# The neostriatum and response selection in overt sentence production: An fMRI study

Georgios P. Argyropoulos [a,*], Pascale Tremblay [b], Steven L. Small [a]

[a] Department of Neurology, University of California Irvine, Irvine, CA, USA
[b] Rehabilitation Department, Université Laval, Centre de recherche de l'Institut universitaire en santé mentale de Québec, Quebec, Canada

## ARTICLE INFO

## ABSTRACT

A number of premotor and prefrontal brain areas have been recently shown to play a significant role in response selection in overt sentence production. These areas are anatomically connected to the basal ganglia, a set of subcortical structures that has been traditionally involved in response selection across behavioral domains. The putamen and the caudate, the two major inputs to the basal ganglia, have been shown to undertake motor— as well as non-motor-related selection operations in language processing. Here we investigate the role of these basal ganglia structures in sentence repetition and generation in healthy adults. Although sentence generation is known to activate prefrontal and premotor cortical areas that reciprocally connect with these two neostriatal structures, their specific contributions are not known. We present evidence suggesting that that the putamen undertakes articulation-related aspects across tasks, while the caudate selectively supports selection processes in sentence generation.

© 2013 Elsevier Inc. All rights reserved.

## Introduction

A fundamental aspect of spoken language production is selection, both of linguistic units (e.g., words) that convey a particular meaning and of sequences of motor programs that instantiate these units as articulatory gestures. Constraints in response selection may vary dramatically across tasks. In contrast to word repetition, for instance, where the linguistic response is externally pre-selected, word generation involves internally imposed constraints in the selection of the correct response among competing alternatives (e.g., Crosson et al., 2001). Despite the importance of this process, its neural underpinnings have not been well integrated into current neurobiological models of language (see Tremblay and Small, 2011b for references and discussion). Recent evidence suggests that a number of cortical areas engaged in the production of words and oral motor gestures, including the pre-supplementary motor area (pre-SMA), the ventral premotor cortex, and the pars opercularis and triangularis of the inferior frontal gyrus, are involved in selecting motor and/or lexical responses during word (Tremblay and Gracco, 2009a,b; Tremblay et al., 2008) or sentence production (Tremblay and Small, 2011b). A fundamental property of these areas is their participation in multiple segregated frontal–basal-ganglionic–thalamic loops (e.g., Middleton and Strick, 2000). Each loop includes projections from the cerebral cortex, through the basal ganglia (BG), to the thalamus, and back to the cerebral cortex. The neostriatum, consisting of the caudate nucleus and putamen, receives the main input from the cerebral cortex to the BG: the

putamen from motor and premotor cortices, while the caudate from various prefrontal structures (Hoover and Strick, 1999; Parent, 1990). The caudate and putamen each project to distinct segments of the medial globus pallidus, and, via projections to the thalamus, reach the cortical regions to which they are reciprocally connected. Both tract tracing studies in primates and non-invasive imaging in humans (e.g., resting-state functional connectivity, white matter tractography with diffusion tensor imaging) have shown that the pre-SMA as well as the dorsolateral and ventrolateral prefrontal cortices connect with the caudate head and the anterior putamen (i.e., associative cortico-striatal loop), while the motor and premotor cortices connect (primarily) with the posterior and dorsolateral anterior putamen (i.e., sensorimotor cortico-striatal loop) (e.g. Akkal et al., 2007; Chan et al., 2013; Di Martino et al., 2008).

The differentiation of these cortico-cortical-BG-thalamic loops is strongly suggestive of relative functional specialization within the BG, promoting the idea that different aspects of language processing rely more on certain BG components than others. However, while involvement of the BG in language is well established (e.g., Chan et al., 2013; Ketteler et al., 2008), its role remains unclear. Selective BG lesions do not consistently replicate classical aphasic symptoms (Crosson and Haaland, 2003; Crosson et al., 2007) and there is some thought that the resulting language deficits are more related to cortical hypoperfusion caused by the BG lesion than to the lesion per se (e.g., Hillis et al., 2002). Nevertheless, the architectural parallels among the different cortico-BG loops have suggested that the BG functions in a unitary fashion across behavioral domains. Two popular proposals on the role of the BG are (i) action selection among competing alternatives (e.g., Jueptner and Weiller, 1998); and (ii) suppression

of undesired actions and facilitation–initiation of desired ones (e.g., Gerfen, 1992). Studies on monolingual speakers have demonstrated BG involvement in the controlled process of syntactic integration (Friederici and Kotz, 2003; Friederici et al., 2003), while studies on bilingual speakers have highlighted the significance of the BG in second language comprehension and in the control of switching between languages (e.g., Abutalebi et al., 2000; Friederici, 2006; Lehtonen et al., 2005). The BG may thus play a role in cognitive control, assisting multi-level language processes by enhancing selected actions while suppressing competing ones (Crosson et al., 2003, 2007). Of particular relevance for sentence generation and production (Tremblay and Small, 2011b) is the recent finding implicating the left caudate in single word suppression (Ali et al., 2010).

In the present study, we aim to build on these results that have demonstrated (i) the roles of the BG in enhancement and suppression during single word processing and in cognitive control during sentence processing, and (ii) the reciprocal connectivity of the BG with cortical areas involved in response selection in overt production of both single words and sentences. The question that we address is whether structures of the neostriatum (caudate and putamen) are involved in the production of larger strings of words, such as sentences, in the same fashion as that they participate in single word selection and, more broadly, action selection. To this aim, we compare neostriatal activation during sentence repetition (externally constrained selection) with that during sentence generation (volitional selection). Based on the above, we hypothesize, first, that the caudate nucleus would be more active during sentence generation than repetition, given its involvement in the prefrontal-associative loop and its significance in aspects of response selection, cognitive control, and semantics; and second, that the putamen would be similarly active in both repetition and generation, based on its involvement in the motor-attentional cortico-striatal loop.

## Materials and methods

### Participants

The present study represents a reanalysis of data collected previously (Tremblay and Small, 2011a) and here we briefly repeat the methods that are described fully in that paper. Twenty-one healthy right-handed (Oldfield, 1971) native speakers of English (mean: 25 ± 4.4 years of age; 10 males) with a mean of 15.4 years of education participated. All had normal hearing, as assessed by normal pure-tone thresholds and normal speech recognition scores (92.3% accuracy on the Northwestern University auditory test number 6). The study was approved by the Institutional Review Board of the Division of Biological Sciences of The University of Chicago.

### Experimental procedures

Participants completed 5 different tasks inside the scanner: (1) passive sentence listening, (2) passive picture observation, (3) sentence repetition, (4) sentence generation, and (5) passive observation of short action movies. The individual trials for each of these tasks were grouped together in separate runs and, within each of these runs, experimental trials were alternated with periods of "rest" during which participants were asked to relax. For each run, the order of the conditions and number of rest trials was optimized using OPTseq2 (http://surfer.nmr.mgh.harvard.edu/optseq/). In the present article, we focus on the first four tasks.

A detailed description of stimulus construction and presentation may be found in Tremblay and Small (2011a,b). The sentence listening run consisted of 110 total trials, including 80 active auditory sentence listening trials (0.9 to 1.3 s each) and 30 visual fixation control trials (crosshairs) in a pseudorandom sequence. Half of these sentences described manual object-directed actions and the other half described

visual properties of the same set of objects. The sentence stimuli were presented while the MRI gradients were shut off, which ensured ease of auditory processing for participants ("sparse sampling" MRI acquisition (Gracco et al., 2005)). The picture observation run involved 77 trials, consisting of 40 simple black-and-white line drawings and 37 visual fixation control trials (1 s each). The pictures represented common man-made objects selected from the International Picture Norming Project corpus from the Center for Research in Language at the University of California San Diego (Bates et al., 2003; Szekely et al., 2003). Participants were instructed to attend to the pictures. The sentence repetition run consisted of 110 trials, including 80 auditory sentence trials (40 action, 40 object sentences) and 30 visual fixation control trials; participants were instructed to repeat the sentence. Both stimulus presentation and response occurred while the gradients were shut off for a 4.5 second period of silence. At the beginning of the silent interval, a 'Go' cue was presented, instructing participants to start repeating the sentence. All responses were recorded. The sentence generation run consisted of 108 trials. In the 80 active trials, participants were asked to generate sentences (40 action, 40 object), with 28 visual fixation trials pseudorandomly interspersed in the run. In each experimental trial, a picture was presented for 1 s and was followed, after 500 ms, by a visual 'Go' cue instructing participants to start generating a sentence. All responses occurred while the MR gradients were shut off. The listening and picture observation tasks provided control conditions for the sentence repetition and sentence generation tasks, respectively.

### Image acquisition and analysis

#### Image acquisition

Functional data were collected on a 3 T General Electric Signa HDx MRI scanner with EXCITE parallel acquisition capability. Subjects wore MR-compatible headphones and goggles (Nordic NeuroLabAudio/Visual system). 34 axial slices (3.125 mm × 3.125 mm × 3.6 mm, no gap, FOV = 256 mm × 256 mm, matrix = 64 × 64) were acquired in 1.5 s using a multi-slice EPI sequence with parallel imaging (ASSET = 2; TE = 26 ms; FOV = 20 cm; 64 × 64 matrix; Flip angle: 73). A sparse image acquisition technique (Gracco et al., 2005) was used for the three language tasks (sentence generation, repetition, and listening), to eliminate movement artifacts associated with speaking, and to ensure satisfactory audition. A silent period (1.5 s for listening, 4.5 s for repetition and generation) was interleaved between each volume acquisition. High-resolution T1-weighted volumes were also acquired for anatomical localization.

#### Time series pre-processing

We first segmented each individual's high-resolution structural image, using the FreeSurfer parcellation of white and gray matter (e.g. Dale et al., 1999). The functional images were co-registered to each other and then to the structural volume (Saad et al., 2009), and the functional data were motion-corrected (within and across runs), de-spiked, and mean-normalized using AFNI (Cox, 1996). A linear least squares model was used to establish a fit to each time point of the hemodynamic response function for each condition. We modeled the entire trial duration (i.e., 6 s), which included stimulus presentation and speech production.

#### First level (subject) analysis

Event-related signals were deconvolved by linear interpolation, beginning at stimulus onset, and continuing for 12 s, using AFNI's tent function (i.e., a piecewise linear spline model). For sentence generation and sentence repetition, we examined the fit at two different time lags (0–6 s, and 6–12 s) to identify the time point showing the strongest hemodynamic response both across the brain as well as in all of our regions of interest (left and right caudate and putamen; see First level (subject) analysis section below). All subsequent analyses focused on the activation from the first 6 s post-stimulus onset.