The Effects of Task Condition on L2 Oral Performance: Exploring Complexity Measures

Ayşen Tuzcu and Şebnem Yalçın

Abstract

Syntactic complexity has been used as a dependent variable in studies investigating the effects of task features. Most studies measured only subordination and overall complexity and eschewed clausal and phrasal level complexity (Bulté & Housen, 2012). The present study investigated how the selection of syntactic complexity measures as well as specific task designs affect L2 oral performance. The study analyzed the oral narratives of 64 university-level participants in two groups which differed in terms of online planning opportunities in task demands. Syntactic complexity was assessed with four measures focusing on the supraclausal, clausal, and phrasal levels. The results did not show any statistically significant differences between the two groups. The findings were discussed in the light of the syntactic complexity measures, the nature of the task, and L2 proficiency.

Keywords: Syntactic complexity, online planning, L2 oral performance

Introduction

Complexity has received a considerable amount of attention as a construct to assess second language (L2) proficiency, L2 performance, and L2 development, along with the related constructs of accuracy and fluency (see Bulté & Housen, 2012; Norris & Ortega, 2009 for a review). One group of researchers investigated complexity as an outcome and focused on determining what linguistic complexity entails (e.g. Bulté & Housen, 2012; Norris & Ortega, 2009; Pallotti, 2009, 2015) and how it captures L2 proficiency (e.g. Bulté & Housen, 2018; De Clercq & Housen, 2017; Lambert & Kormos, 2014). Another group of researchers have examined how complexity is affected by various independent variables such as task features, modality, genre, L2 instruction, and learner variables (e.g. Kormos & Trebits, 2011, 2012; Kuiken & Vedder, 2011; Révész, Sachs, & Mackey, 2011; Yoon & Polio, 2017).

Ayşen Tuzcu, Second Language Studies, Michigan State University, tuzcuays@msu.edu, ORCID: 0000-0002-3783-3183

Şebnem Yalçın, Foreign Language Education Department, Boğaziçi University, sebnem.yalcin@boun.edu.tr, ORCID: 0000-0003-2143-8055

The latter group of researchers were particularly interested in examining if specific task designs lead to higher syntactic complexity (SC) levels. These researchers, however, focused mostly on the changes in the overall complexity and subordination levels (Bulté & Housen, 2012) and made generalizations about L2 performance based on these measures. However, there is now an agreement in SLA research that SC is multidimensional (Bulté & Housen, 2012; Norris & Ortega, 2009; Pallotti, 2009, 2015) and that SC should be measured accordingly. In other words, researchers are encouraged to use measures that not only assess overall complexity and subordination but also clausal and phrasal complexity (Norris & Ortega, 2009). Failing to use measures reflecting different subcomponents of SC may lead to inefficient understanding and interpretation of the language data.

This study was intended to help address some of these limitations in the previous studies which investigated the effects of task manipulations on L2 performance and to answer Norris and Ortega's (2009) call by using four SC measures that represent distinct subcomponents of SC. The oral narrative task used in the current study was manipulated along online planning. We investigated how manipulation of online planning in an oral narrative task affected the performance of university-level L2 learners of English and if distinct SC measures indicated different results for the same language data.

L2 Complexity

Complexity is a construct that has been widely investigated by researchers interested in different lines of SLA research. However, investigating this construct poses some challenges for researchers since the same term is used to refer to various cognitive and linguistic aspects of language and performance (Bulté & Housen, 2012; Housen & Simoens, 2016; Pallotti, 2009, 2015). Complexity is considered as an independent variable when it refers to the features of communicative tasks or linguistic structures. It can also be employed as a dependent variable when it is used to describe language performance.

Several researchers (Bulté & Housen, 2012; Housen & Simoens; 2016; Miestamo, 2009) distinguished between two types of language complexity: relative complexity (RC) and absolute complexity (AC). RC focuses on the language user and it is defined based on the cognitive cost in learning and using a linguistic phenomenon by a language learner in a language context. RC is subjective in its nature since the use of mental resources and cognitive mechanisms in processing and using a linguistic feature differs depending on the learner. Therefore, RC is associated with difficulty and cognitive complexity (Bulté & Housen, 2012; Housen & Simoens, 2016; Miestamo, 2009). In contrast, AC refers to the inherent properties of a language and it is more quantitative in nature compared to RC. AC (also called structural complexity or syntactic complexity) has to do with the structuring of linguistics systems such as the number and variety of elements in a linguistic feature and the number and variety of connections between these elements. AC may contribute to RC but there is no one to one correspondence between them. In other words, the AC of a linguistic feature may not coincide with the mental difficulty of processing and learning of that feature (Bulté & Housen, 2012; Housen & Simoens, 2016).

The research investigating L2 performance has primarily focused on AC because this approach to complexity is more objective, it does not differ depending on the language learner (Pallotti, 2015), and it excludes any theoretical assumptions regarding the difficulty of linguistic structures (Miestamo, 2009). However, because of the difficulty in assessing the overall complexity o5f the language system, Pallotti (2015) further argued for a simple view of complexity in L2 performance research. This view focuses on the SC of the texts produced by learners instead of the language system. Thus, he defined SC as "the number of different elements and their interconnections (i.e. their systematic, organized relationships), which both produce a longer description of the text's structure" (p. 120).

SC is argued to be a multidimensional construct consisting of various subcomponents (Norris & Ortega, 2009). To fully capture this multidimensionality, Norris and Ortega called for a more organic approach and advised measuring SC at three different levels, sentence level, clausal level, and phrasal level. The researchers underscored the necessity of using "distinct and complementary" measures that "can be used and interpreted together" (p. 562). Moreover, they also argued that different subcomponents of structural complexity can predict different levels of L2 proficiency. Coordination is argued to be a better predictor for beginning level L2 proficiency whereas at intermediate level proficiency, complexity is asserted to be represented through subordination. Phrasal complexity, on the other hand, is predicted to be a better indicator of advanced level proficiency since advanced level learners use modifications and nominalizations more in their productions.

Studies investigating complexity as a dependent variable to assess L2 performance have used this construct as one of the three constructs in CAF triad, namely complexity, accuracy, and fluency. This line of research has mostly investigated the effects of various variables such as task complexity (e.g. Robinson, 2007), task modality (e.g. Kuiken & Vedder, 2011), task type (e.g. Michel, Shi, & Li, 2019), genre (e.g. Yoon & Polio, 2017), L2 proficiency (e.g. Bulté & Housen, 2018), and target language (De Clercq & Housen, 2017). However, these studies, particularly the ones with a task complexity focus like the current study, have reported inconsistent findings. There can be several reasons for these conflicting results.

Firstly, despite Norris and Ortega's (2009) suggestion for a more organic approach to SC of L2 performance, Bulté and Housen (2012) argued that most studies adopted a reductionist approach and employed only one or two measures of complexity, particularly the ones targeting syntactic length and subordination. Mean length of T-unit and mean length of AS-unit, measuring syntactic length, have been the main measures of complexity in many writing and speaking studies. In addition to length-based measures, subordination measures such as number of clauses per unit (i.e. c-unit, T-unit, and AS-unit) have also been widely employed in empirical research. However, measuring SC only with length-based measures or subordination measures and avoiding other SC measures such as the phrasal and clausal level measures is undesirable because focusing only on one or two aspects of SC can result in incomplete representation of SC. Bulté and Housen (2012) further asserted that this reductionist approach to complexity is particularly problematic since most of these studies attempt to make generalizations about learners' L2 complexity levels and the effects of independent variables such as the

manipulations of tasks, so not using measures tapping into all subcomponents of complexity may lead to incomplete implications.

Secondly, even when studies use more than one or two measures of complexity, some of these measures target the same subcomponent of the complexity and cause redundancy (see Table 2 in Bulté and Housen, 2012 for a list of studies). Norris and Ortega (2009) problematized measuring the same subcomponent with more than one measure stating that these measures are not independent and the correlations between these measures may cause problems in the multivariate analysis used by most researchers investigating L2 performance.

Complexity in Oral Performance

L2 performance has been argued to be affected by various factors such as task features (Kuiken & Vedder, 2011; Robinson, 2007; Tuzcu, 2018), task condition (Michel, Kuiken, & Vedder, 200; Tuzcu, 2018; Tuzcu & Yalçın, 2019), modality (Biber, Gray, & Poonpon, 2011), L2 instruction (Révész, Sachs, & Mackey, 2011), L2 proficiency (Lambert & Kormos, 2014), and individual differences (Kormos & Trebits, 2011, 2012). This line of research has examined the changes in learners' complexity, accuracy, and fluency levels and reported various findings depending on the nature of the task as well as the learner profile. Modality, particularly, has been found to be a factor influencing the SC levels. For example, Biber et al. (2011) demonstrated that SC in written modality is generally characterized by phrasal embedding such as modifications and nominalizations whereas in spoken modality SC relies more on subordination clauses. These differences in the integration of novel complexity subcomponents in two modalities emerge from the differences in the cognitive demands of each modality (Trebits, 2016). Oral production requires simultaneous focus on the content, formulation, and monitoring of speech, so this cognitive load may prevent learners from producing complex language. Written production, on the other hand, allows for more online planning, monitoring, and reviewing opportunities since it puts less time pressure on learners. Although previous studies investigating the changes in SC levels have mostly focused on the written production, the results of these studies are not generalizable to oral production. Moreover, the higher cognitive demands of oral production may provide researchers better understanding of the changes in SC (Vercelotti, 2019).

To understand the changes in SC in spoken production, it is essential to clarify the processes involved in speech production. One of the most influential speech production models has been proposed by Levelt (1989) for L1 speech production and later on this model has been extended to L2 speech production (Kormos, 2006, 2011). Levelt (1989) argued that speech production is modular consisting of three basic mechanisms, conceptualizer, formulator, and articulator. Conceptualization involves conceptual preparation for speech. During the conceptualization stage, communicative goals, topic, focus, and content of a speech are determined through macro planning and micro planning stages. Since determining communicative intentions require attention, the conceptualization stage is highly controlled. The formulation stage is responsible for converting the pre-verbal messages created by the conceptualizer into linguistic messages by retrieving lexical items from the mental lexicon. During the articulation

stage, linguistic messages are articulated as overt speech. Unlike conceptualization, formulation and articulation are automatic. Processing of all mechanisms in the model is incremental, and all mechanisms work in parallel. For example, when a syllable of a word is encoded in the formulator, it can be articulated without waiting for the whole sentence to be completed. Time pressure may affect the conceptualization stage in L1 oral production since this stage requires controlled attention, however the formulation and articulation stages are not influenced by time pressure because of the autonomous nature of these stages (Wang, 2014).

Kormos (2006, 2011) posited that L1 and L2 speech production is similar. L2 speech production consists of the same three mechanisms, namely conceptualizer, formulator, and articulator, and these mechanisms work incrementally and in parallel as in L1 production. However, differently from L1 speech production, parallel processing can only be achieved after a certain proficiency level. Differently from the L1 production, in the bilingual model, when L2 proficiency is low, formulating and even articulating stages as well as conceptualizing and monitoring need conscious attentional control since most of the syntactic and phonological rules of L2 are stored in the form of declarative knowledge (Ullman, 2001). When proficiency increases, the regularities in grammatical and phonological rules become proceduralized, and only then the stages of formulation and articulation become autonomous (Kormos, 2011). Because of this conscious attentional control in formulation and sometimes articulation stages in addition to conceptualization and monitoring, time pressure affects L2 speech production even more than L1 speech production (Wang, 2014).

In the light of these speech production models, some researchers have investigated if decreasing the time pressure leads to increased L2 oral performance. Ellis (2005) distinguished two planning types, namely pre-task planning (planning before the task) and online planning (planning during the task), which can be helpful in decreasing the pressure in L2 oral performance. Online planning, in particular, has been argued to contribute to L2 performance (Ellis, 2005). Yuan and Ellis (2003) defined online planning as "the process by which speakers attend carefully to the formulation stage during speech planning and engage in pre-production and post-production monitoring of their speech" (p.6). Ellis and Yuan (2005) differentiated unpressured online planning from pressured online planning. In the unpressured online planning learners are provided with sufficient time that allows them to pay attention to conceptualizing, formulating, articulating, and monitoring of their speech. The pressured online planning, however, provides learners a limited time to produce their speech, so it puts learners under pressure. Therefore, learners may not allocate enough equal attention to the conceptualization, formulation, and monitoring stages.

Following Yuan and Ellis (2003), several researchers investigated the effects of availability of online planning on L2 performance. Table 1 summarizes the characteristics and results of these studies. The table specifies target L2, participants, their proficiency levels in L2, type of tasks, mode of tasks, SC measures used in the task, and results of each study for the SC construct. All studies employed narrative tasks based on either a video or a set of pictures. Most of these studies reported increased SC levels in unpressured online planning. Despite Norris and Ortega's (2009) call for a more organic approach to SC, most of these studies took a reductionist approach (Bulté & Housen, 2012; Norris & Ortega, 2009) to L2 syntactic complexity. They all measured

SC with subordination measures. Four studies also included a syntactic variety measure which indicated the number of distinct verb forms used in the task. However, it is not certain how this measure represents SC. De Clercq and Housen (2017) stated that this syntactic variety measure may not be compatible with the absolute view of SC since the verb forms used in the task can be related to developmental stages. Wang (2014) examined the total number of words in the task as a measure of SC. However, this measure can be problematic for measuring SC since it only provides information about the length of speech which can increase due to extended speaking time in the unpressured online planning condition. Except for Wang (2014), none of the researchers employed a length based SC measure. Moreover, none of the studies investigated the possible changes in the complexity at phrasal and clausal levels.

Table 1. Summary of studies manipulating tasks along pressured/unpressured online planning

	n	Proficiency	Tasks	Mode	Complexity Measures	Finding
Yuan & Ellis (2003)	42	100–120 (max.150) on HEB Examination	Narrative (picture-based)	Oral	Subordination (Clauses per T unit), Syntactic variety (the total number of different grammatical verb forms used)	+ / =
Ellis & Yuan (2005)	42	100 – 120 (max. 150) on HEB Examination	Narrative (picture-based)	Oral	Subordination (Clauses per T unit), Syntactic variety (the total number of different grammatical verb forms used)	+
Ahmadian & Tavakoli (2010)	60	Intermediate	Narrative (video-based)	Oral	Subordination (Clauses per AS unit), Syntactic variety (the total number of different grammatical verb forms used)	+
Wang (2014)	77	540 - 630 on TOEFL and 6 - 7.5 on IELTS	Narrative (video-based)	Oral	Total Words, The average number of morphemes per AS unit, Subordination (clauses and verb infinitives per AS unit)	+
Ahmadian, Tavakoli & Dastjerdi (2015)	60	Intermediate	Narrative (video-based)	Oral	Subordination (Clauses per AS unit), Syntactic variety (the total number of different grammatical verb forms used)	+

Note. All of the studies were conducted with adult learners of English. *n* = number of participants; + shows that unpressured online planning had a positive effect, = shows a neutral effect.

Following the suggestions of Norris and Ortega (2009) and Bulté and Housen (2012), the current study aims to address the limitations of the previous studies by using

distinct and complementary measures of SC tapping into different subcomponents. In the light of these aims, the present study intends to answer the following research questions:

1. How does manipulating task complexity along online planning in an oral narrative task affect the SC levels of advanced L2 speakers?

2. Do distinct SC measures (overall complexity, complexity by subordination, and phrasal complexity) indicate the predicted variation between the pressured and unpressured online planning groups differently?

Method

Participants

Participants of this study were 64 undergraduate students majoring in English language education program at a public university in Turkey. The participants were 49 females and 15 males aged 19-24 from a fairly homogeneous group in terms of educational background and exposure to English. All the participants studied English intensively in high school. Moreover, they all had a pass mark from the institutional proficiency exam on which the minimum point accepted as successful is equivalent to 550 in TOEFL PBT (Paper-based), 79 in TOEFL IBT (Internet-based), and 6.5 in Academic IELTS. These participants had also taken several courses aimed at improving their written and oral skills in English during their first year in their program.

Research Design and Tasks

This study has a single-factor between-participant design with two randomized groups: pressured online planning group ($n = 32$) and unpressured online planning group ($n = 32$). All participants in the two groups were asked to tell a story based on a set of six picture frames. Oral tasks are argued to put language learners under more pressure than written tasks (Ellis, 1987), therefore these tasks may provide better opportunities to observe the impacts of online planning. Moreover, the task in the current study was a narrative task. There were several reasons for choosing an oral narrative task. Firstly, Foster and Skehan (1996) reported that narrative tasks are more likely to stimulate the use of complex language than other tasks such as decision-making tasks. Secondly, most of the previous online planning studies used oral narrative tasks, so employing an oral narrative task can make the results of this study more comparable to the previous studies.

As in previous online planning studies, we used pre-organized narrative tasks which did not require creating the storyline. The tasks included two wordless picture strips (Heaton, 1966, 1975) that had been frequently used in SLA research. The first picture is about two children (one girl and one boy) going on a picnic. They are unaware of their dog hiding in the picnic basket. They realize this situation when they open the picnic basket and see that there is no food left (the picnic story). In the second picture there is a boy going to his home without knowing that he has dropped one of the boxes that he is carrying. A man chases him to give him his box back (the chase story). The

selected picture strips are similar in terms of their storyline complexity, story structure, and code complexity. Both pictures include foreground and background events in their storylines. According to Tavakoli and Foster (2008), this type of storyline leads to more complex language since it requires the use of several subordinating conjunctions such as *because*, *therefore*, *after*, *before*, *when*, and *while* to explain the events happening in the background. Furthermore, both of the picture strips have a tight story structure. In other words, it is not possible to change the order of the six picture frames and the events in the story. Moreover, the code complexity in both picture strips are relatively easy. Both of the pictures require the use of easy vocabulary related to real life events. Although our participants were quite proficient speakers of English, we wanted to use picture strips with easy vocabulary since vocabulary related to specific subjects such as fishing terms in English is argued to affect participants' performances negatively (Sasayama, 2015).

Task Conditions and Procedure

The participants were randomly assigned to the two groups which differed in online planning opportunities. As in the previous online planning studies (Ahmadian & Tavakoli, 2010; Ellis & Yuan, 2005; Yuan & Ellis, 2003), the participants in the pressured online planning group were given limited time to tell their stories. The appropriate time limit for this group was determined with a pilot study conducted with 14 participants from the same university (Tuzcu & Yalçın, 2019). It was observed that all the participants in the pilot study completed the task between 90–120 seconds, therefore the time limit in the current study was set at 90 seconds. The participants in the unpressured online planning group, on the other hand, were allocated unlimited time to tell the story depicted on the picture strip.

Before the data collection, all the participants completed a background questionnaire about their personal and educational background and a consent form. Regardless of their group, all the participants were given one minute to look at the picture strip to ensure their understanding of the story. They were instructed to start narrating the story right after examining the picture strip for one minute. Depending on their group, they were either given unlimited time or 90 seconds to complete the task. The picture was present when the participants were telling the story. Each participant told only one story, either the picnic story or the chase story.

Measures

Time on Task. Following the previous online planning studies (Ahmadian & Tavakoli, 2010; Ahmadian, Tavakoli, & Dastjerdi, 2015; Ellis & Yuan, 2005; Yuan & Ellis, 2003), total time that each participant spent on task was measured and the two groups were compared to ensure that participants followed the instructions given to them. Participants in the unpressured online planning group were expected to spend more time on the task since they were given unlimited time to tell their stories. Participants in the pressured online planning group, on the other hand, were predicted to perform the task in a shorter time since they were given only 90 seconds to complete the task.

SC Measures. We adopted the absolute approach and considered SC as an objective and quantitative construct. Our working definition followed Pallotti's (2015) simple view of SC. AS-unit which is argued to be a more reliable and consistent unit for spoken data (Foster, Tonkyn, & Wigglesworth, 2000) was used as the base unit for the data. Similar to a T-unit, in an AS-unit, an independent clause refers to a clause with a finite verb and its dependent clauses. However, differently from a T-unit a subordinate clause may include a finite verb or a non-finite verb and at least one other clausal element (Foster et al., 2000). When coding the data into AS- units, in addition to syntactic criterion, intonation and pause information are taken into consideration. For example, some subordinate conjunctions such as *because* may function as an elided version of an independent clause. While this kind of clauses is always coded as a dependent clause in T-units, it can be regarded as an independent clause in AS-units depending on intonation and pausing. The following is an example of AS-unit coding.

Example: |probably John forgot something at the bus|
|so he brought back to him|
|because he was giving him another box|

In the example above, the upright slash indicates the boundaries of an AS unit. The sentence *because he was giving him another box* is coded as a separate AS unit instead of a dependent clause since there is a pause before it and it functions as an elided version of an independent clause such as *I say this because he was giving him another box*.

Following the suggestions of Norris and Ortega (2009) and Pallotti (2009, 2015), SC levels of participants' performances were measured with four different measures targeting different SC forms, i.e. the supra-clausal level, the clausal level, and the phrasal level. The measures and their computations are indicated in Table 2.

All measures were calculated from the pruned narratives. This means that only the final version of all repetitions, false starts, and self-corrections were taken into account when measuring the SC.

Table 2. The SC measures used in the study

Subtype	Measure	Computation
Overall (supra-clausal)	Mean length of AS-unit (LenAS)	Total number of words / Total number of AS-units
Sentential -Subordination	Subordination ratio (SR)	Total number of clauses / Total number of AS-units
Subsentential (Clausal + Phrasal)	Mean length of clauses (LenC)	Total number of words/ Total number of clauses
Phrasal	Mean length of noun phrases (LenNP)	Total number of words in NPs / Total number of NPs

Data Analysis

The narratives were transcribed by one of the researchers. The data were coded by one of the researchers and a second coder independently. Any disagreement between the coders was resolved with discussions. For all SC measures, the inter-coder reliability was more than 85%. For statistical analysis, the data were entered into SPSS version 21.0. The data were examined with descriptive statistics, inferential statistics, and effect sizes. Before conducting any analysis, the normality of the data was checked using some statistical tests and graphics such as skewness and kurtosis values, Shapiro-Wilk test, histograms, Q-Q plots, and box plots. Three measures, mean length of AS-unit, mean length of clauses and mean length of noun phrases, yielded normal distribution whereas subordination ratio measure showed non-normal distribution. For the three measures that showed normal distribution, an independent samples t-test was conducted. For the non-normally distributed measure, a Mann-Whitney U test for nonparametric data was performed. In addition to the inferential statistics, the effect size (Cohen's *d*) for each measure was also calculated following the suggestions of Cumming (2012) and Larson-Hall and Plonsky (2014). To answer the first research question, we examined the differences between the pressured and unpressured online planning groups. To answer the second research question, we examined the results of each SC measure separately.

Results

Time on Task

To see whether the two groups behaved similarly or differently under the instructions provided to them, for each group the length of time spent on the task was checked. According to the results, the unpressured online planning group (M = 87.60, SD = 44.56) spent more time on tasks than the pressured online planning group (M = 76.15, SD = 26.19). However, the difference between two groups was not statistically significant, U = 471.00, p = .58, d = .31.

Syntactic Complexity

As indicated earlier, the SC of participants' performances was assessed with four syntactic measures tapping into a different subcomponent. Table 3 indicates the descriptive statistics of the results for the four measures.

Table 3. Descriptive statistics for complexity measures

	Unpressured (Careful)			Pressured		
	Mean (SD)	Min.	Max.	Mean (SD)	Min.	Max.
Mean Length of AS-unit	9.34 (1.90)	6.60	12.63	9.70 (1.88)	5.45	13
Subordination Ratio	1.62 (0.38)	1.00	2.89	1.71 (0.37)	1.06	2.50
Mean Length of Clause	5.91 (0.86)	4.35	8.25	5.82 (0.91)	4.38	8.40
Mean Length of Noun Phrase	2.15 (0.33)	1.68	2.86	2.05 (0.38)	1.38	2.82

The four SC measures showed somewhat different results for the effects of online planning on L2 performance. For the overall complexity, the pressured online planning group had a higher mean score than the unpressured online planning group. However, an independent samples t-test did not reveal any statistically significant difference between the two groups, $t(62) = .748$, $p = .46$. The Cohen's d for overall complexity was found to be $d = .19$, indicating a small effect size. Similar results were obtained for the clause ratio. The pressured online planning group was found to have higher clause ratio scores than the unpressured online planning group although this difference was not statistically significant, $U = 433.5$, $z = -1.05$, $p = .29$, $d = .24$.

The clausal and phrasal level SC measures presented distinct results. The unpressured online planning group had higher mean scores than the pressured online planning group for both the mean length of clause and the mean length of noun phrases. However, these differences were not statistically significant, $t(62) = -.410$, $p = .68$ for the mean length of clause and $t(62) = -1.029$, $p = .31$ for the mean length of noun phrase. The effect sizes of the mean length of clause and the mean length of noun phrase were small, Cohen's ds were .10 and .28, respectively.

Discussion

This study aimed to investigate two questions: the effects of online planning on SC levels and the possible differences between four SC measures tapping into different subcomponents of SC. The findings regarding these issues will be discussed respectively.

Online Planning and Syntactic Complexity

Following the previous online planning studies (Ahmadian & Tavakoli, 2010; Ahmadian, et al. 2015; Ellis & Yuan, 2005; Yuan & Ellis, 2003) and Kormos' (2006) L2 speech production model, unpressured online planning was expected to lead to increased SC scores because of its positive effects on the conceptualizing and formulating stages of speech production. Unlike our predictions, there were no statistically significant differences between the pressured and unpressured online planning conditions for none of the four SC measures. In other words, SC levels did not

increase when the participants were provided with online planning opportunities. This result may show that regardless of the availability of online planning, all participants used their readily available resources.

The results of the present study were different from the findings of the previous studies (Ahmadian & Tavakoli, 2010; Ahmadian et al. 2015; Ellis & Yuan, 2005; Yuan & Ellis, 2003). The researchers reported increased SC levels in the unpressured online planning conditions. Their results conflict with the results of the current study. There can be two reasons for this difference in the findings.

Firstly, L2 proficiency can be a factor affecting L2 performance. The participants of the current study were majoring in English language education to be English teachers. Thus, they had high proficiency in English. Moreover, all participants had taken courses aiming at improving their oral skills during their first year at the department before participating in the current study. This difference in proficiency between the participants of the current study and the previous studies may have affected their L2 performance. As Kormos (2011) pointed out, when learners' L2 proficiency increases, the morphological, syntactic, and phonological rules that had been learned in the declarative form become proceduralized. Therefore, the retrieval of knowledge that once required controlled processing becomes more automatic. Wang (2014) also argued that time pressure affects the retrieval of "newly learned but not yet automatized linguistic structures" in L2 (p. 32). This automatization in the formulator of the speech production system can be a reason for the insignificant differences between the groups in the current study. In other words, as a result of automatization in the formulation stage, the time pressure in the pressured online planning condition might not have affected the formulation of their speech. Moreover, all the participants in the present study were given one minute to understand the pictures before the task. However, Wigglesworth (1997) reported that even one-minute planning time before the task was associated with higher SC levels for the advanced group of participants. Therefore, the one-minute pre-task planning time given to the participants might have decreased the time pressure on the conceptualizer, and this ease in the conceptualizing stage might have allowed the participants in both pressured and unpressured online planning groups to pay more attention to formulating their speech.

Secondly, in the current study the absence or presence of online planning was operationalized through the amount of time given to the participants. In the previous studies, on the other hand, the researchers not only put their participants under time pressure, but also asked them to produce at least four sentences for each picture frame. Therefore, the main pressure in the previous studies can be due to the combination of time limit and the number of sentences. This means that when used alone, time limit may not lead to decreased SC levels as predicted.

Syntactic Complexity Measures

Norris and Ortega (2009) and Bulté & Housen (2012) asserted that SC is a multidimensional construct and that the studies investigating L2 performance should use different and complementary measures reflecting each subcomponent. Norris and Ortega further argued that measures of phrasal complexity can be more suitable for assessing advanced learners' L2 performances since this group of learners generally use

nominalizations and modifications which increase the complexity at the phrasal level. In the light of this argument, we predicted that if the availability of online planning does result in increased SC, the difference between the pressured and unpressured online planning groups will be reflected in the phrasal complexity measures, namely, the mean length of clause and the mean length of noun phrases. However, since using the mean length of clause is posited to be a hybrid measure of phrasal and clausal complexity rather than being a pure measure of phrasal complexity (for further explanation on this issue see Bulté and Housen, 2012), we expected the length of noun phrases measure to reflect the difference better than the mean length of clause measure.

The descriptive statistics showed different patterns of change for each SC measure. Participants in the pressured online planning group were found to have greater SC at the supraclausal level measured as the mean length of AS-units and subordination ratio than participants in the unpressured online planning group. For SC measures at the clausal and phrasal complexity levels, on the other hand, participants in the unpressured online planning group had higher mean scores than the ones in the pressured online planning group. However, as stated previously, none of these differences were statistically significant. In other words, even the phrasal complexity level which is posited to be the best predictor for advanced learners was not affected by the manipulations of online planning. The important point here, related to the previous research question, is that although researchers generally manipulate various aspects of a task hoping to find differences in L2 performance, these manipulations may not reflect the differences even when complementary SC measures are used. The interaction between the tasks and L2 performance is far from straightforward and task demands are still in need of exploration. As argued by Pallotti (2009), "linguistic complexity grows when this is specifically required by the task and its goals, and not for the sake of it, as if learners aimed at complexification by default, as if they were eager to produce a subordinate clause or a rare lexical item whenever they have a pinch of free attentional resources" (p. 596).

Limitations and Future Research

The task employed in this study has been used in the field extensively (e.g. De Clercq & Housen, 2017; Ellis & Yuan, 2005; Tavakoli & Foster, 2008; Yuan & Ellis, 2003) and has met the requirements of the larger research study undertaken (for the larger study see Tuzcu, 2018). It was an oral pre-organized narrative task which required participants to narrate a story based on a series of six picture frames. Although the task required participants to integrate foreground and background events, participants did not need to manipulate the information since both pictures had a tight structure. Therefore, participants only needed to tell the story depicted in the pictures. Moreover, the task required the use of relatively easy vocabulary related to real life activities with preferably cause and effect. These features might have made the task rather easily manageable for the target group of participants, i.e. highly proficient speakers. Moreover, manipulating one aspect of the task, in this case the availability of online planning, might not have been enough to make the task cognitively challenging. Future studies of online planning with tasks other than narratives may report different findings.

Moreover, Ellis (2009) underscored the role of learners' attitude towards a task and how it may lead to differences in performance. The low cognitive demands of the task used in the study might have reduced learners' need to resort to available cognitive resources. In other words, low cognitive demands might have affected the participants' perceptions of the task, which in turn might have influenced their SC levels. However, this argument should be taken cautiously since there is not compelling evidence to suggest that SC is dependent on perceptions of task difficulty. In addition to attitude, Ellis (2009) stated that working memory, language aptitude, willingness to communicate, and anxiety are potential individual learner factors that could mediate the role of planning. The role of learners' perceptions of task difficulty and the effects of individual differences in SC levels should be investigated in future research.

The results of this study imply that preparing tasks which promote learners' SC levels is difficult. In other words, only providing extra time on task may not improve L2 performance. Teachers may provide learners pre-task planning opportunities in addition to decreasing time pressure through unpressured speaking conditions, since the combination of pre-task planning and online planning has been reported to be more effective (Wang, 2014). Planning can also be combined with form-focused instruction (Hulstijn & Hulstijn, 1994) to increase learners' accuracy. Moreover, instead of manipulating a feature of a task and expecting it to promote SC, teachers are encouraged to design tasks that will attract learners' attention to specific linguistics structures such as past tense and stimulate the use of those structures.

Conclusion

The main aim of this study was to examine how different task demands affect L2 learners' SC at sentential, clausal, and phrasal levels. Following Pallotti (2015), a simple view of SC was adopted and based on the suggestions of Norris and Ortega (2009) and Bulté and Housen (2012), SC was measured with four distinct but complementary measures tapping into different SC subcomponents. An oral narrative task was employed and the complexity of the task was manipulated along the presence and absence of online planning. There were no significant differences between the pressured and unpressured online planning groups for none of the four SC measures. This statistical insignificance in the results could be related to the proficiency level of the participants and nature of the task used. The findings indicated that instead of making generalizations about the impacts of more complex tasks on L2 performance, it is important to observe how specific task designs lead to specific L2 performance.

This study has several limitations. Firstly, the task used in the study may not be cognitively challenging enough for the participants and these low cognitive demands may have obscured the effects of online planning opportunities. Secondly, this study measured SC only with syntactic elaboration measures. However, syntactic elaboration measures may not capture all aspects of SC (Pallotti, 2009; Lambert & Kormos, 2014), therefore using syntactic diversity measures can lead to a more fine-grained analysis of syntactic complexity.

Authors' Note

References

Ahmadian, M. J., & Tavakoli, M. (2010). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research, 15*(1), 35–59.

Ahmadian, M. J., Tavakoli, M., & Dastjerdi, H. V. (2015). The combined effects of online planning and task structure on complexity, accuracy and fluency of L2 speech. *The Language Learning Journal, 43*(1), 41-56.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5-35.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (eds.). *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA* (pp. 21–46). Amsterdam: John Benjamins.

Bulté, B., & Housen, A. (2018). Syntactic complexity in L2 writing: Individual pathways and emerging group trends. *International Journal of Applied Linguistics, 28*, 147–164.

De Clercq, B., & Housen, A. (2017). A cross-linguistic perspective on syntactic complexity in L2 development: syntactic elaboration and diversity. *The Modern Language Journal, 101*(2), 315–334.

Cumming, G. (2012). *Understanding the new statistics. Effect sizes, confidence intervals, and Meta-Analysis*. New York, NY: Routledge.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style in the use of past tense. *Studies in Second Language Acquisition, 9*, 12-20.

Ellis, R. (2005). Planning and task-based performance: Theory and research. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 3- 34). Amsterdam: John Benjamins.

Ellis, R., & Yuan, F. (2005). The effects of careful within-task planning on oral and written task performance. In R. Ellis (ed.), *Planning and task performance in a second language* (pp. 167–92). Amsterdam: John Benjamins.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*(3), 299-323.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 11*(2), 354-375.

Heaton, J. B. (1966). *Composition through pictures*. Essex: Longman.

Heaton, J. B. (1975). *Beginning composition through pictures*. London: Longman.

Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition, 38*, 163–175.

Hulstijn, J.. & Hulstijn, W. (1984). Grammatical errors as a function of processing constraints and explicit knowledge. *Language Learning, 34*, 23–43.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kormos, J. (2011). Speech production and the cognition hypothesis. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 39-60). Amsterdam: John Benjamins.

Kormos, J., & Trebits, A. (2011). Working memory capacity and narrative task performance. In Peter Robinson (ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 267–285). Amsterdam: John Benjamins Publishing.

Kormos, J., & Trebits, A. (2012). The role of task complexity, modality and aptitude in narrative task performance. *Language Learning*, 1-34.

Kuiken, F., & Vedder, I. (2011). Task complexity and linguistic performance in L2 writing and speaking: The effect of mode. In Peter Robinson (ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 91-104). Amsterdam: John Benjamins Publishing.

Lambert, C., & Kormos, J. (2014). Complexity, accuracy, and fluency in task-based L2 research: Toward more developmentally based measures of second language acquisition. *Applied Linguistics, 35*, 606-614.

Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning, 65*(S1), 127-159.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Michel, M., Shi, D., & Li, Y. (2019). The effects of task demands on linguistic complexity and accuracy across task types and L1/L2 speakers. In Z. Wen & M. J. Ahmadian (eds.) *Researching L2 task performance and pedagogy: In honour of Peter Skehan*, (pp. 133-152). Amsterdam: John Benjamins.

Miestamo, M. (2009). Implicational hierarchies and grammatical complexity. In G. Sampson, D. Gil, & P. Trudgill (eds) *Language complexity as an evolving variable* (pp. 80–97). Oxford: Oxford University Press.

Norris, J. M., & Ortega, L. (2009). Measurement for understanding: An organic approach to investigating complexity, accuracy, and fluency in SLA. *Applied Linguistics, 30*(4), 555–578.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*(4), 590–601.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research, 31*, 117–134.

Révész, A., Sachs, R., & Mackey, A. (2011). Task complexity, uptake of recasts, and L2 development. In Peter Robinson (ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 203-235). Amsterdam: John Benjamins Publishing.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics, 45*, 193-213.

Sasayama, S. (2015). *Validating the assumed relationship between task design, cognitive complexity, and second language task performance*. Unpublished doctoral dissertation. Georgetown University, Washington, DC.

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 58*(2), 439–473.

Trebits, A. (2016). Sources of individual differences in L2 narrative production: The contribution of input, processing, and output anxiety. *Applied Linguistics, 37*(2), 155-174.

Tuzcu, A. (2018). *Task type and online planning as mediating factors in the oral performance of L2 learners*. Unpublished master's thesis. Boğaziçi University, İstanbul.

Tuzcu, A., & Yalçın, Ş. (2019). The combined effects of manipulating tasks in two dimensions on L2 speech performance. In H. Wilson et al. (eds.), *Selected Proceedings of the 2017 Second Language Research Forum*, (pp. 175-184). Somerville, MA: Cascadilla Proceedings Project.

Vercellotti, M. L. (2019). Finding variation: assessing the development of syntactic complexity in ESL Speech. *International Journal of Applied Linguistics, 29*(2), 233-247.

Wang, Z. (2014). On-line time pressure manipulations: L2 speaking performance under five types of planning and repetition conditions. In Peter Skehan (ed.), *Processing perspectives on task performance* (pp. 27-62). Amsterdam: John Benjamins.

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14*(1), 85–106.

Yoon, H. J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly, 51*, 275-301.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity, and accuracy in L2 monologic oral production. *Applied Linguistics, 24*(1), 1-27.

Ödev Koşullarının İkinci Dil Performansına Etkileri: Karmaşıklık Ölçütlerinin İncelenmesi

Öz

Sözdizimsel karmaşıklık, ödev özelliklerinin etkilerini incelemek üzere çeşitli çalışmalarda bağımlı değişken olarak kullanılmıştır. Önceki çalışmaların büyük bir çoğunluğunda genel karmaşıklık ve yan tümceleme karmaşıklığı incelenmiş olup, söz öbeği karmaşıklığı araştırılmamıştır (Bulté & Housen, 2012). Bu çalışma, sözdizimsel karmaşıklık ölçülerinin seçimi ve ödev tasarımının ikinci dil öğrenenlerin sözlü performanslarını nasıl etkilediğini incelemiştir. Çalışmada çevrimiçi planlama fırsatlarına göre iki gruba ayrılmış 64 üniversite öğrencisinin sözlü olarak anlattıkları sözel hikayelere bakılmıştır. Sözdizimsel karmaşıklık; tümce, yantümce ve söz öbeği karmaşıklıklarını gösteren dört farklı ölçüt ile incelenmiştir. Çalışmanın sonucunda iki grup arasında istatiksel olarak anlamlı bir fark gözlemlenmemiştir. Sonuçlar; kullanılan sözdizimsel karmaşıklık ölçüleri, çalışmada kullanılan ödevin özellikleri ve katılımcıların ikinci dildeki yeterlilikleri göz önünde bulundurularak tartışılmıştır.

Anahtar kelimeler: Çevrimiçi planlama, sözdizimsel karmaşıklık, ikinci dilde sözlü performans