T.C.
ERCIYES UNIVERSITY
INSTITUTE OF EDUCATIONAL SCIENCES
DEPARTMENT OF FOREIGN LANGUAGES EDUCATION

‘’WE WILL BE DISCUSSING…’’ A LEXICAL BUNDLE APPROACH TO EXPLORING TEACHING PRACTICES IN EMI LECTURES IN TÜRKİYE

Prepared by
ÖZGÜR REHA ALICI

Master’s Thesis

July 2024
KAYSERİ

T.C.
ERCIYES UNIVERSITY
INSTITUTE OF EDUCATIONAL SCIENCES
DEPARTMENT OF FOREIGN LANGUAGES EDUCATION

‘’WE WILL BE DISCUSSING…’’ A LEXICAL BUNDLE APPROACH TO EXPLORING TEACHING PRACTICES IN EMI LECTURES IN TÜRKİYE

Prepared by
ÖZGÜR REHA ALICI

Supervisor
ASSOC. PROF. DR. ERDEM AKBAŞ

July 2024
KAYSERİ

BİLİMSEL ETİĞE UYGUNLUK

Erciyes Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel etik kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin bütününü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir bilimsel çalışması olarak sunmadığımı

beyan ederim.

Özgür Reha ALICI

İmza

YÖNERGEYE UYGUNLUK

''We will be discussing'': A lexical bundle approach to exploring teaching practices in EMI classes in Türkiye adlı Yüksek Lisans tezi, Erciyes Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi ’ne uygun olarak hazırlanmıştır.

Hazırlayan	**Danışman**
Özgür Reha ALICI	Doç. Dr. Erdem AKBAŞ

KABUL VE ONAY

Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Özgür Reha ALICI'nın hazırladığı **''We will be discussing'': A lexical bundle approach to exploring teaching practices in EMI classes in Türkiye** başlıklı bu çalışma jürimiz tarafından **Yabancı Diller Ana Bilim Dalı, İngiliz Dili Eğitimi Bilim Dalında Yüksek Lisans Tezi** olarak kabul edilmiştir.

Jüri Başkanı	Doç.Dr. Elif TOKDEMİR DEMİREL	……..
Jüri Üyesi (Danışman)	Doç. Dr. Erdem AKBAŞ	……..
Jüri Üyesi	Dr. Öğr. Üyesi Ahmet ŞAHAN	……..

Bu tez Erciyes Üniversitesi Lisansüstü Eğitim, Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından 17 / 07 / 2024 tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca … / … / …… tarihi ve …….. sayılı karar ile kabul edilmiştir.

Prof. Dr. Hüseyin ARAK
Eğitim Bilimleri Enstitüsü Müdürü

ACKNOWLEDGEMENTS

First, I would like to start by expressing my gratitude to my thesis supervisor, Assoc. Prof. Dr. Erdem AKBAŞ, who has always been an inspiration for me during my MA career. It has been a privilege for me to be one of his students.

Second, I am grateful to Sinem Bilgin YÜCEL for her contribution to transcribing our corpus data.

Besides, I am grateful to my dear parents Elif ALICI and Halil İbrahim ALICI who have always supported me in every stage of my life.

Moreover, I would like to send my deepest thanks to my lovely son, Can Pars ALICI, his presence has been a constant source of joy and motivation for me.

Finally, I would like to express my heartfelt thanks to my dear wife, Neriman ALICI, for helping me with her love and patience throughout this thesis adventure. It would have been impossible for me to complete this thesis without her precious support.

Özgür Reha ALICI

July 2024, KAYSERİ

ÖZ

"…TARTIŞIYOR OLACAĞIZ" TÜRKİYE'DE EĞİTİM DİLİ İNGİLİZCE (EDİ) DERSLERİNDE SÖZCÜK ÖBEKLERİNİN İNCELENMESİ

Özgür Reha ALICI

Erciyes Üniversitesi, Eğitim Bilimleri Enstitüsü

Yüksek Lisans Tezi, Temmuz, 2024

Danışman: Doç. Dr. Erdem AKBAŞ

İngilizcenin uluslararası bir dil olarak tüm dünyada yaygınlaşmasına paralel olarak Eğitim Dili İngilizce (EDİ) programlarının önem kazandığı yaygın olarak kabul edilmektedir. Bu bağlamda, bu tez, Türkiye'deki yükseköğretim düzeyindeki EDİ derslerindeki sınıf söylemini analiz etmeyi ve öğretim elemanlarının EDİ derslerinde sözcük öbeklerinin kullanımını araştırmayı amaçlamaktadır. Bu amaçla araştırmacı tarafından, çevrimiçi bir platform olan ODTÜ Açık Ders Malzemelerinden, sözcük öbeklerinin kullanımını analiz etmek amacıyla bir derlem hazırlanmıştır. Bu anlamda derlemdeki sözcük öbekleri hem yapısal ve işlevsel özellikleri hem de kullanım sıklıkları açısından analiz edilmiştir. Bu amaçla öncelikle Biber ve diğerlerinin (2004) yapısal ve fonksiyonel çerçeveleri temel alınarak sözcük öbekleri belirlenmiş ve kategorize edilmiştir. Çevrimiçi platformdan manuel olarak hazırlanan derlem, AntConc (Anthony, 2024) programı kullanılarak analiz edilmiştir. İkinci olarak, en sık kullanılan sözcük grupları belirlenip analiz edilmiştir. Son olarak, belirlenen sözcük öbeklerinin yapısal ve işlevsel özellikleri, üniversite düzeyinde EDİ eğitmenleri için pedagojik çıkarımlar oluşturmak amacıyla incelenmiştir. Bu anlamda, bu çalışma mevcut literatüre katkıda bulunmayı amaçlamakta ve üniversite düzeyinde EDİ eğitmenlerinin sınıflarında dili nasıl kullandıklarını anlamamıza yardımcı olmaktadır. Buna ek olarak, bu derlem çalışması, EDİ eğitmenlerinin sınıflarında kullandıkları sözcük öbeklerini desteklemeyi ve onlar için pedagojik çıkarımlar sağlamayı amaçlamaktadır.

Anahtar sözcükler: EDİ dersleri, sözcük öbekleri, sınıf söylemi

ABSTRACT

"WE WILL BE DISCUSSING…" A LEXICAL BUNDLE APPROACH TO EXPLORING TEACHING PRACTICES IN EMI LECTURES IN TÜRKİYE

Özgür Reha ALICI

Erciyes University, Institute of Educational Sciences

Master's Thesis, July, 2024

Supervisor: Assoc. Prof. Dr. Erdem AKBAŞ

It is widely acknowledged that English medium instruction (EMI) programmes have gained importance in parallel with the spread of English as an international language all over the world. In this regard, this study involved the analysis of the classroom discourse in EMI lectures in Türkiye at tertiary level education and investigated the use of lexical bundles in EMI lectures by instructors. To achieve this, the researcher collected data manually from an online platform, METU Open Course Ware, and a corpus was compiled in order to analyse the use of lexical bundles in respect of their structural and functional features as well as their frequency. For this purpose, first, lexical bundles were identified and categorized based on Biber et al.'s (2004) structural and functional frameworks. The manually compiled corpus of the research, Disciplinary English Medium Instruction (DEMI), was analysed by using the concordance programme AntConc (Anthony, 2024). Second, the most commonly used lexical bundles were identified and analysed in terms of their frequency. Finally, the identified lexical bundles' structural and functional features were explored in order to establish pedagogical implications for EMI lecturers at university level. Through this process, it is intended that the findings of the study will contribute to the existing literature and help us to understand the ways in which EMI lecturers use language in their lectures at university level. In addition, this corpus-based study is designed to support EMI lecturers' use of lexical bundles in their lectures and offer them pedagogical implications.

Keywords: *EMI lectures, lexical bundles, classroom discourse*

TABLE OF CONTENTS

BİLİMSEL ETİĞE UYGUNLUK ..... **i**
YÖNERGEYE UYGUNLUK ..... **ii**
KABUL VE ONAY ..... **iii**
ACKNOWLEDGEMENTS ..... **iv**
ÖZ ..... **v**
ABSTRACT ..... **vi**
TABLE OF CONTENTS ..... **vii**
LIST OF TABLES ..... **ix**
LIST OF FIGURES ..... **x**
LIST OF ABBREVIATIONS ..... **xi**
CHAPTER I **1**
INTRODUCTION **1**
CHAPTER II **8**
LITERATURE REVIEW **8**
2.1. Introduction 8
2.2. English Medium Instruction (EMI) 8
2.3. Classroom Discourse in EMI Settings 11
2.3.1. Teacher Talk 12
2.4. Lexical Bundles 12
2.5. Earlier studies on EMI and lexical bundle use in EMI classrooms 14
2.5.1. 'If you look at..': Lexical Bundles in University Teaching and Textbooks studied by Biber et al. (2004) 15
2.6. Summary and Research Gap 17
CHAPTER III **19**
METHODOLOGY **19**
3.1. Introduction 19
3.2. Research Design 19
3.3. Data 20
3.4. Data collection 21
3.5. Data Analysis 22
3.5.1. The structural and functional frameworks of Biber et al. (2004) 25
CHAPTER IV **28**

FINDINGS **28**
4.1. Introduction 28
4.2. An overview of the most frequent lexical bundles in the DEMI Corpus 28
4.3. The structural classification of the lexical bundles 33
4.3.1. Verb phrase fragments 35
4.3.2. Noun and prepositional phrase fragments 37
4.3.3. Dependent clause fragments 39
4.4. The functional classification of lexical bundles 41
4.4.1. Discourse organisers 41
4.4.2 Stance expressions 44
4.4.3 Referential expressions 46
CHAPTER V **48**
DISCUSSION, CONCLUSION AND IMPLICATIONS **48**
5.1 Introduction 48
5.2. Discussion of the findings 48
5.3. Conclusion 51
5.3.1 Limitations of the research and suggestions for future studies 51
5.3.2. Pedagogical implications 52
REFERENCES **53**
APPENDICES **58**
ÖZGEÇMİŞ **59**

LIST OF TABLES

Table 1.DEMI corpus main fields 21
Table 2.DEMI corpus subjects 21
Table 3.Structural taxonomy of Biber et al. (2004) 25
Table 4.Functional framework of Biber et al. (2004) 26
Table 5. The most frequently used 20 four-word lexical bundles in the DEMI corpus 30
Table 6. Frequency of lexical bundles in hard and soft sciences 31
Table 7.Lexical bundles that incorporate verb phrase fragments 37
Table 8.Lexical bundles that incorporate noun and prepositional phrase fragments 38
Table 9.Lexical bundles that incorporate dependent clause fragments 40
Table 10. Discourse organisers in the DEMI corpus 44
Table 11. Stance expressions in the DEMI corpus 45
Table 12. Referential expressions in the DEMI corpus 47

LIST OF FIGURES

Figure 1. Common and unique lexical bundles in the hard and soft sciences 32
Figure 2. The structural classification of four-word lexical bundles 34
Figure 3. Structural types of lexical bundles in hard and soft sciences 35
Figure 4. Functional types of lexical bundles in hard and soft sciences 41
Figure 5. Functional classification of four-word lexical bundles..... 43

LIST OF ABBREVIATIONS

EMI	English as Medium of Instruction
GISAM	Audio-Visual Systems Research and Applications Centre
HE	Higher Education
HEI	Higher Education Institution
METU	Middle East Technical University

CHAPTER I

INTRODUCTION

The present study took a new perspective on English Medium Instruction (EMI) in the Turkish higher education context. The study specifically focused on how lexical bundles are being used by lecturers in the hard and soft sciences in EMI lectures. This opening chapter contains a presentation of the problem statement, the significance of the study, the purpose of the study, the research questions, the assumptions and limitations, and the key definitions.

1.1. Problem of the study

Higher education institutions (HEI) around the world have witnessed a number of changes as a result of the process of internationalization in terms of teaching methods, materials and different mediums of instruction (Akbaş & Bal-Gezegin, 2022). As part of this internationalization process, English started to be more significant in the academic world. As a result of this increased use of English, HEIs have started to implement more EMI programmes to convey academic content. With this in mind, analysing classroom discourse has gained more importance and academics have started to investigate EMI lectures more than before. It can be stated that investigating how the language used in EMI classrooms is useful for both lecturers and students in order to have an insight into the efficiency of the content conveyed in L2.

Alongside growing interest in EMI over recent decades, classroom discourse and the investigation of classroom conversation have gained more importance. Discourse is defined broadly as written or spoken texts which have been produced in a particular context or for a specific purpose (Walsh, 2013, p.23), and discourse analysis is the investigation of the language used in a specific context (Rymes, 2015, p.12). The study of classroom discourse can therefore be defined as exploring the language-in-use during a lecture. Discourse analysis is important to understand how the language is actually used in a classroom context. As a consequence of this rapid

increase, exploring the efficacy of the lecture delivered in L2 has become more significant for lecturers and students.

Focusing on recurring word sequences is one of the ways to investigate the classroom discourse. Sykes (2017) highlights that to learn and use a language it is essential to use formulaic language which includes specified word sequences with prefabricated forms. It is therefore important to explore how formulaic language is used in the lectures. In the literature, researchers have labelled recurring word combinations under different names; "prefabs" (Bolinger, 1976), "lexicalized sentence stems" (Pawley & Syder, 1983), "lexical phrases" (Nattinger & DeCarrico, 1988, 1992), "lexical bundles" (Biber et al., 1999) and "chunks" (Boers & Lindstromberg, 2009). In the current study, the term 'lexical bundle' (Biber et al., 1999) is preferred to use to refer to all such usages.

Lexical bundle studies have gained considerable attention due to their role in facilitating language production and comprehension. Lexical bundles, which are defined as recurring sequences of words (Biber et al., 1999), are essential for communicative competence and discourse organisation. Research on the use of lexical bundles in EMI classrooms have revealed that lexical bundles help lecturers and learners to express complex content knowledge clearly and coherently. Studies have investigated the frequency, distribution and functions of lexical bundles in academic discourse and also aimed to identify the language use in order to use them in teaching practices more effectively.

For instance, Molino (2019) examined the occurrence of lexical bundles in EMI lectures at university setting in Italy. Molino indicated that lexical bundles are worth exploring as they contribute to the fluency of the lecture in academic discourse. The findings of the study revealed that EMI lectures include various structural categories. It is also indicated by the researcher that some of the structural types and discourse functions, which are associated with university teaching contexts such as imprecision bundles, are not used in EMI lectures. Molino (2019) suggests that teacher training programmes could include explicit teaching of lexical bundles to improve the quality of the EMI programmes.

Despite the existing body of research on lexical bundles, the literature on lexical bundle use in spoken academic contexts remains limited. Specifically, there have been few studies that have explored the use of lexical bundles across contexts and disciplines. The research to date has tended to focus on written contexts such as

textbooks, academic written papers and books. This indicates a need to understand the use of lexical bundles in spoken academic discourse, specifically within the EMI contexts. The spoken discourse in EMI classrooms differs significantly from written text in terms of the language used by lecturers to convey complex academic content clearly to facilitate students' learning. In order to develop pedagogical strategies and effective teaching practices, understanding how lexical bundles are used in the spoken academic context can provide useful information. However, literature review on lexical bundles indicates that there is a lack of research on lexical bundles in EMI lectures within the Turkish academic context. As Türkiye is a country which expands its EMI programmes to attract more international students and improve its global academic environment, it is important to fully understand the specific linguistic strategies employed by lecturers in Turkish HEIs. Understanding the use of lexical bundles can help to develop better instructional practices and improve the effectiveness of EMI programmes in Türkiye and in similar educational settings.

Considering these factors, further research is necessary to explore the use of lexical bundles in spoken academic contexts. The purpose of the present study is to fill in the aforementioned gaps in the existing literature by providing a deeper understanding of lexical bundle use in EMI classes. The present study also aims to offer useful information on the use of lexical bundles to imrpove academic communication skills of EMI lecturers and learners. In addition, the current study adopted a novel perspective by analysing the lexical bundles used by EMI lecturers in the hard science and soft science classes within the Turkish higher education context. Through a corpus-based analysis of lexical bundles in classroom discourse, this research seeks to inform language educators, curriculum developers and EMI students by offering recommendations for curriculum design and instructional practices.

1.2. Significance of the study

Researchers have recently shown an increased interest in exploring EMI programmes and EMI lectures. In the literature on EMI, a number of studies were conducted by many researchers but it is seen that these studies have mostly focused on exploring a systematic review of EMI programmes (Macaro et al., 2018), teachers' and students' beliefs and perspectives (Corrales et al., 2016) and the benefits and drawbacks of EMI programmes (Wilkinson, 2013). The research to date has tended to focus on the effectiveness and the pedagogical aspects of EMI programmes rather than

the classroom discourse in EMI settings. In these studies, opinions of teachers and students regarding EMI have been investigated. Furthermore, many researchers have focused on exploring the advantages and disadvantages of EMI programmes. Previous research has established that EMI programmes have some benefits such as improving language proficiency and leading to better international job opportunities. However, the difficulties faced by lecturers and students such as comprehending academic content in L2 and linguistic proficiency problems have been pointed out in some other studies.

Moreover, there are studies in the literature which have examined lexical bundles used by lecturers in EMI settings. It is important to understand structures and functions of lexical bundles in EMI classes that they have significant importance for improving the effectiveness of the lecture. Investigating the ways lexical bundles are used in EMI classes can provide data on the linguistic needs of EMI lecturers and students. Similarly, it can be helpful for both lecturers and students to recognise the linguistic requirements of EMI programmes across various disciplines. With this information in mind, lexical bundles can be used in EMI classrooms more effectively by providing students with the specialised linguistic resources they need to be successful in their subject of study. Lexical bundle analysis can also provide useful information on the language competency levels of EMI students and their capacity for effective participation in academic discourse. Language teachers can also benefit from the results of this research to improve the productivity of academic speech in EMI classrooms. Teachers can effectively support their instruction by understanding the role that lexical bundles play in classroom discourse. Using lexical bundles and providing students with specifically designed vocabulary to improve their language development can be beneficial for both lecturers and learners (Biber & Barbieri, 2007).

Researchers can explore the ways of language use that are specific to various academic contexts by building and investigating specifically designed EMI corpora. Research in corpus linguistics continues to grow due to the increasing number of corpus-based lexical bundle investigations conducted in EMI contexts. In the literature, the majority of previous research has focused on the use of lexical bundles in written contexts. However, the current study focuses on spoken discourse at the university level to fill this gap in the literature. This research aims to map the links between discourse, education and language by using a specifically compiled EMI corpus and analyse the lexical bundles use during EMI classes. It is expected that the

results of this study will make significant contributions to the field of applied linguistics by providing more productive communication in academic contexts.

In summary, the significance of the present research on lexical bundles in EMI classes comes from several key aspects. Initially, it helps us better understanding of the actual language use in classrooms where English is used as the medium of instruction. The study is significant in that it uses a specifically designed corpus from the Turkish higher education context. By using a corpus-based methodology, the present study explores the language use of EMI lecturers in terms of lexical bundle usage. In addition, this study aims to analyse how disciplinary contexts influence language use in EMI classrooms by highlighting similarities and differences in the use of lexical bundles between the hard and soft sciences. This data can be beneficial for the development of specialised instructional tools that focus on the specific needs of both lecturers and students across various academic disciplines.

The studies in the existing literature mostly focus on the written discourse, for this reason it is important to investigate to spoken discourse in academic contexts. By focussing on spoken academic discourse, this research also fills a gap in the literature that has previously focused merely on written discourse. This shift in emphasis is necessary because language used during lectures directly affects students' understanding and attention to the lesson material. Specifically in EMI classes where English is used as the medium of interaction, the way the language used is of importance as the students has different mother tongues and comes from various linguistic backgrounds. Finally, conducting the study within the Turkish higher education context makes the study significant, as there are few studies in the literature. It also contributes to the global understanding of EMI practices and make it possible to compare the Turkish higher education context with different linguistic and cultural settings.

1.3. Purpose of the study

Considering the issues and problems set out above, the purpose of this investigation was to explore the classroom discourse in EMI lectures in the Turkish higher education context. For this purpose, the lexical bundles used by the lecturers were analysed. The research questions which were devised for the study were therefore:

Research Question 1. What are the types and frequencies of lexical bundles used in English Medium Instruction (EMI) lectures at the Turkish university level?

Research Question 2. What are the structural and functional characteristics of the lexical bundles which occurred in EMI lectures in the hard and soft sciences?

Research Question 3. What are the differences between the structural and functional characteristics of the lexical bundles used in the hard and soft sciences?

1.4. Assumptions

Three assumptions were made within the scope of the present study. First, it is expected that the objectives of the study are compatible with the research design which is detailed in the methods section. Second, it is presumed that the data transcription method is valid and reliable. And third, it is considered that the method of analysis and the instruments employed are appropriate for the design of the study.

1.5. Limitations

There are some limitations of this study. First, the study was limited to four major disciplines from the hard and soft sciences: Geometry, Probability and Random Variables, Law and Institutions of the European Union, and History of Art and Architecture. It was, therefore, beyond the scope of the study to comment on other disciplines. For a deeper investigation, the number of the disciplines could be increased. Another limitation of this study is that the data were drawn from lecturers' speech during class and students' comments were not included in the corpus. For this reason, it is limited to teacher talk. In future studies, researchers could include students' comments and incorporate them into their research.

1.6. Definitions of key terms

The discussion in the following chapters can be better understood by referring to the definitions of key concepts.

English Medium Instruction (EMI): The use of the English language to teach academic subjects (other than English itself) in countries or jurisdictions where the first language (L1) of the majority of the population is not English (Macaro et al., 2018, p.37).

Discourse: Written or spoken texts which have been produced in a particular context or for a specific purpose (Walsh, 2013, p.23)

Classroom Discourse: Verbal communication and interaction between students and teachers that occur within the context of a classroom setting (O. Sert, 2015).

Discourse Analysis: Discourse analysis is the investigation of the language use in a specific context (Rymes, 2015, p.24).

Lexical Bundle: Recurrent expressions, regardless of their idiomaticity and regardless of their structural status (Biber et al., 1999, p.990).

CHAPTER II

LITERATURE REVIEW

2.1. Introduction

In this chapter, I shall provide background information by focusing on earlier studies of EMI, classroom discourse in EMI settings and lexical bundles. First, EMI will be discussed in terms of its growing use across the world and also in Türkiye. Then classroom discourse in EMI settings will be examined. Finally, the definition of lexical bundles and the use of lexical bundles in EMI settings will be scrutinized.

2.2. English Medium Instruction (EMI)

As stated in the previous chapter, EMI can be defined simply as ”the use of the English language to teach academic subjects (other than English itself) in countries or jurisdictions where the first language (L1) of the majority of the population is not English” (Macaro et al., 2018, p.37). Toth and Paulsrud (2017, p.189) defined EMI as “is a model of content instruction in contexts where English is not a majority language but is nonetheless used as a language for teaching and learning”. It therefore refers to the use of English when teaching academic subjects rather than teaching English as a subject. The main purpose of EMI is to expand students’ academic subject knowledge by using English as the medium of communication (Taguchi, 2014, p.89). These definitions are important for understanding the conceptual difference between EMI and Content and Language Integrated Learning (CLIL). To make the scope of EMI clear, Pecorari and Malmström (2018, p.499) identified the characteristics of an EMI setting:

1) English is the language used for instructional purposes.

2) English is not itself the subject being taught.

3) Language development is not a primary intended outcome.

4) For most participants in the setting, English is a second language (L2).

The spread of English is inseparable from globalization (Hüppauf, 2004). In recent years, EMI has gained much popularity and importance as a result of the use of

English as the medium of instruction instead of being taught as a foreign language because of its evolving global status (Dearden, 2014). Globally, a rapidly growing number of HEIs are implementing EMI programmes and offering education through the medium of English (Earls, 2016; Lasagabaster et al., 2014). The reasons why EMI has been becoming popular among HEIs can be listed as "internationalization, student exchanges, teaching and research materials, graduate employability and market in the international students" (Coleman, 2006). Similarly, N. Sert (2008) stated that for sociological and economic reasons, the number of private HEIs implementing EMI programmes has increased rapidly. With the globalization of English as a lingua franca, EMI has started to attract more HEIs, lecturers and students than ever. The rapid growth of EMI has made the stakeholders of education focus on this area more. As a consequence of this growth all around the world, HEIs started to see EMI as a profitable idea for both economic and educational reasons. Lecturers wanted to keep up with the educational trend of the time by teaching in the academic lingua franca and students also desired to gain the advantage of being taught in English. Taking into account these reasons, HEIs view EMI as a way to internationalize the education which they provide to their potential students and advance them by using the benefits listed above. Galloway and Rose (2015, p.230) stated that many HEIs seek to apply an EMI policy which enables:

- an increase in the international student population;
- an increase in the internationalization of faculties;
- an increase in the reputation and presence of the institution in the global market; and
- an increase in student mobility options.

It can be found in a number of studies in the literature on EMI that implementing an EMI programme has positive motivations for both lecturers and students. Lasagabaster (2022) stated that many lecturers express strong enthusiasm for implementing EMI programmes considering it as a way to internationalize their universities. In addition, EMI offers lecturers easier access to content materials, better English language competency and more international student involvement (Lasagabaster, 2022). It was pointed out by Corrales et al. (2016) that EMI helps students to prepare themselves for their future professional careers, improve their language skills and interact with international students in their own universities and

universities in other countries. N. Sert (2008) similarly commented that EMI helps students to improve their language proficiency. Even so, it has also been noted by researchers that EMI can have some drawbacks for students in addition to the benefits mentioned above, such as insufficient linguistic knowledge, inability to understand the academic content in L2 and lack of motivation (Coleman, 2006; Corrales et al., 2016; N. Sert, 2008). Similarly, Bradford (2019) commented that EMI lecturers spend much more effort teaching in English than in their L1, and Lasagabaster (2022) said that academic English can also be a problem even for students with higher English proficiency. It can therefore be stated that EMI lecturers and students have both advantages and disadvantages of EMI classes. All in all, it can be said that EMI is becoming increasingly popular around the world with the help of the factors discussed above.

As EMI is growing rapidly in popularity all over the world, it started to be investigated by researchers. Dearden (2014) conducted research on EMI in 55 countries with the aim of creating a global map of EMI. The research examined the reasons behind the use of EMI and the effects of teaching and learning through EMI. The results showed that there is a rapidly expanding general trend towards EMI and that this trend is also supported by policymakers. However, the study also revealed some drawbacks of EMI programmes: many of the countries did not have enough linguistically qualified teachers and there was a lack of teacher training programmes for EMI.

Macaro et al. (2018) stated that EMI is a rapidly growing phenomenon, especially in higher education contexts all over the world. In their systematic review of EMI, they analysed 83 studies in HEIs in terms of the growth of EMI, teacher and student beliefs on EMI and the effects of teaching academic subjects through EMI. The findings of their study showed that the use of EMI is rapidly growing in all parts of the world, especially among HEIs. In addition, they highlighted that teachers and students have positive motivations for EMI such as improving their linguistic skills, accessing lesson materials more easily and having better job opportunities. However, the results also showed that teachers and students have negative motivations for EMI, such as insufficient academic language proficiency for both lecturers and students and extra workload for lecturers.

As for the implementation of EMI programmes, it can be stated that the rapid growth of EMI is nearly the same in Türkiye as in other countries. There has been a

significant increase in the number of higher education institutions which apply EMI programmes in Türkiye as well. As a consequence of this rapid growth, a number of studies have been carried out to explore the effectiveness and implementation of EMI programmes. N. Sert (2008) conducted a case study to analyse the effectiveness of EMI in the Turkish context. The results of the research indicated that in terms of teching English EMI is more effective than the formal English instruction. In addition, EMI is quite effective in terms of teaching academic content and improving the language skills of the students at the same time. Similarly, Ekoç (2020) indicated that EMI students have an instrumental motivation as they want to be part of the international community and find better jobs in the international market with the help of English language proficiency and content knowledge. However, both studies also underlined that students have some linguistic problems and difficulties in understanding the academic content in English and in understanding lecturers.

2.3. Classroom Discourse in EMI Settings

Classroom discourse refers to the verbal communication and interaction between students and teachers which occur within the context of a classroom setting (O. Sert, 2015). It encompasses the exchange of spoken language between instructors and students during the process of teaching and learning. Classroom discourse involves various linguistic features, communication strategies and interactional patterns which shape the dynamics of classroom interactions. Classroom discourse in EMI classrooms represents dynamic environments where academic content is delivered and discussed in English, fostering both language acquisition and content learning simultaneously. Understanding the characteristics of classroom discourse in EMI settings is essential for unravelling the intricate dynamics of language use and communication strategies employed by both instructors and students. Classroom communication is complex and difficult to comprehend in terms of the language being used and can have different functions such as checking comprehension, assessing learning and giving advice (Walsh, 2011). It is therefore reasonable to argue that any effort to enhance teaching and learning should start by examining classroom interaction given its complexity and importance because language is used in every situation that takes place in the classroom (Walsh, 2011).

Instructors in EMI classrooms employ various linguistic features to convey information effectively and engage students in the learning process. Discourse markers

such as ‘however’, ‘therefore’ and ‘in conclusion’ are commonly used to signal transitions between ideas and highlight key points (Jenkins, 2006). Metadiscourse markers, including expressions of certainty, probability and doubt, serve to guide students’ comprehension and facilitate interaction in academic discourse (Hyland, 2005). In a classroom, the language in use shapes how we learn new information, gain new abilities, comprehend issues, handle breakdowns in communication, and build and preserve relationships (Walsh, 2011). In order to ensure that students comprehend what is being taught and to promote efficient instruction, classroom conversation plays an essential role in EMI classrooms. There has, however, been relatively little research on how lecturers create and manage classroom discourse in the Turkish EMI environment (Akbaş & Bal-Gezegin, 2022).

2.3.1. Teacher Talk

The language used by the teacher during a lesson is known as teacher talk. In EMI contexts, learners are exposed to teacher talk in a language which is not their mother tongue. In EMI classrooms in which the primary teaching focus is on the disciplinary content, the way the teacher uses the language is important as it was noted by Sinclair and Coulthard (1975) that teacher talk covers two-thirds of the lesson. Especially in EMI classrooms, where a language other than students’ L1 is used, teacher talk serves as the primary medium of instruction. Nicaise (2020) stated that teacher talk is an effective pedagogical tool to promote learning opportunities when it is used strategically. Teacher talk plays an important role in a classroom as it shapes students’ level of understanding and motivation during a lesson (Gharbavi & Iravani, 2014). In order to create an efficient learning atmosphere, teachers should be aware of the importance of their language use. Teacher awareness of the function of specific teacher talk strategies and their potential for supporting students’ conceptual development will assist teachers in both planning lessons and acting responsively in teacher/student interactions (Sharpe, 2008).

2.4. Lexical Bundles

The study of multi-word units, including collocations, fixed phrases and other frequently occurring word combinations which display particular grammatical, semantic and lexical characteristics is known as phraseology. Phraseology is defined as “the co-occurrence of a form or a lemma of a lexical item and one or more additional

linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance" (Granger & Meunier, 2008). The study of multi-word expressions can be traced back to the beginning of the last century. Jespersen (2013) and Firth (1951) were the first researchers to conduct research on fixed multi-word expressions. During the early period of phraseology research, analytical tools were limited and as a result, researchers could explore relatively small amounts of data, but researchers have more recently been able to explore huge amounts of data by the advances in technology over recent decades (Salazar, 2014). Since that change, there has been a growing interest in studies exploring how words are used together (Salazar, 2014). As already explained, researchers have labelled these multi-word expressions under different names: 'collocations' (Firth, 1951), 'lexicalized sentence stems' (Pawley & Syder, 1983), 'lexical phrases' (Nattinger & DeCarrico, 1988, 1992), 'lexical bundles' (Biber et al., 1999), 'formulaic language' (Schmitt & Carter, 2004) and 'chunks' (Boers & Lindstromberg, 2009).

The term 'lexical bundles' was first introduced by Biber et al. (1999, p.990) and described as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status". Similarly, Cortes (2004) defined lexical bundles as "extended collocations of three or more words that statistically co-occur in a register". In addition, Biber and Conrad (1999) defined lexical bundles as collocations of the most frequent recurring lexical items which are used statistically more frequently than others. Biber et al. (2004) investigated the use of lexical bundles in tertiary-level education and compared the use of lexical bundles in classroom teaching and textbooks and found that lexical bundles are used in classroom teaching more than in academic writing or textbooks and during the classroom teaching period, many lexical bundles are used. That study also developed a functional framework for the identification of lexical bundles in discourse.

This current study benefits from both Biber et al.'s (2004) structural and functional frameworks for the categorisation of lexical bundles. Biber et al. (2004) showed that lexical bundles have particular structural functions which determine how they are categorized under three main categories; *verb phrase*, *dependent clause* and *noun/prepositional phrase lexical bundles.* In addition to the structural categories of lexical bundles, a framework for their functions was also developed by Biber et al.

(2004). In this framework, lexical bundles are categorized under three functional groups; *stance expressions*, *discourse organisers* and *referential expressions.*

2.5. Earlier studies on EMI and lexical bundle use in EMI classrooms

Cortes (2004) investigated the use of four-word lexical bundles in academic writing by comparing published works from the history and biology disciplines. The research first focused on history and biology journals to identify the most frequently used four-word lexical bundles and created a target lexical bundle list. Then students' use of the identified target lexical bundles in their academic writing was explored. The findings indicated that the target lexical bundles were rarely used by the students in their own academic writings and that the lexical bundles used by the students and authors of the journals did not match at all. Considering the findings, Cortes (2004) stated that, simple exposure to frequently used lexical bundles is not sufficient alone and that students should be introduced to the functions of lexical bundles in academic contexts. In addition, students should be able to use the academic lexical bundles in an appropriate way if they are explicitly introduced to them.

Nesi and Basturkmen (2006) conducted a study on four-word lexical bundles and examined 160 university lectures to explore the cohesive role of lexical bundles. The corpus of the study compiled from two different online corpora: the British Academic Spoken English (BASE) corpus and the Michigan Corpus of Academic Spoken English (MICASE). The aim of the study was to investigate the discourse signalling role of lexical bundles. According to the findings of their study, Nesi and Basturkmen (2006) emphasized that lexical bundles help creating cohesion during lectures and contribute to facilitate comprehension. They noted that it is essential for language learners to be informed of the significance of discourse signalling role of lexical bundles. In order to facilitate learning in the classroom contexts, it would be useful to make students aware of the lexical bundles.

Kashiha and Chan (2013) investigated the use of four-word lexical bundles in academic lectures in the hard and soft sciences from the British Academic Spoken English (BASE) corpus. The aim of their study was to identify the four-word lexical bundles in the BASE corpus and explore the frequency, structure and function of the identified bundles. The selected data from the BASE corpus included two broad fields of sciences: hard sciences and soft sciences. The researchers adopted Biber et al.'s (2004) structural and functional frameworks to classify the lexical bundles as those

frameworks were well designed to analyse spoken data. The findings of the study showed that lexical bundles are frequently employed in academic lectures. Additionally, lecturers in the hard and soft sciences used lexical bundles in a variety of forms and functions. It was found that lecturers in hard sciences used more lexical bundles than the lecturers in soft sciences. In addition to the frequency, the researchers also identified some similarities and differences. Dependent clause fragments were most common in the hard sciences whereas noun and prepositional phrase fragments were most common in the soft science corpus. The functional analysis of the lexical bundles used in those academic lectures showed that whereas the lecturers in the hard sciences tended to use more discourse organisers, there was more emphasis on referential expressions in the soft sciences. The findigs highlighted the differences between hard and soft sciences in terms of using lexical bundles. The results indicated that it is important to raise students' awareness about lexical bundles as they are frequently used in academic lectures.

In conclusion, this literature review has shown that earlier studies of EMI were mostly about general pedagogical aspects. They mostly highlighted lecturers' and students' perceptions and attitudes, as well as the positive and negative effects of EMI programmes on lecturers and students. Furthermore, they mostly outlined the difficulties and shortcomings of EMI programmes and made suggestions for resolving the problems and deficiencies. Furthermore, the previous studies mostly outlined the difficulties and shortcomings of EMI programmes and made suggestions to solve the problems and deficiencies. In addition to these, from the lexical bundle use perspective, the existing literature shows that most of the studies focused on the written context and few studies investigated the use of lexical bundles in spoken academic discourse specifically in EMI contexts. Literature review also showed that there are few studies in terms of lexical bundle use in EMI lectures in the Turkish higher education context. In this sense, the present study aims to analyse the language use in EMI classrooms by applying a corpus-based methodology.

2.5.1.'If you look at..': Lexical Bundles in University Teaching and Textbooks studied by Biber et al. (2004)

Biber et al. (2004) investigated the use and functions of lexical bundles in university teaching and textbooks across various disciplines. The purpose of the study

was to explore the occurrence and usage patterns of lexical bundles in university contexts. Biber et al. (2004) stated that classroom discourse is important in academic contexts but that very little is known about it. Moreover, few studies have been conducted to investigate the linguistic features of spoken academic discourse. According to Biber et al. (2004), our daily language use consists of a number of fixed linguistic items so the researchers sought to investigate the use of lexical bundles in university contexts in order to identify the characteristics of language use by adopting a frequency-driven approach. The study was conducted on data from the university context to analyse the use of lexical bundles; it was based on the T2K-SWAL (TOEFL 2000 Spoken and Written Academic Language Corpus). T2K-SWAL is a specially designed corpus to exemplify the spoken and written language use of university students in the United States. The researchers compiled a large corpus of academic discourse samples across many disciplines, including the natural sciences, social sciences and humanities, and professional fields.

In analyses of lexical bundles, there is no consensus among researchers on the frequency cut-off point for their use. Biber et al. (2004) set the frequency cut-off point to 40 times per million words. In addition to considering frequency, it is important to establish a range of criteria for the analysis in order to detect and ignore idiosyncratic use by speakers (Biber et al., 2004). In the study, a word sequence that is represented in at least five texts was counted as a lexical bundle. They stated that spoken university registers are different from written registers in terms of grammatical characteristics. For instance, verbs and pronouns are much more common in the spoken register than in the written register. Likewise, spoken registers have a wider range of lexical bundles than written registers. They also mentioned that classroom discussions contain a wider variety of lexical bundles.

As a consequence of the analysis of lexical bundles, the researchers identified three principal structural types of lexical bundle category; verb phrase fragments, dependent clause fragments and phrasal expressions. Verb phrase fragments can be a subject pronoun followed by a verb phrase (for example *it's going to be*), a discourse marker followed by a verb phrase (for example *you know this is*), a verb phrase (for example *is going to be*) and question fragments (for example *what do you think*). Dependent clause fragments are lexical bundles with the main clause followed by a complementizer or beginning with a complementizer such as *I want you to*, *if we look at* and *what I want to*. On the other hand, phrasal expressions are noun phrase

components and prepositional phrase components such as *of the things that* and *the end of the*. In addition to the structural characteristics, the researchers introduced a framework for the functional categorisation of lexical bundles. They identified three functional categories of lexical bundles; stance expressions, discourse organisers and referential expressions. Stance bundles reflect beliefs or levels of confidence. The relationships between previous and upcoming discourses are reflected through discourse organisers. Referential bundles directly refer to concrete or intangible objects, or the textual material itself, either to name an object or to highlight a specific, crucial quality of it. In addition, these three categories have more specific sub-categories. A table showing the categories and sub-categories of structural and functional taxonomies is shown in the methodology session below.

It was found by the researchers that lexical bundles are widely used in university teaching and textbooks across all disciplines, indicating their importance in academic discourse. In addition, the researchers stated that lexical bundles serve various functions in academic communication, such as organizing information, signalling transitions, summarizing key points and expressing stance or evaluation. The frequency and distribution of lexical bundles were found to vary across disciplines, reflecting disciplinary differences in language use and discourse conventions. The findings showed that some lexical bundles are more common in specific academic fields. Another finding is that the study have several implications for language teaching, curriculum design and academic writing in terms of understanding the structures and functions of lexical bundles. The findings also revealed that the explicit teaching of lexical bundles could help in enhancing the instruction quality with respect to teaching complex academic discourse. Integrating lexical bundles into language instruction can enhance learners' academic language proficiency and communicative competence. The study noted that discipline-specific language awareness in academic writing courses is of crucial importance. To sum up, the findings of the study pointed out the significant role of lexical bundles in academic discourse and their integration in the academic settings.

2.6. Summary and Research Gap

In this chapter, previous studies of the lexical bundles used by lecturers in EMI lectures at the university level have been discueed. These studies explored the use of lexical bundles and their usage in EMI contexts. Through these studies, it is

highlighted that the use of lexical bundles helps lecturers to structure the flow of a lesson and learners to comprehend the academic content more effectively. Previous studies by Biber et al. (2004), Cortes (2004), Kashiha and Chan (2013) emphasized the significant contribution of lexical bundles to the flow of lectures. The previous studies highlighted that lexical bundles facilitate comprehension and interaction during lectures. Moreover, Nesi and Basturkmen (2006) stated that lexical bundles contribute to the cohesion of the lesson as they represent a crucial discourse signalling role.

Hence, the study aims to adress several research gaps. First, while several studies have investigated English-Medium Instruction (EMI) classrooms, they mainly focused on the perceptions of teachers and learners rather than the actual language used in these contexts. Second, existing studies tend to emphasize written academic discourse, leaving a significant gap in our understanding of spoken academic discourse, particularly in EMI settings. Third, there are very few studies investigating the use of lexical bundles in spoken academic discourse, especially in the Turkish higher education context.

Although there is an increasing number of research on lexical bundles, there are few studies exploring the lexical bundles in EMI contexts. In this sense, the present research aims to explore the use of lexical bundles by EMI lecturers in both hard and soft science classes in the Turkish higher education context. The research was designed to investigate the structural and functional characteristics of lexical bundles and their occurrence in the lectures. By focusing on a specifically compiled corpus an academic context, the findings are expected to contribute to the comprehension of the lexical bundle use in EMI contexts.

CHAPTER III

METHODOLOGY

3.1. Introduction

In this chapter, all the processes involved in this study will be described in a detailed manner. First, the research design used for this study will be explained in detail. Following this, the data collection and analysis procedures will also be outlined. Lastly, the frameworks used to categorise lexical bundles will be described.

3.2. Research Design

The current study examines the use of lexical bundles by lecturers in EMI classes by using a corpus-based methodology. Corpus-based research refers to the analysis of large data collections to explore the language use. Conrad (1999) identifies the three key features characterize corpus-based research. First, it is based on corpora, which must be carefully selected to provide representativeness and diversity. The size and diversity of a corpus are crucial concerns in order to obtain reliable results. Second, corpus-based studies use computers for analyses to examine large datasets and complex linguistic features. Third, corpus-based research involves both quantitative analyses to identify patterns and qualitative interpretations to understand the communicative functions of language.

Conrad (1999) noted that conducting corpus-based research enables researchers to investigate multiple factors simultaneously, such as frequency, grammatical structures and semantic features. The corpus-based approach also enables researchers to explore how these factors interact with each other, which can be challenging to examine a huge amount of data manually. Corpus-based studies offer the benefit of providing concrete support and examples to address difficult questions about language use. Rather than relying solely on intuition, educators can find empirical evidence to support their explanations or discover unexpected findings.

Before describing the procedures of the study, it is important to outline the research questions again. The first research question aims to investigate the language

use in the EMI lectures by analysing the most frequently used lexical bundles. The second question aims to explore the structural and functional differences of lexical bundles in EMI lectures. The third question aims to analyse the disciplinal differences of lexical bundles between the hard and soft science. In order to answer these research questions, a data-driven methodology was employed to identify the most frequently used lexical bundles and scrutinize the structural and functional role of these bundles within a spoken corpus of academic English. In the following sections, a more detailed explanation of the methodology applied in this study will be introduced.

Research Question 1. What are the types and frequencies of lexical bundles used in English-medium instruction (EMI) lectures in Turkish higher education institutions?

Research Question 2. What are the structural and functional characteristics of the lexical bundles that occurred in EMI lectures in the hard and soft sciences?

Research Question 3. What are the differences between the structural and functional characteristics of the lexical bundles used in the hard and soft sciences?

3.3. Data

Conrad (1999) recommended that a corpus-based study needs to have a well-designed corpus in order to make it appropriate for the research questions. With this in mind, the corpus of the present study was carefully compiled in order to make it suitable for the purpose of the study. This study was based on a specialised corpus which was compiled from academic lectures at the university level in Türkiye. As stated by Tokdemir Demirel (2023), compiling a specialized corpus is essential for conducting large-scale analyses of English texts or speech with native Turkish speakers. In order to undertake research on spoken English in EMI lectures, it was important to find appropriate data for the purpose of the study. The data for the study were therefore drawn from the Disciplinary English Medium Instruction (DEMI) corpus. In terms of corpus size, Biber (2006b) suggested that "A corpus must be large enough to adequately represent the occurrence of the features being studied". Considering this, the current research corpus was made up of 91 lectures from the DEMI corpus, and the total corpus size was 493,762 words. The DEMI corpus consists of four different subjects: Law and Institutions of the European Union; History of Art and Architecture; Geometry and Probability; and Random Variables, all from two main fields: the hard and the soft sciences. The lecturers in these four disciplines were

native speakers of Turkish and used English as the medium of instruction in their classes. Details of the DEMI corpus, including the total number of files and the number of words, are shown in Tables 1 and 2.

Table 1

DEMI corpus main fields

Domain	*Event type*	*Number of files*	*Number of words*
DEMI_Soft	Lecture	30	229.131
DEMI_Hard	Lecture	61	264.634
Total		91	493.762

Table 2

DEMI corpus subjects

Domain	*Event type*	*Number of files*	*Number of words*
DEMI_LIEU	Lecture	12	82.532
DEMI_HAA	Lecture	18	146.599
DEMI_GEO	Lecture	21	95.530
DEMI_PRV	Lecture	40	169.101
Total		91	493.762

3.4. Data collection

The corpus for the study was compiled from open-access university-level EMI lectures at the Middle East Technical University shared on a publicly available online platform (METU Open-Courseware), which is on YouTube. In order to use the videos

for the purposes of the present study, the required permissions were obtained by contacting the university's related centre, the METU Audio-visual Systems Research and Applications Centre (GISAM). In addition, ethical permissions were sought and obtained in order to use the videos and their transcriptions as the data for the research. Having obtained the necessary permissions, lectures in a single term (fourteen weeks) from four different lecturers were used to compile the corpus of the study. Four lectures from two main fields –two in hard science and two in soft science – were chosen in line with the purpose of the study. They were transcribed manually and added to the corpus of the study by the researcher. In total, the data consisted of 91 lectures and approximately half a million words. The transcriptions were converted into plain text in order to analyse them with a concordance programme named AntConc 4.3.0 (Anthony, 2024). After converting the texts and making the data ready to analyse, all of the files were loaded into the AntConc 4.3.0 software programme. Initially, all 91 files were uploaded and analysed to see the most frequently used lexical bundles in the data. The hard and soft science files were then uploaded and analysed separately in order to see the differences between them in the use of lexical bundles. The results obtained from the AntConc 4.3.0 analysis were then transferred into an Excel file for a detailed analysis and categorisation of the lexical bundles.

3.5. Data Analysis

For the analysis, first, lexical bundles were identified according to the frequency and range criteria which had been set. They were then investigated in terms of their structural and functional features. In order to classify the lexical bundles regarding their structural and functional features, the taxonomy devised by Biber et al. (2004) was used. It is a widely held view that in order for a word group to be regarded as a lexical bundle, it needs to meet various criteria, such as frequency and range in the corpus. For the identification of lexical bundles, different studies have suggested a number of criteria. In the literature, there are different cut-off points for the frequency of lexical bundles ranging from ten (Biber et al., 1999) and twenty (Cortes, 2004; Hyland, 2008a, 2008b) to 40 (Biber & Barbieri, 2007) times per million words. In addition to the frequency of lexical bundles, another criterion is that a lexical bundle should occur in different texts more than a specified number in the related corpus, such as five times (Biber & Barbieri, 2007), or should occur at least in 10% of the texts in

the corpus (Hyland, 2008a). It can be stated that in order to avoid the authors' idiosyncratic uses, it is necessary to set the range criterion.

Samraj (2024) indicated that the identification of the lexical bundles in the corpus varies according to the threshold and dispersion criteria and corpus size. Besides this, when various frequency and range criteria are used for manual bundle extraction, it was seen that the proportion of the lexical bundles can vary in the corpus. In this study, I have carefully considered these factors by using a consistent set of criteria for frequency, range, and bundle extraction. By doing so, the present research aims to offer a more accurate and comprehensive analysis of how lexical bundles are used across various academic disciplines in EMI contexts. This approach enhances the validity of our findings and contributes to a better understanding of the role of lexical bundles in academic discourse. Thus, the way lexical bundles are used in different fields can be greatly affected by the specific methods used to compile and analyse corpora. This emphasizes the need to entirely explain and take into account these methodological choices when conducting corpus-based research, in order to ensure the validity and reliability of any comparisons and conclusions.

It is "somewhat arbitrary" (Biber et al., 2004, p.376) to decide on the actual frequency and dispersion cut-off point for lexical bundle research, but in the light of the previous views, in the present study, the frequency cut-off number was set to the relatively high point of 40 times per million words. In addition, bundles which occurred at least five times in different texts in the corpus were included. Having decided on the frequency cut-off number and the range of the bundles in the corpus, another point to be decided was the length of the lexical bundle. Hyland and Jiang (2018) stated that two-word lexical bundles are too common to analyse, whereas five-word and six-word lexical bundles are too rare to analyse. They therefore suggested that four-word lexical bundles should be analysed in corpus-based studies. Similarly, Cortes (2004) stated that four-word bundles include three-word bundles in their structures and that four-word bundles are more commonly used than five-word bundles. Taking into account the dispersion and frequency criteria of previous studies, the current study focused on analysing four-word lexical bundles in the corpus.

After determining the scope of the research and identifying the lexical bundles by using a software programme, a long list of lexical bundles was obtained. In order to make this list more manageable and pedagogically more effective, I manually excluded some of the lexical bundles following Chen and Baker (2010) and Salazar

(2014). Salazar (2014) stated that there is a need for the manual exclusion of irrelevant items in order to have a better-organised set of lexical bundles for someone who wants to use them for pedagogical purposes. In other words, the manual removal of some items is needed to make the list more effective in terms of reaching the objectives of the research. For these reasons, lexical bundles which were text-dependent (for example *European Court of Justice*), discipline-specific (for example *perpendicular bisector of*) and proper nouns (*the Forum of Trajan*) were manually excluded. In addition to these exclusion criteria, to guard against inflated results caused by lexical bundles ending with articles *a*, *an* and *the* were manually excluded as they represented three-word bundles more than four-word bundles. Salazar (2014) pointed out that four-word lexical bundles ending with articles do not provide any additional information to the three-word bundles, as in the case of *in addition to the* and *in other words the*. After applying these criteria and manual deletions, there remained 122 four-word lexical bundles.

Once the cut-off point, range criteria and bundle length were decided, it was necessary to decide on the framework to categorise the bundles. In order to analyse the lexical bundles in the corpus, the structural and functional taxonomies of Biber et al. (2004) were used because their study was well-designed for the purposes of the present research. According to the structural taxonomy of Biber et al. (2004), there are three main categories and seventeen sub-categories. In the functional taxonomy, they proposed three main categories: stance expressions, discourse organisers and referential expressions. I classified the lexical bundles manually according to the taxonomies shown in detail in Tables 3 and 4.

In order to ensure the reliability of the lexical bundle categorisation, two independent researchers categorised the lexical bundles. To investigate the reliability of the categorisation of the lexical bundles a statistical test was conducted in this study. Cohen's kappa test is a statistical measure that evaluates the level of agreement between raters. Kappa results range from 0 to 1, while 0 represents no agreement between raters, 1 represents a perfect agreement. In this study, Cohen's kappa result showed a high level of agreement between raters with 0.85 value. This high level of agreement between raters ensured the reliability of the lexical bundle categorisation.

3.5.1. The structural and functional frameworks of Biber et al. (2004)

Biber et al. (2004) developed frameworks for the categorisation of lexical bundles according to their structures and functions in a specific discourse. Biber et al.'s (2004) frameworks were designed to investigate the structural and functional characteristics of lexical bundles in academic spoken discourse. Structural framework helps to identify the most frequently used bundle types in academic lectures and the functional framework helps to see the functions of these lexical bundles. The frameworks of Biber et al. (2004) were employed in this study as their study specifically focused on university classroom teaching and were well-designed for the categorisation of spoken data. As the DEMI corpus consists of spoken academic discourse, the present study employed these structural and functional frameworks developed by Biber et al. (2004) which are shown in Tables 3 and 4.

Table 3

Structural taxonomy of Biber et al. (2004)

Structural categories	*Sub-categories*	*Sample bundles*
Verb phrase fragments	a. 1st/2nd person pronoun + VP	*i would like to*
	b. 3rd person pronoun + VP	*it is going to*
	c. Discourse marker + VP	*so that is why*
	d.Verb phrase with (non-passive)	*is going to be*
	e. Verb phrase with passive verb	*can be found by*
	f. yes-no question	*do you know that*
	g. WH-question	*what does that mean*
Dependent clause fragments	a. 1st/2nd person pronoun	*i want you to*
	b. WH- clause	*when you look at*
	c. If-clause	*if you look at*
	d. To-clause	*to be able to*
	e. That-clause	*that you have to*
Noun phrase and prepositional phrase fragments	a. Noun phrase with of-phrase	*the centre of gravity*
	b. Noun phrase with post-modifier	*the point in which*
	c. Other noun phrase	*all the way to*
	d. Prepositional phrase	*at the same time*
	e. Comparative expressions	*bigger than or equal*

The data in Table 3 show three primary structural lexical bundle categories: verb phrase fragments, dependent clause fragments and phrasal expressions. The first type includes verb phrase fragments, starting with a subject pronoun and followed by a verb phrase. Additionally, these bundles can start directly with a VP, such as *let us look at* or *is going to be*, or with question fragments, such as *what does that mean*. The second major structural type is similar to the first in incorporating verb phrase elements, but also includes dependent clause fragments as in *to be able to*, *if you look at*. The third type only includes phrasal components, with most bundles consisting of noun phrase components ending with the beginning of a post modifier such as *at the same time* and *in the case of.*

Table 4

Functional framework of Biber et al. (2004)

Functional categories	*Sub-categories*	*Sample bundles*
Stance expressions	A. Epistemic stance	*I do not know*
	B. Attitudinal/ Modality Stance	
	a. Desire	*I do not want*
	b. Obligation/Directive	*I want you to*
	c. Intention/ Prediction	*is going to be*
	d. Ability	*to be able to*
Discourse organisers	A. Topic Introduction/Focus	*okay so let us*
	B.TopicElaboration/Clarification	*on the other hand*
Referential Expressions	A. Identification/ Focus	*in fact it is*
	B. Imprecision	*or something like that*
	C. Specification of Attributes	
	a. Quantity specification	*less than or equal*
	b. Intangible framing	*in the form of*
	c. Tangible framing	*in the case of*
	D. Time/Place/Text Reference	
	a. Place reference	*in the United States*
	b. Time reference	*at the same time*
	c. Text-deixis	*as shown in figure*
	d. Multi-functional reference	*at the end of*

Biber et al. (2004) identified three primary discourse functions for lexical bundles: stance expressions, discourse organisers and referential expressions. Stance bundles are used to express attitude or certainty and can be divided into two categories: epistemic or attitudinal/modality (Biber et al., 1999). Discourse organiser bundles help to structure the flow of ideas and relationships between previous and upcoming discourse, as well as introduce, expand on and clarify topics. Referential bundles consist of a wide variety of lexical bundles which typically refer to the textual context, physical or abstract entities, such as *the nature of the* and *that's one of the*.

CHAPTER IV

FINDINGS

4.1. Introduction

In this chapter, I shall set out the findings of the research considering the research questions of the study. In order to address the research questions, the findings will be organised into two sections. In the first section, an overview of the most frequently occurring lexical bundles in the DEMI will be analysed. In the second section, the functions and structural characteristics of these bundles will be investigated and interpreted. Before stating the findings of the analysis of this research, a restatement of the research questions which guided the study will be useful:

Research Question 1. What are the types and frequencies of lexical bundles used in English Medium Instruction (EMI) lectures at the Turkish university level?

Research Question 2. What are the structural and functional characteristics of the lexical bundles which occurred in EMI lectures in the hard and soft sciences?

Research Question 3. What are the differences between the structural and functional characteristics of the lexical bundles used in the hard and soft sciences?

To be able to answer these questions, the findings of the research will be organised into two sections. In the first section, an overview of the most frequently used lexical bundles in the DEMI corpus will be analysed. In the following section, the structural and functional categorisation of the identified lexical bundles will be assessed.

4.2. An overview of the most frequent lexical bundles in the DEMI Corpus

As explained in the methodology chapter, the structural and functional frameworks of Biber et al. (2004) were implemented to analyse the lexical bundles used by the lecturers in EMI classrooms. The structural and functional taxonomies consisted of main and sub-categories. The structural framework consists of three main headings: verb phrase fragments, dependent clause fragments, noun and prepositional phrase fragments. The functional framework has four main headings: stance expressions, discourse organisers, referential expressions and special conversational

functions. The frequency cut-off was 40 times per million and the range criterion was at least five different lectures. In the initial analysis, 345 individual instances were identified, but after the results had been manually checked by the researcher, 223 instances were manually excluded as they did not correspond with the aim of the study. Taking into account the exclusion criteria mentioned in the methodology section, lexical bundles which are text-dependent, discipline-specific, including proper nouns and ending with articles manually excluded so as to prevent the inflation in the number of the results. As a result of these exclusions, a total of 122 four-word lexical bundles were identified in the DEMI corpus. Sykes (2017) points out the that manual exclusion of lexical bundles is somewhat risky when deciding which ones constitute a bundle or not. However, exclusion is essential to compile a list of pedagogically valuable lexical bundles to use for educational purposes. Considering this, the present study carried out an elimination process for providing a list of pedagogically useful bundles in the EMI context.

The current investigation revealed that the DEMI corpus, compiled from classroom teaching contexts, consists of 122 four-word lexical bundles with a total of 5,040 individual occurrences. In accordance with the present results, Biber et al. (2004) demonstrated that classroom teaching uses much more lexical bundles than all the other spoken and written register types. Among the identified bundles, the most frequent lexical bundle was *is going to be*, with 196 occurrences. Biber et al. (1999) also found that *is going to be* is one of the most common four-word lexical bundles in conversation. Following it, *okay so let us* and *i would like to* were the next most frequent lexical bundles in the corpus. The highly frequent use of these bundles in academic lectures indicates that university lecturers rely heavily on them to introduce topics and sustain learners' attention. The frequency cut-off criteria was 40 times per million words for the analysis of lexical bundles in the DEMI corpus. However, twenty of the 122 lexical bundles identified in the corpus occurred more than 50 times, which reflects a highly frequent use of these recurrent expressions.

In the literature, thirteen of the twenty most commonly used lexical bundles identified in this study had also been recognized by previous researchers (Biber et al., 2004; Cortes, 2004; Kashiha & Chan, 2013; Sykes, 2017). Table 5 shows the most frequently occurring lexical bundles in the corpus along with their frequencies both in the whole corpus and in the two sub-corpora. In total, 35 of the identified lexical bundles in the DEMI corpus were also identified in previous studies. As Table 5 shows,

some of the lexical bundles were used much more frequently in a specific discipline, such as *okay so let us* which was used 194 times in the hard sciences but just once in the soft sciences. Similarly, *when we look at* was used 75 times in soft sciences whereas lecturers in the hard sciences used it just four times. This difference in using lexical bundles in different disciplines shows that lecturers need specific recurrent word combinations according to the needs of the lecture which they are delivering.

Table 5

The twenty most frequently used four-word lexical bundles in the DEMI corpus

Lexical Bundle	*Overall Frequency*	*Frequency in Hard sciences*	*Frequency in Soft sciences*
1. *is going to be*	196	61	135
2. *okay so let us*	195	194	1
3. *i would like to*	137	126	11
4. *okay so that is*	136	136	0
5. *less than or equal*	127	127	0
6. *if you look at*	114	49	65
7. *okay so let me*	108	108	0
8. *it is going to*	95	34	61
9. *to be able to*	90	11	79
10. *okay so this is*	86	86	0
11. *at the same time*	85	13	72
12. *i do not know*	83	56	27
13. *when we look at*	79	4	75
14. *so that is why*	74	7	67
15. *it is a very*	65	35	30
16. *on the other hand*	63	27	36
17. *bigger than or equal*	60	60	0
18. *as you can see*	59	17	42
19. *when you look at*	59	5	54
20. *let us call this*	56	56	0

The findings showed that lexical bundles were used significantly more frequently in the hard sciences than in the soft sciences in EMI classrooms at the university level, as shown in Table 6. Lecturers in the hard sciences used a greater number of lexical bundles with 118 different types in 3,149 individual cases, making up 6% of the sub-corpus. On the other hand, lecturers in the soft sciences used only 73

different lexical bundle types in 1,891 individual cases, which was 4% of the soft science sub-corpus. These findings align with those of Kashiha and Chan (2013) who identified 121 lexical bundles in their corpus of spoken academic lectures, with 118 of these bundles used in the hard sciences and only 96 identified in the soft science lectures.

A log-likelihood test was conducted to examine the differences in the use of lexical bundles between the hard and soft sciences. The hard science corpus is made up of 264,631 words with 3,149 individual occurrences of lexical bundles, while the soft science corpus consists of 229,131 words with 1,891 individual occurrences. The log-likelihood (LL) score calculated for this comparison is 162.26. This high LL score indicates a statistically significant difference in the frequency of lexical bundles between the two sub-corpora, suggesting that the use of lexical bundles is more prevalent in the hard sciences compared to the soft sciences. These results emphasize the differences in the use of lexical bundles across academic disciplines, revealing a greater reliance on these lexical bundles in hard science lectures compared to those in the soft sciences.

Table 6

Frequency of lexical bundles in hard and soft sciences

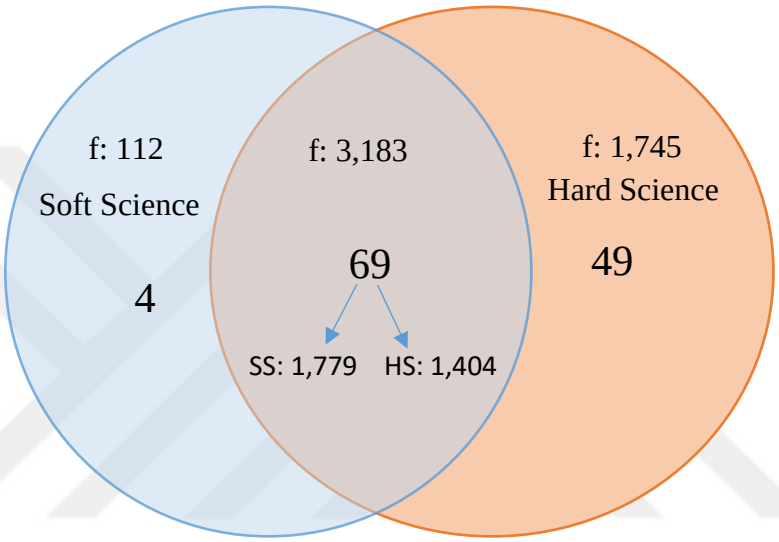
Sciences	Bundle types	Total No. of bundles	% of total words	Log-likelihood
Hard sciences	118	3.149	0.6	162.26
Soft sciences	73	1.891	0.4	

Of the 122 lexical bundles identified in the current corpus, only four were not used in the hard sciences (*on the one hand*, *that we saw in*, *to make sure that* and *that we see in*), whereas 49 of the identified lexical bundles were not used in the soft sciences as seen in Figure 1. Except for these bundles, all the other 69 bundles were shared between the two sub-corpora. These shared bundles occurred relatively more frequently in the soft sciences than in the hard sciences. This finding is also consistent with Kashiha and Chan (2013), who observed that only three lexical bundles were not

identified in hard science lectures, compared to 23 bundles that were not used in soft science classes. This finding shows that the lecturers in the hard sciences tended to use lexical bundles much more frequently than the lecturers in the soft sciences. This higher usage in the hard sciences reflects hard sciences lecturers' need for precision, clarity and the detailed presentation of findings and procedures, as these disciplines often require explicit communication of experimental results and technical details.

Figure 1

Common and unique lexical bundles in the hard and soft sciences



Kashiha and Chan (2013) note that despite some disciplinary variations, the number of shared bundle types across both hard and soft sciences highlights the significant dependence of university lectures on multi-word combinations and specifically on lexical bundles. This dependence shows the essential function these bundles serve in organizing lectures, facilitating student understanding, and maintaining a flow of information in academic discourse. This dependence also indicates that lecturers in university contexts frequently rely on lexical bundles to enhance the clarity and coherence of their lectures, ensuring the effective instruction of complex academic knowledge.

(1) as we will define later on. *okay so let us* go to the next definition.

(2) it can be used directly and that is exactly what I have done *okay so that is* a formula by Euler

(3) yes okay first *I would like to* talk about a very simple fact okay so let me do it on this part of the blackboard

Commonly used lexical bundles in hard sciences, such (1) as *okay so let us*, *(2) okay so that is* and *(3) i would like to* are used to attract students' attention, clarify findings and highlight significant points. In contrast, the soft sciences lecturers used lexical bundles less frequently, with a focus on expressing opinions, contextualizing information and discussing broader implications, as seen in bundles such as (4) *all the way to, (5) when we look at, (6) at the same time*. This distinction in the use of lexical bundles indicates that lecturers in the hard and soft sciences have different communicative needs and purposes while conveying academic knowledge in their classes. As one of the purposes of the present study was to analyse the structural and functional features of the lexical bundles in the corpus, in the next sections, the structural and functional analyses of the study will be presented.

(4) including the Egypt and Libya today *all the way to* the Spanish coast were under Roman rule

(5) so this is a remarkable achievement because *when we look at* the evolution of Greek sculpture

(6) to able to achieve peace but *at the same time* to be able to maintain that peace

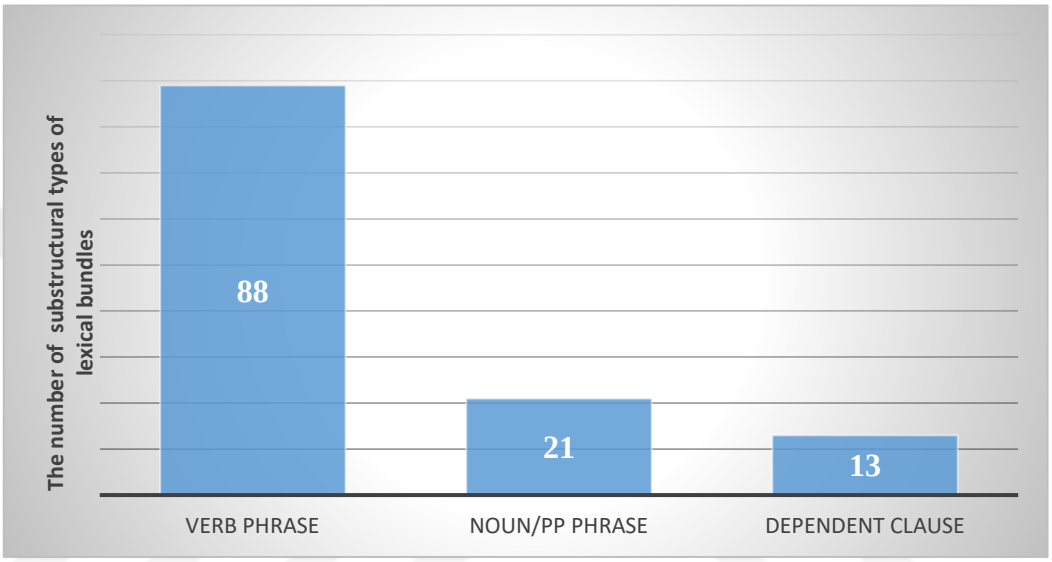
4.3. The structural classification of the lexical bundles

The structural analysis of the lexical bundles was based on the taxonomy of Biber et al. (2004) which contains three main categories: verb phrase fragments, dependent clause fragments, noun and prepositional phrase fragments. First, this is in line with the statement of Biber et al. (1999) that lexical bundles found in the DEMI corpus are not grammatically complete units. Examples such as *is going to be*, *i would like to* and *okay so this is* prove that lexical bundles are not complete units. Altogether, 122 lexical bundles were identified in the corpus, comprising 88 verb phrase bundles, 21 noun and prepositional phrase bundles and 13 dependent clause bundles, as shown in Figure 1. The most frequently used type of lexical bundle was verb phrase fragments, which formed 72% with 88 individual cases. Following this, noun and prepositional phrase bundles were used at 18% with 21 individual occurrences and dependent clause bundles were the least used type of lexical bundles at 10% and thirteen different appearances. As can be seen, lecturers used a large number of lexical bundles, especially verb phrase bundles, to convey academic knowledge in EMI classrooms. The structural analysis of the DEMI corpus showed that lecturers of both

disciplines relied more on verb phrase bundles to maintain the flow of communication throughout the lesson. Biber et al. (2004) said that classroom teaching uses far more different lexical bundles than both conversation and textbooks. The higher tendency to use lexical bundles in the lectures could be the result of the nature of academic lectures.

Figure 2

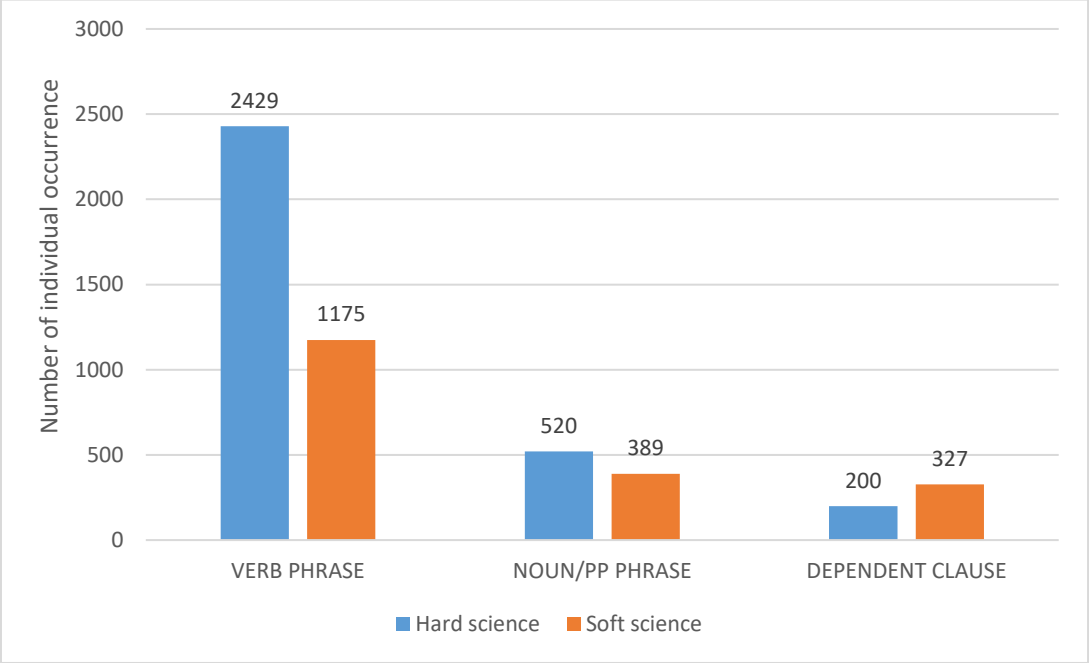
The structural classification of four-word lexical bundles



In the analysis of lexical bundles in EMI classrooms, distinct structural patterns were found between the hard and the soft sciences. From the disciplinary point of view of the structural classification of lexical bundles, it was found that verb phrase fragments were the most frequently used structure in both sub-corpora. As shown in Figure 2, in the hard sciences verb phrase fragments were the most frequent with 2,429 instances, followed by dependent clause fragments at 520 instances and noun phrase fragments at 200 instances. Similarly, the soft sciences showed lower overall frequencies, with verb phrase fragments at 1,175 instances, dependent clause fragments at 389 instances and noun clause fragments at 327 instances. These findings show that although the overall frequency of lexical bundles varied between disciplines, verb phrase fragments were consistently predominant, reflecting a crucial role in conveying messages in academic lectures. Dependent clause and noun phrase fragments followed in frequency, highlighting their importance in providing context and specifying key concepts across both the hard and the soft sciences.

Figure 3

Structural types of lexical bundles in hard and soft sciences



4.3.1. Verb phrase fragments

The analysis of the structural classification of the lexical bundles in the corpus showed that the most commonly used type was lexical bundles which included verb phrase fragments in both disciplines. Verb phrase fragments, which typically include a verb and one or more accompanying elements, play a significant role in academic discourse to construct academic arguments and convey messages clearly to the audience. In the DEMI corpus, there were 88 verb phrase fragments out of the 122 lexical bundles which made 72% of the whole corpus. Biber et al. (2004) state that almost ninety percent of the frequently occurring lexical bundles in conversation consist of verb phrases. This indicates that verb phrases are highly important in spoken interactions and are critical for maintaining the flow of communication. With reference to the findings of Biber et al.'s (2004), the result of this study revealed that university classroom teaching is clearly closer to conversation.

In terms of disciplinary differences, the hard sciences used verb phrase bundles much more than the soft sciences. In the analysis, significant disciplinary distinctions were found in the use of lexical bundles, particularly verb phrase fragments. It was found that the hard sciences showed a higher frequency of verb phrase bundles than

the soft sciences as shown in Figure 2. Specifically, lecturers in the hard sciences tended to employ lexical bundles beginning with a discourse marker followed by a verb phrase, for example (7) *okay so let us*, *(8) okay so let me*, (9) *let us call this* and *(10) let us look at.* These bundles served to facilitate the management of discourse flow in hard science lectures, enabling transitions, highlighting key points and facilitating student comprehension of complex explanations and procedures. On the other hand, there was less reliance on such directive verb phrase bundles in the soft sciences, which frequently employed more exploratory and discussion-based teaching techniques. This discipline-specific variance draws attention to the differences in the communication needs and teaching approaches used in EMI classrooms for hard and soft science lectures.

(7) okay any question about this *okay so let us* do an example

(8) answers are also available on the website of the book in case you did not know *okay so let me* repeat the solutions of the problems

(9) *okay so let us* start giving names so *let us call this* sphere sigma *let us call this* sphere sigma prime

(10) you just have to be careful about the regions in which the density function is valid any questions about this *let us look at* another example today is all about examples

As already stated, the most common sub-type was *discourse marker + VP fragment* with 38 occurrences. Following this, the second most frequent type was *first/second person pronoun + VP fragment* with 28 occurrences, and third was *third person pronoun + VP fragment* with ten occurrences. This finding is in line with that reported by Biber et al. (2004) that verb phrase-based bundles are the most frequent type of lexical bundles in both conversation and academic lectures. This could indicate that lecturers use verb phrase-based lexical bundles to initiate their speech or to reflect their intention, as in:

(11) Let us start today *I would like to* first make you go through all the EU institutions so you will have a general idea how many EU institutions are actually out there.

Table 7

Lexical bundles that incorporate verb phrase fragments

Structural categories	*HS*	*SS*	*Sub-categories*	*Sample bundles*
Verb phrase fragments	2.429	1.175	1a. 1st/2nd person pronoun + VP	*i would like to*
			1b. 3rd person pronoun + VP	*it is going to*
			1c. Discourse marker + VP	*so that is why*
			1d.Verb phrase (with non-passive verb)	*is going to be*
			1e. Verb phrase with passive verb	*can be found by*
			1f. yes-no question	*
			1g. WH-question	*what does that mean*

4.3.2. Noun and prepositional phrase fragments

In the analysis of lexical bundles within EMI classrooms, noun phrase fragments were identified as a significant category. These bundles typically consist of a noun phrase which frequently recurs in academic discourse, providing cohesion and clarity to the communication. As mentioned earlier, noun phrase and prepositional phrase fragments are less frequently used than verb phrase fragments in the DEMI corpus. This finding is in contrast with the findings of previous studies such as that of Kashiha and Chan (2013), who found that in their spoken academic corpus the most commonly used structural type was noun and prepositional phrase fragments. They stated that the highly frequent use of noun and prepositional phrases could be the result of the lecturers' tendency to convey their messages by using various nouns and prepositions.

In the DEMI corpus, lecturers used this type of lexical bundle 17% with 21 occurrences. Common noun phrase fragments identified in the corpus included (12) *at the same time*, (13) *on the other hand* and (14) *in the form of*. These bundles were found to occur frequently in academic lectures, serving as a component in constructing

academic arguments and explanations, as in:

(12) it does not matter if they all pick *at the same time* or they do it one by one

(13) he did not want to imitate what Trajan had done *on the other hand* even if he did there was simply no space left in Rome

(14) in order to be able to have a roof and internal support *in the form of* a colonnade was initially put inside because

The current findings indicated that the hard sciences used more noun and prepositional phrase lexical bundles in terms of the number of individual cases compared with the soft sciences. However, when considering the relative frequency within each sub-corpus, these bundles made up 0.1% of the total words in both the hard and the soft sciences. This shows that although the hard sciences employed a greater number of these lexical bundles, the density in each sub-corpus was nearly the same in both disciplines.

This similar frequency proportion of noun and prepositional phrase lexical bundles suggests that despite the differing communicative needs of the hard and soft sciences, noun and prepositional phrase lexical bundles play a similarly important role in both contexts. These bundles help in introducing and elaborating on key concepts, specifying details and conveying information within a particular context. This similar frequency highlights their significant role in structuring academic discourse and facilitating clear and coherent communication in EMI classrooms across the hard and soft sciences. By examining the usage patterns of these lexical bundles, lecturers can gain valuable insights for developing targeted teaching strategies and materials. This can aid lecturers in both hard and soft sciences in effectively employing these linguistic structures to enhance the clarity of their lectures.These findings also highlight the importance of considering whole numbers and frequencies when analysing lexical bundle use across different academic disciplines.

Table 8

Lexical bundles which incorporate noun and prepositional phrase fragments

Structural categories	*HS*	*SS*	*Sub-categories*	*Sample bundles*

Table 8 continued				
Noun phrase and prepositional phrase	520	389	3a. Noun phrase with of-phrase	*the centre of gravity*
			3b. Noun phrase with post-modifier	*the point in which*
			3c. Other noun phrase expressions	*all the way to*
			3d. Prepositional phrase expressions	*at the same time*
			3e. Comparative expressions	*bigger than or equal*

4.3.3. Dependent clause fragments

Dependent clause fragments were identified as the least frequently used structural type in the analysis of lexical bundles within EMI classrooms. Dependent clause fragments play a significant role in adding complexity and depth to academic discourse. For instance, common dependent clause fragments identified in the corpus included (15) *if you look at*, *(16) to be able to* and *(17) i want you to*. These bundles frequently occur in academic discourse, providing essential contextual and explanatory information. The findings showed that the dependent clause fragment *if you look at* occurred 114 times in the whole corpus. This suggests that dependent clause fragments are one of the important factors in focusing on points, providing examples and explaining complex ideas within different academic disciplines. Example usages from the DEMI corpus are:

(15) *if you look at* it from a great distance and *if you look at* it from very far away it really becomes impossible to distinguish

(16) Turkey's economy will be able to become strong enough *to be able to* compete with the European counterparts

(17) This pair of straight lines will generate a cone okay okay and *I want you to* imagine it as not as your

The findings also showed that the soft sciences used more dependent clause

lexical bundles in terms of the number of individual cases compared with the hard sciences. Specifically, there were 200 individual occurrences in the hard sciences sub-corpus, whereas there were 327 occurrences in the soft sciences sub-corpus. However, when considering the frequency within each sub-corpus, these bundles constituted 0.2% of the total words in both the hard and the soft sciences.

Table 9

Lexical bundles that incorporate dependent clause fragments

Structural categories	*HS*	*SS*	*Sub-categories*	*Sample bundles*
Dependent clause fragments	200	327	2a.1st/2nd person pronoun + dependent clause fragment	*i want you to*
			2b. WH- clause fragments	*when you look at*
			2c. If-clause fragments	*if you look at*
			2d. To-clause fragment	*to be able to*
			2e. That-clause fragment	*that you have to*

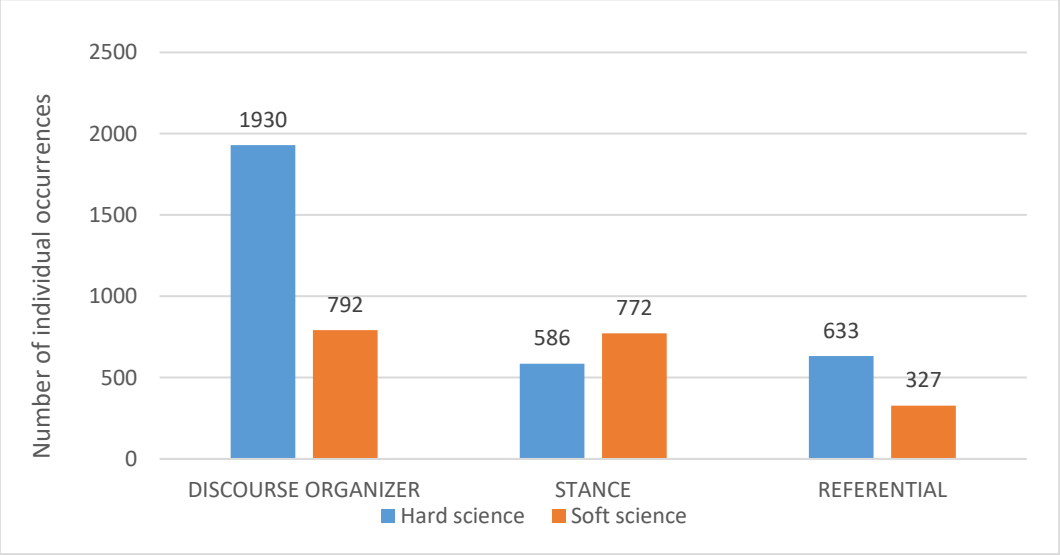
This indicates that although soft sciences employed a greater number of these lexical bundles, the relative frequency was nearly the same in both disciplines. This similar proportion suggests that despite the differing communicative needs of the hard and soft sciences, noun and prepositional phrase lexical bundles play a similarly important role in both contexts. These bundles are used for introducing and elaborating on key concepts, specifying details and explaining a topic within a particular context. This similar frequency highlights their fundamental role in structuring academic discourse and facilitating clear and organized communication in EMI classrooms across both scientific and humanities-based disciplines.

4.4. The functional classification of lexical bundles

As explained in the methodology chapter, the functional framework developed by Biber et al. (2004) was used to analyse the lexical bundles. The functional classification of lexical bundles developed by Biber et al. (2004) divided lexical bundles into three main functional categories: discourse organisers, stance and referential expressions. Each category and subcategory serves a distinct communicative purpose, contributing to the overall efficacy of academic discourse. The main three categories of the lexical bundles are shown with respect to the number of individual occurrences in each sub-corpus in Table 6. The findings of the present study showed that discourse organisers were by far the most frequently used lexical bundle type in the DEMI corpus, followed by stance expressions, with referential bundles being the least used functional type of lexical bundles.

Figure 4

Functional types of lexical bundles in hard and soft sciences



4.4.1. Discourse organisers

Discourse organiser bundles play a significant role in building cohesion and conveying academic messages as they bridge prior and upcoming discourse (Biber et al., 2004). Discourse organisers have two essential functions in classroom teaching: topic introduction/focus and topic elaboration/clarification. The findings of the present study showed that discourse organisers were the most frequently preferred type of lexical bundle among lecturers in EMI classes in academic discourse. This finding contrasts with Biber and Barbieri (2007), who found that stance bundles are the most

preferred functional type in classroom teaching. However, they also noted that discourse organisers are extensively employed in classroom teaching among the other university-spoken registers. This indicates that while there may be variations in the preferred types of lexical bundles across different studies and contexts, the role of discourse organisers remains significant in facilitating effective communication and structuring in academic lectures.

More than half of the lexical bundles in the DEMI corpus represent the discourse organiser function, with 55% and 2.722 individual occurrences. In the DEMI corpus, 67 discourse organiser bundles were identified, among these bundles the three most frequently used discourse organisers are *okay so let us, I would like to* and *okay so that is.* Studies by Nesi and Basturkmen (2006) and Biber et al. (2004) have emphasized the importance of discourse organisers in creating cohesion and clarity in spoken academic discourse. As seen in the example (19), *okay so let us* is frequently used to transition to a new topic or section, helping to signal to students that a shift in focus is occurring. This highly frequent use of discourse organisers could be the result of lecturers' tendency to initiate a new sentence or a new topic while avoiding ambiguity. It is also possible that the reason why lecturers used discourse organisers two or three times more than the other functional types is the intention of the lecturers to attract the students' attention before introducing a new topic or asking for their participation.

(18) *okay so let us* return to the question what is the probability that a signal was sent given that detector counted…

(19) in this example it is not our current topic but *I would like to* stop here for a second because I know that

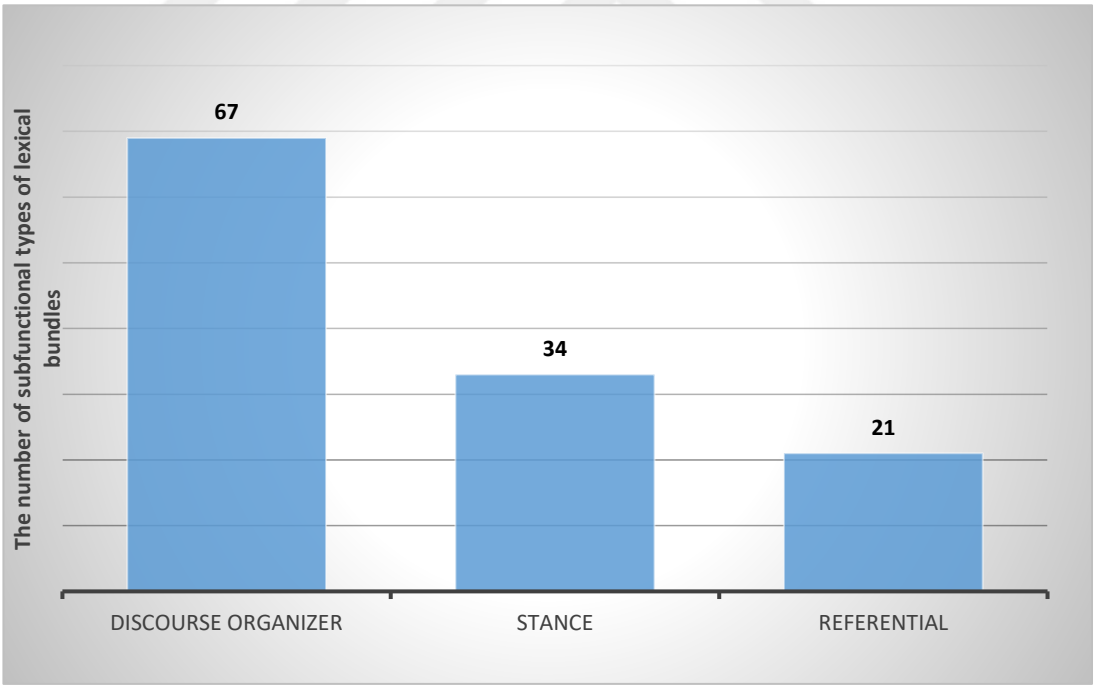
(20) *okay so that is* the end of our discussion on discrete probability models now we can switch to continuous probability models and let me try to zoom in up

(21) these were important in the daily life of the Hittites as well so *when we look at* the political structure of the Hittites…

From the disciplinary point of view, the findings showed that discourse organiser bundles were the most frequent functional type in both disciplines. In the hard sciences, lecturers used discourse organiser bundles in 1,930 individual occurrences compared with 792 occurrences in the soft sciences. Despite this numerical difference, discourse organiser bundles were still the most preferred functional type in the soft sciences. Overall, discourse organisers were two to three times more preferred than other functional categories across both disciplines. This highlights the critical role of discourse organiser bundles in academic lectures. Phrases such as *i would like to*, *okay so let us*, *when we look at* and *if you look at* are crucial for transitioning between topics, emphasizing key points and guiding the audience through complex lesson topics. The higher frequency of these bundles in the hard sciences suggests a need for clear and specific organization to guide the content topics of scientific disciplines.

Figure 5

Functional classification of four-word lexical bundles



In contrast, although the soft sciences make use of these bundles, their usage patterns reflect the different communicative strategies and instructional styles specific to these fields. The significance of discourse organiser bundles in both the hard and the soft sciences highlights their importance in facilitating effective communication

and comprehension in academic lectures. This preference for discourse organisers across the disciplines shows their critical role in improving clarity and coherence in academic discourse, making them essential tools for instructors in EMI settings.

Table 10

Discourse organisers in the DEMI corpus

Functional category	HS	SS	Sub-categories	Sample bundles
1.Discourse organisers	1.930	792	1a. Topic introduction/focus	*okay so let us*
			1b. Topic elaboration/clarification	*on the other hand*

4.4.2 Stance expressions

According to Biber (2006a), stance expressions are specific phrases which convey the speaker's attitudes, judgments or evaluations about the information being presented. Stance expressions have significant importance in academic discourse, especially in EMI classrooms, as they enable lecturers to express certainty, doubt, importance and personal opinions. In this section, I shall examine the frequency, distribution and functions of stance expressions in EMI lectures, based on the analysis of the DEMI corpus. The analysis showed that stance expressions such as *i do not know, we do not know* and *if you want to* were used by the EMI lecturers to facilitate the expression of the speaker's opinions and attitudes, direct the audience's understanding of the information and improve the conversation by promoting interaction and engagement. In EMI classrooms, stance expressions play a crucial role in organising and presenting academic discourse.

(22) okay the probability that the fifth toss is tails. *I do not know* which coin I picked remember I start the experiment by throwing by tossing a coin

(23) so in this case we have a model for the lifetime of the lightbulb as an exponential but *we do not know* the parameter *we do not know* the rate

of the exponential distribution we will infer the rate of the distribution

(24) and sometimes *if you want to* give a very clear message and if you want to have in that sense a more powerful action about a certain topic then you have to

Table 11

Stance expressions in the DEMI corpus

Functional category	HS	SS	Sub-categories	Sample bundles
2.Stance expressions	586	772	2a. Epistemic stance	i do not know
			2b. Attitudinal/ Modality Stance	do not have to

Another key finding was that stance expressions were mostly used in the soft sciences, with 772 individual occurrences, compared with 586 occurrences in the hard sciences. Stance expressions, which include phrases such as *i do not know*, *i want you to* and *I think it is* are crucial for conveying the speaker's attitude or opinion on a particular topic. In the soft sciences, where subjective instruction plays a significant role, the higher frequency of these bundles reflects the necessity of expressing ideas, beliefs and uncertainties. These expressions help lecturers in the soft sciences to highlight critical points and encourage critical thinking among students.

(25) in the official document itself the reason why I got certain sentences because *I want you to* focus on those sentences but you also would use the whole text that is related

(26) what kind of actions you are taking *I think it is* a very valid principle that we sometimes forget the proportionality what you are doing and what you are achieving

In contrast, although stance expressions were also present in the hard sciences, their lower frequency suggests a more objective and factual approach to delivering information. Hard sciences tend to use empirical data and concrete materials, resulting

in less frequent use of stance expressions. However, the presence of 586 individual occurrences indicates that even in the hard sciences, conveying the speaker's evaluation and interpretation remains important for explaining experimental results, discussing implications and guiding the students' understanding. The distinctive use of stance expressions between the hard and soft sciences highlights the variable communicative strategies across disciplines, reflecting their different instructional goals.

4.4.3 Referential expressions

The final key finding was that referential bundles were the least preferred functional type of lexical bundle in EMI classrooms. This finding supports evidence from Biber and Barbieri (2007) that referential bundles are used more frequently in written registers than in spoken university registers. As shown in Table 12, referential bundles occurred 633 times in the hard sciences and 327 times in the soft sciences, indicating that the hard sciences used nearly twice as many referential bundles as the soft sciences. The finding of the study is also consistent with Molino (2019) that imprecision bundles are not used frequently in EMI lectures.The analysis of EMI lectures from Turkish higher education contexts indicates that the lecturers do not tend to use imprecision bundles frequently. Referential bundles, which include phrases such as *at the same time, less than or equal* and *in the form of* are essential for referring to specific entities, concepts or activities within the discourse. In the hard sciences, the higher frequency of these bundles reflects the need for accurate and clear references to experimental results, specific procedures and described concepts. This usage supports the objective and detailed nature of the hard sciences in which clarity is essential.

(27) okay great so we did two things *at the same time* we show the splitting result for the Poisson and we also talked about how we handle a random sum of random variables

(28) okay let us compute that the probability that x squared is *less than or equal* to zero point five okay what do we do we integrate

(29) you often need a transitional element you have these triangular elements and here it is shown *in the form of* steps so these transitions are often done

In the soft sciences, the lower occurrence of referential bundles suggests a different approach to discourse, in which the focus might be more on context and

discussion than on specific referencing of specific topics. The use of 327 referential bundles in the soft sciences still highlights their importance in the soft sciences, but their relatively lower frequency shows the nature of these fields. Overall, the preference for referential bundles in the hard sciences over the soft sciences emphasises the different communicative needs of these disciplines. It reflects how different fields prioritize and use language to achieve their instructional goals in EMI classrooms.

Table 12

Referential expressions in the DEMI corpus

Functional category	HS	SS	Sub-categories	Sample bundles
Referential expressions	633	327	3a. Identification/ Focus	*in fact it is*
			3b. Imprecision	*
			3c. Specification of Attributes	*less than or equal*
			3d.Time/place/text reference	*at the same time*

CHAPTER V

DISCUSSION, CONCLUSION AND IMPLICATIONS

5.1 Introduction

In this concluding chapter, the findings from the analysis of the DEMI corpus compiled from EMI classroom discourse at the university level are discussed. In addition, this chapter defines the limitations of the present study and offers suggestions for further research. Finally, the pedagogical implications which can be beneficial for the EMI lecturers and learners are presented.

5.2. Discussion of the findings

In the scope of the study, EMI lectures at the university level in METU were analysed by using the DEMI corpus which contains lectures from both the hard and the soft sciences. The first objective of the study was to determine the most frequently used lexical bundles in EMI lectures and the second was to analyse the structural and functional characteristics of the lexical bundles employed in general and in both disciplines separately. The analysis employed the structural and functional frameworks of Biber et al. (2004) to categorize the lexical bundles and examine their frequency and distribution across different academic disciplines. Data were collected from a corpus of academic lectures delivered in EMI settings and then the corpus was divided into two sub-corpora representing hard sciences and soft sciences.

The analysis of the DEMI corpus revealed that 122 four-word lexical bundles were apparently used in EMI lectures. This widespread use of lexical bundles highlights the critical role of lexical bundles in academic discourse in EMI settings. Extensive use of lexical bundles may help lecturers organise the flow of the lesson and convey complex information more effectively to students. There were 5.040 individual occurrences at least 40 times per million words in the corpus as a whole. The results indicates that EMI lecturers in the Turkish higher education context frequently used lexical bundles during their classes. Hyland (2008b) emphasises the significance of lexical bundles as essential components of fluent linguistic production and the crucial role of these multiword expressions in facilitating fluent and coherent

language use. As it is stated by Biber and Barbieri (2007), lexical bundles play a critical role in structuring discourse by acting as the major component of larger phrases and clauses and they serve as '' discourse frames for the expression of new information''. Proficiency in formulaic language could offer three main advantages for second language speakers: first, it could make them appear more fluent; second, it could provide them with a broader range of expression; and finally, it could lead to them being perceived as more accurate speakers (Boers et al., 2006). Considering the extensive use of lexical bundles in the DEMI corpus, it is important to highlight that lexical bundles are important parts of an academic discourse.

The findings showed some disciplinary differences in the use of lexical bundles in EMI classrooms. Overall, the hard sciences employed more lexical bundles than the soft sciences, with 118 bundles identified in the hard sciences compared with 73 in the soft sciences. In addition, although only four of the identified 122 lexical bundles were absent from the hard sciences, 49 of them were not used in the soft sciences. This finding suggests that the exact and technical nature of the hard sciences needs more significant reliance on formulaic language to convey complicated information effectively. In both the hard and the soft sciences, verb phrase fragments were the most preferred type, followed by noun phrase fragments, and finally dependent clause bundles. Verb phrase fragments, such as *is going to be*, *okay so let us* and *i would like to* were found to be important for defining actions and processes, which are essential to the instructional scope of academic lectures.

A log-likelihood test was conducted in order to statistically analyse the use of lexical bundles across disciplines. The stastical analysis revealed that the hard science corpus, consisting of 264,631 words and 3,149 individual occurrences of lexical bundles, significantly differs from the soft science corpus, which comprises 229,131 words and 1,891 individual occurrences. The score of the log-likelihood (LL) test was 162.26 which indicates a strong statistical significance in terms of the differences in lexical bundle frequency across disciplines. The result of the log-likelihood test confirmed that there is a statistically significant difference in terms of lexical bundle usage across the hard and soft sciences.

A significant finding of the study was the extensive use of verb phrase lexical bundles starting with discourse markers, such as *let us move on*, *okay so let us* and *so it means that*. In the structural analysis, 88 verb phrase bundles were identified and 33 of them were verb phrase lexical bundles starting with *let*, *so* and *okay*. This highly

frequent usage of these three discourse markers contributed to the increased presence of verb phrase bundles in the corpus. Lecturers used these phrases to attract students' attention, guide them through the lecture content, signal transitions and encourage participation. Increased use of verb phrase bundles starting with discourse markers might be the result of the nature of the EMI contexts where clear and structured communication is crucial for non-native English lecturers. EMI lecturers may need to use these phrases to maintain the flow of the academic discourse in an effective manner.

In the functional analysis, discourse organisers were found to be the most frequently employed type of lexical bundle. In the hard sciences, discourse organisers appeared 1,930 times, compared with 792 times in the soft sciences. Discourse organisers were found to be the most frequently used functional type in both disciplines. The findings also showed that EMI lecturers used discourse organisers starting with *let us*, *okay* and *so far* more than the others. This significant finding indicated that lecturers relied much more on some specific discourse organisers. As a result of this tendency, discourse organisers were found to be the most commonly used functional type. This finding also indicated that discourse organisers play a critical role in organizing a lecture, transitioning between topics and guiding the audience through the material in EMI classrooms. Biber and Barbieri (2007) note that "teacher-centred" spoken registers tend to use more discourse organiser functions. The highly frequent reliance on these specific discourse organisers in the DEMI corpus highlights their pedagogical value. Considering this, lecturers and students should be aware of and intentionally use these bundles to improve the effectiveness of their teaching and learning. The results of this study indicated that EMI lecturers do not tend to use impresision bundles in the lectures. In accordance with Molino (2019) that imprecision bundles are not frequently used in EMI classes. A possible explanation for this might be that the nature of EMI classes does not require to use imprecision bundles.

The findings showed that lexical bundles were widely employed in EMI classrooms in academic discourse. In addition, the results of this study indicated that the hard sciences relied more heavily on lexical bundles to convey detailed and technical information and that both disciplines employed these bundles to improve clarity, coherence and engagement in academic lectures. In light of these findings, it is important to state that understanding these different practices between disciplines can help to develop pedagogical strategies. In order to support lecturers to improve the

effectiveness of their lectures and support student comprehension in EMI settings, the use of lexical bundles might be beneficial.

5.3. Conclusion

The influence of internationalization and globalization has significantly impacted HEIs. As a result, these institutions have increasingly implemented EMI programmes to attract more international students and to internationalize the education which they offer. As a consequence of the widespread implementation of EMI programmes across the world, English has become more prominent in academic contexts. Despite this shift, there have been relatively few studies focusing on spoken academic discourse, particularly in terms of lexical bundles. With these issues in mind, the present study focused on the classroom discourse in EMI lectures at the university level in order to identify the lexical bundle usage.

The findings of the study showed that lexical bundles are widely used by lecturers in EMI classrooms. The findings also showed that EMI lecturers employed bundles starting with *let us, okay* and *so* far more than all the other bundle types. This highly frequent employment of these bundles indicated that lecturers rely on them to structure the flow of the lesson and to attract learners' attention. Additionally, EMI lecturers preferred using verb phrases to using the other structural types and discourse organisers to the other functional types.

The distinctiveness of the present study originates from some specific aspects: it focused on Turkish EMI classroom discourse, investigated various disciplines and applied a corpus-driven method. To summarize, the present study sheds light on the critical role of lexical bundles in the Turkish EMI context and provides invaluable insights into the ways in which lexical bundles are used to structure classroom discourse. By recognizing the significance of lexical bundles, this research is intended to support pedagogical practices which can benefit the potential of lexical bundles to improve understanding, attract attention and improve the overall quality of EMI classroom discourse.

5.3.1 Limitations of the research and suggestions for future studies

There are of course some limitations in this study. The first limitation is that the study was limited to only hard and soft science lectures in the EMI context. The results were therefore considered within the framework of these specific disciplines.

Further studies could be carried out focusing on different disciplines in order to have a broader understanding of the use of lexical bundles in EMI classrooms. Future studies could also focus on different contexts to deepen the understanding of lexical bundles in general rather than specifically in EMI contexts.

5.3.2. Pedagogical implications

The findings of this research have significant implications for the understanding of how lexical bundles are employed at spoken university contexts. The insights gained from this study may be of assistance to the understanding of the lexical bundles in EMI lectures. It would be beneficial for students to be aware of the significance of lexical bundles in academic discourse. Being aware of the most frequently used lexical bundles during the lectures could help students to better follow the lectures and understand the content. Furthermore, intentional exposure to the lexical bundles could allow students to comprehend the organisation of lectures and significant details. Explicit exposure to lexical bundles through direct teaching or lesson materials designed with lexical bundles in mind may be an efficient way to improve students' overall success in the lecture.

Since the lexical bundles are frequently used in EMI lectures, it can be stated that lecturers should enhance their awareness of the lexical bundles used in a class. The deliberate integration of lexical bundles into the EMI lectures may have a positive effect on improving clarity and coherence in the lectures. Moreover, lecturers can include the most frequently used lexical bundles in the process of conveying complex content subjects to facilitate learning among students. This may lead to foster effective communication during the EMI lectures. It can, therefore, make EMI lectures more accessible to students with various linguistic backgrounds. As a result, making students and lecturers more aware of lexical bundles could support the learning process and contribute to the overall academic success in EMI classrooms.

REFERENCES

Akbaş, E., & Bal-Gezegin, B. (2022). 'Exploring the functions of okay as a discourse marker in an English-medium instruction class.' In Y. Kırkgöz & A. Karakaş (Eds), *English as the medium of instruction in Turkish higher education: policy, practice and progress* (pp.233-256). Springer International Publishing. https://doi.org/10.1007/978-3-030-88597-7_12

Anthony, L. (2024). AntConc (4.3.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). Longman grammar of spoken and written English (Vol. 2). https://doi.org/10.1162/089120101300346831

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In *Out of corpora* (pp. 181-190). Brill. https://doi.org/10.1163/9789004653689_017

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied linguistics*, *25*(3), 371-405. https://doi.org/10.1093/applin/25.3.371

Biber, D. (2006a). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, *5*(2), 97-116. https://doi.org/10.1016/j.jeap.2006.05.001

Biber, D. (2006b). University language. *University Language*, 1-269.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes*, *26*(3), 263-286. https://doi.org/10.1016/j.esp.2006.08.003

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language teaching research*, *10*(3), 245-261. https://doi.org/10.1191/1362168806lr195oa

Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition*. Springer. https://doi.org/10.1057/9780230245006

Bolinger, D. (1976). Again-one or two subjunctives? *Hispania*, *59*(1), 41-49. https://doi.org/10.2307/339370

Bradford, A. (2019). It's not all about English! The problem of language foregrounding in English-medium programmes in Japan. *Journal of Multilingual and Multicultural Development*, *40*(8), 707-720. https://doi.org/10.1080/01434632.2018.1551402

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. http://dx.doi.org/10125/44213

Coleman, J. A. (2006). English-medium teaching in European higher education. Language teaching, 39(1), 1-14. https://doi.org/10.1017/S026144480600320X

Conrad, S. M. (1999). The importance of corpus-based research for language teachers. *System*, *27*(1), 1-18. https://doi.org/10.1016/S0346-251X(98)00046-3

Corrales, K. A., Rey, L. A. P., & Escamilla, N. S. (2016). Is EMI enough? Perceptions from university professors and students. *Latin American Journal of Content & Language Integrated Learning*, *9*(2). https://doi.org/10.5294/7094

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for specific purposes*, *23*(4), 397-423. https://doi.org/10.1016/j.esp.2003.12.001

Dearden, J. (2014). *English as a medium of instruction-a growing globalphenomenon.* British Council. https://doi.org/10.13140/RG.2.2.12079.94888

Earls, C. W. (2016). *Evolving agendas in European English-medium higher education: Interculturality, multilingualism and language policy.* Springer. https://doi.org/10.1057/9781137543127

Ekoç, A. (2020). English Medium Instruction (EMI) from the perspectives of students at a technical university in Turkey. *Journal of Further and Higher Education*, *44*(2), 231-243. https://doi.org/10.1080/0309877X.2018.1527025

Firth, J. R. (1951). Modes of meaning. In Papers in Linguistics, 1934-1951, 118–149. London: OUP.

Galloway, N., & Rose, H. (2015). Introducing global englishes. Routledge. https://doi.org/10.4324/9781315734347

Gharbavi, A., & Iravani, H. (2014). Is teacher talk pernicious to students? A discourse analysis of teacher talk. *Procedia-Social and Behavioral Sciences*, *98*, 552-561. https://doi.org/10.1016/j.sbspro.2014.03.451

Granger, S., & Meunier, F. (Eds.). (2008). *Phraseology: An interdisciplinary perspective*. John Benjamins Publishing. https://doi.org/10.1075/z.139

Hüppauf, B. (2004). Globalization–Threats and opportunities. *Gardt, a*, 3-25. https://doi.org/10.1515/9783110197297

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse studies*, *7*(2), 173-192. https://doi.org/10.1177/1461445605050365

Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*, *27*(1), 4-21. https://doi.org/10.1016/j.esp.2007.06.001

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International journal of applied linguistics*, *18*(1), 41-62. https://doi.org/10.1111/j.1473-4192.2008.00178.x

Hyland, K., & Jiang, F. (2018). Academic lexical bundles: How are they changing?. *International Journal of Corpus Linguistics*, *23*(4), 383-407. https://doi.org/10.1075/ijcl.17080.hyl

Jenkins, J. (2006). Points of view and blind spots: ELF and SLA. *International Journal of Applied Linguistics*, *16*(2), 137-162. https://doi.org/10.1111/j.1473-4192.2006.00111.x

Jespersen, O. (2013). *The philosophy of grammar*. Routledge. https://doi.org/10.4324/9780203716045

Kashiha, H., & Chan, S. H. (2013). An exploration of lexical bundles in academic lectures: Examples from hard and soft sciences. *Journal of Asia TEFL*, *10*(4).

Lasagabaster, D., Doiz, A., & Sierra, J. M. (Eds.). (2014). *Motivation and foreign language learning: From theory to practice* (Vol. 40). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.40

Lasagabaster, D. (2022). *English-medium instruction in higher education*. Cambridge University Press. https://doi.org/10.1017/9781108903493

Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching*, *51*(1), 36-76. https://doi.org/10.1017/S0261444817000350

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, *22*(3), 276-282. https://doi.org/10.11613/BM.2012.031

Molino, A. (2019). Lexical bundles in EMI lectures: An exploratory study. *Textus*, *32*(1), 119-140. https://doi.org/10.7370/93186

Nattinger, J. R., & DeCarrico, J. S. (1988). Lexical phrases for the comprehension of academic lectures. *English for specific purposes*, *7*(2), 91-102. https://doi.org/10.1016/0889-4906(88)90027-0

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press. https://doi.org/10.1017/S0272263100013000

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International journal of corpus linguistics*, *11*(3), 283-304. https://doi.org/10.1075/ijcl.11.3.04nes

Nicaise, E. (2020). *Native and non-native teacher talk in the EFL classroom: A corpus-informed study*. Routledge. https://doi.org/10.4324/9780429263026

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: native like selection and native like fluency. In J. Richards & R. Schmidt (Eds.), Language and communication (pp. 191-226). London: Longman.

Pecorari, D., & Malmström, H. (2018). At the crossroads of TESOL and English medium instruction. *Tesol Quarterly*, *52*(3), 497-515. https://doi.org/10.1002/tesq.470

Rymes, B. (2015). Classroom discourse analysis: A tool for critical reflection. Routledge. https://doi.org/10.4324/9781315775630

Salazar, D. (2014). Lexical bundles in native and non-native scientific writing. https://doi.org/10.1075/scl.65

Samraj, B. (2024). Disciplinary differences in lexical bundles use: A cautionary tale from methodological variations. Journal of English for Academic Purposes, 70, 101399. https://doi.org/10.1016/j.jeap.2024.101399

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action. *Formulaic sequences: Acquisition, processing and use*, 1-22. https://doi.org/10.1075/lllt.9

Sert, N. (2008). The language of instruction dilemma in the Turkish context. *System*, *36*(2), 156-171. https://doi.org/10.1016/j.system.2007.11.006

Sert, O. (2015). *Social interaction and L2 classroom discourse*. Edinburgh University

Press. https://doi.org/10.1515/9780748692651

Sharpe, T. (2008). How can teacher talk support learning? *Linguistics and education*, *19*(2), 132-148. https://doi.org/10.1016/j.linged.2008.05.001

Sinclair, J., & Coulthard, R. M. (1975). Towards an Analysis of Discourse: the English Used by Teachers and Pupils Oxford University Press. https://doi.org/10.1017/S0047404500007363

Sykes, D. L. (2017). *An investigation of spoken lexical bundles in interactive academic contexts* (Doctoral dissertation, Carleton University). https://doi.org/10.22215/etd/2017-12110

Taguchi, N. (2014). English-medium education in the global society: Introduction to the special issue. *International Review of Applied Linguistics in Language Teaching*, *52*(2), 89-98. https://doi.org/10.1515/iral-2014-0004

Tokdemir Demirel, E. (2023). Variation in Disciplinary Academic Writing: A Corpus-Based Investigation. In E. Tokdemir Demirel, B. C. Cengiz & I. G. Kaçar (Eds.). Disciplinary Academic Writing: Opportunities and Challenges, Chapter: 1, 1-29, ISBN: 978-625-367-289-8.

Toth, J., & Paulsrud, B. (2017). Agency and affordance in translanguaging for learning: Case studies from English-medium instruction in Swedish schools. *New perspectives on translanguaging and education*, 189-207. https://doi.org/10.21832/PAULSR7814

Walsh, S. (2011). *Exploring classroom discourse: Language in action*. Routledge. https://doi.org/10.4324/9780203827826

Walsh, S. (2013). *Classroom discourse and teacher development*. Edinburgh University Press. https://doi.org/10.1075/etc.8.1.09meu

Wilkinson, R. (2013). English-medium instruction at a Dutch university: Challenges and pitfalls. *English-medium instruction at universities: Global challenges*, *324*(10.21832), 9781847698162-005. https://doi.org/10.21832/9781847698162-005

APPENDICES

Appendix 1.

Corpus metadata sample

	Major Disciplines	Minor Disciplines	Lecturer	Lecture Week	Lecture Duration	words	Identification Code	Accessible link
25	Mathematics	Geometry	Cem TEZER	1.1.	46.17	5846	Geo_Wk_1.1.	https://www.youtube.com/watch?v=AJvtiK2mmpU&list=PLuiPz6iU5SQ82888mWVxNMC
26	Mathematics	Geometry	Cem TEZER	1.2.	46.17	5519	Geo_Wk_1.2.	https://www.youtube.com/watch?v=1hNR-ICuw7g&list=PLuiPz6iU5SQ82888mWVxNMC
27	Mathematics	Geometry	Cem TEZER	3.1.	44.12	5427	Geo_Wk_3.1.	https://www.youtube.com/watch?v=Le4cup_8tbY&t=7s
28	Mathematics	Geometry	Cem TEZER	3.2.	48.46	6119	Geo_Wk_3.2.	https://www.youtube.com/watch?v=svdOtznxOQU&t=1249s
29	Mathematics	Geometry	Cem TEZER	3.3.	44.08	5174	Geo_Wk_3.3.	https://www.youtube.com/watch?v=Nne1GBvj5ig&t=6s
30	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	2.1.	46.26	4726	PRV_WK_2.1.	https://www.youtube.com/watch?v=0o0tLfvyMM&list=PLuiPz6iU5SQ8ra5kixx770k_far
31	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	2.2.	46.57	4661	PRV_WK_2.2.	https://www.youtube.com/watch?v=sRNqK899ZTQ&list=PLuiPz6iU5SQ8ra5kixx770k_fe
32	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	2.3.	40.30	3910	PRV_WK_2.3.	https://www.youtube.com/watch?v=kuCXRRIPmY0&list=PLuiPz6iU5SQ8ra5kixx770k_fe
33	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	3.1.	46.03	4739	PRV_WK_3.1.	https://www.youtube.com/watch?v=zKO1MAyCzSg&list=PLuiPz6iU5SQ8ra5kixx770k_fe
34	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	3.2.	48.50	4695	PRV_WK_3.2.	https://www.youtube.com/watch?v=eUVwIKGm-2Q8&list=PLuiPz6iU5SQ8ra5kixx770k_f
35	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	3.3.	41.34	4098	PRV_WK_3.3.	https://www.youtube.com/watch?v=IbBYDSUZEOc&list=PLuiPz6iU5SQ8ra5kixx770k_
36	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	4.1.	53.00	5324	PRV_WK_4.1.	https://www.youtube.com/watch?v=-4RvlUkANDg&list=PLuiPz6iU5SQ8ra5kixx770k_fa
37	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	4.2.	52.30	5086	PRV_WK_4.2.	https://www.youtube.com/watch?v=-8865N6e5vw&list=PLuiPz6iU5SQ8ra5kixx770k_fa
38	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	4.3.	29.29	2689	PRV_WK_4.3.	https://www.youtube.com/watch?v=hxFlNGuwvA8&list=PLuiPz6iU5SQ8ra5kixx770k_f
39	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	5.1.	41.47	4148	PRV_WK_5.1.	https://www.youtube.com/watch?v=-I3pMJ7XQcI&list=PLuiPz6iU5SQ8ra5kixx770k_f
40	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	5.2.	39.29	3872	PRV_WK_5.2.	https://www.youtube.com/watch?v=kia5AUuztDc&list=PLuiPz6iU5SQ8ra5kixx770k_fan
41	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	5.3.	35.36	3537	PRV_WK_5.3.	https://www.youtube.com/watch?v=MRqVvnVY8nM&list=PLuiPz6iU5SQ8ra5kixx770k_f
42	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	6.1.	47.58	4782	PRV_WK_6.1.	https://www.youtube.com/watch?v=hOyc8IOWaIM&list=PLuiPz6iU5SQ8ra5kixx770k_fa
43	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	6.2.	50.09	4873	PRV_WK_6.2.	https://www.youtube.com/watch?v=mlK8dxd1sRM&list=PLuiPz6iU5SQ8ra5kixx770k_fa
44	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	6.3.	42.46	4264	PRV_WK_6.3.	https://www.youtube.com/watch?v=oTLGf9ePzY8&list=PLuiPz6iU5SQ8ra5kixx770k_fan
45	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	7.1.	43.17	4339	PRV_WK_7.1.	https://www.youtube.com/watch?v=VjMBVq7Eh5Q&list=PLuiPz6iU5SQ8ra5kixx770k_f
46	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	7.2.	39.24	3802	PRV_WK_7.2.	https://www.youtube.com/watch?v=JB6ZriGKNuo&list=PLuiPz6iU5SQ8ra5kixx770k_fa
47	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	7.3.	32.50	3306	PRV_WK_7.3.	https://www.youtube.com/watch?v=T9oA05FiSUE&list=PLuiPz6iU5SQ8ra5kixx770k_fa
48	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	8.1.	42.08	4091	PRV_WK_8.1.	https://www.youtube.com/watch?v=9GiE-x0kelk&list=PLuiPz6iU5SQ8ra5kixx770k_fan
49	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	8.2.	32.14	3309	PRV_WK_8.2.	https://www.youtube.com/watch?v=LKd14aJK_AU&list=PLuiPz6iU5SQ8ra5kixx770k_fa
50	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	8.3.	56.38	5478	PRV_WK_8.3.	https://www.youtube.com/watch?v=EdN_Cvimr88&list=PLuiPz6iU5SQ8ra5kixx770k_fa
51	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	9.1.	49.32	4887	PRV_WK_9.1.	https://www.youtube.com/watch?v=HIY9UM3Cjre&list=PLuiPz6iU5SQ8ra5kixx770k_fa
52	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	9.2.	54.39	5218	PRV_WK_9.2.	https://www.youtube.com/watch?v=9QD7KK2hkFo&list=PLuiPz6iU5SQ8ra5kixx770k_fa
53	Electrical and Electronics Engineering	Probability and Random Variables	Elif Uysal BIYIKOĞLU	9.3.	34.17	3194	PRV_WK_9.3.	https://www.youtube.com/watch?v=OwgzrnHMtcI&list=PLuiPz6iU5SQ8ra5kixx770k_f

ÖZGEÇMİŞ

Adı Soyadı:	Özgür Reha ALICI
Uyruğu:	Türkiye Cumhuriyeti

EĞİTİM

Derece	**Kurum**	**Mezuniyet Tarihi**
Yüksek Lisans	Erciyes Üniversitesi	2024
Lisans	Erciyes Üniversitesi	2017
Lise	Kocasinan Anadolu Lisesi	2012

İŞ DENEYİMLERİ

Yıl	**Kurum**	**Görev**
2018-	Milli Eğitim Bakanlığı	İngilizce Öğretmeni
2017-2018	Kayseri Bilfen İlkokulu	İngilizce Öğretmeni

YABANCI DİL

İngilizce (96.25)