T.C. KOCAELİ ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

YABANCI DİLLER EĞİTİMİ ANABİLİM DALI

İNGİLİZ DİLİ EĞİTİMİ BİLİM DALI

ANALYZING THE DISCURSIVE FEATURES AND SYNTACTIC COMPLEXITY OF L2 ENGLISH LEARNERS ACROSS THE CEFR LEVELS: A CORPUS BASED RESEARCH

(YÜKSEK LİSANS TEZİ)

Ayşe Gizem ÇİFTÇİ

KOCAELİ 2021

T.C. KOCAELİ ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

YABANCI DİLLER EĞİTİMİ ANABİLİM DALI

İNGİLİZ DİLİ EĞİTİMİ BİLİM DALI

ANALYZING THE DISCURSIVE FEATURES AND SYNTACTIC COMPLEXITY OF L2 ENGLISH LEARNERS ACROSS THE CEFR LEVELS: A CORPUS BASED RESEARCH

(YÜKSEK LİSANS TEZİ)

Ayşe Gizem ÇİFTÇİ

Dr. Öğretim Üyesi Mehmet ALTAY

Tezin Kabul Edildiği Enstitü Yönetim Kurulu Karar ve No:
13.01.2021/02

KOCAELİ 2021

Anneme…

ACKNOWLEDGEMENTS

Throughout the writing of this thesis, I have received a great deal of support and assistance from several people, without whom I would not have been able to complete this research, and without whom I would not have made it through my MA degree!

I acknowledge with a deep sense of reverence, my gratitude to my thesis advisor, Dr. Mehmet Altay, who constantly showed his unfailing support and assistance in the process of formulating the research topic and methodology of my thesis. I would also like to thank him for always being there when I cannot wrap my mind around statistics.

I would like to show my appreciation to my thesis committee members, Dr. İhsan Ünaldı and Dr. Doğan Yüksel for their insightful feedback on my research. I wish to extend my special thanks to Dr. Yüksel who made us feel that he was always there with his guiding support. He is a great example to those who wish to become an academic, and I appreciate him for setting an outstanding example of a teacher.

I am more than grateful to my biggest advisor in life, Utku Tönel, for always being there when I need him for anything one can think of. Seda! Thank you for being the best roommate who supports my work no matter what. I would also like to thank Fatma Ege for being the best study-buddy I can think of, helping me in any way she can in the process of my thesis. She also picked me up when I fell, and I appreciate her for it. My statistics-nerd friend Merve Bayram has also helped me in the tough times when I thought I would not be able to finish this thesis, and I am grateful for her insightful conversations and support. Hürrem and Ceren! I would like you thank you for always believing in me and say sorry for the times we could not see each other because I was busy writing a thesis. And to all my other friends whom I spent time with, without you I would be able to finish writing this thesis three months ago.

Ayrıca canım annem! Bana sarf ettiğin bunca senelik emeği bu paragraf içine sığdırmam çok da mümkün değil. Ama beni her daim desteklediğin ve bu yaşıma kadar getirdiğin için sana minnettarım. Bu tezi yazmamdaki en büyük motivasyon kaynağım emeklerini boşa çıkarmamak ve Oğulcan'a iyi bir örnek olmaktı. İyi ki varsın.

Lastly, this thesis is ostensibly supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) within the scope of 2210-A National Scholarship Program for Master's Students. I would like to thank them for their support.

TABLE OF CONTENTS

ABSTRACT

The Common European Framework of Reference (CEFR) was created out of a necessity of a new education policy that arose due to Europe becoming a multilingual and multicultural continent and citizens from various backgrounds starting to interact with each other. Though being a crucial framework in language education, the CEFR research has put linguistic and discursive aspects of learner language across levels under scrutiny rather less compared to other areas of second language literature. Motivated by the research lacuna on these aspects, the current research primarily aims to investigate syntactic complexity and discursive features in learner speech within and across the CEFR levels of proficiency. To achieve this aim, 314 texts of learner speech from an open-access corpus at different CEFR levels were analyzed in terms of syntactic complexity and discursive features. Qualitative content analysis and qualitative analysis with an automated tool were employed in the research to analyze the data. The analysis of the learner speech revealed that (i) most syntactic complexity indices were non-normally distributed within the CEFR levels, (ii) syntactic complexity of the texts differ significantly across the CEFR levels, (iii) not a normal distribution of the discursive features was identified in most CEFR levels, and (iv) connectors and discourse markers were identified to be the features that were used the most in the whole corpus. Several pedagogical implications were drawn based on the findings of the research.

Keywords: CEFR, syntactic complexity, discursive features, Coh-Metrix, NVivo

ÖZET

Avrupa'nın çok dilli ve çok kültürlü bir kıta haline gelmesi ve farklı sosyal çevrelerden gelen vatandaşların birbirleriyle etkileşime girmesinden ötürü yeni bir eğitim politikasının gerekliliği ortaya çıkmıştır ve Avrupa Dilleri Ortak Çerçeve programı da bu sebeple oluşturulmuştur. Dil eğitiminde çok önemli bir çerçeve olmasına rağmen, CEFR araştırmaları, öğrenen dilinin dilbilimsel ve söylemsel yönlerini diğer yabancı dil alanlarına kıyasla daha az inceleme altına almıştır. Bu yönlerle ilgili araştırma boşluğundan yola çıkarak mevcut araştırma, CEFR yeterlilik seviyeleri içinde ve farklı CEFR seviyeleri arasında öğrenen konuşmasındaki sözcüksel karmaşıklığı ve söylemsel özellikleri araştırmayı amaçlamaktadır. Bu amaca ulaşmak için açık erişimli bir derlemden alınan farklı seviyelerdeki 314 öğrenci konuşma metni sözcüksel karmaşıklık ve söylemsel özellikler açısından analiz edilmiştir. Araştırmada verileri analiz etmek için nitel içerik analizi ve bilgisayar destekli bir araçla nitel analiz yöntemi kullanılmıştır. Öğrenci konuşmasının analizi, (i) sözcüksel karmaşıklık indekslerinin çoğunun CEFR seviyeleri içinde normal olarak dağılmadığını, (ii) metinlerin sözcüksel karmaşıklığının CEFR seviyeleri arasında önemli ölçüde farklılık gösterdiğini, (iii) birçok CEFR seviyelerinde söylemsel özelliklerin normal bir dağılımı olmadığını ve (iv) bağlaçlar ve söylem belirteçlerinin tüm derlemde en çok kullanılan özellikler olduğunu ortaya koymuştur. Araştırmanın bulgularına dayanarak çeşitli pedagojik çıkarımlar yapılmıştır.

Anahtar kelimeler: CEFR, sözcüksel karmaşıklık, söylemsel özellikler, Coh-Metrix, NVivo

ABBREVIATIONS

CAF: Complexity, Accuracy, and Fluency

CAQDAS: Computer-Assisted Qualitative Data Analysis Software

CEFR: Common European Framework of Reference for languages

CLEC: CEFR-Labeled English Corpus

CLIL: Content and Language Integrated Learning

CS: Code-switching

DC: Dependent Clause

DM: Discourse Marker

EAP: English for Academic Purposes

ELF: English as a Lingua Franca

ELP: European Language Portfolio

EP: English Profile

L1: First Language

L2: Second Language

LLP: The Linguistics Language Program

L2SCA: The L2 Syntactic Complexity Analyzer

MLC: Mean Length of Clause

SC: Syntactic Complexity

SLA: Second Language Acquisition

SPSS: Statistical Package for the Social Sciences

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Since language, discourse, and communication are considered to be valuable constructs with which one may catch a glimpse of several processes that learners go through, characteristics of language and discourse in learner language has a significant value in language pedagogy. Hence, analyzing language and discourse related constructs can be of help in language education. Additionally, the CEFR is a highly significant document that is utilized in language teaching not only in Europe but also in other parts of the world. However, although the CEFR is one of the most influential frameworks in language education, little research has addressed the linguistic and discursive aspects of learner language across the CEFR levels. Correspondingly, the current thesis was designed to analyze linguistic and discoursal features that language learners employ at different proficiency levels. To this wake, the present research attempted to explore (i) the syntactic complexity of learners' productions at specific the CEFR levels, (ii) the syntactic complexity of learners' productions at across the CEFR levels, (iii) how learners at specific levels of the CEFR employ discursive features, and (iv) how learners across the CEFR levels employ discursive features.

Three hundred and fourteen texts from different levels of the CEFR constitute the sampling of the current research. In order to analyze the syntactic complexity of the learners, an automated analysis tool, Coh-Metrix, was utilized as the data analysis tool. To be able to capture the discursive features of the learners, qualitative content analysis was utilized as data analysis. In that vein, a computerized qualitative analysis tool, NVivo, was employed. The findings revealed that syntactic complexity of the learners differ significantly across the levels. Additionally, several discursive features are utilized across the levels, and the most frequent features were identified to be connectors and discourse markers.

CHAPTER 1

1. INTRODUCTION

This present section presents the background to the research. It then explains the statement of the problem, the purpose of the research as well as the significance of the research.

1.1. BACKGROUND TO THE RESEARCH

Those involved in education circles have been striving for defining the things language learners should learn and how they should be described so that everyone can understand and benefit from (Figueras, 2012: p. 477). With the European Union, Europe has become a multilingual and multicultural continent, which led policymakers to change the policies in education. To this end, The Common European Framework of Reference (hereafter, CEFR) was written by the Council of Europe, indicating the probable expansion of the language and culture policy document (Barni & Salvati, 2019: p. 418). The CEFR is a broadly referred and used document in language proficiency and language syllabi (Figueras, 2012: p. 477), and it is arguably one of the most influential frameworks in language education nowadays.

As it is a widely referred document, it has been questioned by many scholars (e.g., Valax, 2011: p. 1; Weir, 2005: p. 281) in terms of its applicability, validity, suitability, etc. Based on the concern with regard to the validity of the descriptive scales of language proficiency in the document, Hulstijn, Alderson, and Schoonen (2010: p.17) address five research questions, one of which is immediately relevant to the concern of the present research:

> What are the linguistic profiles at every CEFR level for the two productive language skills (speaking and writing) and what are the linguistic features typical of the two receptive skills (listening and reading) at every CEFR level?
>
> (Hulstijn, Alderson & Schoonen, 2010: p. 17)

Furthermore, considering the fact that capturing the characteristics of language and discourse in learner language has very significant practical and theoretical value in education, as language, discourse, and communication have been considered as a motherlode providing potent insights into various processes that learners go through such as cognitive, affective, motivational, and social processes as well as other learning-related phenomena (Dowell, Graesser, & Cai, 2016: p. 72). Syntactic complexity is one aspect of language that has received a considerable amount of scholarly attention in SLA research; however, there is not a consensus on the definition to operationalize the construct. Norris and Ortega (2009: p. 556) state that one agreement on the construct is that it is a rather complex concept that encompasses several levels and dimensions. Moreover, complexity as a term has been repeatedly mentioned in the CEFR; however, it is not defined thoroughly in the document. Other than the problems in the definition of the concept, it could be stated that it has been employed in L2 research from different perspectives such as task modalities, different groups of learners, and proficiency levels.

As discourse also has a significant value in education, analyzing discursive features and how they are used by learners of L2 is of significance. Cohesion, coherence, discourse markers, and code-switching are among the features which have received scholarly attention in the L2 research. From the CEFR point of view, it would be fair to state that the document does not handle discursive features separately; instead, it deals with the concept discourse competence referring to learners' ability "to arrange sentences in an order in order to produce coherent stretches of language" (Council of Europe, 2001: p. 123). The document assumes that as the proficiency level of the learners' increases, they are expected to develop good discourse competence, organizing their language in a cohesive, coherent, and effective manner (Council of Europe, 2001: p. 123).

Keeping the abovementioned research question (Hulstijn, Alderson & Schoonen, 2010: p. 17) in mind, the present research aims to examine how a set of indices of syntactic complexity as well as a set of discursive features in spoken data of learner English differ between assessed CEFR levels.

1.2. STATEMENT OF THE PROBLEM

In terms of the linguistic aspects, syntactic complexity has been recognized as an important construct in L2 teaching and research since the growth of syntactic complexity in learner language is a fundamental part of their development in the target language (Ortega, 2003: p. 492). Several measures of syntactic complexity (e.g., length of the production unit, amount of subordination, degree of sophistication of specific syntactic structures) have been employed and explored in the characterization of the construct in L2 research with the aim of finding valid and reliable syntactic complexity indices to determine and define learners' proficiency level (Wolfe-Quintero, Inagaki, & Kim 1998: p. 119). However, the manual analysis of learner language is quite labor-intense, hence, most previous studies only investigated one or two indices of syntactic complexity in relatively small data. Furthermore, although it could be relatively easier for learners to master the vocabulary and grammar of a target language, they may still have issues when producing acceptable compositions because of the problems resulting from the use of discourse. Additionally, the CEFR includes some references to syntactic complexity and discourse; however, Khushik and Huhta, (2019: p. 3) assert that the references are mostly unsystematic and ambiguous, and not linkable with specific levels.

To this wake, an analysis of linguistic and discursive features is of significance. However, although the CEFR is one of the most influential frameworks in language education, little research has addressed the linguistic and discursive aspects of learner language across the CEFR levels.

1.3. PURPOSE OF THE RESEARCH

Considering the lack of studies on linguistic and discursive features in terms of learner language development, more studies are needed to better understand how these features are distributed across proficiency levels, specifically across the CEFR levels in

this research. To address the research lacuna, the present research aims to investigate syntactic complexity indices (i.e., left embeddedness, modifiers per noun phrase, sentence syntax similarity, and minimal edit distance) with the help of an automated tool (i.e., Coh-Metrix) and discursive features (i.e., backward reference, code-switching, connectors, discourse markers, high information load, reference to context, weak coherence) that are employed in learner language across the CEFR levels. In addition, how these features were distributed within the CEFR levels was also the concern of this research to better understand learner language at specific levels. To achieve this purpose, learner data were selected from a learner corpus and analyzed with a robust procedure (see Section 3.7 for further explanations of the procedure).

1.4. SIGNIFICANCE OF THE RESEARCH

In view of the abovementioned brief background, the significance of the present research can be explained in many respects. To begin with, considering the lack of research in learner language across the CEFR levels, the research sheds light on how syntactic complexity and discursive features are distributed across the CEFR levels. In addition, which features are used by learners at specific levels of proficiency is also of great significance to better understand learner language. In that vein, the present research also provides insights into how these features are distributed within specific CEFR levels in learner speech. Furthermore, the research also contributes to the design of potential curricula for language learners including syntactic complexity and discursive features as such curricula would bring learners closer to the levels as defined on the CEFR. Finally, in terms of syntactic complexity, far too few studies on learner speech have employed an automated tool to investigate the indices across the CEFR levels. Additionally, yet again little attention has been paid to the discursive features other than connectors and discourse markers across the CEFR levels of proficiency. Hence, the current research yields a broader picture of learner speech across different levels of proficiency with the help of Coh-Metrix to investigate syntactic complexity and NVivo to analyse discursive features.

1.5. RESEARCH QUESTIONS

Taken all together, the purpose of the research is twofold; first to examine the learner language in terms of syntactic complexity. Second, to investigate the learner language in terms of discursive features. More specifically, the current research addressed the following research questions:

1. Is there consistency within the same CEFR level distribution in a spoken learner corpus in terms of syntactic complexity?
2. Is there a significant difference among CEFR level distributions in a spoken learner corpus in terms of syntactic complexity?
3. Is there consistency within the same CEFR level distribution in a spoken learner corpus in terms of discursive features?
4. Is there consistency among CEFR level distributions in a spoken learner corpus in terms of discursive features?

1.6. CONCLUSION

In short, this research examines the growth trajectories of spoken syntactic complexity and discursive features in learner language across different levels of the CEFR. The research builds on these features in two ways. To begin with, an analysis of a corpus of 314 samples produced by learners at different proficiency levels were analyzed rather than using experimental data, which is common in previous research on, especially syntactic complexity. Polat et al. (2019) include that nonexperimental data are of great significance in understanding L2 development considering research on ecological validity and research-driven pedagogy. Second, the current research is exploratory since it is not based on specific hypotheses that assume how syntactic complexity and discursive features might characterize specific CEFR levels. Instead, a range of indices taken from Coh-Metrix and a set of discursive features taken from the corpus used in this research have been investigated. For a more in-depth explanation of the procedures, please see the section 3.7.

CHAPTER 2

2. THEORETICAL FRAMEWORK AND LITERATURE REVIEW

2.1. INTRODUCTION

This chapter aims to cast light on the theoretical constructs and relevant literature of the CEFR, syntactic complexity, and discursive features in diverse contexts. First, it presents the CEFR and historical background of it along with the introduction of the Common Reference levels. Next, an overview of syntactic complexity and relevant empirical studies are demonstrated. Also, the relationship between syntactic complexity and the CEFR, and the bulk of current research on these concepts are reviewed. Finally, discursive features and related empirical studies are presented based on the aims set as well as the relationship between discursive features and the CEFR.

2.2. THE COMMON EUROPEAN FRAMEWORK OF REFERENCE

Educators and policymakers have been attempting to define what language learners should learn and how it should be described in a fashion that everyone involved can understand and benefit from (Figueras, 2012: p. 477). People involved in language studies in the twentieth century were actively working to describe the language itself, how it is learned, what processes learners go through, how it should be taught, and how it should be assessed. With the emerge of the European Union, Europe has become a multilingual and multicultural continent, and the development of technology has helped to accelerate this process. What was once thought to be far was not considered that far any more thanks to technology, which increased the interaction between the citizens from different backgrounds accordingly. All these required changes in every aspect of life, consequently, the Council of Europe particularized the necessity of a new policy on education. However, it must also be noted that the authors of the document clarified from the very beginning that the CEFR is not aspired to establish a uniform pan-

European system. Instead, they strongly state that the chief purpose of the document is to encourage reflection and discussion as well as presenting ways to describe the diversity.

In 1991, an intergovernmental symposium in Switzerland requested the Council of Europe to undertake the construction of "a Common European Framework of Reference for language learning at all levels" and a European Language Portfolio (ELP) (Council of Europe, 1992: p. 37). CEFR is a groundbreaking language policy document developed by the Council of Europe in 1992. The aim of developing the CEFR at all levels was to support and encourage collaboration among educational institutions in different countries, to provide a foundation for shared acknowledgment of language qualifications, and to assist parties involved in the learning process to position and harmonize their efforts (Trim, 2007: p. 38). The Council of Europe drew up CEFR as an open-access document in 1996, which hinted the probable expansion of the language and culture policy document (Barni & Salvati, 2019: p. 418). From 1998 to 2000, several pilot projects have further explored possible forms the document might take in different educational contexts (Little, 2012: p. 1) and revised it based on the wide-ranging feedback from users and discussions. Finally, CEFR was officially published in English and French in 2001 to coincide with the opening of the European Year of Languages.

One of the main aims of the document is aligning the language learning, teaching, assessment and testing so that learning outcomes across language, countries, and contexts are accessible for comparison. Put differently, the CEFR intends to be utilized for learners' needs analysis, to stipulate language learning outcomes, to assist in developing L2 learning materials and assessment schemes for language learning outcomes (Little, 2006: p. 167). The document is quite comprehensive and descriptive in nature referring to all languages, encouraging language practitioners to utilize the document at their discretion. North (2007: p. 656) calls the document a 'concertina-like reference tool' stating that it is not something to be applied in a prescriptive manner, rather, users should utilize it as it is appropriate to their context.

Since 2001, the European Commission has adopted the CEFR in its language policies such as EC Action Plan (Barni & Salvati, 2019: p. 418). Having been translated into over 30 languages, the framework has strongly and widely influenced language teaching and examining and has been employed and consulted by several countries around the world to develop foreign language policies (Figueras, 2012: p. 477). To gather more information about the use of the CEFR in more than 40 member states of the Council of Europe, Martyniuk and Noijons (2007: p. 2) conducted research, and the results demonstrated that the CEFR was predominantly employed and considered as a beneficial tool for all aspects of language teaching. In a similar vein, Broek & van den Ende (2013: p. 71) reported similar results of a research conducted to provide an insight into the extent to which the CEFR is implemented in six European countries.

Today, it is prevalently acknowledged that the CEFR is a symbol of globalization in education and a significant example of working together internationally in educational policy and practice (Normand-Marconnet & Bianco: p. 282). The framework has been called 'one of the most influential publications of the last decade in the field of language learning and language testing in Europe' (Figueras et al. 2005: p. 261). Furthermore, Byrnes (2007: p. 645) adds that the CEFR has gained such an undeniable momentum that all people involved in language education in Europe will need to refer to it at a certain point.

2.2.1. The Common Reference Levels

To date, the principal contribution of the CEFR seems to be the Common Reference Levels predominantly used by education stakeholders in order to outline standardized definitions of levels of language proficiency. Heyworth (2006: p. 181) includes that the best-known feature of the CEFR is the Common Reference Levels, being at the heart of the framework. The levels soon outshone the other aspects of the document and became the gem of policymakers, practitioners, researchers, and all the other people involved in language teaching (Figueras, 2012: p. 479).

The CEFR level descriptors provide a thick description of what learners are capable of in terms of language knowledge and skills, positively drafted in 'can-do statements'. From the CEFR point of view, learners progress through six stages under three main user-level categories. The Common Reference Levels in the CEFR characterize this learner progress in terms of communicative activities and communicative competences (North, 2005: p. 30). The levels are composed of a thorough description of levels along the global proficiency scale at six levels. The three broad user levels, the basic user (A1-A2), the independent user (B1-B2), and the proficient user (C1-C2) have been developed with the intention of reinforcing the usability of the framework. Furthermore, to comply with the learning outcomes of the European language learners, six broad levels, Breakthrough (A1) and Waystage (A2), Threshold (B1) and Vantage (B2), and Effective Operational Proficiency (C1) and Mastery (C2) have been outlined by the Council of Europe (Council of Europe, 2001: p. 23). In tables and scales of descriptors, the CEFR constantly displays C2 (Mastery) at the top and A1 (Breakthrough) at the bottom.

The 'can-do' statements describe what each level should be considered to subsume the level below on the scale. Put differently, a learner at the C1 level is presumed to be capable of what the descriptors include in the A1, A2, B1, and B2 levels of the CEFR. Furthermore, it must also be noted that C2 level learners are not characterized to be native speakers in the same manner that A1 level learners are not considered to be a complete beginner. The level descriptors only intend to label the language skills assumed to be typical for successful learners. Thanks to the CEFR levels, language educators are guided in terms of identifying the language competency levels of learners, which helps develop curriculum and courses with the aim of promoting communicative competence (Tannenbaum & Wylie, 2005: p. 1).

2.3. COMPLEXITY IN SLA RESEARCH

Within the SLA literature, the construct of complexity takes two significant roles: figuring as an independent variable and as a dependent variable (Bulté & Housen,

2012: p. 21). In the former, the influence of complexity on aspects of L2 performance or proficiency is investigated. To exemplify, Spada and Tomita (2010, p: 263) conducted a meta-analysis to investigate the effects of instruction on simple and grammatical features of English. In the latter, on the other hand, complexity is investigated to see whether it describes L2 performance and indicates L2 proficiency. In this line of research, the complexity is measured with the aim of displaying the effect of other variables such as age, level, etc. on L2 acquisition. However, the construct of complexity was not defined clearly or if it was defined, it was done vaguely in the current research (Bulté & Housen, 2012: p. 22). Consequently, the results of the studies presented mixed and contradictory results (Bulté & Housen, 2012: p. 22). In that vein, the more explicit characterization of the construct is of great importance to help interpret the results of the studies on complexity measurements.

2.3.1. Defining Complexity

Complexity has been researched comprehensively in several fields of research ranging from natural sciences to social sciences, and it was investigated in second language research as well (Khushik & Huhta, 2019: p. 2). However, researchers have not been able to reach a consensus on the definition of complexity (Bulté & Housen, 2012: p. 22), and there is a dearth of scrupulous definition to operationalize the construct except the acknowledgment that the concept is very complex encompassing several levels and dimensions (Norris & Ortega, 2009: p. 556). To address this issue, Bulté and Housen (2012: p. 22) suggested a framework detailing complexity reposing on the theoretical discussions by previous researchers such as Dahl (2004: p. 2) and Miestamo (2008: p. 8). The framework distinguishes a relative and an absolute approach to the construct of complexity. The former refers to the complexity in connection with language users' mental efforts whereas the latter defines complexity in objective and quantitative terms as the number of distinct components and connections between several linguistic features (Bulté & Housen, 2012: p. 24). Absolute complexity consists of linguistic, propositional, and discourse-interactional complexity.

Linguistic complexity is further divided into two components: grammatical complexity and lexical complexity (Bulté & Housen, 2012: p. 26), and grammatical complexity also consists of two major sources: syntactic complexity and morphological complexity. For the purpose of the current research, syntactic complexity and related research studies will be presented in the following sections.

2.3.2. Syntactic Complexity

Skehan (1996: p. 46) defines syntactic complexity (SC, henceforth) as "the stage and elaboration of the underlying interlanguage system". According to Ortega (2003: p. 492), SC is the set of forms appearing in language production and the extent of complexity those forms reach. Ryshina-Pankova (2015: p. 2) challenges this definition and presents a theoretically sound argument of "meaning dimension of complexity", or "discourse semantic motivations syntactic complexity". She argues that SC is one feature of language production driven by the practical necessity to use complex language within discourse (Ryshina-Pankova, 2015: p. 2). Ortega (2015: p. 5) adds that from an educational perspective, Ryshina-Pankova's approach to SC with a linguistic, communicative, and rhetorical intent is more likely to create practical connections between SC as objective and specific pedagogical matters and uses to cater to educational necessities.

The construct has received a considerable amount of scholarly attention in second language attainment research in recent years (Bulté & Housen, 2014: p. 43; Polio & Yoon 2018: p. 2). Polat, Mahalingappa, and Mancilla (2019: p. 3) assert that since SC is regarded as a robust indicator of L2 competence, researching on it can enhance the understanding of a wide variety of matters regarding L2 learning, use, development, and assessment. In a similar vein, Crossley and McNamara (2014: p. 67) argue that SC output of L2 learners equals to L2 progress as development in L2 necessitates the attainment and production of less frequent and larger range of syntactic features.

In SLA research, various measures have been suggested to assess the linguistic performance of L2 learners (Ortega 2003: p. 492). The SC measures most widely

utilized in language-related fields relied on length of units, and the same measures were also employed in SLA research (Norris & Ortega, 2009: p. 558). Furthermore, Biber, Gray, and Poonpon, (2011: p. 7) include that SC was traditionally measured with T-units. T-unit stands for a terminable unit, videlicet, a main clause, and all related dependent clauses (Renkema, 2004: p. 226). Wolfe-Quintero, Inagaki, and Kim (1998: p. 119) reported the reliance on T-unit-based measures in early research studies concluding that they are "the best complexity measures so far". Consequently, Biber et al. (2011: p. 8) assert that it is probable that research on L2 development depended greatly on these measures because of the recommendation of Wolfe-Quintero et al. (e.g., Brown, Iwashita, & McNamara, 2005: p. 11; Gyllstad, Granfeldt, Bernardini, & Källkvist, 2014: p. 11; Larsen-Freeman, 2006: p. 595; Jiang, 2012: p. 11). The use of T-units as measures of SC in SLA research has yielded contradictory results, with some research claiming no relationship between the measure and syntactic growth (e.g., Casanave, 1994: p. 179; Ishikawa, 1995: p. 59) and some claiming a strong relationship (Ortega, 2003: p. 509; Stockwell & Harrington, 2003: p. 342).

Other than T-units, the mean number of clauses per Analysis of Speech unit (AS-unit) has been frequently utilized in SLA research since Foster et al. (2000: p. 365) proposed it as a better option for oral discourse segmentation. Moreover, Bulté and Housen (2012: p. 21-46) counted more than 40 different complexity measures in research studies published between 2005 and 2008, including measures of T-unit, clauses/sentences, dependent clauses/ total clauses, etc. (Bulté & Housen, 2014: p. 44). Crossley and McNamara (2014: p. 68) also state that other SC measures employed in L2 research include the embeddings, the coordination, phrasal units, and the frequency of clauses and phrases used.

2.3.3. Empirical Studies on Syntactic Complexity

Researchers of L2 development are incrementally focusing on measures of complexity as well as accuracy and fluency to assess learners' written and oral output (Polat & Kim, 2013: p. 186). In that vein, SC has been widely researched on the writing

performance of L2 learners. Some of this research have investigated the acquisition and the use of SC measures' particular aspects in asynchronous online discussions (e.g., Mancilla et al. 2015: p. 112), whereas others focused on the variability of development of SC (e.g., Verspoor et al., 2008: p. 214). Overall, several aspects and measures of SC have been paid attention to in L2 writing research.

One aspect of research on SC in L2 writing is instructional effectiveness in several L2 programs. For instance, Cooper (1976: p. 176) investigated the development of SC between five groups of different proficiency levels and reported significant differences between levels beyond one year of instruction. Moreover, Serrano et al. (2012: p. 138) analyzed 14 Spanish-speaking learners of English from a longitudinal perspective. They analyzed the data in terms of SC as well as other aspects and found out linear improvements in the complexity of L2 writing. Likewise, the findings reported by Mazgutova and Kormos (2015: p. 3) suggest that instruction plays an important role in SC development over time.

Another aspect of research on SC and L2 writing is the correlation between learners' proficiency and SC measures. Several research studies on proficiency and SC have been conducted on the SC indices. One commonly researched index is the mean length of clause (MLC). In their research, Cumming et al. (2005: p. 5) assessed the discourse of texts written for TOEFL with reference to lexical and SC and grammatical accuracy as well as several other aspects. They reported a significant positive relationship between SC and proficiency levels of the learners. Similarly, in her synthesis of college-level L2 writing, Ortega (2003: p. 496) focused on the studies investigating the relationship between SC and proficiency differences of learners. The analysis focused on the six most prevalently used measures of SC including MLC. The quantitative analysis on MLC revealed that there may be a statistically significant difference between proficiency level and MLC provided that the differences are slightly over a word and the sample is large. However, Knoch, Rouhshad, and Storch (2014: p.1) examined learners' writing proficiency in terms of accuracy and syntactic and lexical complexity and found out that there were no observed gains in terms of MLC in relation to learners' proficiency.

Another favorable measure employed in SLA research of SC is T-units. To exemplify, Benzehaf (2017: p. 43) investigated the correlation between English proficiency and complexity, accuracy, and fluency measures in writing. To measure complexity, they utilized the number of dependent clauses per T-unit, and the results yielded a strong correlation between proficiency and complexity. In another study, Becker (2010: p. 406) investigated complexity differences across different levels and found differences in the number of clauses per T-unit and words per T-unit. However, he observed no differences with regard to the number of T-units across groups.

In another writing research, Díez-Bedmar and Pérez-Paredes (2020: p. 4) examined noun phrase SC in the writing of Spanish secondary school learners in the International Corpus of Crosslinguistic Interlanguage by using manual parsing and an automated analysis tool. The results of their study revealed that SC in noun phrases develops in premodifying slots, and they argue that nouns and modifiers, and determiner + multiple premodification + head can be used as measures of SC in learner language development.

Research on SC of L2 writing has provided significant results to the field. On the other hand, from both empirical and theoretical perspectives, it is widely acknowledged that spoken and written output are complex in different ways (Biber et al., 2011: p. 10). Some features associated with complexity are far more common in speech than in writing (see Biber et al., 2011: p. 1). To this end, the research on SC in L2 speaking should also be presented.

Although complexity is widely associated with L2 writing research, there is a plethora of research on SC and L2 speaking as well. Several researchers investigated SC through analyzing speech samples. To exemplify, Ortega (1999: p.109) sought to investigate the impact of pre-task planning opportunity and SC, lexical range, accuracy and fluency as well as other aspects regarding the planning process of learners. The results demonstrated that the mean number of words per utterance was significantly higher in the planned output. In another study, Iwashita, McNamara, & Elder (2001: p. 401) investigated the association between task characteristics and performance

conditions and different levels of fluency, complexity, and accuracy. They measured complexity by calculating the number of clauses per C-unit. The results yielded that task conditions did not have an impact on the scores of learners regarding complexity and other aspects. Furthermore, Ferrari (2012: p. 277) investigated the development of SC, accuracy and fluency in four L2 learners of Italian and two native speakers of Italian in relation to task types over the course of three years. The data were collected with the help of dialogic and interactive tasks, and they were measured with AS-unit (a unit specifically designed for spoken production) and the average number of words per clause. In terms of SC, her findings suggested that learners' scores increased for clause length, but no increase was observed for subordination.

More recently, by addressing the research lacuna of advanced naturalistic learning, Polat and Kim (2013: p. 187) examined the language development of an untutored Turkish immigrant over one year in comparison with three native speakers. After they converted the data to another format to be compatible with the program they utilized, the data were analyzed for SC measures (viz., mean length of AS-units, clauses per AS-unit, mean length of clauses), lexical diversity, and accuracy. The results demonstrated that SC in learner's speech showed "potential but unverifiable gains" (Polat & Kim 2013: p. 184). Furthermore, Lahmann et al. (2015: p. 354) investigated the factors influencing grammatical and lexical complexity with the help of L2 spontaneous oral interviews with a sample of 102 participants. The measures they employed in the research consisted of the mean number of words per AS-units, the mean number of dependent clauses, the mean number of nonfinite adverbial dependent clauses (DCs) per AS-unit, and mean number of words per noun phrase. According to the results, gender and level of education have a significant impact on syntactic and lexical complexity.

In another study, Nippold et al. (2017: p. 1) examined SC in L2 speaking with the help of interviews consisting of different tasks. They analyzed the data for mean length of communication unit and clausal density. The results yielded that SC was greater in narrative and critical-thinking tasks than the conversational task. Also, De Clercq and Housen (2017: p. 315) analyzed cross-sectional data from adolescent native speakers of four different proficiencies. The results revealed a progressive increase in

SC in both groups. In addition, Vercellotti (2018: p. 233) examined the development and variation of SC in L2 speech of 66 learners over three academic semesters. All measures of SC showed significant and meaningful progress in the speeches of L2 learners over time. More recently, Lambert and Nakamura (2018: p. 1) compared the discourse of Japanese learners of English at three levels (native, advanced, and intermediate) with their native English-speaking peers in completing six communication tasks to better comprehend the emergence of SC in terms of developing (L2) proficiency. The results revealed that some features of SC (the four types of clause combination strategies) vary with proficiency level.

Taking stock of the research carried out on SC, one may observe that the concept has been employed in several research studies from the perspectives of task modalities (written or spoken), groups of L2 learners and proficiency levels, and various measures have been utilized in this line of research (Kuiken et al., 2019: p. 3).

2.3.4. Syntactic Complexity and Coh-Metrix

Researchers conducted studies on SC measuring from various perspectives. However, less attention has been directed toward a discussion of how automated measures were employed in L2 research (Polio & Yoon, 2018: p. 2). Automated tools have started to provide a more available and theoretically sound approach to scrutinizing SC (Azadnia, Lotfi & Biria, 2019: p. 235-236). One automated tool that can be used for complexity measures included in research extensively is Coh-Metrix which was first introduced as a part of a research study by Graesser et al. (2004: p. 193). Coh-Metrix enables researchers to gauge deeper levels of textual features and characteristics including SC, and it has been widely employed in linguistics and applied linguistics fields.

Of those researchers using Coh-Metrix to measure SC, McNamara, Crossley, and McCarthy (2010: p. 57) aimed to detect linguistic features of writing quality in English as the first language. To this end, they examined 120 argumentative essays from undergraduate students using Coh-Metrix measures of the mean number of higher-level

constituents per word and number of words before the main verb. Findings revealed that the biggest significant difference was in the use of left embeddedness measure between high- and low-proficiency essays. Furthermore, Crossley and McNamara (2011: p. 271) investigated the intergroup homogeneity within the high intermediate and advanced L2 writers from four different first language backgrounds in order to determine which features distinguished among L1 groups. They investigated the left embeddedness structure in the groups and found differences between some groups but did not contribute much to discriminate the groups. Based on this result, the researchers suggest that some aspects of L2 writing may rely on linguistic knowledge as a result of language experience and proficiency level. Kormos (2011: p. 148), in a similar vein, examined how SC differs in two different task types. By using left embeddedness and modifiers per noun phrase measures on Coh-Metrix, the research found that there were no differences in complexity on the two Coh-Metrix measures.

In a later study, Banerjee et al. (2015: p. 5) set out to identify which text features predict writing proficiency and distinguish among levels of proficiency. By operationalizing SC as the number of modifiers per noun phrase by Coh-Metrix, the researchers found that SC did not distinguish the levels of proficiency. However, it must be noted that the researchers only used one measure of SC, which, as also they claim, might have affected the results of the research. In a later study, Riazi (2016: p. 15) explored readability based on word length and sentence length, left embeddedness, sentence syntax similarity, passive density, and the average number of modifiers per noun phrases with an aim of comparing learners' writings on three tasks. The tasks were observed to be similar in terms of left embeddedness and sentence syntax similarity.

More recently, Shooraki, Barati, and Moinzadeh (2020: p. 84) examined the linguistic and discoursal differences in TOEFL-iBT essays of Iranian learners. Using Coh-Metrix, they analyzed the texts in terms of text easability, cohesion, lexical sophistication, and syntactic complexity. The findings of the study revealed that analyzing discourse qualities such as cohesion can help anchor the scores of test-takers empirically.

2.3.5. Syntactic Complexity and the CEFR

In the literature review, several studies that have measured SC from different aspects of language were perused. Extending the literature review of SC in L2 writing and speaking research, this section will include the relationship between the CEFR and SC and current research carried out to investigate it.

Housen et al. (2012: p. 302) state that some significant questions about the link between complexity and proficiency rating scales (e.g., the American Council on the Teaching of Foreign Languages proficiency scale, CEFR proficiency scale) remain unanswered to this day. Consequently, these two important concepts in language education circles should be considered an important aspect, and research related to the relationship between the CEFR and SC should be reviewed.

The concept of complexity is repeatedly mentioned in the CEFR document referring to complex language, complex speech, simple syntax, etc. (Gyllstad et al. (2014: p. 5). Accordingly, one may observe that the CEFR is broadly representative of the concept of complexity and its antonym simplicity. However, regardless of being frequently referred to, the concept is not defined thoroughly in the document (Gyllstad et al., 2014: p. 5). By analyzing the terminology used in all scales of the CEFR, Gyllstad et al. (2014: p. 5) observed four issues. First, the terms complex, simple and basic occur highly frequently in the document. Second, the terms specifically repeatedly occur at the beginning levels of the CEFR. Third, the development starts in simple structures and progresses into incrementally more complex structures. Finally, in line with the third observation, the researchers included that at the beginning levels only adjectives simple and basic are included whereas, at B1 level, descriptors start including the adjective complex alongside simple and basic, and at levels above B1, only the adjective complex is included. Consequently, it is fair to assume that the CEFR considers the B1 level appropriate for learners to start producing complex sentences.

2.3.6. Syntactic Complexity, the CEFR, and Empirical Studies

Researchers of L2 have endeavored to distinguish linguistic features of L2 production across different levels of L2 proficiency (Bulté & Roothooft, 2020: p. 1). English Profile (EP) project is one of the significant attempts aiming to match different levels of the CEFR with English lexical and grammatical features (Hawkins & Filipović, 2012) (For an overview of research on EP project see Kurteš & Saville, 2008: p. 2-4; Salamoura, 2008: p. 5-7). Other than the EP project, several studies have analyzed the relationship between L2 complexity and the CEFR proficiency scales.

In L2 studies, the definition of proficiency is rather vague or implicit, and different operationalizations arise, which also applies to the studies investigating the correlation between linguistic features and L2 proficiency levels (Bulté & Roothooft, 2020: p. 1). A number of studies have operationalized proficiency in terms of the CEFR levels. An example of work relying on the CEFR proficiency levels and SC can be found in Kim (2004: p. 31) who investigated CEFR-rated scripts obtained from 33 Chinese learners of English on an English for Academic Purposes (EAP) course. The study aimed to describe how learners at different CEFR levels (viz., A2, B1, B2) vary in grammatical complexity in writing. The SC measures used in the research were a variety of structures, the number of subordinate clauses, and shift from clauses to phrases. The results revealed a progression from A2 to B2 levels in all SC measures except for nominal clauses per clause and gerund phrases per clause, and the difference in SC was far clearer between B1 and B2 compared to A2 and B1.

Furthermore, Kuiken, Vedder & Gilabert (2010: p. 81) examined the relationship between communicative adequacy and linguistic complexity operationalized as SC, lexical diversity and accuracy to provide learner performance at a particular CEFR level and contribute to the portrayal of learner language at a specific proficiency level. The researchers analyzed 200 short essays written by three groups of university students consisting of international students learning Dutch as L2, Dutch students learning Italian as L2, and Dutch students learning Spanish as L2. The participants varied in the CEFR proficiency level from A2 to C1. The results did not demonstrate a significant

correlation between communicative adequacy and measured SC, and the SC measures did not seem to guide the rated CEFR levels for linguistic complexity.

Another example of research based on CEFR proficiency levels and linguistic features, as mentioned above, can be found in the EP project (Hawkins & Filipović, 2012). The results have demonstrated that one measure of SC, mean length of utterance, progresses in line with learners' CEFR levels (Hawkins & Filipović, 2012: p. 23). The researchers conclude that learners produce more syntactically complex utterances as they progress into upper CEFR levels. Similarly, Green (2012: p. 124) aimed to distinguish between texts at the CEFR B2, C1, and C2 levels. By using three automated text analysis tools (Coh-Metrix, Wordsmith tools, and RANGE), Green (2012: p. 124) investigated lexical and syntactic complexity across CEFR levels. Significant differences in the noun phrase incidence and the number of modifiers per noun between B2 and C1 levels were reported in the study. In addition, the results revealed that sentence syntax similarity differs in C1 and C2 levels.

Moreover, Verspoor et al. (2012: p. 1) aimed to investigate the indices of language development in L2 written data based on the Complexity, Accuracy, and Fluency (CAF) model. To do so, the researchers investigated 437 Dutch learners of English. They collected descriptive written data from the participants and rated them in line with CEFR levels A and B. The results obtained from the data revealed an increase in the mean T-unit length across levels. Additionally, researchers concluded that learners at the beginning levels are more involved with words rather than syntactic complexity. Also, they found complex sentences to be a reasonably good separator of levels, especially between A1.2 and A2, dependent clauses to be a notably good separator and relative clauses to increase continually across all levels (Khushik & Huhta, 2019: p. 5).

In another study, Gyllstad et al. (2014: p. 1) investigated the correlation between linguistic features and the CEFR proficiency levels, though the texts in the data set ranged preponderantly between A1 and B2. The data consisted of 120 learners' written texts in three different languages L2 English, L3 French, and L4 Italian and were analyzed through SC measures (length of T-unit, subclause ratio, and mean length of

clause). The results yielded that overall and in each language, three measures of SC significantly correlate to assigned CEFR levels. Moreover, the researchers found out that learners at CEFR level A do not significantly differ in SC irrespective of language while learners at CEFR level B do differ across languages.

In a similar vein, Kang and Yan (2018: p. 24) also examined linguistic features distinguishing CEFR levels. Applying a quantitative/corpus-based approach, the researchers analyzed speaking performance in several linguistic features across four CEFR levels. The researchers coded the data for linguistic features for grammatical and lexical complexity, discourse management, and pronunciation. They measured grammatical complexity through verb-phrase complexity per T-unit, the number of clauses, and the number of dependent clauses. Though the research did not operationalize SC as a measure, the measures of complexity were the measures used in SC research, as the literature review suggests. The findings suggest that all measures of SC, except the number of T-units, increased significantly between B1 and C2 levels. Moreover, the total number of clauses, total number of dependent clauses, and clauses per T-unit were found to be the measures to distinguish between B1, B2, and C2 (Bulté & Roothooft, 2020: p. 4)

More recently, Alexopoulou et al. (2017: p. 180) investigated how learner language developed in a longitudinal learner corpus-based on linguistic complexity and accuracy analysis across proficiency levels. Researchers analyzed texts on an open-access corpus consisting of writings submitted to an online school whose curriculum contains CEFR levels from A1 to C2. Based on three SC indices (viz., average sentence length, mean length of clause, and subordinate clause per T-unit), the researchers reported an increase in sentence length across all CEFR levels as well as an increase in the length of clause from A2 to B2 and in subordinate clause per T-unit from A1 to B2. In addition, Lahuerta Martínez (2018: p. 1) examined differences in SC in L2 English writing by means of SC measures between stated A2 and B1 levels and compared the scores of SC measures with learners' overall writing quality. The researcher reported a significant correlation between SC and the writing quality of learners. Moreover, the study reported a significant difference between the levels in terms of sentence length,

compound and complex sentence ratios, coordinate and dependent clause ratios, and noun phrases per clause. However, the placement of learners at CEFR levels in this research is only based on the researchers' statements and learners' grade levels.

Different from other empirical studies, in a research paper investigating two linguistically different groups of English learners in two different countries, Khushik and Huhta (2019: p. 8) examined how the CEFR levels differ in SC in the writings of learners from two different backgrounds. The data were gathered from 868 Pakistani and 287 Finnish learners who wrote the same argumentative essay rated on a CEFR-based scale. By means of automated analysis tools, the texts were analyzed through 28 SC indices. The indices with the most significant difference between groups were length measures and phrasal density. Also, the CEFR levels A1, A2, and B1 were observed to be distinguished by the length of production units, subordination, and phrasal density indices. However, A2 and B1 levels were found to be not comparable in terms of SC between the two groups. Finally, the researchers included that as the proficiency of learners' progress, linguistic differences increase as well.

Taking together the other aspects that have been researched on in the previous studies, Kuiken and Vedder (2019: p. 193) examined the extent of variation and regularities in SC across CEFR levels between L2 and L1, and across different languages. The researchers carried out the analysis on the basis of written argumentative texts from 32 L2 learners of Dutch, 39 of Italian, and 23 of Spanish, and the proficiency level of the participants ranged from CEFR A2 to B1. Assessing the SC by means of the mean number of clauses per T-unit, the number of dependent clauses per clause, and other specific measures used for subordination, coordination, and phrasal complexity, the researchers found out that as the proficiency increase, learners use more coordinate and subordinate clauses. Also, the results implied a variation in the process of progressive complexification across proficiency levels, across different languages, and across L1 and L2 (Kuiken & Vedder, 2019: p. 192)

2.4. DISCURSIVE FEATURES

Many linguists define 'discourse' as anything independently of the notion of the sentence (Schiffrin, Hamilton & Tannen, 2011: p. 1), focusing on context and social functions rather than operating on grammar level (Mohamadi & Rahimpour, 2018: p. 19). Though learners may master the vocabulary and grammar of a target language, it is still a possibility that they produce compositions that are not acceptable due to problems stemming from the use of discourse. Mohamadi and Rahimpour (2018: p. 19) also include that the process of comprehension and getting the message across require several things, namely, grammatical, and phonological elements, context, situation, purpose, pitch, intonation, and gesture. Therefore, it is fair to say that analyzing discursive features, the efficacy of these features, and how they are used by learners of L2 is significant.

One aspect of discourse that has been widely researched in L2 circles is the use of cohesion in learner language. Halliday (1994: p. 309) defines cohesion as the set of resources for constructing relations in discourse beyond grammar. Halliday and Hasan (1976: p. 14) organized the inventory of cohesive resources as 'reference' (anaphoric, cataphoric, and exophoric) referring to recoverable resources such as pronouns, definite article, adverbs here, there, etc. in a context, 'ellipsis' referring to resources for omitting a structure in a context where it is possible to assume what is omitted, 'substitution' referring to resources indicating to a group of items for nominals, verb groups, and clauses – e.g., so, not, do, one, etc., 'conjunction' referring to resources comprising linkers connecting sentences to each other, and 'lexical cohesion' including the repetition of lexical items, synonymy, collocation, etc.

Another aspect of discourse that has been prevalently researched is the concept of coherence. Coherence is considered to be an underlying phenomenon concerning semantico-pragmatic links between parts of the text that can be interpreted deciphered via the background of particular world knowledge (Gómez González, 2013: p. 128). Traditionally, coherence has been regarded as the formal criteria distinguishing texts from non-texts; however, more recently, it has been viewed rather as a cognitive process

in which learners build a mental representation of the information given, making the concept as a more mental representation of the discourse (Gómez González, 2013: p. 129). Though coherence is much easier to detect *prime facie* in a well-written text, it can be located in the spontaneous oral production, the most revealing medium for coherence (Givón, 1995: p. 59). Also, one cannot deny that conversation is comprehended and mostly rely on the coherence as, without the coherence, the speech would be hard to comprehend and follow. In discourse, coherence exists through meaningful and coherent units; however, the connections may occasionally be unclear. Cohesive markers help to make such connections between the sentences of discourse overt (Hatch, 1992: p. p. 209). In that sense, cohesion and coherence are two terms that are highly associated with each other. However, one should not assume that cohesion necessarily results in coherence since coherence can exist even in the case of no cohesion (Gómez González, 2013: p. 129).

In addition, discourse markers (hereafter, DMs) in L2 output are also a widely researched aspect of discourse. Starting from the 1980s, the study of DMs has evolved into an expanding sector in language studies; however, no one operationalization has been made (Fraser, 1999: p. 932). Research on DMs has been conducted under several labels including discourse connectives (e.g., Blakemore, 1992: p. 139), pragmatic expressions (e.g., Erman, 1992: p. 1), pragmatic markers (e.g., Fraser, 1988: p. 19) and so many others (for a broader set of references see Jucker & Ziv, 1998: p. 1).

The multiplicity of labels surrounding DMs demonstrates varying interest in research, which causes difficulties for operationalizing them sufficiently in theoretical terms (Fung & Carter, 2007: p, 411). However, regardless of the label they were researched on, DMs usually include expressions such as *I mean, so, oh, then, well, you know*, and so on. Schiffrin (1987: p. 326) included that DMs functioned to "add to discourse coherence", providing contextual coordinates for an utterance. Fraser (1999: p. 938) also included that regardless of the name or label, DMs function in two places: one in the segment they introduce, and the other in the prior discourse. In the current research, the label "discourse marker" was chosen as a convenient term as Jucker and

Ziv (1998: p. 1) state, it is the term that is used it in the literature with the widest currency and acts as an umbrella term including a variety of elements.

Furthermore, information load "refers to the variety of stimuli (in type and number) to which the receiver must attend" (McCormick, 1970: p. 114). In the presence of information overload, the receiver is not able to assimilate and process information at a given time, and human performance becomes less accurate and less effective (Jacoby, 1977: p. 569). Not much research has been conducted regarding information load in second language research.

Finally, code-switching (CS) is an aspect of discourse having been broadly investigated in language-related fields. Different researchers have made various definitions of the phenomenon without achieving any consensus about the terminology. Some of the earlier definitions of the phenomenon include “the alternate use of two or more languages, varieties of a language, or even speech styles” (Hymes, 1977: p.103) or “the use of more than one linguistic variety, by a single speaker in the course of a single conversation” (Heller & Pfaff, 1996: p. 594). There are various functions of CS in learner language such as gap-filling, expressing ethnic identity, and achieving specific discursive goals (Bullock & Toribio, 2009: p. 2). CS includes more than one language, and CS occurrences betoken underlying reasons for CS. To this end, CS utterances have been approached from various angles. Several models of code-switching have been developed to answer the many questions uttered by theoretical linguists (see Woolford, 1983: p. 522 for a more detailed listing of the models). Besides, Woolford (1983: p. 529) asserts reasons for switching between languages one of which is the length of utterances. Moreover, Hamers and Blanc (2000: p. 267) also included reasons for the switch and assert that just as CS could be used by bilinguals with high competence in both languages, it could also be used as a reparation tool for inadequacy in the target language. In that vein, Song and Andrews (2009: p. 59) include that by code-switching, learners “attempt to keep the conversation flowing without having to pause or abandon the message”.

2.4.1. The CEFR and Discursive Features

The CEFR does not directly handle discursive features on their own, rather, it handles the concept of discourse in the term 'discourse competence'. In the document, the term discourse competence is used to refer to learners' ability "to arrange sentences in an order in order to produce coherent stretches of language" (Council of Europe, 2001: p. 123). According to the CEFR, as learners become more proficient in the language (as they progress into the more advanced levels, e.g., C2) they are expected to develop good discourse competence, organizing their language in a cohesive, coherent, and effective manner (Council of Europe, 2001: p. 123).

One aspect of discourse coherence mentioned in the document is coherence and cohesion which are discursive features of language as aforementioned. In the illustrative scale for several aspects of discourse competence (Council of Europe, 2001: p. 125), the CEFR makes specific predictions about the use of coherence and cohesion which seem to have been primarily designed for speech but are applicable to written texts. To exemplify, when describing A1 learners in terms of the use of coherence and cohesion, the document states that "they can link words or groups of words with very basic linear connectors like and or then" (p. 125). In addition, the framework predicts that C1 level learners can "produce clear, smoothly flowing, well-structured speech, showing controlled use of organizational patterns, connectors, and cohesive devices" (p. 125).

Moreover, in the illustrative scale for general linguistic knowledge, it is included that C1 level learners "can select an appropriate formulation from a broad range of language" (Council of Europe, 2001: p. 110). The word choice of 'appropriate' is curious as in the previous levels, the word 'sufficient' was preferred. In that vein, it can be assumed that C1 is the threshold for learners to be able to choose the appropriate one from a range of language items to be coherent in discourse. Such a distinction of the role of discourse in more advanced levels of the CEFR delineates it as the defining features of these levels (Waller, 2015: p. 69). Consequently, it is fair to assume that as learners progress into the upper levels, they are required to produce more coherent structures.

In terms of code-switching, the CEFR (2001) has been contributory in recognizing the merit of code-switching as the framework promotes "plurilingualism in

response to European linguistic and cultural diversity". Piccardo and North (2020: p. 284) state that according to the CEFR, plurilingualism is the ability to call "call flexibly upon a holistic, integrated, inter-related, uneven, plurilinguistic repertory in which all linguistic abilities have a place." Piccardo and North (2020: p. 284) relate some characteristics of the plurilingualism as described in the CEFR to code-switching which describe the process of crossing the boundaries between languages. Hence, code-switching might be related to the CEFR though it is not explicitly stated in the document. Moreover, although being perceived as damaging elements in language pedagogy in the literature (Malakoff & Hakuta, 1991, cited in Cook, 2001: p. 417), L1, translation and interpreting were considered relevant in the framework since these elements enable learners to mediate between languages and cultures (Gutiérrez Eugenio, 2013: p. 444). The CEFR aims to encourage the valorization of plurilingualism, however, the CEFR (2001) did not provide any descriptors for plurilingualism (Piccardo & North, 2020: p. 283-284), hence, there are also no descriptors for code-switching to demonstrate how learners' deployment of the feature alters with the change of proficiency level.

On the other hand, though having been included in the literature as a construct adding to discourse coherence, DMs and information load have not been included in the CEFR. Consequently, the framework does not state how they may evolve as language proficiency increases.

2.4.2. The CEFR, Discursive Features, and Empirical Studies

Several studies have investigated the discourse features of different learners in various contexts. To exemplify, the construct of cohesion is one of the prevalently investigated sub-fields of second language writing. Although some researchers came to similar findings, the findings of research on the construct of cohesion have been rather contradictory. Of those researchers investigating the difference in the use of cohesive devices in high-rated and low-rated writings, Jafarpur (1991: p. 459) found that cohesive elements can be the defining features of the proficiency of learners as high-rated essay

differed from low ones in their use of cohesive devices by using more cohesive devices. In a similar vein, Norment (1995: p. 561; 2002, p. 98) examined the cohesive devices in three types of essays of African American students at low and high proficiency levels in a research study focusing on the correlation between the use of connectives and the level of proficiency. The results yielded a positive correlation between the use of cohesive devices and students' proficiency levels.

In a later study, Mohammed (2015: p. 74) analyzed the use of various forms of connectives in the writings of English language learners. The analysis revealed that students only utilized eleven connectives in their writings, and a significant difference was observed in the use of 'and' between high and low rated texts. For the types of connectives, temporal and demonstratives were found to be the ones hardly employed by learners. However, Zhang (2000: p. 61) found out that there was no significant difference between high and low-rated writings in their use of cohesive devices. Similarly, Castro (2004: p. 215) investigated the relationship between writing quality and use of cohesive devices between low mid and high rated essays and observed no significant differences in grammatical cohesive device use, conjunction use, and in the frequency of occurrence of reference and conjunction.

As the CEFR makes strong predictions about cohesion in texts, research on cohesion across CEFR levels is of great significance to search for the practical applications of the document. However, though some studies do exist, the use of cohesion in L2 learners of different CEFR levels is not widely investigated compared to other areas of research. One example of those studies can be found in Carlsen (2010: p. 191) who investigated whether the predictions made in the CEFR about learners' use of discourse connectives are supported by authentic learner data. Referring to conjunctions as discourse connectives, Carlsen (2010: p. 193) investigated the use of *and, but, because, however, despite, furthermore* in a computer learner corpus of written Norwegian with ten different first languages. The results indicated a high degree of correlation between the CEFR levels and the use of discourse connectives, and lower levels and higher levels of proficiency differed in their use of discourse connectives.

Moreover, Springer (2012: p. 55) put lexico-grammatical discourse devices (viz., attention-getting/focusing devices, backgrounding/clause combining devices, and cohesive devices) under scrutiny in Dutch learners' academic prose in English in the light of the C1 and C2 levels of the CEFR. The researcher aimed to establish the instances of overuse and underuse of such devices by comparing a learner corpus with a native corpus. The results revealed a significant overuse of the connectives in the categories of addition, enumeration, result, inference, and purpose in C1 and C2 levels, and revealed significant underuse of the elaboration category. Also, the use of connectives 'and' and 'but' appeared to be not homogenous in both learner and native corpus.

Similarly, Zarco-Tejada et al. (2016: p. 215) aimed to analyze which and how conjunctions vary among the CEFR levels of A2, B1, and B2 in CLEC corpus (CEFR-levelled English Corpus). The results indicated that upper proficiency levels show a greater level of connectives uses compared to the lower levels. In terms of the connective categories, causal and temporal connectives are mostly seen in upper levels while additive connectives demonstrate higher results in lower levels (A2 and B1), and the highest use of adversative connectives is observed in the B1 level of proficiency. In a later study, Iwashita, May, and Moore (2017, p. 12) investigated discoursal and lexical performance across different CEFR levels in spoken test performances. The results revealed that B2, C1, and C2 levels used logical and adversative/contrastive conjunctions more frequently than the other levels whereas no pattern was observed in the use of temporal, expanded temporal and additive conjunctions. Additionally, the only significant difference between the CEFR levels and frequency of conjunction was observed in adversative/causative and additive conjunctions.

Another discourse feature that needs to be addressed is DMs. Although having been ignored in research in the past, there has been a proliferation of research on DMs in the last three decades in several fields of languages (Torres & Potowski, 2008: p. 263). Various aspects of DMs have been studied in several contexts and language proficiencies. To exemplify, DMs have been studied to investigate whether and how they affect comprehension. One exemplary work can be found in Chaudron and

Richards (1986: p. 113) who examined the ways the extent of DMs affecting foreign students' understanding of university lectures. The research specifically focused on the effects of micro markers (e.g., fillers) and macro markers (those indicating overall organization). Macro markers were found to help in the recall of the text material better than micro markers. In a study inspired by Chaudron and Richards (1986: p. 113), Pérez and Macià (2002: p. 7) conducted exploratory research aiming to explore to what extent the presence or absence of DMs affect comprehension. The results indicated that students' proficiency in English and the types of DMs are key factors to affect the comprehension of listening by learners.

Other than how DMs affect comprehension; some specific DMs have also been put under scrutiny to investigate how they are employed in learner language. For instance, Fuller (2003: p. 23) examined the use of *you know, like, oh, well, yeah,* and *I mean* in interviews and casual conversations with an aim to identify their role in marking and negotiating speaker roles. The results revealed that *oh* and *well* are most frequently used in conversations whereas *you know, like* and *I mean* did not indicate any significant differences between contexts. In a similar vein, Müller (2004: p. 1157) specifically investigated the use of *well* by German learners of English in comparison with its use by American native speakers. Analyzing data from a paired silent film retelling, the researcher found out that non-native speakers used *well* more frequently than native speakers. In more recent research, House (2013: p. 57) examined how learners of English improve their pragmatic competence by using the DMs *yeah/yes, so,* and *I know*. The learners in the study were found to achieve pragmatic fluency by creatively re-interpreting *yeah/yes, so,* and *I know* for their own discourse structuring purposes.

Though previous research has investigated the use of DMs by learners, little is known about the DM use across a wide spectrum of language learners. In addition, most studies of DMs have focused on a single proficiency level or by no means considered the impact of proficiency. Hellermann and Vergun (2007: p. 157) note that learners' proficiency level takes part in how frequent and diverse these types of expressions are. Consequently, the relationship between DMs and learners' proficiency should be dealt

with to gain more insight into DM research. However, it must also be noted that the research does not define or identify the markers by using the same terminology though most use the same expressions.

Of those few studies that addressed proficiency in relation to DMs, Hasselgreen (2004: p. 160) examined the role of 'smallwords' (e.g., *I think, you know, sort of, right, well, ok*) and how they contribute to three levels of proficiency (native-speaking high school students from the United Kingdom, high-proficiency Norwegian students, and lower proficiency Norwegian students). She found out that the frequency of the use of smallwords was significantly higher in native speakers than that of nonnative speakers. Moreover, *I think* was observed to be the most frequently used DMs in nonnative groups whereas *just* was most frequently used by the native group. In a similar vein, Hellermann and Vergun (2007: p. 157) investigated the relationship between DMs *well, you know,* and *like* and proficiency levels from absolute beginner (Level A) to advanced (Level D). The findings indicated that the average frequency of DMs increases as learners' proficiency level increases although the lower levels did not employ them as frequently as higher levels. Using the list provided by Hasselgreen (2004: p. 160), Neary-Sundquist (2014: p. 637) examined the rate and range of DMs by analyzing data from learners at varying proficiency levels in comparison with native speakers. The results demonstrated an increase in DM use as proficiency level, and there were significant differences between levels although the pattern was not utterly straightforward. Moreover, *I think, so,* and *also* were found to be the three most frequently used DMs in Level 3, 4, and 5 whereas *just, so* and *you know* were most frequently employed by Level 6 learners.

In another research study, Fernández, Gates Tapia, and Lu (2014: p. 150) examined the use, frequency, and functional differences of markers *pues* and *bueno* in Spanish as a foreign language at two levels of proficiency. The study investigated whether learners' proficiency level would have an impact on the use of markers. The proficiency level was found to be significantly relevant to the overall frequency and variety of markers. Overall, research on the relationship between proficiency level and

DMs has revealed that the frequency and variety of DMs depend heavily on the proficiency level of learners.

The research investigating the correlation between proficiency level and DMs have mostly relied on either their own proficiency scales or used other examinations to understand learners' levels. Few studies correlated DMs to the proficiency level by using CEFR scales. One exemplary work is found in Corsetti and Perna (2017: p. 302) who examined the most common discourse marking adverbs used by Brazilian learners of English at the CEFR B1 level with an aim of tackling the intermediate level of the CEFR. Overall, it was revealed that participants produced a limited array of discourse marking adverbs though they used *really* consistently and did not use *well* and *actually* to the same degree. Another CEFR and DMs related research was carried out by Kang, Larson, and Koo (2019: p. 1). By analyzing 58 video-files of CEFR B1 to C2 in high-stakes contexts, the researchers investigated four interactive features: co-operation, coherence operationalized as DMs, turn-taking, and strategy use. The results of the study revealed that DMs did not significantly differ among groups, only distinguishing between B2 and C2, but not B1 and C1. From a different perspective, Jones and Carter (2014: p. 37) examined the effect of two explicit teaching frameworks on teaching DMs to two Chinese learner groups compared to a control group at the CEFR B2 level. According to the results, the experimental groups outperformed the control group with increased use of DMs. While research has been conducted in relation to CEFR levels and the use of DMs, less attention has been paid to how frequency and use of DMs vary across levels.

2.5. CONCLUSION

The literature suggests that learner language has been investigated from several perspectives including syntactic complexity and discursive features. Furthermore, CEFR, a symbol of globalization in education, has also been investigated to gain a more in-depth understanding of the document and also to investigate whether the document represents authentic learner data. Though various research studies have contributed to

the field, CEFR-related studies are relatively limited from several aspects. Most of the studies presented in the literature suggest that as the proficiency level of learners increases so do the linguistic and discursive features. However, how these features vary across levels is of the under-explored areas of research and raises important questions that need to be addressed. Hence, in order to gain insight into the nature of learner language across proficiency levels, there is a need for empirical studies revealing more about the linguistic and discursive features of learner language based on the CEFR levels. The aim of the current research is to fill the aforementioned literature gap to shed light on learner language on specific CEFR levels.

CHAPTER 3

3. METHODOLOGY

3.1. INTRODUCTION

The present section presents the methodology of the study. Firstly, the chapter describes the data used in this research. Following this, it elaborates on the discursive features chosen for the research. Moreover, the methods used and procedures of analyzing the data are presented.

3.2. RESEARCH DESIGN

As it is indicated in the previous chapters, there is a plethora of research focusing on linguistic and discursive features of learner language, yet the link between the CEFR and these features is relatively scarce. Hence, there is a need for research studies investigating the relationship between the CEFR and linguistic and discursive features to gain better insights into learner language at different levels. Correspondingly, the primary aim of the present research is twofold: first to examine the discursive features in a spoken learner corpus both within and across CEFR levels, second, to assess syntactic complexity in spoken learner corpus both within and across CEFR levels. More specifically, the current research addressed the following research questions:

1. Is there consistency within the same CEFR level distribution in a spoken learner corpus in terms of syntactic complexity?
2. Is there a significant difference among CEFR level distributions in a spoken learner corpus in terms of syntactic complexity?
3. Is there consistency within the same CEFR level distribution in a spoken learner corpus in terms of discursive features?

4. Is there consistency among CEFR level distributions in a spoken learner corpus in terms of discursive features?

By addressing the abovementioned research questions, the present research aims to present a bulky description of discursive and linguistic features used in learner language. Because the data obtained from the BACKBONE corpus does not have a representative nature, the thesis does not claim the generalizability of the results. Rather, it aims to cast light on discursive features and syntactic complexity of learner language across CEFR levels. A more thorough description of how the present research was carried out is presented in the following sections.

3.3. THE CORPUS

The central data source utilized in this research were the BACKBONE corpus which was a project developed by Professor Kurt Kohn in the European LLP/Languages project funded with support from the European Commission. The educational content needs and the challenges language teachers encounter were addressed in the project with reference to integrating CLIL and e-learning into pedagogy. The BACKBONE corpus consists of sub-corpora containing video interviews with speakers from different L1 backgrounds. Lesser taught languages such as Polish and Turkish were prioritized in the corpora in addition to more frequently used and taught ones such as English, German, French, and Spanish, as well as English as a lingua franca (ELF).

The data chosen for this research were gathered from the ELF section of the corpus. The data consist of 50 interviews in total conducted with 10 non-native speakers of English from five different language backgrounds including Polish, Turkish, German, French, and Spanish with an average length of 10 minutes per interview. The interviews had been annotated according to several categories referring to thematic, grammatical, lexical, and textual characteristics in addition to CEFR levels. The data utilized in this research were selected based on the stated CEFR levels on the website. In the ELF section of the corpus, there are 317 texts gathered from the speakers of different

backgrounds and walks of life. Of these 317 texts, eight of them are A1, 7 of them are A2, 30 of them are A2-B1, 121 of them are B1, 30 of them are B1-B2, 46 of them are B2, one of them are B2-C1, and six of them are C1 level texts. The difference in the sample distribution of the CEFR levels stems from the corpus used in the research. Due to the heterogeneity in the level distribution as well as other factors, the non-parametric analysis was implemented in the quantitative data analysis.

The interviews were conducted between two people and have "a native narrative character" (Kohn, 2012: p. 6). Kohn (2012: p. 6) stated that though coming from different professional backgrounds, the participants included in ELF interviews were all used to speaking English in their professional environments. The questions asked to the participants do not necessarily require a conversational interaction, rather primarily, they aim to get the interviewee to relax and speak. In that sense, the main purpose of the questions is to promote longer descriptions, explanations, and opinions rather than conversational interaction (Kohn, 2012: p. 6). Both the interviewer and the interviewee were physically present when being recorded. A wide variety of topics were included in the interviews including education, health, culture, business, urban and rural life, social issues, health, and social security, environment and government and politics.

As the primary aim of the corpus is to be utilized in learning contexts, the transcription of the interviews used orthographical notation in accordance with the nature of spoken discourse. A number of pre-defined mark-up codes had been included in the transcriptions including breaks or comments, etc. Other discourse elements such as fillers, repetitions, and hesitation phenomena had only been included if they were found meaningful by the annotator-teacher.

3.4. DISCURSIVE FEATURES

One of the categories in the BACKBONE Corpus applied to the texts gathered from English speakers was discourse categorized under the title of interpreting challenges. The discursive features chosen in this research were selected from the ones

that were already included in another section of the corpus. Those elements included connectors, discourse markers, high information load, reference to context, weak coherence, and backward reference. Other than those elements, when coding the data, several code-switching elements were also detected, thus, code-switching was also added in the discursive features in the research.

3.4.1. Connectors

The bulk of the analysis in the present research relies on the connectors as a discursive feature of the language. The present research is concerned with the cohesive devices that are included in Halliday and Hasan's (1976: p. 14) definitions of conjunctions referring to linkers that connect ideas. Some examples to the connectors that were coded in the data are *and, but, or, so,* etc. An example of which conjunctions were coded can be seen in Excerpt 1.

Excerpt 1

Connectors in the text B1.32

J: It's rather quite good life. People shown in these films ***or*** in these TV shows usually don't work a lot they can afford everything ***but*** their work is not shown in the shows they have wife successful beautiful ***and*** the clever children ***and*** they probably vote Republicans ***and*** they have for sure at least one pet ***and*** it's nice calm type of life with no bigger problems with no troubles.

3.4.2. Discourse Markers

In order to begin coding DMs across the CEFR levels, first it was essential to determine which DMs within the corpus are acting as DMs and which ones are the words that were being used in their literal meaning. An adapted version of the criteria of characteristic features of DMs as included in a broad description of DMs in pedagogic settings by Fung and Carter (2007: p. 412-415) were employed in the research. Some examples of the DMs coded in the data are *right, well, I mean, sort of, kind of, so, I*

think, I don't know and others. An example of which DMs were coded can be seen in Excerpt 2.

Excerpt 2

Discourse markers in the text A2.15

Yes, of course this is why the government takes one third of my income every month and I believe that since I have paid enormous amounts of money they should ***you know*** at least invest in my own health because I am their source of income as each ***you know*** working person so I believe that ***yeah*** they should ***you know*** provide us with free health services.

3.4.3. High Information Load

One other discursive feature included in this research is the high information load observed in learner language. This feature mostly relates to the coherence of the texts. This feature was coded when learner speech included too long utterances, or the overall meaning was lost in the sentences due to the length of conservation.

3.4.4. Reference to Context

Reference is a subcategory of cohesive resources defined by Halliday and Hasan (1976: p. 14). One observable reference resource is the use of adverbs *here* and *there* which refer to the context of the conversation. An example of how reference to context was coded can be seen in Excerpt 3.

Excerpt 3

Reference to context items in the text A1. 5

My name is Dieter Müller, I am working for Festo, a producer of pneumatics, in the area of the IT department and ***there*** I was — ***there*** I'm responsible for the whole IT and the whole business process things. The other part of the questions was the education. I was study informatics a long time ago and I use all this information on my daily job and I think it's very interested to do this in this area.

3.4.5. Backward Reference

As mentioned above, reference is a subcategory of cohesive resources. There are three types of references as included in Halliday and Hasan (1976: p. 14): anaphoric, cataphoric, and exophoric reference. In the BACKBONE corpus, backward reference had been characterized including the phrases such as *as I said before, as I mentioned,* etc. referring back to the ideas introduced by the speakers. Consequently, the frequency of these structures across the CEFR levels was also investigated. An example of how backward reference was coded can be seen in Excerpt 4.

Excerpt 4

Backward reference in the text B2.3.

Joaquín**:** Well, ***as I said before*** I am a doctor, but I only practiced medicine for about a year when I finished Medicine. Then I went into the university to pursue my academic career and then since that time I have no clinical practice…

3.4.6. Weak Coherence

Another discursive feature investigated in the spoken learner corpus is coherence. As recently the concept has been characterized as building a mental representation of the information, weak coherence was coded in the texts where no mental representation of the text occurred. In other words, if the text was viewed inapprehensible and did not make any sense, weak coherence was detected.

3.4.7. Code-switching

Although code-switching had not originally been included in the BACKBONE corpus, several occurrences of code-switching in the learners' texts were detected. Hence, how CEFR levels differ in their use of code-switching was also investigated. An example of how code-switching was coded can be seen in Excerpt 5.

Excerpt 5

Code-switching in the text A2.14

Kunduzhan: And then I can continue at the university. I would like to go to school of the — ***intrepreture*** or to — or just to do LEA, the language ***les langues étrangères appliquées, les*** […]
Bernard: Yes, applied linguistics, yes.

3.5. SYNTACTIC COMPLEXITY MEASURES

Research on the measurement of L2 SC has concentrated on various sets of indices (Norris & Ortega (2009): p. 556). Traditionally, a bulk of research on SC was based on human judgment. However, the advent of technology allowed for computational tools to be developed and utilized in SC research (e.g. Coh-Metrix: Graesser et al., 2004; L2SCA: Lu, 2010). All the SC measures used in this research were provided by Coh-Metrix. Coh-Metrix is an advanced computational tool providing over 600 language and discourse measures related to cohesion, lexical difficulty, SC, and so on (see Appendix B for Coh-Metrix output). Coh-Metrix reports on different indices of SC such as left embeddedness which calculates the mean number of words before the main verb, the average number of modifiers per noun phrase, minimal edit distance, and sentence syntax similarity (see McNamara et al., 2014 for the theoretical basis and definitions for measures). These four indices of SC were included in this research in order to provide answers to the research questions related to SC.

3.6. QUALITY CRITERIA

As Dörnyei (2007: p.54) includes, it is easier to define quality criteria in quantitative research than qualitative research as the findings of a qualitative inquiry are relative and dependent upon individual discernment. To comply with the stated reasons,

qualitative research requires its own procedures in providing validity and reliability (Dörnyei, 2007: p.54). Lincoln and Guba (1985: p. 301-327) characterized the criteria of validity and reliability in terms of dependability, credibility, transferability, and confirmability.

Because the data in this research are text-based and software is used as an analysis tool, in lieu of reliability, a dependability check was conducted. Dependability refers to the consistency between the data and results gained from the data are dependable (Lincoln & Guba, 1985; as cited in Merriam, 2009: p. 221). Researcher triangulation is one technique used to fulfill the dependability criteria to achieve 'reliability'. In that vein, a colleague with an MA degree was asked to process the data on Coh-Metrix and code randomly selected 50 texts on NVivo to fulfill the dependability criterion as an external reviewer. To increase representativeness, the data were selected from all levels. Before she started uploading the texts on Coh-Metrix, she was provided with the necessary information on how the software works and which indices she should look at. Also, before she started coding the texts on NVivo, each discourse feature was explained to her thoroughly and she was explained how each one was represented in the literature. When she finished the analysis on both tools based on the aims set, we compared and contrasted the gained outputs. As a result of this process, an inquiry audit was conducted, and in the event of a discrepancy, either one of us revised the coding through discussion until no difference was found in the outputs.

As a counterpart of internal validity, credibility (Bryman, 2012: p. 390) addresses the "fit" between respondents' views and the researcher's representation of them (Tobin & Begley, 2004: p. 391). Credibility deals with the matter of whether the explanations provided in research corresponds to the descriptions (Janesick, 2000: p. 379-400). The credibility criterion was fulfilled by providing clear descriptions in the methodology section in quest of providing a thorough description of the subject matter. Also, examples of transcripts were cited and used in the methodology as well as in the appendix (see Appendix A) to ameliorate to clarify the findings. Furthermore, peer debriefing was also utilized especially in the methodology of the research with the help

of the thesis advisor to ensure that the process of research is "logical, traceable and clearly documented" (Schwandt, 2001: p. 230).

Transferability refers to the generalizability of the research. However, as qualitative research is highly dependent on the researcher and the context, Tobin and Begley (2004: p. 392) assert that transferability is only concerned with the case-to-case transfer. The researcher should be aware of the fact that they cannot estimate the context in which findings are desired to be transferred and there is no one cure-for-all interpretation in qualitative research. Thick descriptions should be provided for the readers (Nowell et al., 2017: p. 3) to meet this challenge so that those desiring to transfer the findings would be able to judge transferability. To this end, clear descriptions of the data, data analysis, and results as well as example data were provided.

Confirmability is regarded with reporting the investigation process in a thorough and clear manner by researcher's providing a clear rationale for how they reached conclusions and interpretations (Nowell et al., 2017: p. 3). Guba and Lincoln (1989) assert that the researcher can ensure confirmability when credibility, transferability, and dependability are all achieved. Achieving the mentioned criteria as mentioned above were attempted to be achieved. In addition, audit trail technique was utilized in every phase of the research. Put differently, all the phases of implementation and data analysis were noted down in a systematic and precise manner. Also, examples from the corpus were provided both in the methodology and the results section with an aim of enabling other researchers to follow similar procedures.

3.7. DATA ANALYSIS

Qualitative content analysis, qualitative analysis with an automated tool, and quantitative analysis are the main methods used in the present research. These methods helped to investigate syntactic complexity and discursive features in the BACKBONE corpus. Before starting the manual coding process, in order to validate the transcriptions of the texts, 50 texts were randomly selected, and the researcher watched the related

video recordings to observe any misconduct of the transcriptions. Révész (2012: p. 207) also suggests that at least part of the data be checked second time in qualitative research, to decrease and control the errors in transcribing, and to verify transcriber reliability. Thus, a colleague with an MA degree was asked to check part of the transcriptions.

Corpora are usually 'cleaned' to prevent any distortion of the results. Dowell, Graesser, & Cai (2016: p. 85) included that regardless of the collection of corpora, whether collected by the researcher or borrowed from other research, etc., each corpus has 'dirtiness', each of which has the potential to distort the validity of Coh-Metrix analyses. Since in this research, both Coh-Metrix and NVivo were used as research tools, the cleaning process started before both analyses. Even though there are no guidelines are existing to clean the corpora, two practices were included by Dowell, Graesser, & Cai (2016: p.86):

"1. If there is no good reason to take it out, the researcher should leave it in.

2. What the researcher does to one text should be done to all."

Similarly, McNamara, Graesser, McCarthy, & Cai (2014: p. 156) assert that when a researcher corrects something, it should be conducted in a systematic manner. After the effect of possible problematic issues was examined, minor spelling mistakes were corrected, and missing sentence punctuation marks were added. Apart from these mistakes, no other corrections were made.

After the validation of the transcripts and cleaning process, the texts were filed according to the CEFR levels stated on the corpus website. A number of texts had been categorized into two levels at the same time by the annotator of the corpus. With the purpose of obtaining more comprehensible justification of the data, texts categorized under the levels as A2-B1, B1-B2, and B2-C1 were included in one upper level as the annotator already included them in there as well. That is, if a text was included in both A2 and B1 levels, they were coded in the B1 level. In the end, the corpus consisted of eight A1 level texts, 72 A2 level texts, 151 B1 level texts, 76 B2 level texts, and seven C1 level texts. To help with interpreting and discussing the results; word count, sentence length, and sentence counts of each text were also gathered from the Coh-Metrix. The

data were then uploaded to IBM SPSS 22.0 and descriptive statistics were obtained. The calculated means of word count, sentence length, and sentence count can be seen in table 1.

Table 1. *Calculated means of word count, sentence count, and sentence length in the CEFR levels of the corpus*

	Word count	Sentence count	Sentence length
A1	78.37	4	23.59
A2	118.43	8.31	15.69
B1	172.38	12.33	17.16
B2	251.23	13.51	20.64
C1	387.57	18.42	21.64

3.7.1. Quantitative Data Analysis

To investigate any possible significant differences within and among CEFR levels, the outputs obtained from the Coh-Metrix analysis were uploaded on IBM SPSS 22.0. A number of preliminary steps were achieved to obtain results. The first step was coding the data on IBM SPSS software, in which the variables were written on the software. Since the CEFR level is a categorical variable, the type of the variable was chosen as a string variable to be able to label the levels with their own names rather than giving numbers. The other variables namely the left embeddedness, modifiers per noun phrase, minimal edit distance, and sentence syntax similarity were coded in numeric format. In the second step, each output of each text was uploaded on the software and was categorized according to the level. The entries were double-checked by a colleague with more experience in SPSS applications. Subsequently, inferential statistics were elicited from the data to ascertain any statistically significant differences within and among levels. To begin with, Kolmogorov-Smirnov and Shapiro-Wilk were used to test the normal distribution of data with the aim of determining the statistical method to be used. Below is the table of the obtained outcome for the items. (Table 2)

Table 2. *The obtained outcome of Test of Normality for the items*

		Shapiro-Wilk		
	CEFR	Statistic	df	Sig.
Left Embeddedness	A1	.72	8	.005
	A2	.75	72	.000
	B1	.35	151	.000
	B2	.70	76	.000
	C1	.89	7	.279
Modifiers per NP	A1	.83	8	.065
	A2	.99	72	.823
	B1	.99	151	.905
	B2	.99	76	.821
	C1	.93	7	.578
Minimal Edit Distance	A1	.51	8	.000
	A2	.53	72	.000
	B1	.54	151	.000
	B2	.92	76	.000
	C1	.94	7	.641
Sentence Syntax Simil.	A1	.86	8	.120
	A2	.89	72	.000
	B1	.95	151	.000
	B2	.93	76	.001
	C1	.99	7	.998

As clearly seen in Table 2, the normality tests provided a ranging value of significance, mostly indicating that there is no normal distribution in the data. As a result, nonparametric tests were employed to investigate any significant differences. Based on this finding, the nonparametric equivalent of one-way MANOVA, Kruskal-Wallis tests were employed for the comparison of groups. The tests were performed on each SC measure separately for each level. After the Kruskal-Wallis tests, post-hoc Mann-Whitney U tests were performed to examine the group differences. The results provided answers to the second research question. To answer the first research question, the data were coded on IBM SPSS 22.0. Shapiro-Wilk normality tests were administered to investigate how the data were distributed within each level. However, normality tests only provide superficial normality results and can be misleading. To this end, other descriptive statistics such as standard deviation, mean, etc. should be taken into consideration. Therefore, descriptive statistics on each level were employed in order to investigate how the data were distributed within the level.

3.7.2. Qualitative Data Analysis

To seek answers for the third and fourth research questions, the qualitative coding process started. Unlike quantitative analysis, qualitative analysis entails "non-numerical examination and interpretation of observations, for the purpose of discovering underlying meanings and patterns of relationships" (Babbie, 2007: p. 379). Because it involves interpretation, qualitative analysis is quite subjective and requires interrogative and investigative skills. This subjective process can be conducted with qualitative content analysis, a research method involving individual discernment of the content with the help of coding and determining themes or patterns (Hsieh & Shannon, 2005: p. 1278). In the present research, qualitative content analysis was employed on the transcripts taken from the corpus website.

Baralt (2012: p.222) stated that SLA researchers are using "computer-assisted qualitative data analysis software (CAQDAS)" incrementally to carry out a qualitative inquiry as they provide several advantages such as facilitated data management and the ability to deal with all data in the same workspace. NVivo is one example of CAQDAS which is widely used in second language research. Keeping the stated benefits in mind, the qualitative analysis tool, NVivo, was used in this process. To begin with, the texts were uploaded on NVivo separately according to the levels. Ellis and Barkhuizen (2005: p. 253) assert that coding enables researchers to interpret the data and draw conclusions by grouping "thick, rich, and deep qualitative data" into themes. In that vein, the texts in each level were coded with the discursive features of connectors, discourse markers, high information load, reference to context, weak coherence, and backward reference, as well as code-switching. In particular situations, the researcher encountered challenges such as not being certain of which code to use. To exemplify, the word *so* could be used both as a connector or as a discourse marker depending on the context. In such situations, the researcher watched the related video again to make sure of the context of the transcription. Baralt (2012: p.223) states that some aspects of the qualitative data can be quantitative. In that vein, the qualitative analysis of the data in this research provided quantitative results as the number of instances was divided by the sample size at each level to better understand the distribution of features within and across levels. In

addition, the output gathered from Coh-Metrix also provided quantitative results as it was used in the IBM SPSS analysis.

After the manual coding of the data, the texts were separately filed for each level. Each file was named according to the level of the text and the order of appearance on the corpus website. That is, if a text was categorized in B1 level, and that text was the third text appearing on B1 level, it was named as B1.3. Each text was uploaded on the Coh-Metrix 3.0 website separately. The output of each text was downloaded and saved in folders separated by the levels. For the purpose of the current research, the output related to SC for each text was separated and copied on an Excel file.

3.8. CONCLUSION

Table 3. *Summary of the methodology used in the research*

Language to be analyzed	Indices/Features	Within/Across the Levels	Data Analysis
Syntactic Complexity	Left embeddedness, number of modifiers per noun phrase, minimal edit distance, sentence syntax similarity	Within	Coh-Metrix, SPSS (normality tests and descriptive statistics)
		Across	Coh-Metrix, SPSS (Kruskal-Wallis test and Mann-Whitney U test)
Discursive Features	Backward reference, code switching, connectors, discourse markers, high information load, reference to context, weak coherence	Within	NVivo, SPSS (normality tests)
		Across	NVivo (the number of instances), SPSS (the calculated means of instances)

Table 3 summarizes the methodology used in the current research. In this chapter, the methodological issues of the present research are presented at length. First, the research design along with the research questions are displayed. Following this, the corpus from which the data were amassed is introduced as well as the syntactic complexity measures and discursive elements. Then, quality criteria are explained in detail. Finally, the chapter is concluded by demonstrating the data analysis procedures.

CHAPTER 4

4. FINDINGS

4.1. INTRODUCTION

This chapter displays the findings of the data analysis employed to scrutinize the syntactic complexity and discursive features in a spoken learner corpus. The findings of the current research that emerged from the qualitative analyses are reported in four sections to answer each research question, respectively. To begin with, the distribution of SC measures within levels is presented with the help of descriptive statistics as well as the test of normality results. Following this, how SC measures vary across levels is displayed with the help of inferential statistics. These are followed by an in-depth analysis of discursive features in learner language within levels. Finally, an analysis of discursive features across the CEFR levels is presented.

4.2. SYNTACTIC COMPLEXITY

4.2.1. Syntactic Complexity Measures within Levels

The first research question was asked aiming to investigate how the data were distributed within the CEFR levels. The following section will present the results of the quantitative analysis presenting the test of normality results along with displaying the results separately on each SC measure.

4.2.1.1. A1 Level Results

Table 4. *Test of normality results of syntactic complexity measures in the A1 level*

		Shapiro-Wilk		
	CEFR	Statistic	df	Sig.
Left embeddedness	A1	.72	8	.00
Modifiers per noun phrase	A1	.83	8	.06
Minimal edit distance	A1	.51	8	.00
Sentence syntax similarity	A1	.86	8	.12

Table 4 demonstrates the results of the test of normality on syntactic complexity measures at the A1 level. According to the results of the test, the only normal distribution can be seen in modifiers per noun phrase and sentence syntax similarity indices. No normal distribution of the other indices was observed at the A1 level ($p < .05$).

Table 5. *Descriptive statistics of syntactic complexity measures in the A1 level*

INDEX	Min	Max	*M*	*SD*	Skewness		Kurtosis	
					Statistic	Std. Error	Statistic	Std. Error
LEFT EMB	1.80	9.00	3.94	2.76	1.41	.79	.61	1.58
MOD NP	.73	1.00	.81	.08	1.74	.75	3.80	1.48
MED	.00	.93	.76	.31	-2.76	.75	7.74	1.48
SEN SYN SML	.00	.26	.09	.07	1.42	.75	3.19	1.48

Table 5 demonstrates the descriptive statistics of the syntactic complexity measures within the A1 level. The scores of left embeddedness of A1 level texts ranged from 1.80 to 9.00 (*M*= 3.94, *SD* = 2.76). The left embeddedness scores of A1 texts were non-normally distributed, with right skewness of 1.41 (*SE* = .79) and kurtosis of .615 (*SE* = 1.59). Furthermore, the scores of the modifiers per noun phrase of A1 level texts ranged from .73 to 1.00 (*M* = .82, *SD* = .083). Though the normality test results revealed there was a normal distribution of the data, skewness (1.75, *SE* = .75) and kurtosis (3.80, *SE* = 1.49) revealed that there was not a normal distribution in the modifiers per noun phrase use in A1 level texts. In terms of the minimal edit distance index, the scores in A1 level texts ranged from .00 to .93 (*M* = .76, *SD* = .31). Skewness (-2.77, *SE* = .75) and kurtosis (7.74, *SE* = 1.48) results revealed that the data regarding minimal edit distance within A1 level were not normally distributed. Finally, the sentence syntax similarity average in A1 level texts was .09 (*SD* = .080). The scores ranged from .00 to

.26 and were non-normally distributed according to the skewness of 1.43 (*SE* = .75) and kurtosis of 3.20 (*SE* = 1.49) although the normality test results revealed the opposite.

4.2.1.2. A2 Level Results

Table 6. *Test of normality results of syntactic complexity measures in the A2 level*

		Shapiro-Wilk		
	CEFR	Statistic	df	Sig.
Left embeddedness	A2	.75	72	.00
Modifiers per noun phrase	A2	.99	72	.82
Minimal edit distance	A2	.53	72	.00
Sentence syntax similarity	A2	.89	72	.00

Table 6 demonstrates the results of the test of normality on syntactic complexity measures at the A2 level. The results of the test revealed that the only normal distribution can be seen in the modifiers per noun phrase measure. No normal distribution of the other indices was observed at the A2 level ($p < .05$).

Table 7. *Descriptive statistics of syntactic complexity measures in the A2 level*

INDEX					Skewness		Kurtosis	
	Min	Max	*M*	*SD*	Statistic	Std. Error	Statistic	Std. Error
LEFT EMB	.67	10.50	2.29	1.41	3.05	.28	15.16	.55
MOD NP	.11	1.35	.65	.22	.24	.28	.24	.55
MED	.00	1.00	.86	.11	-5.55	.28	40.37	.55
SEN SYN SML	.00	.30	.10	.04	1.43	.28	4.95	.55

Table 7 demonstrates the descriptive statistics of the syntactic complexity measures within the A2 level. The average score of left embeddedness of A2 level texts was 2.30 (*SD* = 1.41). The scores ranged from .67 to 10.50 and were non-normally distributed, with skewness of 3.05 (*SE* = .28) and kurtosis of 15.16 (*SE* = .56). Moreover, the average score of modifiers used per noun phrase was .66 in A2 level texts (*SD* = .23). The scores ranged from .11 to 1.35 and were normally distributed with

skewness of .24 (*SE* = .28) and kurtosis of .24 (*SE* = .56) as also the normality test results suggested. For A2 level texts, the mean score of minimal edit distance was .87 (*SD* = .12), and the scores ranged from .00 to 1.00. The data were non-normally distributed with skewness of -5.56 (*SE* = .28) and kurtosis of 40.38 (*SE* = .56). Finally, sentence syntax similarity output gathered from the Coh-Metrix analysis ranged from .00 to .30 (*M* = .10, *SD* = .05) in A2 level texts. The data were not normally distributed with skewness of 1.43 (*SE* = .28) and kurtosis of 4.96 (*SE* = .56).

4.2.1.3. B1 Level Results

Table 8. *Test of normality results of syntactic complexity measures in the B1 level*

		Shapiro-Wilk		
	CEFR	Statistic	df	Sig.
Left embeddedness	B1	.35	151	.00
Modifiers per noun phrase	B1	.99	151	.90
Minimal edit distance	B1	.54	151	.00
Sentence syntax similarity	B1	.95	151	.00

Table 8 demonstrates the results of the test of normality on syntactic complexity measures at the B1 level. The results of the test revealed that the only normal distribution can be seen in the modifiers per noun phrase measure. No normal distribution of the other indices was observed at the B1 level ($p < .05$).

Table 9. *Descriptive statistics of syntactic complexity measures in the B1 level*

INDEX					Skewness		Kurtosis	
	Min	Max	*M*	*SD*	Statistic	Std. Error	Statistic	Std. Error
LEFT EMB	.33	57.00	3.46	5.10	8.22	.19	82.29	.39
MOD NP	.18	1.03	.59	.16	.04	.19	-.13	.39
MED	.00	.98	.87	.08	-6.88	.19	68.16	.39
SEN SYN SML	.00	.29	.09	.04	.91	.19	3.13	.39

Table 9 illustrates the descriptive statistics of the syntactic complexity measures within the B1 level. For B1 level texts, the average score of left embeddedness was 3.46, and the scores ranged from .33 to 57.00. The data were not normally distributed, with skewness of 8.23 (*SE* = .19) and kurtosis of 82.30 (*SE* = .40). Moreover, B1 level text scores of modifiers per noun phrase ranged from .18 to 1.03 (*M* = .59, *SD* = .16). The data were normally distributed with skewness of .04 (*SE* = .20) and kurtosis of -.13 (*SE* = .39). The minimal edit distance scores of B1 level texts ranged from .00 to .98 (*M* = .87, *SD* = .08). Similar to the previous levels, the data were not normally distributed with skewness of -6.88 (*SE* = .20) and kurtosis of 68.169 (*SE* = .39). The average use of sentence syntax similarity output was .10 (*SD* = .04) in B1 level texts. The scores ranged from .00 to .29 and were not normally distributed with skewness of .92 (*SE* = .20) and kurtosis of 3.12 (*SE* = .39).

4.2.1.4. B2 Level Results

Table 10. *Test of normality results of syntactic complexity measures in the B2 level*

		Shapiro-Wilk		
	CEFR	Statistic	df	Sig.
Left embeddedness	B2	.70	76	.00
Modifiers per noun phrase	B2	.99	76	.82
Minimal edit distance	B2	.92	76	.00
Sentence syntax similarity	B2	.93	76	.00

Table 10 demonstrates the results of the test of normality on syntactic complexity measures at the B2 level. The results of the test revealed that the only normal distribution can be seen in the modifiers per noun phrase measure. The other measures did not signal a normal distribution at the B2 level ($p < .05$).

Table 11. *Descriptive statistics of syntactic complexity measures in the B2 level*

INDEX	Min	Max	*M*	*SD*	Skewness		Kurtosis	
					Statistic	Std. Error	Statistic	Std. Error
LEFT EMB	.83	21.50	3.92	3.01	3.3	.27	15.66	.54
MOD NP	.13	1.10	.61	.17	.13	.27	.62	.54
MED	.76	.95	.88	.03	-1.14	.27	1.79	.54
SEN SYN SML	.01	.20	.07	.02	1.05	.27	4.12	.54

Table 11 demonstrates the descriptive statistics of the syntactic complexity measures within the B2 level. B2 level left embeddedness scores ranged from .83 to 21.50 (*M* = 3.92, *SD* = 3.01). Skewness (3.31, *SE* = .28) and Kurtosis (15.66, *SE* = .54) results revealed that the data were not normally distributed. The average score of modifiers per noun phrase in B2 level texts was .61 (*SD* = .18). The scores ranged from .13 to 1.10, and the data were normally distributed with skewness of .134 (*SE* = .28) and kurtosis of .62 (*SE* = .54). In addition, the average score of minimal edit distance was .89 (*SD* = .04) for B2 texts, and according to the skewness (-1.14, *SE* = .27) and kurtosis (1.79, *SE* = .54) results, there was no normal distribution of data of minimal edit distance. Finally, B2 level text scores of sentence syntax similarity ranged from .01 to .20 (*M* = .08, *SD* = .03). There was no normal distribution of data according to the results of skewness of 1.06 (*SE* = .28) and kurtosis of 4.12 (*SE* = .54).

4.2.1.5. C1 Level Results

Table 12. *Test of normality results of syntactic complexity measures in the C1 level*

	CEFR	Shapiro-Wilk		
		Statistic	df	Sig.
Left embeddedness	C1	.89	7	.27
Modifiers per noun phrase	C1	.93	7	.57
Minimal edit distance	C1	.94	7	.64
Sentence syntax similarity	C1	.99	7	.99

Table 12 demonstrates the results of the test of normality on syntactic complexity measures at the C1 level. The results of the test revealed that all the measures in the C1 learner language were normally distributed ($p > .05$).

Table 13. *Descriptive statistics of syntactic complexity measures in the C1 level*

INDEX	Min	Max	*M*	*SD*	Skewness		Kurtosis	
					Statistic	Std. Error	Statistic	Std. Error
LEFT EMB	1.04	5.00	2.70	1.57	.44	.79	-1.84	1.58
MOD NP	.41	.80	.60	.14	.25	.79	-1.20	1.58
MED	.83	.91	.87	.03	.03	.79	-1.31	1.58
SEN SYN SML	.06	.09	.07	.01	.07	.79	.19	1.58

Table 13 demonstrates the descriptive statistics of the syntactic complexity measures within the C1 level. The mean score of left embeddedness in C1 texts was 2.71 (*SD* = 1.57). The scores raged from 1.04 to 5.00 (*SD* = 1.57), and Skewness (.44, *SE* = .79) and Kurtosis (-1.84, *SE* = 1.6) results revealed that the data were normally distributed though the kurtosis is high. The scores of modifiers per noun phrase in C1 level texts ranged from .41 to .80 (*M* = .61, *SD* = .15). The data were normally distributed with skewness of .25 (*SE* = 79) and kurtosis of -1.20 (*SE* = 1.59). The scores of minimal edit distance in C1 level texts ranged from .83 to .91 (*M* = .88, *SD* = .03). The data were normally distributed with skewness of .03 (*SE* = .79) and kurtosis of -1.31 (*SE* = 1.59). Sentence syntax similarity output from Coh-Metrix analysis for C1 level texts ranged from .06 to .09 (M = .07, SD = .01). Skewness (.07, *SE* = .79) and kurtosis (.19, *SE* = 1.59) results revealed that the data were normally distributed.

4.2.2. Syntactic Complexity Measures across Levels

The second research question was asked aiming to investigate any significant differences across the CEFR levels. The following section will present the results of the quantitative analysis separately on each SC measure.

4.2.2.1. Left Embeddedness

Table 14. *The Kruskal-Wallis Test results of the use of Left embeddedness across CEFR levels*

	Chi-Square	df	Asymp. Sig.
Left Embeddedness	24.56	4	0.00

As table 14 demonstrates, a Kruskal-Wallis test showed that there is a significant difference across levels in terms of the use of left embeddedness, $H(4) = 24.56$, $p < 0.05$. The use of left embeddedness was found to be the most in B2 level texts (*Mdn* = 3.22, *M* = 3.92) compared to A1 (*Mdn* = 2.58, *M* = 3.66), A2 (*Mdn* = 1.88, *M* = 2.29), B1 (*Mdn* = 2.30, *M* = 3.46), and C1 level texts (*Mdn* = 2.00, *M* = 2.70). Post-hoc Mann-Whitney tests indicated that the difference between A1 and A2 level texts was not statistically significantly different in terms of the use of left embeddedness, $U(N_{A1}=8, N_{A2}=72,) = 183.50$, $z = -1.67$, $p > .05$ as can be seen in table 15. The table also demonstrates that when A1 and B1 level texts were compared using Mann-Whitney post-hoc tests, not a statistically significant difference was found in the number of words they use before the main verb, $U(N_{A1}=8, N_{B1}=151,) = 504.50$, $z = -.78$, $p > .05$. Furthermore, the post-hoc also revealed that there was not a statistically significant difference between A1 and B2 level texts in the use of left embeddedness, $U(N_{A1}=8, N_{B2}=76,) = 265.00$, $z = -.59$, $p > .05$. Finally, A1 level texts were compared to C1 level tests to check any statistically significant difference in terms of the number of words before the main verb, and not a significant difference was found $U(N_{A1}=8, N_{C1}=7,) =$

20.00, z = -.92, p > .05. All in all, the post-hoc results revealed that A1 levels do not differ from other CEFR levels in their use of left embeddedness.

Table 15. *Mann-Whitney Post-hoc results between A1 and other CEFR levels in terms of left embeddedness*

		CEFR Level A1	
CEFR Level	Mean rank	Z-value	*p*
A2	183.50	-1.67	.09
B1	504.50	-.78	.43
B2	265.00	-.59	.5
C1	20.00	-.92	.35

Mann-Whitney post-hoc tests were also performed on A2 level texts comparing them to the other CEFR levels. As table 15 illustrates, not a statistically significant difference was observed between A1 and A2 level texts. A2 level texts were also compared to B1 level texts, and as the table 16 illustrates, the results revealed that there was a statistically significant difference between A2 and B1 level texts, $U(N_{A2}=72, N_{B1}=151,) = 4265.00$, z = -2.60, p < .05. Furthermore, a statistically significant difference was also observed between A2 and B2 level texts, $U(N_{A2}=72, N_{B2}=76,) = 1452.50$, z = -4.92, p < .05. Finally, A2 level texts were compared with C1 level texts, and the results indicated that A2 levels do not significantly differ from C1 levels in their use of left embeddedness, $U(N_{A2}=72, N_{C1}=7,) = 215.50$, z = -.63, p > .05.

Table 16. *Mann-Whitney Post-hoc results between A2 and other CEFR levels in terms of left embeddedness*

		CEFR Level A2	
CEFR Level	Mean rank	Z-value	*p*
A1	183.50	-1.67	.09
B1	4265.00	-2.60	.00
B2	1452.50	-4.92	.00
C1	215.50	-.63	.52

B1 level texts were also compared to the other CEFR levels, and as presented above, though there was no significant difference between A1 and B1 levels, a statistically significant difference was observed between A2 and B1 level texts in their use of left embeddedness. Similarly, table 17 illustrates that B1 level texts significantly differ from B2 level texts with regard to their use of left embeddedness, $U(N_{B1}=151, N_{B2}=76,) = 4329.00$, $z = -3.01$, $p < .05$. Finally, B1 level texts were compared to C1 level texts, and no statistically significant difference was observed, $U(N_{B1}=151, N_{C1}=7,) = 486.50$, $z = -.35$, $p > .05$.

Table 17. *Mann-Whitney Post-hoc results between B1 and other CEFR levels in terms of left embeddedness*

		CEFR Level B1	
CEFR Level	Mean rank	Z-value	*p*
A1	504.50	-.78	.43
A2	4265.00	-2.60	.00
B2	4329.00	-3.01	.00
C1	486.50	-.35	.72

Finally, post-hoc Mann-Whitney test was performed on B2 level texts comparing them to the other CEFR levels. As mentioned above, B2 level texts significantly differ from A2 and B1 level texts in their use of left embeddedness. However, no statistically significant difference was observed between A1 and B2. Similarly, when B2 level texts were compared to C1 level texts, no statistically significant difference was found, $U(N_{B2}=76, N_{C1}=7,) = 193.00$, $z = -.1.19$, $p > .05$, as shown in table 18.

Table 18. *Mann-Whitney Post-hoc results between B2 and other CEFR Levels in terms of left embeddedness*

	CEFR Level B2		
CEFR Level	Mean rank	Z-value	*p*
A1	265.00	-.59	.55
A2	1452.50	-4.92	.00
B1	4329.00	-3.01	.00
C1	193.00	-1.19	.23

4.2.2.2. The Number of Modifiers per Noun Phrase

As shown in table 19, a Kruskal-Wallis test showed that the number of modifiers used per noun phrase significantly differs across the CEFR levels, $H(4) = 16.01$, $p = .003$. The number of modifiers used per noun phrase was observed to be the most in A1 level texts (*Mdn* = .79, *M* = .81) compared to A2 (*Mdn* = .63, *M* = .65), B1 (*Mdn* = .59, *M* = .59), B2 (*Mdn* = .62, *M* = .61), and C1 level texts (*Mdn* = .60, *M* = 60).

Table 19. *The Kruskal-Wallis Test results of the use of Number of Modifiers per Noun Phrase across CEFR levels*

CEFR Level	Modifiers per Non Phrase
Chi-Square	16.017
df	4
Asymp. Sig.	.003

Post-hoc Mann-Whitney tests indicated that A1 level texts significantly differ from A2 level texts in terms of the number of modifiers used per noun phrase $U(N_{A1}=8, N_{A2}=72,) = 158.00$, $z = -2.08$, $p < .05$ as can be seen in Table 20 Similarly, there was a statistically significant difference found between A1 level texts and B1 level texts $U(N_{A1}=8, N_{B1}=151,) = 130.50$, $z = -3.73$, $p < .05$ when compared. Furthermore, A1 level texts were compared to B2 level texts with regard to their use of modifiers per noun phrase, and the results revealed that A1 level texts significantly differ from B2 level texts, $U(N_{A1}=8, N_{B2}=76,) = 76.00$, $z = -3.47$, $p < .05$. Finally, a statistically significant difference was observed between A1 and C1 level texts in their use of modifiers per noun phrase, $U(N_{A1}=8, N_{C1}=7,) = 8.50$, $z = -2.25$, $p < .05$.

Table 20. *Mann-Whitney Post-hoc results between A1 and other CEFR levels in terms of modifiers per noun phrase*

		CEFR Level A1	
CEFR Level	Mean rank	Z-value	*p*
A2	158.00	-2.08	.03
B1	130.50	-3.73	.00
B2	76.00	-3.47	.00
C1	8.50	-2.25	.02

A2 level texts were also compared to other CEFR level texts to investigate any significant differences among levels in terms of the use of modifiers per noun phrase. As

mentioned above, there was a significant difference between A1 and A2 levels. As table 21 demonstrates, the results of the Mann-Whitney U test revealed that A2 level texts significantly differ from B1 level texts with regard to their use of modifiers per noun phrase, $U(N_{A2}=72, N_{B1}=151,) = 4482.50$, $z = -2.11$, $p < .05$. However, when A2 was compared to B2, not a statistically significant difference was observed, $U(N_{A2}=72, N_{B2}=76,) = 2424.00$, $z = -1.19$, $p > .05$. Similarly, there was not a statistically significant difference between A2 and C1 level texts according to the Mann-Whitney U test results, $U(N_{A2}=72, N_{C1}=7,) = 218.50$, $z = -.57$, $p > .05$.

Table 21. *Mann-Whitney Post-hoc results between A2 and other CEFR levels in terms of modifiers per noun phrase*

		CEFR Level A2	
CEFR Level	Mean rank	Z-value	*p*
A1	158.00	-2.08	.03
B1	4482.50	-2.11	.03
B2	2424.00	-1.19	.23
C1	218.50	-.57	.56

Mann-Whitney post-hoc tests were performed on B1 level texts comparing them to other CEFR levels. As aforementioned, there was a statistically significant difference between A1 and B1 level texts and A2 and B1 levels. Table 22 shows that not a significant difference was observed between B1 and B2 level texts in terms of modifiers used per noun phrase, $U(N_{B1}=151, N_{B2}=76,) = 5372.500$, $z = -.783$, $p > .05$. Similarly, B1 level texts do not significantly differ from C1 level texts in their use of modifiers per noun phrase, $U(N_{B1}=151, N_{C1}=7,) = 498.000$, $z = -.258$, $p > .05$.

Table 22. *Mann-Whitney Post-hoc results between B1 and other CEFR levels in terms of modifiers per noun phrase*

		CEFR Level B1	
CEFR Level	Mean rank	Z-value	*p*
A1	130.50	-3.73	.00
A2	4482.50	-2.11	.03
B2	5372.50	-.78	.43
C1	498.00	-.25	.79

When B2 texts were compared to A1, A2, and B1, as mentioned above, the only significant difference was observed with A1 level texts. Besides, Table 23 demonstrates that Mann-Whitney post-hoc tests revealed not a statistically significant difference between B2 and C1 level texts.

As aforementioned, the only statistically significant difference between C1 and other CEFR level texts was found between A1 and C1 levels; no other significant difference was observed, $U(N_{B2}=151, N_{C1}=7,) = 258.000$, $z = -.131$, $p > .05$.

Table 23. *Mann-Whitney Post-hoc results between B2 and other CEFR levels in terms of modifiers per noun phrase*

		CEFR Level B2	
CEFR Level	Mean rank	Z-value	*p*
A1	76.00	-3.47	.00
A2	2424.00	-1.19	.23
B1	5372.50	-.78	.43
C1	258.00	-.13	.89

4.2.2.3. Minimal Edit Distance

Table 24. *The Kruskal-Wallis Test results of the use of Minimal Edit Distance across CEFR levels*

	Minimal Edit Distance
Chi-Square	3.82
df	4
Asymp. Sig.	.43

A Kruskal-Wallis test revealed that the use of minimal edit distance does not significantly differ across the CEFR levels, $H(4) = 3.823$, $p > 0.05$ as can be seen in table 24. Minimal edit distance rankings of B2 level texts (*Mdn* = .8925, *M* = 8865) were higher than A1 (*Mdn* = .8595, *M* = 7644), A2 (*Mdn* = .8845, *M* = 8682), B1 (*Mdn* = .8800, *M* = 8729), and C1 level texts (*Mdn* = .8710, *M* = 8769). Post-hoc Mann-Whitney tests were used to compare all pairs of groups.

Table 25. *Mann-Whitney Post-hoc results between A1 and other CEFR levels in terms of minimal edit distance*

		CEFR Level A1	
CEFR Level	Mean rank	Z-value	*p*
A2	229.00	-.94	.34
B1	472.50	-1.03	.30
B2	197.50	-1.62	.10
C1	23.00	-.57	.56

As can be seen in table 25, the post-hoc test revealed that A1 level texts do not significantly differ from A2 level texts in terms of their use of minimal edit distance, $U(N_{A1}=8, N_{A2}=72,) = 229.000$, $z = -.946$, $p > .05$. Similarly, not a statistically significant difference was observed between A1 and B1 level texts, $U(N_{A1}=8, N_{B1}=151,) = 472.500$,

z = -1.036, p > .05. Furthermore, there was not a statistically significant difference between A1 and B2 level texts in terms of their use of minimal edit distance, $U(N_{A1}=8, N_{B2}=76,) = 197.500$, z = -1.623, p > .05. Likewise, not a statistically significant difference was observed between A1 and C1 level texts, $U(N_{A1}=8, N_{C1}=7,) = 23.000$, z = -.579, p > .05.

A2 level texts were also compared to other CEFR level texts to explore possible significant differences among levels in terms of the use of minimal edit distance. As mentioned above, A1 level texts do not differ from A2 level texts. Similarly, table 26 shows that not a statistically significant difference was observed between A2 and B1 level texts with regard to their use of minimal edit distance, $U(N_{A2}=72, N_{B1}=151,) = 5416.500$, z = -.043, p > .05. Moreover, when A2 and B2 level texts were compared, there was not a statistically significant difference found, $U(N_{A2}=72, N_{B2}=76,) = 2483.500$, z = -.969, p > .05. Finally, A2 level texts were compared to C1 level texts, and not a significant difference was observed, $U(N_{A2}=72, N_{C1}=7,) = 245.000$, z = -.121, p > .05.

Table 26. *Mann-Whitney Post-hoc results between A2 and other CEFR levels in terms of minimal edit distance*

		CEFR Level A2	
CEFR Level	Mean rank	Z-value	*p*
A1	229.00	-.94	.34
B1	5416.50	-.04	.96
B2	2483.50	-.96	.33
C1	245.00	-.12	.90

Mann-Whitney U tests were also performed on B1 level texts, and they were compared with other CEFR levels. As previously mentioned, not a significant difference was observed between A1 and B1, and A2 and B1 level texts in terms of minimal edit

distance. Table 27 also illustrates that there was not a statistically significant difference between B1 level and B2 level texts as well, $U(N_{B1}=151, N_{B2}=76,) = 5046.000$, $z = -1.482$, $p > .05$. Finally, B1 level texts were compared to C1 level texts in terms of their use of minimal edit distance, and the results revealed not a statistically significant difference, $U(N_{B1}=151, N_{C1}=7,) = 500.500$, $z = -.237$, $p > .05$.

Table 27. *Mann-Whitney Post-hoc results between B1 and other CEFR levels in terms of minimal edit distance*

	CEFR Level B1		
CEFR Level	Mean rank	Z-value	*p*
A1	472.50	-1.03	.30
A2	5416.50	-.04	.96
B2	5046.00	-1.48	.13
C1	500.50	-.23	.81

Finally, the results of the Mann-Whitney U test revealed that there was no statistically significant difference between B2 and A1, B2 and A2, and B2 and B1 level texts, as mentioned above. B2 level texts were also compared to C1 level texts in terms of their use of minimal edit distance, and the table 28 demonstrates that the results revealed not a statistically significant difference, $U(N_{B2}=76, N_{C1}=7,) = 200.500$, $z = -1.073$, $p > .05$.

As previously mentioned, C1 level texts were compared to other CEFR levels as well, and not a statistically significant difference was observed between C1 and any other level in terms of the use of minimal edit distance.

Table 28. *Mann-Whitney Post-hoc results between B2 and other CEFR levels in terms of minimal edit distance*

	CEFR Level B2		
CEFR Level	Mean rank	Z-value	*p*
A1	197.50	-1.62	.10
A2	2483.50	-.96	.33
B1	5046.00	-1.48	.13
C1	200.50	-1.07	.28

4.2.2.4. Sentence Syntax Similarity

Table 29. *The Kruskal-Wallis Test results of the use of Sentence Syntax Similarity CEFR levels*

	Sentence Syntax Similarity
Chi-Square	20.09
df	4
Asymp. Sig.	.00

Finally, a Kruskal-Wallis test was performed to investigate any significant differences across CEFR levels on sentence syntax similarity. As shown in table 29, the results revealed that sentence syntax similarity of the texts does not significantly differ across the CEFR levels, $H(4) = 20.097$, $p < 0.05$. The group with the highest ranking of sentence syntax similarity was found to be A2 level (*Mdn* = .0975, *M* = .1010) compared to A1 (*Mdn* = .0855, *M* = .0910), B1 (*Mdn* = .0940, *M* = .0960), B2 (*Mdn* = .0770, *M* = .0782), and C1 (*Mdn* = .0750, *M* = .0746). Post-hoc Mann-Whitney tests revealed that A1 level texts do not differ from A2 level texts in terms sentence syntax similarity, $U(N_{A1}=8, N_{A2}=72,) = 225.500$, $z = -1.002$, $p > .05$ as can be seen in table 30.

Similarly, there was no significant difference observed between A1 and B1 level texts in terms of sentence syntax similarity, $U(N_{A1}=8, N_{B1}=151,) = 505.500$, $z = -.776$, $p > .05$. Moreover, not a significant difference was found when A1 level texts were

compared to B2 level texts, $U(N_{A1}=8, N_{B2}=76,) = 277.000$, $z = -.412$, $p > .05$. Likewise, A1 level texts do not significantly differ from C1 level texts with regard to sentence syntax similarity, $U(N_{A1}=8, N_{C1}=7,) = 22.000$, $z = -.696$, $p > .05$.

Table 30. *Mann-Whitney Post-hoc results between A1 and other CEFR levels in terms of sentence syntax similarity*

		CEFR Level A1	
CEFR Level	Mean rank	Z-value	*p*
A2	225.50	-1.00	.31
B1	505.50	-.77	.43
B2	277.00	-.41	.68
C1	22.00	-.69	.48

Post-hoc Mann Whitney tests were also performed on A2 level texts comparing them to other CEFR levels. As mentioned above, not a significant difference was observed between A1 and A2 texts. As shown in table 31, similarly, there was not a statistically significant difference between A2 and B1 level texts with regard to sentence syntax similarity, $U(N_{A2}=72, N_{B1}=151,) = 5144.000$, $z = -.648$, $p > .05$. However, A2 level texts significantly differ from B2 level texts in terms their sentence syntax similarity, $U(N_{A2}=72, N_{B2}=76,) = 1763.000$, $z = -3.733$, $p < .05$. Likewise, A2 level texts were compared to C1 level texts, and statistically significant difference was observed, $U(N_{A2}=72, N_{C1}=7,) = 132.500$, $z = -2.062$, $p < .05$.

Table 31. *Mann-Whitney Post-hoc results between A2 and other CEFR levels in terms of sentence syntax similarity*

		CEFR Level A2	
CEFR Level	Mean rank	Z-value	*p*
A1	225.50	-1.00	.31
B1	5144.00	-.64	.51
B2	1763.00	-3.73	.00
C1	132.50	-2.06	.03

As mentioned above, no significant difference was observed between A1 and B1 level texts, and A2 and B1 level texts. As the table 32 illustrates, the results of the post-hoc Mann-Whitney tests revealed that there was a statistically significant difference between B1 and B2 level texts with regard to their sentence syntax similarity, $U(N_{B1}=151, N_{B2}=76,) = 4013.000$, $z = -3.695$, $p < .05$. Finally, B1 level texts were compared to C1 level texts, and the difference was not statistically significant, $U(N_{B1}=151, N_{C1}=7,) = 317.500$, $z = -1.783$, $p > .05$.

Table 32. *Mann-Whitney Post-hoc results between B1 and other CEFR levels in terms of sentence syntax similarity*

		CEFR Level B1	
CEFR Level	Mean rank	Z-value	*p*
A1	505.50	-.77	.43
A2	5144.00	-.64	.51
B2	4013.00	-3.69	.00
C1	317.50	-1.78	.07

Post-hoc Mann-Whitney test was performed on B2 level texts and compared them to the other CEFR levels. As mentioned above, even though not a statistically significant difference was observed between A1 and B2 level texts, the difference between A2 and B2, and B1 and B2 were statistically significant with regard to their sentence syntax similarity. Moreover, table 33 also demonstrates that B2 level texts do not differ from C1 level texts in terms of sentence syntax similarity, $U(N_{B2}=76, N_{C1}=7,) = 247.500$, $z = -.303$, $p > .05$.

Table 33. *Mann-Whitney Post-hoc results between B2 and other CEFR levels in terms of sentence syntax similarity*

	CEFR Level B2		
CEFR Level	Mean rank	Z-value	*p*
A1	277.00	-.41	.68
A2	1763.00	-3.73	.00
B1	4013.00	-3.69	.00
C1	247.50	-.30	.76

As previously mentioned, post-hoc Mann-Whitney test was performed on C1 level texts, and the only statistically significant difference was identified between C1 and A2 level texts.

4.3. DISCURSIVE FEATURES

4.3.1. Discursive Features within the CEFR Levels

The third research question was addressed with the aim of investigating how the data were distributed within the same CEFR level. The following section will present the

results of the quantitative analysis presenting the test of normality results as well as displaying the results separately on each discursive feature.

4.3.1.1. Backward Reference

Backward reference was identified at all levels except the A1 level. As Table 34 demonstrates, a Shapiro-Wilk test showed a significant departure from normality, $W(72) = .30$, $p < .05$ for A2 level texts in terms of backward reference. Likewise, for B1 level texts, the test of normality results revealed that the backward reference data within the B1 level were not normally distributed, $W(151) = .43$, $p < .05$. Neither of the upper levels were found to be approximately normally distributed: B2 level texts ($W = .35$, $p < .05$), C1 level texts ($W = .60$, $p < .05$).

Table 34. *Test of normality results of backward reference within the CEFR levels*

	Shapiro-Wilk		
CEFR	Statistic	df	Sig.
A2	.30	72	.00
B1	.43	151	.00
B2	.35	76	.00
C1	.60	7	.00

4.3.1.2. Code-Switching

Code-switching was observed in all the levels of the CEFR in the corpus. Table 35 illustrates the test of normality results for code-switching within the CEFR levels. A Shapiro-Wilk test revealed that none of the data of code-switching within the levels of the corpus were normally distributed: A1 level texts ($W = .41$, $p < .05$), A2 level texts ($W = .45$, $p < .05$), B1 level texts ($W = .24$, $p < .05$), B2 level texts ($W = .14$, $p < .05$), and C1 level texts ($W = .64$, $p < .05$).

Table 35. *Test of normality results of code-switching within the CEFR levels*

	Shapiro-Wilk		
CEFR	Statistic	df	Sig.
A1	.41	8	.00
A2	.45	72	.00
B1	.24	151	.00
B2	.14	76	.00
C1	.64	7	.00

4.3.1.3. Connectors

Connectors were also identified in all the levels of the corpus. As can be seen in Table 36, the data related to connectors were normally distributed only in the levels A1, $W(8) = .87$, $p > .05$, and C1 $W(7) = .93$, $p > .05$. A Shapiro-Wilk test showed a significant departure from normality for the other levels: A2 ($W = .87$, $p < .05$), B1 ($W = .84$, $p < .05$) and B2 ($W = .88$, $p < .05$).

Table 36. *Test of normality results of connectors within the CEFR levels*

	Shapiro-Wilk		
CEFR	Statistic	df	Sig.
A1	.87	8	.15
A2	.87	72	.00
B1	.84	151	.00
B2	.88	76	.00
C1	.93	7	.56

4.3.1.4. Discourse Markers

As shown in Table 37, a Shapiro-Wilk test demonstrated that the data in all the levels of the CEFR in the corpus except the C1 level significantly departed from normality: A1 ($W = .72$, $p < .05$), A2 ($W = .74$, $p < .05$), B1 ($W = .65$, $p < .05$), and B2 ($W = .86$, $p < .05$). The only normal distribution of discourse markers data were identified in the C1 level of the corpus, $W(7) = .72$, $p < .05$).

Table 37. *Test of normality results of discourse markers within the CEFR levels*

	Shapiro-Wilk		
CEFR	Statistic	df	Sig.
A1	.72	8	.00
A2	.74	72	.00
B1	.65	151	.00
B2	.86	76	.00
C1	.93	7	.55

4.3.1.5. High Information Load

Table 38 illustrates the test of normality results for high information load within the CEFR levels of the corpus. High information load was only identified in the A2 and B1 levels, and neither of the data within these levels showed a normal distribution: A2 level texts ($W = .15$, $p < .05$) and B1 level texts ($W = .14$, $p < .05$).

Table 38. *Test of normality results of high information load within the CEFR levels*

	Shapiro-Wilk		
CEFR	Statistic	df	Sig.
A2	.15	72	.00
B1	.14	151	.00

4.3.1.6. Reference to Context

Reference to context was identified in all levels of the corpus except the C1 level. None of the reference to context data showed normal distribution within the C1 level: A1 ($W = .41$, $p < .05$), A2 ($W = .39$, $p < .05$), B1 ($W = .39$, $p < .05$), and B2 ($W = .86$, $p < .05$), as Table 39 demonstrates.

Table 39. *Test of normality results of reference to context within the CEFR levels*

	Shapiro-Wilk		
CEFR	Statistic	df	Sig.
A1	.41	8	.00
A2	.39	72	.00
B1	.39	151	.00
B2	.86	76	.00

4.3.1.7. Weak Coherence

Similar to high information load, weak coherence has only been identified in the A2 and B1 levels of the corpus. A Shapiro-Wilk test showed a significant departure from normality for both levels: A2 ($W = .32$, $p < .05$) and B1 ($W = .17$, $p < .05$) as Table 40 demonstrates.

Table 40. *Test of normality results of weak coherence within the CEFR levels*

	Shapiro-Wilk		
CEFR	Statistic	df	Sig.
A2	.32	72	.00
B1	.17	151	.00

4.3.2. Discursive Features Across the CEFR Levels

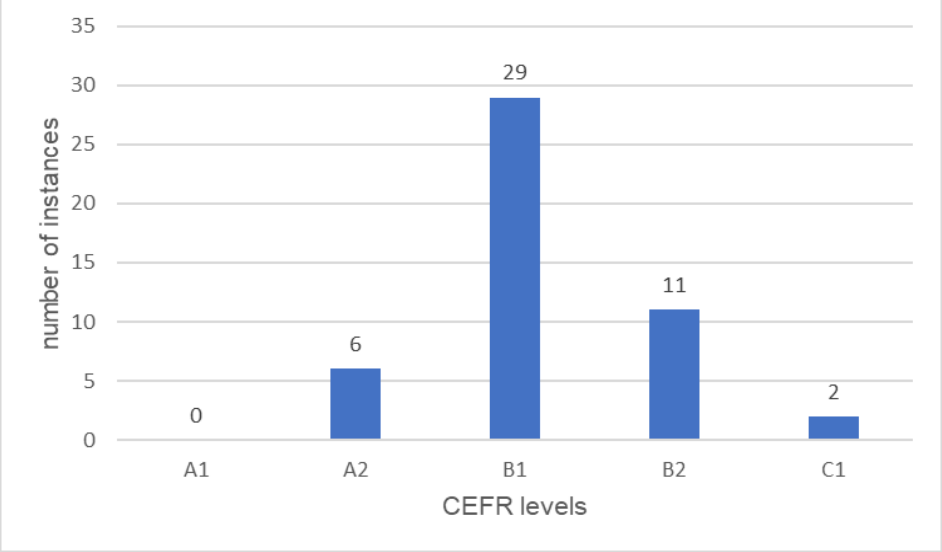

Figure 1. *Backward reference coding distribution across the CEFR levels*

Figure 1 illustrates how the backward reference feature was distributed across the CEFR levels. As can be seen in the figure, backward reference instances were observed in all the CEFR levels except the A1 level. B1 level appears to be the level that

backward reference was used the most in the speeches, with 29 instances in the texts. B2 level follows with 11 instances in the texts. The third level in which backward reference was observed the most was the A2 level with six instances, and C1 follows with two instances of backward reference. However, though the number of instances is of significance for the understanding of how the feature was distributed across the levels, the calculated means of occurrences within the levels and how they differ should also be considered. Table 41 illustrates the calculated means of backward reference occurrences across the CEFR levels. As can be seen, though the number of instances was the highest in the B1 level (M = .185) in terms of backward reference use, calculating the means of backward reference instances revealed that the highest use was in the C1 level texts (M = .285) considering the sample size. B1 (M = .185), B2 (M = .144), and A2 (M = .083) follow respectively, and because no instances were observed in the A1 level, the calculated means of backward reference was zero for that level.

Table 41. *The calculated means of backward reference occurrences across the CEFR levels*

CEFR Level	*n*	*M*
A1	8	0
A2	72	.08
B1	151	.18
B2	76	.14
C1	7	.28

Figure 2 shows how code-switching was distributed across the CEFR levels. As can be seen, in all levels of the CEFR, the code-switching feature of discourse was observed. The most frequent use was found in the B1 level texts with 30 instances of code-switching. In addition, a great number of code-switching was also observed in the A2 level texts with 19 instances of the feature. In the C1 level texts, the number of instances of code-switching was six, and B2 level texts follow with two instances of code-switching. Finally, the A1 level follows with one instance making it the level that code-switching was least used in the speeches.

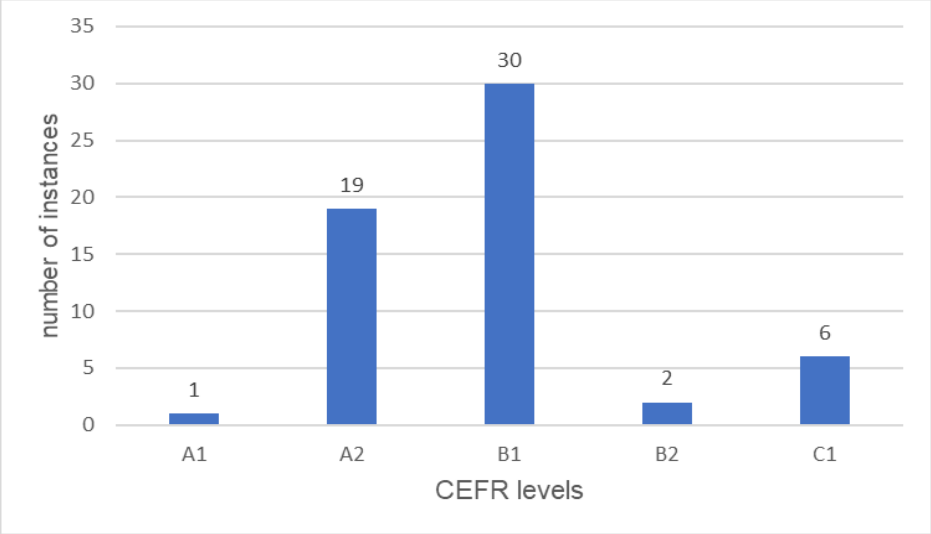

Figure 2. *Code-switching coding distribution across the CEFR levels*

Table 42 demonstrates the calculated means of code-switching occurrences across the CEFR levels. The analysis of the calculated means revealed that the level that code-switching was used the highest in the speeches was the C1 level (M = .857). A2 (M = .263), B1 (M = .178), and A1 (M = .125) follow, respectively. The level that the code-switching feature was used the least was the B2 level texts (M = .026).

Table 42. *The calculated means of code-switching occurrences across the CEFR levels*

CEFR Level	*n*	*M*
A1	8	.12
A2	72	.26
B1	151	.17
B2	76	.02
C1	7	.85

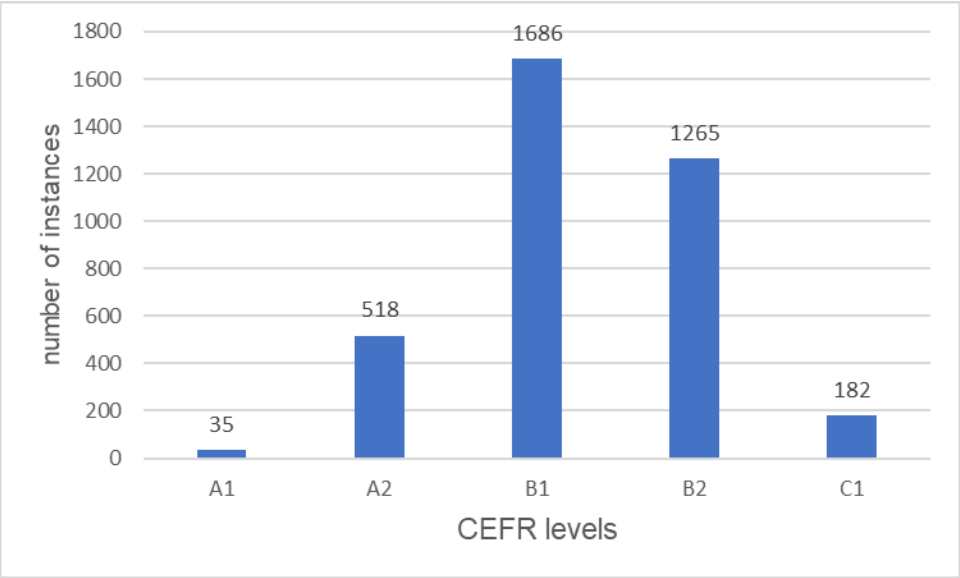

Figure 3. *Connectors coding distribution across the CEFR levels*

How connector use was distributed across the CEFR levels can be seen in Figure 3. The highest frequency of connector use was observed in the B1 level texts with 1686 instances. The B2 level texts follow with 1265 instances of connectors. The A2 level was the third level in which connectors were used most frequently with 518 instances. In the C1 level texts, the number of connectors was 182, and at the A1 level, it was 35.

Table 43. *The calculated means of connectors occurrences across the CEFR levels*

CEFR Level	*n*	*M*
A1	8	4.37
A2	72	7.18
B1	151	11
B2	76	16.64
C1	7	26

Table 43 illustrates the calculated means of connector use across the CEFR levels. Though the number of occurrences was highest in the B1 level texts, considering

the sample size, the highest calculated mean was observed in the C1 level texts (*M* = 26) in terms of connectors use. B2 (*M* = 16.644), B1 (*M* = 11), and A2 (*M* = 7.180) follow, respectively. The calculated means of connector use in the A1 level texts was 4.375, making the level the one that connectors were used the least.

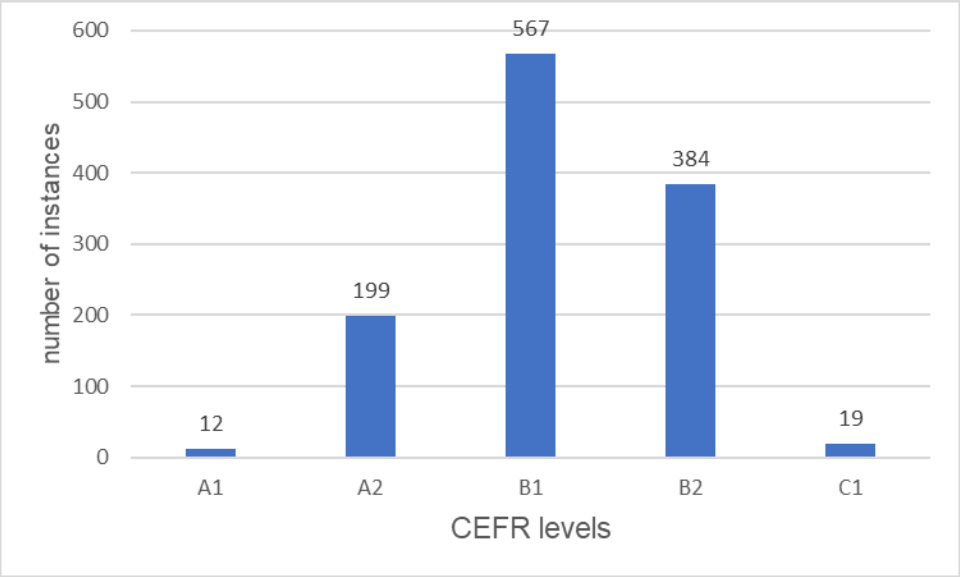

Figure 4. *Discourse markers coding distribution across the CEFR levels*

Figure 4 demonstrates how discourse markers were distributed across the CEFR levels. The B1 level appears to be the level in which discourse markers were used in the texts the most with 567 instances. The second level in which discourse markers were used the most was the B2 level with 384 instances. A2 level texts follow with 199 instances of discourse markers. C1 level was found to be the fourth level in which discourse markers were used the most with 19 instances. Finally, discourse markers were observed to be employed in the A1 level texts the least with 12 instances.

Moreover, as can be seen in Table 44, the calculated means of discourse markers use revealed that the level that discourse markers were used the most in the speeches was the B2 level (*M* = 5.052) although the B1 level was the one that the greatest number of them were used. B1 (*M* = 3.741), A2 (*M* = 2.763), C1 (*M* = 2.714) level texts follow,

respectively. The calculated means of discourse markers use also revealed that the A1 level was the level that discourse markers were used the least in the texts.

Table 44. *The calculated means of discourse markers occurrences across the CEFR levels*

CEFR Level	*n*	*M*
A1	8	1.5
A2	72	2.763
B1	151	3.741
B2	76	5.052
C1	7	2.714

In terms of high information load, the only texts it was found in were A2 level texts and B1 level texts, as can be seen in Figure 5. In the A2 level, the number of instances was two whereas five instances were found in the B1 level texts. A1, B2, and C1 level texts did not appear to include items that would result in high information load; hence, no instances were observed in the mentioned levels.

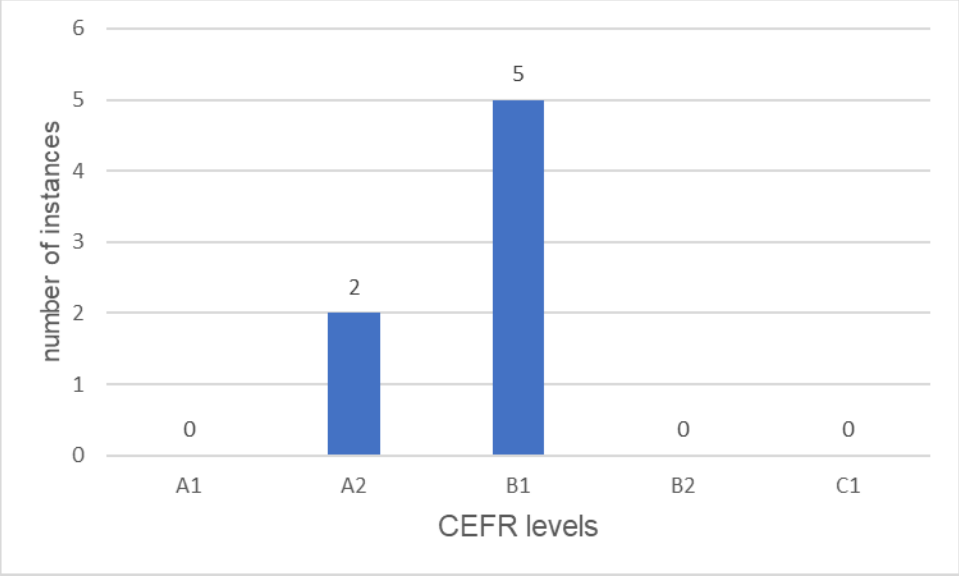

Figure 5. *High information load coding distribution across the CEFR levels*

As Table 45 illustrates, the calculated means of high information load occurrences revealed that although the number of times high information load was detected in the A2 and B1 levels were different, the means of the feature in the A2 (M = .027) and B1 (M = .026) levels were quite close to each other. Because no instances of high information load were identified in the A1, B2, and C1 level texts, the mean was calculated to be zero for these levels.

Table 45. *The calculated means of high information load occurrences across the CEFR levels*

CEFR Level	*n*	*M*
A1	8	0
A2	72	.027
B1	151	.026
B2	76	0
C1	7	0

Figure 6 demonstrates how reference to context items were distributed across the CEFR levels. Except for the C1 level, reference to context was coded in all levels of the CEFR in the corpus. In the B1 level, the items were found to be the most with 39 occurrences. The B2 level follows with 35 instances, making it the second level in which reference to context was used the most. In the A2 level, the number of reference to context items were found to be 12, and in the A1 level, only one reference to context item was found.

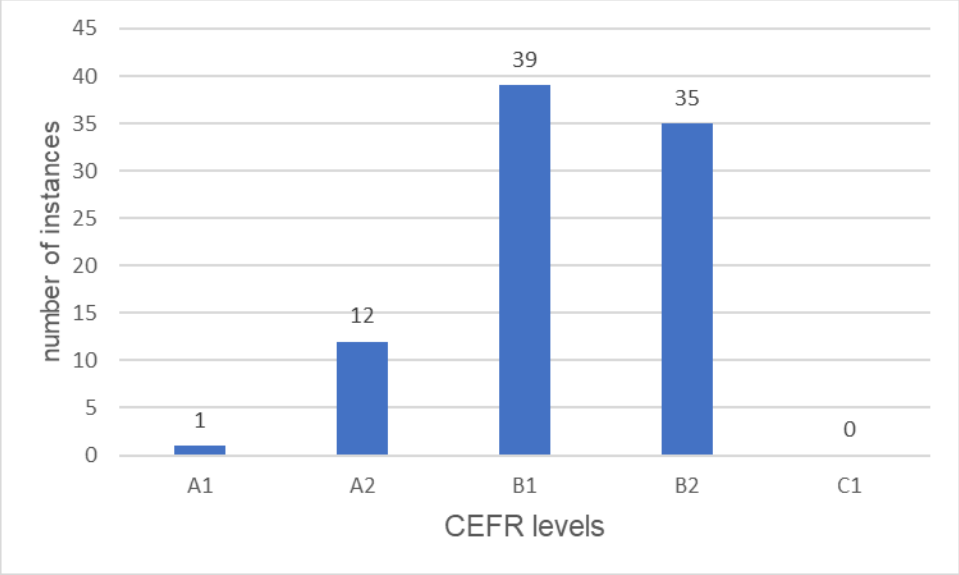

Figure 6. *Reference to context coding distribution across the CEFR levels*

Moreover, according to the calculated means of reference to context occurrences, the level that reference to context items were used the most was the B2 level (M = .460) in the corpus, as shown in Table 46. B1 level (M = .258), A2 level (M = .180), and A1 level (M = .125) follow, respectively. Since no instances were identified in the C1 level, the calculated mean of reference to context items was zero for this level in the corpus.

Table 46. *The calculated means of reference to context occurrences across the CEFR levels*

CEFR Level	*n*	*M*
A1	8	.125
A2	72	.180
B1	151	.258
B2	76	.460
C1	7	0

Finally, figure 7 illustrates how weak coherence was observed in the CEFR levels in the corpus. Weak coherence was only found in the A2 level and B1 level texts. In the A2 level, the number of instances of weak coherence was found to be 13 whereas in the B1 level the number was five. In the A1, B2, and C1 levels, no instances of weak coherence were identified.

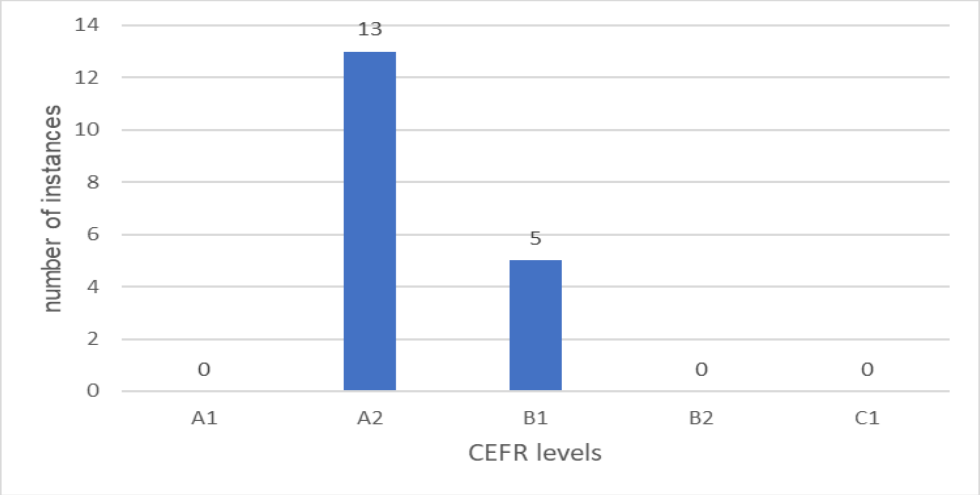

Figure 7. *Weak coherence coding distribution across the CEFR levels*

According to Table 47, the calculated mean of weak coherence was found to be .180 for the A2 level texts whereas it was found to be .033 for the B1 level texts. Since no instances of weak coherence were identified in the A1, B2, and C1 levels, the calculated means of reference to context items were zero for these levels in the corpus.

Table 47. *The calculated means of weak coherence occurrences across the CEFR levels*

CEFR Level	*n*	*M*
A1	8	0
A2	72	.180
B1	151	.033
B2	76	0
C1	7	0

4.4. CONCLUSION

The summary of the most relevant findings reported above is:

1. In the A1 level texts, all measures of syntactic complexity were non-normally distributed within the level. In the A2 level texts, only modifiers per noun phrase measure was found to be normally distributed. Likewise, only modifiers per noun phrase measure was normally distributed in the B1 level and the B2 level. All measures of syntactic complexity were found to be normally distributed in the C1 level texts.

2. The left embeddedness measure significantly differs across the CEFR levels. Statistically significant differences were observed between A2-B1, A2-B2, and B1-B2 levels in terms of the left embeddedness measure.

3. Modifiers per noun phrase measure also significantly differs across the CEFR levels. Statistically significant differences were identified between A1-A2, A1-B1, A1-B2, A1-C1, and A2-B1 levels in terms of the modifiers per noun phrase measure.

4. Minimal edit distance was the only syntactic complexity measure that does not significantly differ across the CEFR levels. None of the levels significantly differs in terms of the modifiers per noun phrase measure.

5. Sentence syntax similarity measure significantly differs across the CEFR levels as well. A2-B2, A2-C1, and B1-B2 were identified to significantly differ from each other in terms of sentence syntax similarity.

6. According to the tests of normality, the only normal distribution was found in the connector use in the A1 level, connector use in the C1 level, and discourse markers use in the C1 level. The other discursive features were non-normally distributed within the levels.

7. 48 references of backward reference were identified in the whole corpus. Backward reference was used the most in the C1 level texts. B1, B2, and A2 follow, respectively.

8. 58 references of code-switching were found in the whole corpus. Code-switching was used the most in the C1 level texts. A2, B1, A1, and B2 level texts follow, respectively.

9. Connectors were the discursive features that were used the most in all levels in the corpus. In total, 3686 references were identified. In terms of connectors, the highest number of occurrences were found in the C1 level texts, and B2, B1, A2, and A1 follow, respectively.

10. Discourse markers were the second discursive features that were used the most in all levels in the corpus. 1181 references were identified in the texts in total. Discourse markers were used the most in the speeches in the B2 level of the corpus. B1, A2, C1, and A1 level texts follow, respectively.

11. High information load was the discursive feature that was used the least in the speeches. It was identified for six times in the whole corpus. It was only found in the A2 level and B1 level texts.

12. In all of the levels, 87 references of reference to context were found in total. In the B2 level of the corpus, reference to context items were found the most. The second level that they were identified the most was the B1 level, and A2, A1, and C1 level texts follow, respectively.

13. 18 references of weak coherence were identified at all levels in total. No instances of weak coherence were identified in the A1, B2, and C1 level texts of the corpus. It was only observed in the A2 level and B1 level texts.

CHAPTER 5

5. DISCUSSION

5.1. INTRODUCTION

In this chapter, the findings unveiled from the analysis of a spoken corpus in terms of syntactic complexity and discursive features are interpreted. In order to gain a better insight into the possibilities and justification of the current findings, the research questions of 'within' and 'across' levels are combined when discussing the findings. In addition, the findings are discussed based on the existing body of literature on syntactic complexity and discursive features across different levels of learners. Moreover, this chapter touches upon the limitations of the research as well as the pedagogical implications of the findings. Finally, relevant recommendations for further research studies are provided.

5.2. DISCUSSION OF THE FINDINGS

5.2.1. SYNTACTIC COMPLEXITY

With respect to the research questions about the syntactic complexity, both similar and different findings with the existing literature are observed. The current research investigated the use of left embeddedness, modifiers per noun phrase, minimal edit distance, and sentence syntax similarity in a spoken corpus consisting of 314 conversations. Overall, significant differences across different CEFR levels were observed in terms of the use of left embeddedness, modifiers per noun phrase, and sentence syntax similarity indices. This finding supports the results of other studies on syntactic complexity proficiency levels of learners (e.g., Benzehaf, 2017: p. 43; Cumming et al., 2005: p. 5; Ortega, 2003: p. 496) which revealed significant differences between different levels of proficiency in syntactic complexity. In a similar vein, the finding that syntactic complexity differs across the CEFR levels is also in line with the studies conducted by researchers who operationalized proficiency in terms of the CEFR

as well (e.g., Alexopoulou et al., 2017: p. 180; Hawkins & Filipović, 2012: p. 23; Kang and Yan, 2018: p. 24; Khushik & Huhta, 2019: p. 8; Lahuerta Martínez, 2018: p. 1; Verspoor et al., 2012: p. 1). In these research studies, the common finding was that syntactic complexity differs in different CEFR proficiency levels.

5.2.1.1. Left embeddedness across Levels

One of the syntactic complexity measures that sought to be investigated across the CEFR levels was left embeddedness in learner speech. Since the literature suggests that syntactic complexity increase as the proficiency level increases (Hawkins & Filipović, 2012: p. 23), left embeddedness is predicted to increase with the proficiency level. The results yielded that there is a significant difference ($p < .05$) across the CEFR levels with regard to the use of left embeddedness. This result is consistent with the study conducted by McNamara, Crossley, and McCarthy (2010: p. 57) which found significant differences between high-level and low-level proficiency learners. However, in the corpus utilized in this research, the highest use of left embeddedness was observed in the A1 level, which does not support the previous research (e.g., Benzehaf, 2017: p. 43; Cumming et al., 2005: p. 5; Kim, 2004: p. 31; Ortega, 2003: p. 496; Verspoor et al., 2012: p. 1) which revealed that syntactic complexity increases as the learners' proficiency level increases.

The explanation for this finding might be due to the fact that the highest sentence length means in the whole dataset is observed at the A1 level. As left embeddedness is concerned with how many words there are before the main verb, the sentence length might have affected the results. However, the generalizability of this finding is questionable since the number of samples in the A1 level is eight which is the second least number of samples in the whole dataset. Moreover, the fact that not a significant difference was observed between the A1 level and the other levels raises questions about the classification of sample texts in terms of CEFR levels. In addition, the C1 level texts of the corpus did not provide results as the CEFR would suggest since the second-lowest mean of left embeddedness was found in the C1 level. Although the sentence length

mean increases as the levels increase in the dataset (except for the A1 level), the mean of left embeddedness is low in the C1 level. Furthermore, there is only one C1 level text whose mean is higher than the mean of B2 level texts, which raises questions about the categorization of the texts in the correct CEFR levels in the corpus. Finally, the A2-B2 continuum progresses as the CEFR level increases in terms of mean scores of sentence length, word count, sentence count, and left embeddedness, which suggests that these data yielded results as expected. As for the significant differences, the common levels that point to no significant differences were A1 and C1 levels which support the aforementioned arguments for the samples in the corpus for these levels. This finding might result from the number of samples in the levels or the incorrect categorization of the texts in these levels.

5.2.1.2. Modifiers per Noun Phrase across Levels

Another syntactic complexity measure used in this research was modifiers per noun phrase use in learner speech across different levels. Similar to the left embeddedness index, modifiers per noun phrase use is expected to increase as learners progress into upper levels. Across different CEFR levels, modifiers per noun phrase differed significantly in the current research, which is in line with the notion held by Green (2012: p. 124) who found significant differences between B and C levels and different from Banerjee et al. (2015: p. 5) who found that modifiers per noun phrase did not distinguish the levels of proficiency. The syntactic complexity results in this dataset demonstrate no logical progress as the CEFR suggests. The fact that level A1 significantly differs from all the other levels, and that A2 differs from the B level is in line with what the literature suggests as they are the lowest levels in the CEFR. However, the highest use of modifiers per noun phrase was observed in the A1 level texts which in turn spoils the findings of significant differences between A1 level and the other levels as it conflicts with the literature. Word count of the texts might have an effect on the use of modifiers per noun phrase, and word count increases as the levels increase in the dataset. This could be considered appropriate for this to lower the mean

of modifiers per noun phrase, however, since there is no logical distribution in the means, the result might not stem from this. Similar to the left embeddedness measure, this finding raises questions about the categorization of the texts in the correct CEFR level in the corpus.

5.2.1.3. Minimal Edit Distance across Levels

Minimal edit distance was another measure utilized in the current research to investigate syntactic complexity differences in learner speech across the CEFR levels. The analysis revealed that different CEFR levels did not differ in their use of minimal edit distance. Out of four indices of Coh-Metrix, this is the only one which did not differ across levels. This finding is not in line with the previous research studies which suggested an increase in syntactic complexity with the proficiency levels (e.g., Alexopoulou et al., 2017: p. 180; Benzehaf, 2017: p. 43; Cumming et al., 2005: p. 5; Hawkins & Filipović, 2012: p. 23; Kang and Yan, 2018: p. 24; Khushik and Huhta, 2019: p. 8; Lahuerta Martínez, 2018: p. 1; Ortega, 2003: p. 496; Verspoor et al., 2012: p. 1). This finding also contradicts with the CEFR according to which there need to be differences between levels in terms of complexity. However, it must also be noted that the means of minimal edit distance are quite high in all levels, that is, the similarity between sentences is little, which can be considered good for complexity as the more the output of minimal edit distance, the better it is for the complexity. While there was an expected increase in the mean values for each level from A1 to B2, there was a slight decrease in terms of the minimal edit distance in the C1 level. The data I analyzed does not signal the source of the decrease in the C1 level. In that vein, analyzing the minimal edit distance of the texts is not an indicator of level distribution for this corpus.

5.2.1.4. Sentence Syntax Similarity across Levels

The last syntactic complexity measure utilized in this research was sentence syntax similarity differences across different levels of proficiency. Similar to the left

embeddedness and modifiers per noun phrase measure, sentence syntax similarity also significantly differed across the CEFR levels in the corpus. According to the expectations based on the CEFR, as learners progress into upper levels, they are supposed to produce more syntactically complex utterances. The analysis of the texts in terms of sentence syntax similarity revealed that the A1 level did not significantly differ from the other levels of the CEFR in the corpus. This finding contradicts both the expectations based on the CEFR and the previous literature such as the study conducted by Verspoor et al. (2012: p. 1) which suggested that complex sentences are a good separator of levels, especially between A1 and A2 levels. The number of samples and the non-normal distribution of the data within the A1 level might account for A1 level texts to produce unexpected results. Furthermore, although no significant difference was observed between A1 and A2 levels, the analysis revealed that the means of sentence syntax similarity is lower in the A1 level texts than the A2 level texts. To consider the sentences more complex, the sentence syntax similarity value obtained from the Coh-Metrix needs to be low. In that vein, the fact that A1 level texts were more syntactically complex than A2 level texts in terms of sentence syntax similarity contradicts the expectations based on the CEFR and the syntactic complexity research. This finding could be explained by the fact that the samples in the A1 level are far fewer than those of the A2 level. Other than these, though no significant differences were observed across many of the levels, sentence syntax similarity means from A2 to C1 levels meet the expectations based on the CEFR.

5.2.2. DISCURSIVE FEATURES

Concerning the research questions about the discursive features, there are findings similar to, as well as different from the results of previous research. To seek answers for the research questions related to the discursive features, the current research investigated the discursive features in a spoken corpus consisting of 314 conversations. I operationalized discursive features as backward reference, code-switching, connectors, discourse markers, high information load, reference to context, and weak coherence. In

the literature, as the studies investigating these specific features are scarce and even non-existent in some. Thus, the results regarding the discursive features of the 314 texts in the corpus are predominantly discussed based on the expectations according to the CEFR. In the following sections, the discussion about each discursive feature will be presented.

5.2.2.1. Backward Reference

The only level that backward reference items was not observed was the A1 level in the corpus, which is an expected finding considering backward reference requires handling the discourse quite well by referring to the previous utterances. Council of Europe (2001) assumes that the range of cohesion expands across proficiency levels, hence, the fact that no backward reference items were found in the A1 level is in line with the CEFR's predictions. Moreover, the results in the A2 and B1 level is also expected as the number of backward reference items increases by the proficiency level. However, a decrease in the backward reference use was observed in the B2 and C1 levels, which contradicts the expectations. The reason for this decrease might be explained with the number of samples in each level. If the number were even among the levels, backward reference items would have been increased as expected. It must also be noted that the distribution of backward reference items in each level is not normal, which might have affected the results as well.

5.2.2.2. Code-switching

Another discursive feature analyzed in learner speech across levels was code-switching. The number of occurrences of code-switching should decrease with the increase of proficiency as learners are required to handle the discourse better in the upper levels of proficiency according to the Council of Europe (2001: p. 123). The fact that fewer occurrences were observed in the A1 level than the A2 level contradicts the assumptions of the CEFR. On the other hand, the calculated means of code-switching occurrences decrease starting from the A2 level to the B2 level, which is a probable finding. However, the highest calculated mean of occurrences amongst levels was

observed in the B2 level rather than the C1 level. When the word counts of the texts and levels are taken into consideration, the greatest number of words is observed at the C1 level. Because the data were preliminarily based on speech and the speeches were the longest in the C1 level, the fact that those texts revealed more code-switching examples might be an expected result as Woolford (1983: p. 529) asserts, code-switching could occur more in longer conversations. However, this finding is different from what Hamers and Blanc (2000: p. 267) stated as they believed code-switching could occur in situations where speakers' knowledge in the L2 is insufficient. To this end, the finding could be due to the length of the sentences, however, considering the fact that C1 is the highest level of the CEFR in the corpus, and learners are expected to have sufficient knowledge in the target language as the CEFR requires, the categorization of the texts in the correct level is questionable. Finally, the fact that code-switching findings fluctuate across the levels might stem from whether or not the learners who participated in the corpus had received any education about strategic competence.

5.2.2.3. Connectors

The texts in the corpus are relatively richer in terms of connector use, in fact, connectors are the discursive features that were most observed in the corpus across all levels. Based on the Council of Europe's (2001: p. 125) assumptions, it would be fair to say that the number and range of connectors in learner language should increase as learners become more proficient in the language. The connector use calculated means in learner language in the corpus increased in direct proportion to the word counts and levels. This finding is in compliance with the studies in the literature such as that of Jafarpur (1991: p. 459), Norment (1995: p. 561; 2002, p. 98), Mohammed (2015: p. 74), Carlsen (2010: p. 193), Zarco-Tejada et al. (2016: p. 215) who found that upper proficiency levels show a greater level of connectives use compared to the lower levels, yet different from the studies conducted by Zhang (2000: p. 61) and Castro (2004: p. 215) who did not find any significant differences between levels. As mentioned above, the discursive feature observed the most in all the levels were connectors, and it is the feature that is most aligned with the CEFR's assumptions as the number of samples and

the levels increase. On the other hand, while it would be realistic to expect typical sampling in terms of all features, only connector use demonstrates typicality across the A1 to C1 continuum. This result might mean that the corpus sampling could be unrealistic, or it might not be typical.

5.2.2.4. Discourse Markers

The calculated means of discourse markers occurrences from the A1 to the B2 level is aligned with the CEFR's assumptions, which correlates with what Hellermann and Vergun (2007: p. 157) found whose findings indicated that the average frequency of discourse markers increases as learners' proficiency level increases. This finding also supports the work of Neary-Sundquist (2014: p. 637) who found that discourse marker use increased as the proficiency level of the learners increased. However, although the findings are aligned with these research studies, it must be noted that the studies did not use CEFR scales to measure learner proficiency.

While the A1 to B2 continuum meets the assumptions of the CEFR, the C1 level demonstrates a decrease in the calculated means of discourse markers in learner speech. Though the other discursive features are used to a good degree in C1 level texts, the calculated means of discourse markers are very close to those of A2 level texts, which does not meet the CEFR expectations and differs from what was stated in the literature. Furthermore, the number of instances of discourse markers use in the A1 and C1 levels are very close to each other. This might either be explained with the A1 level texts' being exceptionally good or C1 level texts' exceptionally bad. However, because there was a normal distribution within the C1 level in terms of discourse markers use, it might be stated that the texts in the A1 level are exceptionally good in terms of discourse markers use. Finally, the reasons for the decrease in the C1 level might stem from issues such as topics, speech environment, and interaction between the interlocutor and the speakers. Also, it might stem from the fact that texts in the C1 level do not demonstrate a typical sampling, which may point to problems in the formation of the corpus.

5.2.2.5. High Information Load

High information load is the discursive feature that was observed the rarest in the analysis. In the whole corpus, it was only identified in the texts from A2 and B1 levels with two and five number of occurrences, respectively. Presumably, it could be expected to be seen in B2 and C1 levels as learners at these levels might tend to overuse the language. The possible reasons for this finding cannot be explained with the data analyzed in this research as the data are not enough to draw any conclusions.

5.2.2.6. Reference to Context

Council of Europe (2001: p. 125) assumes that the use of cohesive devices enriches with the proficiency level. Besides, Halliday and Hasan (1976: p. 14) categorizes 'reference' under cohesive resources. Hence, the use of reference to context might be expected to increase with the increase in proficiency levels. The calculated means of reference to context use increases as the proficiency level of the texts increases, which confirms the CEFR (Council of Europe, 2001: p. 125) assumptions. However, no instances of reference to context were observed in the C1 level, which might spring from the range of topics selected in the level.

5.2.2.7. Weak coherence

Waller (2015: p. 69) states that because the CEFR chooses wording which could mean C1 is the threshold for learners to be able to choose the appropriate language item to be coherent in discourse, discourse coherence might be the defining feature of the C1 level and above. In that vein, weak coherence might be expected to be seen in the A1 level texts the most, however, no instances were observed in the A1 level of the corpus. Moreover, while there is none in the A1 level, it is odd that the greatest number of occurrences were detected in the A2 level as the A2 level is supposed to be more proficient than the A1 level. The reason for this finding could stem from the differences in the topics, the categorization of the texts in the correct level in the corpus, or due to the fact that the speeches in the A1 level are shorter than the speeches in the A2 level. Moreover, it is also an expected finding that the instances decrease in the B1 level and

supported by the current research findings. Finally, it could also be considered normal that no instances were observed in the B2 and C1 levels though the length of speech increases incrementally, which supports the beliefs of Waller (2015: p. 69) who asserted that discourse coherence is the defining feature of the upper levels.

CHAPTER 6

6. CONCLUSION

6.1. INTRODUCTION

This section includes a summary of the research based on the research questions addressed. Furthermore, the pedagogical implications of the research limitations as well as suggestions for further research were presented.

6.2. SUMMARY OF THE FINDINGS

The present research was designed (1) to identify whether there is consistency in terms of syntactic complexity within the same CEFR level, (2) to examine any significant differences in terms of syntactic complexity across the CEFR levels, (3) to identify whether there is consistency in terms of discursive features within the same CEFR level, and (4) to analyze how discursive features were distributed across the CEFR levels. To this wake, the texts were analyzed in terms of syntactic complexity and discursive features. The syntactic complexity of the texts was analyzed with the help of an automated analysis tool, Coh-Metrix, and how they were distributed within and across the levels was analyzed with the help of IBM SPSS 22.0. For discursive features end of the research, qualitative content analysis was conducted, and the data were analyzed on a qualitative analysis tool, NVivo. In that vein, four research questions were addressed:

Research Question 1: Is there consistency within the same CEFR level distribution in a spoken learner corpus in terms of syntactic complexity?

Normality tests and means scores of the levels unveiled that all measures of syntactic complexity were non-normally distributed in the A1 and C1 level texts. However, in the A2, B1, and B2 level texts, only modifiers per noun phrase measure was found to be normally distributed. The fact that A1 and C1 level samples sizes were quite

close to each other, and that there were few texts in the levels might have affected the distribution of the data within these levels.

Research Question 2: Is there a significant difference among CEFR level distributions in a spoken learner corpus in terms of syntactic complexity?

Kruskal-Wallis tests revealed that three out of four syntactic complexity measures significantly differ across the CEFR levels. Left embeddedness, modifiers per noun phrase, and sentence syntax similarity measures were the ones that demonstrated a statistically significant difference. Statistically significant differences were observed between A2-B1, A2-B2, and B1-B2 levels in terms of the left embeddedness measure. Moreover, statistically significant differences were identified between A1-A2, A1-B1, A1-B2, A1-C1, and A2-B1 levels in terms of the modifiers per noun phrase measure. A2-B2, A2-C1, and B1-B2 were also identified to significantly differ from each other in terms of sentence syntax similarity. However, minimal edit distance was the only syntactic complexity measure that does not significantly differ across the CEFR levels, and none of the levels does not significantly differ in terms of the modifiers per noun phrase measure.

Research Question 3: Is there consistency within the same CEFR level distribution in a spoken learner corpus in terms of discursive features?

The tests of normality revealed that most of the data within the levels were not normally distributed. The only normal distribution was found in the connector use in the A1 level, connector use in the C1 level and discourse markers use in the C1 level. The other discursive features were non-normally distributed within the levels. This finding might have affected the other findings related to the discursive features across the levels.

Research Question 4: Is there consistency among CEFR level distributions in a spoken learner corpus in terms of discursive features?

In total, 48 references of backward reference, 58 references of code-switching, 3686 references of connectors, 1181 references of discourse markers, six references high information load, 87 references of reference to context, and 18 references of weak

coherence were identified in all of the levels of the corpus. Connectors were the discursive features that were used the most in all levels in the corpus. The greatest number of instances were identified in the C1 level texts, and B2, B1, A2, and A1 follow, respectively. The mean scores of connectors use in the levels of the corpus increased in line with the increase of word counts and proficiency levels.

The second discursive feature that was used the most in all the levels was discourse markers. B2 level texts were the ones that discourse markers were used the most in the speeches, and B1, A2, C1, and A1 level texts follow, respectively. The mean scores of discourse markers instances from the A1 to the B2 increased as the proficiency levels did, however, a decrease in the mean scores of discourse markers in the C1 level of the corpus was identified, which could spring from the topic, environment, and interaction-related issues, or due to C1 level not demonstrating a typical sampling.

Reference to context was the third discursive feature that was used the most in the learner speeches across the levels. B2 level of the corpus was the one that reference to context items were identified the most. B1, A2, A1, and C1 level texts follow, respectively. The mean scores of reference to context items within the levels increase with the proficiency level, nonetheless, C1 level texts did not appear to include any reference to context items, which could result from the range of topics selected in the level.

Code-switching was used the most in the C1 level texts. A2, B1, A1, and B2 level texts follow, respectively. The continuum A2-B2 revealed probable findings according to the mean scores of code-switching in the levels. Since the data were based on speech, and the speeches were the longest in the C1 level, the matter of mean scores of code-switching being the highest among the levels could be considered normal.

Furthermore, backward reference was also used the most in the C1 level texts, and B1, B2, and A2 level texts follow, respectively. The fact that the number of instances increases in the A2 and B1 levels might be considered normal, nevertheless, there should not be a decrease in the B2 and C1 levels. The decrease in these levels could be due to the sample size or the data being non-normally distributed.

Finally, high information load and weak coherence were the discursive features that were rather rarely identified in the levels. Both high information load and weak coherence were only identified in the A2 level and B1 level texts. Weak coherence might have been expected to be found in the A1 level texts, however, no instances were observed in the A1 level of the corpus. The reason for the finding regarding weak coherence could result from the differences in the topics, the categorization of the texts in the correct level in the corpus, or the speeches being shorter in the A1 level than in the A2 level. The fact that no instances were identified in the upper levels of the corpus supports what was stated in the previous research.

6.3. LIMITATIONS OF THE RESEARCH

It is plausible that a number of limitations are present in the research and could have influenced the results obtained. To begin with, the sample size of the corpus is not evenly distributed (n = 8 for A1 level, n = 72 for A2 level, n = 151 for B1 level, n = 76 for B2 level, and n = 7 for C1 level). Although several validity and reliability checks were conducted in the research, caution must be exercised inasmuch as the generalizability of these results in broader contexts is subject to certain limitations due to the inequality of the sample size.

Besides, syntactic complexity extends beyond just the linguistic form and involves various other constituents, namely, the cognitive load of the speakers, the environment that the data were gathered, and factors arising from interaction, which limits the generalizability of the results. Furthermore, the procedure of the classification and level identification of the data was not clearly stated, hence, the validity and correctness of the corpus levels may be questionable, which in turn might affect the results of the present research. Finally, most of the research included in the literature and the discussion sections used either one or some of the indices and discursive features as well as using different statistical methods, which could be limitation of the discussion

section of the current research. The aforementioned limitations mean that one should interpret the findings of the research cautiously.

6.4. PEDAGOGICAL IMPLICATIONS

Notwithstanding the limitations, the present research offers valuable pedagogical implications based on the findings from the analysis of spoken texts in a learner corpus. First, the present research yielded that there could be problems in the classification of the levels in the correct CEFR level in the corpus. Considering BACKBONE was designed as pedagogic corpora for both learners to self-study and teachers to create tasks and activities, problems may arise resulting from the incorrect categorization of the levels. Hence, those wishing to include the corpus in their learning and teaching must be cautious when using the corpus. Also, based on the findings, it would be suggested that language curricula be prepared taking syntactic complexity and discursive features into consideration since such curricula would bring learners closer to the levels as defined on the CEFR. To exemplify, several instances of code-switching were identified in the levels. For this feature to be used correctly by the learners, curricula can be prepared so that learners can achieve the predicted goals of the specific levels. Also, strategies to improve learners' use of minimal edit distance could be included in language curricula as the learners did not statistically differed across the levels in the current research.

6.5. RECOMMENDATIONS FOR FURTHER RESEARCH

The findings unveiled from the analysis provide some insights for further investigation. First, further experimental investigations could be conducted exploring the cognitive processes of the learners in order to have a greater understanding of the learners' motivations to use the language as they do. Moreover, researchers could use data that were categorized in a proficiency level with the help of a valid rating scale. Suchlike data can provide better insights into the learner language at specific proficiency

levels. Finally, on a wider level, research is also needed to explore the teachability of syntactic complexity and discursive features by dint of explicit instruction. Such data demand a longitudinal research design investigating the change in these features in learner language over a period of time, thereby enabling to reach better conclusions for preparing language curricula.

REFERENCES

1. Books

Baralt, Melissa (2012). "Coding qualitative data." In Eds Alison Mackey & Susan. M. Gass. Research Methods in Second Language Acquisition. West Sussex: Blackwell, 222–244.

Barni, Monica & Luisa Salvati (2019). "The Common European Framework of Reference (CEFR)." In Eds Elana Shohamy, Iair G. Or, & Stephen May. Language Testing and Assessment (3rd ed.). Cham, New Zealand: Springer International Publishing. 417-426

Blakemore, Diane (1992). Understanding utterances. Oxford: Blackwell.

Bryman, Alan (2012). Social research methods (4th edition). Oxford: Oxford University Press.

Bullock, Barbara E. & Almeida Jacqueline Toribio (2009). "Themes in the study of code-switching." In Eds Barbara E. Bullock & Almeida Jacqueline Toribio. The Cambridge handbook of linguistic code-switching. Cambridge: Cambridge University Press. 1-17.

Bulté, Bram & Alex Housen (2012). "Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA." In Eds Alex Housen, Folkert Kuiken, & Ineke Vedder. Amsterdam: John Benjamins Pub, 21-46.

Carlsen, Cecilie (2010). "Discourse connectives across CEFR-levels: A corpus based study." In Eds Inge Bartning, Maisa Martin, & Ineke Vedder. Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research [Eurosla Monographs Series 1] European Second Language Association, 191-206.

Dahl, Östen (2004). The growth and maintenance of linguistic complexity. John Benjamins.

Ellis, Rod & Gary Barkhuizen (2005). Analysing Learner Language. Oxford: Oxford University Press.

Erman, Britt (1987). Pragmatic expressions in English: A study of you know, you see, and I mean in face-to-face conversation. Stockholm: Almqvist and Wiksell.

Ferrari, Stefania (2012). "A longitudinal study of complexity, accuracy and fluency variation in second language development." In Eds Alex Housen, Folkert Kuiken, & Ineke Vedder. Dimensions of L2 performance and proficiency. Amsterdam: John Benjamins, 277-298.

Givón, Thomas (1995). “Coherence in text vs. coherence in mind.” In Eds Morton Ann Gernsbacher & Thomas Givón. Coherence in spontaneous text. Amsterdam: John Benjamins, 59-115.

Green, Anthony (2012). Language functions revisited: Theoretical and empirical bases for language construct definition across the ability range (1st ed.). Cambridge: Cambridge University Press.

Guba, Egon G., & Yvonna Lincoln (1989). Fourth generation evaluation. Newbury Park, CA: Sage.

Gutiérrez Eugenio, Esther (2013). La traducción y la interpretación como herramientas para fomentar el aprendizaje en el aula de ELE: Principios básicos y propuestas didácticas [Translation and Interpreting as Tools to Promote Learning in the Spanish as a Foreign Language Classroom: Basic Principles and Didactic Proposals]. In Actas del I Congreso Internacional de Didáctica de Español como Lengua Extranjera. Budapest: Instituto Cervantes.

Hamers, Josiane F. & Michael H. A. Blanc (2000). Bilinguality and bilingualism. Cambridge: Cambridge University Press.

Hatch, Evelyn (1992). Discourse and Language Education. Cambridge: Cambridge University Press.

Halliday, Michael Alexander (1994). An introduction to functional grammar. London: Edward Arnold.

Halliday, Michael Alexander & Ruqaiya Hasan (1976). Cohesion in English. London: Longman.

Hasselgreen, Angela (2004). Testing the spoken English of young Norwegians: A study of test validity and the role of "smallwords" in contributing to pupils' fluency. Cambridge: Cambridge University Press.

Heller, Monica & Carol W. Pfaff (1996). “Code-switching.” In Eds Hans Goebl, Peter H. Nelde, Zdenek Stary & Wolfgang Wölck. Contact linguistics, Vol. 1, Berlin/New York: de Gruyter, 594–609

Housen, Alex, Folkert Kuiken & Ineke Vedder (Eds.). (2012). “Epilogue.” Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA. Amsterdam, Philadelphia: John Benjamins Publishing Company, 299-302

Hulstijn, Jan H., Charles J. Alderson & Rob Schoonen (2010). “Developmental stages in second-language acquisition and levels of second-language proficiency: Are there links between them?” In Eds Inge Bartning, Maisa Martin and Ineke Vedder. Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research [Eurosla Monographs Series 1]. European Second Language Association, 11–20. http://eurosla.org/monographs/EM01/EM01home.html

Hymes, Dell (1977). Foundations in sociolinguistics. London: Tavistock.

Janesick, Valerie J. (2000). "The choreography of qualitative research design: Minuets, improvisations, and crystallization." In Eds Norman K. Denzin & Yvonna S. Lincoln. Handbook of qualitative research (2nd ed.). Thousand Oaks, CA: Sage Publications, 379-400.

Jucker Andreas H. & Yael Ziv (1998). "Discourse markers: Introduction." In Eds Andreas H. Jucker & Yael Ziv. Discourse Markers: Descriptions and theory. Amsterdam/Philadelphia: John Benjamins Publishing, 1-12

Kuiken, Folkert, Ineke Vedder & Roger Gilabert (2010). "Communicative adequacy and linguistic complexity in L2 writing." In Eds Inge Bartning, Maisa Martin, & Ineke Vedder. Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research [Eurosla Monographs Series 1] European Second Language Association, 81–100.

Lincoln, Yvonna S. & Egon G. Guba (1985). Naturalistic inquiry. Beverly Hills, CA: Sage.

McCormick, Ernest (1970). Human Factors Engineering. New York: McGraw-Hill Book Company.

McNamara, Danielle S., Arthur C. Graesser, Philip M. McCarthy & Zhiqiang Cai (2014). Automated evaluation of text and discourse with Coh-Metrix. Cambridge, MA: Cambridge University Press.

Merriam, Sharan B. (2009). Qualitative research: A guide to design and implementation. San Francisco: Jossey-Bass.

Miestamo, Matti (2008). "Grammatical complexity in a cross-linguistic perspective." In Eds Matti Miestamo, Kaius Sinnemaki, & Fred Karlsson. Language complexity: typology, contact, change. John Benjamins, 23–41.

North, Brian (2008). "The CEFR levels and descriptor scales." In Eds Lyanda Taylor & Cyril J. Weir. Multilingualism and Assessment: Achieving transparency, assuring quality, sustaining diversity. Cambridge, UK: Cambridge University Press, 21-55.

Piccardo, Enrica & Brian North (2020). "The Dynamic Nature of Plurilingualism: Creating and Validating CEFR Descriptors for Mediation, Plurilingualism and Pluricultural Competence." In Eds Sunny Man Chu Lau & Saskia Van Viegen. Plurilingual Pedagogies Critical and Creative Endeavors for Equitable Language in Education. Switzerland: Springer, 279-301

Polat, Nihat, Tammy Gregersen & Peter MacIntyre (eds) (2019). Research-Driven Pedagogy: Implications of L2A Theory and Research for the Teaching of Language Skills. Routledge.

Renkema, Jan (2004). Introduction to discourse studies. Philadelphia, PA, USA, Amsterdam: John Benjamins Publishing Company.

Révész, Andrea (2012). "Coding Second Language Data Validly and Reliably." In Eds Alison Mackey & Susan M. Gass. Research Methods in Second Language Acquisition. West Sussex: Blackwell, 222–244.

Schiffrin, Deborah (1987). Discourse markers. Cambridge: Cambridge University Press.

Schiffrin, Deborah, Deborah Tannen & Heidi E. Hamilton (2001). The handbook of discourse analysis. Malden, Massachusetts: Blackwell.

Schwandt Thomas A. (2001). Dictionary of Qualitative Inquiry. Sage, Thousand Oaks, CA.

Song, Yanan & Stephen Andrews (2009). The L1 in L2 learning: Teachers' beliefs and practices. München: LINCOM.

Trim, John (2007). Modern languages in the Council of Europe 1954-1997: International co-operation in support of lifelong language learning for effective communication, mutual cultural enrichment and democratic citizenship in Europe. Strasbourg: Council of Europe

Wolfe-Quintero, Kate, Shunji Inagaki & Hae-Young Kim (1998). Second language development in writing: Measures of fluency, accuracy, and complexity (Technical Report No. 17). Honolulu, HI: Second Language Teaching & Curriculum Center, University of Hawaii.

2. <u>Articles, Papers, Other Printed Publications</u>

Alexopoulou, Theodora, Marije Michel, Akira Murakami & Detmar Meurers (2017). "Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques". Language Learning, 67(S1), 180-208. doi:10.1111/lang.12232

Azadnia, Masoud, Ahmad Reza Lotfi, Reza Biria. (2019). "A Study of Syntactic Complexity via Coh-Metrix: Similarities and Differences of Ph.D. Dissertations Written by Iranian University Students and English Native Speakers". Research in English Language Pedagogy, 7(2), 232-254. doi: 10.30486/relp.2018.663453

Banerjee, Jayanti, Xun Yan, Mark Chapman & Heather Elliott (2015). "Keeping up with the times: Revising and refreshing a rating scale". Assessing Writing, 26(1), 5–19.

Becker, Anthony (2010). "Distinguishing linguistic and discourse features in ESL students' written performance". Modern Journal of Applied Linguistics, 2(5), 406-424.

Benzehaf, Bouchaib (2017). "Comparing learners' general proficiency levels with their writing productive ability: How correlated are they?". Eurasian Journal of Applied Linguistics, 3(2), 43-58. doi:10.32601/ejal.460961

Biber, Douglas, Bethany Gray & Kornwipa Poonpon (2011). "Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?". TESOL Quarterly, 45(1), 5-35. doi:10.5054/tq.2011.244483

Broek, Simon & Inge van den Ende (2013). "The implementation of the Common European Framework for Languages in European education systems – Study". European Union. Retrieved from https://op.europa.eu/en/publication-detail/-/publication/e5586004-7f80-4431-88fc-b028f28e891a/language-en

Brown, Annie, Noriko Iwashita & Tim McNamara (2005). "An examination of rater orientations and test-taker performance on English-for-academic-purposes speaking tasks". TOEFL Monograph Series MS-29. Princeton, NJ: Educational Testing Service.

Bulté, Bram & Alex Housen (2014). "Conceptualizing and measuring short-term changes in L2 writing complexity". Journal of Second Language Writing, 26, 42–65. doi:10.1016/j.jslw.2014.09.005

Bulté, Bram & Hanne Roothooft (2020). "Investigating the interrelationship between rated L2 proficiency and linguistic complexity in L2 speech". System, 91, 102246. doi:10.1016/j.system.2020.102246

Byrnes, Heidi (2007). "Perspectives". The Modern Language Journal, 91(4), 641-645.

Casanave, Christine Pearson (1994). "Language development in students' journals". Journal of Second Language Writing, 3, 179-201. http://dx.doi.org/10.1016/ 1060-3743(94)90016-7

Castro, Carolyn D. (2004). "Cohesion and the social construction of meaning in the essays of Filipino college students' writings in L2 English". Asia Pacific Education Review, 5, 215–225.

Chaudron, Craig. & Richards Jack C. (1986). "The effect of discourse markers on the comprehension of lectures". Applied Linguistics, 7, 113-127.

Clercq, Bastien De & Alex Housen (2017). "A cross-linguistic perspective on syntactic complexity in L2 development: Syntactic elaboration and diversity". The Modern Language Journal, 101(2), 315-334. doi:10.1111/modl.12396

Cook, Vivian (2001). "Using the first language in the classroom". The Canadian Modern Language Review, 57, 402–23. doi:10.3138/ cmlr.57.3.402.

Cooper, Thomas C. (1976). "Measuring written syntactic patterns of second language learners of German". The Journal of Educational Research, 69(5), 176-183. doi:10.1080/00220671.1976.10884868

Corsetti, Cristiane Ruzicki & Cristina Lopes Perna (2018). "A corpus-based study of pragmatic markers at CEFR level B1". Letras de Hoje, 52(3), 302-309.

Council of Europe (1992). "Transparency and Coherence in Language Learning in Europe: Objectives, Evaluation, Certification". Report on the Rüschlikon Symposium. Strasbourg: Council of Europe.

Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press

Crossley, Scott A. & Danielle S. Mcnamara. (2011). Shared features of L2 writing: Intergroup homogeneity and text classification. Journal of Second Language Writing, 20(4), 271–285.

Crossley, Scott A. & Danielle S. Mcnamara (2014). "Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners". Journal of Second Language Writing, 26, 66-79. doi:10.1016/j.jslw.2014.09.006

Cumming, Alister, Robert Kantor, Kyoko Baba, Usman Erdosy, Keanre Eouanzoui, & Mark James (2005). "Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL". Assessing Writing, 10(1), 5-43.

Díez-Bedmar, María Belén & Pascual Pérez-Paredes (2020). "Noun phrase complexity in young Spanish EFL learners' writing: Complementing syntactic complexity indices with corpus-driven analyses". International Journal of Corpus Linguistics, 25(1), 4-35

Dowell, Nia M. M., Arthur C. Graesser, & Zhiqiang Cai (2016). "Language and discourse analysis with Coh-Metrix: Applications from educational material to learning environments at scale". Journal of Learning Analytics, 3(3), 72–95. doi: 10.18608/jla.2016.33.5

Fernández, Julieta, Anna Gates Tapia & Xiaofei Lu (2014). "Oral proficiency and pragmatic marker use in L2 spoken Spanish: The case of pues and bueno". Journal of Pragmatics, 74, 150-164.

Ferreira, Fernanda (1991). "Effects of length and syntactic complexity on initiation times for prepared utterances". Journal of Memory and Language, 30(2), 210-233. doi:10.1016/0749-596x(91)90004-4

Figueras, Neus (2012). "The impact of the CEFR". ELT Journal, 66(4), 477-485. doi:10.1093/elt/ccs037

Figueras, Neus, Brian North, Sauli Takala, Norman Verhelst & Piet Van Avermaet (2005). "Relating examinations to the Common European Framework: a manual". Language Testing, 22(3), 261-279

Fraser, Bruce (1988). "Types of English discourse markers". Acta Linguistica Hungarica 38(1-4): 19-33.

Fraser, Bruce (1999). "What are discourse markers?". Journal of Pragmatics, 31(7), 931-952. doi:10.1016/s0378-2166(98)00101-5

Fuller, Janet M (2003). "The influence of speaker roles on discourse marker use". Journal of Pragmatics 35, 23–45.

Fung, Loretta & Ronald Carter (2007). "Discourse markers and spoken English: Native and learner use in pedagogic settings". Applied Linguistics, 28(3), 410–439. doi:10.1093/applin/amm030

Gomez Gonzalez, Maria de Los Angeles. (2013). "A reappraisal of lexical cohesion in conversational discourse". Applied Linguistics, 34(2), 128-150. doi:10.1093/applin/ams026

Graesser, Arthur C., Danielle S. McNamara, Max M. Louwerse, & Zhiqiang Cai (2004). "Coh-Metrix: Analysis of text on cohesion and language". Behavior Research Methods, Instruments, & Computers, 36(2), 193-202. doi: 10.3758/bf03195564

Gyllstad, Henrik, Jonas Granfeldt, Petra Bernardini, & Marie Källkvist (2014). "Linguistic correlates to communicative proficiency levels of the CEFR: The case of syntactic complexity in written L2 English, L3 French and L4 Italian". EUROSLA Yearbook, 14, 1-30. doi:10.1075/eurosla.14.01gyl

Hellermann, John & Andrea Vergun (2007). "Language which is not taught: The discourse marker use of beginning adult learners of English". Journal of Pragmatics, 39, 157-179.

Heyworth, Frank (2006). "The Common European Framework". ELT Journal, 60(2), 181-183. doi:10.1093/elt/cci105

House, Juliane (2013). "Developing pragmatic competence in English as a lingua franca: Using discourse markers to express (inter) subjectivity and connectivity". Journal of Pragmatics, 59, 57-67.

Hsieh, Hsiu-Fang, & Sarah E. Shannon (2005). "Three approaches to qualitative content analysis". Qualitative Health Research, 15(9), 1277-1288. doi:10.1177/1049732305276687

Ishikawa, Sandra (1995). "Objective measurement of low-proficiency EFL narrative writing". Journal of Second Language Writing, 4, 51–69.

Iwashita, Noriko, Annie Brown, Tim McNamara & Sally O'Hagan (2005). "What features of language distinguish levels of learner proficiency?". Applied Linguistics, 29(1), 24-49.

Iwashita, Noriko, Lyn May & Paul Moore (2017). Features of discourse and lexical richness at different performance levels in the APTIS speaking test. British Council. ISSN 2057-5203.

Jacoby, Jacob (1977). “Information Load and Decision Quality: Some Contested Issues”. Journal of Marketing Research, 14(4), 569-773. doi:10.2307/3151201

Jafarpur, Abdoljavad (1991). “Cohesiveness as a basis for evaluating compositions”. System, 19(4), 459-465.

Jiang, Wenying (2012). “Measurements of development in L2 written production: The case of L2 Chinese”. Applied Linguistics, 34(1), 1-24. doi:10.1093/applin/ams019

Jones, Christian & Ronald Carter (2014). “Teaching spoken discourse markers explicitly: A comparison of III and PPP”. IJES, 14(1), 37-54.

Kang, Okim & Xun Yan (2018). “Linguistic features distinguishing examinees' speaking performances at different proficiency levels”. Journal of Language Testing & Assessment, 1(1), 24-39. doi:10.23977/langta.2018.11003

Kang, Okim, Garrett M. Larson & Soo-Hyun Koo (2019). “Interaction features in predicting Common European Framework of Reference level”. Journal of Language Testing & Assessment, 2, 1-12.

Khushik, Ghulam Abbas & Ari Huhta (2019). “Investigating syntactic complexity in EFL learners' writing across Common European Framework of Reference levels A1, A2, and B1”. Applied Linguistics, 1-28. doi:10.1093/applin/amy064

Kim, S (2004). “A study of development in syntactic complexity by Chinese learners of English and its implications on the CEF scales” [Unpublished master’s thesis]. Lancaster University

Knoch, Ute, Amir Rouhshad, & Neomy Storch (2014). “Does the writing of undergraduate ESL students develop after one year of study in an English-medium university?”. Assessing Writing, 21, 1– 17.

Kohn, Kurt (2012). “Pedagogic corpora for content and language integrated learning. Insights from the BACKBONE Project”. The EuroCALL Review, 20(2), 3–22. doi: 10.4995/eurocall.2012.11374

Kormos, Judit (2011). “Task complexity and linguistic and discourse features of narrative writing performance”. Journal of Second Language Writing, 20(2), 148–161.

Kuiken, Folkert & Ineke Vedder (2019). “Syntactic complexity across proficiency and languages: L2 and L1 writing in Dutch, Italian and Spanish”. International Journal of Applied Linguistics, 29, 192-210. doi:10.1111/ijal.12256

Kuiken, Folkert, Ineke Vedder, Alex Housen, & Bastien De Clercq (2019). “Variation in syntactic complexity: Introduction”. International Journal of Applied Linguistics. doi:10.1111/ijal.12255

Kurteš, Svetlana, & Nick Saville (2008). “The English Profile Programme: An overview”. Research Notes, 33, 2–4.

Lahmann, Cornelia, Rasmus Steinkrauss, & Monika S Schmid (2015). "Factors affecting grammatical and lexical complexity of long-term L2 speakers' oral proficiency". Language Learning, 66(2), 354–385. doi:10.1111/lang.12151

Lahuerta Martínez, Ana Cristina (2018). "Analysis of syntactic complexity in secondary education EFL writers at different proficiency levels". Assessing Writing, 35, 1-11. doi:10.1016/j.asw.2017.11.002

Lambert, C., & Nakamura, S. (2018). "Proficiency-related variation in syntactic complexity: A study of English L1 and L2 oral descriptive discourse". International Journal of Applied Linguistics, 29(2), 1-17. doi:10.1111/ijal.12224

Larsen-Freeman, Diane (2006). "The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English". Applied Linguistics, 27, 590–619. doi:10.1093/applin/aml029.

Little, David (2006). "The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact". Language Teaching, 39(3), 167-190.

Little, David (2012). "The Common European Framework of Reference for Languages, the European Language Portfolio, and language learning in higher education". Language Learning in Higher Education, 1(1), 1-21. doi:10.1515/cercles-2011-0001

Lu, Xiaofei (2010). "Automatic analysis of syntactic complexity in second language writing". International Journal of Corpus Linguistics, 15(4):474-496.

Mancilla, Rae L., Nihat Polat, & Ahmet O. Akcay (2015). "An investigation of native and nonnative English speakers' levels of written syntactic complexity in asynchronous online discussions". Applied Linguistics, 38(1), 112-134. doi:10.1093/applin/amv012

Martyniuk, Waldemar & José Noyons Noijons (2007). "Executive summary of results of a survey on the use of the CEFR at national level in the Council of Europe member states". Retrieved from https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168069b7ad

Mazgutova, Diana, & Judit Kormos (2015). "Syntactic and lexical development in an intensive English for Academic Purposes programme". Journal of Second Language Writing, 29, 3-15. doi:10.1016/j.jslw.2015.06.004

McNamara, Danielle S., Scott A Crossley & Philip M. McCarthy (2010). "Linguistic features of writing quality". Written Communication, 27(1), 57-86.

Mohamadi, Zohre & Massoud Rahimpour (2018). "Task types and discursive features: mediating role of meta-talk in focus". Iranian Journal of Language Teaching Research, 6(1), 17-40.

Mohammed, Asabe Sadiya (2015). "Conjunctions as cohesive devices in the writings of English as second language learners". Procedia - Social and Behavioral Sciences, 208, 74–81. doi:10.1016/j.sbspro.2015.11.182

Müller, Simone (2004). "'Well you know that type of person': functions of well in the speech of American and German students". Journal of Pragmatics, 36(6), 1157–1182. doi:10.1016/j.pragma.2004.01.008

Neary-Sundquist, Colleen (2014). "The use of pragmatic markers across proficiency levels in second language speech". Studies in Second Language Learning and Teaching, 4(4), 637-663.

Nippold, Marilyn A., Megan W. Frantz-Kaspar & Laura M. Vigeland (2017). "Spoken language production in young adults: Examining syntactic complexity". Journal of Speech Language and Hearing Research, 60(5), 1-9.

Normand-Marconnet, Nadine & Joseph Lo Bianco (2015). "The Common European Framework of Reference down under: A survey of its use and non-use in Australian universities". Language Learning in Higher Education, 5(2), 281-307. doi:10.1515/cercles-2015-0014

Norment, Nathaniel Jr. (1995). "Discourse features of African American students' writings". Journal of Black Studies, 25, 558-576.

Norment, Nathaniel Jr. (2002). "Quantitative and qualitative analyses of textual cohesion in African American students' writing in narrative, argumentative, and expository modes". CLA Journal, 46(1), 98.

Norris, John. M. & Lourdes Ortega (2009). "Towards an organic approach to investigating CAF in instructed SLA: The case of complexity." Applied Linguistics, 30(4), 555-578. doi:10.1093/applin/amp044

North, Brian (2007). "The CEFR illustrative descriptor scales". The Modern Language Journal, 91(4), 656-659.

Nowell, Lorelli S., Jill M. Norris, Deborah E. White, et al. (2017). "Thematic analysis: Striving to meet the trustworthiness criteria". International Journal of Qualitative Methods, 16(1), 1-13.

Ortega, Lourdes (1999). "Planning and focus on form in L2 oral performance". Studies in Second Language Acquisition, 21(1), 109-148.

Ortega, Lourdes (2003). "Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing". Applied Linguistics, 24(4), 492-518. doi:10.1093/applin/24.4.492

Ortega, Lourdes (2015). "Syntactic complexity in L2 writing: Progress and expansion". Journal of Second Language Writing, 29, 82–94. doi:10.1016/j.jslw.2015.06.008

Pérez, Marta Aguilar & Elisabeth Arnó Macià (2002). "Metadiscourse in lecture comprehension: Does it really help foreign language learners?". Atlantis, 7-21.

Polat, Nihat, Laura Mahalingappa & Rae L. Mancilla (2019). "Longitudinal growth trajectories of written syntactic complexity: The case of Turkish learners in an intensive English program". Applied Linguistics, 1-25. doi:10.1093/applin/amz034

Polat, Brittany & Youjin Kim (2013). "Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development". Applied Linguistics, 35(2), 184–207. doi:10.1093/applin/amt013

Phuong, Bui Hue (2018). "Can using picture description in speaking sessions help improve EFL students' coherence in speaking?". European Journal of Foreign Language Teaching, 3(3), 33-51.

Riazi, A. Mehdi (2016). "Comparing writing performance in TOEFL-iBT and academic assignments: An exploration of textual features". Assessing Writing, 28(1), 15–27.

Salamoura, Angeliki (2008). "Aligning English Profile research data to the CEFR". Research Notes, 33, 5–7.

Serrano, Raquel, Elsa Tragant & Àngels Llanes (2012). "A longitudinal analysis of the effects of one year abroad". Canadian Modern Language Review, 68(2), 138-163. doi:10.3138/cmlr.68.2.138

Shooraki, Farah, Hossein Barati & Ahmad Moinzadeh (2020). "Assessing writing performance in TOEFL-iBT: A closer look at linguistic and discoursal features". English Text Construction, 13(1), 84

Skehan, P. (1996). "A framework for the implementation of task-based instruction". Applied Linguistics, 17, 38-62.

Spada, Nina & Yasuyo Tomita (2010). "Interactions between type of instruction and type of language feature: A meta-analysis". Language Learning, 60(2), 1–46.

Springer, Philip Ernest (2012). Advanced Learner Writing: a corpus-based study of the discourse competence of Dutch writers of English in the light of the C1/C2 levels of the CEFR. Doctoral dissertation, Vrije Universiteit, Amsterdam.

Stockwell, Glenn & Michael Harrington (2003). "The incidental development of L2 proficiency in NS-NNS email interactions". CALICO Journal, 20, 337–359.

Tannenbaum, Richard J. & E. Caroline Wylie (2005). "Mapping English language proficiency test scores onto the Common European Framework". ETS Research Report Series, 2005(2), I-46. doi:10.1002/j.2333-8504.2005.tb01995.x

Tobin, Gerard A. & Cecily M. Begley (2004). "Methodological rigour within a qualitative framework". Journal of Advanced Nursing, 48, 388–396. doi:10.1111/j.1365-2648.2004.03207.x

Torres, Lourdes & Potowski, Kim (2008). "A comparative study of bilingual discourse markers in Chicago Mexican, Puerto Rican, and MexiRican Spanish". International Journal of Bilingualism, 12(4), 263-279. doi:10.1177/1367006908098571

Urkun, Zeynep (2008). "Re-evaluating the CEFR: an attempt". In The Common European Framework of Reference for Languages (CEFRL): Benefits and Limitations. IATEFL TEA SIG Croatia Conference Proceedings. Canterbury: 10-14.

Valax, Philippe (2011). The Common European Framework of Reference for Languages: A critical analysis of its impact on a sample of teachers and curricula within and beyond Europe. Doctoral dissertation, University of Waikato, Hamilton, New Zealand. Retrieved from https://hdl.handle.net/10289/5546

Vercellotti, Mary Lou (2018). "Finding variation: Assessing the development of syntactic complexity in ESL Speech". International Journal of Applied Linguistics, 29(2), 233-247. doi:10.1111/ijal.12225

Verspoor, Marjolijn, Wander Lowie & Marijn Van Dijk (2008). "Variability in second language development from a dynamic systems perspective". The Modern Language Journal, 92(2), 214-231. doi:10.1111/j.1540-4781.2008.00715.x

Verspoor, Marjolijn, Monika S. Schmid & Xiaoyan Xu (2012). "A dynamic usage based perspective on L2 writing". Journal of Second Language Writing, 21(3), 239-263. doi:10.1016/j.jslw.2012.03.007

Waller, Daniel (2015). Investigation into the features of written discourse at levels B2 and C1 of the CEFR. Doctoral dissertation, University of Bedfordshire, Luton.

Weir, Cyril (2005). "Limitations of the Common European Framework for developing comparable examinations and tests". Language Testing, 22(3), 281-300. http://dx.doi.org/10.1191/0265532205lt309oa

Woolford, Ellen (1983). "Bilingual Code-Switching and Syntactic Theory". Linguistic Inquiry, 14(3), 520-536.

Zarco-Tejada, Ángeles, Carmen Noya Gallardo, M. Carmen Ferradá & Isabel Calderón López (2016). "Analysing corpus-based criterial conjunctions for automatic proficiency classification". Journal of English Studies, 14, 215-237. doi:10.18172/jes.3090

Zhang, Meisuo. (2000). "Cohesive features in the expository writing of undergraduates in two Chinese universities". RELC Journal, 30 (1), 61-95.

APPENDICES

Appendix A. Examples of excerpts of all the CEFR levels.

Excerpt 1. *An example excerpt of A1 level*

Richard: Yes. My name is Richard, I'm 49 years old. I am living **here** ***(reference to context)*** in Tübingen a long time ago, 20 years, **and** (***connector***) my education. First, I was at the middle school, I don't know the exactly name, middle school **then** (***connector***) I make a education for profession like an electronic ***and then*** (***connector***) I go studying sport education **and** (***connector***) theology for to become **or** (***connector***) to get a teacher **but** (***connector***) I don't — I didn't go **and** (***connector***) teaching on this profession **because** (***connector***) another profession was offered. I go to an **Unternehmen oder Firma** (***code-switching***) I go to another company in health insurance, to AOK, the name it's called in German.

Excerpt 2. *An example excerpt of A2 level*

Markus: Ok. First, I explain something about me. My name is Markus Weber. I live — I come from Germany, I live in a small village nearby the borderline to Switzerland, it's about **I think** ***(discourse marker***) maybe 10 to 15 kilometres. It's an — **yeah** ***(discourse marker***), 10 to 15 kilometres. I have different ways which I can drive to my workplace. **And** ***(connector***) **so** ***(discourse markers***), **ok** ***(discourse markers***), I have studied in Ravensburg-Weingarten, I was **there** ***(reference to context)*** on a high school of applied science. I studied **there** ***(reference to context)*** mechanical engineering, my main topics **or** ***(connector***) my study — my major field in study was in research and development, especially, or for example I have some — I have the German words **Vorlesungen** ***(code-switching***)?

Excerpt 3. *An example excerpt of B1 level*

Mustafa: … **before** (***connector***) **like** (***discourse marker***) working in a pension **but** (***connector***) you need **like** (***discourse marker***) you **just** (***discourse marker***) tell rooms, prices **and** (***connector***) you slowly **like** (***discourse marker***) try to explain your own hometown. **But** (***connector***) **when** (***connector***) … I work in my own pension **and** (***connector***) hotel **and** (***connector***) **you know** (***discourse marker***), always, **you know** (***discourse marker***), talk with the tourists. That's help, really help **like** (***discourse marker***) both — **like** (***discourse marker***) three countries I've been to **like** (***discourse marker***) …

Excerpt 4. *An example excerpt of B2 level*

OK (*discourse marker*), in Spain, **well (*discourse marker*)**, I can tell my experience. I've

been working in several shops **like (*discourse marker*)** selling clothes **or** (***connector***) — **yeah (*discourse marker*)** … **Here** (***reference to context***), in Spain, job opportunities for students is — are very reduced **and (*connector*) yeah (*discourse marker)*** you can only work on shops, working for long hours **and** (***connector***) getting little money. **Yeah (*discourse marker)***, that's the way it works **and** (***connector***) we accept it **because** (***connector***) there is nothing else.

Excerpt 5. *An example excerpt of C1 level*

… there's very very very few students who go to upper-secondary education, **I mean**, (***discourse marker***) in fact in my school we don't do that, we just do up to 16. **If** (***connector***) they want to do those last 2 years of school, they have to go to a different village **and** (***connector***) finish **there (*reference to context*)**. **But (*connector*)** it's onl I think last year we only sent 3 or 4 students maximum to study **Bachillerato** (***code-switching***), upper-secondary education …

APPENDIX B. Example output on the Coh-Metrix web-tool

Number	Label	Label V2.x	Text	Full description
Descriptive				
1	DESPC	READNP	1	Paragraph count, number of paragraphs
2	DESSC	READNS	13	Sentence count, number of sentences
3	DESWC	READNW	215	Word count, number of words
4	DESPL	READAPL	13	Paragraph length, number of sentences in a paragraph, mean
5	DESPLd	n/a	0	Paragraph length, number of sentences in a pragraph, standard deviation
6	DESSL	READASL	16.538	Sentence length, number of words, mean
7	DESSLd	n/a	15.25	Sentence length, number of words, standard deviation
8	DESWLsy	READASW	1.363	Word length, number of syllables, mean
9	DESWLsyd	n/a	0.662	Word length, number of syllables, standard deviation
10	DESWLlt	n/a	3.949	Word length, number of letters, mean
11	DESWLltd	n/a	2.388	Word length, number of letters, standard deviation
Text Easability Principle Component Scores				
12	PCNARz	n/a	1.288	Text Easability PC Narrativity, z score
13	PCNARp	n/a	89.970	Text Easability PC Narrativity, percentile
14	PCSYNz	n/a	-0.086	Text Easability PC Syntactic simplicity, z score
15	PCSYNp	n/a	46.810	Text Easability PC Syntactic simplicity, percentile
16	PCCNCz	n/a	-0.157	Text Easability PC Word concreteness, z score
17	PCCNCp	n/a	44.040	Text Easability PC Word concreteness, percentile
18	PCREFz	n/a	0.725	Text Easability PC Referential cohesion, z score
19	PCREFp	n/a	76.420	Text Easability PC Referential cohesion, percentile
20	PCDCz	n/a	0.443	Text Easability PC Deep cohesion, z score

CURRICULUM VITAE

Contact Information: aysegizemciftci@gmail.com

Work Experience:

2020-Present	İstanbul Medipol University English Preparatory School (English Language Instructor)
2017-2020	İstanbul Gelişim University English Preparatory School (English Language Instructor- Material Design and Curriculum Development Unit member)

Education:

Master's Degree:	Kocaeli University English Language Teaching (2018-2021)
Bachelor's Degree:	Anadolu University English Language Teaching (2013-2017)
Erasmus+ Program:	Janusz Korczak Pedagogical University English Language Teaching (2015)
High School:	Edirne 80. Yıl Cumhuriyet Anadolu Lisesi (2009-2013)

Research:

2019	Teacher Beliefs and Practices on Grammar Teaching, and Reasons Behind it: A Study of Teachers at English Preparatory Classes at University
2019	Examining the Effect of Rater Training: A Study of Teachers of English at a Preparatory School in Turkey
2019	Building Awareness of World Englishes Among Adult Students in Higher Education

Certifications:

2020	Cambridge University Press Strategies for teaching grammar online by Gaby Lawson

2020	Cambridge University Press Methodologies and materials in English for Academic Purposes by Helen Basturkmen
2020	Macmillan Education Feedback that empowers
2019	INGED International ELT Conference (Paper presentation)
2012	European Union Certificate European Union Youth Pass Certificate for the project "Dünyamız Ağlamasın"