

System

Multiword sequences in English language learners' speech: The relationship between trigrams and lexical variety across development --Manuscript Draft--

Manuscript Number:	
Article Type:	Full Length Article
Keywords:	constructions; EAP; multiword expressions; MWS; learner corpus; lexical variety; longitudinal data; L2 speech; TESOL; trigrams
Corresponding Author:	Mary Lou Vercellotti, PhD Ball State University Muncie, IN UNITED STATES
First Author:	Mary Lou Vercellotti
Order of Authors:	Mary Lou Vercellotti Alan Juffs Ben Naismith
Abstract:	Languages have formulaic multiword sequences (MWSs) that occur repeatedly in speech and writing (e.g., Nattinger & DeCarrico, 1992; Siyanova-Chanturia & Pellicer-Sanchez, 2018). For learners, then, the production of MWSs is an important element in developing spoken language which is complex, accurate, and fluent. However, though the use of MWSs are important for achieving spoken proficiency, it is unclear whether the production of MWSs supports or hinders another aspect of proficiency, lexical variety. This paper is an exploration of the production of MWSs and the development of lexical variety, found in two-minute speeches (n = 294) from English L2 learners (n = 66) over time in an intensive English program (IEP). In this study, MWSs were operationalized as recurrent three-word trigrams. Using hierarchical linear modeling and correlation analysis, we found different patterns of development for the two measures. The use of MWSs increased and then decreased while the lexical variety scores slightly decreased and then sharply increased over time in the IEP. Although the impact of MWSs on oral fluency has been studied, this is the first study to consider how MWSs affect lexical variety.
Suggested Reviewers:	Aziz Yuldashev Pennsylvania State University ayuldashev@psu.edu Author of relevant paper, Second Language Learners' Contiguous and Discontiguous Multi-Word Unit Use Over Time. Søren Wind Eskildsen, PhD Associate Professor, University of Southern Denmark swe@sitkom.sdu.dk Author of relevant paper, Are recurring multi-word expressions really syntactic freezes? Second language acquisition from the perspective of usage-based linguistics.

Highlights

- Use of multi-word-sequences increased and then decreased over time
- Lexical variety scores slightly decreased and then sharply increased over time
- A negative relationship was found generally, but not for all learners

Multiword sequences in English language learners' speech: The relationship between trigrams and lexical variety across development

Mary Lou Vercellotti^{a*}, Alan Juffs^b, and Ben Naismith^b

^a 2000 W University Ave, 297 Robert Bell Building, Ball State University, Muncie, IN 47306, USA

^b 4200 Fifth Avenue, 2817 Cathedral of Learning, Department of Linguistics, University of Pittsburgh, Pittsburgh, PA 15260, USA

*Corresponding author. 2000 W University Ave, 297 Robert Bell Building, Ball State University, Muncie, IN 47306, USA, +1 765-285-8389

E-mail addresses: mlvercellott@bsu.edu (M. L. Vercellotti), juffs@pitt.edu (A. Juffs), bnaismith@pitt.edu (B. Naismith)

Acknowledgments: The initial data collection (speeches) for this research was funded by the National Science Foundation, Grant Number SBE-0836012 to LearnLab

<http://www.learnlab.org>.

Multiword sequences in L2 English learners' speech: The relationship between trigrams and lexical variety across development

Abstract

Languages have formulaic multiword sequences (MWSs) that occur repeatedly in speech and writing (e.g., Nattinger & DeCarrico, 1992; Siyanova-Chanturia & Pellicer-Sanchez, 2018). For learners, then, the production of MWSs is an important element in developing spoken language which is complex, accurate, and fluent. However, though the use of MWSs are important for achieving spoken proficiency, it is unclear whether the production of MWSs supports or hinders another aspect of proficiency, lexical variety. This paper is an exploration of the production of MWSs and the development of lexical variety, found in two-minute speeches ($n = 294$) from English L2 learners ($n = 66$) over time in an intensive English program (IEP). In this study, MWSs were operationalized as recurrent three-word trigrams. Using hierarchical linear modeling and correlation analysis, we found different patterns of development for the two measures. The use of MWSs increased and then decreased while the lexical variety scores slightly decreased and then sharply increased over time in the IEP. Although the impact of MWSs on oral fluency has been studied, this is the first study to consider how MWSs affect lexical variety.

Keywords

constructions; EAP; multiword expressions; MWE; MWS; learner corpus; lexical variety; longitudinal data; L2 speech; TESOL; trigrams

1. Introduction

It is well documented that language production includes formulaic language, and that *multiword sequences* (MWSs; also known as *multiword expressions*) occur repeatedly in speech and writing. (See Siyanova-Chantura & Pelicer-Sanchez (2018) for a complete discussion.) Despite (or because of) their ubiquity, researchers have proposed multiple definitions and uses of MWSs. These MWSs have been called “prefabricated routines” (Brown, 1973), lexical phrases (Nattinger & DeCarrico, 1992), and chunked sequences (Ellis, 1996). In her influential paper, Wray (2000, p. 464) defined formulaic sequences broadly as set or semi-set MWSs, such as idioms, collocations, and sentence frames but cautioned that such utterances are not “a single phenomenon, but ... a set of more and less closely related ones...”. The use of many different terms (and the same terms being used divergently) has created some confusion when trying to understand their role in language learning (Wray, 2012). MWSs, as any construction of more than one word, is the broadest term, and we will use it here, in line with Tavakoli and Uchiyama (2020).

In L2 learning, MWSs have attracted increased interest in the past twenty years both theoretically and pedagogically (e.g., Ellis et al., 2008; Martinez & Schmitt, 2012). Researchers have discussed three possible functions of formulaic language: as minimal communication, as a means to ease production, and as a learning strategy (Weinert, 1995). First, minimal communication MWSs (e.g., *how are you?*) serve pragmatic functions (e.g., Girard & Sionis, 2003). Second, since MWSs are expected to be partially or wholly stored, speakers may use MWSs for efficient language performance (Wray, 2000) and faster language processing (Dabrowska, 2004). In other words, language learners benefit from MWSs, particularly in oral language performance of fluency, because a “chunk” of language is retrieved together (Boers et

al., 2006; Skehan, 1998). Additionally, MWSs allow second language (L2) learners to sound more native-like (Pawley & Syder, 1983) and more proficient (Boers et al., 2006).

Understandably, much research on MWSs (e.g., Boers et al., 2006; Kuiper, 2004; Tavakoli & Uchihara, 2020; Wood, 2007) has considered the effect on speaker fluency because of the expectation of improved fluency with the production of MWSs. Nevertheless, fluency is but one aspect of language production. Less research has considered how the production of MWSs affect other aspects of language performance, so understanding how MWSs affect language performance beyond fluency is of interest. Specifically, the relationship between MWSs and lexical variety (also referred to as *lexical diversity*) might shed light on the role of MWSs in another aspect of language production: complexity. This paper is an exploration of how MWSs potentially affect lexical variety across development with longitudinal corpus data from English-as-a-Second-Language (ESL) learners in an intensive English program (IEP) in the United States. Specifically, we investigated the development of MWSs and lexical variety in oral language production (speech) from a usage-based, constructionist approach.

2. Background

2.1 Aspects of Language Performance

L2 researchers have identified complexity, accuracy, and fluency (CAF) as the three main broad constructs of language production (Norris & Ortega, 2009). In other words, L2 researchers (and language instructors) are interested in how complex, how accurate, and how fluent the learners' language production is. Given that the CAF constructs reflect very different aspects of language production and that learners have limited attentional resources, Skehan (1998) proposed that an improvement in one area could hinder performance in another, a "trade-off" effect. Certainly, this possibility would be of interest to language researchers and a concern in

language teaching. Accordingly, L2 researchers have studied CAF performances during specific tasks (e.g., Awwad, et al., 2017; Robinson, 1995; Vasylets et al., 2017) as well as across development (e.g., Polat & Kim, 2014; Vercellotti, 2017) to investigate possible interactions between these aspects of language performance. Generally, research has found that lexical variety increases with proficiency (e.g., Bulté & Roothoof, 2020; Polat & Kim, 2014). So far, researchers have not considered the possible benefits of MWSs on lexical variety.

Researchers have investigated whether information in the CAF language performance itself can identify the status of MWSs. For instance, the fluency of the speech (e.g., Dahlmann & Adolphs, 2007; Girard & Sionis, 2003) as well as the complexity of the language produced (e.g., Wood, 2007) has been used to investigate the MWSs' status. With the tools of corpus linguistics (Granger et al., 2015), research with large datasets of learner language is more practical. For instance, lexical variety, within the construct of complexity, can be automatically calculated to provide information about the relationship between types and tokens produced (Jarvis, 2013).

2.2 Multiword expressions in language

With the use of corpus linguistics, researchers have been able to quantify how repetitive human language is, and MWS have been shown to be fundamental and pervasive in language (Nattinger & DeCarrico, 1992; Siyanova-Chanturia & Pellicer-Sanchez, 2018; Weinert, 1995). MWSs may be fixed utterances or linguistic “frames” where various lexical items can be placed into slots (Bohn, 1986; Nattinger & DeCarrico, 1992). For instance, variable MWSs allow the speaker to insert a morpheme (e.g., a tense marker) or replace a word (e.g., the noun) or multiple words of the MWS. On the other hand, “polywords” are fixed, continuous, short utterances that function as words, such as *by the way*. These short fixed MWSs seem to share relevant features of individual lexical items; they perform the same linguistic functions (Boers & Lindstromberg,

2012), are stored in long-term memory (Wood, 2002), and are “retrieved whole from memory” (Wray, 2000, p. 465). Researchers have also explored whether MWSs are stored and processed holistically with experimental data, such as grammaticality judgments (e.g., Jiang & Nekrasova, 2007) and dictation tasks (e.g., Schmitt et al. 2004), though with conflicting results. Regardless of how MWSs are stored, they may still be compositional and analyzable (Boers et al., 2006) by L2 language learners, perhaps more so for adult learners.

To complicate matters further, since most MWSs are short, speakers can use multiple pre-constructed units combined with grammatical knowledge to create a larger sentence (Dabrowska, 2004; Ellis, 1983). Measures of lexical variety could therefore increase with the production of MWSs given the advantages MWSs have over single lexical items: these “polywords” would be counted as multiple words, MWSs’ variability would add lexical variety, and MWS can be combined to make longer stretches of language.

MWSs are also expected to be high-frequency phrases (Jiang & Nekrasova, 2007). In fact, Wong-Fillmore (1976) included two frequency-based criterion: repeated use by the speaker and community-wide use. Considering the importance of frequency and computational methods in linguistic analysis, researchers have used statistical methods for finding MWSs, which has the added benefit of being free from *a priori* assumptions (Hyland, 2012). One simple quantitative method of identifying contiguous MWSs in a given corpora is reporting ngram results, where *n* represents the number of words in the MWS.

Corpus analysis, with quick and consistent methodology, has been suggested as a possible avenue of identifying MWSs for use in language acquisition research (Gries, 2013) but not all agree (e.g., Wray, 2000). Which MWSs emerge is a function of both the size and type of the corpus (Church, 2011), and the statistical methodology (e.g., collecting ngrams) may identify

only “common syntactic construction involving extremely common words” (Manning & Schütze, 1999, p. 31; Warren, 2009). In other words, the MWSs found in corpora may not be psycholinguistically valid, especially for non-native speakers (Schmitt et al., 2004). In their analysis of classroom teaching and textbooks, Biber et al. (2004), however, found that high-frequency 4-gram MWSs served important discourse functions despite not being typically idiomatic nor complete grammatical constituents. As MWS frequency relates to lexical variety then, if the MWSs are repeatedly produced within a single speech, we might expect the speech to have lower lexical variety scores. Overall, however, it is unclear whether the production of MWSs supports lexical variety (in that more words can be spoken) or hinders lexical variety (in that the same words are repeated).

2.3 MWSs and language learning

In usage-based approaches to language learning, including Construction Grammar, learning is thought to take place through exposure to language from which patterns are then induced (Ellis & Wulff, 2014). From this perspective, MWSs are more than just lexical items, i.e., “big words” or “polywords”, and in Construction Grammar specifically, there is not even a division between the lexicon and grammar. Rather, all form-meaning pairings (e.g., morphemes, words, syntactic patterns) are “constructions” (Diessel, 2013; Hoffmann & Trousdale, 2013), and both lexical items and grammatical patterns have certain items that are specific, applicable to a few instances, and some that are general, applicable to many instances (Zernik & Dyer, 1987). And, Ellis (1996) has stated that language learning is the learning of sequences: sequences of sounds to learn words and the sequence of words to learn MWSs.

Learners can use MWSs to extract the individual words and the syntactic patterns (Tomasello, 2003). Therefore, although MWSs may enter a learner’s lexicon as a non-

compositional “chunk”, they do not remain unanalyzed. With increased proficiency, language is more variable and abstract, with specific words being replaced showing “slots” that can be filled with words of the same syntactic category (Diessel, 2013). Although grammatical patterns may enable infinite ways words can be combined into sentences, language often repeats lexical combinations by way of semi-preconstructed phrases (Sinclair, 1987). Some (e.g., Goldberg, 2013; Sinclair, 1987) have suggested that language users compose sentences with grammatical patterns only when forced to do so and rely on formulaic language most of the time. In fact, the developmental U-curves described in first language (L1) development of specific constructions may be explained by such a relationship – first learners can accurately produce an analyzed stretch of language, but may make errors as the abstract categories are being learned, then return to accurate production when the system is figured out (Dabrowska, 2004). Some empirical studies have found support for such a relationship in L2 development (e.g., Bolander, 1989), even when the MWSs used by learners are not linguistic constituents, such as a phrase. Ellis reported that MWSs aided both performance and development of creative speech in child L2 learners of English (1983) and that analysis of the components in functional formulaic language can lead to more abstract structures being used creatively (2013).

Even if child language learners exploit MWSs to extract linguistic patterns, Wray (2012) has cautioned that researchers should not assume that the adult L2 learners do so, or do so in the same way. In fact, research has found that MWSs are difficult for adult L2 learners (e.g., Yorio, 1989), possibly because adult learners use different strategies to learn language (Skehan, 1998). Wood (2002) has suggested that some adult learners might use MWSs to infer rules whereas others may not, as Dabrowska (2004) has said of L1 learners. Nattinger and DeCarrico (1992) surmised that adults could glean patterns from MWSs more readily than children because they

know that the MWS can be segmented and analyzed from their knowledge of their first language. Generally, it is unclear if adult learners do segment MWSs effectively (Wood, 2002). An unresolved question for researchers working from a functional perspective is whether language learning is a single process of learning of sequences or if it is a learning of sequences *and* the ability to extract the abstract patterns (Ellis, 1996). Skehan (1998) suggested that for L2 learning at least, these are different abilities.

L2 research has demonstrated that adult learners can, of course, combine individual familiar words, showing creative use of language (Weinert, 1995). Interestingly, Bolander (1989) observed that adult L2 learners of Swedish in an instructed context did use MWSs, but suggested that the source of the MWSs was more likely informal, spoken language. Focusing on the development of MWS production, Tavakoli and Uchihara (2020) reported a linear increase in the production of spoken MWS across proficiency levels based on group means with cross-sectional data, although the lower proficiency levels had substantial individual variation. In contrast, Siyanova-Chanturia and Spina (2020) found that L2 (Italian) learners had a decreased production of MWSs in writing after six months.

Importantly, learners may produce ungrammatical MWSs in naturalistic contexts (Bahns et al., 1986) and even in instructed contexts (Parkinson, 2015; Weinert, 1995). Yorio (1989) argued that since L2 writing has ungrammatical MWSs, learners do use patterns to get to MWSs in the target language. Alternatively, although learners may use MWSs that do not match the target language, learners may not have used grammatical patterns to create them (Weinert, 1995). This seeming contradiction may be attributed to the “double role” of MWSs as both acquired sequences and means to grammatical patterns (Wood, 2002). Myles et al. (1998) suggested that

overextended (i.e., incorrect) use of MWSs is actually evidence that the MWS is an unanalyzed chunk, a feature of formulaic language.

Research investigating MWSs in language development is not new (e.g., Ellis, 1983), but it is still understudied. Many studies have been conducted with English as a foreign language written language samples, often collected at a single point in time (Callies, 2015), rather than the use of recurrent language in ESL speech (Paquot & Granger, 2012). Moreover, several studies are case studies (e.g., Dahlmann & Adolphs, 2007; Eskildsen & Cadierno, 2007; Yuldashev et al., 2013) or cross-sectional research (Forsberg Lundell & Lindqvist, 2012; Tavakoli & Uchihara, 2020); fewer investigations attempted to investigate the production of MWSs by the same learners over time. In fact, Paquot and Granger (2012) have called for more research on spoken MWSs and more longitudinal studies.

Furthermore, little work has been done to study MWSs and lexical complexity. One study, Forsberg Lundell and Lindqvist (2012), found that MWSs and production of less common words in advanced L2 learners of French did not develop in tandem in their cross-sectional oral data. In their study, the L2 learners who had lived in the target language environment for over fifteen years used more low-frequency vocabulary words in their speech, while the learners who had lived in the L2 environment for less time (5-15 years) produced more MWSs. Specifically, there has been little research which has investigated the development of MWSs and lexical variety in ESL speech. It is unclear whether the production of MWSs results in higher lexical variety, i.e. in a supportive role, as MWSs have with fluency. Alternatively, despite their connection to higher proficiency, MWSs may inadvertently hinder lexical variety through the recycling of the same words. After identifying MWSs (defined as trigrams) in these English language learner data, we ask the following research questions:

1. What is the developmental trajectory of the use of MWSs in the speech of ESL learners?
2. What is the developmental trajectory of lexical variety scores in the speech of ESL learners?
3. What is the relationship between the use of MWSs and lexical variety in ESL learners' speech across development?

3. Methodology

3.1 Participants

The participants ($n = 66$) were adult (18 - 35 years) L2 learners of English, 34 males and 32 females, who entered an IEP in the United States during 2010. They had various L1s: Arabic ($n = 43$), Chinese ($n = 16$), and Korean ($n = 7$). All participants had studied English in their home country, but studying in the IEP was their first time in an English-speaking country. Upon enrollment in this IEP, the students completed multiple placements tests. At the time of the data collection, this IEP had three proficiency levels: low-intermediate (CEFR B1), high-intermediate (CEFR B1+/B2), and advanced (CEFR B2/C1); all participants were placed at either the low- or high-intermediate level upon enrollment. Most students were simultaneously enrolled in speaking, listening, grammar, reading, and writing classes at the same level but could have been assigned to different proficiency levels per skill.

3.2 Data

The participants gave two-minute monologues on a curriculum-determined topic as part of the IEP's speaking course. (See McCormick & Vercellotti, 2013 for a description of the activity.) The two-minute monologues ($n = 294$) were given multiple times per semester, and these data

were collected over three consecutive semesters. The speeches were transcribed by a native English speaker and coded for use in CLAN (MacWhinney, 2000).

Three-word MWSs (trigrams) were chosen as the interest for this study because four-grams might be too infrequent (e.g., Hyland, 2012) while bigrams were expected to include many meaningless combinations resulting from common words in common structures (e.g., *do not, it is*), which would not be fruitful to analyze. Potential MWSs were extracted from the data using COOCUR, a program in CLAN which, unsurprisingly, compiles cooccurrences of words (MacWhinney, 2000), resulting in a list of all of the contiguous trigrams in each speech (e.g., Dahlmann & Adolphs, 2007). Contracted forms were considered two words, and the contracted and full form were aggregated.

Given the difficulty of identifying MWSs in data in a consistent and principled way, we chose a conservative two-pronged approach. First, a list of the most frequent (more than 100 tokens in a corpus of four million words) trigrams was created based on written data from the same IEP, to find community-wide MWSs, following Wong-Fillmore (1976). The use of independent written data to identify MWSs in the spoken data gave a principled criterion for identifying MWSs independently of the spoken data that were the target of analysis. Second, any potential MWS had to be produced at least twice in a single speech because an MWS is expected to be a readily-available chunk in the speaker's linguistic repertoire (Wray, 2000). As a corpus-derived list, these data would be considered "recurrent clusters" (Schmitt et al., 2004), polywords, or continuous non-variable phrasal constraints (Nattinger & DeCarrico, 1992). We might expect such non-variable polywords to be used earlier by learners as compared to variable expressions which may not be recognizable as a salient MWS.

Lexical variety was measured by D, using the metric *vocd* based on word (McKee et al., 2000), which calculates mean type-token ratios of random samples (McCarthy & Jarvis, 2007). This measure can reliably compare texts of different lengths, even relatively short texts (McKee et al., 2000) and has been shown to be a useful measure for L2 data (e.g., Vasylets et al, 2017; Vercellotti, 2017; Yu, 2010).

3.3 Analyses

First, these longitudinal data were analyzed using a mixed effects modeling, specifically Hierarchical Linear and Non-linear Modeling (HLM), to understand the change trajectory of the production of MWSs and of lexical variety as well as the relationship between MWSs and lexical variety. Longitudinal data, with multiple observations from each individual is more robust and can reveal how change occurs (Verspoor & Behrens, 2011), and a method with multilevel (e.g., observations nested within individuals) modeling is “extremely powerful” (Gries, 2013, p. 108). An HLM analysis estimates a growth curve for each participant (Singer & Willett, 2003), which can include nonlinear curves with four or more data points. Crucially, HLM allows participants to have a varying number of observations, and it considers the distance between observations in the model (i.e., observations farther apart can be expected to be more different than closer observations). Mixed-effects modeling, such as HLM, has been increasingly used in L2 research (e.g., Kuriscak, 2015; Siyanova-Chanturia & Spina, 2020; Vercellotti, 2017, 2018). We created an HLM model for the production of MWSs (tokens) and another for the lexical variety scores in the ESL speeches. We used the full maximum likelihood method of estimation and report the final estimation of fixed effects based on robust standard errors.

In this study, instruction level was not held constant because we were interested in the use of IEP-specific MWSs and the development of lexical variety in an instructed context.

Differences in initial proficiency, however, could affect the production of MWSs or the lexical variety in the speeches and should be considered as a possible predictor in the model. The in-house listening placement test was most highly correlated with actual placement into instruction levels, $r = .838$ ($p < .001$), so it was chosen as the measure of initial proficiency ($M = 19.3$; $SD = 4.5$, range 9-27). The initial proficiency scores were centered (subtracted from the person-level mean) to create a meaningful score in the analysis.

Second, the relationship between MWS production and lexical variety was evaluated with two correlational analyses. A correlation analysis can reveal the relationship (or lack thereof) between measures. After separating the data by observation, Pearson correlation analyses (one-tailed) were done at each observation. One-tailed correlations were appropriate because the HLM results showed a specific direction of the relationship of the measures. By comparing the correlations at each observation, we can observe whether the relationship between the measures changes. The aggregated data also mitigate any unplanned topic effects, which has been found to influence lexical variety in writing tasks (Yu, 2010) and in speech (De Jong & Vercellotti, 2016) as well as ngrams use in speech (Tavakoli & Uchihara, 2020).

We also wanted to further explore the relationship between observations within individual learners. Therefore, another Pearson's correlation analysis (two-tailed) was completed on each participant's observations. A two-tailed correlation is required when the relationship between the measures may be positive or negative.

4. Results

4.1 MWSs in the data

We first share the descriptive results before the inferential statistics which answer the research questions. The methodology identified 570 tokens of 85 different three-word, fixed

MWSs (Appendix A). Notably, the most commonly produced (10+ tokens) MWSs in the speeches (Table 1) and a review of the entire list shows that several of the identified MWSs have a similar syntactic structure of *in my* N_{PLACE} (e.g., *in my country, in my city*) which aligns with the common prepositional phrase structure PREP DET N (e.g., *in the world, in my opinion, with my friends, around the world, in the past*). (The syntactic category of *Determiner* includes articles and genitives.) Interestingly, one MWS in the data, *there is many*, is traditionally considered ungrammatical.

4.2 Production of MWSs in ESL speech

The mean for any specific MWS in any given monologue was 2.43 ($SD = .87$). In other words, if a given MWS is in a monologue, it was produced nearly 2.5 times on average. Across all observations, the mean MWS count was 1.95 ($SD 2.56$) per speech. These descriptive statistics are given to describe the data generally, but include between-participant variation and potentially intra-individual variation over time.

To explore change over time, we examined the change trajectory of the production of MWSs using HLM. Since a review of the raw data indicated that the change over time might be non-linear, a quadratic model was considered and acceleration rate was found to be significant in the model ($p = .009$).

Results of the best-fitting growth model (Table 2) specify that the expected number of MWSs at the first observation (one month in the IEP) for an average student was estimated to be 1.63 MWSs. The mean growth rate was estimated to be +4.59, and the mean acceleration was estimated to be -7.24. All parameters were significant, indicating that all parameters are necessary for describing the mean growth trajectory. Students did not vary significantly in their initial scores ($\chi^2_{65} = 55.46, p > .500$), in their growth rate ($\chi^2_{65} = 72.09, p = .255$), or in their

acceleration rate, ($\chi^2_{65} = 74.99, p = .186$), which indicates the model would not be substantially improved by adding any predictors. In other words, there was no systematic variation across development to investigate. In summary, the average student had less than two MWSs at the first observation, which at first increased and then decreased across proficiency (Figure 1).

4.3 Lexical Variety in ESL speeches

Across all observations, the mean for the lexical variety scores (D) per speech was 52.19 ($SD = 15.39$), which includes variation from between-participant performance and potentially intra-individual variation from change over time. We examined the development of lexical variety scores using HLM.

We first compared the unconditioned linear growth model with the unconditioned non-linear growth model because again the review of the raw data indicated that the growth might be non-linear. The non-linear model was found to better fit the data ($\chi^2_4 = 14.64, p = .006$). Based on the results of the unconditioned quadratic model, which showed significant variation in initial scores, initial proficiency was added as a predictor in initial scores.

Results of the conditional quadratic growth model (Table 3) showed that the expected lexical variety scores at one month for an average student was estimated to be 53.40. The coefficient for initial proficiency indicates that for every one point increase in centered initial proficiency, there was a corresponding 1.02 increase in lexical variety score. The mean change rate at one month was estimated to be -25.96, and the mean acceleration of 61.13. All parameters were significantly different from zero, indicating that all three parameters are necessary for describing the mean growth trajectory. Students still varied significantly in their lexical variety scores at one month ($\chi^2_{64} = 88.871, p = .021$). Students did not vary significantly in their growth rate ($\chi^2_{65} = 65.60, p = .456$), or in their acceleration rate, ($\chi^2_{65} = 69.52, p = .326$), which indicates

the model would not be substantially improved by adding any time-constant predictors. In summary, students with higher proficiency, had higher lexical variety scores, but additional unexplained variation remained at the initial observation, and across development, lexical variety scores initially decreased but then increased significantly (Figure 2).

Since it is theoretically plausible that the use of MWSs might influence lexical variety scores, the time-varying MWS score was tested as a predictor in the model. When MWS count was added to the model as a potential time-varying predictor, it was a significant predictor in that each increase in MWS count *lowered* lexical variety. In this model, however, time was no longer a significant predictor. Given that change over time was the focus of the study, the conditioned model without the time-varying predictor of MWS was chosen as the best-fitting model. It is noteworthy, however, that higher MWS production was associated with lower lexical variety scores.

4.4 Relationship between MWS and lexical variety

Two correlational analyses were performed. First, the relationship between the production of MWSs and lexical variety per observation was investigated with Pearson's correlation (one-tailed) analyses. The correlation analyses per observation (Table 4) showed that the MWS counts and lexical variety scores were significantly negatively correlated at the second ($r = -.211, p = .045$) and sixth observation ($r = -.570, p = .043$). The correlation at the second observation is considered a weak negative relationship, whereas the sixth correlation is a strong negative relationship. At the other observations, these measures of lexical complexity were not correlated. In other words, at only two points in across development, one early and one later, was there a negative relationship between the production of MWSs and lexical variety in ESL speech.

Second, the relationship between the production of MWSs and lexical variety within individual learners was investigated with Pearson's correlation (two-tailed) analyses. For three participants, a correlation could not be calculated because their MWS scores were zero at each observation. Most within-participant correlations were not significant, and several coefficients were near zero. The scores of three individuals, however, showed a significant correlation between these measures across observations. One participant had a statistically significant negative correlation, $r = -.892$ ($n = 5$), $p = .042$ while two participants showed a significant positive relationship, $r = .999$ ($n = 3$), $p = .028$ and $r = .999$ ($n = 3$), $p = .032$.

5. Discussion

5.1 MWSs in the data

Several MWSs in our data matched the MWSs found in previous research on spoken language, such as *I do not* (Dahlmann & Adolphs, 2007) and *a lot of* (Dahlann & Adolphs, 2007; Tavakoli & Uchihara, 2020), and appear in the 4-word written MWSs in Biber et al. (2004), such as, *and I think*, *you want to*, and *would like to*. Many of the MWSs could be considered a phrasal unit (e.g., *in my country*), as expected, but at the same time, they were often very frequent words in common phrases, a risk predicted by Manning and Schütze (1999). Some MWSs crossed phrasal boundaries (e.g., *after that I*, *for example if*), as mentioned in other studies (Bolander, 1989; Tavakoli & Uchihara, 2020) and so these MWSs are not likely a coherent chunk, another risk of corpus-based MWS research (Schmitt et al., 2004).

The production of MWSs (measured by trigrams) was rather low, less than three tokens per speech, and the use of MWSs varied considerably across participants. In fact, three participants did not produce two tokens of any identified trigrams, despite the collection of frequent words in generic phrases. As stated, the methodology to identify MWSs was

conservative, excluding any trigram that did not occur at least twice in the same observation and excluding any variable MWSs. If learners produced variations of these MWSs, they would not have been counted.

In addition, the MWS *there is many* could suggest that the learners were not yet able to adhere to subject-verb agreement with postponed subjects with the processing demands of speech. Alternatively, this ungrammatical MWS could be the result of a combination of smaller MWSs, (i.e., *there is* with a noun phrase such as *many reasons*), as suggested by Dabrowska (2004). In fact, the use of *there's* with plural subjects is increasing in many English dialects (Collins, 2012). It is plausible that *there's* is a fixed two-word MWS for both expert speakers and learners.

5.2 Production of MWS in ESL speech

Regarding the change of the production of MWSs, the model showed that after an increase in MWSs, the production of the MWSs steeply decreased across development, making an upside-down U-shape (Figure 1). This increase of identified MWSs soon after enrollment in the IEP could be the result of formulaic language learned by the learners. Given that MWSs are linked to higher fluency (Wray, 2000), the participants may have been producing MWSs to gain fluency in this speaking task. Alternatively, the initial increased production of MWSs may be explained by the participants learning to apply grammatical patterns with common words. Since the current study considered only trigrams common in the IEP in both the speeches and in the community-wide written corpus, this increase in production early in their time the IEP could reflect the learning the common phrases in the context, or simply these MWSs reflect trigrams that are highly functional in language generally.

After the increase, the production of MWSs decreased with more time spent in the IEP, which echoes the findings by Siyanova-Chanturia and Spina (2020) who found a decrease in MWSs after six months in an intensive language program (L2 Italian) with written data. The reason for this decrease in community-wide MWSs with common words is unclear, and we can only speculate based on the current data. The preposition *in* is an early-learned English preposition and *my country* is a relevant and salient noun phrase for newly arrived students. In fact, several of the MWSs which were prepositional phrases were formed with a single preposition (i.e., *in*) and a physical place noun phrase (e.g., *my city*, *the USA*). The nouns seemingly fall into predictable semantic categories, aligning with Hoey's (2005) usage-based notion of lexical priming in which a word sequence, i.e., a MWS, may associate with a specific semantic context. A usage-based framework would then predict that learners use a prototype (e.g., *in my country*) to create other similar MWSs. Therefore, a possible explanation for the subsequent decrease in MWSs is that with increasing proficiency, the learners may have started producing more variable MWSs. This explanation aligns with previous research (Siyanova-Chanturia & Spina, 2020; Tavakoli & Uchirara, 2020) which found that lower proficiency speakers used fixed MWSs while higher proficiency speakers substituted words in the common MWSs. The number of fixed MWSs would decrease when the learner has begun to create novel constructions based on the pattern, filling "slots" in the MWSs. Further, the production of novel constructions would likely increase lexical variety, which was, indeed, found in these data.

The findings reveal a nonlinear trajectory of the use of MWSs in speech in this context, rather than the linear increase found by Tavakoli and Uchihara (2020) with cross-sectional data. Moreover, despite the expectation of U-shaped developmental patterns in learning (Carlucci & Case, 2013), the production of these MWSs actually had the opposite developmental pattern.

Finally, the analysis found no difference in the production of MWSs from the learners' differences in initial proficiency. Although it was plausible that the use of MWSs might vary by initial proficiency, the production of MWSs in this study would be tied closely to time in the IEP because the identified MWSs were fixed community-wide trigrams.

5.3 Lexical Variety in ESL speeches

In the best-fitting HLM model for lexical variety, initial proficiency was a predictor, which means that participants who entered the IEP with higher proficiency used a larger variety of words at the first observation. This finding is not unexpected; vocabulary measures are often connected to general language proficiency (e.g., Author & Coauthor, *forthcoming*; Gass & Selinker, 2008; Vasylets et al., 2017). The model for lexical variety showed a slight decrease in lexical variety scores after the first observation followed by a steeper increase over time in the IEP (Figure 2). The slight decrease is unexpected in an instructed learning context because the students learn new vocabulary each semester. But, given that productive use of vocabulary lags (Author & Coauthor, *forthcoming*), the decrease in lexical variety might be understood as a delay between learning new vocabulary and knowing how to use it successfully in oral production. The decrease in lexical variety is contrary to findings from cross-sectional research (Bulté & Roothoof, 2020) and longitudinal case-study research (Polat & Kim, 2014), but those studies also reported substantial variation. It is plausible that as the learners realize how MWSs can be used to convey different ideas, they substitute different lexical items. Moreover, the ability and the confidence to use variable MWSs would be attained with substantial individual variation, not necessarily captured at the group means, used in cross-sectional research.

5.4 Relationship between MWS and lexical variety

MWS was tested as a possible time-varying predictor in the model for lexical variety. Since the results of that model showed that an increase in MWSs decreased lexical variety scores, we can presume that lowered lexical variety scores were driven (at least in part) by the increased production of the MWSs. This finding of a trade-off effect between MWSs and lexical variety seems to be the first empirical evidence that the production of fixed MWSs lowers the lexical variety score (measured by D). Considering the differing patterns of development shown by the HLM models for these measures, it seems that as instructed learners increase in proficiency, the use of fixed MWSs initially increases which lowers lexical variety scores, but then when the learners produce fewer fixed MWS, the lexical variety scores increase. The finding of a relationship between the MWSs production and lexical variety seems contrary to Siyanova-Chanturia and Spina (2020) who found no relationship, but that study was limited to only two observations which were six months apart. The production of variable MWSs would both decrease fixed MWSs and increase the lexical variety in the language sample. One implication for the language classroom is that the production of MWSs will likely only temporarily decrease the variety of vocabulary in ESL speech.

The correlation analysis by observation confirmed a significant negative relationship at only two observations, one soon after enrollment and one during the third semester of instruction. Considering the non-linear growth curves of the two measures, we see that the correlations reach significance when MWS production was rising at the same time lexical variety scores were falling and then later when the production of the fixed MWSs were falling and the lexical variety was increasing in the speeches. At the other observations, the measures were not correlated, which is consistent with Forsberg Lundell and Lindquist's (2012) study of lexical richness and collocations in L2 French.

The findings of the correlations within-individuals across development were complicated, with most individual correlations near zero and insignificant. This lack of a relationship and coefficients near zero reflect a changing or inconsistent relationship. The production of MWSs and lexical variety scores for one participant showed a significant negative relationship across development, echoing the significant relationship at two observations at the group level and the finding of MWS as a potential significant predictor in the HLM model for lexical variety. Since the production of MWSs was associated with a decrease in lexical variety across development, the findings suggest a reliance on fixed utterances means more recycling of vocabulary.

Surprisingly, the two measures were positively correlated for two individuals. The finding of positive correlations across development means that the production of fixed MWSs can boost lexical variety for at least some learners, perhaps for a specific period of development. These findings are tentative, but nonetheless, the individual differences are striking for two at least two reasons. First, for these three learners, there was a more consistent relationship between MWS production and lexical variety across development unlike the majority of learners. Second, the use of MWSs was associated with lower *and* higher lexical variety scores within the individual's development in an IEP. Since different conclusions might be drawn when considering only observations at a single snapshot of development, this longitudinal study hints at possible reasons for the divergent findings in the literature.

6. Conclusion

Research on the effect of MWSs in oral language production has focused more on how MWSs can improve the fluency of the speech. This study expanded the discussion to how MWSs may influence lexical complexity in adult ESL speech across development. This longitudinal study found divergent developmental trajectories of the oral production of MWSs and lexical

variety. The production of MWSs initially increased with time in the IEP and then steadily decreased while lexical variety slightly decreased and then increased sharply with development. Further, the findings suggested a negative influence of fixed MWSs on lexical variety scores in the speech of most adult learners in an IEP. It seems that this potential trade-off effect is short lived since the lexical variety in these oral data (speeches) improved over time in the IEP. The current study did not investigate how MWSs might influence the accuracy of the language performance; longitudinal research between MWSs and accuracy would give further insight on how MWSs influence aspects of language performance and language learning.

With a large data set, we relied on statistical methodology which quickly processes the data without confirmation that the MWS was meaningful or produced as a semantic or phonological unit. The statistical methodology (e.g., collecting ngrams) did seem to identify common words in general patterns as Manning and Schütze (2001, p. 31) warned. Overall, although researchers expect formulaic sequences to be frequent as well as meaningful, this additional criterion is hard to establish. It is quite likely that some of the identified MWSs were not, in fact, produced with a fluid delivery in a single intonational unit, as would be expected if it were stored and retrieved as a whole. Since this study was interested in the relationship with lexical variety, we considered this statistical methodological decision justified. Likewise, the methodology employed in this study (requiring frequent use by the speaker and community-wide) did not capture creative use of grammatical patterns. Future research, therefore, may want to include a phraseology criteria to identify only MWS which may be a meaningful unit. Additionally, future research could closely follow the production of fixed and related variable MWSs of a subset of students across time in an IEP.

This research is limited to instructed learners in a single IEP; the findings are not necessarily generalizable to other contexts or languages. In fact, Parkinson (2015) has suggested the production of MWSs differ by learning context (i.e., ESL vs English as a Foreign Language), so the production of MWSs may be lower in when not in the target language environment. Given the ubiquity of MWSs in language, research of MWS production and lexical variety in the speech of naturalistic English language learners would be an interesting comparison study. With the dearth of longitudinal studies on the production of MWSs, this paper expands upon existing case-study and cross-sectional research.

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at

<http://talkbank.org/browser/index.php?url=SLABank/English/Vercellotti/>.

References

Author & Coauthor. (*forthcoming*).

Awwad, A., Tavakoli, P., & Wright, C. (2017). "I think that's what he's doing": Effects of intentional reasoning on second language (L2) speech performance. *System*, 67, 148–69.
<http://dx.doi.org/0.1016/j.system.2017.002>

Bahns J., Burmeister, H., & Vogel, T. (1986). The pragmatics of formulas in L2 learner speech: Use and development. *Journal of Pragmatics*, 10(6), 693–723.
[https://doi.org/10.1016/0378-2166\(86\)90147-5](https://doi.org/10.1016/0378-2166(86)90147-5)

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>

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. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, 32, 83–110.
<https://doi.org/10.1017/S0267190512000050>

Bohn, O-S. (1986). Formulas, frame structures, and stereotypes in early syntactic development: Some new evidence from L2 acquisition. *Linguistics*, 24(1), 185–202.
<https://doi.org/10.1515/ling.1986.24.1.185>

Bolander, M. (1989). Prefabs, patterns and rules in interaction? Formulaic speech in adult learner' L2 Swedish. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism across the*

- lifespan: Aspects of acquisition, maturity, and loss* (pp.73–86). Cambridge University Press.
- Brown, R. (1973). *A first language: The early stages*. Harvard University Press.
- Bulté, B. & Roothoof, H. (2020). Investigating the interrelationship between rated L2 proficiency and linguistic complexity in L2 speech. *System*, 91, <https://doi.org/10.1016/j.system.2020.102246>
- Callies, M. (2015). Learner corpus methodology. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 35-56). Cambridge University Press.
- Carlucci, L., & Case, J. (2013). On the necessity of U-shaped learning. *Topics in cognitive science*, 5(1), 56–88. <https://doi.org/10.1111/tops.12002>
- Church, K. (2011). *How many multiword expressions do people know?* Paper presented at the Proceedings of the Workshop on multiword expressions: from parsing and generation to the real world, Portland, Oregon.
- Collins, P. (2012). Singular agreement in there-existentials: An intervaretal corpus-based study. *English World-Wide*, 33(1), 53–68. <https://doi.org/10.1075/eww.33.1.03col>
- Dabrowska, E. (2004). *Language, Mind and Brain*. Georgetown University Press.
- Dahlmann, I. & Adolphs, S. (2007). Pauses as an indicator of psycholinguistically valid multiword Expressions (MWEs)? *Proceedings of the workshop on a broader perspective on multiword expressions, Association for Compositional Linguistics*, 48–56.
- De Jong, N., & Vercellotti, M. L. (2016). Similar prompts may not be similar in the performance they elicit: Examining fluency, complexity, accuracy, and lexis in narratives from five picture prompts. *Language Teaching Research*, 20(3), 387–404. <https://doi.org/10.1177/1362168815606161>

- Diessel, H. (2013). Construction grammar and first language acquisition. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 247–364). Oxford University Press.
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18, 91–126.
<https://doi.org/10.1017/S0272263100014698>
- Ellis, N. C. (2013). Construction grammar and second language acquisition. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 365–378). Oxford University Press.
- Ellis, N. C., Simpson-Vlach R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375–396. <https://doi.org/10.1002/j.1545-7249.2008.tb00137.x>
- Ellis, N. C., & Wulff, S. (2014). Usage-based approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction (2nd ed.)* (pp. 76-93). Routledge.
- Ellis, R. (1983). Formulaic speech in early classroom second language development. In *On TESOL '83: The question of Control. Selected papers from the seventeenth annual convention of Teachers of English to Speakers of Other Languages*, Toronto, Canada. 15-20 March 1983, pp. 53-65.
- Eskildsen, S. W. & Cadierno, T. (2007). Are recurring multi-word expressions really syntactic freezes? Second language acquisition from the perspective of usage-based linguistics. *Collocations and Idioms, 1, Papers from the first Nordic conference on syntactic freezes*. Joensuu University Press.

- Forsberg Lundell, F. F., & Lindqvist, C. (2012). Vocabulary aspects of advanced L2 French: Do lexical formulaic sequences and lexical richness develop at the same rate? *Language, Interaction and Acquisition*, 3(1), 73–92. <https://doi.org/10.1075/lta.3.1.05>
- Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course (3rd ed.)*. Routledge.
- Girard, M., & Sionis, C. (2003). Formulaic speech in the L2 classroom: An attempt at identification and classification. *Pragmatics*, 13(2), 231–251. <https://doi.org/10.1075/prag.13.2.02gir>
- Goldberg, A. E. (2013). Constructionist approaches. In T. Hoffmann & G. Trousdale (Eds.) *The Oxford handbook of Construction Grammar* (pp. 15-31). Oxford University Press.
- Granger, S., Gilquin, G., & Meunier, F. (2015). Introduction: Learner corpus research – past, present and future. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (Cambridge Handbooks in Language and Linguistics, pp. 1-6). Cambridge University Press. <https://doi.org/10.1017/CBO9781139649414.001>
- Gries S. Th. (2013). Data in Construction Grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 93-108). Oxford University Press.
- Hoffmann, T., & Trousdale, G. (2013). Construction Grammar: Introduction. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 1-12). Oxford University Press.
- Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32, 150–169. <https://doi.org/10.1017/S0267190512000037>

- Jarvis, S. (2013). Capturing the Diversity in Lexical Diversity. *Language Learning*, 63, 87–106.
<https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91(3), 433–445. <https://doi.org/10.1111/j.1540-4781.2007.00589.x>
- Kuiper, K. (2004). Formulaic performance in conventionalised varieties of speech. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (Vol. 9, pp. 37–54). Benjamins.
- Kuriscak, L. (2015). Examination of learner and situation level variables: Choice of speech act and request strategy by Spanish L2 learners. *Hispania* 98(2), 300-318.
- McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, 24(4), 459–488. <https://doi.org/10.1177/0265532207080767>
- McCormick, D. E. & Vercellotti, M. L. (2013). Examining the impact of self-correction notes on grammatical accuracy in speaking. *TESOL Quarterly*, 92, 410–420.
<https://doi.org/10.1002/tesq.92>
- McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, 15(3), 323–337.
<https://doi.org/10.1093/lc/15.3.323>
- MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk. 3rd Edition.
 Lawrence Erlbaum Associates.
- Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing* (2nd Printing ed.). MIT Press.
- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299–320.

- 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), 323–363. <https://doi.org/10.1111/0023-8333.00045>
- Nattinger, J. R. & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press.
- Parkinson, J. (2015). Noun-noun collocation in learner writing. *Journal of English for Academic Purposes*, 20, 103–113. <https://doi.org/10.1016/j.jeap.2015.08.003>
- Pawley, A., & Synder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 191-226). Longman.
- Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, 32, 130–149. <https://doi.org/10.1017/S0267190512000098>
- Polat, B., & Kim, Y. (2014). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics*, 34(2), 184–207. <https://doi.org/10.1093/applin/amt013>
- Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning*, 45(1), 99–140. <https://doi.org/10.1111/j.1467-1770.1995.tb00964.x>
- Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? Formulaic Sequences: Acquisition, processing and use. In N. Schmitt (Ed.), *Formulaic Sequences* (pp. 127-151). John Benjamins.
- Sinclair, J. (1987). Collocation: A progress report. In R. Steele & T. Threadgold (Eds.), *Language Topics: Essays in Honour of Michael Halliday* (pp. 319-331). John Benjamins.

- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: modeling change and event occurrence*. Oxford University Press.
- Siyanova-Chanturia, A. & Pelicer-Sanchez, A. (Eds.). (2018). *Understanding formulaic language: A second language acquisition perspective*. Routledge.
- Siyanova-Chanturia, A. & Spina, S. (2020). Multi-word expressions in second language writing: A large-scale longitudinal learner corpus study. *Language Learning*, 70(2), 420–463.
- Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford University Press.
- Tavakoli, P., & Uchihara, T. (2020). To what extent are multiword sequences associated with oral fluency? *Language Learning* 70(2), 506–547. <https://doi.org/10.1111/lang.12384>
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Vasylets, O., Gilabert, R., & Manchón, R. M. (2017). The effects of mode and task complexity on second language production. *Language Learning* 67(2), 394–430.
- Vercellotti, M. L. (2017). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistics*, 38(1), 90–111. <https://doi.org/10.1093/applin/amv002>
- Vercellotti, M. L. (2018). Finding variation: Assessing the development of syntactic complexity in ESL speech. *International Journal of Applied Linguistics*, 1–15. <https://doi.org/10.1111/ijal.12225>
- Verspoor, M., & Behrens, H. (2011). Dynamic systems theory and a usage-based approach to second language development. In M. Verspoor, de K. Bot, & W. Lowie (Eds.), *A dynamic approach to second language development* (pp. 25-38). John Benjamins Publishing Company.
- Warren, M. (2009). Why concgram? In C. Greaves (Ed.), *ConcGram 1.0: A phraseological search engine* (pp. 1-11). John Benjamins.

- Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics*, 16(2), 180–205. <https://doi.org/10.1093/applin/16.2.180>
- Wong-Fillmore, L. (1976). *The second time around: Cognitive and social strategies in second language acquisition*. PhD dissertation, Stanford University.
- Wood, D. (2002). Formulaic language in acquisition and production: Implications for teaching. *TESL Canada Journal*, 20(1), 1–15. <https://doi.org/10.18806/tesl.v20i1.935>
- Wood, D. (2007). Mastering the English formula: Fluency development of Japanese learners in a study abroad context. *JALT Journal*, 29(2), 209–230.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, 21(4), 463–489. <https://doi.org/10.1093/applin/21.4.463>
- Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, 32, 231–254. <https://doi.org/10.1017/S026719051200013X>
- Yorio, C. A. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism across the lifespan: Aspects of acquisition, maturity, and loss* (pp.55–72). Cambridge University Press.
- Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, 31(2), 236–259. <https://doi.org/10.1093/applin/amp024>
- Yuldashev, A., Thorne, S. L., & Fernandez, J. (2013). Second language learners' contiguous and discontiguous multi-word unit use over time. *Modern Language Journal*, 97, 31–45. <https://doi.org/10.1111/j.1540-4781.2012.01420.x0026-7902/13/31-45>
- Zernik, U., & Dyer, M. G. (1987). The self-extending phrasal lexicon. *Computational Linguistics*, 13(3-4), 309–327.

Appendix A: Multiword Sequences (trigrams) in Data by Total Tokens

Rank	N	n-1	n-2	n-3	Rank	N	n-1	n-2	n-3
1	62	I	do	not	44	4	they	do	not
2	60	in	my	country	45	4	with	my	friend
3	29	a	lot	of	46	4	you	do	not
4	23	when	I	was	47	3	has	a	meeting
5	19	do	not	like	48	3	is	a	big
6	15	in	the	world	49	3	we	like	to
7	13	there	are	many	50	3	you	have	a
8	13	there	is	a	51	3	you	have	to
9	12	we	have	to	52	2	a	long	time
10	11	and	I	think	53	2	after	that	I
11	10	in	my	city	54	2	and	I	was
12	10	it	is	very	55	2	and	it	is
13	10	when	he	was	56	2	because	it	is
14	9	I	think	the	57	2	for	example	if
15	9	I	want	to	58	2	I	have	a
16	9	one	of	the	59	2	I	wanted	to
17	8	do	not	have	60	2	I	was	in
18	8	go	to	the	61	2	in	my	opinion
19	8	has	a	lot	62	2	in	order	to
20	8	I	did	not	63	2	in	the	future
21	8	it	is	not	64	2	in	the	U_S_A
22	8	they	did	not	65	2	is	a	famous
23	8	would	like	to	66	2	is	a	traditional
24	7	have	a	good	67	2	is	a	very
25	7	I	think	that	68	2	is	the	best
26	7	there	are	a	69	2	is	very	important
27	6	I	like	to	70	2	it	is	important
28	6	I	think	it	71	2	it	is	not
29	6	I	would	like	72	2	it	is	the
30	6	it	is	a	73	2	most	important	thing
31	6	to	be	a	74	2	the	most	important
32	6	with	my	friends	75	2	the	other	hand
33	5	after	that	he	76	2	there	is	no
34	5	are	a	lot	77	2	there	is	some
35	5	around	the	world	78	2	they	have	to
36	5	in	front	of	79	2	to	go	to
37	4	*there	is	many	80	2	to	have	a
38	4	have	a	lot	81	2	to	make	a
39	4	I	am	going	82	2	to	study	English
40	4	if	you	have	83	2	we	do	not
41	4	in	our	country	84	2	with	my	family
42	4	in	the	past	85	2	you	want	to
43	4	in	the	same					

Table 1

List of the Most Frequent MWSs.

ranking	tokens	n-1	n-2	n-3
1	62	I	do	not
2	60	in	my	country
3	29	a	lot	of
4	23	when	I	was
5	19	do	not	like
6	15	in	the	world
7	13	there	are	many
8	13	there	is	a
9	12	we	have	to
10	11	and	I	think
11t	10	in	my	city
11t	10	it	is	very
11t	10	when	he	was

Table 2

Unconditioned Quadratic Model of Growth of MWSs.

Fixed Effects	Coefficient	SE	<i>t</i>	<i>p</i>
Mean Initial status, β_{00}	1.63	.21	7.64	<.001
Mean Growth Rate, β_{10}	4.59	1.83	2.51	.015
Mean Acceleration Rate, β_{20}	-7.24	2.67	-2.71	.009
Random Effects	Variance Component	<i>df</i>	X^2	<i>p</i>
Initial status, r_0	0.01	65	55.46	>.500
Change Rate, r_1	5.02	65	72.09	.255
Acceleration Rate, r_2	7.48	65	74.99	.186
Level-1 error, <i>e</i>	6.31			

Table 3

Model of Growth of Lexical Variety Scores (D).

Fixed Effects	Coefficient	SE	<i>t</i>	<i>p</i>
<i>Model for Initial Status, π_{0i}</i>				
Mean Initial status, β_{00}	53.40	1.38	38.61	<.001
Initial Proficiency (centered), β_{10}	1.02	.24	4.31	<.001
<i>Model for Growth Rate, π_{1i}</i>				
Mean Growth Rate, β_{10}	-25.96	10.82	-2.40	.019
Mean Acceleration Rate, β_{20}	61.13	19.69	3.12	.003
Random Effects	Variance Component	<i>df</i>	<i>X</i> ²	<i>p</i>
Initial status, r_0	47.72	64	88.87	.021
Change Rate, r_1	565.90	65	65.60	.456
Acceleration Rate, r_2	2135.10	65	69.52	.326
Level-1 error, e	167.62			

Table 4

Correlation (One-Tailed) of MWS and Lexical Variety scores by Observation.

Observation	Correlation Coefficient	p-value
1 (n = 66)	-.056	.328
2 (n = 66)	-.211	.045
3 (n = 66)	-.084	.251
4 (n = 51)	.036	.402
5 (n = 26)	-.074	.359
6 (n = 10)	-.570	.043
7 (n = 9)	-.525	.073

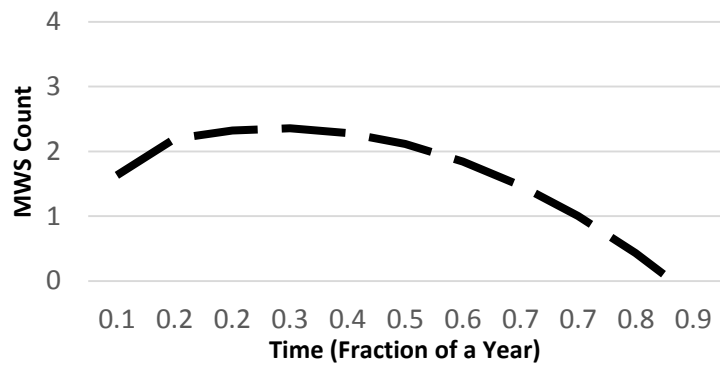


Figure 1
Production of MWSs by Time in an IEP.

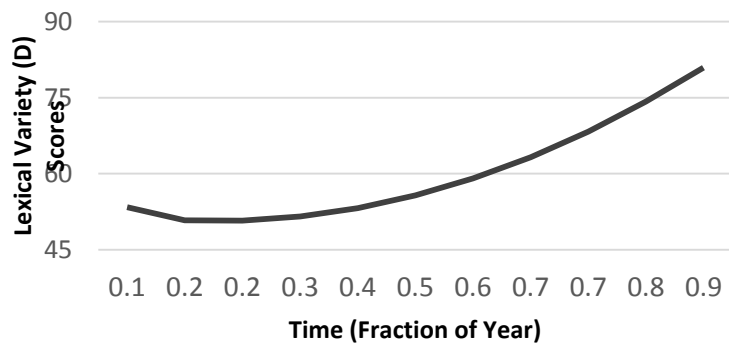


Figure 2
Lexical Variety (D) by Time in an IEP.