The simultaneous manipulation of task complexity along planning time
and +/- Here-and-Now: effects on L2 oral production

Roger Gilabert

Blanquerna Communication Studies Department
Universitat Ramon Llull
C/ Valldonzella 23
08001 Barcelona
Spain
rogergg@blanquerna.url.edu
1. Introduction

The objective of this study is to analyze the effects of manipulating the cognitive complexity of tasks on L2 narrative oral production, and it specifically addresses the issue of how the three dimensions of production (i.e. fluency, complexity, and accuracy) compete for attention during L2 task performance. It does so by establishing four levels of cognitive complexity which are achieved by simultaneously manipulating two widely researched variables: pre-task planning time and the degree of displaced, past time reference.

In the last few years, considerable attention has been devoted to task design and task-based syllabus design. Basically, two different agendas have inspired research into task features. The first one is an interactionist perspective which has been concerned with establishing what modifications can be applied to tasks in order for them to generate specific conversational episodes which, generally, have been regarded as negotiation of meaning. These studies have been particularly interested in whether task design can lead to interactive production episodes that have been referred to as clarification requests, confirmation checks, and comprehension checks. These episodes have been claimed to lead to second language acquisition (Long, 1985, 1989, 2000). Apart from their interest in negotiation of meaning during production, in these kinds of studies researchers have also looked into the consequences of task manipulation on the amount of production and the level of participation of learners.

From an information-processing perspective concerned with performance, questions have been asked as to how task manipulation can lead to differentials in the areas of fluency, complexity, and accuracy. A number of researchers have investigated
the effects of task on production along their degree of familiarity (Bygate, 1999, 2001; Foster & Skehan, 1996; Plough & Gass, 1993; Robinson, 2001a); their number of elements (Kuiken & Vedder, 2004; Robinson, 2001a); single and dual task performance (Niwa, 2000); the pre-task and on-line planning time allotted to them (Crookes, 1989; Ellis, 1987; Foster & Skehan, 1996; Mehnert, 1998; Ortega, 1999; Skehan & Foster, 1997; Wigglesworth, 1997; Yuan & Ellis, 2003); and their degree of complexity along displaced, past time reference (Iwashita et al. 2001; Robinson, 1995; Rahimpour, 1997).

These studies have been concerned with how a balanced performance in the three areas of production can potentially lead to more effective language use and acquisition, as well as with how such information can be used to make sequencing decisions in syllabus design.

It is the preoccupation with sequencing tasks in a syllabus in a principled way that has given rise to the concept of task complexity. A great divide has traditionally existed between models which argue that decisions in syllabus construction should be motivated by findings in second language acquisition (SLA) (Long, 1985; Long & Crookes, 1992; Robinson, 1998; Skehan, 1998) and those who suggest criteria which are not necessarily informed by SLA (Ellis, 1997; Nunan, 1989; Willis, 1990). In synthetic syllabi, whether structural, lexical, or functional-notional, grading of syllabus units is quite an intuitive activity which depends on various notions of ‘difficulty’, ‘usefulness’ or ‘frequency’. Structural syllabi have traditionally used sequencing criteria such as ‘difficulty’, ‘usefulness’, or ‘general agreement’ to decide on the order in which linguistic material should be presented to learners. In most cases, if not all, the various concepts of difficulty or usefulness have been left unexplained. Ellis (1993) suggested the teaching of marked features, since unmarked features can be learned naturally by learners. As Robinson (1998) points out, the definition of markedness and the implicit
idea that unmarked features do not need focused attention remains a problem. Ellis (1993) also suggested on-line modification of the syllabus to correct errors made by learners during course implementation. These, however, are non-complementary criteria. In lexical syllabi, frequency has been the basic criterion for sequencing units, with most frequent items being taught first. Recycling (e.g. spiral syllabi) has been used as a solution to one-time presentation of linguistic items, and in notional-functional syllabi the relative frequency of functions was used as a sequencing criterion.

The only non-intuitive, data-driven proposal for the sequencing of structural syllabi was advanced by Pienemann (1998) who suggested the use of what we know about the stages in which structures are learned to organize a syllabus which is coherent with acquisition. This idea, although compelling, is limited by the fact that we only have information about a small number of structures and in only a few languages, which makes reasoned structural sequencing difficult. Within analytic syllabi with an almost exclusive focus on meaning, procedural syllabi (Prabhu, 1987) have had quite a random selection and sequencing of tasks, which were taken from other subjects. Finally, in process syllabi (Breen, 1984; Candlin, 1984) tasks are jointly negotiated between teachers and learners and therefore organized according to learners’ wants and needs. In content-based syllabi, sequencing is the result of incorporating the intuition of experts in the subject matter into syllabus design.

However, criteria are needed on which to base prospective decisions about syllabus design and to advance predictions about interlanguage development. In this context, the concept of task complexity is born.
2. Task complexity and L2 oral production

The concept of task complexity, then, springs from the need to establish criteria for sequencing tasks in a syllabus from easy/simple to difficult/complex in a reasoned way that will foster interlanguage development. Rather than looking at the linguistic features of language activities, syllabi that have used tasks as their units have focused on task design in order to find out how tasks impose cognitive demands on learners. In this way, task design has allowed researchers to speculate about the effects that increasing task difficulty or complexity may have on L2 task performance. With some exceptions (Kuiken & Vedder, 2004), the majority of these studies have been concerned with oral production.

2.1. The concept of task complexity

Beyond early attempts at sequencing tasks from simple to complex in a syllabus (Brown et al., 1984; Prahbu, 1987), Skehan (Skehan, 1998; Skehan & Foster, 2001) and Robinson (2001a; 2001b; 2003; 2005a) have identified a series of task design factors that can be manipulated along a continuum in order to achieve different levels of difficulty or complexity.

From a communicative approach to language teaching, which has been concerned, among other issues, with how task and syllabus design can contribute to interlanguage development, Skehan (Skehan, 1998; Skehan & Foster, 2001) suggests that both task manipulation and sequencing for syllabus design should be based not just on intuitions about difficulty but on empirical findings. For Skehan and Foster (2001:196): “Task difficulty has to do with the amount of attention the task demands from the participants.
Difficult tasks require more attention than easy tasks”. Having evidence of the effects of task demands on production can be used to direct learners’ efforts toward different areas of performance separately or simultaneously. In addition to that, if links are established between production and acquisition, research evidence can be used to manipulate tasks to maximize the effectiveness of language learning.

Skehan (Skehan, 1998; Skehan & Foster, 2001) suggests a three-way distinction of difficulty, to which learner factors can also be added: code complexity (vocabulary load and variety; linguistic complexity and variety); cognitive complexity (familiarity topic, discourse or task; amount of computation and organization, and sufficiency of information); communicative stress (time pressure; scale; number of participants; length of text; modality; stakes; opportunity for control); and learner factors (intelligence; breadth of imagination; personal experience).

He suggests that information should be collected regarding the effects of task manipulation on the areas of fluency, accuracy, and complexity. He takes linguistic complexity to be a ‘surrogate’ of learners’ willingness to stretch their interlanguage by experimenting with more difficult forms and by trying out more elaborate language. In his view, the information obtained from the manipulation of task features can be used to establish longer-term pedagogic goals in which both meaning and form can be attended to, and in which interlanguage development can be integrated into fluent performance. Regarding sequencing, Skehan and Foster (2001: 193-194) propose that: “the individual task has to be located, in a principled way, in longer-term instructional sequences which seek to promote balanced development, such that improvement in one area will be consolidated by improvements in others.”

Skehan and Foster (2001), however, make no specific suggestions as to which dimension should be used for making prospective sequencing decisions. Their starting
point is language instruction, which in their view should foster a balanced improvement in the three areas of production. In their view, information about how fluency, complexity, and accuracy as affected by increasing task demands should be used to arrange individual tasks in a principled way in a long-term instructional sequence. This will consequently promote such balanced development in the three areas. However, Skehan and Foster’s ‘principled way’ is in need of further exploration.

Acknowledging the rich research tradition in the interactive dimension of tasks, Robinson has shifted the focus to the cognitive processes involved in task production. Robinson (2001a:28) says that: “task complexity is the result of the attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner. These differences in information processing demands, resulting from design characteristics, are relatively fixed and invariant”.

Robinson proposes a three-dimensional model that distinguishes between three different types of factors: cognitive complexity factors (resource-directing ones such as +/- few elements, +/- Here-and-Now, and +/- no reasoning demands; resource-dispersing ones such as +/- planning, +/- single task, and +/- prior knowledge); interactive factors (participation variables such as one way/two way, convergent/divergent, open/closed; participant variables such as gender, familiarity, power/solidarity); and learner factors (affective variables such as motivation, anxiety, and confidence; ability variables such as aptitude, proficiency, and intelligence).

In Robinson’s view, task complexity should be the sole basis for making prospective sequencing decisions, since task conditions (participation and participant variables) and task difficulty (affective and ability variables) often cannot be predicted before a course starts and can therefore only inform on-line decisions. In his view, task performance conditions are determined by a needs analysis. Information about the
effects of task complexity on production should help syllabus designers to organize pedagogic tasks from simple to complex so that they progressively approximate real world target tasks. According to Robinson (2001b:301), increasing the cognitive complexity of tasks “will facilitate the ‘means’ of language learning, and therefore lead to a transition in the learner’s knowledge states.”

2.2. Production as mediated by complexity and attention

These two different but complementary conceptions of complexity make different predictions about the fluency, complexity, and accuracy of L2 production during task performance.

As we have seen, Skehan (Skehan, 1998; Skehan & Foster, 2001) understands difficulty as the amount of attention the task demands from the participants. Skehan’s view of how the three areas of production interact during performance is based on a limited-capacity view of attention in which more difficult tasks demand more attention than easy tasks, and such higher demands for attention have specific consequences for performance. Skehan bases his predictions on a limited-capacity conception of attention, which suggests that when task demands are high, attention can only be allocated to certain aspects of performance to the detriment of others. His conception of attention resembles the models of early selection and limited capacity, like the one advanced by Kahneman’s (1973) in which there is a single volume of attention that ‘runs out’ of resources. Of the three dimensions of performance, Skehan believes that complexity and accuracy are in competition for attention, a statement that he has supported with evidence from a number of studies on pre-task planning, which will be reviewed in the following section.
Robinson (2001a; 2001b; 2003; 2005a), however, has a contrasting view regarding cognitive complexity. He proposes that attention, as suggested by models such as Wickens’ (1989, 1992), can draw on multiple resources. Wickens’ (1989) model of dual-task performance proposes breaking the single volume of attentional resources into a series of dichotomical dimensions. The dichotomy affecting processing stages opposes the perceptual/cognitive dimension (i.e. encoding and central processing) to the response dimension. For codes of processing, the spatial is opposed to the verbal. In the case of modality, auditory perception is opposed to visual perception, and responses can be manual or vocal. Wickens (1989, 1992) claims that because these different dimensions draw on different resource pools, competition for attention may not necessarily happen. There is competition when two tasks or two dimensions of the same task feed on the same resource pool. For example, having two conversations simultaneously would lead to competition of resources and poor performance since they would both draw on the same ‘verbal’ resource pool. On the other hand, driving home while simultaneously singing or verbalizing thoughts would draw on the ‘manual’ and ‘vocal’ dimensions of performance, and therefore no competition for attention would occur. Thus, the model predicts that the amount of interference will depend on the number of shared levels on all three dimensions (i.e. codes, modalities, and responses). Critics of this model have suggested that Wickens’ resource pools may have been too broadly defined, since evidence exists that two tasks defined as drawing on the verbal dimension, like simultaneous spelling and arithmetic calculation, have been shown not to necessarily enter into competition.

Robinson believes that manipulating task complexity by increasing the cognitive demands of tasks can lead to simultaneous improvement of accuracy and complexity. In his view, then, differentials in performance may be better explained by concepts such as
interference and confusion rather than by limited resources. As was briefly mentioned above, Robinson distinguishes between resource-directing and resource-dispersing dimensions. Manipulating Task Complexity along the first group of task variables (+/- elements, +/- here-and-now, +/- reasoning demands) directs attention to a wide range of functional and linguistic requirements. Increasing complexity along resource-dispersing dimensions (+/- planning time, +/- prior knowledge, +/- single task) reduces attentional and memory resources with negative consequences for production, a position which is in agreement with Skehan’s. Despite such negative consequences, progressively increasing complexity along resource-dispersing variables is also important in order to approximate the complexity conditions under which real-world tasks are performed.

3. Planning time and +/- Here-and-Now studies

This section reviews the literature on the variables of planning time and +/- Here-and-Now, as they are crucial for this study. I specifically focus on the issue of attention during oral performance and discuss the different findings. For a more detailed review of planning time studies see Ellis (2005) and for a review of both planning time and +/- Here-and-Now studies see Gilabert (2005).

3.1. Planning time studies

What research evidence has shown so far is that giving learners extended planning time before task performance seems to have beneficial effects for fluency and complexity, while the picture for accuracy is not so clear. As we will see below,
providing extended ‘on-line’ (i.e. during performance) planning time has shown to have positive effects for accuracy (Yuan & Ellis, 2003).

From a variationist stance, Ellis (1987) looked into how different levels of planned discourse affect learners’ written and oral performance. He operationalized three different degrees of planned discourse, and hypothesized that access to forms that have not been fully automatized, such as the third person ‘–s’ or regular past ‘-ed’, would benefit from planning time. Ellis found that the accuracy in the use of irregular past forms was not affected by the different levels of planning. His main conclusion was that increased planning time leads to higher accuracy of rule-based language, while unplanned discourse is more lexical in nature.

Foster and Skehan (1996) manipulated planning time on three different task types: a personal information gap task, a narrative task, and a decision-making task. They predicted that planning (which was operationalized as 10-minute pre-task planning, and 10-minute pre-task planning plus guidance) would make language more fluent, more structurally complex, and more accurate, and even more so when guidance as to how to plan was provided. Foster and Skehan found a significant effect for two measures of fluency, number of pauses and total silence, and reported that the personal task triggered the most fluent speech; complexity, in terms of sentence nodes per C-units, was higher for the detailed planning group than for the undetailed planning group. The undetailed planning group, in turn, triggered higher levels of complexity than the group without planning. The researchers, however, obtained mixed results for accuracy, the undetailed planning group obtaining more error-free clauses. The narrative task triggered the most complex speech but the lowest levels of accurate language. Foster and Skehan concluded that there are ‘tradeoff’ effects between complexity and accuracy, especially with narrative tasks, in which attention devoted to complexity has negative effects for
accuracy. This was not the case with decision-making tasks, in which accuracy and complexity were more balanced.

In a subsequent study, Skehan and Foster (1997) used the same kind of task types as in their 1996 study, and with the same kind of predictions as to how planning time would affect performance. This time, however, they operationalized a post-task requirement in which learners were told that they would go public, which they predicted would result in higher levels of accuracy. This second time their results showed that planning could be associated with greater fluency and accuracy in the narrative task, while they did not find greater complexity associated with increased planning. Skehan and Foster suggested once again that accuracy and complexity are in competition for attentional resources when task demands are increased along planning time. They therefore concluded that planning time can only be channeled to one of the aspects (either accuracy or complexity), and not to the two dimensions simultaneously.

Mehnert (1998) confirmed Skehan and Foster’s (1997) limited capacity model of attention. In a study with two task types varying in complexity, a simple instruction task and a complex exposition task, Menhert compared the effects of allotting no planning time, one minute, five minutes, and ten minutes to four different groups. Mehnert followed Foster and Skehan (1996) and Skehan and Foster (1997) and hypothesized higher fluency, complexity, and accuracy for planned tasks. His overall results showed that engaging in extended planning time before carrying out a task had positive effects for performance. Regarding each specific measure, he found that planners who had 10 minutes were more fluent, more accurate, and more lexically dense than non-planners, with no significant differences in structural complexity. He found, however, that when learners had a short time for planning (one minute), they focused more on accuracy, whereas when they had longer time to prepare (10 minutes), they tried to produce more
complex speech at the expense of accuracy, suggesting that any gains in accuracy and complexity are not achieved simultaneously.

In contrast to previous suggestions, Ortega (1999) questioned the limited processing capacity model in which the three dimensions of production enter into competition, arguing that previous studies had neglected the investigation of the planning process. In a study of oral narrative discourse under a 10-minute planning condition and a no-planning condition with learners of Spanish, Ortega found results similar to previous studies regarding production. Her results showed that complexity and fluency were enhanced by pre-task planning, whereas mixed results were found for accuracy. She went beyond previous studies and included retrospective interviews in order to find out more about the quality of pre-task planning, in term of focus on form and strategic planning. Ortega concluded that a number of factors affect the quality of planning. Firstly, she suggested that task complexity may play a role in the sense that cognitively complex tasks may benefit more from planning than simple ones. Secondly, she suggested that the operationalization of planning is important in the sense that developmental readiness and task essentialness need to be taken into account. Thirdly, learner orientation towards form or meaning also plays a role. In the fourth place, learner proficiency needs to be brought into the picture since she speculated that higher level students may benefit more and differently from planning than lower-level students. Finally, she suggested that proficiency should be a moderating factor in the limited processing capacity model.

Springing from a concern with the mixed findings regarding accuracy, Yuan and Ellis (2003) operationalized the construct of ‘online’ planning, which they presented in contrast with no planning and pre-task planning. Using an oral narrative in their study, Yuan and Ellis found that whereas pre-task planning time promoted higher complexity
and lexical variety, it did not have significant effects on accuracy, in line with what several previous studies had also reported. Extended on-line planning with no pre-task planning, despite having negative effects for lexical variety, also had a beneficial effect for complexity and, most importantly, for accuracy. Learners who were given unlimited time during performance were less fluent but reformulated and self-corrected their speech more, by drawing on their explicit knowledge, which as a consequence led to a more accurate performance. They also confirmed the trade-off effect in language production, especially of learners with limited L2 proficiency. Since they found that both pre-task planning and on-line planning promote higher accuracy, they concluded that the main trade-off effect is between fluency and accuracy. If learners are given time to plan prior to task performance, they prioritize fluency. If they are given time to plan on-line, they may pay more attention to accuracy at the expense of fluency. Finally, Yuan and Ellis detailed the trade-off effect further by showing that pre-task planning increases lexical variety but not grammatical accuracy, whereas on-line planning improves grammatical accuracy over lexical variety.

3.2. Here-and-Now studies

In general, previous studies have shown that tasks in the There-and-Then are more cognitively demanding than tasks performed in the Here-and-Now, with specific consequences for production.

Robinson (1995) investigated the impact of manipulating Here-and-Now on three different narratives. In the Here-and-Now condition, learners were asked to narrate a comic strip in the present tense while looking at it. The There-and-Then was operationalized by having the students narrate the story in the past tense and without
visual support during performance. Such operationalization was based on both L1 and SLA findings that had shown that displaced, past time reference is more complex and therefore appears later than present, context-supported reference. Robinson predicted less fluent speech but higher lexical and structural complexity as well as accuracy for There-and-Then tasks. Robinson found that the most complex narrative, performed in displaced past time reference, elicited more accurate speech and more lexical complexity than the narrative performed in the Here-and-Now. It also showed a trend for greater dysfluency but showed no significant differences for structural complexity.

Rahimpour (1997) extended Robinson’s research by crossing a complexity variable (Here-and-Now) with a condition variable (open vs. closed). Rahimpour operationalized three levels of complexity by including a narrative in the Here-and-Now, one in the There-and-Then, and one in the Here-and-Now/There-and-Then. Rahimpour hypothesized that the Here-and-Now/There-and-Then narrative would His results showed that learners who carried out the most complex versions of the task were significantly less fluent, with no significant differences regarding either structural or lexical complexity, and with significant improvements with regard to error-free units but not target-like use of articles.

From an interest in language testing, Iwashita et al. (2001) investigated the effects of manipulating complexity on L2 learners’ fluency, complexity, and accuracy. They established 8 levels of complexity along four dimensions: i) +/- perspective, that is, whether the learner was speaking as if the story had happened to her or not; ii) +/- immediacy, that is, in the Here-and-Now or in the There-and-Then; iii) +/- adequacy, that is, whether the set of pictures was complete or incomplete; iv) +/- planning time, which they operationalized as either 3.5 minutes or 0.5 minutes. Following Skehan’s (1996, 1998) predictions for task difficulty, the researchers hypothesized that less
difficult versions of tasks would trigger more fluency and accuracy but less complex speech. They found that there were no significant differences between easy and difficult versions of tasks except for accuracy. In the case of immediacy, they found that the more difficult version of tasks, that is, in There-and-Then, triggered higher levels of accuracy, which went against their prediction.

3.3. Questions and hypotheses

This study is motivated by, firstly, the fact that research evidence is needed regarding the synergistic effects of crossing planning time and the degree of displaced, past time reference, two widely researched task factors which so far have been investigated in isolation. Secondly, it is also motivated by the need to explore the different, and sometimes contradictory, conceptions of attention and predictions regarding the effects of increasing the cognitive demands of tasks on production. This study therefore aims at answering the following general question:

How does increasing the cognitive complexity of tasks along simultaneously planning time and the Here-and-Now variable affect the fluency, complexity, and accuracy of production?

On the basis of previous studies, I entertain the following hypotheses:

Hypothesis 1: Increasing the cognitive complexity of tasks along planning time will have negative effects for all three areas of production, while reducing cognitive
complexity by providing pre-task planning time will have positive effect for fluency and structural complexity but not for lexical complexity or for accuracy.

Hypothesis 2: Increasing the cognitive complexity of tasks along the degree of displaced, past time reference will generate dysfluency but higher structural and lexical complexity and accuracy.

In addition, this research will investigate the synergistic effects of manipulating both dimensions of complexity (+/-planning time and +/- Here-and-Now) simultaneously. No hypotheses will be advanced regarding the manipulation of the two variables since there is a lack of studies that combine them.

4. Experimental design

This study draws on data that are part of a larger study on the effect of manipulating task complexity on L2 narrative oral production (Gilabert, 2005) and addresses the issue of competition for attention. In what follows I provide information regarding participants and experimental design.

4.1. Participants

Forty-eight (48) first- and second-year university students with a lower-intermediate proficiency level of English participated in the study on a volunteer basis. They had been placed in the same level of English class by an internal placement test of the Blanquerna Communication Studies program at Universitat Ramon Llull, in
Barcelona, Spain. However, since the reliability of such a test has never been statistically tested, homogeneity regarding proficiency levels was also controlled for by means of a C-test\(^1\). Learners’ years of instruction ranged from 6 to 12 and their ages were between 18 and 22.

4.2. Design

A repeated-measures design was used in which the within-subjects factor was task complexity. As far as the independent variable is concerned, four levels of Task Complexity were analyzed:

- Condition 1: Planned Here-and-Now
- Condition 2: Unplanned Here-and-Now
- Condition 3: Planned There-and-Then
- Condition 4: Unplanned There-and-Then

Repeated-measures analyses of variance (ANOVA) of the dependent variables were carried out. These include: pruned speech rate for fluency, the Guiraud’s Index of lexical richness for lexical complexity, the S-Nodes per T-units for structural complexity, and the percentage of self-repairs for accuracy.

The design used for data collection assumed that stories were similar to one another and that what made a difference in performance was the condition under which each story was performed. In other words, if a specific condition were to have an effect on performance, it should have happened regardless of the story type. Besides story type, the sequence of conditions under which the tasks were performed was also thought
to potentially affect performance. In order to counterbalance any carryover effects, the following measures were taken:

- Subjects were only given two stories in each session; stories 1 and 2 in the first session, and stories 3 and 4 in the second one. Sessions were two days apart.
- All students narrated the four stories in the same order, but a Latin square design was used to counterbalance the effects of sequence of condition presentation.

Repeated measures ANOVAs for each measure displayed neither differences among stories nor interaction between conditions and sequence of presentation.

4.3. Tasks and procedures

The four stories used in this research were thought to be especially useful for data collection because they were all wordless comic strips, they all contained a small number of characters who were involved in the action, and they had a clear climax and resolution. All plots worked in a similar way: in the first vignettes a number of expectations were generated which were reversed towards the end of the story, with the aim of achieving a humorous effect. Finally, despite the fact that all the social events represented by the comic strips were exaggerated and distorted by the artist’s sense of humor, they were all thought to refer to situations all students could be familiar with within their cultural parameters (See Appendix A).

Following several studies (Foster & Skehan, 1996; Skehan & Foster, 1997; Mehnert, 1998; Ortega, 1999), operationalization of planning time was 10 minutes for
planned narratives and 50 seconds for unplanned ones (enough to understand the story). When planning time was available, subjects were encouraged to take notes on what to say and how to say it as they planned, but were told they would not be allowed to keep their notes during task performance.

Regarding the Here-and-Now/There-and-Then distinction, this research followed Robinson’s (1995a) operationalization. For Here-and-Now, learners were asked to narrate the story in the present while they looked at the strips. For There-and-Then, learners were asked to narrate the story in the past tense, and they were not allowed to look at the pictures as they performed the task.

4.4. Production measures

There has been a wide variety of approaches to measuring fluency. Planning studies have reported significant differences in favor of planning when calculating the number of replacements, repetitions and hesitations (Foster & Skehan, 1996); the number of pauses and total silence (Foster & Skehan, 1996; Mehnert, 1998); and unpruned and pruned speech rates (Mehnert, 1998; Ortega, 1999; Yuan & Ellis, 2003). Studies that have manipulated task complexity along the Here-and-Now variable have used measures such as the total number of 2-second pauses (Robinson, 1995a) and the number of words per pausal unit (Robinson, 1995a; Rahimpour, 1997), albeit with no significant difference. In this study, the rate of pruned speech was chosen to code and measure each narrative. The main advantage of this kind of measure is that it in fact includes both the amount of speech and the length of pauses, since it takes into account the number of syllables and the total number of seconds in the narrative (Griffiths, 1991). In pruned speech rate, as opposed to unpruned speech rate, repetitions,
reformulations, false starts, and asides in the L1 are eliminated from the calculation (Lennon, 1991). The formula used for the calculation of pruned speech rate is the number of syllables divided by the total number of seconds and multiplied by 60.

The type/token ratio has been shown to be extremely sensitive to differences in text length, since the higher the number of tokens, the lower the ratio (Vermeer, 2000). Several alternatives\textsuperscript{3} to the type/token ratio have been advanced by a number of researchers in order to correct the negative correlation existing between type/token results and the number of tokens. After carefully considering Vermeer’s (2000) analyses of the different variations of the type/token measure, this study uses the Guiraud’s index of lexical richness. The advantage of such a measure is that by including the square root of the tokens it compensates for differences in text length. Hence, the Guiraud’s Index of lexical richness was calculated by dividing the number of types by the square root of the number of tokens.

Three basic units of analysis have been used by production researchers for measuring syntactic complexity in oral production: the C-unit, the T-unit, and the utterance\textsuperscript{4}. In this study the T-unit was preferred over the C-unit, since it dealt with one-way, monologic narratives which were expected to trigger no elliptical answers. In this study, then, syntactic complexity was measured by counting the number of S-Nodes (a term which is interchangeable with ‘clause’) and dividing it by the total number of T-units.

Regarding accuracy, this research operationalizes a new measurement: the percentage of self-repairs. There are a number of arguments that can be advanced in order to justify its use: firstly, self-repairs, whether other-initiated or self-initiated (Schegloff \textit{et al.}, 1977), denote students’ awareness of form and can be interpreted as learners’ attempts at being accurate (Kormos, 1999). Lyster and Ranta (1997:57), for
example, suggest repairs generated by learners as a result of corrective feedback lead them both to automatize the retrieval of target language knowledge they already have and to revise their hypotheses about the target language. Swain (1998:66) has also hypothesized that noticing a hole in their own interlanguage may lead learners to notice the gap by directing their attention to relevant input. All these functions of self-repairs have been said to potentially lead to acquisition, and they have been pointed out in order to defend the benefits of certain types of corrective feedback. Self-initiated repairs, on the other hand, serve the same purposes as other-initiated repairs only that they are not the result of corrective feedback but, rather, are spontaneously generated by learners themselves or, in Levelt’s terms (1989), they are the result of the speakers’ monitoring of their own speech.

Secondly, while the percentage of error-free units and the target-like use (TLU) of articles, which have been traditionally used by both planning time and Here-and-Now studies, provide information about the accuracy of the ‘finished’ product of learner’s performance, this new measure presents accuracy ‘in process’ as learners try to correct and improve their own speech. Thirdly, at least one study has reported a higher proportion of self-repairs under certain planning conditions. Yuan and Ellis’s (2003:17), when analyzing the effect of on-line planning on learners’ production, have reported a higher frequency of reformulations and self-corrections when they are given sufficient time ‘during’ (as opposed to ‘before’) performance. About the subjects in their study, they suggest that on-line planners “engaged more fully in searching their linguistic repertoires and in monitoring their speech production”. In the present study, the percentage of self-repairs is calculated by taking the number of self-repairs and dividing it by the total number of errors and multiplying the results by 100.
4.5. Statistical instruments

This study has used repeated-measures analyses of variance (ANOVA) for the comparison of stories and conditions and post hoc Scheffe’s comparisons to identify the exact location of differences. Outliers were detected by means of box plots and eliminated from the calculation in order to achieve the sphericity of the data, which was confirmed by means of Mauchly’s test.

Both intrarater and interrater measures were used in the transcription and coding of the narratives. The transcription of the narratives was carried out by the researcher and a research assistant. Intrarater reliability reached 97%, and interrater agreement, which was calculated by means of percentage agreement, out of a randomly selected sample of 10% percent of the data, reached 93.7%.

5. Results

Results in this section are presented according to the order of the hypotheses advanced in Section 3.3. Table 1 below shows the descriptive statistics for all the measures. After that, the effects of manipulating planning time on the three dimensions of production are presented first, and the effects of increasing complexity along the Here-and-Now variable are presented second. Section 5.3. considers the synergistic effects of manipulating both variables simultaneously.

(Table 1 around here)
5.1. Effects of increasing task complexity along planning time on production.

There was a reliable main effect for pruned speech rate $F(117,3) = 14.767$, $p < .01$, which suggests that fluency was affected by the different degrees of complexity (See Table 2). Post hoc Scheffe’s tests showed that both simple Here-and-Now and complex There-and-Then tasks generated significantly higher speech rate when performed under conditions of 10 minutes planning time. Planned Here-and-Now tasks triggered significantly more fluent speech ($p < .01$) than unplanned Here-and-Now tasks. There-and-Then tasks performed under planned conditions were also significantly more fluent ($p < .05$) than tasks performed under unplanned conditions. (See Table 3 and Figure 1).

Regarding lexical and structural complexity, there was a significant main effect for the Guiraud’s index of lexical richness $F(114,3) = 18.873$, $p < .01$ (See Table 2). Post hoc Scheffe’s tests showed that Here-and-Now tasks performed under planned conditions generated a significantly higher level of lexical richness ($p < .05$) than under unplanned conditions. There-and-Then narratives also generated a higher lexical richness ($p < .05$) when performed under planned conditions than under unplanned ones. Contrary to what Hypothesis 1 predicted, results show that increasing planning time generates significantly higher levels of lexical complexity (See Table 3 and Figure 2).

As for structural complexity, there was no significant main effect for structural complexity $F(123,3) = 1.711$, $p = .168$). In both simple and complex tasks, providing time caused a slightly higher level of structural complexity. Nevertheless, these differences were not significant for either Here-and-Now or There-and-Then narratives. It can therefore be concluded that none of the four combinations of the +/- planning and
Here-and-Now variables had any significant impact on structural complexity (See Table 2 Figure 3).

With regard to the measure of accuracy, there was a significant main effect for both the percentage of self-repairs $F(123,3) = 5.617, p<.01$. Simple tasks performed in the Here-and-Now with 10 minutes’ planning generated a slightly higher percentage of self-repairs than tasks performed with minimal planning time. This was similar for There-and-Then tasks, which again caused a lower proportion of self-repairs in unplanned tasks as compared to planned ones. None of these differences, however, reached statistical significance (See Table 2 and Figure 4).

(Tables 2 and 3 around here)

We can therefore conclude that the manipulation of planning time again had a significant impact on fluency and lexical complexity but not on structural complexity or accuracy, hence only partially confirming Hypothesis 1.

5.2. Effects of increasing task complexity along +/- Here-and-Now on production.

Hypothesis 2 was devised to investigate the impact of increasing complexity along the +/- Here-and-Now variable under both planned and unplanned conditions. It was predicted that such increase would reduce fluency but would have a positive impact on the complexity, both lexical and structural, and accuracy of learners’ production.

Regarding pruned speech rate, learners were significantly more fluent ($p<.05$) when narrating tasks in the Here-and-Now than when doing so in the There-and-Then under planned conditions. This behavior was the same when planning time was 50 seconds, which caused learners to be significantly more fluent ($p<.05$) when producing
Here-and-Now narratives than when producing narratives in the There-and-Then (See Table 4 and Figure 1).

With regard to lexical complexity, the results of the Guiraud’s Index of lexical richness show that complexity is reduced by increasing task demands along the +/- Here-and-Now variable, hence contradicting what was hypothesized in Hypothesis 1 (See Table 4 and Figure 2). As for structural complexity, this was the same between simple and complex tasks performed with 50 seconds’ planning time. Hypothesis 2, predicted that tasks in the There-and-Then would also generate higher levels of structural complexity than those in the Here-and-Now. Hypothesis 2 was not confirmed (See Table 4 and Figure 3).

The results for accuracy regarding Hypothesis 2 differ considerably from the ones obtained for Hypothesis 1. While providing time had a limited, non-significant effect on learners’ accuracy, increasing complexity along the +/- Here-and-Now variable had a strong, positive effect on learners’ accuracy. The percentage of self-repairs showed higher levels of attention to form when tasks were performed in the There-and-Then than when produced in the Here-and-Now. Hence, complex tasks in the There-and-Then triggered a significantly (p<.05) higher proportion of self-repairs than Here-and-Now tasks when performed after 10 minutes of planning. This was also the case when task demands were made higher by reducing planning time to less than a minute, which caused more episodes of self-repair when learners spoke in the past and without looking at the pictures than when narrating the stories in the Here-and-Now. Hypothesis 2 was therefore confirmed for accuracy (See Table 4 and Figure 4).

(Table 4 near here)
5.3. Effects of increasing task complexity simultaneously along planning time and +/- Here-and-Now on production.

Results show that reducing task complexity along resource-depleting dimensions and increasing it along resource-directing ones may have the largest benefits of all. This conclusion can be drawn if we compare Condition 3 to the three other conditions. While under Condition 1 attention was drawn towards fluency and lexical complexity, under Condition 3 fluency was significantly reduced but accuracy and lexical complexity were both attended to. As compared to Condition 2, in which lexical complexity and self-repairs were low, Condition 3 which combined increased planning time with increased displaced, past time reference, learners could use a variety of vocabulary while they also monitored their speech for errors. This is shown by the fact that both lexical complexity and accuracy were significantly higher under Condition 3 than under Condition 2. Finally, while Condition 4 shows similar levels of accuracy as Condition 3, lexical complexity was significantly lower.

6. Discussion

The first part of the discussion deals with the results of the different dimensions of production. It does so by considering production as affected by the different degrees of task complexity. The second part focuses on the issue of competition for attention during task performance.
6.1. Production results

Certainly, manipulating task features along both pre-task planning time and context-unsupported, displaced past time reference has displayed interesting synergies between the two variables. In general, increasing complexity along pre-task planning time brings learners closer to the ‘real’ processing conditions under which narrative discourse in conversation often takes place, that is, without previous preparation and often using displaced, past time reference. However, the results reported in this chapter lead us to draw the conclusion that these increases in complexity do not seem to direct learners’ attention to any particular grammatical features of their language. If, on the contrary, pre-task planning time is provided, as it is often the case in instructional contexts, learners will achieve a more lexically rich and varied performance as well as a more fluent one. On the other hand, increasing complexity along the +/- Here-and-Now variable makes learners reduce their rate of speech but also pushes them to focus on how they code their messages. Reducing complexity along this latter variable only produces a minor improvement in fluency.

Under Condition 1, that is, with the support of the context (i.e. comic strips) and in the present, pre-task planning time had the effect of ensuring a fluent and lexically complex performance but did not particularly draw the learners’ attention to how they encoded their messages. In this sense I would agree with Yuan and Ellis (2003:7) who conclude that: “It follows that pre-task planning does not greatly assist formulation, especially of grammatical morphology. Thus, the linguistic correlate of effort put into conceptualizing what to say is enhanced complexity and fluency rather than accuracy.” Furthermore, the fact that planning time improves lexical complexity but does not improve accuracy does not necessarily have to be explained in terms of limited attention
or competition for attention. It may also be explained by the fact that providing planning time *per se* does not guide learners towards more accurate speech. Although it has been shown that during pre-task planning time, attention to form can and does take place, the results described in this study as well as in some previous ones show that pre-task planning time does not focus learners on form during performance in any particular way.

For Condition 2, which keeps tasks simple at the level of contextual support and present time reference but makes them complex by reducing planning time, all dimensions of production are negatively affected, which confirms both Skehan’s and Robinson’s predictions. Speech rate is significantly slower than in that of its planned counterpart. If anything, and as shown by protocol analysis reports, under this condition, learners use their resources to find the words they need to communicate their message, which still generates a low percentage of lexical words and low ratio of lexical to function words, as well as a minimal level of monitoring and self-repairs.

Condition 3, which makes tasks simple along planning time but complex along displaced, past time reference, triggers lexically complex language as well as increased attention to form, with only fluency being affected negatively. Under planned conditions, the effects of increasing task complexity along the +/- Here-and-Now on fluency were stronger than when doing so along unplanned ones. This means that when comparing Condition 1 to Condition 3, in Condition 3 attention was geared towards both complexity and accuracy with negative consequences for fluency. Going from Condition 2 to Condition 4, that is, from an unplanned Here-and-Now task to an unplanned There-and-Then one, led learners only to focus on self-repairs under Condition 4, and hence the mean difference in fluency was lower than between planned tasks. This also conforms to Robinson’s predictions for monologic tasks, which
suggests that if tasks are kept simple along resource-dispersing variables (e.g. planning time) but are made more complex along resource-directing variables (e.g. +/- Here-and-Now), attention may be allotted to complexity and accuracy simultaneously. These findings are also coherent with those of Yuan and Ellis (2003).

Finally, Condition 4 makes tasks complex at both levels, which has negative effects for fluency and complexity and only has positive effects for monitoring of learners’ own speech. This conforms to both Skehan’s and Robinson’s prediction that if tasks are made more complex along planning time, attention will be drawn to either complexity or accuracy, but not both simultaneously.

6.2. The competition for attention issue

During the review of both planning time and Here-and-Now studies, we saw that trade-off effects have been hypothesized to exist between the different dimensions of production. Hence, Foster and Skehan (1996) and Skehan and Foster (1997) suggested that trade-off effects exist between accuracy and complexity, a conclusion that was also reached by Mehnert (1998). In both cases, it was concluded that any gains in complexity are achieved at the expense of accuracy and vice versa. Yuan and Ellis (2003) found that trade-off effects took place between fluency and accuracy.

First of all, despite the fact that the metaphor that has prevailed is one of limited capacity in which the three dimensions of production compete equally for attention, I would argue that fluency should be considered to be separate from complexity and accuracy. This research has shown that fluency is reduced when processing demands are high. If processing load is reduced, for example, by the effect of providing pre-task planning time, fluency increases. These two facts show fluency to be clearly sensitive to
processing. But the results of this experiment present quite a complex picture of how fluency interacts with complexity and accuracy, a picture which depends on the direction in which planning time is manipulated. When extended pre-task planning time is provided, significant gains are found for fluency and lexical complexity at the expense of attention to structural complexity or accuracy. When task demands are made higher by reducing planning time, resources do not seem to be employed to improve performance in any of the other areas, as seen by the results of lexical and structural complexity, and accuracy. Nevertheless, I would like to suggest that fluency does not require attention in the same way that complexity and accuracy do. Differences in fluency are the consequence of what happens with complexity and accuracy, but fluency does not diminish because of the lack of attentional control. In other words, higher fluency is not the consequence of attention allocation policies, as complexity and accuracy would be, but the consequence of more efficient message planning and faster lexical access and selection (Levelt, 1989).

Wickens (1989:73) suggests that when two tasks are being carried out simultaneously, confusion between the tasks may lead to poor performance. However, a number of strategies can be applied, and he adds: “One such strategy may be to continue to try to perform the two activities in parallel but at some lower rate of performance, perhaps by being more careful, in a manner that will slow the rate of responses. This strategy will be effortful and extract a toll on resources, but will reduce the degree of confusion that results.” In contrast, planning utterances, finding the words to express meaning, and coding them grammatically and phonologically, all involve conscious attention allocation, at least in the case of L2 speakers, with detrimental effects for fluency.
Secondly, I believe that the results obtained for Condition 3 disprove the argument that there are not enough attentional resources to focus on complexity and accuracy simultaneously. As we saw in the previous section, if tasks are made simple along resource-dispersing dimensions but made complex along resource-directing ones, a simultaneous focus on accuracy and complexity is possible. Even people who have drawn on multiple source models have come to the conclusion that complexity and accuracy compete for attention. Kormos (2000:348), who uses Wickens’ (1989) model as a reference, has argued that the different dimensions of production draw on the same resource pool: “Upon processing their speech, L2 learners need to rely on the same verbal resource pool; therefore the various phases of speech production need to compete with each other for attentional resources.” Maybe the explanation lies in the fact that the concept ‘resource pool’ has not been sufficiently defined. In this sense, Robinson (2003:646) indicated that: “codes would have to be representationally specified, as would resource pools.” I would add that, regarding performance, we also need to clarify the nature of the dimensions of accuracy and complexity. If we see them as just two dimensions of a single task (i.e. speaking) that draw on a single pool of resources and therefore compete, as Skehan (1998) and Kormos (2000) suggest, then we need to define what the limits of such capacity are and account for how time-sharing can be achieved, as in Condition 3 of this experiment where learners paid attention to both meaning and morphosyntactic forms. If we take a multiple-resource approach, it remains to be proven that what happens during speech is that accuracy and complexity are two tasks (i.e. in a dual-task conception like Wickens’, 1989) that draw on different resource pools and can therefore be attended to simultaneously.
One of the strongest arguments against the limited-capacity conception of attention has been advanced by Robinson (2003, 2005). Based on attention models that go beyond the limited capacity idea, Robinson suggests that (2003: 646):

[...] These trade-off effects (form vs. function, accuracy vs. fluency) may be better explained not in terms of a priori capacity limits on a single pool of attention but in terms of control functions during central processing (allocation policy, time constraints on scheduling attention allocation), and interference occurring during resource allocation to those specific task demands which central processing responds to. From the perspective of interference theory, explanations linking relative ease or difficulty of L2 comprehension, or different characteristics of L2 production, to task demands may be more legitimately framed in terms of confusion and cross-talk between codes (of L1, interlanguage, and L2 syntax, morphology, semantics, and phonology/orthography) within specific resource pools during task performance, rather than in terms of global capacity limitations.

As a way of conclusion, this study has been an attempt at understanding how different degrees of cognitive complexity affect narrative oral production. In my view, task complexity, both as operationalized in this study as well as in previous ones, stands out as a robust and testable construct for task and syllabus design. Findings obtained from task-based research on production and acquisition lend themselves to not just task-based syllabus construction but also to other approaches such as process or content-based teaching. Besides that, the experimental operationalization and manipulation of different task features can be easily transferred to pedagogic contexts in order to achieve specific effects on production and, possibly, learning.

Finally, further research should bring individual differences into the picture. Some task-based studies have started to cross complexity and learner factors in interesting
ways. For example, Niwa (2000) has looked into how individual differences such as aptitude, intelligence, and working memory can differentiate complex task performance. For a detailed review of such studies see Robinson (2005b).
7. References


Begin the story like this: YESTERDAY Mr. Brown was shopping in the supermarket. He was checking his shopping list and looking at prices. An employee was putting price tags on the products.
C-tests have been described as having high reliability and validity indices and have often been adopted as instruments for testing learner proficiency. See Klein-Braley (1997), Jafarpur (1999), and Babaii and Ansary (2001).

In a Latin square design, treatments are assigned at random within rows and columns, with each treatment once per row and once per column, in order to control for variation in two directions. In this case the two directions are condition and sequence (Steel & Torrie, 1980).


Other alternatives have been suggested more recently. Foster et al. (2000) suggest the AS-Unit (the Analysis of Speech Unit) which corrects some of the shortcomings of T-Units, C-Units, and utterances. In this study the T-Unit was chosen because of the one-way, non-interactional nature of the narratives used, and for the sake of comparability to previous studies.

Self-repairs exclusively refers to error-repairs in this study since no protocol analysis was used to neither detect different repairs nor appropriateness repairs (Kormos, 1999). Also, phonological self-repairs were not considered either, even if they are relatively easy to detect. The reason is to be found in the difficulty to reach an agreement between raters as to what constitutes a phonological error. An example of a lexical self-repair would be: “OK one day eh this eh man #men eh saw hair eh in in a comb a combing comb a combing combing comb REPAIR”. An example of a morphosyntactic self-repair would be: “eh the man go to eh go went the man went REPAIR in a a doctor eh in a doctor's room”.

The definition of error is adopted from Lennon (1991:182): “a linguistic form or combination of forms, which in the same context and under similar conditions of production would, in all likelihood, not be produced by the speakers’ native speaker counterparts”.

---

1 C-tests have been described as having high reliability and validity indices and have often been adopted as instruments for testing learner proficiency. See Klein-Braley (1997), Jafarpur (1999), and Babaii and Ansary (2001).

2 In a Latin square design, treatments are assigned at random within rows and columns, with each treatment once per row and once per column, in order to control for variation in two directions. In this case the two directions are condition and sequence (Steel & Torrie, 1980).


4 Other alternatives have been suggested more recently. Foster et al. (2000) suggest the AS-Unit (the Analysis of Speech Unit) which corrects some of the shortcomings of T-Units, C-Units, and utterances. In this study the T-Unit was chosen because of the one-way, non-interactional nature of the narratives used, and for the sake of comparability to previous studies.

5 Self-repairs exclusively refers to error-repairs in this study since no protocol analysis was used to neither detect different repairs nor appropriateness repairs (Kormos, 1999). Also, phonological self-repairs were not considered either, even if they are relatively easy to detect. The reason is to be found in the difficulty to reach an agreement between raters as to what constitutes a phonological error. An example of a lexical self-repair would be: “OK one day eh this eh man #men eh saw hair eh in in a comb a combing comb a combing combing comb REPAIR”. An example of a morphosyntactic self-repair would be: “eh the man go to eh go went the man went REPAIR in a a doctor eh in a doctor's room”.

6 The definition of error is adopted from Lennon (1991:182): “a linguistic form or combination of forms, which in the same context and under similar conditions of production would, in all likelihood, not be produced by the speakers’ native speaker counterparts”.