CHAPTER 6

The effects of task demands on linguistic complexity and accuracy across task types and L1/L2 speakers

Marije Michel, Groningen University, NL, and Lancaster University, UK
Andrea Révész, University College London, UK
Danni Shi, University College London, UK
Yanmei Li, University of Barcelona, Spain

Abstract

The relationship between task variables and linguistic production has been the object of much second language (L2) research. This study contributes to this line of research by investigating the extent to which increasing cognitive task demands affects the syntactic complexity and accuracy of L2 performance across different speaker groups (first and second language) and task types. An additional goal of the study was to explore how various measures of linguistic accuracy and complexity pattern together. The participants were English L1 speakers (N=16) and German L2 speakers of English (N=16), who performed cognitively less and more demanding versions of three task types (decision-making, map, and narrative). Syntactic complexity was evaluated in terms of measures of overall, subordination and phrasal complexity, and accuracy was assessed with error-free clause and weighted clause ratios. Reflecting predictions by Skehan (2009, 2015), the results showed that the effects of cognitive task demands on syntactic complexity and accuracy varied according to task type and speaker status. Subordination complexity had a strong positive relationship with overall complexity, but correlated negatively with phrasal complexity. A limited number of trade-offs between the syntactic complexity and accuracy measures was also attested.

Introduction

The past three decades have seen a growing interest in task-based language teaching (TBLT) and learning (e.g., Bygate, Skehan, & Swain, 2000; Ellis, 2003; Long, 1985, 2015; Skehan, 1998, 2014; Samuda & Bygate, 2008). In light of the problems identified with traditional synthetic syllabuses where the content and order of instruction are primarily defined in terms of linguistic difficulty, second language (L2) researchers have been calling for a shift towards analytic approaches where syllabus construction is more learner-centred and informed by psycholinguistic rather than linguistic criteria alone (e.g., Long, 1985; Long & Crookes, 1992; Skehan, 1998). Among the analytic approaches, task-based language teaching has been argued to be especially promising, given that adopting tasks as a unit of analysis has the potential to make instruction relevant to students' specific needs, and, as a result, raise their interest and motivation in language learning in a classroom setting (Robinson, 2001a). A number of different approaches to TBLT (e.g., Breen, 1987; Candlin, 1987; Prabhu, 1987, see van den Branden, Bygate & Norris, 2009, for an overview) have been put forward over the years, grounded in areas such as education, general psychology, and/or second language acquisition (SLA) research. The present study is conceptualised in terms of Skehan's proposal for task-based instruction (Skehan, 1996, 1998; Skehan & Foster, 2001), an approach primarily motivated by SLA theory and research findings (see also, Long, 1985, 2015; Robinson, 2001a, Robinson & Gilabert, 2007).

Skehan's task-based approach reflects a cognitive, information processing perspective. Its principal goal is to create a psycholinguistic environment that fosters balanced L2 development in the linguistic performance areas of complexity (i.e., the ability to produce advanced and elaborate interlanguage constructions), accuracy (i.e., the ability to use the L2 correctly), and fluency (i.e., the ability to use the L2 in real time). Skehan contends that this can be achieved by carefully selecting, manipulating and sequencing pedagogic tasks, given that various task characteristics exert differential cognitive and linguistic demands and thereby affect the three areas of learner production in distinct ways. Partly motivated by this proposal (see also, Robinson, 2001a; Robinson & Gilabert, 2007), a considerable amount of research has set out to identify task variables that can exhort learners to prioritize either of these performance areas when engaged in task-based work.

The aim of this study was to contribute to this research domain by investigating the effects of increased task demands on two areas of linguistic performance: syntactic complexity and accuracy. In other words, we intended to explore how the cognitive complexity of tasks may influence the extent to which a speaker focuses their attention to challenging language (complexity) or avoidance of errors (accuracy). Although there is abundant research on the relationship between task manipulations and linguistic performance, so far relatively few studies have considered the extent to which these links may generalise across various task types and participants with native or non-native mastery of the target language. Our goal was to help address these gaps in the literature by comparing first language (L1) and L2 users' performance on less and more cognitively demanding versions on three task types (decision-making, map, and narrative). In addition, we were interested in exploring how various measures of linguistic accuracy and complexity might pattern together, given that there are indications in the literature that task factors might affect different subconstructs associated with general linguistic measures in distinct ways (Skehan, in press). In the section that follows, we first describe Skehan's Limited Attention Capacity Model (2009, 2015), which constitutes the theoretical basis for Skehan's notion of task-based development. Then, we turn to a brief review of recent empirical work relevant to the current investigation.

The Limited Attention Capacity Model

Theoretical Assumptions

The fundamental assumption underlying the Limited Attention Capacity Model is that working memory capacity and attentional resources are limited, and therefore there will be a competition for attentional resources during speech production processes. In making predictions about how capacity limitations will affect the relationship between task manipulations and L2 linguistic performance, Skehan draws on Levelt's (1989) L1 speech production model and its application to L2 speaking by Kormos (2006, 2011). These models describe four stages of speech production. The first stage, *conceptualisation,* involves the macro-planning of a message and micro-planning of information transmission, in other words, it entails the planning of what one wants to say. The resulting pre-verbal message then enters the second stage, *formulation*, which transforms the pre-verbal message into linguistic form through the processes of lexical retrieval, and syntactic and phonological encoding. Next, at the *articulation* stage, the phonemic representation activates phonological forms and retrieves articulatory gestures to prepare the actual speech utterance. An additional stage, *monitoring*, occurs at each of the previous stages operating in multiple feedback loops. In principle, all four stages can work in a parallel, incremental fashion. Indeed, for L1 speakers, formulation and articulation operations tend to be automatic, thus can work in parallel with conceptualisation processes. For L2 learners, on the other hand, formulation will likely be less automatic and take up greater attentional capacity, given that L2 users have a less developed mental lexicon and less advanced grammatical encoding skills (Kormos, 2006). In turn, fierce competition for cognitive resources will ensue, yielding trade-offs in the amount of attention L2 users can dedicate to conceptualisation versus formulation processes, which might become visible as trade-offs between complexity and accuracy measures (Skehan, 2009).

Building on these models, Skehan (2009, 2015) argues that task factors can make separate demands on conceptualisation and formulation, leading to “complexifying/pressuring” and “easing/focusing” influences in relation to various stages of speech production. For example, conceptualisation operations are expected to come under more pressure when a task involves a larger number of steps; more complex manipulation and transformation of information; greater need to justify actions, address challenges, or link different pieces of information; and deal with unfamiliar and abstract information. In contrast, formulation processes will be challenged more if, for instance, the task requires the use of more sophisticated and diverse lexis, lacks a clear structure and online planning opportunities, and allows for little choice about what is said. Against these theoretical ideas, a key practical issue is how increased pressure on the conceptualiser and/or formulator will be reflected in performance, expressed in terms of the linguistic areas of complexity, accuracy, and fluency.

As Skehan explains, at this stage of research, it is impossible to make encompassing predictions about how these three linguistic areas will be affected during task performance. This is so because "production is influenced by many factors, and one can only understand such speech production by attempting to integrate these diverse influences within a wide and accommodating framework" (Skehan, 2015, p. 137). However, drawing on Levelt's speech production model and the assumption that there are limitations to attentional capacity, some predictions can be made for how linguistic areas will be affected during oral performance for specific learner populations and task manipulations (Skehan, 2015). For example, it is expected that, with increasing proficiency, there will be diminished trade-off effects among various linguistic areas, because formulator processes will demand less attentional resources, thereby decreasing the competition between formulator and conceptualiser influences. Turning to task effects, we can also hypothesize that pressure on the conceptualiser will lead to greater linguistic complexity, whereas easing influences on the conceptualiser (e.g., greater structure) might release attention for formulator operations and, hence, promote accuracy (Skehan, 2009). In the next paragraph we turn to a review of previous empirical research testing the predictions of the Limited Attention Capacity Model, with a focus on task variables, speaker status (L1 versus L2), and linguistic areas (syntactic complexity and accuracy) that are directly relevant to this investigation.

Empirical findings: Task effects

As predicted, the overall findings of previous empirical research seem to suggest that task manipulations that put increased demands on the conceptualiser will, indeed, lead to greater syntactic complexity, when complexity is operationalised as a subordination ratio (e.g., clause per AS-unit). For example, tasks that prompt speakers to make more connections between various pieces of information were found to push speakers towards more extensive use of subordinate clauses (Foster & Tavakoli, 2009; Tavakoli & Foster, 2008; Tavakoli & Skehan, 2005). Also, in a study by Skehan and Foster (1997) greater need to transform information during task performance was found to lead to higher subordination ratios.

There is also empirical evidence in the literature demonstrating that, as predicted by the Limited Attention Capacity Model, decreasing conceptual demands will increase accuracy. For instance, greater topic familiarity (Bui, 2014) and task structure (Skehan & Shum, 2014; Wang & Skehan, 2014) were observed to yield more accurate performance in a number of studies, probably because lower demands on the conceptualiser freed up resources for allocating attention to formulation processes. Interestingly, narrative tasks with greater structure also promoted the incidence of subordination in Skehan and Shum (2014) and Wang and Skehan (2014). As Wang and Skehan explain, clearer macrostructure might have afforded speakers the opportunity to elaborate on details more, given that the speakers needed to grapple less with conceptualising the main storyline.

Another intriguing finding emerging from these two studies is that the results for phrasal (e.g., words per clause) and subordination (e.g., clause per AS-unit) complexity do not overlap. In Skehan and Shum (2014), higher task structure was associated with increased levels of subordination, but participants produced longer clauses when retelling less structured video narratives. The study by Wang and Skehan (2014) yielded no significant links between task structure and phrasal complexity. In neither of the studies were significant correlations detected between the phrasal and subordination complexity measures for L2 speakers. As Skehan (in press) argues, a possible explanation for the differential findings for the two measures may be that, depending on task condition, speakers adopt a broader (e.g., in structured narratives) or more local (e.g., in less structured narratives) discourse emphasis, reflected in greater subordination or phrasal complexity respectively.

Empirical findings: Speaker status

To date, little research has considered how L1/L2 speaker status might influence the relationship between task manipulations and linguistic performance. However, to separate the effects of task-related factors from effects that are due to the incomplete mastery of the L2, it is essential to examine how task manipulations influence the linguistic performance of L1 versus L2 speakers (De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012; Foster & Tavakoli, 2009). Among the few studies that have investigated L1 and L2 differences in relation to task effects (e.g., De Jong et al., 2014; Foster, 2001; Foster & Tavakoli, 2009; Michel, 2011), Foster and Tavakoli's (2009) research is particularly relevant to this study. One of the aims of Foster and Tavakoli (2009), similar to this investigation, was to determine the extent to which manipulating task factors may differentially affect the syntactic complexity of task performances across native and non-native speakers. The researchers found that increasing the complexity of the storyline in narrative tasks led to greater syntactic complexity (measured by ratio of clauses to AS-units and words per AS-unit) in both native and non-native performances. This had led the researchers to conclude that increased syntactic complexity on tasks with more complex storylines seems to be, at least partly, a task effect. On the other hand, Michel's (2011) study, comparing performance on less versus more cognitively demanding decision-making tasks, showed that the latter increased lexical complexity of both L1 and L2 speakers, but task demands had an impact on neither syntactic complexity nor accuracy. Clearly, more research is needed on links between speaker status, task manipulations, and linguistic performance.

Research Questions

The current study aimed to address the following research questions:

1. To what extent does increasing the cognitive demands of tasks affect the syntactic complexity and accuracy of task-based production? Do these relationships vary by task type and speaker status?

2. To what extent are the various measures of syntactic complexity (phrasal, subordination and overall complexity) related to each other? Do these relationships vary by cognitive task demands and task type?
3. To what extent are measures of syntactic complexity and accuracy related to each other? Do these relationships vary by cognitive task demands and task type?

Methodology

Design

The data set for the present study comes from a larger project described in Révész, Michel, and Gilabert (2016). For the purposes of this research, we analysed data from 16 first language (L1) speakers of English (henceforth, L1 English group) and 16 German L1 speakers of English L2 (henceforth, L2 English group). All the participants performed a more and a less cognitively demanding version of three oral tasks – a decision-making, a map and a narrative task. Cognitive task demands and task type were counterbalanced across the participants. The validity of the task demand manipulations was confirmed in Révész, Michel, and Gilabert (2016). In other words, the versions designed to be less cognitively demanding were found, as intended, to pose lower cognitive demands as compared to the version of each task designed to be more cognitively demanding (Norris, 2010; Révész, 2014).

Participants

The mean age of the L1 and L2 English groups was 22.8 (SD = 4.4) and 23.7 (SD = 6.2) respectively. Like age, the gender distribution was similar across the groups (L1 English: 8 females; L2 English: 9 females). The majority of the participants were undergraduate students at an English (11 BA; 4 MA; 1 PhD) or German (12 BA; 3 MA; 1 PhD) university. The L2 speakers were at B2 to C1 levels of proficiency according to the Common European Framework of Reference, as established by a paper-and-pencil version of the Oxford Placement Test (OPT, mean score = 75.3, SD = 9.9, min/max = 63/91). Participants received a small monetary reward for taking part in the research.

Tasks

We adapted the decision-making tasks from Gilabert and colleagues' work (Gilabert, 2007; Gilabert, Barón, & Llanes, 2009). Participants were told that they volunteered for the university's fire emergency team, and their task was to describe the actions they would take in case of an emergency based on a picture of a building on fire. Under the less cognitively demanding condition, the people trapped in the building looked similar and did not seem vulnerable. There were also sufficient resources to extinguish the fire, and the participants only had to consider a few unrelated factors. Under the more cognitively demanding condition, on the other hand, participants had to save a group of vulnerable people, had fewer resources available, and needed to take into account several dynamic factors (e.g., wind direction, fire moving). Drawing on Skehan (2009, 2015), the latter version of the task was anticipated to generate more pressure on conceptualisation processes, as completing the task entailed a larger amount of manipulation and transformation of information, and a greater need to solve challenges and link different pieces of information.

We also adapted the map tasks from Gilabert and colleagues' work (Gilabert, 2007; Gilabert et al., 2009). They required participants to leave a voice mail providing directions on how to get around the city center to perform some errands. Under the less cognitively challenging condition, participants were asked to describe a route along one lateral axis (left, right, and straight) involving unique, easily distinguishable points of reference. Under the more demanding condition, participants had to give directions not only along the lateral (left, right, straight) but also the sagittal (i.e., front, back) and vertical (up, down) axes. In addition, the complex map task included several similar reference points. Following Skehan (2009, 2015), we expected that the version designed to be more challenging would lead to increased demands on conceptualisation; because it involved more steps, required making more links between various pieces of information, and exposed students to task input that was less clear.

Following Robinson (2007), the narratives were taken from the picture arrangement subtest of the Wechsler Adult Intelligence Scale, Third Revised version (WAIS-III). We used Story 3 (LAUNDRY), one of the narratives with lesser complexity, and Story 11 (SHARK), the most complex narrative. The version with lower cognitive demands asked participants to describe simple activities, which were carried out by a single character doing his laundry. The version with higher cognitive demands required the narration of a story involving a surfer and included a set of unpredictable events. Based on Skehan (2009, 2015), we assumed that the more cognitively demanding version would impose greater processing load both in terms of conceptualisation and formulation. Having to deal with less familiar information was expected to increase demands on the conceptualiser, whereas lack of clear structure was predicted to exert greater strain on the formulator.

Procedure

The students participated in one individual session. After providing informed consent, they completed a background questionnaire. The L2 participants were also administered a paper and pencil version of the Oxford Placement Test (Dave, 2004), which lasted a maximum of 40 minutes. The psycholinguistic software E-Prime (Schneider, Eschman, & Zuccolotto, 2002) was utilised to deliver the rest of the experiment. First, participants received the task instructions and took part in a practice phase, as part of which they listened to a sample task performance, carried out the practice task, and completed an example task perception questionnaire. The aim of the practice phase was to familiarise students with the computer-administered procedure employed in the experimental tasks.

Then, the six experimental tasks followed, each preceded by a 30-second planning period, and followed by a task perception questionnaire. In total, participants took 40-60 minutes to complete the practice and experimental tasks.

Linguistic analyses

All speech recordings were transcribed verbatim. Next, we removed repairs, repetitions and false starts from the transcripts, which were then segmented into analysis of speech (AS) units (Foster, Tonkyn & Wigglesworth, 2000), clauses, and subordinate clauses. Following Norris and Ortega (2009), we operationalised syntactic complexity using multiple measures to capture the multidimensional nature of the construct. Overall complexity was assessed by mean length of AS-unit (words per AS-unit); subordination complexity was expressed in terms of clauses per AS-unit and subordinate clauses per AS unit; and phrasal complexity was gauged by calculating the mean length of clauses (words per clause). To assess accuracy, we utilised two global measures: percentage of error-free clauses and Weighted Clause Ratio (WCR), an index recently proposed by Foster and Wigglesworth (2016). The calculation of WCR involved determining whether a clause was entirely correct (1.0 point), included minor errors not compromising meaning (0.8 points) or errors interfering with meaning (0.5 points), or was incomprehensible (0.1 points). The ratio was computed by dividing the sum of points obtained with the total number of clauses.

Ten percent of the data were double coded for AS-units and number of clauses/subordinate clauses by trained research assistants. Percentage agreement was 85.4% for identification of AS units and 95.6% for that of clauses. Intercoder agreement for error-free clauses and WCR were, respectively, 83.4% and 89.7%.

Statistical analyses

First, descriptive statistics were calculated for the linguistic complexity and accuracy measures across task types and speaker groups. Due to the non-normal distribution of the data, we opted for non-parametric inferential statistics. To address the first research question, a series of Wilcoxon Signed Ranks were conducted. An alpha level of $p < .05$ was set for all tests, and r was calculated to measure effect sizes. Following Plonsky and Oswald (2014), r values of .25, .40, and .60 were considered small, medium, and large, respectively. The second research question was addressed by computing Spearman correlations between the various measures of syntactic complexity.

Results

Cognitive demands, task type and speaker status

Table 1 summarises the descriptive statistics for syntactic complexity and accuracy across task conditions, task types and different speaker groups. Out of the 36 Wilcoxon Signed Ranks tests, which were conducted to compare the syntactic complexity and accuracy of participants' performances across the less and more cognitively demanding task conditions, 11 yielded significant relationships (see Table 2).

On the more cognitively demanding version of the decision-making task, L2 users produced more subordinate clauses per AS-unit, but average clause length was found to be shorter. For the L1 group, task demands had no effects on these indices. Neither were the rest of the measures affected by the task manipulation regardless of speaker status.

The more cognitively demanding map task yielded lower number of clauses per AS-unit and error-free clauses for the L1 speakers. Task demands, however, had no impact on the remaining measures for the L1 group. The L2 groups' linguistic performance was not affected by the task manipulation, as reflected in non-significant differences in the syntactic complexity or accuracy measures across the two task conditions.

On the narrative task, increased cognitive demands led to greater number of subordinate clauses per AS-unit, lower phrasal complexity, and higher number of error-free clauses for the L2 participants. L1 speakers produced more clauses per AS-unit, shorter clauses but less accurate language when task demands were increased. In addition, the L1 group demonstrated greater overall complexity. The task manipulation had no impact on the other indices.

Table 1. Descriptive statistics for syntactic complexity and accuracy by task type, task demands and speaker status

		Decision Making				Map				Narrative			
		Low demand		High demand		Low demand		High demand		Low demand		High demand	
		L2	L1	L2	L1	L2	L1	L2	L1	L2	L1	L2	L1
		n=16	n=16	n=16	n=16	n=16	n=16	n=16	n=16	n=16	n=16	n=16	n=16
Clauses/AS	Median	2.17	2.88	2.31	3.00	1.93	2.68	1.78	2.31	2.13	2.38	2.73	3.14
	IQR	0.74	1.06	1.32	1.14	1.29	1.11	0.69	0.52	0.60	0.96	1.36	1.14
Subordinate clause/AS	Median	0.73	1.60	0.89	1.47	0.47	1.11	0.37	0.74	0.54	0.80	0.85	1.44
	IQR	0.70	1.23	1.12	0.92	0.94	0.89	0.38	0.69	0.38	0.50	0.82	1.22
Mean length of AS	Median	17.08	21.75	17.24	21.72	15.23	18.06	13.33	17.02	15.98	16.69	16.75	18.67
	IQR	4.73	9.71	6.85	10.95	8.14	7.54	3.68	4.67	4.29	7.56	7.25	5.28
Mean length of clause	Median	8.08	7.31	7.28	6.95	7.27	6.82	7.64	7.18	7.79	6.97	6.77	6.15
	IQR	2.08	1.45	1.42	1.18	1.42	0.88	1.94	1.33	1.29	1.44	1.13	0.91
Error free clauses (%)	Median	0.74	1.00	0.82	0.97	0.86	1.00	0.87	0.99	0.67	1.00	0.77	0.96
	IQR	0.29	0.03	0.18	0.06	0.16	0.00	0.13	0.05	0.28	0.00	0.21	0.06
Weighted clause ratio	Median	0.86	1.00	0.91	0.99	0.93	1.00	0.94	1.00	0.88	1.00	0.89	0.99
	IQR	0.12	0.01	0.10	0.02	0.06	0.00	0.08	0.01	0.10	0.00	0.14	0.02

Note. IQR: Interquartile Range; AS=Analysis of Speech unit

Table 2. Results for Wilcoxon Signed Ranks tests assessing the effects of task complexity by task type and language group

			Clause/AS	Subclause /AS	Mean length of AS	Mean length of clause	Percent of error-free clauses	Weighted clause ratio
Decision Making	L2	Z	-1.758[a]	-2.017[a]	-.052[a]	-2.896[b]	-1.590[a]	-1.874[a]
		p	.079	.044	.959	.004	.112	.061
		r	-0.311	-0.357	-0.009	-0.512	-0.281	-0.331
	L1	Z	-.052[a]	-.362[b]	-.310[b]	-.827[b]	-1.412[b]	-.628[b]
		p	.959	.717	.756	.408	.158	.530
		r	-0.009	-0.064	-0.055	-0.146	-0.250	-0.111
Map	L2	Z	-1.913[b]	-1.396[b]	-1.603[b]	-.931[a]	-.259[a]	-.362[a]
		p	.056	.163	.109	.352	.796	.717
		r	-0.338	-0.247	-0.283	-0.165	-0.046	-0.064
	L1	Z	-2.430[b]	-1.371[b]	-1.241[b]	-1.655[a]	-2.073[b]	-1.718[b]
		p	.015	.171	.215	.098	.038	.086
		r	-0.430	-0.242	-0.219	-0.293	-0.366	-0.304
Narrative	L2	Z	-2.159[a]	-1.913[a]	-.879[a]	-2.689[b]	-1.988[a]	-.931[a]
		p	.031	.056	.379	.007	.047	.352
		r	-0.382	-0.338	-0.155	-0.475	-0.351	-0.165
	L1	Z	-1.758[a]	-2.741[a]	-.414[a]	-2.637[b]	-2.580[b]	-2.535[b]
		p	.079	.006	.679	.008	.010	.011
		r	-0.311	-0.485	-0.073	-0.466	-0.456	-0.448

Note. AS = AS unit; [a] = based on positive ranks; [b] = based on negative ranks.

Relationships between measures of linguistic complexity and accuracy

Table 3 displays the results of the Spearman correlational analyses, which were conducted to examine the extent to which the various measures of syntactic complexity and accuracy correlated with each other. The syntactic complexity measures show similar patterns across the different task versions and task types. The subordination indices of clause per AS-unit and subclause per AS-unit are strongly correlated, and show strong relationships with overall complexity. Except for the more demanding version of the map task, the analyses also yielded small- to medium-size negative correlations between the subordination and phrasal complexity measures. Turning to accuracy, for all task versions and task types, the weighted clause ratios and percentages of error-free clauses demonstrate very high correlations. Finally, the relationships between the accuracy and syntactic complexity measures are relatively consistent across task conditions and task types. The analyses revealed small- to medium-size correlations between the subordination complexity and accuracy measures, but few links were found between the accuracy and the overall complexity and phrasal complexity measures. These were observed, respectively, on the less demanding version of the decision-making and map tasks.

Table 3. Spearman Correlation Coefficients for all tasks and participants (N=32)

Spearman Rho			Subclause /AS	Mean length of AS	Mean length of clause	Error-free clause %	Weighted clause ratio
Complex	Decision Making	Clause/AS	.949***	.915***	-.372**	.206	.208
		Subclauses/AS		.886***	-.369**	.221	.226
		Mean length of AS			-.051	.299	.319
		Mean length of clause				-.120	-.041
		Error-free clause %					.979***
	Map	Clause/AS	.899***	.791***	-.326	.474***	.437**
		Subclauses/AS		.747***	-.241	.495***	.492***
		Mean length of AS			.254	.341	.255
		Mean length of clause				-.191	-.233
		Error-free clause %					.961***
	Narrative	Clause/AS	.804***	.874***	-.516***	.210	.335
		Subclauses/AS		.733***	-.466***	.413**	.477***
		Mean length of AS			-.115	.167	.292
		Mean length of clause				-.209	-.250
		Error-free clause %					.944***
Simple	Decision Making	Clause/AS	.934***	.750***	-.592***	.373**	.377**
		Subclauses/AS		.643***	-.608***	.302	.308
		Mean length of AS			-.041	.357**	.328
		Mean length of clause				-.042	-.069
		Error-free clause %					.988***
	Map	Clause/AS	.875***	.896***	-.459***	.360**	.414**
		Subclauses/AS		.737***	-.544***	.511***	.563***
		Mean length of AS			-.085	.252	.290
		Mean length of clause				-.414**	-.428**
		Error-free clause %					.986***

Narrative	Clause/AS	.631***	.798***	-.435**	.371**	.384**
	Subclauses/AS		.427**	-.498***	.025	.083
	Mean length of AS			.127	.210	.181
	Mean length of clause				-.283	-.337
	Error-free clause %					.975***

Note. AS = AS unit; *** =<.001; ** =<.01;

Discussion

Our first research question addressed the extent to which increasing the cognitive demands of tasks affect the syntactic complexity and accuracy of task-based production, and whether these relationships vary by task type and speaker status. A series of Wilcoxon Signed Ranks tests, which were conducted to gauge the effects of cognitive task demands on linguistic performance, revealed different patterns across task types as well as speaker groups. In light of this and given that task factors are central to this study, the discussion for the first research question is organised around the three task types.

Task-related effects

The results obtained for the L2 group on the decision-making task are partly consistent with our expectations. Following Skehan (2009, 2015), we anticipated that, under the more demanding condition, the increased need to manipulate and transform information and to link different pieces of information would put greater pressure on conceptualisation processes, which, in turn, would result in greater syntactic complexity. This projection was borne out in the results for subordination complexity, given that participants showed greater ratio of subordinate clauses per AS-unit when they performed the more cognitively demanding task version. Interestingly, however, the opposite trend was observed for phrasal complexity; participants produced shorter clauses when task complexity was increased. These results are similar to those obtained by Skehan and Shum (2014), where task structure had a differential impact on subordination and phrasal complexity. As Skehan and Shum speculate, task changes may differentially affect these two indices, with subordination complexity being pushed by the need "to express connections between elements through clause relations" (p. 203), whereas phrasal complexity might be sensitive to other task manipulations. The results of the present study additionally suggest that these two measures might have a negative relationship (see below). A question that arises is why these task effects were only evidenced in the L2 speakers' production. A possible explanation is that L1 speakers, given their more automatic linguistic encoding processes, were less affected by the increased conceptual demands and faced no difficulty producing increased subordination and longer clauses.

The findings for the map task run counter to our expectations. Based on Skehan (2009, 2015), we anticipated that the more challenging map task, given that it involved more steps, required making more links, and entailed less clear task input, would lead to increased conceptual demands and, as a result, lead to increased linguistic complexity. To the contrary, however, the L2 groups' performance was not affected by the task manipulation, and the L1 group displayed lower subordination on the more cognitively demanding task version. Similar results were obtained by Gilabert, Barón, and Levkina (2011) and Robinson (2001b): In neither study was the syntactic complexity of L2 participants' performance affected by cognitive task demands, regardless of whether they carried out dialogic (Gilabert et al., 2011, Robinson, 2001b) or monologic map tasks (Gilabert et al., 2011). A possible explanation for the lack of impact on L2 learners might lie in the fact that, although giving directions based on the more complex map posed greater cognitive demands, the nature of the syntax required to successfully complete the task was similar across the two task versions, requiring the use of short and clear imperative constructions. The lower complexity observed for the L1 speakers on the more cognitively demanding task version might have been due to the L1 speakers' attempt to make their language more concise in order to relieve the processing burden on the listener. This, in turn, might have led to slightly lower accuracy, as they might have been a little less able to balance these two performance areas.

Moving onto the narrative task, the results can, in large part, be explained in terms of Skehan's model (2009, 2015) and they align with Luo and Xin (this volume). We assumed that the more cognitively challenging narrative would exert greater processing pressure on both conceptualisation and formulation. Having to produce a narrative without a familiar schema was anticipated to lead to enhanced demands on the conceptualiser and, hence, promote syntactic complexity. On the other hand, lack of clear structure was expected to put more strain on formulation processes, leading to decreased accuracy. As expected, the L1 participants demonstrated lower accuracy as well as greater subordination and overall complexity on the more demanding narrative. The more challenging task condition also led to more extensive subordination in the L2 group, but, contrary to our prediction, it also resulted in greater accuracy. In line with what was found for the decision-making tasks, clause length was negatively affected by increased cognitive demands for both speaker groups. This provides further evidence for a potential inverse link between these two measures. The difference in accuracy results for the L1 and L2 groups is intriguing, and calls for further exploration.

Relationships between linguistic complexity and accuracy measures

Our second research question was concerned with the relationships among the various syntactic complexity indices and whether the nature of these links might vary according to task-related factors. The correlational analyses, overall, yielded consistent patterns across the various task types and task conditions. As expected, the two subordination complexity indices were found to be strongly correlated, and had a strong link with overall complexity. Interestingly, however, the correlational analyses yielded small- to medium-size negative correlations between the subordination and phrasal complexity measures (see Bui, this volume, for similar findings). This means that, when participants produced more subordinate clauses, they used shorter clauses; or vice versa, when there was a decrease in subordination, there was a simultaneous increase in clause length. Following Skehan (in press), one possible explanation for this finding is that students might adopt either a broader or more local discourse emphasis, the latter leading to higher incidence of subordination and the former to greater clause length. Based on the current dataset, these two approaches do not seem compatible: Participants did not, in general, push for greater subordination and phrasal complexity concomitantly. Taking these results together with the findings for the task complexity manipulations, it would appear that cognitive task demands might be one factor that influences whether participants prioritise broader or local discourse-oriented processing. On the narrative and decision-making tasks, greater demands on the conceptualiser resulted in higher levels of subordination but generated lower clause length. As discussed above, the lack of effects on the map task might have been due to the fact that this task type, due to its inherent nature, did not elicit syntactically complex constructions.

In response to our third research question, the extent to which complexity and accuracy measures relate to each other, we observed small to medium positive relationships on most tasks, with the strongest correlations on the more cognitively demanding version of the map task. This might have been due to the L1 speakers in the data set, who could generate both syntactically complex and accurate language. Interestingly, we also observed a negative correlation between mean length of clause and both accuracy measures on the less cognitively demanding version of the map task. It seems that, on this task, participants who produced longer clauses were more prone to making errors. This suggests that there might have been trade-off effects between this complexity measure and accuracy, as proposed by Skehan (2009) and earlier found by Sample and Michel (2014). A possible reason why only length of clause (and not the other complexity measures) shows this relationship might be that this measure is more sensitive to differences among learners at higher levels of proficiency than subordination or general complexity indices (Norris & Ortega, 2003).

Finally, there are several limitations to this study that need to be acknowledged and considered in future research. First, our sample size was small. A larger sample size would have allowed for using more sophisticated statistical procedures to address our first research question, and running separate correlations for the L1 and L2 groups in an attempt to answer the second and third research questions. Second, we only focused on a single level of proficiency. In future research, it would be interesting to explore whether the patterns found here would transfer to lower and higher proficiency L2 learners. Lastly, the study included a limited number of task types and only two levels of cognitive complexity manipulations. It would seem worthwhile to replicate this study with additional task types and levels of task demands, as our results were not uniform for all measures across the three task types we examined.

Conclusion

One aim of this study was to examine the effects of increasing task demands on the complexity and accuracy of linguistic performance, and the extent to which these links may be influenced by task type and speaker status (L1 versus L2). We found that the impact of cognitive task demands on linguistic complexity and accuracy indeed varied according to task type and speaker status. Drawing on Skehan's (2009, 2015) work, we were able to interpret these results considering the inherent nature of the task types and task manipulations. Our second goal was to explore how various measures of linguistic complexity might relate to each other. Subordination complexity, as expected, correlated strongly with overall complexity. However, it was found to have a negative relationship with phrasal complexity, suggesting that the two measures are associated with differential discourse processing (Skehan, in press). Finally, we were interested in investigating the extent to which measures of linguistic complexity and accuracy correlated. The positive correlations observed between the complexity and accuracy measures could be attributed to the proficiency level of the participants, while the negative correlations between clause length and accuracy might be linked to trade-off effects (Skehan, 2009).

Acknowledgements

This research was in part supported by a Trinity Research Internship awarded to Marije Michel by the Linguistics and English Language Department at Lancaster University. We thank David Stutt for his support with transcribing and coding the data. We would also like to thank Mohammad Ahmadian and Wen Zhisheng (Edward) for initiating this volume and the series editors for their feedback. Finally, we are grateful to Peter Skehan for his contribution to the field and for his unwavering support of and inspiring engagement with students and early career researchers.

References

Breen, M. P. (1987). Learner contributions to task design. In C. N. Candlin & D. F. Murphy (Eds.), *Language learning tasks. Lancaster Practical Papers in English Language Education, Vol. 7* (pp. 23-46). Englewood Cliffs NJ: Prentice-Hall International.

Bui, H. Y. G. (2014). Task readiness: Theoretical framework and empirical evidence from topic familiarity, strategic planning and proficiency levels. In P. Skehan (Ed.). *Processing perspectives on task performance* (pp. 63-93). Amsterdam/Philadelphia: John Benjamins.

Bygate, M., Skehan, P., & Swain, M. (2000). Introduction. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: second language learning, teaching and testing*. Harlow, Essex: Longman.

Candlin, C. N. (1987). Towards task-based language learning. In C. N. Candlin and D. F. Murphy (Eds.), *Language learning tasks* (pp. 5-22). London: Prentice-Hall International.

Dave, A. (2004). Oxford placement test. Oxford: Oxford University Press.

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., Hulstijn, J. H. (2012). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In Housen, A., Kuiken, F., Vedder, I. (Eds.), *Dimensions of*

L2 performance and proficiency: Complexity, accuracy and fluency in SLA (pp. 121–142). Amsterdam/Philadelphia: John Benjamins.

Ellis, R. (2003). *Task-based language teaching and learning*. Oxford, UK: Oxford University Press.

Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language learning*, *59*(4), 866-896.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21(3),* 354-375.

Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics*, *36*, 98-116.

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching*, *45* , 215 – 240.

Gilabert, R., Barón, J., & Levkina, M. (2011). Manipulating task complexity across task types and modes. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp.105-138). Amsterdam/Philadelphia: John Benjamins.

Gilabert, R., Barón, J., & Llanes, À. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *IRAL-International Review of Applied Linguistics in Language Teaching*, *47*(3-4), 367-395.

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

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* (p.39-59). Amsterdam/Philadelphia: John Benjamins.

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

Long, M. H. (1985). A role for instruction in second language acquisition: task-based language teaching. In K. Hyltenstam & M. Pienemann (Eds.), *Modeling and assessing second language development* (pp. 77-99). Clevedon, UK: Multilingual Matters.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley-Blackwell.

Long, M. H. & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly*, *24*, 27-56.

Michel, M. (2011). Effects of task complexity and interaction on L2 performance. In P. Robinson *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (p.141-173). Amsterdam/Philadelphia: John Benjamins.

Norris, J. M. (2010, September). *Understanding instructed SLA: Constructs, contexts, and consequences.* In Plenary address delivered at the annual conference of the European Second Language Association (EUROSLA), Reggio Emilia, Italy.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30(4)*, 555-578.

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning, 64(4)*, 878-912.

Prabhu, N. S. (1987). *Second language pedagogy*. Oxford: Oxford University Press.

Révész, A. (2014). Towards a fuller assessment of cognitive models of task-based learning: Investigating task-generated cognitive demands and processes. *Applied Linguistics, 35*, 87–92.

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments: A validation study. *Studies in Second Language Acquisition*, *38(4),* 703-737.

Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for investigating task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). New York: Cambridge University Press.

Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27-57.

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.

Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *International Review of Applied Linguistics, 45*, 161–176.

Sample, E., & Michel, M. (2015). An exploratory study into trade-off effects of complexity, accuracy, and fluency on young learners’ oral task repetition. *TESL Canada Journal, 31*, 23-46.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning.* Basingstoke, UK: Palgrave Macmillan.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime (Version 2.0). [Computer software and manual]. Pittsburgh, PA Psychology Software Tools Inc.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied linguistics*, *17*(1), 38-62.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, UK: Oxford University Press.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30(4)*, 510-532.

Skehan, P. (2014) (Ed.). *Processing perspectives on task performance*. Amsterdam/Philadelphia: John Benjamins.

Skehan, P. (2015). Limited attention capacity and cognition: Two hypotheses regarding second language performance on tasks. In M. Bygate (Ed.). *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (pp. 123-156). Amsterdam/Philadelphia: John Benjamins.

Skehan, P. (in press). Second language task-based performance: Theory, research, pedagogy. New York: Routledge.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*, 185-211.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press.

Skehan, P., & Shum, S. (2014). Structure and processing condition in video-based narrative retelling. In P. Skehan (Ed.). *Processing perspectives on task performance* (pp. 187-210). Amsterdam/Philadelphia: John Benjamins.

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.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239 – 277). Amsterdam/Philadelphia: John Benjamins.

Van den Branden, K., Bygate, M., & Norris, J. M. (2009). *Task-based language teaching: A reader*. Amsterdam/Philadelphia: John Benjamins.

Wang, Z., & Skehan, P. (2014). Structure, lexis, and time perspective: Influences on task performance. In P. Skehan (Ed.). *Processing perspectives on task performance* (pp. 155-185). Amsterdam/Philadelphia: John Benjamins.