Applied Research on English Language

V. 10 N. 3 2021 pp: 137-160 http://jare.ui.ac.ir

DOI: 10.22108/ARE.2021.126390.1666 Document Type: Research Paper

An Analysis of the Multiword Units Presented in IELTS Speaking Preparation Books Published by Iranian Authors

Mansoor Ganji 1* , Peyman Nasrabady 2

¹ Assistant Professor, Department of English, Faculty of Management and Humanities, Chabahar Maritime University, Chabahar, Iran ² PhD Candidate, Department of English, Faculty of Management and Humanities, Chabahar Maritime University, Chabahar, Iran

Received: 2020/12/09 Accepted: 2021/04/21

Abstract: This study compared the use of lexical bundles in authentic spoken language data with those of two English learning textbooks developed for IELTS preparation courses in Iran. The aim was to see if the language chunks presented in the said textbooks were representative of the real-life language for which learners need to prepare. To achieve this aim, firstly, a list of lexical bundles was compiled based on the Michigan Corpus of Academic Spoken English (MICASE) and British Academic Spoken English (BASE). Further, the list was used as a reference tool to analyze the language data of two IELTS preparation books published and widely used in Iran. The results highlighted some considerable differences between MICASE and BASE with regard to the use of lexical bundles in terms of their frequency, structure, and function. Moreover, it was revealed that the books under investigation shared only a scarce number of similar lexical bundles with MICASE and BASE as a whole. Therefore, they failed to be representative of the authentic language that people use in the real world. Finally, the implications of the study are discussed, and several practical suggestions are made in order to inform teachers, material developers, and syllabus designers of the importance of related corpus linguistic studies.

Keywords: IELTS, Lexical Bundles, Speaking Skill, Spoken Corpora.

^{*} Corresponding Author.

¹ Mansoor Ganji (ganji@cmu.ac.ir)), ² Peyman Nasrabady (peyman.nasrabady1994@gmail.com)

Introduction

The International English Language Testing System (IELTS) is one of the main English proficiency tests individuals take for both migration and higher education purposes. Governments and universities expect individuals to demonstrate their English proficiency levels by taking a language proficiency test such as TOEFL, CAE, GRE, and IELTS; however, IELTS scores are one of the most highly reliable criteria in this regard. This test consists of listening, reading, speaking, and writing sub-tests (O'Sullivan, 2018), and test takers are given an average score of their scores on the four sub-tests, where 1 shows a 'non-user' while 9 is awarded to an 'expert user' (IELTS, 2019a).

Given the importance of IELTS for individuals with migration or higher education aims, it is necessary for teachers and material developers to carefully take all aspects of language into account. Instructionally comprehensive and pedagogically relevant activities should target the needs of learners in real contexts. Accordingly, they can help learners develop essential language skills they need when taking an IELTS exam or when living in English speaking countries (Farid & Saifuddin, 2018).

Real-life language is comprehensively and repeatedly shown to be comprised of large proportions of fixed multiword units (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Martinez & Schmitt, 2012; Wang, 2019). Interrupted (e.g., It is *that) or uninterrupted multiword units (e.g., It is important that) are basically explored on the basis of frequency and distribution in linguistic data by researchers under different terms, for instance, "memorized sequences" (Pawley & Syder 1983), "lexical bundles" (Biber et al., 1999), "n-grams" (Stubbs, 2007), "phrase-frames" (Fletcher, 2007), "formulaic frames" (Biber, 2009), "formulas" (Simpson-Vlach & Ellis, 2010), or "lexical frames" (Gray & Biber, 2013). Additionally, the importance of learning and using these units is frequently addressed in the literature for different language skills (see Wang, 2019 for multiword units in writing skill; Tavakoli & Uchihara, 2020 for the association of multiword units with speaking skill; Kim & Kim, 2012 for the effects of multiword units on reading skill; Tang, 2013 for the effectiveness of chunk acquisition for listening comprehension).

In the light of the foregoing, teachers and materials developers usually tend to rely on introducing multiword units of different functions for instructional and learning purposes (see Oshima & Hogue, 2006). However, some of the existing materials seem not to be based on reallife language corpora but on writers' intuitions (Bhatia, 2002; Swales, 2002). Additionally, they often fail to convincingly operationalize research findings in effective ways within their contents

(Harwood, 2005; Kuo, 1993). As a result, what some authors suggest in their books seems to fail to address the authentic language needs of learners (McEnery & Kifle, 2002).

Given the importance of using authentic language in teaching English and the lack of research on the books published for teaching IELTS speaking test, this study aims to explore the possible differences (if any) between learning materials which are put forward by experienced IELTS teachers and authors of IELTS books, and what actual real-world language encompasses. More particularly, adopting a corpus-based approach, this study aims to compare the most frequent multiword units in two IELTS preparation textbooks with those in an English reference corpus. To achieve this, a corpus of IELTS materials that target speaking skill is compiled based on the IELTS preparation books nationally published in Iran (*IELTS Speaking Tests* by Iravani, 2003, and *IELTS Speaking Ultimate* by Borhani & Hashemi, 2016). After that, a list of the most frequent multiword units suggested in these textbooks is generated from the corpus for further comparison. Following this, another multiword unit list is made based on two freely available English speech corpora, namely British Academic Spoken English (BASE) and Michigan Corpus Academic Spoken English (MICASE). Next, the two multi-word-unit lists are compared structurally and functionally to see if the multiword units in the textbooks are representative of real-life language, and further can fulfill learners' real-context language needs, particularly in the realm of speaking skill.

Review of the Literature

IELTS Exam

The importance and the widespread use of IELTS (Pearson, 2019) have made it one of the most standard tests. Besides, its fairness and predictive ability are already ensured (IELTS, 2015b) and confirmed through several studies in the literature (Schoepp, 2018; Thorpe, Snell, Davey-Evans, & Talman, 2017). The test has two versions: academic and general training. The former is taken by individuals applying for higher education, and the general version is considered a necessity for migration to Australia, Canada, and the United Kingdom (IELTS, 2019c). The test provides an assessment of listening, reading, writing, and speaking language skills. General and academic versions of the test are the same with regard to the speaking and listening skills, whereas they differ in how the reading and writing skills are assessed. For example, those applying for an academic program of study sit the IELTS with academic reading and writing modules, and those intending to migrate to an English-speaking country for other purposes take the general reading and writing modules (IELTS, 2019d; Wilson, 2010).

The speaking test is face-to-face interaction with a trained examiner. There are three parts in the speaking test of IELTS. In part one, candidates are asked general questions, intended to elicit personal information regarding their interests, families, and homes. In part two, with which this study is concerned, a verbal prompt is given to candidates on a topic card, which is designed to bring up a description of a particular topic. The candidates are given one minute to prepare and are then asked to talk on the topic for one to two minutes. In part three, the candidates are engaged in a discussion with the examiner on more abstract levels of the topic previously prompted. More details regarding the design and development of the IELTS speaking test can be found in Seedhouse and Nakatsuhara (2018). Some relevant studies in this regard have explored the linguistic characteristics of the oral language communicated within IELTS speaking tests. In a mixed-methods study, Read and Nation (2002) explored several lexical features (lexical sophistication, for example) followed by a qualitative analysis of the formulaic language used by IELTS candidates. They found that higher values of the lexical statistics were present in the transcriptions of the oral language of the candidates with higher IELTS bands. Besides, the use of formulaic expressions showed an increasing pattern from low-proficiency to high-proficiency candidates. In another study, Mirzaei, Hashemian, and Azizi Farsani (2016) analyzed the effect of teaching formulaic language to three intact classes comprising of 40 learners on IELTS speaking performance. The results showed that the method could to a large extent assist the development of speaking proficiency. However, the results were in favor of dialogic tasks more than monologic ones. Similarly, in a recent study, for developing tasks for the acquisition of the formulaic language, Goncharov (2019) gathered data through pre-and-post speaking tests administered in agreement with the IELTS speaking format. Not only were the results encouraging with respect to the learners' IELTS speaking performance, but also to their general speaking skill.

For one thing, the difficulty of gaining a high score on IELTS language subtests and the universities' requirements have placed a heavy demand on IELTS candidates. Thus, researchers and material developers are doing their best to design IELTS-oriented books which help students improve their performance in taking the test. Farid and Saifuddin (2018) conducted a study to identify the needs of low proficiency language users as the basis of IELTS writing material development. Instruments used to obtain data in this study were both a questionnaire and writing tests. The latter consisted of two tasks similar to those of IELTS: task 1 presented test takers with a figure to summarize the information, and task 2 was in an essay-elicitation format, requiring the candidates to write an essay in response to a certain topic. Further, the researchers found 10 and 8 common writing problems of the test takers performing the writing tasks 1 and 2 respectively. As a result, the materials they designed included activities that targeted the writing problems approached in this study to fulfill the language learners' needs.

IELTS Preparation Books, Speaking Fluency, and the Need for Considering Multiword Units

Furthermore, there have been books specifically designed for IELTS preparation purposes: New Insights into IELTS (Jakeman & McDowell, 2009); Objective IELTS Advanced (Capel & Black, 2006); IELTS Graduation (Allen, Powell, & Dolby, 2007) to name some. These books and some others are reviewed on the basis of what they cover, and target audience levels and needs by Wilson (2010). For example, *New Insights into IELTS* is "ideal for those teachers who have mixed IELTS classes of varying needs and levels" and also provides "a teacher who is new to IELTS with clear information about the test and the tasks" (p. 223).

In order to help learners with speaking skill, a number of books have been published in Iran, namely, *IELTS Speaking Tests* (Iravani, 2003), and *IELTS Speaking Ultimate* (Borhani & Hashemi, 2016). The former includes three chapters, specifically targeting parts 1, 2, and 3 of the IELTS speaking test respectively. The book covers a range of common topics for the IELTS exam followed by conversation tips and sample answers. Similarly, the latter contains three sections with the same purposes as the former. This book also covers a range of the most common topics for the IELTS Speaking test followed by categorized samples. These two books have remained unevaluated in the literature and are used by many IELTS candidates.

Theoretical Framework

Speaking performance and fluency is shown to be associated with the use of multiword units in the literature (Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; Stengers, Boers, Housen, & Eyckmans, 2011; Tavakoli, 2011; Tavakoli & Uchihara, 2020; Thomson, Boers, & Coxhead, 2017; Wood, 2009, 2010). This relationship can be explained through a psycholinguistic research point of view. This line of research suggests that multiword units (e.g., *in the middle of the*) are dealt with differently from novel language strings (e.g., *association is not a matter of*), with the former units being more advantageous than the latter when one processes them in both productive and receptive linguistic tasks (Siyanova-Chanturia & Van Lancker Sidtis, 2018). This increase in the language processing speed, which can enable

speakers to communicate language items more fluently, is shown to liberate the attentional resources speakers need to activate in favor of other aspects of language production resources such as articulation and monitoring (Kormos, 2006; Skehan, 2009). In other words, multiword units provide cost-effective and ready access to acceptable lexico-grammatical linguistic elements for learners, enabling them to move beyond their current language production capacity and creativity (Myles, Hooper, & Mitchell, 1998).

Further theoretical support for the relationship between the use of multiword units and fluency and their importance in language learning is the speech production model proposed by Levelt (1989), which was further refined to take L2 speakers into account by Kormos (2006). According to this model, three stages are, at least, involved in oral language production: conceptualization, formulation, and articulation. A pre-articulatory message which carries the speaker's communicative aim is generated and further encoded into an orderly abstract plan. Simultaneously, the about-to-send message is also monitored as far as the conceptualization stage is involved. The formulation stage is where the lexical considerations and grammatical encodings occur. The preverbal message moves into this stage to activate appropriate lexical items in the mental lexicon and place the items into appropriate grammatical surface structures. These linguistic items are further morpho-phonologically and phonetically encoded in this stage. At the articulation phase, the product from the previous stages is executed in a phonetic plan, and the speech is produced.

Compared to its L1 counterpart, the L2 mental lexicon is "smaller, less organized, likely slower in access, less elaborated with syntactic and collocational information, and contains a narrower repertoire of formulaic language" (Skehan, Foster, & Shum, 2016, p. 98). Accordingly, one way to free up attentional resources used at stages of oral language production processes is to develop a reasonable command of formulaic language (Kormos, 2006; Skehan, 2014).

It is apparent that the formulation stage or more specifically the lexical selection phase in speech production can benefit from the use of multiword units for speaking fluency (Kormos, 2006; Levelt, 1992). In the lexical selection phase, speakers rely on the mental lexicon to retrieve appropriate lemmas from the alternatives available in it. Longer multiword units, as opposed to single-word linguistic items with a similar processing cost, can be retrieved by speakers who have a large repertoire of multiword units at this phase. Doing so, they can save processing time in favor of other syntactic and message generation processing (Boers et al., 2006; Skehan, 1998). On the contrary, speakers who have a small amount of multiword

units in their mental lexicon may not benefit from this processing advantage since they need more cognitive resources when retrieving every constituent of the whole multiword units.

Based on the considerations stated above, it seems reasonable that learning materials targeting speaking skill need to pay careful attention to the multiword units. Material writers can achieve this either explicitly by giving lists of relevant multiword units or implicitly by using them with high frequency in spoken language samples provided. Therefore, this study aims at answering the following research questions:

- 1) What are the most frequent multiword units in MICASE and BASE Corpora as examples of spoken real-life language?
- 2) To what extent are the multiword units found in MICASE and BASE present in IELTS speaking preparation books nationally published in Iran, namely IELTS Speaking Tests and IELTS Speaking Ultimate?
- 3) How similar or different are these two sources of spoken English in terms of frequency of the multiword units, their structural characteristics, and their functional characteristics?

Methods

Corpus

Two existing corpora plus one compiled corpus are used in this study: The British Academic Spoken English corpus (BASE), the Michigan Corpus of Academic Spoken English (MICASE) (Simpson, Briggs, Ovens, & Swales, 2002), and IELTS Speaking Sample Answers Corpus (ISSAC). BASE corpus is developed under the directorship of Hilary Nesi and Paul Thompson. It is a 1.5 million-word corpus including the transcriptions of a variety of lectures and seminars recorded in different departments of the universities of Warwick and Reading. MICASE is a collection of transcribed speech (approx. 1.8 million words) from the University of Michigan comprising of a wide range of academic events such as seminars, advising sessions, and lectures. The researcher-compiled corpus, ISSAC, is based on two widely used IELTS preparation books in Iran. The first one, IELTS Speaking Tests with Answers and Sample Interviews, includes 40 'Items', each concerned with a sample IELTS speaking topic together with a one-paragraph answer. The latter, IELTS Speaking Ultimate, covers categorized samples of the IELTS speaking task (50 items) followed by definite answers to each.

The reasons for choosing BASE and MICASE as authentic sources of spoken English over others are: 1) Different studies exploring spoken language have referred to these corpora throughout the literature (Dang & Webb, 2014; Grant, 2011; Lee & Ziegeler, 2006; Lindemann & Mauranen, 2001; Nesi, 2002; Pastizzo & Carbone, 2007; Simpson & Mendis, 2003; & Yang, 2014) 2) BASE as a sample of British English and MICASE as a sample of American English were chosen to avoid bias in favor of each side, and 3) BASE and MICASE were freely available for language exploration (through Sketch Engine), and downloading for further analysis. ISSAC is compiled based on two books: IELTS Speaking TESTS and IELTS Speaking *Ultimate*, namely. The texts in ISSAC were written as intuited responses to IELTS speaking part 2 topics. For the most part, the language in these textbooks is introduced as oral language in the form of monologues. These two books are published in Iran as IELTS speaking test preparation guides for the candidates. The books are published in 2003 and 2016 respectively. They are two of the most frequently used textbooks in different IELTS preparation courses and centers held at several private language institutes in Iran. Table 1 presents more information regarding the transcripts and tokens of each corpus.

Table 1. Constituents of the Three Spoken Corpora (MICASE, BASE, and ISSAC)

Representation	Corpus	Word count	No. of texts
Authentic Academic Speech	BASE	1,511,207	200
Authentic Academic Speech	MICASE	1,761,511	152
Non-authentic Speech Samples	ISSAC	17,520	90

It seems that ISSAC is a relatively small corpus with regard to general research in the field of corpus linguistics. This is especially because we are handling a very particular type of discourse in a specific domain (intuited spoken texts in answer to sample IELTS speaking part 2 topics). Only the most frequently used textbooks were selected to represent an overall view of the spoken English discourse IELTS candidates in Iran are exposed to. The language presented in these textbooks is the English spoken register that students encounter the most often in IELTS courses in Iran. Therefore, ISSAC seemed more suitable for the identification of the relevant linguistic aspects.

Data Analysis Criteria

The primary purpose of this study is to make a comparison between the use of lexical bundles in authentic spoken English in academic contexts and those of spoken English prepared for IELTS preparation courses in Iran. To achieve this, BASE and MICASE were selected as the samples of authentic spoken language. Next, the most frequent lexical bundles used in BASE and MICASE were identified and analyzed with regard to their frequency, structure, and function. Following this, the n-gram counterpart of the lexical bundles previously found in BASE and MICASE were manually searched for in the ISSAC. Subsequently, the n-grams' structural and functional characteristics in ISSAC were compared to those of the lexical bundles in BASE and MICASE. It should be noted that, as ISSAC is relatively small, we compared the n-grams found in this corpus with their lexical bundle counterparts in two (instead of one) reliable reference corpora in order to ensure more reliable results.

Three basic criteria have been indicated in the previous literature concerning the analysis of lexical bundles. The first criterion considers the length of word sequences. To identify lexical bundles, researchers need to first decide on the length of the word sequences. Usually, 2, 3, 4, 5, 6, or 7-word sequences are considered for analysis in the literature, and this factor varies from study to study. The present study focuses only on the four-word lexical bundles because of three reasons: 4-word lexical bundles often contain 3-word lexical bundles within their structure as well, and offer more variation for analysis than 5-word lexical bundles (Cortes, 2004), they offer a more straightforward range of functional characteristics (Hyland, 2008), and they are perceived to bring forward a more manageable list for further analyses (Chen & Baker, 2010).

The next criterion is the cut-off frequency. This factor determines the number of times a 4-word sequence must occur repeatedly in a corpus data to be considered as a lexical bundle in further analysis. This threshold ranges between 20-40 times per million words in studies dealing with large corpora (e.g., Biber, Conrad, & Cortes, 2004; Hyland, 2008). It should also be mentioned that, for spoken corpora that are relatively small, a non-normalized cut-off frequency ranging from 2 to 10 is commonly used (e.g., De Cock, 1998). Accordingly, in order to adopt a conservative approach, the cut-off frequency was set to 30 times per million words to consider 4-word sequences as lexical bundles in this study.

The last criterion is called the range criterion which requires lexical bundles to occur, regardless of their frequency, in at least 3-5 different texts (e.g., Biber & Barbieri, 2007; Cortes, 2008). The same concern is also expressed by Hyland (2008). He believed that, in order to avoid individual writers' idiosyncratic tendencies of language use, a lexical bundle needs to occur in at least 10% of the texts. The range criterion was applied for lexical bundle identification in MICASE corpus since the text files were available for downloading and

manual analysis. Similarly, the criterion was also applied when analyzing the n-grams in ISSAC (which were the counterparts of the lexical bundles found in both BASE and MICASE) because the texts were available for one-by-one manual analysis. However, we were unable to include range criterion in the identification of the lexical bundles in BASE corpus due to the fact that the texts were only available for automated analysis through Sketch Engine (Kilgarriff et al., 2014) interface.

Data Collection and Analysis Procedure

To extract lexical bundles, the sketch engine (an online corpus linguistic tool) was employed. This interface allows the researcher to conduct corpus linguistic explorations in language data through a range of different functions. The n-gram function was used to yield a list of fourword lexical bundles out of the BASE corpus, which is one of the available corpora in the Sketch Engine by default. Following this, to identify four-word lexical bundles in the MICASE, the text files, after downloading from the corresponding MICASE website, were uploaded into the sketch engine as a user corpus. Later, the raw list of four-word lexical bundles compiled through the BASE and MICASE corpora were used as a reference list to enable a comparison between the lexical bundles used in samples of authentic spoken English (in BASE and MICASE) and those intuited in samples of non-authentic spoken English in ISSAC. To do so, the lexical bundles found in the reference corpora were manually searched for in ISSAC to find their 4-gram counterparts. This task was applied for all of the BASE and MICASE four-word lexical bundles in ISSAC. In case, they were also present in ISSAC, for every 4-gram counterpart of the four-word lexical bundles, further structural and functional comparisons were made through qualitative analyses of the related concordance lines. For example, at the same time is a lexical bundle found in both BASE and MICASE corpora. This lexical bundle was further manually searched for in ISSAC to see if its 4-gram counterpart (at the same time with no frequency and range criteria applied) was also present in this corpus. If yes, additional qualitative concordance analyses were applied to see if the n-gram found in all the three corpora demonstrated the same structural and functional characteristics in ISSAC as its lexical bundle counterpart does in BASE and MICASE or not. The reason for calling them 'n-gram' (instead of a lexical bundle) in ISSAC is due to the fact that ISSAC is a relatively small corpus; therefore, frequency and range criteria were not applied to the word sequences analyzed in this corpus.

The structural analysis of the lexical bundles was carried out based on the structural types identified by Biber et al. (2004). The classification divides lexical bundles into three structural types: (1) lexical bundles that carry verb phrase fragments like is based on the, have a lot of, (2) lexical bundles that contain dependent clause fragments like *if you look at, to be able to*, and (3) lexical bundles that carry noun phrase and prepositional phrase fragments like *a little bit more*, *at the end of*. Additionally, as presented in Table 2, each major type involves different structural sub-types.

Table 2. Structural Types of Lexical Bundles (Biber et al., 2004, p. 381).

Structural types	Structural sub-types	Examples
1. Lexical bundles that incorporate verb phrase fragments	1a. (connector +) 1st/2nd person pronoun + VP fragment	you don't have to
-	1b. (connector +) 3rd person pronoun + VP fragment	that's one of the
	1c. Discourse marker + VP fragment	you know it was
	1d. Verb phrase (with a non-passive verb)	is going to be
	1e. Verb phrase (with a passive verb)	can be used to
	1f. Yes-no question fragments	do you want to
	1g. WH-question fragments	how many of you
2. Lexical bundles that incorporate dependent clause fragments	2a. 1st/2nd person pronoun + dependent clause fragment	I don't know why
	2b. WH-clause fragments	what I want to
	2c. If-clause fragments	if you want to
	2d. (verb/adjective+) To-clause fragment	to come up with
	2e. That-clause fragment	that there is a
3. Lexical bundles that incorporate noun phrase and prepositional	3a. (connector +) Noun phrase with of- phrase fragment	the end of the
phrase fragments	3b. Noun phrase with other post-modifier fragment	a little bit about
	3c. Other noun phrase expressions	a little bit more
	3d. Prepositional phrase expressions	of the things that
	3e. Comparative expressions	as far as the

As regards the functional analysis of the lexical bundles, the taxonomy developed by Biber et al. (2004) was applied. According to this taxonomy, lexical bundles can serve three primary functions: (1) stance bundles that convey attitudes or assessments of other propositions

like are more likely to, the fact that the (2) discourse organizers that demonstrate the relationships between the texts of discourses like take a look at, what I want to, and (3) referential bundles that refer to physical or abstract things, or to other textual contexts like that's one of the, the rest of the. Besides, each functional category entails different subcategories conveying specific functions and meanings. Table 3 presents more information and examples regarding the functional taxonomy of the lexical bundles. It should be noted that, in order to apply a functional analysis on the lexical bundles and n-grams of this study, the concordance function of the sketch engine tool was used. This function provides more textual contexts for the lexical bundles/n-grams found in the corpora on which this study is based and, hence enabled a manual and more in-depth analysis of the word sequences of interest in this study.

Table 3. Functional taxonomy of lexical bundles proposed by Biber et al. (2004, pp. 386-388)

Functional	Sub-categories	Examples
I. Stance bundles	Epistemic stance (personal/impersonal) Attitudinal/modality stance	I don't know if, are more likely to
	(B1) Desire (personal)	do you want a
	(B2) Obligation/directive (personal/impersonal)	and you have to, it is necessary to
	(B3) Intention/prediction (personal/ impersonal)	I was going to, it is going to be
	(B4) Ability (personal/impersonal)	to be able to, can be used to
II. Discourse	A. Topic introduction	going to talk about
	B. Topic elaboration/clarification	has to do with
III. Referential bundles	A. Identification/focus	and this is a
	B. ImprecisionC. Specification of attributes	and things like that
	(C1) Quantity specification	have a lot of
	(C2) Tangible framing	the size of the
	(C3) Intangible framing D. Time/place/text reference	the nature of the
	(D1) Place reference	in the United States
	(D2) Time reference	at the time of
	(D3) Text-deixis	as shown in figure
	(D4) Multi-functional reference	at the end of

Results and Discussion

The above threshold resulted in the identification of a total of 58 and 49 lexical bundles in BASE and MICASE, respectively (Appendix A is a full list of lexical bundles in each corpus). As can be seen in Appendix A, the lexical bundles in MICASE are fairly more frequent in comparison to those of BASE. In other words, the sum of the lexical bundle frequencies in MICASE is far greater than that of the BASE (7312 vs. 4835).

Before presenting and discussing the results of the analysis of this study, it should be noted that, as referred to before, two academic spoken corpora are compared with a less formal spoken corpus; therefore, the results of this study should be interpreted cautiously. They are mainly used to enable a comparison between the lexical bundles that constitute the discourse of spoken in English in authentic versus intuited samples of oral English communication. The structural distribution of the patterns of the 4-word lexical bundles in BASE and MICASE are shown in Figure 1.

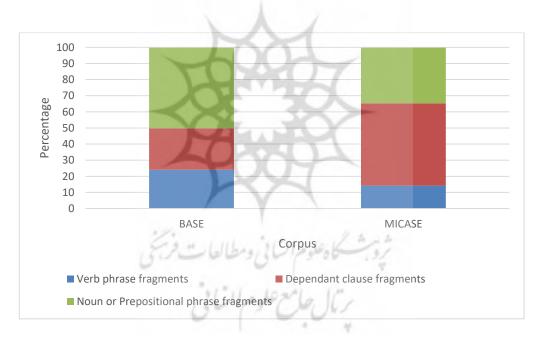


Figure 1. Structural Distribution of Lexical Bundles in BASE and MICASE

As can be seen in Figure 1, the distribution patterns of structural categories in BASE and MICASE are different in all but one way. The only similarity between the two corpora exists in the fact that lexical bundles with verb phrase fragments contribute the least, in comparison to other structural categories, to the formulaic discourse of spoken English in both corpora. On the other hand, while lexical bundles with the noun or prepositional phrase fragments are the most frequent type in BASE (50%), they are used with less frequency in MICASE (almost

37%). Further, the distributions of the lexical bundles with dependent clause fragments go in opposite directions in BASE and MICASE. While lexical bundles with dependent clause category are the most frequent structural type in MICASE, it has ranked the second type in BASE, with 51%, and almost 26% distribution in MICASE and BASE, respectively.

Comparing the lexical bundles between ISSAC and the two said corpora, it was found that fifty-four (93%) of the four-word lexical bundles found in BASE do not appear at all in ISSAC. Of the remaining 4 items, is one of the appears in the top 10 of the lexical bundles in ISSAC, while it has ranked the 26th most frequent one in BASE. Further lexical bundles that are shared by both BASE and ISSAC include the end of the, at the same time, and, and one of the. Additionally, not only was there a small number of shared lexical bundles, there was also a lack of enough variation with regard to the structural types in ISSAC. Hence, while ISSAC includes at least one lexical bundle with verb phrase fragments (is one of the), and 3 lexical bundles with the noun or prepositional phrase fragments (the end of the, at the same time, and (and one of the), no lexical bundle with dependent clause fragments appeared in ISSAC. Altogether, all the lexical bundles found in this study share the same major characteristics the lexical bundles are believed to have. In fact, most of them are not idiomatic, their meanings are perceptually salient from their individual words, and they usually do not consist of a complete structural unit (they cannot stand alone as complete sentences) (Biber & Barbieri, 2007). The only complete structural unit found as a lexical bundle (including the frequency and the range criteria) was does that make sense which appeared in MICASE. Similarly, this finding is in line with what was found by Biber, et al. (1999). They asserted that a very low proportion of the lexical bundles found in conversation (15%) can be judged as a complete structural unit. The functional distribution of the lexical bundles found in both BASE and MICASE are ربال حامع علوم الساني presented in Figure 2.

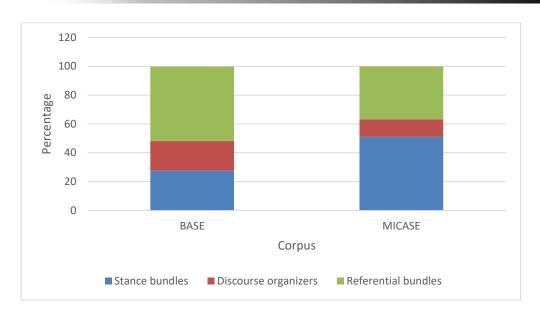


Figure 2. Functional Distribution of Lexical Bundles in BASE and MICASE

As shown in Figure 2, functional distribution patterns of lexical bundles in BASE and MICASE do not exhibit any similarity except for one way. In both of them, lexical bundles that carry discourse organizing functions are the least frequent type. On the other hand, while referential bundles are the most frequent functional category used in BASE (almost 52%), they are used with less frequency in MICASE. Finally, lexical bundles that carry stance features are the most frequently used ones in MICASE (just above half of all lexical bundles in this corpus), while they are considerably less frequent in BASE (27.5%).

As stated before, only four lexical bundles are shared between ISSAC and the two other corpora. Among them, at the same time is the only lexical bundle with discourse organizing function, while the other three (the end of the, is one of the, and one of the) belong to the referential category. Consequently, the lack of enough variation, which was present with regard to the structural distribution of lexical bundles in ISSAC, is similarly noticeable in the functional distribution of them. In particular, stance lexical bundles which are the most frequent in MICASE do not appear at all in ISSAC. In comparison with BASE and MICASE, neither of the authors of the books the content of which are compiled into the ISSAC (IELTS Speaking TESTS and IELTS Speaking Ultimate) seem to have recognized the importance of stance bundles in English spoken language. Biber and Barbieri (2007), Chen and Baker (2010), and Jablonkai (2010) argue stance bundles can perform different functions in the discourse of the English language. According to Biber et al. (2004), stance bundles can further demonstrate

epistemic stance, desire, obligation, intention, and ability-related identities. Some examples of these bundles extracted from BASE are given below:

(1) and if you want to be really definite you can blow some oxygen through it because what we tend to do

In this extract, if you want to is used to display a personal desire with regard to a tendency to be 'definite'.

(2) pay attention you guys cause you're going to have to remember all these names

In extract (2), going to have to as a frequent lexical bundle in BASE is used to demonstrate a sense of obligation, and to assert that the addressees should remember something.

(3) we have to know the previous two numbers to be able to work out the next one In extract (3), to be able to is used to refer to an ability that is needed in order to move to another task.

As shown in the examples above, stance bundles can contribute with considerable variation to the discourse of spoken English. Surprisingly, they are absent in the ISSAC.

Based on the findings, the absence of any significant similarity between the lexical bundles used in ISSAC and those used in BASE/MICASE can question the validity of the books which provided the language data for ISSAC. Additionally, as Biber et al. (1999) asserted that lexical bundles contribute considerably to the discourse of English, this lack of enough attention to the lexical bundles in preparing the books IELTS Speaking TESTS and *IELTS Speaking Ultimate* as sources of IELTS preparation materials seems to be inexcusable.

Therefore, the findings of this study carry implications for material developers as well as English teachers. First of all, the structural and functional gaps revealed through the present study suggest that the textbooks under investigation seem not to be representative of the authentic spoken English IELTS candidates possibly encounter. More particularly, Biber et al. (1999) maintain that both oral and written languages utilize a large body of lexical bundles, with oral language containing more lexical bundles than written form. Therefore, it is necessary for material developers to consult with corpora of authentic language or corpus linguists in order to gather data and produce books based on frequent language patterns that learners are more likely to encounter in real-life contexts. Second, it is essential for language teachers to model authentic conversations based on the structural and functional characteristics of frequent lexical bundles which are found to be of high frequency in oral language, personal pronoun followed by a lexical verb phrase (+ complement clause) or I don't know what (Biber et al.,

1999), for example. This is comparable with related findings in the literature. For example, Zipagan, and Lee (2018), exploring the use of lexical bundles in speaking by Korean language learners, found that language learners need more proper and explicit guidance with regard to the correct use of lexical bundles with different functions. Additionally, explicit teaching and more frequent exposure to lexical bundles with particular structures and functions will help language learners to build a large repertoire of specific lexical bundles based on the specific needs of particular contexts. This will help learners, while speaking, save more mental processing time and develop their speaking fluency (Boers et al., 2006). However, the results of this study need to be interpreted with some caution since the resource corpora (BASE and MICASE) and the compiled corpus (ISSAC) on which this study is based are not perfectly comparable. In particular, BASE and MICASE contain language data based on the oral language used in academic contexts, in university lectures, for example, while ISSAC is composed of the oral language specifically intuited as possible answers for IELTS speaking part 2 topics. This limitation is because of the fact that the researchers did not have access to an exactly similar linguistic resource for the analysis. However, as referred to before, in order to ensure more reliable results, the researchers have chosen two samples of authentic oral language (BASE and MICASE) instead of one.

Conclusion

This study has matched the lexical bundles, their structural, and functional characteristics found in two IELTS speaking preparation books published in Iran with those of BASE and MICASE as samples of authentic language. It contributes to the related line of research in spoken corpus linguistic research, shedding light on structural and functional aspects of the frequent lexical bundles which have influenced/are possibly influencing the form of the discourse of authentic spoken English.

Lexical bundles with different structural and functional characteristics were identified in the language data of BASE, MICASE, and ISSAC corpora. Firstly, with regard to frequency, the findings revealed that lexical bundles captured a higher proportion of MICASE texts compared to the texts in the BASE corpus. This suggests that the spoken language in the Warwick and the Reading universities is less formulaic than that of the University of Michigan. Secondly, as far as the structural analysis of the lexical bundles is concerned, the findings showed that MICASE texts made higher use of dependent clause fragments, while BASE

language data contained more verb phrase fragments. The comparable difference was further noticed concerning the functional analysis of the lexical bundles. Referential bundles were the most frequent in BASE data, but stance bundles were mainly recurrent in MICASE. This finding is in line with and can confirm an assertion previously made in the literature. Taking a genre and disciplinary variation view towards the use of lexical bundles in spoken academic language, Wang (2017) stated that lexical bundles can be regarded as a useful lens to look through in order to capture genre and disciplinary variations in different language data. Further, he notes that lexical bundles "may be used to help the newcomers get familiar with the conventions of their own field of study and achieve fluency more quickly" (p. 208). Additionally, a more noticeable difference was perceived when the lexical bundles identified in BASE and MICASE were looked for in ISSAC. A very small number of lexical bundles (only 4 items) in ISSAC matched those of the BASE and MICASE. This, more significantly, indicates that the books upon which ISSAC is compiled seem to lack several crucial elements. According to the communicative language teaching approach, to help the development of communicative competence in language learners, instructional materials need to represent reallife language to fulfill learners' communicative needs (Richards & Rodgers, 2014).

Therefore, based on the results of this study, it seems that material designers in the field of English language teaching need to pay more attention to the features of their target context and audience. In other words, lexical bundles, according to the findings of this study, were used with high frequency and particular structural and functional characteristics in BASE (as a sample of English in a British context) and MICASE (as a sample of English in an American context). This further can imply that textbook authors need to consider more carefully the context their book is going to represent as well the needs of their audience (textbook readers or language learners). To do so, consulting the findings of corpus linguistic explorations such as the present study can actively help material designers to be able to base their textbooks on real-life language data.

The first limitation of this study is that the books under investigation contained a limited number of tokens; therefore, it seems unrealistic to expect such a small amount of language data parallel linguistically with the considerable amount of language data available in BASE and MICASE in terms of the frequency with which lexical bundles recur. However, one would expect, at least, various types of lexical bundles, regardless of their frequency, in a book designed for language teaching/learning aims. Such a linguistic shortage caused by an insufficient number of authentic multiword units can raise material developers' awareness of the need to enrich their books with adequate and more suitable data representative of the kind of language they target. In addition, syllabus designers and teachers should focus on materials that are developed and designed more carefully with regard to the needs of the students.

Moreover, it can be asserted that lexical bundles can be used as a descriptive tool to analyze oral language in books as well as real-life contexts. Thus, similar research can offer interesting insights into spoken English both at structural and functional levels. For example, further research can undertake the analyses of authentic language data based on the stance and engagement interaction model proposed by Hyland (2005). This area of research can consistently inform teachers, material developers, and students of the conventionalized ways of making meaning in real-life language.

References

- Allen, M., Powell, D., & Dolby, D. (2007). *IELTS Graduation*. Oxford: Macmillan.
- Bhatia, V. K. (2002). Applied Genre Analysis: A Multi-Perspective Model. *Ibérica*, 4, 2-19.
- Biber, D. (2009). A Corpus-Driven Approach to Formulaic Language in English: Multi-Word Patterns in Speech and Writing. *International Journal of Corpus Linguistics*, 14 (3), 275-311.
- Biber, D., & Barbieri, F. (2007). Lexical Bundles in University Spoken and Written Registers. *English for Specific Purposes*, 26(3), 263-286.
- Biber, D., Conrad, S., & Cortes, V. (2004). If You Look at...: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, 25(3), 371-405.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman Grammar of Spoken and Written English*: MIT Press Cambridge, MA.
- Boers, F., Eyckmans, J., Kappel, K., Stengers, H., & Demecheleer, M. (2006). Formulaic Sequences and Perceived Oral Proficiency: Putting a Lexical Approach to the Test. Language Teaching Research, 10, 245-261.
- Borhani, A., & Hashemi, Z. (2016). *IELTS Speaking Ultimate*. Tehran: Hadafe-Novin Publication.
- Capel, A., & Black, M. (2006). *Objective IELTS Advanced Self-Study Student's Book with CD ROM*. Cambridge: Cambridge University Press.
- Chen, Y. H., & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning and Technology*, 14(2), 30-49.

- Cortes, V. (2004). Lexical Bundles in Published and Student Disciplinary Writing: Examples from History and Biology. English for Specific Purposes, 23(4), 397-423.
- Cortes, V. (2008). A Comparative Analysis of Lexical Bundles in Academic History Writing in English and Spanish. *Corpora*, 3(1), 43-57.
- Dang, T. N. Y., & Webb, S. (2014). The Lexical Profile of Academic Spoken English. English for Specific Purposes, 33, 66-76.
- De Cock, S. (1998). A Recurrent Word Combination Approach to the Study of Formulae in the Speech of Native and Non-Native Speakers of English. *International Journal of Corpus* Linguistics, 3(1), 59-80.
- Farid, A., & Saifuddin, M. (2018). Designing IELTS Writing Material for Learners with Low Level of English Proficiency Based on Needs Analysis. Journal of Research in Foreign Language Teaching (JRFLT), 1(1), 49-61.
- Fletcher, W. H. (2007). Concordancing the Web: Promise and Problems, Tools and Techniques. In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), Corpus Linguistics and the Web (pp. 7–24). Amsterdam: Rodopi.
- Goncharov, G. (2019). The Effect of Direct Instruction in Formulaic Sequences on IELTS Students' Speaking Performance. Advanced Education, 11, 30-39.
- Grant, L. E. (2011). The Frequency and Functions of Just in British Academic Spoken English. Journal of English for Academic Purposes, 10(3), 183-197.
- Gray, B., & Biber, D. (2013). Lexical Frames in Academic Prose and Conversation. *International Journal of Corpus Linguistics*, 18(1), 109-136.
- Harwood, N. (2005). What Do We Want EAP Teaching Materials for? Journal of English for *Academic Purposes*, 4(2), 149-161.
- Hyland, K. (2005). Stance and Engagement: A Model of Interaction in Academic Discourse. *Discourse Studies*, 7(2), 173-192.
- Hyland, K. (2008). As Can Be Seen: Lexical Bundles and Disciplinary Variation. English for Specific Purposes, 27(1), 4-21.
- IELTS. (2019a). IELTS website. Retrieved from: https://www.ielts.org.
- IELTS. (2019b). IELTS website. Retrieved from: https://www.ielts.org.
- IELTS. (2019c). IELTS website. Retrieved from: https://www.ielts.org.
- IELTS. (2019d). IELTS website. Retrieved from: https://www.ielts.org.
- Iravani, H. (2003). IELTS Speaking Tests with Answers and Sample Interviews. Tehran: Zabankadeh Publication.

- Jakeman, V., & McDowell, C. (2009). *New Insight into IELTS: Student's Book with Answers*. Cambridge: Cambridge University Press.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten Years on. *Lexicography*, 1(1), 7-36.
- Kim, S., & Kim, J. (2012). Frequency Effects in L2 Multiword Unit Processing: Evidence from Self-Paced Reading. *TESOL Quarterly*, 46(4), 831-841.
- Kormos, J. (2006). *Speech Production and Second Language Acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Kuo, C. H. (1992). Problematic Issues in EST Materials Development. English for Specific Purposes, 12(2), 81-171.
- Lee, S. & Ziegeler, D. (2006). Analyzing a Semantic Corpus Study across English Dialects: Searching for Paradigmatic Parallels. In A. Wilson, D. Archer and P. Rayson (Eds), *Corpus Linguistics around the World* (pp. 121–39). Amsterdam: Rodopi.
- Levelt, W. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press.
- Levelt, W. J. (1992). Accessing Words in Speech Production: Stages, Processes and Representations. *Cognition*, 42, 1-22.
- Lindemann, S., & Mauranen, A. (2001). It's Just Real Messy": The Occurrence and Function of Just in a Corpus of Academic Speech. *English for Specific Purposes*, 20, 459-475.
- Martinez, R., & Schmitt, N. (2012). A Phrasal Expressions List. *Applied Linguistics*, 33(3), 299-320.
- McEnery, T., & Kifle, N. (2002). Epistemic Modality in Argumentative Essays of Second Language Writers. In J. Flowerdew (Ed.), *Academic Discourse* (pp. 182–195). London: Longman, Pearson Education.
- Mirzaei, A., Hashemian, M., & Azizi Farsani, M. (2016). Lexis-Based Instruction and IELTS Candidates' Development of L2 Speaking Ability: Use of Formulaicity in Monologic Versus Dialogic Task. *Journal of Teaching Language Skills*, 35(2), 69-98.
- Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or Rule? Exploring the Role of Formulaic Language in Classroom Foreign Language Learning. *Language Learning*, 48(3), 323-364.

- Nesi, H. (2002). An English Spoken Academic Wordlist. Paper presented at the EURALEX 2002, Copenhagen, Denmark. Retrieved from: http://www.euralex.org/elx_proceedings/Euralex2002/036_2002_V1_Hilary%20
- Oshima, A., & Hogue, A. (2006). Writing Academic English. White Plains, NY: Pearson Longman.
- O'Sullivan, B. (2018). IELTS (International English Language Testing System). In J. I. Liontas and M. DelliCarpini (Eds.), The TESOL Encyclopedia of English Language Teaching. International Association.
- Pastizzo, M. J., & Carbone, R. F. (2007). Spoken Word Frequency Counts Based on 1.6 Million Words in American English. *Behavior Research Methods*, 39(4), 1025-1028.
- Pawley, A., & Syder, F. H. (1983). Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency. In J. C. Richards and R. W. Schmidt (Eds.), Language and Communication (pp. 191–225). London: Longman.
- Pearson, W. S. (2019). Critical Perspectives on the IELTS Test. ELT Journal, 73(2), 197-206.
- Read, J., & Nation, P. (2006). An Investigation of the Lexical Dimension of the IELTS Speaking Test. IELTS Research Reports 6. IELTS Australia/British Council207–231.
- Richards, J. C., & Rodgers, T. S. (2014). Communicative Language Teaching. Approaches and Methods in Language Teaching. Third Edition. Cambridge: Cambridge University Press.
- Schoepp, K. (2018). Predictive Validity of the IELTS in an English as a Medium of Instruction Environment. Higher Education Quarterly, 72(4), 271-285.
- Seedhouse, P., & Nakatsuhara, F. (2018). The Discourse of the IELTS Speaking Test: Interactional Design and Practice. Cambridge: Cambridge University Press.
- Simpson, R. C., Briggs, S. L., Ovens, J., & Swales. J. M. (2002). The Michigan Corpus of Academic Spoken English. Ann Arbor, MI: The Regents of the University of Michigan.
- Simpson, R., & Mendis, D. (2003). A Corpus-Based Study of Idioms in Academic Speech. TESOL Quarterly, 37(3), 419-441.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, 31(4), 487–512.
- Siyanova-Chanturia, A., & Van Lancker Sidtis, D. (2018). What on-Line Processing Tells us about Formulaic Language? In A. Siyanova-Chanturia and A. Pellicer-Sánchez (Eds.) *Understanding Formulaic Language: A Second Language Acquisition Perspective* (pp. 38-61). London, New York: Routledge.

- Skehan, P. (2009). Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Applied Linguistics*, 30(4), 510-532.
- Skehan, P. (2014). *Processing Perspectives on Task Performance*. Amsterdam: John Benjamins Publishing Company.
- Skehan, P., Foster, P., & Shum, S. (2016). Ladders and Snakes in Second Language Fluency. *International Review of Applied Linguistics in Language Teaching*, 54(2), 97-111.
- Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2011). Formulaic Sequences and L2 Oral Proficiency: Does the Type of Target Language Influence the Association?. *International Review of Applied Linguistics in Language Teaching*, 49(4), 321-343.
- Stubbs, M. (2007). An Example of Frequent English Phraseology: Distributions, Structures and Functions. In R. Facchinetti (Ed.), *Corpus Linguistics 25 Years on* (pp. 89–105). Amsterdam/New York: Rodopi.
- Swales, J. M. (2002). Integrated and Fragmented Worlds: EAP Materials and Corpus Linguistics. In J. Flowerdew (Ed.), *Academic Discourse* (pp. 150–164). London: Longman, Pearson Education.
- Tavakoli, P. (2011). Pausing Patterns: Differences Between L2 Learners and Native Speakers. *ELT Journal*, 65(1), 71-79.
- Tavakoli, P., & Uchihara, T. (2020). To What Extent Are Multiword Sequences Associated with Oral Fluency?. *Language Learning*, 70(2), 506-547.
- Thomson, H., Boers, F., & Coxhead, A. (2019). Replication Research in Pedagogical Approaches to Spoken Fluency and Formulaic Sequences: A Call for Replication of Wood (2009) and Boers, Eyckmans, Kappel, Stengers, and Demecheleer (2006). *Language Teaching*, 52(3), 406-414.
- Thorpe, A., Snell, M., Davey Evans, S., & Talman, R. (2017). Improving the Academic Performance of Nonnative English Speaking Students: The Contribution of Pre Sessional English Language Programmes. *Higher Education Quarterly*, 71(1), 5-32.
- Wang, Y. (2017). Lexical Bundles in Spoken Academic ELF: Genre and Disciplinary Variation. International Journal of Corpus Linguistics, 22(2), 187-211.

- Wang, Y. (2019). A Functional Analysis of Text-Oriented Formulaic Expressions in Written Academic Discourse: Multiword Sequences vs. Single Words. English for Specific Purposes, 54, 50-61.
- Wilson, J. (2010). Recent IELTS Materials. ELT Journal, 64(2), 219-232.
- Wood, D. (2009). Effects of Focused Instruction of Formulaic Sequences on Fluent Expression in Second Language Narratives: A Case Study. Canadian Journal of Applied Linguistics, 12, 39-57.
- Wood, D. (2010). Formulaic Language and Second Language Speech Fluency: Background, Evidence, and Classroom Applications. London, New York: Continuum.
- Yang, S. (2014). Interaction and Codability: A Multi-Layered Analytical Approach to Discourse Markers in Teacher's Spoken Discourse. In J. Romero-Trillo (Ed.), Yearbook of Corpus Linguistics and Pragmatics 2014 (pp. 291–314). Dordrecht: Springer.
- Zipagan, M. N., & Lee, K. R. (2018). Korean English Learners' Use of Lexical Bundles in Speaking. Journal of Asia

TEFL, 15(2), 276-291.