verspoor et al., 2012; Ortega, 2012). Dans un premier stade initial, le développement syntaxique passe par une complexification au sein de la proposition d'une organisation nominale à un emploi de formes verbales non-finies et finies. La complexification peut ensuite être caractérisée comme un développement continu des phrases simples à des phrases complexes, et l'apparition de constructions infinitives et participiales marque surtout les niveaux avancés. Au long de ce développement, les apprenants maîtrisent progressivement des constructions telles que la négation, mais diversifient aussi la gamme de structures employées d'un point de vue formel ainsi que fonctionnel. Le développement aux niveaux les plus avancés serait ensuite marqué par une réorganisation du syntagme nominal, qui intègre des structures de nature précédemment propositionnelle. Afin d'examiner le développement de la complexité linguistique, il sera donc non seulement nécessaire de représenter la complexité des structures, mais aussi le degré de diversité de ces structures au niveau du texte.

Ces approches ont été révélatrices même si les résultats doivent être considérés avec prudence. L'approche quantitative envers la complexité syntaxique a surtout été adoptée pour l'anglais et l'étude de la complexité en FLS bénéficierait également de l'examen de différents contextes d'apprentissage. Il est aussi essentiel de vérifier dans quelle mesure les tendances antérieurement observées s'appliquent au développement à l'oral, car une large partie des études se concentre sur l'écrit, de sorte que certains développements observés, tels que le développement avancé au sein du SN, ne seraient pas caractéristiques de l'oral. De plus, l'étude de productions orales entraîne certaines problématiques méthodologiques. Le démarquage des énoncés au niveau multipropositionnel, par exemple, est compliqué par le caractère plus elliptique et fragmenté de l'oral, raison pour laquelle d'autres définitions des unités multipropositionnelles ont été proposées pour l'oral (par exemple, l'unité AS de Foster et al., 2000) et pour l'écrit (par exemple, la 'T-unit' de Hunt, 1965). L'étude de l'oral a également révélé un autre rôle du connecteur *et* dans le développement : si, à l'écrit, la coordination constitue parfois une étape intermédiaire dans la complexification syntaxique entre l'utilisation d'énoncés simples et de phrases complexes (par exemple, Bardovi-Harlig, 1992), le *et* à l'oral fonctionne fréquemment comme outil d'enchaînement narratif utilisé dès les premiers stades d'apprentissage, créant de la cohésion au niveau du texte plutôt qu'au niveau de l'énoncé (Beaman, 1986; Kerr-Barnes, 1998). Finalement, jusqu'à

maintenant, peu d'études ont intégré les différentes dimensions de complexité à différents niveaux syntaxiques de façon systématique. Surtout la diversité syntaxique, à notre connaissance, a été négligée dans cette approche, même si d'autres approches moins explicitement liées à la complexité touchent néanmoins à cette dimension.

La démarche de cette étude abordera cette problématique en intégrant la dimension structurelle ainsi que la diversité syntaxique dans une exploration du développement oral du FLS à différents niveaux syntaxiques et introduira pour ce faire une approche plus systématique envers la mesure de la diversité syntaxique.

3 Méthodologie

3.1 Questions de recherche et hypothèses

L'objectif de cette contribution est d'examiner le développement de la complexité syntaxique dans un corpus oral de FLS, représentant quatre niveaux de compétence linguistique, afin de répondre aux questions de recherche suivantes :

- a) Comment se développe la complexité syntaxique au niveau structurel ?
- b) Comment se développe la complexité syntaxique au niveau de la diversité ?
- c) Quelles mesures de complexité syntaxique peuvent fonctionner comme indice de compétence ?

Nous émettons les hypothèses suivantes :

H1) La complexification structurelle dans les textes des apprenants aura d'abord lieu au niveau du syntagme nominal et à l'intérieur des propositions. Ensuite nous nous attendons à une complexification structurelle par l'utilisation de phrases complexes.

H2) La diversification syntaxique, surtout au niveau multipropositionnel, se manifestera plutôt aux niveaux plus avancés.

H3) Finalement, nous nous attendons à ce que les mesures de complexification structurelle au niveau multipropositionnel fonctionneront le plus comme indice de compétence, car celles-ci représentent également le développement aux niveaux sous-jacents, i.e. les syntagmes nominaux et les propositions.

3.2 Participants et données

Cette étude utilise un corpus composé de récits oraux, racontés par des apprenants du FLE dans l'enseignement secondaire néerlandophone en Flandre, Belgique. Le corpus consiste de données tirées de trois projets de recherche menés à la Vrije Universiteit Brussel (VUB), ayant comme but l'évaluation de la compétence linguistique dans l'enseignement secondaire dans différentes écoles flamandes. Les projets de recherche représentaient originellement un total de 286 apprenants au début (N : 189 ; âges : 12-13 ans), au milieu (N = 28 ; âges : 14-15 ans) et à la fin (N = 87 ; âges : 17-18 ans). Ces apprenants sont des locuteurs natifs et monolingues du néerlandais dont le contact principal (et souvent unique) avec la langue cible se passe par l'enseignement scolaire. En Flandre, l'enseignement du français L2 débute généralement à la fin de l'école primaire, mais il s'agit d'un contact limité dont l'apport a été mis en doute (Spoelders, 1997). Le nombre d'heures d'enseignement pour les groupes inclus dans l'étude revient à environ 4 à 5 heures par semaine et comprend enseigner la compétence écrite et orale dans la même mesure, dans les six ans d'enseignement secondaire. Si la langue maternelle des apprenants est encore utilisée comme support au début de l'enseignement, les cours ont entièrement lieu dans la langue cible à la fin. Un groupe de locuteurs natifs du français a également été inclus dans l'étude. Il s'agit de 32 locuteurs âgés de 12 à 13 ans, suivant l'enseignement francophone à Bruxelles. Ce groupe a été choisi afin de correspondre à l'âge des apprenants les plus jeunes et servait en tant que groupe de référence dans

les projets de recherche originaux. Le groupe de référence était estimé nécessaire afin de mieux pouvoir interpréter les résultats des apprenants : non seulement tâchons nous de vérifier si les textes des apprenants se complexifient sur le plan syntaxique, mais nous sommes également intéressés à voir si le degré de complexité correspond à celui observé dans des textes de locuteurs natifs. Les données ont été recueillies lors d'entretiens individuels dans lesquels les participants devaient raconter « l'histoire de la grenouille » (Mayer, 1969), une histoire en images d'un garçon qui part à la recherche de sa grenouille.

Les productions ont ensuite été réparties en quatre groupes de compétence linguistique croissante, d'après l'année scolaire des participants et de trois mesures de précision. Ces mesures de précision étaient estimées nécessaires car les distinctions sur la base de l'année scolaire aboutissaient à des regroupements particulièrement hétérogènes quant à la compétence des apprenants. Chaque texte a donc été évalué à l'aide d'une mesure de précision générale (nombre de phrases correctes) et de deux mesures spécifiques (pourcentage de verbes conjugués correctement, pourcentage d'emplois prépositionnels corrects). Afin d'homogénéiser les groupes pour les analyses statistiques, nous les avons ensuite réduits à 25 participants par groupe, utilisant une sélection aléatoire. Bien que nous perdions ainsi une grande partie des données, nous estimons la grandeur des groupes suffisante pour les analyses prévues.

Les transcriptions ont alors été préparées pour les analyses en format CHAT (Macwhinney, 2000). A cause du caractère spontané, fragmenté et parfois interactif des données, certaines parties des transcriptions ont été exclues des analyses. D'abord, les répétitions, hésitations ou corrections n'ont pas été incluses (exemple 1). Ensuite, les énoncés où un constituant obligatoire manquait ou dont le verbe ne se trouvait pas dans la langue cible ont été supprimés (exemple 2). Enfin, vu que le niveau d'interaction des entretiens est variable, les énoncés de nature dialogique ont été supprimés (exemple 3). Ces énoncés sont souvent elliptiques et donc structurellement différents des énoncés monologiques.

- (1) et ils voient
[que le que] que la grenouille n(e) est pas là.
- (2) le garçon euh dans le steen [= nl pierre].
- (3) - pourquoi est-ce-qu(e) il le fait?
- pour appelE le petit animal.

3.3 Mesures de complexité syntaxique

Un éventail de mesures de complexité syntaxique a été appliqué aux transcriptions des données, représentant d'une part la complexité structurelle et d'autre part la diversité syntaxique. Les mesures de complexité structurelle suivent les recommandations d'Ortega (2012) et Pallotti (2015) et se basent sur la longueur de différentes unités syntaxiques, en comptant le nombre de mots à plusieurs niveaux syntaxiques, i.e. le syntagme nominal (LonSN), la proposition (LonProp), et l'unité AS (LonAS), et le nombre de propositions par unité AS (Prop/AS). L'unité AS ("Analysis of Speech Unit") est une unité multipropositionnelle proposée par Foster et al. (2000) pour l'analyse de données orales et consiste, dans nos données, en une proposition matrice indépendante et chaque proposition qui dépend syntaxiquement de celle-ci. Seuls les mots dans la langue cible, ou dont la structure phonologique ou morphologique correspondent à la langue cible, ont été inclus.

Contrairement à la complexité structurelle, la diversité syntaxique n'a pas été étudiée systématiquement, de sorte que les mesures proposées dans cette étude sont exploratrices. Trois types de mesures de diversité ont été employés. Premièrement, l'écart type des longueurs d'unité est utilisé comme approximation de la diversité syntaxique, mathématiquement proche des mesures structurelles. Cette mesure s'appuie sur le raisonnement que, si la longueur de phrase peut fonctionner comme indice de complexité structurelle, plus un texte contient des phrases de longueur variable, plus la variation mesurée par l'écart type sera haute. Par conséquent, l'écart type pourrait distinguer entre des passages syntaxiquement divers que les mesures de longueur de phrase traiteraient comme identiques (exemples 4 et 5).

(4) Longueur moyenne de proposition : 7 mots ; Variation : 0.

et le chien regardait dans le bocal.
mais sa tête était dans le bocal.
alors le garçon il ouvre la fenêtre.

(5) Longueur moyenne de proposition : 7 mots ; Variation : 2.

et puis il a crié.
et alors il s(e) était tenu à une branche.
qui était une corne d(e) un cerf.

Or, il est évident que ces mesures sont limitées et superficielles, car la longueur de phrase ne peut pas identifier les différences structurelles entre les exemples (6) et (7). Comme mentionné ci-dessus, l'emploi de telles mesures est fondé sur leur fonctionnement en tant qu'*approximations* de la complexité syntaxique et sur leur forte corrélation avec d'autres mesures plus spécifiques. L'avantage de cette mesure est qu'elle est facile à calculer et à interpréter, en plus d'être comparable aux mesures de longueur. Ici, elle sera appliquée au syntagme nominal (DivSN), à la proposition (DivProp) et à l'unité AS (DivAS).

(6) mais l(e) animal ne vient pas. (6 mots)

(7) le garçon regarde dans un trou. (6 mots)

Deuxièmement, nous utilisons comme indice plus traditionnel et général la distribution de différents types de proposition, dont les trois types généraux sont la matrice (%mat), la subordonnée (%sub) et la coordonnée (%coor). Suivant Cosme (2007) et Biber et al. (2011), les propositions subordonnées sont ensuite étiquetées selon des catégories formelles (finie/non-finie) et fonctionnelles (complétive, adverbiale, infinitive). Les phrases coordonnées constituent dans notre analyse une catégorie particulière car elles ne représentent que la coordination avec ellipse. Surtout dans des données orales et narratives, les conjonctions de coordination sont surutilisées comme outils d'enchaînement narratif, fonctionnant plus comme tics de langage, caractéristiques de l'oral, que comme de véritables connecteurs syntaxiques (Beaman, 1986 ; Kerr-Barnes, 1998) (exemple 8).

(8) [mat] puis le petit garçon va dans l(e) arbre.
[mat] et il voit dans un trou.
[mat] et le chien pousse contre l(e) arbre.
[mat] et la ruche des abeille tombe sur la terre.
[mat] et les abeille parte.

La surreprésentation de tels marqueurs nous a donc menés à considérer ces énoncés comme enchaînements de phrases matrices. Or, nous avons distingué les cas de pseudo-coordinations, ou coordonnées avec ellipse, comme catégorie séparée (cf. Bartning & Kirchmeyer, 2003), car celles-ci sont clairement dépendantes des matrices, mais ne correspondent toutefois pas aux catégories de subordonnées. Parmi les propositions en pseudo-coordination, nous avons inclus celles avec ellipse du sujet (exemple 9) ou avec ellipse d'un autre ou plusieurs constituants (exemple 10). Pour chacune des catégories parcourues ci-dessus, le taux (en pourcentage) sera rapporté.

(9) [mat] et le jeune homme emporte un petit crapaud
[coor] et va à la maison.

(10) [mat] et puis le garçon a tombé dans l(a) eau
[coor] et le chien aussi.

Troisièmement, nous proposons une mesure qui représente la diversité de combinaisons propositionnelles et qui a été développée par analogie à la mesure de diversité morphologique proposée par Pallotti (2015). Cette mesure, l'indice de diversité syntaxique (IDS), se fonde sur une analyse de chaque unité AS en tant que combinaison linéaire des différents types de proposition distingués ci-dessus (exemple 11). Après avoir pourvu chaque unité AS d'un code représentant sa structure sous-jacente, l'IDS est calculé en comparant le nombre de structures AS uniques dans et à travers des segments de cinq unités AS. La segmentation aléatoire des textes en cinq unités AS ainsi que le calcul de la mesure ont été effectués à

l'aide de l'outil d'analyse de Brezina & Pallotti (2015). Puisque cet outil exige la présence d'au moins 10 unités AS, 6 textes du niveau 1 ont dû être exclus de cette analyse.

- (11) [sub. adv. fin.] et quand le petit se réveille le matin .
 [mat] il voit .
 [sub. compl. fin.] que son grenouille est parti .
 = S.ADV-F+M+S.C-F

3.4 Analyses statistiques

Les analyses statistiques ont été faites à l'aide du logiciel SPSS (IBM, 2015) et consistent d'une combinaison de tests paramétriques (ANOVA) et non-paramétriques (Kruskal-Wallis), ainsi que de tests de corrélation linéaire, qui ne seront appliqués qu'aux groupes d'apprenants (ρ de Spearman). L'emploi de tests non-paramétriques est surtout incité par les résultats des catégories spécifiques, dont l'apparition dans les récits n'est parfois que sporadique. Les résultats des tests inférentiels et des tests post-hoc seront accompagnés de statistiques descriptives et seront résumés en sous-ensembles homogènes, indiquant des différences statistiquement significatives ($p < 0,05$) par des chevrons. Finalement, les tailles d'effet des tests paramétriques seront présentées sous forme de la valeur eta-carré partiel (η_p^2).

4 Résultats

Les résultats seront présentés en trois parties traitant, respectivement, de la complexité structurelle, la diversité syntaxique et une troisième partie comparant les deux dimensions. La partie sur la diversité contiendra d'une part les résultats des indices simples et d'autre part ceux des mesures de distribution.

4.1 Complexité structurelle

Les mesures de complexité structurelle, essentiellement des mesures de longueur d'unité syntaxique, révèlent un développement statistiquement significatif ($p < 0,05$) uniquement au niveau des unités AS, où l'on observe une augmentation progressive du nombre de propositions par unité AS (Prop/AS) du niveau 1 à 3 (Tableau 1). Cette augmentation s'observe également pour la longueur d'unité AS en mots (LonAS), dont l'augmentation n'est statistiquement significative qu'entre les niveaux 1 et 2, d'une part, et 3 et 4, d'autre part ($p < 0,05$). La longueur au niveau des SN (LonSN) ou de la proposition (LonProp) demeure stagnante dans les groupes d'apprenants. Comparés aux natifs, tous les groupes d'apprenants atteignent des scores significativement plus bas à chaque niveau syntaxique excepté la longueur de proposition ($p < 0,05$). Les plus grandes différences observées au niveau de l'unité AS sont également reflétées dans les tailles d'effet considérables (0,535 pour LonAS et 0,566 pour Prop/AS).

Groupe	LonSN			LonProp			LonAS			Prop/AS		
	<i>N</i>	<i>Moyenne</i>	<i>E.T.</i>	<i>N</i>	<i>Moyenne</i>	<i>E.T.</i>	<i>N</i>	<i>Moyenne</i>	<i>E.T.</i>	<i>N</i>	<i>Moyenne</i>	<i>E.T.</i>
1	25	2,49	0,43	25	5,38	0,72	25	5,79	0,92	25	1,08	0,11
2	25	2,47	0,27	25	5,41	0,61	25	6,36	0,92	25	1,18	0,13
3	25	2,57	0,21	25	5,72	0,68	25	7,80	1,56	25	1,36	0,20
4	25	2,58	0,20	25	5,54	0,43	25	7,92	0,97	25	1,43	0,18

Natifs	25	2,92	0,34	25	5,62	0,59	25	9,14	1,18	25	1,61	0,20
Post-hoc	1,2,3,4<N		1,2,3,4,N		1,2<3,4<N		1<2<3,4<N					
η_p^2	0,233		0,009		0,535		0,566					

Tableau 1 Résumé des mesures de complexité structurelle

Les corrélations entre ces mesures révèlent que toutes les mesures représentant la longueur en mots sont en corrélation significative les unes avec les autres ($p < 0,05$), fait qui s'explique en large partie aussi par l'unité de base partagée par ces trois mesures (Tableau 2). Or, les fortes corrélations entre les deux mesures d'unité AS (LonAS et Prop/AS), ainsi que l'absence d'une corrélation entre LonProp et Prop/AS indiquent que l'augmentation de LonAS est avant tout le résultat d'une tendance à introduire plus de propositions (subordonnées comme on le verra ci-dessous) plutôt que d'élaborer les structures à l'intérieur des propositions.

	LonSN	LonProp	LonAS	Prop/AS
<i>LonSN</i>	1	0,213*	0,390*	0,329*
<i>LonProp</i>	0,213*	1	0,493*	0,019
<i>LonAS</i>	0,390*	0,493*	1	0,846*
<i>Prop/AS</i>	0,329*	0,019	0,846*	1

Tableau 2 Corrélations entre les mesures de complexité structurelle (* = corrélation statistiquement significative ($p < 0,05$))

En somme, nous observons surtout une complexification structurelle au niveau multipropositionnel chez les apprenants, bien que ceux-ci n'atteignent généralement pas le niveau des natifs en ce qui concerne la complexité des SN et des unités AS.

4.2 Diversité syntaxique

Les résultats des indices simples de diversité syntaxique reflètent en quelque sorte ceux des mesures structurelles (Tableau 3). Les mesures de diversité au niveau de l'unité AS (DivAS et IDS) indiquent une différence statistiquement significative entre les niveaux 1-2 et 3-4 ($p < 0,05$). Si la différence entre les niveaux 1 et 2, respectivement 0,68 et 1,24, n'est pas statistiquement significative pour IDS, elle est toutefois considérable. A nouveau, aucune différence statistiquement significative n'a été observée dans les groupes d'apprenants au niveau des SN (DivSN), même si les scores augmentent légèrement. Contrairement aux résultats de complexité structurelle, la différence entre le niveau 1 et les niveaux 3-4 est statistiquement significative pour la diversité au niveau des propositions (DivProp) ($p < 0,05$). Une comparaison avec les natifs révèle que les apprenants n'atteignent généralement pas la même diversité que les natifs francophones ($p < 0,05$), sauf pour DivProp dont le taux est significativement différent entre les niveaux 1-2 et les natifs.

Groupe	DivSN			DivProp			DivAS			IDS		
	N	Moyenne	E.T.	N	Moyenne	E.T.	N	Moyenne	E.T.	N	Moyenne	E.T.
1	25	1,03	0,48	25	2,04	0,46	25	2,27	0,75	19	0,68	0,67
2	25	1,19	0,50	25	2,13	0,45	25	2,37	0,59	25	1,24	0,86

3	25	1,23	0,38	25	2,37	0,41	25	3,36	1,15	25	2,38	0,93
4	25	1,31	0,42	25	2,44	0,29	25	3,60	0,86	25	2,65	0,85
Natifs	25	2,08	0,75	25	2,51	0,35	25	4,75	1,03	25	3,43	0,67
Post-hoc	1,2,3,4<N			1,2<2,3,4<3,4,N			1,2<3,4<N			1,2<3,4<N		
η_p^2	0,342			0,179			0,513			0,596		

Tableau 3 Résumé des mesures de diversité : indices simples

Le Tableau 4 présente les résultats des analyses de corrélation. Quelques tendances spécifiques se distinguent, bien que toutes les mesures soient en corrélation significative ($p < 0,05$). Tout d'abord, les corrélations sont les plus élevées pour les deux mesures d'unité AS (DivAS et IDS) ($r = 0,798$), ce qui semble indiquer que les deux mesures couvrent effectivement le même type de diversité syntaxique. De plus, les corrélations les plus basses sont observées entre IDS, d'une part, et DivSN ($r = 0,326$) et DivProp ($r = 0,423$), d'autre part. Ceci suggère également qu'IDS mesure la diversité au niveau de l'AS plus indépendamment de ces deux niveaux que DivAS. Toutefois, il faut noter que, même si les deux mesures de diversité au niveau AS parviennent à représenter valablement leur concept sous-jacent, leur opérationnalisation diffère : IDS est la seule à incorporer des différences formelles et fonctionnelles dans les jonctions interpropositionnelles, tandis que, à son tour, DivAS représente également, en théorie, des différences structurelles à des niveaux sous-jacents (ici, la proposition et le SN).

Diversité	DivSN	DivProp	DivAS	IDS
DivSN	1	0,500*	0,503*	0,326*
DivProp	0,500*	1	0,664*	0,423*
DivAS	0,503*	0,664*	1	0,798*
IDS	0,326*	0,423*	0,798*	1

Tableau 4 Corrélations entre les mesures de diversité : indices simples (* = corrélation statistiquement significative ($p < 0,05$))

Les scores obtenus pour les mesures de distribution (Tableau 5) démontrent avant tout une évolution inverse des propositions matrices et subordonnées. Premièrement, il y a une baisse progressive pour les phrases matrices (%mat) qui est statistiquement significative du niveau 1 à 3, et une augmentation progressive du nombre de subordonnées (%sub), également statistiquement significative entre les niveaux 1, 2 et 3 ($p < 0,05$). Les proportions de propositions coordonnées avec ellipse (%coor) ne diffèrent nulle part de façon statistiquement significative, malgré que les scores semblent considérablement plus bas aux niveaux 3 et 4 qu'aux niveaux 1 et 2. Quant aux natifs, le taux de propositions matrices dans leurs textes est significativement plus bas que celui dans chacun des groupes d'apprenants ($p < 0,05$), tandis qu'aucune différence statistiquement significative n'est observée entre les natifs et les apprenants dans le cas des coordonnées, tandis que seuls les apprenants du niveau 4 atteignent un score comparable à celui des natifs dans le cas des subordonnées.

Groupe	%mat			%coor			%sub		
	N	Moyenne	E.T.	N	Moyenne	E.T.	N	Moyenne	E.T.
1	25	93,59%	7,79%	25	3,05%	4,04%	25	3,36%	6,43%
2	25	85,58%	9,33%	25	3,17%	4,69%	25	11,25%	10,02%

3	25	74,76%	9,13%	25	1,70%	2,38%	25	23,54%	9,25%
4	25	70,75%	8,67%	25	1,63%	1,82%	25	27,62%	8,51%
Natifs	25	61,97%	6,58%	25	3,91%	5,62%	25	34,12%	7,35%
Post-hoc	1>2>3,4>N			1,2,3,4,N			1<2<3,4<4,N		
η_p^2	0,648			/			0,647		

Tableau 5 Résumé des mesures de diversité : distribution

Remarquez que le développement des propositions subordonnées est l'image miroir de celle des matrices et que la catégorie additionnelle de pseudo-coordination n'intervient guère dans la distribution des autres catégories. Ce résultat est également confirmé par les analyses de corrélation qui indiquent uniquement une corrélation très forte ($r = -0,969$) et statistiquement significative ($p < 0,05$) entre les subordonnées et les matrices (Tableau 6).

Distribution	%mat	%coor	%sub
%mat	1	-0,12	-0,969*
%coor	-0,12	1	-0,094
%sub	-0,940*	-0,059	1

Tableau 6 Corrélation entre les mesures de diversité : distribution (* = corrélation statistiquement significative ($p < 0,05$))

Une analyse plus spécifique selon les types fonctionnels (Tableau 7) et formels (Tableau 8) de propositions subordonnées révèle que l'augmentation du nombre de subordonnées est surtout déterminée par les subordonnées complétives et celles contenant un verbe fini. Premièrement, si toutes les catégories fonctionnelles deviennent plus nombreuses dans les textes, cette augmentation est la moins marquée pour les relatives, qui sont sous-utilisées par tous les niveaux d'apprenants par comparaison aux natifs ($p < 0,05$). Les résultats de l'ANOVA révèlent des différences statistiquement significatives entre les niveaux 1-2 et 3-4 pour les complétives et pour les adverbiales ($p < 0,05$). Les scores absolus démontrent toutefois une évolution plus constante dans le cas des complétives. Notez également que les apprenants du niveau 1 emploient uniquement des subordonnées complétives et aucune adverbiale ou relative n'a été retrouvée. Le développement semble donc majoritairement déterminé par les complétives, et la diversification des types fonctionnels de sous-phrases est plutôt limitée.

Groupe	% Complétives			% Adverbiales			% Relatives		
	N	Moyenne	E.T.	N	Moyenne	E.T.	N	Moyenne	E.T.
1	25	3,12%	6,44%	25	0,00%	0,00%	25	0,00%	0,00%
2	25	8,04%	8,05%	25	2,04%	3,98%	25	1,17%	1,93%
3	25	14,90%	6,05%	25	6,18%	5,04%	25	2,46%	2,95%
4	25	18,34%	5,64%	25	6,23%	3,61%	25	3,05%	3,14%
Natifs	25	18,59%	6,92%	25	8,17%	4,30%	25	7,35%	4,56%
Post-hoc	1,2<3,4,N			1,2<3,4,N			1,2<2,3,4<N		
η_p^2	0,464			/			/		

Tableau 7 Résumé des mesures de diversité : types fonctionnels

Quant aux types formels, une augmentation linéaire peut être remarquée pour les deux types, i.e. les subordonnées finies et non-finies, bien que cette progression soit plus marquée dans le cas de subordonnées à verbe fini, pour lesquelles nous observons des différences statistiquement significatives du niveau 1 à 3 ($p < 0,05$). Contrairement aux types fonctionnels, dont les apprenants utilisent majoritairement les complétives, le nombre de subordonnées à verbe fini et non-fini semble distribué de façon plus égale à chaque niveau. Remarquez toutefois que la différence entre les groupes 3 et 4, bien que non significative, est plus grande pour les subordonnées non-finies que pour les finies. Ceci pourrait indiquer que les propositions non-finies jouent un rôle plus important dans le développement en L2 d'apprenants avancés, même si, dans les deux cas, la différence entre les apprenants du niveau 4 et les natifs ne peut pas être vérifiée par les tests post-hoc.

Groupe	% Finies			% Non-Finies		
	<i>N</i>	<i>Moyenne</i>	<i>E.T.</i>	<i>N</i>	<i>Moyenne</i>	<i>E.T.</i>
1	25	0,36%	1,82%	25	2,99%	6,35%
2	25	5,07%	5,93%	25	6,18%	6,67%
3	25	13,76%	8,28%	25	9,78%	4,43%
4	25	14,83%	6,45%	25	12,79%	4,95%
Natifs	25	18,62%	4,84%	25	15,50%	6,88%
Post-hoc	1<2<3,4<4,N			1,2<2,3<3,4,N		
η_p^2	0,579			/		

Tableau 8 Résumé des mesures de diversité : types formels

Finalement, une analyse des corrélations entre les mesures simples de diversité et les mesures de distribution au niveau général (Tableau 9) révèle surtout une très forte corrélation entre l'indice de diversité syntaxique et le pourcentage de propositions matrices ($r = 0,955$) et subordonnées ($r = 0,929$). Cette corrélation s'explique surtout par le fait que les textes obtenant des scores de diversité syntaxique bas contiennent surtout des unités AS simples, c'est-à-dire, consistant seulement d'une phrase matrice. Plus les locuteurs atteignent des scores de l'indice de diversité syntaxique élevés, plus ils enchaînent différents types de subordonnées dans les unités AS et le pourcentage de phrases matrices baisse. Les corrélations entre les mesures de distribution et les mesures de variation au niveau du SN ou de la proposition sont également statistiquement significatives ($p < 0,05$), mais sont sans doute de nature secondaire. Dans le cas des SN, par exemple, les corrélations significatives mais basses peuvent en partie s'expliquer par la légère augmentation non-significative dans les données d'apprenants. Même si cette augmentation n'était pas assez prononcée pour être vérifiée par le test ANOVA, elle pourrait être parallèle à celle des mesures de distribution.

	IDS	DivSN	DivProp	DivAS
%mat	-0,955*	-0,334*	-0,497*	-0,802*
%coor	0,076	-0,1	-0,033	0,136
%sub	0,929*	0,366*	0,514*	0,774*

Tableau 9 Corrélations entre les mesures de diversité (* = corrélation statistiquement significative ($p < 0,05$))

En résumé, les mesures de diversité indiquent que les développements les plus prononcés se situent au niveau multipropositionnel et que l'emploi de subordonnées complétives finies détermine en large partie cette diversification. De plus, les productions des natifs sont plus diverses que ceux des apprenants sur le plan des SN et des unités AS, et seuls les apprenants plus avancés emploient des structures d'une diversité comparable à ceux des natifs.

4.3 Comparaison

Le tableau 10 présente les résultats des tests de corrélation entre les deux types de complexité syntaxique et indique plusieurs tendances. Premièrement, les corrélations entre la mesure de longueur de proposition (%prop) et le pourcentage de propositions coordonnées (%coor), d'une part, et les autres mesures, d'autre part, sont généralement basses ou non significatives, fait qui s'explique par l'absence de variation systématique pour ces deux variables. Deuxièmement, les corrélations les plus fortes sont observées pour les mesures du même niveau syntaxique. Dans un sens, ce lien paraît logique et inévitable, car pour certaines mesures (par exemple, LonSN et DivSN : $r = 0,838$) l'opérationnalisation des concepts sous-jacents est partiellement la même. Ce lien est également fort pour des mesures touchant le même niveau syntaxique, i.e. l'unité AS, mais dont l'opérationnalisation est toutefois différente (par exemple, LonAS et IDS: $r = 0,834$). Dans ce cas, le lien peut sembler évident mais reste avant tout un constat empirique. D'un côté, si les textes contiennent davantage de structures longues, on y retrouvera également des phrases courtes, auxquelles la tâche narrative se prête facilement, et la diversité augmentera par conséquent. De l'autre côté, le niveau de la proposition révèle un résultat apparemment aberrant, car les mesures de longueur de proposition (LonProp) et de diversité de proposition (DivProp) n'ont qu'un lien faible ($r = 0,279$). De plus, nous avons également observé une légère augmentation statistiquement significative de DivProp, tandis que LonProp ne démontrait pas de progrès. Donc, les deux mesures présentent en effet deux situations différentes dans les textes : si en moyenne la longueur des propositions ne change guère, un léger développement de diversité a néanmoins lieu dans les groupes d'apprenants.

Finalement, parmi toutes les mesures de longueur, nous observons les corrélations les plus fortes sur le plan de la proportion de propositions par unité AS (Prop/AS). Celle-ci semble couvrir quasiment parfaitement le pourcentage de phrases matrices ($r = -0,987$) et subordonnées ($r = 0,956$), ainsi que l'indice de diversité syntaxique ($r = 0,966$).

Diversité/longueur	LonSN	LonProp	LonAS	Prop/AS
<i>DivSN</i>	0,838*	0,140	0,377*	0,337*
<i>DivProp</i>	0,562*	0,279*	0,565*	0,508*
<i>DivAS</i>	0,506*	0,229*	0,822*	0,814*
<i>IDS</i>	0,351*	0,012*	0,834*	0,966*
<i>%mat</i>	-0,301*	-0,022	-0,846*	-0,987*
<i>%coor</i>	-0,085	-0,026	0,085	0,138
<i>%sub</i>	0,328*	0,012	0,813*	0,956*

Tableau 10 Corrélations entre complexité structurelle et diversité (* = corrélation statistiquement significative ($p < 0,05$))

5 Discussion et conclusion

Dans cette section, nous traiterons consécutivement des trois questions de recherche, portant sur le rôle de la complexité structurelle dans l'ALS, sur le rôle de la diversité et sur la mesure de la compétence linguistique.

Pour ce qui est de la complexité structurelle, nous avons observé en premier lieu une complexification au niveau multipropositionnel, contrairement au niveau de la proposition, où aucun développement n'a été observé. Ceci suggère que les apprenants du niveau 1 ont déjà passé le stade prébasique (Klein & Perdue, 1997), marqué par une organisation nominale. En effet, les transcriptions révèlent que ces textes correspondent plutôt au stade post-initial de Bartning & Schlyter (2004) : nous y retrouvons surtout des phrases simples, avec un verbe qui n'est pas nécessairement accordé au sujet, mais aussi déjà quelques structures négatives. Cependant, les structures multipropositionnelles des apprenants n'atteignent pas le niveau de complexité observé chez les natifs. La complexification progressive au niveau multipropositionnel sera donc surtout provoquée par l'enchaînement de propositions plutôt que par une réorganisation au sein des propositions. Finalement, l'absence de différences au niveau du syntagme nominal semble également indiquer que la structure de base du SN est déjà établie dans le groupe 1, même si les apprenants n'atteignent généralement pas le niveau des natifs. La plus grande complexité des SN des natifs s'explique en premier lieu par un plus grand emploi de relatives, indiqué par les mesures de distribution, plutôt que par l'influence d'adjectifs et de syntagmes prépositionnels (par exemple, « le chien du petit garçon ») qui apparaissent fréquemment dans les données des apprenants. Parallèlement à l'oral, nos résultats sur la complexité des SN avancent donc la complexité à ce niveau syntaxique comme un domaine plus avancé dans le développement. Or, si cette complexification à l'écrit concerne avant tout la nominalisation, il y a surtout un maintien du caractère propositionnel dans les SN à l'oral, du moins dans nos textes narratifs.

En ce qui concerne la diversité, cette dimension semble se développer en tandem avec la complexité structurelle, de sorte qu'il y a non seulement une complexification générale, mais aussi une plus grande diversification, surtout au niveau multipropositionnel. Toutefois, les indices simples ont révélé certains aspects où le développement de la diversité diverge de celui de la complexité structurelle. Premièrement, au niveau des SN, la taille d'effet plus élevée pour la diversité indique que les natifs s'opposent plus aux apprenants sur le plan de la variation de longueur que sur le plan de la longueur moyenne. Deuxièmement, contrairement à la complexité structurelle au sein des propositions, la diversité des propositions augmente légèrement à travers les différents niveaux. Cette observation révèle deux choses. D'une part, elle indique que, même si les mesures de diversité et de complexité structurelle au même niveau syntaxique sont en corrélation élevée, les deux dimensions peuvent quand-même indiquer des développements séparés. D'autre part, elle remet la stagnation de la longueur des propositions en cause. Les mesures de distribution indiquent surtout une augmentation des subordinées complétives à travers les niveaux de compétence. Ces constructions prennent souvent la forme de l'exemple (12), où la matrice est considérablement plus courte que la complétive. Par conséquent, la mesure de longueur de proposition (LonProp) analysera ces phrases comme égales à ceux de (13), tandis la mesure de diversité (DivProp) parviendra à les différencier. Donc, dans des cas de diversité élevée, les moyennes de la longueur cacheraient dans un sens le développement au niveau propositionnel, même si ce développement peut être repéré à d'autres niveaux, tels que l'unité AS (LonAS) dans les exemples (12-13).

- (12) LonProp : 3,5 DivProp : 2,12 LonAS :7
 [mat] il pense
 [sub. comp. fin.] que c'est des branches .
- (13) LonProp : 3,5 DivProp : 0,71 LonAS :3,5
 [mat] et il cherchE.
 [mat] il tombe en arbre.

Finalement, les mesures de distribution ont indiqué que la diversification des unités AS était, en premier lieu, causée par une augmentation de complétives (finies et non-finies) et, en second lieu, par l'emploi de subordinées adverbiales, tandis que les relatives sont une catégorie peu exploitée en comparaison avec leur emploi dans le discours des natifs. De plus, le rôle de la pseudo-coordination dans nos données

semble négligeable pour le développement de la complexité syntaxique. Le traitement de propositions coordonnées et l'exclusion d'autres types (plus prototypiques) de coordination de nos analyses sont sans doute deux des différences importantes par rapport aux études précédentes traitant de la complexité à l'écrit. Peut-être un examen détaillé qui tient compte d'éléments prosodiques et de différents marqueurs de coordination pourrait attribuer une plus grande importance au développement de la coordination.

Quant à notre dernière question de recherche, si nous regardons uniquement les tailles d'effet (qui tiennent compte non seulement des différences entre les groupes d'apprenants mais aussi des natifs) et des différences statistiquement significatives, les mesures de diversité et complexité structurelle au niveau de l'AS, mais surtout aussi la proportion de phrases matrices et subordonnées semblent les plus efficaces en tant que mesures de compétence linguistique. Ces mesures parviennent à indiquer une progression linéaire dans trois des quatre groupes, où les tailles d'effet se situent entre 0,5 et 0,65. Même si la longueur d'unité multipropositionnelle et le nombre de proposition par unité multipropositionnelle ont été suremployés en tant que mesures de complexité linguistique (Norris & Ortega 2009), ces mesures demeurent donc intéressantes comme indices de compétence linguistique. Or, nos résultats soulignent aussi le développement hétérogène des différentes dimensions syntaxiques, surtout illustré par le retard dans la complexification des SN, de sorte que les mesures au niveau multipropositionnel ne parviennent pas à représenter l'entièreté du développement syntaxique. En outre, l'image fournie par nos données n'est pas non plus complète. D'abord, les analyses de corrélation et les tailles d'effet sont calculées à travers tous les groupes, de sorte que le degré des corrélations et des différences observées pourrait varier d'un niveau à l'autre et, par exemple, être plus fort au début du développement qu'à la fin. De plus, les participants de cette étude ne couvrent pas tout le trajet d'apprentissage et, notamment, nous n'avons pas pu vérifier le développement du stade prébasique au stade basique, pour lequel la mesure de longueur de proposition pourrait s'avérer plus intéressante. Nous n'avons non plus pu intégrer des apprenants d'un niveau très avancé ou quasi-bilingue et il sera intéressant de vérifier si ces apprenants comblent les écarts entre apprenants et natifs observés dans cette étude (par exemple, pour l'emploi de relatives, ou la longueur des SN).

Il est également important de considérer les effets potentiels de la tâche de production. Ainsi, le lien fort entre les indices simples de diversité et les mesures de distribution pourrait être dû au genre narratif et à la modalité orale des données. Dans nos données, par exemple, la corrélation entre l'indice de diversité syntaxique et la proportion de phrases matrices s'explique en partie par le fait que la tâche narrative se prête à un emploi élevé de phrases simples. Dans des textes académiques écrits, par contre, le rapport entre les phrases simples, l'indice de diversité syntaxique et la compétence linguistique pourrait être perturbé par un développement de phrases complexes à une réorganisation nominale, de sorte que la présence de phrases simples en elle-même ne suffit plus comme indice de compétence (Norris & Ortega, 2009 ; Byrnes, 2009). Notons d'ailleurs que les mesures de distribution ne sont intéressantes en tant qu'indices de diversité que si elles sont interprétées simultanément. Autrement dit, même si nous avons observé des corrélations élevées entre ces mesures et l'indice de diversité syntaxique, le nombre de phrases matrices ne peut que nous informer indirectement sur la diversité des textes, si elle n'est pas accompagnée par les autres mesures de distribution.

En résumé, les résultats de cette étude corroborent donc en large partie ceux d'études précédentes (par exemple, Bartning & Kirchmeyer, 2003; Welcomme, 2013), en ce que les apprenants emploient progressivement plus de structures subordonnées et en ce qu'ils utilisent des structures formellement et fonctionnellement différentes aux niveaux plus avancés. Les différences observées entre les apprenants avancés et les natifs indiquent toutefois que les participants de cette étude n'ont que rarement atteint les stades avancés décrits par Bartning & Kirchmeyer (2003). L'examen quantitatif a également démontré que, si nous retrouvons les mêmes structures fonctionnels et formels dans la plupart des productions des apprenants, il y a néanmoins des différences considérables quant à la distribution de ces structures. Par exemple, même si les apprenants utilisent des propositions relatives dès le niveau 2, celles-ci demeurent sous-employées en comparaison avec les natifs. Notons, de plus, que nous n'avons pas pu corroborer l'hypothèse de Lambert & Kormos (2014), selon laquelle des experts, ou locuteurs très avancés, emploieraient des structures plus efficaces et simples que des "novices". Bien que nos résultats

n'infirmant pas cette hypothèse, vu que nos données sont limitées à une seule tâche et qu'il ne s'agit pas d'experts natifs, mais d'enfants âgés de 12-13 ans, ils réaffirment toutefois que dans nos données une plus grande compétence linguistique va de pair avec l'emploi de structures plus longues et plus diverses à chaque niveau étudié.

Quant à nos hypothèses, nous n'avons que partiellement pu confirmer notre première hypothèse, selon laquelle le développement dans les groupes moins avancés aurait surtout lieu au niveau des SN et des propositions, vu que le groupe d'apprenants au niveau 1 utilisait déjà des SN et propositions de longueur comparable à ceux des autres groupes. En revanche, la complexification structurelle des unités multipropositionnelles s'est bien avéré comme un domaine de développement important à travers tous les niveaux. Notre deuxième hypothèse, qui posait que la diversité syntaxique suivrait le développement de la complexité structurelle, a été confirmée en ce qui concerne l'apparition et l'emploi plutôt tardifs des différents types de propositions subordonnées, surtout dans le cas des propositions adverbiales et relatives. A l'exception du niveau des propositions, les mesures de longueur et de diversité indiquaient toutefois des développements parallèles de ces deux dimensions. Conformément à notre troisième hypothèse, la longueur d'unité AS était parmi les mesures qui démontraient les plus grandes différences entre les quatre niveaux, même si d'autres mesures touchant le niveau multipropositionnel, telles que l'IDS, et la proportion de phrases matrices, distinguaient plus clairement entre les niveaux 1 et 2.

Certaines limitations de cette étude devraient être prises en compte. Premièrement, les augmentations observées au cours du développement FLS sont sans doute également déterminées par un développement cognitif, car les âges des apprenants varient de 12 à 18 ans. Même pour les natifs, âgés de 12-13 ans, bien qu'ils atteignent souvent des scores plus élevés que les apprenants, on pourrait également s'attendre à une complexification syntaxique dans leur propre langue maternelle à mesure qu'ils deviennent plus âgés (cf. Berman & Verhoeven, 2002).

Deuxièmement, si nous avons élaboré la gamme de mesures typiquement employée pour l'étude de la complexité syntaxique, notre approche n'évite pas entièrement le réductionnisme. En ce qui concerne la diversité, par exemple, l'indice de diversité syntaxique ne prend qu'en compte les jonctions interpropositionnelles, tandis que les autres mesures n'impliquent pas directement la gamme de structures censées complexes et prennent plutôt la longueur des structures comme approximation de leur complexité. D'autres approches, telles que celle de Scarborough (1990) qui attribue des scores de complexité sur la base de l'emploi de structures spécifiques, ont tenté d'offrir une mesure plus détaillée de la complexité syntaxique. Or, la justification pour l'attribution d'une certaine complexité à une structure peut s'avérer subjective, ou basée sur d'autres notions, telles que la difficulté ou la fréquence, qui ne sont qu'indirectement liées à la complexité absolue. De plus, un autre défi consistera alors de homogénéiser les valeurs de complexité syntaxique à travers des phénomènes de différents ordres. Par exemple, peut-on additionner le score attribué à la complexité d'une proposition enchâssée à verbe non-fini à celui de la complexité par inversion ?

D'autres questions sont encore à résoudre aussi. Ainsi, notre approche a délimité a priori le nombre de catégories mesurées par les indices de diversité, tandis que d'autres catégories pourraient également être considérées. Par exemple, l'indice de diversité syntaxique pourrait représenter davantage de spécifications (par exemple, le type de connecteur) ou être appliqué à d'autres types de diversité (par exemple, l'ordre des mots) si ces aspects sont supposés être pertinents.

En somme, dans cette étude, nous avons tenté de faire un premier pas vers une analyse systématique de la complexité syntaxique, en tenant compte de la diversité et de la complexité structurelle. Nous avons également complété les résultats des recherches antérieures en appliquant l'approche quantitative au français et en fournissant des données sur la complexité syntaxique à l'oral. Nos résultats sont généralement conformes à ceux d'études précédentes, en ce qu'ils soulignent l'importance de la complexification des unités multipropositionnelles et de la jonction interpropositionnelle dans le développement linguistique et en ce qu'ils démontrent que certains domaines, tels que l'organisation au sein du syntagme nominal, demeurent sous-développés chez les apprenants. Si cette étude a révélé que la diversité joue également un rôle important dans le développement, elle ouvre par ailleurs de nombreuses

pistes à explorer. Ainsi, il sera nécessaire de confronter les résultats de cette étude à un examen du développement aux stades initiaux (de prébasique à basique), mais aussi aux stades (très) avancés. L'influence de certaines variables contextuelles, telles que la tâche et la modalité, devrait également être étudiée afin de vérifier dans quelle mesure les tendances observées sont spécifiques aux caractéristiques de notre étude. Finalement, il sera intéressant d'intégrer la diversité syntaxique à l'étude longitudinale de l'apprentissage de langues secondes.

Références bibliographiques

- Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence TESOL Quarterly, 26(2), 390-395.
- Bartning, I., & Kirchmeyer, N. (2003). Le développement de la compétence textuelle à travers les stades acquisitionnels en français L2. *Acquisition et interaction en langue étrangère*, 19, 9-39.
- Bartning, I., & Schlyter, S. (2004). Itinéraires acquisitionnels et stades de développement en français L2. *Journal of French language studies*, 14(3), 281-299.
- Beaman, K. (1984). Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse. Dans D. Tannen (Ed.), *Coherence in spoken and written discourse* (pp. 45-80). Norwood, N.J.: Ablex.
- Berman, R. A., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language & Literacy*, 5(1), 1-43.
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5-35.
- Bulté, B. (2013). *The development of complexity in second language acquisition: a dynamic systems approach*. (thèse de doctorat), Vrije Universiteit Brussel, Brussel.
- Byrnes, H. (2009). Emergent L2 German writing ability in a curricular context: A longitudinal study of grammatical metaphor. *Linguistics and Education*, 20, 50-66.
- Cosme, C. (2007). *Clause linking across languages. A corpus-based study of coordination and subordination in English, French and Dutch*. (thèse de doctorat), Université Catholique de Louvain, Louvain-la-Neuve.
- Dahl, Ö. (2009). Testing the assumption of complexity Invariance: the case of Eflalian and Swedish. Dans G. Sampson, D. Gil & P. Trudgill (Eds.), *Language Complexity as an Evolving Variable* (pp. 50-63). Oxford: Oxford University Press.
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354-375.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473.
- Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins.
- Hunt, K. (1965). *Grammatical structures written at three grade levels*. Champaign, Ill: National Council of Teachers of English.
- IBM. (2015). *IBM SPSS Statistics (Version 23.0)*.
- Jarvis, S., & Daller, M. (Eds.). (2013). *Vocabulary knowledge: Human ratings and automated measures*. Amsterdam: John Benjamins Publishing.
- Kerr-Barnes, B. (1998). The acquisition of connectors in French L2 narrative discourse. *Journal of French language studies*, 8(2), 189-208.
- Klein, W., & Perdue, C. (1997). The Basic Variety (or: Couldn't natural languages be much simpler?). *Second Language Research*, 13(4), 301-347.
- Kirchmeyer, N. (2002). *Étude de la compétence textuelle des lectes d'apprenants avancés. Aspects structurels, fonctionnels et informationnels*. Cahiers de la Recherche 20. (Thèse de doctorat), Université de Stockholm, Stockholm.

- Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning*, 62(2), 439-472.
- Kortmann, B., & Szmrecsanyi, B. (Eds.). (2012). *Linguistic complexity, second language acquisition, indigenization, contact*. Berlin, Boston: De Gruyter.
- Lambert, C., & Kormos, J. (2014). Complexity, accuracy, and fluency in task-based L2 research: Toward more developmentally based measures of second language acquisition. *Applied Linguistics*, 2014, 1-9.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Malvern, D., Richards, B. J., Chipere, N., & Duran, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan.
- Mayer, M. (1969). *Frog, Where are you?* New York: Dial Press.
- Miestamo, M. (2008). Grammatical complexity in a cross-linguistic perspective. Dans M. Miestamo, K. Sinnemäki & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 23-41). Amsterdam: John Benjamins.
- Miestamo, M. (2009). Implicational hierarchies and grammatical complexity. Dans G. Sampson, D. Gil & P. Trudgill (Eds.), *Language complexity as an evolving variable* (pp. 80-97). Oxford: Oxford University Press.
- Miestamo, M., Sinnemäki, K., & Karlsson, F. (Eds.). (2008). *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555-578.
- Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. Dans B. Kortmann & B. Szmrecsanyi (Eds.), *Linguistic complexity: Second language acquisition, indigenization, contact* (pp. 127-155). Berlin, Boston: De Gruyter.
- Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30(4), 590-601.
- Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, 31(1), 117-134.
- Paprocka-Piotrowska, U., Martinot, C., & Gerolimich, S. (Eds.). (2013). *La complexité en langue et son acquisition*. Lublin: Towarzystwo Naukowe Kul Katolicki Uniwersytet Jana Pawla II.
- Pierrard, M., & Housen, A. (2013). Complexité linguistique et efficacité d'un enseignement grammatical explicite. Dans U. Paprocka-Piotrowska, C. Martinot & S. Gerolimich (Eds.), *La complexité en langue et son acquisition* (pp. 285-304). Lublin: Towarzystwo Naukowe Kul Katolicki Uniwersytet Jana Pawla II.
- Scarborough, H. S. (1990). Index of productive syntax. *Applied Psycholinguistics*, 11, 1-22.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, P. (2009). Performance by native and non-native speakers on language-learning tasks. Dans In B. J. Richards, M. Daller, D. Malvern, P. Meara, J. Milton & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application* (pp. 107-124). London: Palgrave Macmillan.
- Szmrecsanyi, B. (2004). On operationalizing syntactic complexity. Présenté à la conférence JADT 2004, Louvain la Neuve.
- Szmrecsanyi, B., & Kortmann, B. (2012). Introduction: Linguistic complexity. Dans B. Kortmann & B. Szmrecsanyi (Eds.), *Linguistic complexity: Second language acquisition, indigenization, contact*. Boston: De Gruyter.
- Verspoor, M., Schmid, M. S., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, 21, 239-263.
- Welcomme, A. (2013). Connecteurs et complexité syntaxique dans les récits d'apprenants néerlandophones de FLE : premiers résultats. Dans U. Paprocka-Piotrowska, C. Martinot & S. Gerolimich (Eds.), *La complexité en langue et son acquisition* (pp. 261-284). Lublin: Towarzystwo Naukowe Kul Katolicki Uniwersytet Jana Pawla II.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu: National Foreign Language Resource Center.

¹ Selon Bulté & Housen (2012), ces mesures de subordination sont des mesures hybrides car leur statut en tant qu'indices de complexité absolue n'est pas entièrement clair. Plus spécifiquement, elles seraient en partie choisies, non seulement pour des raisons structurelles, mais aussi parce qu'elles seraient d'une difficulté cognitive plus élevée. Nous adhérons néanmoins au statut absolu de ces mesures, vu que la subordination (et donc la phrase complexe) occupe une position centrale en linguistique et que le lien avec la difficulté cognitive n'est dans ce sens que secondaire.