

PRE-TASK PLANNING, WORKING MEMORY CAPACITY, AND L2 SPEECH PERFORMANCE

Maria da Glória Guará Tavares

Resumo: *As pesquisas acerca do planejamento pré-tarefa apontam para efeitos de troca atencional entre os aspectos da fluência, complexidade e acurácia do desempenho oral em L2 diante da limitação dos recursos atencionais dos aprendizes (Foster; Skehan, 1996; Menhert, 1998). O presente estudo¹ investigou a relação entre planejamento pré-tarefa, capacidade de memória de trabalho e desempenho oral em L2. Os resultados indicam que (1) o planejamento leva a ganhos significativos em acurácia e complexidade e (2) há diferenças significativas entre o desempenho oral de indivíduos com maior e menor capacidade de memória de trabalho em termos de fluência e complexidade. Os resultados são discutidos com base em Engle, Kane, e Tuholski (1999); e também com base em (Foster; Skehan, 1996; Yuan; Ellis, 2003; e Guará-Tavares, 2009)*

Palavras-chave: *memória de trabalho, planejamento pré-tarefa, desempenho oral, fluência, acurácia, complexidade.*

¹ A presente pesquisa foi conduzida com o apoio financeiro de três anos do CNPq na Universidade Federal de Santa Catarina e de quatro meses de apoio financeiro da CAPES através da bolsa de doutorado sanduíche que me permitiu analisar dados na Universidade de Auckland, NZ.

INTRODUCTION

According to Ellis (2005), there has been a substantial body of research on tasks over the last decades. Within the study of tasks, one construct which has attracted considerable attention is planning. Planning seems to have evolved into an area of inquiry in its own right and “has become a burgeoning area of investigation within task-based learning” (ORTEGA, 2005, p. 77).

Researchers have investigated planning from a variety of perspectives, and, in general, studies have shown a positive impact of planning on L2 performance. Several studies have shown that planning leads to gains in fluency (FOSTER; SKEHAN, 1996; MEHNERT, 1998; ORTEGA, 1999). Planning also leads to gains in accuracy, although results have been more mixed in this respect (ELLIS, 1987; MEHNERT, 1998; ORTEGA, 1999; FOSTER; SKEHAN, 1999). Finally, studies have also shown that planning enhances complexity (CROOKES, 1989; FOSTER; SKEHAN, 1996; MEHNERT, 1998; ORTEGA, 1999; YUAN; ELLIS, 2003).

One interesting finding of the studies on the impact of planning on L2 performance is the evidence of attentional trade-off effects among the goals of fluency, accuracy, and complexity. Foster and Skehan (1996), Menhert (1998), as well as Yuan and Ellis (2003) discuss results of their studies in terms of an attentional model of learning and performance. These researchers propose that there are trade-off effects among the goals of fluency, accuracy, and complexity in the context of the use of learners’ limited capacity attentional resources. In other words, because attentional resources are limited, planning benefits cannot be achieved to the same extent simultaneously for fluency, accuracy, and complexity of L2 performance. The trend of research results shows that there are gains in fluency and complexity at the expense of gains in accuracy.

Despite the fact that researchers in task-based planning (e.g., FOSTER; SKEHAN, 1996; MENHERT, 1998; YUAN; ELLIS, 2003) explain results of studies in terms of learners’ limited capacity attentional resources, individual differences in working memory capacity have not been taken into account in any of these studies (e.g., FOSTER; SKEHAN, 1996; MENHERT, 1998; YUAN; ELLIS, 2003) as a feasible variable for affecting learners’ performance under planning conditions.

REVIEW OF THE LITERATURE

Models of L1 and L2 speech production

Levelt's (1989) model of L1 speech production has four specialized components, which underlie speech production: the conceptualizer, the formulator, the articulator, and the speech comprehension system. These components work in a highly automatic way, and automaticity is what allows the components to work in parallel, which is, in turn, "a main condition for the production of uninterrupted speech" Levelt, 1989, p. 2).

De Bot (1992) made a few adaptations to Levelt's (1989) model in order to account for L2 speech production. The first assumption of De Bot's (1992) model is that the speaker has, first of all, to decide what language to speak. This decision takes place in the conceptualizer. As far as the formulator is concerned, De Bot (1992) proposes that it is language-specific, thus, different procedures are applied to the grammatical encoding of L1 and L2 speech. Finally, De Bot (1992) suggests only one articulator for both languages. By assuming only one articulator L1 interferences in L2 can be explained.

Working memory and L2 speech performance

As briefly stated in the introduction, the perspective taken in the present study is that working memory resources are attentional, that is to say, I take Engle's (1999) attention-view perspective of working memory capacity. Engle et al. (1999) view working memory as a cognitive system comprising (a) a store in the form of long-term memory traces active above a threshold, (b) processes for achieving and maintaining this activation, and (c) controlled attention. Nevertheless, when they refer to 'working memory capacity', it is the limited capacity of the element of *controlled attention* that is being referred to. Several studies (DANEMAN, 1991; FORTKAMP, 1999, 2000, just to mention a few) suggest a relationship between working memory capacity and aspects of L1 and L2 speech performance.

Pre-task planning

D'Ely (2004) argues that although planning is essentially a cognitive process which is inherent to the act of speaking, it evolves into a meta-cognitive process when it is used strategically by the learner. Therefore, D'Ely (2004) defines strategic planning as a problem-solving activity

that provides learners the opportunity “to exert some control over what they know towards achieving gains in oral performance” (p.17).

To reiterate, planning is a problem solving activity, and according to Hambrick and Engle (2003), a problem is a goal which is not instantaneously achievable and whose most prominent feature is that although the initial state and the target are clear, how to convert the initial state into the target state is uncertain. Problem solving activities require “the ability to maintain goals, action plans, and other task-relevant information in a highly activated and accessible state, and when necessary, to inhibit activation of irrelevant or distracting information” (HAMBRICK AND ENGLE, 2003 p.179).

Guará-Tavares (2009) is, to the best of my knowledge, the first study to address whether individual differences in working memory capacity mediate the impact of planning of L2 performance. All in all, results are not conclusive and indicate that whenever the task is made more manageable for learners (either due to planning or task familiarity) individual differences in WM capacity tend to emerge.

THE CURRENT STUDY

Based on the preceding review, individual differences in WM capacity seem to be a potential source of individual differences in L2 speech performance. In order to expand the study by Guará-Tavares (2009), the present study pursued two research questions:

- (1) Does pre-task planning opportunity significantly increase fluency, accuracy, and complexity of L2 speech performance?
- (2) Do higher working memory span individuals significantly outperform lower working memory span individuals in terms of fluency, accuracy and complexity of L2 speech performance under pre-task planning condition?

METHOD

Research Design

The study employed a between-subject design, in which participants in the control group completed both first and second narrative tasks

under a no-planning condition, and participants in the experimental group completed the first task under a no-planning and the second task under a planning condition.

Data collection of the present study was divided into three phases. The first phase was the selection of participants which aimed at controlling for proficiency level. Participants performed the proficiency trial task at the language laboratory, and all students of the same class did the task together. Due to participants' time constraints, the task used for selecting participants also served as the first sample of L2 speech performance under no planning condition.

The second phase consisted of the Speaking Span Test to measure participants' working memory capacity. Participants of the experimental and control groups carried out the speaking span test individually with the researcher in a computer lab. The third phase of data collection consisted of the second narrative task. Participants of the control group carried out the second narrative task under the same condition as the first narrative task, that is to say, under a no planning condition. On the other hand, participants of the experimental group carried out the second narrative task under a planning condition.

Context and participants

The participants of the present study were 50 intermediate learners from the Letras² Licenciatura, Letras Secretariado³, and also from the Extracurricular Language Courses at Universidade Federal de Santa Catarina (UFSC). Participants of the Extracurricular Language courses were all undergraduate students at the Universidade Federal de Santa Catarina from a variety of backgrounds. Out of the 50 participants, 30 were female, 20 were male, and their ages ranged between 18 and 29 years. Participation was voluntary and no financial reward was given.

Procedures for the Selection of participants

The selection of participants was conducted using the rating scale proposed by D'Ely and Weissheimer (2004). The level of proficiency chosen for participation in the study was the intermediate level. The choice of the intermediate level was due to two reasons. First, the level

2 Undergraduate Language Teaching and Literature program

3 Undergraduate Bilingual Secretary program

of participants in most planning studies range from pre-intermediate to advanced levels (FOSTER; SKEHAN, 1996; KAUWACHI, 2005; MEHNERT, 1998; ORTEGA, 1999; SANGARUN, 2005). Thus, choosing participants from these levels would allow for comparisons between my study and previous studies on planning.

The first task consisted of a picture-cued narrative. Participants had to look at a set of pictures and organize them in a sequence in order to tell a story. Participants' oral production was recorded and given to three raters who were instructed to evaluate their performance according to the rating scale proposed by D'Ely and Weissheimer (2004). According to this scale, participants are assessed in terms of fluency, accuracy and complexity on a scale from 0 to 5. Raters were all experienced ones who have been trained and have used the scale previously.

The scores given by the raters were submitted to statistical treatment in order to verify the interreliability of the rating procedures. A Cronbach Alpha Coefficient was run in order to find the degree of reliability, the means and the standard deviation of participants' performance. Reliability estimates for the rating procedure were .82, which is considered an acceptable level. The mean of participants' performance was 2.95. In order to be able to select 50 participants at proximal levels of proficiency, 2.54 to 3.5 were selected for the present study. Therefore, out of the 99 students who took part in the proficiency trial, 55 were selected for the present study and 50 actually took part in the study.

After selecting the 50 participants of the present study according to their level of proficiency, they were randomly divided into a *control group* and an *experimental group*.

The Speaking Span Test

A version of Daneman and Green's (1986) Speaking Span Test was used. A training phase (20 words) preceded the testing phase (60 words). The test contained 60 unrelated words organized in three sets each of two, three, four, five and six words.

Each word was presented individually, on the middle line of a computer screen for one second. Participants were instructed to read each word aloud. At the end of each set, question marks appeared. These marks signaled the number of words that had to be stored and the number of sentences to be produced. Participants were instructed to

use the words in the exact form and order they appeared to generate syntactically and semantically acceptable sentences, aloud, in English.

There were no restrictions concerning the length or complexity of the sentences produced. For instance, after being presented a set of three words: *guy - point - train*, a participant produced the following sentences: “I am a guy”, “what’s your point?”; “The train was dirty”. Participants’ speaking span score was defined as the maximum number of words for which they could generate grammatically and semantically acceptable sentences in English.

Following Daneman (1991), in this study, participants’ responses, which were recorded, transcribed and analyzed, generated two different speaking span scores: a speaking span *strict score*, when all the sentences the subject produced contained the target word in the exact form and order of presentation; and a speaking span *lenient score*, when credit was given for sentences that contained the target word in a form other than that of presentation (e.g., target word being ‘guy’ and the word in the sentence produced being ‘guys’), and half credit was given to words recalled in a different order. No credit was given to ungrammatical sentences in terms of syntax and semantics.

The speech generation tasks: ‘there and then narratives’

The two tasks used were both ‘there-and-then’ picture cued narratives (ROBINSON, 1995). In both tasks, participants had 50 seconds to look at the set of pictures and then put pictures away. Narrative tasks have been widely used in previous studies on task-based planning (D’ELY, 2006; ELLIS, 1987; ELLIS & YUAN, 2005; FOSTER; SKEHAN, 1996; KAWAUCHI, 2005; ORTEGA, 1999) thus, allowing for comparison between the present study and previous ones in the field. The order of tasks was counterbalanced among participants for the purpose of controlling practice effects.

Participants of the control group carried out both tasks under a no-planning condition. They were instructed to start telling their stories immediately after looking at the pictures for 50 seconds. On the other hand, participants of the experimental group were instructed to plan the second task for 10 minutes after looking at the pictures for 50 seconds. The pictures were removed from them before they started planning their task. The choice of 50 seconds for looking at the set of pictures

aimed at minimizing planning as much as possible in the no-planning condition. According to Mehnert (1998), one minute planning may be enough for gains in accuracy to take place. Thus, participants were given less than one minute to look at the set of pictures.

Measures of L2 speech performance

The speech samples were analyzed in terms of fluency, accuracy, and complexity. These measures have been extensively used in studies investigating the effects of planning on L2 speech performance (FOSTER; SKEHAN, 1996; MEHNERT, 1998; ORTEGA, 1999; YUAN; ELLIS, 2003; among others).

Fluency - as in Fortkamp (2000), speech rate unpruned was calculated by dividing the total number of semantic units (complete and partial words) produced by the total time in seconds (including pause time), the resulting figure was multiplied by 60 to express the number of semantic units per minute; speech rate pruned was calculated in the same way but excluding: (a) the words that were abandoned before completion, and (b) words that were immediately repeated (except words repeated for rhetorical purposes).

Number of silent pauses per c-unit was calculated by dividing the number of silent pauses in each subject's speech sample by the number of c-units, as in D'Ely (2006). Following Foster and Skehan (1996), and Mehnert (1998), a cut-off point of 1 second was considered optimal in determining silent pauses in L2 speech samples. Pauses were identified and measured using the computer software PRAAT[®] version 4.6.06.⁴ Percentage of total silent pausing time was calculated by dividing the total silent pausing time by the total time participants took to complete the task, the resulting figure was multiplied by 100 (D'ELY, 2006; FOSTER; SKEHAN, 1996).

Accuracy - according to Skehan (1996, 1998), accuracy concerns form in the sense of error-free performance. It was operationalized in terms of number of errors per a hundred words and percentage of error-free clauses. Number of errors⁵ per a hundred words was calculated by dividing participants' total number of errors by the

⁴ This software allows the identification of the precise location and length of speech pauses.

⁵ The criteria for defining errors was based on American English norms since this is the English variety adopted by the textbooks used by the participants.

total number of words produced and multiplying the result by 100 (FORTKAMP, 2000; MEHNERT, 1998). Percentage of error-free clauses was calculated by identifying the number of error-free clauses, which was then divided by the total number of clauses produced, and the resulting figure was multiplied by 100 (FOSTER; SKEHAN, 1996; MEHNERT, 1998).

Complexity - according to Foster and Skehan (1996), subordination is considered a satisfactory measure to assess complexity, which was measured by an index of subordination reflected by the number of clauses per c-unit. It was calculated by dividing the total number of clauses (dependent and independent) by the total number of c-units. The higher the index of subordination obtained the higher the complexity of the speech was.

Data Analysis Procedures

Data was submitted to statistical treatment. The first step was to carry out descriptive statistics analyses in order to give an overview of the seven variables of speech production, and lenient and strict working memory scores. Next, the normal distribution of each group on all variables was tested by examining skewness, kurtosis, and the results of the Kolmogorov-Smirnov test of normality.

In order to verify whether planning led to significant differences in the performance of the experimental group when compared to the control group, one-way ANOVAs were used for this purpose. First, ANOVAs were run to compare the control and experimental groups in terms of a) performance of the *first* task, and (b) lenient and strict scores on the SST. These procedures were followed to verify whether the groups were nearly homogenous in a) performance of Task 1 performance, and b) working memory capacity scores. In case groups were nearly homogeneous, any group differences in the performance of the *second* task could be attributed to pre-task planning, not to group differences in terms of speech performance a priori, or to group differences in terms of working memory capacity scores.

Then, ANOVAs were run to compare the control and experimental groups in the second task in order to check whether planning led to significant differences in the performance of the experimental group (planning condition) when compared to the performance of the control

group (no planning condition) in Task 2. Effect sizes were also calculated in the attempt to verify the magnitude of the effects of planning on L2 speech performance. Effect sizes were calculated using the formula by Cohen (1988, as cited in NORRIS; ORTEGA, 2000).

Following Conway et al. (2005), the present study adopted an extreme group design in the attempt to scrutinize differences between lower and higher spans individuals more precisely. According to Conway et al. (2005, p. 782), “extreme-group designs refer to contexts in which a continuous variable is categorized, and only the lower and upper ends of this variable distribution are represented”.

Although extreme-group designs present problems⁶, they are common in the working-memory literature, and they may be useful in the attempt to scrutinize differences between lower and higher spans individuals (CONWAY et al., 2005). One advantage in using extreme-group designs is that individuals are hardly ever misclassified as lower or higher spans since only the extremes are used. Moreover, it allows further scrutiny of differences between higher and lower spans (CONWAY et al., 2005).

The most common type of extreme-group design is based on quartiles; however, tertiles can also be used when data samples are small (CONWAY et al., 2005). In order to conduct the extreme group design, the cutoff point was established between two percentiles: 33,3% and 66,6%. Based on these percentiles, participants were categorized as having higher working memory span when they fell in the upper tertile (the ones above 66,6%), and lower working memory span when they fell in the lower tertile (below 33,3%).

Out of the 25 participants who belonged to the experimental group, 8 were classified as lower spans, and 8 were classified as higher spans. The remaining 9 participants were classified as intermediate spans and were not included in the analysis which focused specifically on comparing lower and higher spans. In order to verify whether higher spans outperform lower spans in planning performance, one-way ANOVAs were run for the measures of fluency, accuracy and complexity.

⁶ The first problem with extreme-group designs is that information is lost, since only the extremes of the population are included in the analysis. Second, they tend to overestimate effect sizes (Conway et al., 2005).

RESULTS

In this section, results of the study will be briefly reported by addressing the research questions. The first question is: Does pre-task planning opportunity significantly increase fluency, accuracy, and complexity of L2 speech performance?

Due to space constraints, results of the ANOVAs will not be displayed on tables. ANOVA results showed that there were significant differences in the performance of the experimental group when compared to the performance of the control group in Task 1. There were significant differences in accuracy as measured by the percentage of error free clauses ($F(1, 48) = 5.492^*$, $p = 0.023$), with a medium effect size ($d = .66$), and in complexity as measured by the number of clauses per c-unit ($F(1,48) = 5.067^*$, $p = 0.29$), with a medium effect size ($d = .65$).

It is important to highlight that ANOVAs were also computed for performance in task 1 and there were no significant differences between the control and experimental groups terms of (1) fluency as measured by speech rate unpruned, speech rate pruned, number of pauses per c-unit, and percentage of total pausing time; (2) accuracy measured by number of errors per a hundred words, nor as measured by the percentage of error free clauses; and (3) in terms of complexity as measured by the number of clauses per c-unit.

Bearing in mind that there were no significant differences between the two groups in the performance of Task 1, it can be argued that the significant differences in accuracy and complexity in the performance of Task 2 can be attributed to pre-task planning. In general, these results corroborate previous studies in the literature, in which planning also leads to gains in performance (FOSTER; SKEHAN, 1996; MEHNERT, 1998; ORTEGA, 1999, among others). However, the aspects of performance for which planning leads to an increase in most previous studies are *fluency* and complexity. These results will be discussed later in this paper. Having addressed the first research question, I turn to the results of the second research question of the current study: Do higher working memory span individuals significantly outperform lower working memory span individuals in terms of fluency, accuracy and complexity of L2 speech performance under pre-task planning condition?

Results showed that there were significant differences between lower and higher spans when they perform a task under a planning condition.

There were significant differences in terms of fluency as measured by speech rate unpruned ($F(1, 14) = 8.676, p=0.011$) and pruned ($F(1,14) = 9.473, p=0.008$); and there were also significant differences in terms of complexity as measured by number of clauses per c-unit ($F(1,14) = 6.725, p = 0.021$). Differences in accuracy as measured by number of errors/100 words only approached significance ($3.093, p = .068$).

It is important to highlight that ANOVAs were also computed to check differences in the performance of lower and higher spans individuals in the performance of task 1. Results showed that differences did not reach significance in any of the measures of performance. These results suggest that significant differences in the performance of lower and higher spans, in Task 2, may have taken place *not only* because of individual differences in working memory capacity, *but also* due to the opportunity of pre-task planning.

DISCUSSION

This first section deals with the impact of planning on performance of the experimental group as a whole, *regardless* of individual differences in working memory capacity. To reiterate, the first research question of the present study asked whether pre-task planning would lead to significant increase in fluency, accuracy, and complexity in the performance of the experimental group when compared to the control group. As shown in the previous section, all means of L2 speech performance measures in the second narrative task favor the experimental group when compared to the control group. However, only differences in two of these measures achieved statistical significance: accuracy as measured by the percentage of error-free clauses, and complexity as measured by the number of clauses per c-unit.

In most studies on task based planning, results have shown a stronger impact for fluency (FOSTER; SKEHAN, 1996; MEHNERT, 1998; ORTEGA, 1999) and complexity (CROOKES, 1989; FOSTER; SKEHAN, 1996; ORTEGA, 1999; YUAN, ELLIS, 2003). Most studies show that gains in fluency and complexity may be achieved at the expense of accuracy (MEHNERT, 1998). In this sense, the results of the present study do not corroborate previous findings since the impact of planning was stronger for accuracy and complexity.

This stronger impact for accuracy and complexity is intriguing since, according to Crookes (1989, p. 379), “it is unlikely that learners who produce more complex speech than they are normally capable of will at the same time maintain a given level of accuracy”. In other words, as learners take risks in the attempt to produce more complex language, chances are that they will be less prone to avoid errors (CROOKES, 1989).

A tentative explanation for the planning impact on accuracy and complexity may be related to the mental processes learners engage when they plan an oral task. Foster and Skehan (2001) suggest that the activities that take place during pre-task planning and the mental processes in which learners engage in are crucial for understanding the impact of planning on performance.

These researchers claim that efforts allocated towards different mental activities entail distinct benefits to performance. In the case of rehearsal⁷, it tends to be mostly language oriented and is likely to affect accuracy (FOSTER; SKEHAN, 2001). As for efforts allocated towards retrieval operations⁸, they lead to benefits in complexity by making available a wider language repertoire, allowing learners to access the upper limits of their interlanguage (CROOKES, 1989; FOSTER; SKEHAN, 2001). Previous studies on the processes learners engage when they plan have revealed an emphasis on retrieval and rehearsal strategies (ORTEGA, 2005). Therefore, a possible emphasis on retrieval and rehearsal from the part of learners may explain the lack of significant planning effects on fluency. As pointed by Crookes (1989), as learners take more risks they tend to produce more errors. Since learners were able to take risks and still sustain accurate speech, effects on fluency failed to achieve significance.

Skehan (1998) claims that fluency, accuracy, and complexity compete for learners’ attentional resources, and thus trade-off effects take place among these aspects of performance. Possibly, learners attained significantly more complex and accurate speech at the expense of producing significantly more fluent speech. Previous studies also give evidence for trade-off effects (FOSTER; SKEHAN, 1996; MEHNERT,

⁷ Rehearsal strategies represent learners’ attempts to practice performance, such as practicing it mentally or aloud.

⁸ Retrieval strategies represent learners’ attempts to remember what was planned such as writing key words or drawing.

1998; YUAN; ELLIS, 2003) but in a different direction. The research results tend to show that planning impacts predominantly fluency and complexity at the expense of accuracy (Mehnert, 1998). Guará-Tavares (2009), however, indicated that planning yielded significant differences in fluency and accuracy at the expense of complexity. The present study showed benefits of planning for accuracy and complexity at the expense of fluency. How can these conflicting results be explained?

According to Ortega (2005) some learners seem to be more oriented towards form, while others towards meaning. Perhaps, learners' predispositions towards prioritizing fluency, accuracy or complexity plays a role in determining what aspects will be involved in such trade-off effects. Thus, the attentional trade-off effect among fluency, accuracy and complexity may not be a general one, but a situated one.

It seems that planning triggered learners to search for more efficiency in performance, which was possibly reflected in more accurate and complex speech. According to Ellis (2003) it is the learner who decides what kind of 'activity' to engage in during performance and such choices determine what to prioritize. Possibly, learners' 'choices' on what aspects to prioritize are not deliberate conscious choices. Rather, they may reveal reflexlike behavior based on their learning orientation, learning backgrounds and on task performance conditions (e.g., planning). These 'choices' may start in a reflexlike fashion, but it is attentional control that will be necessary to sustain such 'choices' (FELDMAN-BARRETT et al., 2004), that is, learners will 'choose' what to prioritize as they attend to the tasks, make sense of them, and start performing them. Moreover, planning may have impacted more on fluency in the performance of higher spans; thus, it may not have impacted fluency in the performance of the experimental group *as a whole*. This will be further discussed in the next section.

Differences between lower and higher working memory spans in L2 speech performance under planning condition

In brief, results concerning whether higher spans outperform lower spans in L2 speech performance under planning conditions show that: (1) Higher spans significantly outperformed lower spans in terms of fluency as measured by speech rate unpruned and pruned, and (2) higher spans significantly outperformed lower spans in terms of complexity as measured by number of clauses per c-unit.

These significant differences between the performance of higher and lower spans could be attributed to working memory only, regardless of planning. However, results reported in the previous section showed that there were *no* significant differences between higher and lower spans in the performance of the first narrative task under no planning conditions. Therefore, it seems reasonable to argue that performance conditions in Task 1 (having 40 seconds to look at a set of 9 pictures, put them away and start performing immediately) may have been too difficult for all learners (control and experimental group).

In order for individual differences in working memory capacity to emerge, the task has to be complex. Tasks that are either too easy or too complex are not likely to reveal individual differences in working memory capacity (JUST; CARPENTER, 1992). Once the task was made more manageable due to the opportunity to plan, individual differences could more fully emerge; thus, higher spans significantly outperformed lower spans in terms of fluency and complexity.

Interestingly, fluency was the dimension which yielded greater differences between higher and lower spans (and for speech rate unpruned and pruned respectively), for complexity). In other words, it was the ability to produce significantly more fluent speech while still maintaining significantly more complex, and marginally significantly more accurate speech that yielded greatest differences between higher and lower spans under planning conditions.

Now it seems reasonable to bring the discussion on the impact of planning on fluency back into the present scenario. It is important to highlight that there were no significant differences between lower and higher spans under no a planning condition; but fluency was the dimension which yielded the greatest differences between higher and lower spans under a planning condition, which suggests that there was *some* impact of planning on fluency. However, it seems that higher spans were more susceptible to the impact of planning on fluency; thus, the *overall* impact of planning on fluency for the experimental group *as a whole* was reduced.

These results not only lend support to the issue of trade-off effects among the goals of fluency, accuracy, and complexity (FOSTER; SKEHAN, 1996; MEHNERT, 1998; YUAN; ELLIS, 2003; GUARÁ-TAVARES, 2009) but also suggest that trade-off effects seem to be acute for

learners with lower working memory capacity since higher spans significantly outperformed lower spans in terms of fluency when planning opportunity was provided. Again, following Fortkamp (2000), it can be argued that under planning conditions, individuals with more working memory capacity have more attentional resources available to allocate towards the processes involved in L2 speech production as a controlled process activity, which may explain the results obtained.

Based on the findings that, under planning condition, higher spans outperformed lower spans in terms of fluency and complexity and that these differences cannot be attributed to individual differences in working memory *only*, but *also* to planning, it seems that higher spans were more able to benefit from the opportunity to plan performance of an oral task. Hence, one question must not remain unanswered: What is it that planning requires that higher span individuals are better able to cope with and, as a result, they are more able to benefit from planning?

According to Hambrick and Engle (2003, p. 179), problem solving activities require “the ability to maintain goals, action plans, and other task-relevant information in a highly activated and accessible state, and when necessary, to inhibit activation of irrelevant or distracting information”. When planning an oral task, learners need to activate task-relevant information, maintain them activated and accessible until this information can be integrated to subsequent information in a coherent way; learners also need to sustain, maintain, and switch attention from the various components of the task (e.g., from meaning to form and vice-versa), suppress irrelevant L2 and L1 information, and monitor. It is the ability to control attention among the various components of planning that higher spans seem to be better able to cope with, which may explain why higher spans benefited more from planning and, thus, significantly outperformed lower spans in terms of fluency and complexity.

The fact that fluency was the dimension which yielded the greatest differences between lower and higher spans is an interesting finding which merits some reflection. Ortega (1999) claims that the extent to which planning leads to benefits on performance also depends on the ability to execute what was planned into online performance. In other words, it also depends on the ability to *retrieve* what was planned into real time performance.

In this study, fluency was operationalized as the ability to perform in real time communication (SKEHAN, 1996, 1998), and it was measured by speed (speech rate) and silence (pauses) measures. It seems reasonable to argue that successful retrieval may have aided implementation of what was planned into real time performance, thus, reflecting greater differences between lower and higher spans in terms of fluency.

Individual differences in working memory capacity reflect differences in the ability to retrieve information from long term memory (ROSEN; ENGLE, 1997; UNSWORTH; ENGLE, 2007). Unsworth and Engle (2007) provide evidence that higher spans are more effective at retrieving task-relevant information in the face of interference whereas lower spans are more likely to lose access to task-relevant information since they are more susceptible to have their attention captured by distraction and to activate more irrelevant information.

Based on these findings concerning the role of working memory in retrieval, it seems plausible to argue that higher spans were more able to retrieve what was planned into real time performance. Therefore, fluency as assessed by means of real time communication measures was the dimension of speech which yielded the greatest differences between lower and higher spans when performing a task under planning conditions.

Besides the ability to retrieve what was planned into online performance, I believe the ability to implement new ideas online may also have enhanced the benefits of planning on the performance of higher spans. Several learners verbalized that they followed their plans but also implemented new ideas online. According to Feldman-Barrett et al. (2004), changing representations online is achieved by rule-based processing since it requires incorporating new information into existing representations. Rule-based processing is under attentional control and may be affected by individual differences in working memory capacity.

Bearing the preceding discussion in mind, it seems reasonable to argue that a more comprehensive explanation for the relationship between working memory capacity and L2 speech performance under planning condition is that higher spans are *not only* more able to effectively allocate attentional resources towards the processes involved in L2 speech production during task performance, as argues Fortkamp (2003) *but also* more able to cope with the processes involved in planning as a problem solving activity (HAMBRICK & ENGLE, 2003), more able to

retrieve what was planned into performance (ROSEN & ENGLE, 1997; UNSWORTH & ENGLE, 2007b) and more able to implement new ideas online (FELDMAN-BARRETT et al, 2004).

The last question to be pursued in this section is: Why did higher spans significantly outperform lower spans in fluency and complexity but not accuracy? I will put forward two tentative explanations. First, it could be due to trade-off effects. Higher spans were more able to achieve significantly more fluent and complex speech, when compared to lower spans, at the expense of achieving more accurate speech.

Second, Feldman-Barrett et al. (2004) propose that, in complex tasks, lower spans may have a range of goals; however, they lack sufficient attentional resources to maintain goal-relevant processing in complex situations. As a result, they end up devoting attention to efficiency over any other processing goal. Possibly, lower spans view error free performance as efficient performance and pursued a more conservative approach to L2 speech under planning condition. Consequently, the differences between higher and lower spans in terms of accuracy did not reach statistical significance.

CONCLUSION

The current study sought to examine the role of individual differences in working memory capacity within the effects of planning on L2 speech performance. In general, results show that individual differences in working memory capacity mediate the impact of pre-task planning on L2 speech performance.

Obviously, findings of the present study are to be seen as modest and suggestive rather than conclusive due to its several limitations. The present study is limited in its sample size; in the measures to assess working memory (only one test); in the number of participants. Due to this reduced sample size, the extreme-group design was conducted based on tertiles, not quartiles. Moreover, the differences between lower and higher spans were based on a more reduced sample of only sixteen participants, 8 classified as lower and eight classified as higher spans. Future studies need to consider expanding sample size and measures of working memory capacity.

Despite its limitations, the findings of the present study are relevant since they go beyond the general speculation that the effects of plan-

ning are not achieved simultaneously to the same extent for fluency, accuracy, and complexity due to limitation in attentional resources (FOSTER; SKEHAN, 1996; MEHNERT, 1998; YUAN; ELLIS, 2003). This study represents a step forward by providing evidence that individual differences in working memory capacity mediate L2 speech performance under planning conditions. Moreover, the findings of the present study suggest that lower spans tend to be more susceptible to attentional trade-off effects among fluency, accuracy, and complexity.

BIBLIOGRAFIA

- COHEN, J. *Statistical power analysis for the behavioral sciences* (2nd Ed.). New York: Academic Press, 1988.
- CONWAY, A. R. A. et al. Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review*, 12 (5), p.769-786, 2005.
- CROOKES, G. Planning and interlanguage variation. *Studies in Second Language Acquisition*, 11, p.367-383, 1989.
- DANEMAN, M. Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, 20, 445-464, 1991.
- DANEMAN, M., & GREEN, I. Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, 25, 1-18, 1986.
- DE BOT, K. A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics*, 13, p. 1-24, 1992.
- D'ELY, R. *An investigation of learners' metacognitive processes of planning and rehearsal as catalysts of interlanguage development*. Unpublished manuscript, 2004.
- D'ELY, R. *A focus on learners' metacognitive processes: The impact of strategic planning, repetition, strategic planning plus repetition, and strategic planning for repetition on L2 oral performance*. Unpublished Ph.D. dissertation, Universidade Federal de Santa Catarina, Florianópolis, 30 de outubro de 2006.
- D'ELY, R., & WEISSHEIMER, J. *Scale of L2 speaking proficiency*. Unpublished manuscript, 2004.
- ELLIS, R. Interlanguage variability in narrative discourse: Style shift-

- ing in the use of the past tense. *Studies in Second Language Acquisition*, 9, 12-20, 1987.
- ELLIS, R. *Task-based language teaching and learning*. Oxford: Oxford University Press, 2003.
- ELLIS, R. (2005). Planning and task-based performance: Theory and research. In R. Ellis (org.), *Planning and task performance in a second language*. Amsterdam and Philadelphia: John Benjamins, 2005.
- ELLIS, R. and YUAN, F. The effects of careful within-task planning on oral and written task performance. In R. Ellis (org.), *Planning and task performance in a second language*. Amsterdam and Philadelphia: John Benjamins, pp.167-192, 2005.
- ENGLE, R. W.; KANE, M. J. & TUHOLSKI, S. W. Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (orgs.) *Models of working memory: mechanisms of active maintenance and executive control*. New York: Cambridge University Press, p.102-134, 1999.
- FELDMAN-BARRETT, L., TUGADE, M. M., & ENGLE, R. W. Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, 130, p.553-573, 2004.
- FIELD, A. *Discovering statistics using SPSS*. Sage, 2005.
- FORTKAMP, M. B. M. Working memory capacity and aspects of L2 speech production. *Communication and Cognition*, 32, p.259-296, 1999.
- FORTKAMP, M. B. M. *Working memory capacity and L2 speech production: An exploratory study*. Unpublished Ph.D. dissertation, Universidade federal de Santa Catarina Florianpolis, 4 de dezembro de 2000.
- FOSTER, P. & SKEHAN, P. The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, p.299-323, 1996.
- GUAR-TAVARES, M. G. The relationship among pre-task planning, working memory capacity and L2 speech performance: a pilot study. *Linguagem & Ensino*, v. 12, p. 165-194, 2009.
- HAMBRICK, D. Z., & ENGLE, R. W. The role of working memory in problem solving. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* p. 176-206, 2003. London: Cambridge Press.
- JUST, M. A., & CARPENTER, P. A. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, p.122-149, 1992.

- KAWAUCHI, C. The effects of strategic planning on the oral narratives of learners with low and high intermediate levels of L2 proficiency (pp.) In R. Ellis (org.), *Planning and task performance in a second language*. Amsterdam and Philadelphia: John Benjamins, p. 143-164, 2005.
- LEVELT, W. J. M. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press, 1989.
- MEHNERT, U. The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20, p. 83-108, 1998.
- NORRIS, J., & ORTEGA, L. Effectiveness of L2 instruction: A Research Synthesis and quantitative meta-analysis. *Language Learning*, 50 (3), p.417-528, 2000.
- ORTEGA, L. Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21, p.109-148, 1999.
- ORTEGA, L. What do learners plan? Learner-driven attention to form during pre-task planning. In R. Ellis (org.), *Planning and task performance in a second language*. Amsterdam and Philadelphia: John Benjamins, p. 77-110, 2005.
- ROBINSON, P. Task complexity and second language narrative discourse. *Language Learning*, 45 (1), p. 99-140, 1995.
- ROSEN, V. M., & ENGLE, R. W. The role of working memory capacity in retrieval. *Journal of Experimental Psychology*, 126 (3), p.211-227, 1997.
- SANGARUM, J. The effects of focusing on meaning and form in strategic planning. In R. Ellis (org.), *Planning and task performance in a second language*. Amsterdam and Philadelphia: John Benjamins, p. 111-142, 2005.
- SKEHAN, P. A framework for the implementation of task-based instruction. *Applied Linguistics*, 17 (1), p. 38-62, 1996.
- SKEHAN, P. Tasks and Language performance assessment. In M. Bygate, P. Skehan e M. Swain (orgs.) *Researching Pedagogic Tasks, Second Language Learning, Teaching and Testing*, Harlow, Longman, p. 167-185, 2001.
- SKEHAN, P., & FOSTER, P. Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task performance in a second language*. Amsterdam and Philadelphia: John Benjamins, p. 193-218, 2005.

UNSWORTH, N., & ENGLE, R. W. The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114, 104-132, 2007a.

UNSWORTH, N., & ENGLE, R. W. Individual differences in working memory capacity and retrieval: A Cue-dependent search approach. In J. S. Nairne (org.), *The foundations of remembering: Essays in honor of Henry L. Roediger* (pp. 241-258). New York: Psychology Press, 2007b.

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

Endnote: I would like to thank Mailce Borges Mota for her excellent advising and expertise during my doctoral studies at Universidade Federal de Santa Catarina. I am also thankful to Rod Ellis for his excellent advising and insightful comments during my period of data analysis at the University of Auckland, NZ.