Notes for Exp 2:

***Non-linguistic cognitive tasks***

Following previous work on the relationship between language production and cognitive abilities, we included measures of sustained attention, working memory, and inhibition. The choice of these measures was informed by theoretical and empirical arguments (detailed below). We included additional tasks to be used in exploratory analyses, so as to allow a better comparison of our results with previous work, and to determine which measures to be used in subsequent studies.

Each participant performed the following tasks in this order: Orthographic span task, Symmetry span task, Rotation span task, Conjunctive continuous performance task, Stop signal task, Flanker task, Simon task, and the Continuous time expectancy task. These tasks were all implemented in OpenSesame (Mathot, Schreij, and Theeuwes, 2012). We provide a detailed description of each task below.

*Working memory*

Working memory is often evaluated using span tasks (e.g., Conway et al., 2005). In these tasks, participants are asked to remember target items (e.g., letters) whose presentation alternates with the presentation of a secondary and demanding task (i.e., distractor items, e.g., resolving arithmetic problems). As suggested by various authors (e.g., Loehlin, 1998, Engle, Tuholski, Laughlin, & Conway, 1999; Conway et al., 2005; Loehlin, 1998; Shipstead, Redick, & Engle, 2012), participants’ scores in these tasks are not independent from the intrinsic properties of the task. For instance, two participants with the same working memory capacities but different arithmetic skills obtain different scores in the operation span task, where the secondary tasks involves resolving arithmetic problems. To decrease the influence of the participant-specific skills in the secondary task, we administered three different span tasks, namely the operation span task, the symmetry span task, and the rotation span task and computed a composite score.

In implementing these tasks, we followed the procedure described in Unsworth et al. (2005), but used a shorter version (two blocks only, see Foster et al., 2015, for justifications)[[1]](#footnote-1). Each task consists in sequences of target (i.e., to be remembered) items and distractor items. Participants must provide an answer after each distractor item, and at the end of the sequence, must recall all target items, in the correct order. In the *operation span task*, a sequence consists of the alternative presentation of an arithmetic problem and a letter. The letters are the to-be-remembered items while the arithmetic problems are the distractor items. In the *symmetry span task*, a sequence consists in the alternative presentation of a geometrical shape that can be or not be symmetrical relatively to the vertical axis, and a 4x4 grid containing a single square. The positions of the squares in the 4x4 grid are the target items, while the geometrical shapes, symmetrical or asymmetrical, are the distractor items. After each geometrical shape, participants must determine whether the shape was symmetrical or not. After each sequence, participants are asked to recall, the positions of the squares in the 4x4 grid. In the *rotation span task*, a sequence consists in the alternative presentation of a rotated letter that can be vertically reversed or not and of an arrow pointing to one of eight possible directions. Arrows can be long or short. After each distractor item, participants must determine whether the letter was reversed or not and after each sequence, they must recall the lengths and directions of the arrows.

The length of a sequence is usually referred as the set size. Following Draheim, Harrison, Embretson and Engle (2018), set size varied between 3 and 8 items in the operation span task, between 2 and 6 items in the symmetry span task, and between 2 and 5 items in the rotation span task. Each span task had several blocks, with trials of various set size. For each task, participant, and block, a partial unit score was computed (PCU; Conway et al., 2005). We then averaged these measures over blocks and tasks to obtain a unique measure for each participant.

*Inhibition*

Inhibition is described as the ability to resolve conflicts and suppress irrelevant information. Recent studies suggest that inhibition is not a unitary construct (Rey-Mermet et al., 2018; Rouder and Haaf, 2019). Several authors distinguish for instance between selective and non-selective inhibition (e.g. Forstmann et al., 2008; Shao, Meyer and Roelofs, 2013; Shao, Roelofs, Martin and Meyer, 2015). Several authors further describe several types of selective inhibition, depending on the type of conflict that triggers inhibition (Hommel, 1997; Sceratti et al., 2017). Stimulus-stimulus conflicts are assumed to be a consequence of the incompatibility between overlapping relevant and non-relevant features of the stimulus while stimulus-response conflicts are assumed to occur as a consequence of the incompatibility between overlapping and incompatible features of the stimulus and response (Kornblum, 1994). In the present study, we used three different tasks to assess inhibition skills.

We used the Flanker task and the Simon task to assess non-selective inhibition. The Flanker task has been associated with a stimulus-stimulus conflict while the Simon task has been associated with a stimulus-response conflict (Hommel, 1997; Kornblum, 1994; Scerrati et al., 2017; Chmielewski and Beste, 2019). Several studies reported that performance in these two tasks do not correlate (Rey-Mermet et al., 2018; Keye, Wilhelm, Oberauer and van Ravenzwaaij, 2009), and we used both measures in our main analyses. In the Flanker task, an arrow pointing either to the left or to the right is presented at the center of the screen, and participants have to determine the direction of the arrow using the corresponding left or right keyboard key. This central arrow is flanked either by arrows pointing to the same direction (congruent condition), by arrows pointing to the opposite direction (incongruent condition), or by straight lines (neutral condition). Participants started with 18 training trials, six in each condition. There were four experimental blocks, each consisting in 46 congruent trials, 46 incongruent trials and 46 neutral trials (see Hedge, Powell and Sumner, 2018). For each participant, we divided the cumulative distribution of response times into five quantiles (i.e., delta plots, Ridderinkhof, [2002](https://link.springer.com/article/10.3758/s13421-013-0332-7#CR53); Shao, Meyer and Roelofs, 2013). We then computed the slope between the last two quintiles, which we used as predictors in our statistical model. The slope between the last two quintiles has been argued to represent the ability to efficiently deploy inhibition (Ridderinkhof, [2002](https://link.springer.com/article/10.3758/s13421-013-0332-7#CR53)).

In the Simon task, participants started with a familiarization phase of 40 trials. During this phase, a blue (50% of trials) or red circle appeared in the center of the screen. Participants had to press a key on the left hand side of the keyboard upon seeing a blue circle and on the right hand side upon seeing a red circle. They were asked to act as quickly as possible. The familiarization phase was followed by a practice phase with 56 trials. A trial started by the presentation of a fixation cross located at the center of the screen for 500 ms. Then a blue or red circle appeared either on the left-hand or right-hand side of the fixation cross. Trials with the blue circle appearing on the left-hand side or with a red circle appearing on the right hand side are called congruent trials, because the position of the circle and the correct response key are on the same side. Conversely, trials where the blue circle appears on the right-hand side or a red circle on the left-hand side are called incongruent trials. During this practice phase, there were 42 congruent trials (21 blue and 21 red circles), and 12 incongruent trials (6 blue and 6 red circles). After the practice phase, participants performed the same task in two experimental blocks of 120 trials each, with 90 congruent trials and 30 incongruent trials (see Wöstmann et al., 2013 on the selection of the number of trials). For each participant we divided the cumulative distribution of response times in the experimental blocks into five quantiles (Ridderinkhof, [2002](https://link.springer.com/article/10.3758/s13421-013-0332-7#CR53); Shao, Meyer and Roelofs, 2013). We then computed the slope between the last two quintiles, which we used as a predictor in our statistical model.

We additionally included the stop-signal task (Logan & Cowan, 1984), because performance on this task has been linked with performance in a language production task in a previous study (Shao, Meyer, & Roelofs, 2012). The task consists of two kinds of trials, go and no go trials. During the go trials, a fixation cross is displayed for 250 ms at the center of the screen and then replaced by either the symbol "<" or the symbol ">". Participants are instructed to press a left keyboard key after the onset of the "<" symbol and a right keyboard key after the onset of the ">" symbol. They are asked to perform the task as quickly as possible. During the no go trials, a tone (750Hz, lasting 75 ms) is displayed shortly after the onset of the visual symbols. For these trials, participants are instructed to inhibit their responses. The delay between the onset of the visual symbol and the onset of the tone is referred as the stop signal delay (SSD). At the beginning of the procedure, the SSD is set at 250 ms (i.e. the tone is displayed 250 ms after the onset of the visual symbol). The SSD is then adjusted depending on the ability of the participants to inhibit their responses for the no go trials. Following a successful inhibition, the SSD is increased by 50 ms. Following a failure of inhibiting the response, the SSD is reduced by 50 ms. Following the recommendations of Matzke et al. (2018), the task started with a familiarization block of 20 go trials. Then, participants had to perform a practice block consisting in 18 go trials and 6 no-go trials. For this practice block, participants were instructed to not slow down their responses at go trials (Matzke et al., 2018). Three experimental blocks followed, consisting of 48 go trials and 16 no go trials for each of them. We computed the stop signal reaction time (SSRT) using the integration method (Verbruggen et al., 2013) for each experimental block, and computed the mean of SSRT over the three experimental blocks for each participant. This measure was to be included in exploratory analyses rather than in the main analysis. We assume that non-selective inhibition is less likely to be engaged in language production processes than selective inhibition.

*Sustained attention*

Several tasks have been used to assess sustained attention in the literature. In their studies on the role of sustained attention in language production, Jongman and colleagues used the continuous performance task (Jongman, Meyer, et al., 2015) and the digit discrimination task and (Jongman, Roelofs, et al., 2015b). However, this task generates very few errors (for instance, Jongman, Roelofs, et al., 2015a reported 0.5% of false alarms, and 0.3% of misses in their experiment. We therefore selected a different task for our main analyses, the continuous time expectancy task (O'Connell et al. (2009; see also Irrmischer, van der Wal, Mansvelder, and Linkenkaer-Hansen, 2018). O'Connell et al. (2019) reported that only 64% of the target trial (hit) were correctly identified; therefore, ceiling effects are not a concern for this task.

In the continuous time expectancy task participants have to monitor a flow of alternating visual patterns and must detect a pattern that is presented for a longer duration than the other patterns (see Figure 2). We used the four visual patterns used in O'Connell et al. (2009). The standard duration in our task was set at 800 ms, while the target duration was set at 1300 ms. The procedure involved a familiarization phase, a practice phase, and an experimental phase. During the familiarization phase, 32 patterns appeared one by one on the screen. Four of these patterns had a duration of 1300 ms, while the other patterns had a duration of 800 ms. Participants had to detect the longer patterns. They could only take part in the subsequent phase once they had detected all four patterns. During the practice phase, they were instructed to press the space bar as soon as they noticed a pattern lasting longer than the other ones. There were 24 trials and among them, three longer trials (i.e., targets). If participants were unable to detect the 3 deviant patterns during the practice part, they had to perform it again until the detection of every deviant pattern. The experimental part consisted in 674 trials, with 84 deviant trials. Distances between two deviant trials ranged between 4 and 10 trials. Following the presentation of a target, participants had 700 ms to answer. We then computed the hit rate, which we used as a predictor in our main analysis.

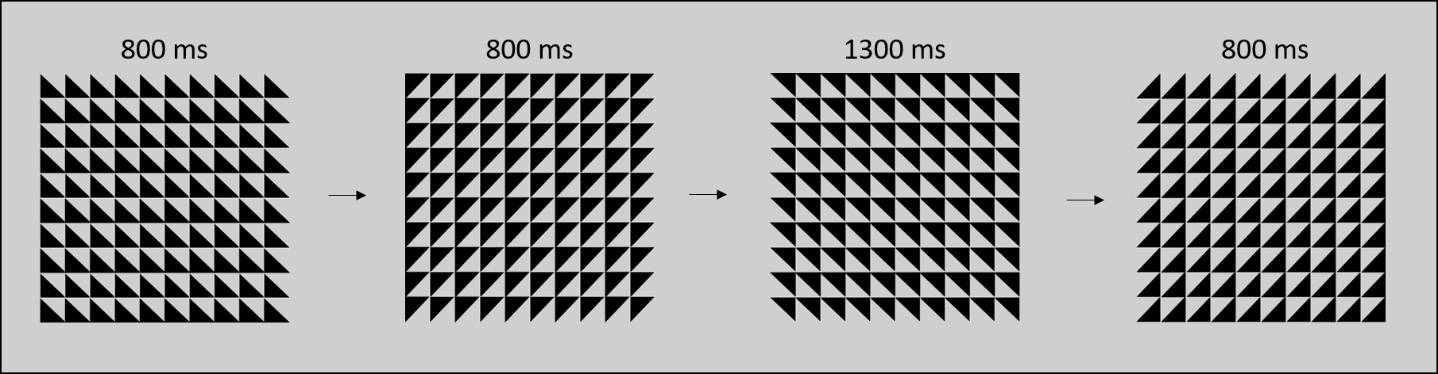


Figure 2: Stimuli used in the continuous time expectancy task. Adapted from O'Connel et al. (2009)

In order to be able to compare our results with previous findings (Jongman, Roelofs, et al., 2015b) we also asked our participants to perform a conjunctive continuous performance task (CCPT; Shalev et al., 2011). This task is similar to the CCP task used by Jongman et al. Unlike in the task used by these authors, however, participants have to attend to two dimensional features (i.e. shape and color) as opposed to only one. A flow a visual symbols is presented on the screen. These symbols vary in shapes (i.e. square, triangle, star, circle) and in color (i.e. blue, green, red, yellow). The task of the participants is to press a button as soon as they see a red square. Because participants have to pay attention to the conjunction of two dimensional features (i.e. shape and color), this task is thought to increase attentional demands as compared to classical continuous performance task. We followed the procedure depicted in Shalev et al. (2011), with participant performing a single block of this task. This block included 320 trials, with 30% of the trials corresponding to the target (i.e. the red square), red non square symbols appearing on 17.5% of the trials, square non red symbols appearing on 17.5% of the trials, and finally non-red and non-square symbols in the remainder of the trials. The inter-trial (between symbols) interval ranged between 1000 ms and 2500 ms by step of 500 ms. For each participant, we computed the hit rate.

*Cognitive functions and language production processes*

We observed that inter-individual differences in sustained attention, as measured in the CTET task, contributes to explaining inter-individual differences in naming times. This finding complements previous studies that reported a link between measures of sustained attention and response times in picture naming tasks. In previous work however, the link between sustained attention and response times was restricted to very slow trials (Jongman, Roelofs, & Meyer, 2015; Jongman, Meyer, et al. 2015). In the present study, we find evidence that the participants’ performance in the CTET tasks (also) influences the mean of the distribution. We speculate that this difference reflects a different sensitivity of the tasks used to assess sustained attention across studies. As discussed in the method section, the task used by Jongman and colleagues generates very few errors (for instance, Jongman, Roelofs, et al., 2015a reported 0.5% of false alarms, and 0.3% of misses in their experiment) suggesting that it does not involve a high attentional load.

Jongman and colleagues conclude that sustained attention is mostly involved after the phonological encoding process, presumably during the phonetic encoding and execution processes. As discussed above, in delayed naming tasks, the naming times reflect mostly execution processes. We reasoned that if the influence of inter-individual differences in sustained attention mostly influence the execution of articulatory gestures, the delayed naming times would capture this variability. The effect of sustained attention on naming times in the immediate task was not affected when the participants naming times in the delayed naming task were entered in the statistical model. Moreover, we found no correlation between the measure of sustained attention and response times in the delayed naming task. This suggests that the effect of sustained attention is not to be found in execution processes, but leave the question open as to whether it influences only the phonetic encoding process or higher level processes.

Our measures of working memory and inhibition, did not modulate the response times (or experimental effects). These are null-results and as such, cannot be taken to suggest that these cognitive functions are not involved in language production tasks. It is possible that our measures of inhibition or working memory do not tap into the working memory or inhibition abilities that are recruited during language production. The literature on inhibition, or on working memory highlights the poor correlation that usually exist between two tasks that are designed to tap into the same underlying processes, hence the same constructs (i.e., inhibition, working memory, attention). This observation is confirmed in the present study (see also Shao et al., 2014; Shao et al., 2013a on inhibition).

1. These tasks are available from the Georgia Tech Attention and Working Memory Lab website (<http://englelab.gatech.edu>), we implemented the scripts available on this website in OpenSesame (Mathot et al., 2012). [↑](#footnote-ref-1)