Task complexity and interaction: (combined) effects on task-based performance in Dutch as a second language[1]

Marije C. Michel[1], Folkert Kuiken[2], and Ineke Vedder[2]

[1]University of Mannheim, [2]University of Amsterdam

Abstract

This paper presents the overarching conclusions of three consecutive investigations into task-based L2 performance. It aims at giving a better understanding of how changes in the number of elements referred to in a task affect L2 production, and how this relates to cognitive task complexity. Furthermore, it evaluates differences between monologic and dialogic tasks, and searches for combined effects of the factors '± elements' and '± monologic'. Analyses examined the oral task performances of 152 participants by using global measures of linguistic complexity, accuracy, and fluency as well as a task specific measure. Results revealed hardly any effects of the manipulation of the number of elements. Dialogic tasks, however, consistently guided L2 performers towards greater accuracy, lexical complexity, and fluency. The discussion compares these findings to native speaker baseline data, reviews the results in light of Robinson's (2005) Cognition Hypothesis, and highlights the cognitive impact of the factor '± monologic'.

Introduction

Research into the cognitive aspects of task-based language production aims at understanding the cognitive and attentional processes during L2 task performance (Skehan 2003). Focused attention has been identified as one of the key prerequisites for L2 development (Long & Robinson 1998; Schmidt 2001). Many investigations manipulate task characteristics that may affect attentional allocation. In this line of research, the cognitive load a task puts forward, i.e., cognitive task complexity, has gained a lot of attention (Robinson 2001, 2005; Skehan 1996; Skehan & Foster 2001). However, within this cognitive strand there exist contrasting hypotheses concerning factors of cognitive task complexity.

Attentional capacity and cognitive task complexity

Skehan and Foster advocate the idea of limited attentional capacity (Skehan 1996, 2009, Skehan and Foster 2001), which means that there is some maximum in the amount of information one can keep active or pay attention to. Also the number or size of attentionally controlled processes that can take place in parallel is limited (Schmidt 2001).

The central point of Skehan and Foster's account is that during task-based L2 performance the different ongoing processes are in competition with each other for attentional resources (Skehan and Foster 2001). Accordingly, when performing a task the available resources have to be shared between the processes a task asks for. The control function of attention will direct where to allocate resources. Only those aspects that receive enough attention will reach optimal performance while other processes fail or become erroneous, that is, a decline in performance results from limited attention. Skehan and Foster argue that during L2 tasks learners first and foremost aim at reaching the communicative goal, i.e., they will prioritize meaning over form (VanPatten 1990). The attention that is left for form needs to be shared between linguistic complexity, accuracy, and fluency. According to the Limited Attentional Capacity model (also known as Trade-off Hypothesis) cognitively complex tasks put L2 learners under attentional pressure such that by default trade-off effects between the dimensions of performance occur. These manifest themselves most obviously between linguistic complexity and accuracy (Skehan 1996, 2009; Skehan & Foster 2001).

In contrast to the Limited Attentional Capacity Model, the Cognition Hypothesis by Robinson (2001, 2005) claims that L2 learners can rely on multiple pools of attention (based on Wickens

2007). Robinson states that not every complex task necessarily induces trade-off effects because different processes may draw on various attentional pools.

Robinson (2005) presents a Triadic Componential Framework that names different factors of task design and explains how they affect task performance. Table 1 presents the cognitive factors of task complexity named in Robinson's (2005) model. Within them he distinguishes (a) *resource-directing* from (b) *resource-dispersing* factors of task complexity. Increases on the latter (e.g., by means of reducing planning time) may similarly to Skehan's model disperse attention away from the task so that L2 performance decreases.

Table 1 about here

However, as long as task complexity is kept simple along resource-dispersing variables increases in complexity along so-called resource-directing variables are expected to focus the learner's cognitive resources towards the language form (Robinson 2001, 2005; Robinson & Gilabert 2007). This may encourage L2 learners to perform in a syntactic mode of processing (Givón 1995). For example, a complex task, which asks to give a description of many elements, will need more complex syntactic structures and different lexical items than a task with only a few elements, where simpler structures suffice. Upon complex tasks overall L2 performance increases because the heightened attention to form will lead language learners to use more accurate speech that is structurally complex and is based on a wide range of vocabulary items. Accordingly, resource-directing cognitively complex tasks can promote linguistic complexity and accuracy in parallel. Crucially, Robinson's model claims that no competition for attentional resources occurs because L2 learners can rely on different attentional resource pools (Robinson 2001; Wickens 2007). Only fluency possibly decreases in cognitively demanding tasks, because fluency is of a more performative nature and, therefore, may suffer from the high processing effort (Levelt 1989).

Earlier work testing the Cognition Hypothesis

Empirical studies investigating the Cognition Hypothesis by means of the factor '± here-and-now' have indeed found that a higher cognitive task complexity promoted linguistic complexity and accuracy as proposed by the Cognition Hypothesis (Gilabert 2005; Ishikawa 2007). With respect to the factor '± elements', findings, however, are less conclusive. This may result from the fact that it is left open in the Triadic Componential Framework how exactly this factor should be operationalized such that studies have chosen different ways to manipulate the number of elements.

Robinson (2001) tested university students, who in a simple map task had to give directions on a well known route on the campus while the complex version of this task addressed an unknown and larger area of town. This study attested the predicted outcome on lexical complexity but parallel increases on accuracy manifested itself as trend effects only. Two studies that used the same manipulations on a map task are Gilabert (2007) for monologic and Gilabert et al. (2009) for dialogic performances, respectively. In the simple task participants had to give directions on a map with few, clearly distinguishable elements. The complex task presented more elements that also were more difficult to distinguish from each other. Gilabert et al. (2011) report on combined analyses of these studies and show that complex tasks yielded a higher accuracy and lexical diversity than simple tasks when reviewing the monologic data – dialogic tasks did not find significant differences.

In a series of written tasks Kuiken et al. (2005) and Kuiken and Vedder (2007) asked L2 learners of French and Italian to take into account three (simple) versus six (complex) different criteria when deciding between five possible holiday destinations. Their data yielded partial support for the Cognition Hypothesis: cognitively complex tasks triggered a higher accuracy while linguistic complexity was influenced in opposite directions in different populations.

Finally, Révész (2011) asked participants to discuss in groups of three or four how much money they would give to a fund organization. The complex tasks required learners to choose among a greater number of proposals while the financial resources where limited (i.e., '+ elements' and

‘+ reasoning demands’). The simple tasks had more money and fewer funds to possibly acknowledge (i.e., ‘– elements’ and ‘– reasoning demands’). Results revealed that complex tasks yielded more accurate language that was more lexically complex compared to the simple version. Yet, structural complexity decreased in complex task performances.

In sum, these empirical studies do generate some support for the Cognition Hypothesis, as accuracy seems to be pushed in complex tasks and most results show that lexical complexity increases too when participants are confronted with a greater number of elements. Crucially, none of these studies found an increasing effect on structural complexity. Furthermore, there seem to be differences between monologic in contrast to dialogic or group performances, such that more work would be welcome.

Influences of task complexity on interaction

The Cognition Hypothesis formulates (in contrast to the Limited Attentional Capacity model) expectations with respect to the impact of task complexity on task-based interaction. Robinson (2001, 2005) argues that complex tasks by their very nature will put up more content and linguistic problems such that they need more clarification. Therefore, he predicts that if L2 learners act on cognitively complex tasks in pairs this will result in more interaction than paired work on simple tasks (Robinson & Gilabert 2007). Following research findings about the effect of interaction (Ellis & He 1999; Gass & Varonis 1994; Long 1990; Mackey 1999; Pica 1994) the Cognition Hypothesis relates cognitively complex tasks to a higher accuracy and complexity of L2 performance. That is, interactional moves like clarification requests, comprehension checks, joint negotiation of form and meaning and other forms of language related episodes (LREs) focus the learners’ attention to the language code, e.g., when interlocutors ask for modified input or are required to modify their own output based on negative feedback by a speaking partner.

As a result, interactive tasks are expected to yield more complex and accurate speech. Robinson (2001) addresses an important issue as he proposes that the nature of interaction possibly mitigates against the effects of focused attention due to task complexity. That is, participants presumably do not show the expected increases in structural complexity on cognitively complex tasks because turn-taking and other interactional moves prevent them to build complex linguistic structures.

Finally, as cognitively complex tasks create a higher need for form-focused interaction than simple tasks they also enhance uptake and intake of information from the input (Long & Robinson 1998). Consequently, the Cognition Hypothesis claims that complex interactive tasks generate more language learning opportunities than simple interactive tasks and as such may promote second language acquisition (Robinson & Gilabert 2007).

A body of research has investigated interactional moves, e.g., clarification requests, comprehension checks, modified output, during task-based L2 production (e.g., Gass & Varonis 1994; Mackey 1999; Shehadeh 2004) and more recent work specifically compared cognitively simple and complex interactive performances (Nuevo 2006; Révész 2009; Robinson 2001; Gilabert et al. 2009). The present study, however, aims at investigating differences in task performance by means of global measures of linguistic Complexity, Accuracy, and Fluency (in short CAF) between simple and complex task performances in a monologic versus a dialogic setting. That is, it aims at investigating the factor '± elements' in combination with the factor '± monologic'.

A monologic versus dialogic task condition

To date, the factor '± monologic' has not received much attention in task-based research such that it remains an empirical question how this factor affects L2 performance. From a cognitive perspective, it is interesting to know how a monologic or dialogic task condition affects attentional processes during task performance. Taking a pedagogical view, it is important to understand how this difference influences e.g., the amount of learning opportunities or how to evaluate individual contributions of learners within a monologic or dialogic setting.

As explained above, a dialogic setting gives L2 learners the possibility to interact such that they can benefit from the interactional processes focusing their attention to form and meaning. On the other hand, dialogues may also distract learners attention from their own performance because interlocutors hold the double role of speaker and hearer. As Rost (2011) explains, listening and comprehension in interaction is a complex and demanding task. In contrast, in a monologic task condition learners can stay with their own knowledge and resources but they also have to rely on these. They do not receive other feedback and no interactional modifications will focus their attention to neither form nor meaning. The only way to generate modified output is by monitoring their own speech, an effortful process that needs time and attentional capacity especially in the L2 (Kormos 2000).

As the presence of an interlocutor may create planning time during the interlocutor's turn Tavakoli and Foster (2008) suggest that task performance in interaction is cognitively simpler than language processing in a monologic situation. In other words, the absence of an interlocutor in a monologue possibly generates an increase in cognitive demands by itself as there is no partner to rely on. This assumption receives support from psycholinguistic research among native speakers. Pickering and Garrod (2004) state that in interaction, interlocutors tend to mirror each other's speech on all linguistic levels such as syntax, semantics, phonology. Their Alignment Hypothesis argues that copying of the other's speech "greatly simplifies production and comprehension in dialogue" (Pickering & Garrod 2004: 169).

First, in a dialogue the speaking partner's turn creates planning time. While listening to the speaker the hearer can conceptualize his own speech act. As a result, the hearer has more cognitive capacity for formulating during his own speaking turn because less attention for online planning and conceptualization is needed. Furthermore, interlocutors tend to *help out* as soon as the partner falls silent in order to keep a constant flow of interaction. Therefore, interactive tasks are more fluent than monologic tasks. In contrast, in a monologue all processes of language production need to be performed at the same time. As this is a complex cognitive task that requires attention, speakers may process information in a mostly serial way, which may cause hesitations (Levelt 1989).

Second, another simplification in dialogues may occur due to priming. Priming is the easier availability of words and syntactic structures that have been pre-activated through related items, e.g., by the interlocutor (Meyer & Schvaneveldt 1971). Mutual priming may again ease language production in dialogues, such that they put up a lower cognitive load than monologues. As Costa et al. (2008) point out, L2 learners possibly do not equally benefit from alignment and priming as L1 speakers do. And, as pointed out before, the role of an L2 listener creates complexities too. Still, it is possible that L2 learners, like L1 speakers, profit from the simplifying processes of alignment and priming in interaction.

Up to now, hardly any investigation looked at the difference between task performances manipulated on the factor '± monologic'. In a meta-analysis Skehan and Foster (2007) compared their data on monologic tasks with data on dialogic tasks and found that dialogues pushed accuracy and complexity but decreased fluency. Even so, this meta-analysis compared L2 performances on various different tasks that manipulated different task factors such that it is difficult to draw general conclusions. First evidence for differential effects of task complexity in monologic versus dialogic tasks have been described by Gilabert et al. (2011, see above). The present studies aim at presenting a systematic comparison of monologic versus dialogic performances on the same simple and complex tasks. Not least, they investigate the combined effects of cognitive task complexity (by means of the factor '± elements') and interaction (manipulating the factor '± monologic').

The present studies

This paper presents the findings of three consecutive empirical investigations manipulating task complexity and interaction (Michel et al. 2007; Michel 2011, in press). It combines the insights from these earlier studies in order to receive a full perspective on the issues under investigation.

The goal is to investigate effects of changes in task design on L2 production by manipulating the factors '± elements' (i.e., task complexity) and '± monologic' (i.e., interaction) both on their own as well as in combination. They evaluate the specific claims of the Cognition Hypothesis (Robinson 2005) regarding these manipulations and add to our understanding of how these factors influence L2 performance.

Research Questions and Hypotheses

The following research questions guide the empirical investigations:

RQ1 What is the effect of increased cognitive task complexity on L2 oral task performance?

RQ2 What is the effect of interaction on L2 oral task performance?

RQ3 Are there any combined effects of increased cognitive task complexity and interaction on L2 oral task performance?

Hypothesis 1, concerning the factor '± elements', follows Robinson's claims about increases in task complexity in monologic and dialogic task performances. Hypothesis 2 is theoretically based on insights from priming (Meyer & Schvaneveldt 1971), the Alignment Hypothesis (Costa et al. 2008; Pickering & Garrod 2004), and assumptions by Tavakoli and Foster (2008). In addition, this paper investigates the empirical question concerning the factor '± monologic', whether dialogic tasks are cognitively less complex than monologic tasks.

Assuming dialogues to be cognitively less complex than monologues would predict, on the one hand, the L2 learners' complexity, accuracy and fluency to increase in interaction because of cognitive ease. On the other hand, turn-taking and alignment/copying would predict dialogic tasks to be structurally and lexically less complex. Similarly, it is hypothesized that in complex interactive tasks, contrasting forces of interaction and cognitive task complexity may mitigate each other.

The third hypothesis follows the Cognition Hypothesis, that is, it predicts more interaction in cognitively complex tasks with its effects on CAF-measures. Data will reveal the nature of these contrasting effects. The hypotheses are summarized as follows (cf. Table 2) :

H1 Increased cognitive task complexity results in higher accuracy and higher linguistic complexity, but lower fluency of L2 oral task performance (cf. Robinson 2005).

H2 In comparison to monologic tasks, interactive tasks raise the accuracy and fluency of L2 oral task performance while it is an empirical question how linguistic complexity is affected.

H3 Increased cognitive task complexity promotes accuracy and interaction. As interaction by itself promotes accuracy, cognitively complex interactive tasks are expected to largely increase the accuracy of L2 oral task performance. Facing the contrasting forces of task complexity and interaction, it is an empirical issue how linguistic complexity and fluency are affected by the combination of these two factors.

Table 2 about here

Method and design

As data are drawn from earlier investigations where the method and design have been thoroughly described (Michel et al. 2007; Michel 2011, in press) this paper sketches these aspects only briefly.

Based on the Triadic Componential Framework (Robinson 2005, Table 1) the studies investigate effects of an increased cognitive task complexity by means of the factor '± elements'. Furthermore, they explore effects of the manipulation of the interactive factor '± monologic'. Both factors are systematically examined on their own as well as in combination. There were two experiments that both followed a 2x2 design where cognitive task complexity was implemented within participants and interaction between participants. That is, all participants did a simple (+ elements) and a complex (– elements) task but half of the participants acted on their own (+ monologic), the other half worked together in pairs (- monologic). In the monologic setting they were told to leave a phone message on the answering machine of a friend. In the dialogic setting, participants were asked to discuss with each other on the phone about the task outcome. The order of presentation of the different versions and settings was counterbalanced over participants.

Participants

The experimental group of participants of the present investigations consisted of 108 adult learners of Dutch as a second language who had their first contact with Dutch after puberty. They were all of Turkish or Moroccan background and were recruited from different language institutes in the Netherlands. On average they had resided in the Netherlands for three and a half years and as they were attending or had just finished higher education most of them were in their twenties. All were students with a higher educational background.

At the moment of testing they were about to take or just had taken the State Examination for Dutch as a second language. Accordingly, they were classified by their teachers to be at an intermediate level of proficiency. Their estimated level was assessed by a written proficiency task.

In the second experiment a control group of 44 native speakers performed the same tasks under the same conditions as the non-native participants. The native speakers were mostly students at the University of Amsterdam and on average a bit younger than the non-natives. They scored at ceiling on the language proficiency task. Table 3 summarizes the background information for all participants.

Table 3 about here

Tasks and design study 1

In the first study L2 learners received a full-color leaflet with two (simple) and six (complex) electronic devices (MP3 players or mobile phones), respectively. The gadgets differed from each other in seven relevant features (e.g., price, color, capacity). Two versions of the same leaflet were created concerning either MP3 players or mobile phones. Participants were asked to formulate a phone message for a friend who needed advice about the electronic device he or she would buy.

All speech samples were transcribed and coded using CLAN (MacWhinney 2000). The output was coded for global CAF-measures of production: two syntactic (subordinate / main clauses per Analysis of Speech, AS-unit, Foster et al. 2000) and two lexical (Guiraud's Index and percentage of lexical words) measures of complexity, three measures of accuracy (number of errors, lexical errors and omissions), and five of fluency (for speed in syllables per minute in pruned and unpruned speech, silence by means of filled pauses, and self-repair of errors and non-errors, cf. Tavakoli and Skehan 2005). All measures were corrected for sample length, for example, by relating them to the AS-unit. More elaborate information on this study can be found in Michel et al. (2007).

Results study 1 by means of global CAF-measures for L2 learners

An analysis of variance (ANOVA) tested for effects of task complexity and interaction, both on their own and in combination. Detailed statistics of this study are reported in Michel et al. (2007). To summarize, complex tasks affected one measure within each CAF dimension, that is, complex performances yielded a higher linguistic complexity and accuracy but a lower fluency (percentage of lexical words: $F(1,42)=4.47$, $p<.05$; total number of errors per AS-unit: $F(1,42)=6.63$, $p<.01$; syllables per minute: $F(1,42)=13.17$, $p<.001$). None of the other measures yielded significant results.

Interaction showed a statistically significant effect for both syntactic measures such that monologues were syntactically more complex than dialogues (total number of clauses per AS-unit: $F(1,42)=29.37$, $p<.001$; number of subordinate clauses per total number of clauses: $F(1,42)=8.87$, $p<.01$;). None of the lexical measures of linguistic complexity were significantly affected. Monologues compared to dialogues yielded less accurate speech (total errors: $F(1,42)=7.63$, $p<.001$; lexical errors: $F(1,42)=5.26$, $p<.05$; omissions: $F(1,42)=3.99$, $p<.05$, all related to the AS-unit). Also fluency increased in dialogic tasks by means of speed and pausing behavior but not regarding self-repairs (unpruned speech: $F(1,42)=10.71$, $p<.01$; pruned speech: $F(1,42)=13.65$, $p<.001$); filled pauses per AS-unit: ($F(1,42)=13.17$, $p<.001$).

There was one significant combined effect of cognitive task complexity by interaction. That is, the pushing effect, that a higher task complexity had on accuracy in monologues, disappeared in the dialogic condition where accuracy showed no difference between simple and complex tasks (total number of errors: $F(1,42)=7.63$, $p<.01$; lexical errors: $F(1,42)=5.26$, $p<.05$; and omissions per AS-unit: $F(1,42)=3.99$, $p<.05$).

Discussion of results study 1

Concerning hypothesis 1, these data are in line with the claims of the Cognition Hypothesis. That is, linguistic complexity and accuracy both increased in cognitively complex tasks while fluency decreased. However, the confirmation may not be very strong as the effects were visible on three out of twelve measures only. Moreover, no combined effects were found that confirm Robinson's (2005) claims. That is, in dialogues the effects of task complexity disappeared, while the Cognition Hypothesis predicts that accuracy and complexity should be promoted in complex interactive tasks and that this effect may be even stronger than in monologic tasks due to the increased attention to form in interaction.

Regarding hypothesis 2, dialogic tasks showed the predicted increasing effects on all three accuracy measures. As they pushed fluency by means of speed and silence measures, these data support the view that dialogues may be cognitively less effortful than monologues (Costa et al. 2008; Meyer & Schvaneveldt 1971; Pickering & Garrod 2004; Tavakoli & Foster 2008). The decreased (syntactic) complexity, however, may highlight the role of turn-taking behavior that possibly prevented participants to build complex structures.

With regard to hypothesis 3 the present data do not give a conclusive picture. There were hardly any combined effects of the factors '± elements' and '± monologic' and the one that evolved contradicted the predictions.

A closer look at the study made us realize that there were some methodological shortcomings, which may account for these findings. First, the amount of speech produced by the participants was sometimes rather limited such that the results possibly suffer from a bias due to sample length. Second, some of the measures seem to be co-linear and redundant (Norris & Ortega 2009). Third, participants had not received pre-task planning time. However, the Cognition Hypothesis states that resource-dispersing variables should be kept simple in order to allow focused attention to form in complex tasks manipulated on a resource-directing factor (Robinson & Gilabert 2007). Most importantly, the study has no proof whether the manipulation by means of the factor '± elements' actually established a difference in task complexity. That is, no external means were used to evaluate the theoretically assumed difference between simple (with two elements) and complex (with six elements) tasks. As this was the first empirical study that systematically looked at (combined) effects of task complexity and interaction, a second experimental study was prepared in order to overcome these problems.

Tasks and design study 2

For the second study measures were adjusted according to recent insights of SLA literature (van Daele et al. 2007; Housen & Kuiken 2009; Norris & Ortega 2009). Moreover, in order to be able to make more valuable interpretations of the L2 data, this second study tested a baseline of native speakers. Finally, two sets of new tasks were designed, the 'study' and the 'dating' tasks, that asked participants to find the best combination of two persons (a study pair in the study task and a love couple in the dating task). Again, task complexity was manipulated on the factor '± elements' (Robinson 2005) but the differences in the amount of information to be processed was footed in psychological research (Halford et al. 2007). Cognitive psychological work argues that human cognitive capacity is limited to a maximum of four items. The tasks in study 2 asked participants to keep possible combinations of dating and study couples active in memory while evaluating their value as a pair. The simple task gave the opportunity to build four combinations of people whereas the complex task asked for a decision among nine possible pairs. Accordingly, the simple tasks in this study addressed a number of elements that were within the assumed human limitations whereas the complex tasks included an amount that should be beyond these cognitive capacity limits. As an external method to evaluate established task complexity, participants were asked to rate their performance after completing the tasks on a 5 point Likert scale regarding difficulty, stress, success, and motivation (Robinson 2001).

Sixty-four L2 learners and forty-four L1 speakers of Dutch were instructed to prepare a simple and a complex phone message for a friend about the best couple in the dating or the study task. They received 2 minutes pre-task planning time in order to keep complexity low along resource-dispersing factors. Again participants performed both tasks either in a monologic or both in a dialogic condition.

The 216 simple and complex speech samples were transcribed and coded for global measures of linguistic complexity, accuracy, and fluency. There were two syntactic and one lexical measures of complexity (words per clause, subordinate clauses per AS-unit, Guiraud's Index). Accuracy was gauged by morphosyntactic, lexical, and determiner errors related to AS-units. Following Tavakoli and Skehan (2005) fluency was coded for by looking at speed (syllables per second in pruned and unpruned speech), pausing (number of filled pauses per AS-units), and breakdown fluency (self-repairs per AS-unit). For details on the method, design and analysis of this study, see Michel (2011).

Results study 2a by means of global CAF-measures for L2 learners and L1 speakers

The statistics included separate mixed general linear model ANOVAs for each CAF-construct and for each language group (L2 versus L1). A detailed report of this analysis is given in (Michel 2011).

To summarize the findings, with regard to the factor '± elements' results revealed in both language groups only one single significant effect of moderate size. That is, the complex task yielded a higher lexical complexity by means of Guiraud's Index (L2 learners: $F(1,62)=6.22$, $p<.05$, $\eta^2=.09$; L1 speakers: $F(1,42)=6.72$, $p<.05$, partial $\eta^2=.14$). No other measure was affected significantly.

Effects of the factor '± monologic' were large in both language groups. In dialogues, L2 learners were syntactically less complex but showed a wider lexical range (words per clause: $F(1,62)=16.71$, $p<.001$, $\eta^2=.21$; number of subordinate clauses per AS-unit: $F(1,62)=12.71$, $p<.001$, $\eta^2=.17$; Guiraud's Index: $F(1,62)=4.40$, $p<.05$, $\eta^2=.01$). L2 learners were more accurate and fluent with respect to all measures of accuracy and fluency (lexical errors: $F(1,62)=35.60$, $p<.001$, $\eta^2=.37$; morphosyntactic errors: $F(1,62)=31.49$, $p<.001$, $\eta^2=.34$; determiner errors: $F(1,62)=31.39$, $p<.001$, $\eta^2=.34$; pruned speech rate: $F(1,62)=8.07$, $p<.01$, $\eta^2=.12$; unpruned speech rate: $F(1,62)=10.28$, $p<.01$, $\eta^2=.14$; pausing: $F(1,62)=20.98$, $p<.001$, $\eta^2=.25$; self-repairs: $F(1,62)=26.96$, $p<.001$, $\eta^2=.30$).

L1 speakers showed similar effects of the factor '± monologic' but not all measures were affected and effect sizes were mostly smaller than for L2 learners. Only the effect of interaction on syntactic complexity was larger for L1 speakers than for L2 learners (words per clause: $F(1,42)=39.18$, $p<.001$, $\eta^2=.48$; subordinate clauses per AS-unit: $F(1,42)=35.31$, $p<.001$, $\eta^2=.45$). Interestingly, the native speakers' lexical complexity decreased in dialogues (Guiraud's Index: $F(1,42)=6.81$, $p<.05$, $\eta^2=.14$). Significant results appeared on two accuracy and two fluency measures, respectively, showing that participants were more accurate and fluent in dialogues than monologues (lexical errors: $F(1,42)=14.43$, $p<.001$, $\eta^2=.25$; determiner errors: $F(1,42)=4.25$, $p<.05$, $\eta^2=.09$; pruned speech rate: $F(1,42)=4.24$, $p<.05$, $\eta^2=.09$; number of pauses: $F(1,42)=39.03$, $p<.001$, $\eta^2=.48$).

Post-hoc pairwise comparisons between L2 learners and L1 speakers regarding complexity measures revealed that the two syntactic measures were significantly influenced in the same direction in both groups while the lexical measure was significantly affected in opposite directions (words per clause: $F(1,104)=4. 46$, $p<.05$, $\eta^2=.04$; subordinate clauses per AS-unit: $F(1,104)=18.80$, $p<.001$, $\eta^2=.15$; Guiraud's Index: $F(1,104)=10.41$, $p<.01$, $\eta^2=.09$)

Finally, the L2 learners did not show any combined effects of the factors '± elements' and '± monologic'. The L1 speakers, however, yielded an interaction effect of moderate size regarding fluency, i.e., task complexity had no effect in monologues, but in dialogues a more complex task slowed them down (pruned speech rate: $F(1,42)=4.56$, $p<.05$, $\eta^2=.10$; unpruned speech rate: $F(1,42)=4.50$, $p<.05$, $\eta^2=.10$).

In sum, study 2a yielded a confirming effect for hypothesis 1 regarding the factor '± elements' on one out of ten measures. That is, the task with more elements promoted a higher lexical complexity. No other significant effects of an increased number of elements emerged from this study. Before discussing this outcome in light of the Cognition Hypothesis and the findings regarding the manipulation of interaction ('± monologic') in relation to hypotheses 2 and 3, the next section will review a discussion about measuring L2 performance as an introduction to the method, design, and results of a follow-up study (2b).

Global versus specific measures of performance

Study 2a has investigated task performance by means of global measures of linguistic complexity, accuracy, and fluency (CAF). Yet, as a reaction to earlier critique of the Cognition Hypothesis Robinson and colleagues doubted whether CAF measures are sensitive enough in order to trace differences in cognitive task complexity. They, therefore, call for task specific measures. “Such specific measures should be more sensitive to conception, task complexity, and its linguistic demands than general measures” (Robinson et al. 2009: 550). Especially, if tasks are designed in such a way that the demands can be met by the use of specific linguistic structures, task specific measures presumably complete the picture that global CAF-measures leave open.

Only a few studies so far did explore data with respect to the use of task specific measures in relation to the claims of the Cognition Hypothesis (Cadierno & Robinson 2009; Révész 2009, 2011; Robinson et al. 2009). These investigations consistently report that while the global CAF-measures did not discriminate between the task manipulations, task specific measures did find supporting results for the Cognition Hypothesis. Therefore, the data of study 2 were subject to an investigation of the effects of task complexity on the use of a task specific measure, i.e., the use of conjunctions. The next section will elaborate on this study 2b.

Study 2b: Investigating the use of conjunctions

The data of the second experiment were analyzed focusing on the use of conjunctions, based on combining theoretical assumptions of earlier research. Kuiken and Vedder (2007) state that the increases of task complexity by means of the factor ‘± elements’ almost automatically induce a higher amount of reasoning. Earlier research by Newton and Kennedy (1996) and Spooren and Sanders (2008) has shown that reasoning is often lexically marked by the use of conjunctions. The present investigation, therefore, tested for the use of conjunctions (see Michel in press for a more elaborate rationale of this choice).

The investigation included a set of 28 different conjunctions including Dutch equivalents of e.g., *but*, *because*, *therefore*, *if…then*. The performances of the L2 learners and L1 speakers acting on the simple and complex study and dating task of study 2a were examined for the frequency (the number of conjunctions per 100 words) and occurrence (the number of conjunctions that was used at least once per performance) of these conjunctions. Two separate multivariate ANOVA's tested for (combined) effects of the factors '± elements' and '± monologic' on the use of conjunctions of L2 learners and L1 speakers respectively.

To sum up the results, these analyses showed no effects in either language group, nor did the univariate analyses disentangle any influences on frequency and occurrence (L2 learners: task complexity $F(1,61)=1.48$, $p = .24$, $\eta^2=.05$; interaction $F(1,61)=1.14$, $p=.39$, $\eta^2=.04$; task complexity X interaction $F(1,61)=2.07$, $p=.14$, $\eta^2=.06$; L1 speakers: task complexity $F(1,41)=.35$, $p=.71$, $\eta^2=.02$; interaction $F(1,41)= 2.97$, $p=.06$, $\eta^2=.13$; task complexity X interaction $F(1, 41)=1.14$, $p=.33$, $\eta^2=.07$). Details of this investigation and an elaborate report of the statistical analyses may be found in Michel (in press).

Summary

This paper presents the overall conclusions of three empirical investigations into task-based L2 performance. The aim was to increase our understanding of how manipulations of the factor '± elements' and the factor '± monologic', both on their own as well as in combination, influence oral L2 production. We analyzed monologic and dialogic speech samples on tasks addressing only a few (simple) versus many (complex) elements of 108 L2 learners and of 44 native speakers performing the same tasks under similar conditions. The analysis evaluated global measures of complexity, accuracy, and fluency as well as a task specific measure, i.e., the use of conjunctions. Before discussing the outcomes in relation to the theoretical background Table 4 summarizes the results of the present studies graphically.

Table 4 about here

General discussion

This section will discuss the findings of the data presented here in relation to the three hypotheses regarding effects of the factor '± elements' (hypothesis 1), the factor '± monologic' (hypothesis 2), and their combined effects (hypothesis 3). It first addresses the implications regarding manipulations of cognitive task complexity, in particular, in relation to the claims of the Cognition Hypothesis (Robinson 2001, 2005; Robinson & Gilabert 2007), which formed the theoretical starting point of the present studies. Second, it discusses the findings on the manipulation of interaction by taking a cognitive perspective referring to psycholinguistic work on alignment (Costa et al. 2008; Pickering & Garrod 2004), priming (Meyer & Schvaneveldt 1971), and task-based research by Tavakoli and Foster (2008). Finally, we highlight the benefits of having a native speaker baseline.

The factor '± elements'

The studies presented in this paper failed at finding confirmatory results for hypothesis 1 that was based on the Cognition Hypothesis (Robinson 2001, 2005). This section will evaluate four possible explanations for these findings.

First, the studies presented here aimed at singling out the factor '± elements'. Therefore, it manipulated the number of elements in a straightforward way: a simple task included a few elements (two elements in study 1; four combinations in study 2) and by adding some more elements a complex version of the same task concerned many elements (six elements in study 1; nine combinations in study 2).

The data suggest, however, that increasing the number of elements in this way did not have the expected effect on cognitive processes. Manipulating the factor '± elements' yielded hardly any significant performance differences. Possibly, just adding more elements did not place "greater functional or conceptual communicative demands" (Robinson & Gilabert 2007: 162) on the learner.

A reason may be that, contrary to expectations, task performers did not evaluate and process all the options given by the task at the same time. That is, when dealing with many elements in the complex tasks at hand an effective strategy could be to consider one element after the other. Such a linear approach to more elements most likely would not affect the cognitive load of a task. Only if the greater number of elements are all taken into account at the same time this presumably produces a higher cognitive load.

This possibly also explains the differences between study 1 and 2. Participants did receive two minutes planning time in the second but not in the first study. In study 2 recordings started after they had made their decision. When talking participants did not show effects of increased processing demands as they addressed one element after another. This manifested itself also in longer speech productions in the 'complex' task. In other words, if talking about nine rather than four options to combine people into pairs, one may just speak more in order to review all possible combinations. The longer speech samples then would be characterized by similar, recurring linguistic structures and forms rather than showing different linguistic means than when talking about a few elements.

This is exactly what we find in the data: the complex tasks generated more speech than the simple tasks. Looking at raw scores suggests that cognitively complex tasks complexify linguistic production. However, using measures of linguistic complexity, accuracy, and fluency or a task specific form that correct for sample length, the differences between the simple and complex task conditions disappear.

Only lexical complexity was significantly raised in both studies when comparing simple to complex tasks. It may be, though, that this is due to the increased number of words in the input of the complex task because the complex tasks inherently had more words (describing the elements) in the input than the simple tasks. Even the data by means of the task specific measure support this explanation. The tasks addressing more elements did not result in participants using substantially more or a different set of conjunctions. It seems that they used the same conjunctions more often.

The studies as a whole, therefore, suggest that participants produced in the complex versions of the tasks *more of the same* language, rather than a linguistically *different* output. Consequently, the factor '± elements' as manipulated in the present studies seems to result in a mere quantitative change of L2 speech production. As the complex tasks did not create a greater cognitive load, L2 learners did not show traces of trying to meet these increased demands, e.g., by means of parallel increases of linguistic complexity and accuracy.

Results of the questionnaire on affective variables that was included in the second study confirm this interpretation. Participants did not perceive the complex tasks to be more difficult than the simple tasks. The average ratings of task difficulty on the simple and complex tasks are 3.6 (SD=0.6) and 3.5 (SD=0.7), respectively in the monologic condition and 3.9 (simple SD=0.6; complex SD=0.5) for both tasks in the dialogic condition (indicated on a 5 point Likert scale from 1=easy to 5=difficult).

The second point of discussion is closely related to this conclusion: How do we know that a task manipulation, e.g., by means of the factor '± elements', indeed induces a higher cognitive load than its simple counterpart? Robinson (2005) explains how changes on different task design factors of this taxonomy influence the cognitive complexity of L2 tasks. He defines cognitive task complexity as the amount of cognitive processing that is needed in order to perform successfully on a task. As such it is dependent on task inherent features and characteristics that increase or decrease the mental effort needed, which in turn affects task performance as established by, for example, CAF-measures. Norris and Ortega (2009) point out the circularity in this definition: How can we objectively determine the cognitive load of a task, the independent variable, if it is defined by successful task performance, the dependent variable?

The operationalization of cognitive task complexity in this paper is based on the Triadic Componential Framework (see Table 1). As such the framework functions as a research agenda for empirical investigations. But indeed, how do we know its assumptions are correct? Future research may aim at having some external means that confirm the theoretically assumed cognitive load even if the foremost goal is to test its effects.

A way to have at least an external point of reference would be to have a native speaker baseline as was done in study 2 of the present paper. Native speakers can rely on mostly automatic processes of language production (Levelt 1989). Therefore, L1 speakers were expected to show hardly any effects of cognitive task complexity. Results of the present studies confirmed this prediction. L1 speakers showed an effect on lexical complexity only but no other measures were significantly influenced. The conclusion for native speakers would be that they were hardly affected by the demands more elements put up in the complex task because their speech processing is highly automatic. Yet, also the L2 learners show hardly any influence on their performance. The similarities between the two populations, therefore, point towards the earlier assumption that also for L2 learners the factor '± elements' as manipulated in the present studies did not directly affect the cognitive load of the task.

Still, this factor affects task performance: It generates more and lexically more complex speech in L2 learners and L1 speakers. Consequently, the factor '± elements' as manipulated in the present studies (i.e., adding some more elements) may be seen as a factor of general task design that pushes a greater amount of speech. But it may not be a resource-directing factor of cognitive task complexity.

Third, it is interesting to see that earlier work investigating the factor '± elements' did find corroborating results for the Cognition Hypothesis. For example, Gilabert (2007) and Gilabert et al. (2009) yielded supporting results by manipulating the number of elements in an instruction giving task. Similarly, Révész (2011), found corroborating results for Robinson's claims when she increased the number of elements in a decision making task. Finally, Kuiken and colleagues (2005,

2007) asked participants to base a decision on a different number of criteria and their data partially confirm Robinson (2005).

As the short review of this earlier work shows most studies use different ways to manipulate the factor '± elements'. Our study used yet another way to operationalize the number of elements. As our data failed to find influences this manipulation on the cognitive processes during task-based performance, it may be that we were unsuccessful in manipulating the factor '± elements' as intended by the Triadic Componential Framework (Robinson 2005). Unfortunately, the literature does not give clear instructions, about how to manipulate this factor. In light of the present data and in relation to the discussion above, future work may aim at following example manipulations of the cited studies rather than the ones reported here.

A last possible explanation needs to be evaluated here: Could it be that the data are in line with the Limited Attentional Capacity Model, which assumes trade-off effects in particular between linguistic complexity and accuracy (Skehan 1996; Skehan & Foster 2001). However, study 1 shows a minor parallel increase of accuracy and complexity while study 2a revealed neither an increase nor a decrease on any accuracy measure. Consequently, the present studies do not support Skehan and Foster's (2001) account either.

To sum up, concerning the factor '± elements' the present studies call for guidelines of how exactly factors in the Triadic Componential Framework (Robinson 2005) should be operationalized. This would be a welcome addition to the model.

The factor '± monologic'

The summary of results (Table 4) shows consistent effects of the factor '± monologic'. Throughout study 1 and 2 interactive tasks raised the accuracy and fluency of L2 oral task performance and decreased structural complexity. Lexical complexity was not affected in the first study but was pushed by a dialogic condition in the second study. Our data, therefore, confirm hypothesis 2 and give us important insights on how changes on the factor '± monologic' affect task-based L2 performance. The following paragraphs will discuss the three CAF-dimensions of task performance separately by using the L1 baseline data to further explain the findings. [2]

In contrast to what one may assume based on joint focus on form in dialogues, interactive tasks resulted in a lower structural complexity than monologues. The fact that L1 speakers show a similar behavior supports the assumption that interactive tasks presumably prevent speakers from building complex syntactic structures (cf. Robinson 2001). In other words, to produce structurally less complex language in dialogues than monologues is what native speakers do. Therefore, this can be characterized as target language use (Pallotti 2009). However, whether this indeed is the reason for the lower structural complexity found in the present data cannot be answered because the analysis did not include counts of, e.g., comprehension checks and clarification requests. Yet, as the average number of AS-units in general was higher in dialogues than monologues this interpretation receives some further support in the data.

Interestingly, lexical complexity was affected differently in L2 learners than in L1 speakers. Processes of alignment and priming can account for these contrasting results though (Costa et al. 2008). It may be that due to copying and mirroring of words L2 interactants could profit from each other's lexical input. In the monologic condition, participants had to rely on their own lexical knowledge while in the dialogic condition speakers could incorporate words of the interlocutor they would not have come up with on their own. Possibly, this kind of priming resulted in the elaborate lexicon we find in L2 dialogues. In contrast, the L1 data show a low lexical complexity in dialogic speech. As explained by Pickering and Garrod (2004) the L1 lexicon is very large by itself. In dialogues routinization and recycling of vocabulary items decreases rather than increases the use of different lexical items in L1 speakers.

Assuming alignment and priming processes can also account for the gains in L2 accuracy from monologic to dialogic tasks. As these processes possibly freed cognitive resources, participants had more attention to focus on form which presumably promoted accuracy in L2 learners and L1 speakers. As native speakers showed a gain in accuracy especially with respect to lexical choices, this explanation receives even more support. Also, the fact that dialogues generated more fluent speech than monologues (in both populations) is in line with these assumptions. Finally, this interpretation is supported by scores on the questionnaire on affective variables where participants rated dialogic tasks to be generally easier than monologic tasks.

As a whole, the data on monologic versus dialogic performances in L2 learners and L1 speakers seem to be in line with the theory that dialogues generate a lower cognitive complexity than monologues (Costa et al. 2008; Pickering & Garrod 2004; Tavakoli & Foster 2008). The present work, therefore, proposes that the factor '± monologic' has a cognitive dimension, which draws on processes of alignment and priming and may be related to the factor '± planning time', that is, a resource-dispersing variable of cognitive task complexity in terms of Robinson's Triadic Componential Framework (2005).

Future work may try to test this assumption by aiming at independent ways of measuring cognitive complexity during monologic and dialogic tasks.

Combined effects of the factors '± elements' and '± monologic'

As summarized in Table 4 there were no combined effects of cognitive task complexity and interaction supporting Robinson's theory. Therefore, our third hypothesis is rejected. In light of the discussion of the factor '± elements' this is not surprising. How could almost no effects (of the cognitive factor of task complexity) create combined effects with the factor interaction?

Based on the discussion about the cognitive aspects of the factor interaction it may be, though, that factors of cognitive task complexity can generate positive synergies with interaction. That is, if interaction reduces task complexity as a resource-dispersing variable and thereby eases the cognitive load during task performances, it makes more attentional resources available. It may be that during interaction more attention can be focused on task relevant linguistic forms by increasing the cognitive task complexity along resource-directing dimensions. Future research may reveal whether such synergies exist.

The benefits of a native speaker baseline

Not least, the present studies show how valuable it is to include a native speaker baseline in research into task-based L2 performance. At all times, the L2 performance could be interpreted in light of task specific performances by L1 speakers. This presumably has more confirmatory strength than evaluating L2 speech by means of an external standard (which is often based on prescriptive written norms). For example, the similarities between the two populations with respect to effects of the factor '± elements' led to the assumption that the way this factor was manipulated may not directly affect the cognitive load of a task. Without the control group, such insights would be hard to defend. Also the disparities between native and non-native speaking performances concerning lexical complexity in L2 and L1 interaction served as an explanation for the effect of the factor '± monologic' in L2 learners. As such, the L1 baseline corroborated tentative explanations of the L2 data and helped understanding the L2 learner results.

Summary and conclusion

The studies presented in this paper reveal hardly any changes in task performance due to a higher cognitive task complexity induced by the manipulation of the factor '± elements' in oral task-based L2 performance (Robinson 2005). However, as the discussion set out, the absence of effects is most likely related to the way the number of elements was operationalized in the present work. It seems that the present manipulation resulted in more but not different speech.

In contrast, the present studies yielded interesting data concerning the factor '± monologic'. Interactive tasks lead L2 learners to increase their lexical complexity, accuracy, and fluency, while monologic tasks yield more complex syntactic structures. With respect to cognitive factors of task complexity as conceptualized in Robinson's Triadic Componential Framework (Table 1) and the Cognition Hypothesis the findings point to some changes regarding the factor '± elements' and the factor '± monologic'. After all, the data suggest that both factors as manipulated in the present studies address additional aspects than expressed in the original framework Robinson (2005). Table 5 displays the proposed additions.

Table 5 about here

First, the factor '± elements' as manipulated here may be characterized as a task design factor that affects the amount of speech while attentional allocation remains unaffected. Adding more elements to a simple task presumably has the result that performers produce more speech. Accordingly, this may create the single CAF-effect of yielding a higher lexical complexity due to the production of more recurring but similar structures and forms. No other effects of this straightforward increase in the number of elements may be found.

Second, the present studies propose that the factor '± monologic' affects L2 performance in at least two ways. (a) dialogic speech allows for frequent interactional moves, which induce a lower structural complexity while the joint attention to language presumably promotes lexical complexity and accuracy, (b) interaction has a cognitive side and as such may be related to the resource-dispersing factor of task complexity '± planning time' because it creates natural pauses for speakers during the turn of the interlocutor (Tavakoli & Foster 2008). Processes of priming and alignment may further decrease the cognitive load in dialogic tasks (Costa et al. 2008; Pickering & Garrod 2004). In sum, dialogues are expected to free attentional capacity, which pushes accuracy and fluency while alignment increases lexical complexity. In contrast, it seems that the turn-taking behavior reduces structural complexity. Importantly, both aspects of the factor '± monologic' are predicted to affect speaking performance.

Finally, it is questionable whether these additions to Robinson's Triadic Componential Framework make the model more feasible. Earlier critique by Kuiken and Vedder (2007) and Ellis (2009) point out that the existing proposals of the framework (e.g., Robinson 2005, 2007) create a large variety of different manipulations and research designs, so that it may be difficult to find consistent results. The present studies possibly support their call for more precise explanations how exactly researchers could operationalize and weight the different factors named in the Triadic Componential Framework. Even so, future research, which uses the Triadic Componential Framework as a research agenda, can benefit from the findings of the present paper concerning the factors '± elements' and '± monologic', i.e., concerning the proposed cognitive impact of interaction.

[1] The authors wish to thank the editor and two anonymous reviewers who gave very valuable feedback on an earlier version of this paper.
[2] As addressed by a reviewer, it is difficult to judge CAF effects in individual performances in a dialogue. This interesting question goes beyond the scope of this paper but future research may investigate it.

References

Cadierno, T. and Robinson, P. 2009. "Language typology, task complexity and the acquisition of L2 lexicalization patterns for reference to motion". Annual Review of Cognitive Linguistics 7(1): 245–276.

Costa, A., Pickering, M.J. and Sorace, A. 2008. "Alignment in second language dialogue". Language and Cognitive Processes 23(4): 528–556.

van Daele, S., Housen, A. Kuiken, F. Pierrard, M. and Vedder, I. (eds). 2007. Complexity, Accuracy, and Fluency in Second Language Use, Learning, and Teaching. Brussel: KVAB.

Ellis, R. 2009. “The differential effects of three types of task planning on fluency, complexity, and accuracy in L2 oral production”. Applied Linguistics 30: 474–509.

Ellis, R. and He, X. 1999. “The roles of modified input and output in the incidental acquisition of word meanings”. Studies in Second Language Acquisition 21(2): 285–301.

Foster, P., Tonkyn, A. and Wigglesworth, G. 2000. “Measuring spoken language: A unit for all reasons”. Applied Linguistics 21(3): 354–375.

Gass, S.M. and Varonis, E. 1994. “Input, interaction, and second language production”. Studies in Second Language Acquisition 16(3): 283–302.

Gilabert, R. 2005. Task Complexity and L2 Narrative Oral Production. Unpublished Ph.D.-thesis, Universitat de Barcelona, Spain.

Gilabert, R. 2007. “Effects of manipulating task complexity on self-repairs during L2 oral production”. International Review of Applied Linguistics 45(3): 215–240.

Gilabert, R., Baron, J. and Llanes, A. 2009. “Manipulating cognitive complexity across task types and its impact on learners’ interaction during oral performance”. International Review of Applied Linguistics 47(3–4): 367–395.

Gilabert, R., Baron, J. and Levkina, M. 2011. “Manipulating task complexity across task types and modes”. In Second Language Task Complexity. Researching the Cognition Hypothesis of language learning and performance, P. Robinson (ed.), 105–138. Amsterdam and Philadelphia: John Benjamins.

Givón, T. 1995. Functionalism and Grammar. Amsterdam and Philadelphia: John Benjamins.

Halford, G.S., Cowan, N. and Andrews, G. 2007. “Separating cognitive capacity from knowledge: A new hypothesis”. Trends in Cognitive Sciences 11(6): 236–242.

Housen, A. and Kuiken, F. 2009. “Complexity, accuracy, and fluency in second language acquisition”. Applied Linguistics 30: 461–473.

Ishikawa, T. 2007. “The effect of manipulating task complexity along the Here-and-Now-dimension on L2 written narrative discourse”. In Investigating tasks in formal language learning, M.P. García Mayo (ed.),136–156. Clevedon: Multilingual Matters.

Kormos, J. 2000. “The role of attention in monitoring second language speech production”. Language Learning 50(2): 343–384.

Kuiken, F., Mos, M. and Vedder, I. 2005. “Cognitive task complexity and second language writing performance”. In EUROSLA Yearbook, S. Foster-Cohen, M.P. García Mayo, and J. Cenoz (eds.), 195–222. Amsterdam and Philadelphia: John Benjamins.

Kuiken, F. and I. Vedder 2007. “Task complexity and measures of linguistic performance in L2 writing”. International Review of Applied Linguistics 45(3): 261–284.

Levelt, W.J. 1989. Speaking: From Intention to Articulation. Cambridge, MA: MIT Press.

Long, M.H. 1990. “Task, group, and task-group interactions”. In Language Teaching Methodology for the Nineties [Anthology series 24]. A. Sarinee (ed.),.

Long, M.H. and Robinson, P. 1998. “Focus on form: Theory, research, and practice”. In Focus on Form in Classroom Second Language Acquisition, C.J. Doughty and J. Williams (eds.), 15–41. New York: Cambridge University Press.

Mackey, A. 1999. “Input, interaction and second language development”. Studies in Second Language Acquisition 21(4): 557–587.

MacWhinney, B. 2000. The CHILDES Project: Tools for Analyzing Talk. Computer Program. Retrieved from http://childes.psy.cmu.edu/.

Meyer, D.M. and R. Schvaneveldt 1971. “Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations”. Journal of Experimental Psychology 90: 227–234.

Michel, M.C. In press. “Effects of Task Complexity on the Use of Conjunctions in L2 Task Performance”. The Modern Language Journal.

Michel, M.C. 2011. “Effects of Task Complexity and Interaction on L2-Performance”. In Second Language Task Complexity. Researching the Cognition Hypothesis of language learning and performance, P. Robinson (ed.), 141–174. Amsterdam and Philadelphia: John Benjamins.

Michel, M.C., Kuiken, F. and Vedder, I. 2007. “Effects of Task Complexity and Task Condition on Dutch L2”. International Review of Applied Linguistics 45(3): 241-259.

Newton, J. and Kennedy, G. 1996. “Effects of communication tasks on the grammatical relations marked by second language learners”. System 24(3): 309–322.

Norris, J. and Ortega, L. 2009. “Measurement for understanding: An organic approach to investigating complexity, accuracy, and fluency in SLA”. Applied Linguistics 30: 555–578.

Nuevo, A.M. 2006. Task Complexity and Interaction. Unpublished Ph.D.-thesis, Georgetown University, Washington D.C., USA.

Pallotti, G. 2009. “CAF: Defining, refining and differentiating constructs”. Applied Linguistics 30: 590–601.

Pica, T. 1994. “Research on negotiation: What does it reveal about second language learning conditions, processes, and outcomes?” Language Learning 4(4): 493–527.

Pickering, M.J. and Garrod, S. 2004. “Toward a mechanistic psychology of dialogue”. Behavioral and Brain Sciences 27(2): 169–226.

Révész, A. 2009. “Task complexity, focus on form, and second language development”. Studies in Second Language Acquisition 31(3): 437–470.

Révész, A. 2011. “Task complexity, focus on L2 constructions, and individual differences: A classroom based study”. The Modern Language Journal 95(4).

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

Robinson, P. 2005. “Cognitive complexity and task sequencing: Studies in a componential framework for second language task design”. International Review of Applied Linguistics 43(1): 1–32.

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(3): 193–213.

Robinson, P., Cadierno, T. and Shirai, Y. 2009. “Mind, time, and motion: Measuring the effects of the conceptual demands of tasks on second language speech production”. Applied Linguistics 30: 533–554.

Robinson, P. and Gilabert, R. 2007. “Task complexity, the Cognition Hypothesis and second language learning and performance”. International Review of Applied Linguistics 45(3): 161–176.

Rost, M. 2011. Teaching and Researching Listening. Applied Linguistics in Action. Harlow, UK: Pearson Education.

Schmidt, R. 2001. “Attention”. In Cognition and Second Language Learning, P. Robinson (ed.), 3–32. New York: Cambridge University Press.

Shehadeh, A. 2004. “Modified output during task-based pair interaction and group interaction”. Journal of Applied Linguistics 1(3), 351–385.

Skehan, P. 1996. “Second language acquisition research and task-based instruction”. In Challenge and Change in Language Teaching, J. Willis and D. Willis (eds.), 17–30. Oxford: Heinemann.

Skehan, P. 2003. “Task-based instruction”. Language Teaching 36(1): 1–14.

Skehan, P. 2009. “Modeling second language performance: Integrating complexity, accuracy, fluency, and lexis”. Applied Linguistics 30, 510–532.

Skehan, P. and P. Foster 2001. “Cognition and tasks”. In Cognition and Second Language Learning, P. Robinson (ed.), 183–205. New York: Cambridge University Press.

Skehan, P. and Foster, P. 2007. “Complexity, accuracy, fluency, and lexis in task-based performance: A meta-analysis of the Ealing Research”. In Complexity, Accuracy, and Fluency in Second Language Use, Learning, and Teaching, S. van Daele, A. Housen, F. Kuiken, M. Pierrard, and I. Vedder (eds.), 207–226. Brussel: KVAB.

Spooren, W. and Sanders, T. 2008. “The acquisition order of coherence relations: On cognitive complexity in discourse”. Journal of Pragmatics 40(12): 2003–2026.

Tavakoli, P. and Foster, P. 2008. “Task design and L2 performance”. Language Learning 58(2): 429–473.

Tavakoli, P. and Skehan, P. 2005. “Strategic planning, task structure, and performance testing”. In Planning and Task Performance in a Second Language, R. Ellis (ed.), 239–273. Amsterdam and Philadelphia: John Benjamins.

VanPatten, B. 1990. “Attention to form and content in the input: An experiment in consciousness”. Studies in Second Language Acquisition 12(3): 287–301.

Wickens, C.D. 2007. “Attention to the second language”. International Review of Applied Linguistics 45(3): 177–191.

Table 1: Cognitive factors of task complexity, adapted from (Robinson 2005)

Task complexity: cognitive factors
(a) resource-directing variables
± here and now
± elements
± reasoning demands
(b) resource-dispersing variables
± planning time
± prior knowledge
± single task

Table 2: Predicted effects of task complexity and interaction

	Performance measure	*Task complexity*	*Interaction*	*Task complexity x interaction*
L2 learners	Linguistic Complexity	↑	↓ ? ↑	↓ ? ↑
	Accuracy	↑	↑	↑ ↑
	Fluency	↓	↑	↓ ? ↑

Note: ↑ = increase; ↓ = decrease; ↑ ↑ = large increase; L2 = second language

Table 3: Background information for all participants

		Age	*Gender*		*Proficiency**	*Stay in Nether-lands***	*Origin*	
	N	Mn (SD)	m	f	Mn (SD)	Mn (SD)	Tur	Mor
Study 1								
L2 learners	44	27.7 (6.4)	17	27	21.5 (9.3)	3.8 (4.5)	15	29
Study 2								
L2 learners	64	27.6 (6.2)	29	35	53.8 (17.2)	3.8 (4.2)	31	33
L1 speakers	44	20.6 (3.5)	9	35	96.3 (3.2)			

Note: N = number of participants; Mn (SD) = Mean (Standard Deviation); m = male, f = female; * = score out of 50 in study 1, score out of 100 in study 2; ** = in years; Tur = Turkish; Mor = Moroccan

Table 4: Summary of results of study 1, 2a, and 2b

Second language learners		*Task complexity*	*Interaction*	*Task complexity X interaction*
Study 1	str. comp.*	≈	↓	≈
Global CAF	lex. comp.	↑ (√)	≈	≈
	accuracy	↑ (√)	↑ √	≈
	fluency	↑ (√)	↑ √	≈
Study 2a	str. comp.*	≈	↓	≈
Global CAF	lex. comp.	↑ √	↑	≈
	accuracy	≈	↑ √	≈
	fluency	≈	↑ √	≈
Study 2b *task specific*	frequency occurrence	≈	≈	≈
*Hypothesis***		1, not confirmed	2 confirmed	3, not confirmed
Native speakers				
Study 2a	str. comp.*	≈	↓	≈
Global CAF	lex. comp.	↑	↓	≈
	accuracy	≈	↑	≈
	fluency	≈	↑	≈
Study 2b *task specific*	frequency occurrence	≈	≈	≈

Note: √ = confirmed prediction; (√) = partially confirmed prediction; CAF = measure of linguistic complexity, accuracy, and fluency; str. comp. = structural complexity; lex. comp. = lexical complexity; task specific = use of conjunctions; *The predictions did not make a difference between structural and lexical linguistic complexity. As the results sometimes revealed different patterns they are listed separately in this table. ** A hypothesis is marked as confirmed if there are more measures

(partially) affected in the predicted direction than measures that are not affected or affected in the opposite direction.

Table 5: Summarizing the effects of the factors '± elements' and '± monologic' as manipulated in the present studies

Factors	± *elements*	± *monologic*	
CAF effects	≈ structural complexity + lexical complexity ≈ accuracy ≈ fluency	- structural complexity + lexical complexity + accuracy + fluency	
EXPLANATIONS	*general* • more speech • more lexical items • similar recurring linguistic structures • hardly affecting attention	*interactional* • interactional moves • interruptions and clarifications • negotiation of meaning and form • joint attention to language	*cognitive* • time for online planning during interlocutor's turn • priming and alignment on all linguistic levels • freed attentional capacity • related to resource-dispersing cognitive factor of task complexity

Note: CAF = measure of linguistic complexity, accuracy, and fluency