

## ORIGINAL ARTICLE

# Commonalities and Asymmetries in the Neurobiological Infrastructure for Language Production and Comprehension

Laura Giglio<sup>1,2</sup>, Markus Ostarek<sup>1,2</sup>, Kirsten Weber<sup>1,2,†</sup> and Peter Hagoort<sup>1,2,†</sup>

<sup>1</sup>Max Planck Institute for Psycholinguistics, 6525 XD Nijmegen, The Netherlands and <sup>2</sup>Donders Institute for Cognition, Brain and Behaviour, Radboud University, 6525 AJ Nijmegen, The Netherlands

Corresponding authors: [laura.giglio@mpi.nl](mailto:laura.giglio@mpi.nl) and [peter.hagoort@mpi.nl](mailto:peter.hagoort@mpi.nl), Wundtlaan 1, 6525 XD Nijmegen, The Netherlands, +31 (0)24-3521911

<sup>†</sup>Kirsten Weber and Peter Hagoort shared last authorship.

## Abstract

The neurobiology of sentence production has been largely understudied compared to the neurobiology of sentence comprehension, due to difficulties with experimental control and motion-related artifacts in neuroimaging. We studied the neural response to constituents of increasing size and specifically focused on the similarities and differences in the production and comprehension of the same stimuli. Participants had to either produce or listen to stimuli in a gradient of constituent size based on a visual prompt. Larger constituent sizes engaged the left inferior frontal gyrus (LIFG) and middle temporal gyrus (LMTG) extending to inferior parietal areas in both production and comprehension, confirming that the neural resources for syntactic encoding and decoding are largely overlapping. An ROI analysis in LIFG and LMTG also showed that production elicited larger responses to constituent size than comprehension and that the LMTG was more engaged in comprehension than production, while the LIFG was more engaged in production than comprehension. Finally, increasing constituent size was characterized by later BOLD peaks in comprehension but earlier peaks in production. These results show that syntactic encoding and parsing engage overlapping areas, but there are asymmetries in the engagement of the language network due to the specific requirements of production and comprehension.

**Key words:** syntax, speaking, sentence, fMRI, constituent structure

Since the association of lesions in the left inferior frontal gyrus (LIFG) and aphasia in the nineteenth century, scientists have tried to understand the relationship between the language faculty and the brain. Early reports called the LIFG or Broca's area a "speech movement centre" and the left superior temporal gyrus (LSTG) or Wernicke's area a "sensory speech centre" (from Wernicke 1892, as described in Levelt 2013). Since then, the field moved forward from a production-comprehension dissociation to the understanding that both areas are critical for language more generally, and that they do not subserve strictly segregated receptive or productive linguistic functions (Tremblay and Dick 2016). A wealth of neuroimaging studies and lesion-symptom mapping studies advanced the characterization of brain

function greatly, which resulted in a general understanding of the contributions of core regions in the language network (e.g., Friederici and Gierhan 2013; Hagoort and Indefrey 2014; Price 2012; Wilson 2017).

The LIFG (i.e., Broca's area and adjacent cortex) has been implicated in sentence-level processes in several neuroimaging studies. These included sentence vs. word list comprehension (Snijders et al. 2009; Fedorenko et al. 2012; Matchin et al. 2017; Zaccarella, et al. 2017b), phrase structure building (Schell et al. 2017; Zaccarella, et al. 2017a; Pallier et al. 2011; Chang et al. 2020), compositional processes in naturalistic language comprehension (Henderson et al. 2016; Bhattasali et al. 2019), and processing of noncanonical sentence structure

(Bornkessel-Schlesewsky et al. 2009; Santi and Grodzinsky 2010; Hirotani et al. 2011; Mack et al. 2013; Europa et al. 2019). In different neurobiological models of language processing the LIFG was thus proposed to have a role in combinatorial (Unification) processes in multiple domains of language and cognition (Hagoort 2005; Hagoort 2013; Hagoort 2019; Jackendoff and Audring 2020); in processing complex syntax (Friederici 2012); or in sentence processing due to its role in working memory (Matchin, 2018; Rogalsky et al. 2008).

Within the temporal lobe, posterior regions have been implicated in several aspects of comprehension, from auditory to phonological and morphological processing along the superior temporal gyrus and sulcus (Hickok and Poeppel 2007; Friederici 2012; Lee et al. 2018). In addition, the posterior middle temporal gyrus (LpMTG) has been associated with the retrieval of lexical-syntactic frames (“Memory” processes, Hagoort, 2005, 2013) and syntactic processes (Flick and Pykkänen 2020; Matchin and Hickok 2020). The anterior temporal lobe (ATL) has been associated with conceptual operations (e.g., Bemis and Pykkänen 2013; Boylan et al. 2017), also based on findings of ATL atrophy leading to semantic dementia (Wilson et al. 2013; Mesulam et al. 2014; Lambon Ralph et al. 2017).

All studies mentioned above, however, are based on linguistic processes in *comprehension*. The involvement of the main nodes of the language network in sentence *production* is less clear. This is mainly for two reasons: (i) the challenge of achieving good experimental control in sentence production studies, which also limited psycholinguistic studies of production processes (Bock 1996), and (ii) the obstacle of motion artifacts in neuroimaging as a consequence of movement during speech, which is however not impossible to overcome with state-of-the-art neuroimaging techniques (Willems and van Gerven 2018). These methodological difficulties have led to far fewer studies on the characterization of brain involvement in production than comprehension. As a consequence, the few meta-analyses that attempted to characterize the language network in the two modalities were severely underpowered in production (Walenski et al. 2019; Indefrey 2018).

The network obtained by meta-analyses of sentence production studies does not fully or consistently overlap with the sentence comprehension network discussed above. Sentence production studies found activity within a left-lateralised fronto-temporal network but not consistently across studies (Indefrey et al. 2001; Indefrey et al. 2004; Haller et al. 2005; Kircher et al. 2005; Golestani et al. 2006; den Ouden et al. 2008; Grande et al. 2012; Collina et al. 2014; Humphreys and Gennari 2014; Pykkänen et al. 2014; Thothathiri and Rattiner 2015; Matchin and Hickok 2016; Thothathiri 2018; Takashima et al. 2020). A recent meta-analysis on some of these studies found left middle frontal gyrus, LpMTG and lateral occipital cortex to be reliably involved in sentence production, but did not find evidence for LIFG involvement (Walenski et al. 2019). Another meta-analysis, instead, found the LIFG to be the only area reliably active in sentence production and for syntactic contrasts across some of those studies, thus lacking temporal lobe involvement (Indefrey 2018). There is thus disagreement on whether inferior frontal areas or temporal areas are reliably engaged during sentence production, while they are both reliably found in sentence comprehension. The contradictory results of these meta-analyses show that more work is needed to robustly determine the neural correlates of sentence production. Interestingly, these results suggest there are some discrepancies in the networks engaged by linguistic processes

in production and comprehension that raise the question whether the same neural resources are used in production and comprehension.

An important line of work addressed the question of a shared or distinct neural infrastructure between sentence production and comprehension. This question is relevant in the context of a long-standing debate in psycholinguistics. There are different views on if, and to what degree, production and comprehension share phonological, lexical, syntactic and semantic representations (Meyer et al. 2016; Momma and Phillips 2018; Phillips 2013; Gambi and Pickering 2017). Support for distinct representations comes from the production/comprehension asymmetries in language acquisition. Comprehension is seen to precede production in many linguistic domains, with some exceptions (Clark and Hecht 1983; Hendriks and Koster 2010). This dissociation in acquisition is more easily accounted for by models that keep production and comprehension representations separate. Accounts of syntactic deficits in agrammatic patients also suggest that different processes are compromised in comprehension and production (i.e., tree pruning in production vs. trace deletion in comprehension) (Grodzinsky 2000). However, there are also views supporting a single processing mechanism that argue that the differences between modalities may be superficial and may instead reflect input differences (Momma and Phillips 2018). Behavioral evidence has shown that syntactic representations are shared between production and comprehension (Kempen et al. 2012). Also, repetition suppression (Grill-Spector et al. 2006) was used in fMRI to understand which areas adapt to the repetition of linguistic material and whether the adaptation occurs only within one language modality or also across modalities (i.e., from comprehension to production and vice versa). The LIFG, precentral gyrus, LMTG and inferior parietal lobule were found to adapt to syntactic and lexical repetition across sentence production and comprehension, suggesting that production and comprehension share neural resources (Menenti et al. 2011; Menenti et al. 2012; Segaert et al. 2012; Segaert et al. 2013). This evidence for shared resources in production and comprehension is, however, challenged by the inconsistent and partly contradictory neuroimaging results in production.

In the current study, therefore, we examined the sentence production network in a high-powered study with the aim to further clarify the brain organization of sentence production. To address this issue, our study investigated language production in analogy to a seminal study on constituent structure building in comprehension (Pallier et al. 2011). Constituents are the syntactic building blocks of sentences. By using a constituent size manipulation, we could focus on the processes that allow for encoding of increasingly larger structures, while keeping lexico-semantic, phonological and articulatory processes constant between conditions. Following Pallier et al. (2011), we expected neural activity to gradually increase with the addition of each new node to the constituent structure of the stimuli. We used visual prompts to elicit the production of utterances that had three levels of constituent structure which differed in complexity. The simplest one consisted of one- and two-word sequences; the intermediate condition consisted of intransitive sentences; the version with the most complex structure had participants produce a sentence with a complementizer phrase embedded in the main clause. In their comprehension study, Pallier et al. (2011) showed that LIFG and the left anterior and posterior temporal lobe were responsive to constituent size. Based on previous comprehension evidence we therefore expected to find

Table 1 Example sentences used for each condition

Condition	Stimuli (in Dutch)	English translation
C1	klappen, slapen, de jongen, het meisje	clap, sleep, the boy, the girl
C2	de jongen slaapt, het meisje praat	the boy sleeps, the girl talks
C4	de jongen hoort dat het meisje klapt	the boy hears that the girl claps
Filler	de man helpt de vrouw	the man helps the woman

a gradual involvement of at least LIFG and LMTG with increasing constituent size.

Additionally, we directly compared the sentence production and comprehension networks with the aim to further clarify to what extent they overlap. Few studies so far used both production and comprehension in the same experiment, including a direct comparison between modalities. In particular, it is still unclear whether sentence production and comprehension rely on core regions of the language network to the same or to a different extent. Humphreys and Gennari (2014) found that frontal and subcortical regions were more engaged in production, while the LpMTG was more engaged in comprehension. Indefrey et al. (2004) found the LIFG to be responsive to syntactic processing in production but not in comprehension. Matchin and Wood (2020) instead found similar activity in LIFG for syntactic production and comprehension, and larger activity in LMTG for syntactic comprehension than production. We therefore selected LIFG and LMTG as regions of interest to better characterize their involvement in sentence processing across modalities. In short, there is no clear answer to the question whether production and comprehension recruit frontal and temporal regions similarly or differently. In this study we attempted to answer this question.

## Materials and Methods

### Participants

Forty-six right-handed native Dutch participants (28 females, mean = 23.8 years, range 19–35 years) participated in the experiment in return for monetary compensation after giving written informed consent. The study was approved by the ethical committee for Region Arnhem-Nijmegen. Participants had no history of neurological or language-related disorders, and reported having normal or corrected-to-normal vision and hearing. Six participants were excluded for the following reasons: technical problems during preprocessing of the MRI data ( $n=1$ ); failing to complete the experiment ( $n=2$ ); too many motion artifacts ( $n=3$ ). Forty participants were included in the analyses. This number was based on an *a priori* power calculation for the detection of an effect for the production of passive vs. active sentences in the LIFG and LMTG in a previous study (Segaert et al. 2012), using fMRIpower (Mumford and Nichols 2008). Even though the specific manipulation was different, it allowed us to estimate the number of participants needed for the detection of a syntactic effect in production in the two regions of interest.

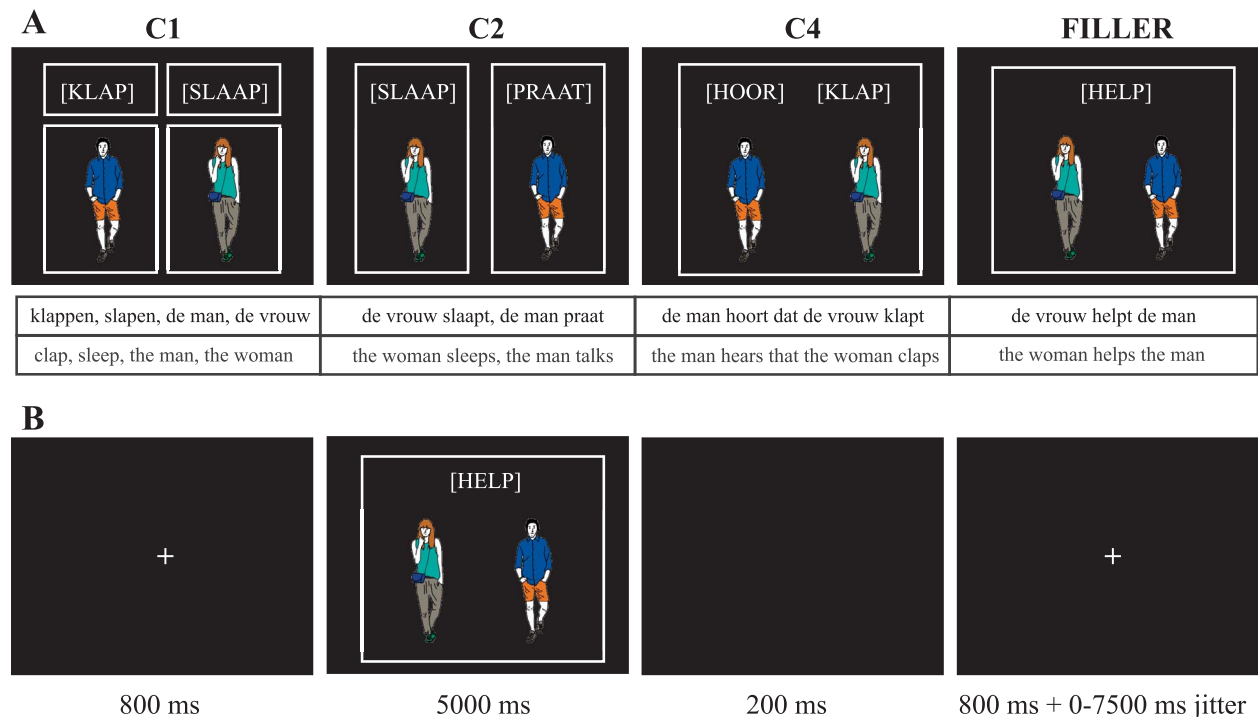
### Materials

In our study, we had three levels of constituent structure (see Table 1). The condition with the smallest constituent size (C1) consisted of two verbs and two noun phrases leading to four constituents with one (content) word (C1: “klappen, slapen, de jongen, het meisje”, “to clap, to sleep, the boy, the girl”). The intermediate condition (C2) involved the combination of a verb

and a noun phrase leading to two constituents with two content words forming intransitive sentences (C2: “de jongen slaapt, het meisje praat”, “the boy sleeps, the girl talks”).<sup>1</sup> The most complex sentence condition (C4) consisted of the combination of the four content words into a complementizer phrase embedded in the main clause (C4: “de jongen hoort dat het meisje klapt”, “the boy hears that the girl claps”). Critically, the conditions were almost identical in the total number of words to be produced, but they differed in constituent structure. The embedded sentence condition included the additional word “dat” (*that*), which in Dutch is obligatory in complementizer sentences.<sup>2</sup> We did not expect function words to affect sentence planning but they might involve articulation-related processes (Ferreira 1991). An additional filler condition was added to avoid too many verb repetitions. Filler sentences consisted of a sentence with one transitive verb (“de man helpt de vrouw”, “the man helps the woman”).

To induce the production of the sentences in the different conditions, participants were shown pictures with written verbs (see Fig. 1). Crucially, the conditions differed in the configuration of boxes around the verbs and the pictures of human figures. The boxes instructed the participants about the production output that was expected. In condition C1, there were four boxes, one around each item, signaling that the production of four separate items was expected. In this condition the actors and verbs should not be combined to form a sentence. In condition C2, there were two boxes, each around a verb and an actor, indicating that two separate sentences had to be produced. In condition C4, there was one box around all items on the screen, indicating that one single sentence was expected with the first verb heading an embedded clause formed by the second noun and verb. For filler sentences, there was only one box around all items on the screen. In this case there was only one verb, indicating that a transitive sentence was expected. Participants had no problems understanding the task and producing the correct output. By eliciting sentence production in this way, we could minimize the visual differences between conditions: pictures or videos would lead to very large differences in the visualization of word sequences vs. complementizer phrases. This type of speech elicitation paradigm is not unusual in the neuroimaging sentence production literature (cf. Matchin and Hickok 2016; Takashima et al. 2020).

- 1 The one provided is a literal translation of the Dutch sentence. A more natural translation would be “the boy is sleeping, the girl is talking”, since the present tense in Dutch is also used for continuous events.”
- 2 The example sentence provided “the boy hears that the girl claps” would also work with a bare infinitival complement construction (“the boy hears the girl clap”, “de jongen hoort het meisje klappen”), but many other verbs used in condition C4 do not support this construction (e.g., “klagen”, to complain, “dromen”, to dream). The word “dat” was thus required for all sentences in condition C4.



**Figure 1.** Stimulus presentation. **A:** Example of the screen that participants would see for each condition (identical in production and comprehension) with the corresponding expected output. The boxes clarified the type of output that was required. **B:** Screen sequence for each trial. The length of the fixation cross presentation was based on jittering optimized for contrast detection. In comprehension, during picture and verb presentation, a sound recording of the sentence started after 1000 ms.

The verbs were always presented in their root form, so that the production of the syntactically correct inflections was required in all conditions. In the C1 condition, the verb had to be produced in its infinitival form (generally, by addition of “en”: *help* to *helpen*); in the other conditions, the verb had to be inflected in the third person singular of the present tense (generally, by addition of “t”: *help* to *helpt*).

Since verbs allowing for a complementizer phrase (CP-verbs) and intransitive verbs (INT-verbs) are inherently different in their use and meaning, we selected a few verbs of each type that were repeated 8 times across the experiment. The verbs were matched in frequency (mean  $\pm$  std: INT-verbs =  $1.38 \pm 0.88$ , CP-verbs =  $1.46 \pm 0.77$ ,  $t = 0.59$ ,  $p = 0.56$ ) based on SUBTLEX-NL values (Keuleers et al. 2010), and concreteness (mean  $\pm$  std: INT-verbs =  $3.26 \pm 0.67$ , CP-verbs =  $3.21 \pm 0.47$ ,  $t = 0.27$ ,  $p = 0.79$ ) (Brybaert et al. 2014). Each condition consisted of 80 trials. In C4, we used 20 CP-verbs, repeated 4 times. The CP-verbs were always in first position to allow for the embedded sentence production; each of the 40 INT-verbs in C4 was repeated twice. In C2, we used the same 40 INT-verbs, each presented twice in first position and twice in second position. In C1, there was always one CP-verb and one INT-verb, with alternating first and second positions. Each CP-verb was repeated 4 times in this condition, and each INT-verb was repeated twice. The filler verbs consisted of 80 transitive verbs, each shown only once. We created 4 lists of stimuli that consisted of the same verb combinations for each condition, but for each list the verb was paired with a different picture. Across lists each verb combination was paired with each actor. The actors could be “the boy”, “the girl”, “the man”, “the woman”, with each presented 160 times in total.

In addition to the production condition, we included a comprehension condition that included half of the materials used for production but from a different list (hence with different actor-verb pairings). In the comprehension condition, each verb was repeated only 4 times in total, with 40 trials per condition. Instead of producing the sentences, participants had to listen to recorded stimuli, which started 1 sec after picture onset and lasted a maximum of 4 seconds (mean duration (in seconds): C1 = 3.14, C2 = 2.46, C4 = 2.46, Fillers = 1.79). The absence of an explicit task during the comprehension runs kept the production and comprehension runs as similar as possible without the introduction of effects unrelated to constituent size.

### Experimental Procedure

The experiment started with a behavioral practice session to familiarize participants with the task. They read instructions for each condition and had to practise producing the sentences. The experimenter gave feedback to make sure that the participant understood the task correctly. After the practice session was concluded, the fMRI experiment started.

The production lists were divided into 8 runs of 40 trials, each including 10 trials per condition, with as few verb repetitions as possible (per block 5–6 verbs were repeated once out of the 60 verbs presented (excluding fillers)). The comprehension lists were divided into 4 runs of 40 trials each. Production and comprehension runs alternated with two production runs always followed by a comprehension run. There were 12 acquisition runs in total. Each run lasted about 5 minutes. A fixation cross was presented for at least 800 ms before the picture screen was presented (Fig. 1). Participants had 5 seconds



to produce the answer. This was followed by a blank screen for 200 ms. We jittered the onset of trials by 0–7500 ms (mean 1500 ms), by varying the length of presentation of the fixation cross. The order of conditions and length of jitter was based on design optimization for contrast detection, made with optseq2 (Dale 1999).

### fMRI Acquisition

MR data were acquired in a 3 T MAGNETOM PrismaFit MR scanner (Siemens AG, Healthcare Sector, Erlangen, Germany) using a 32-channel head coil. The MRI protocol included a T1-weighted MRI scan for anatomical reference and several fMRI scans. The T1-weighted scan was acquired in the sagittal orientation using a 3D MPRAGE sequence with the following parameters: repetition time (TR)/inversion time (TI) 2300/1100 ms, echo time (TE) 3 ms, 8° flip angle, field of view (FOV) 256 mm × 216 mm × 176 mm and a 1 mm isotropic resolution. Parallel imaging (iPAT = 2) was used to accelerate the acquisition resulting in an acquisition time of 5 min and 21 sec. Whole-brain functional images were acquired using a multi-band (accelerator factor of 3) multi-echo T2\*-weighted sequence with the following parameters: TR 1500 ms, TEs 13.4/34.8/56.2, flip angle 75°, FOV 84 mm × 84 mm × 64 mm, voxel size 2.5 mm isotropic. Fieldmap images were also acquired to correct for distortions. We acquired 12 fMRI runs per participant.

### Data Analysis

#### Behavioural Analysis

Speech output in the production fMRI runs was analyzed for accuracy and response times. A Dutch native speaker rated the speech for accuracy. Speech was considered correct when the correct actors and determiners were used, the verb was inflected in the correct way, and the correct sentence structure was used. Self-corrections and word repetitions during hesitations were considered as errors. Speech onset and offset times were coded using Praat, after scanner noise removal. We analyzed onset and durations with linear mixed-effects models (Pinheiro and Bates 2000; Bates et al. 2015) and accuracy data using mixed-effects logit models (Jaeger 2008) with the *lme4* package (version 1.1–21, R version 3.6.2). We used the maximal effect structure that allowed for convergence (Barr et al. 2013). For accuracy, the model contained the factor Condition (C1, C2, C4) and by-participant and by-item (specifically, verbs) random intercepts. For onset and duration analysis, the model contained the factor Condition and by-participant random slopes for Condition and by-item random intercepts, with log-transformed onset and duration times.

#### fMRI Preprocessing

Preprocessing was performed using fMRIPrep 1.2.6–1 (Esteban et al. 2018; Esteban et al. 2018). The T1-weighted (T1w) image was corrected for intensity non-uniformity and skull-stripped. Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, Dale, et al. 1999). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al. 2009) was performed through nonlinear registration using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid, white-matter and gray-matter was performed on the brain-extracted T1w using fast (FSL 5.0.9, Zhang et al. 2001).

For each of the BOLD runs per subject, the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. A deformation field to correct for susceptibility distortions was estimated based on a field map that was co-registered to the BOLD reference, using a custom workflow of fMRIPrep. Based on the estimated susceptibility distortion, an unwarped BOLD reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using *bbregister* (FreeSurfer). Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using *mcflirt* (FSL 5.0.9, Jenkinson, et al. 2002). BOLD runs were slice-time corrected and resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. Multi-echo combination was performed by estimating a T2\* map from the preprocessed BOLD by fitting to a monoexponential signal decay model with log-linear regression. For each voxel, the maximal number of echoes with reliable signal in that voxel were used to fit the model. The calculated T2\* map was then used to optimally combine preprocessed BOLD across echoes following the method described in (Posse et al. 1999). Estimation of motion artifacts using independent component analysis (ICA-AROMA, Pruim et al. 2015) was performed on the preprocessed BOLD on MNI space time-series after removal of non-steady state volumes and spatial smoothing with an isotropic, Gaussian kernel of 6 mm FWHM (full-width half-maximum). The AROMA noise-regressors were later used as confound regressors. The BOLD time-series were resampled to MNI152NLin2009cAsym standard space. Confounding time-series were calculated based on the preprocessed BOLD for framewise displacement (FD) and DVARS (following the definitions by Power et al. 2014). We excluded subjects that had FD values above 2.5 (these were also the subjects that showed highest mean FD and the largest number of volumes with FD values above 1). Additionally, a set of physiological regressors were extracted to allow for anatomical component-based noise correction (aCompCor, Behzadi, et al. 2007).

#### Motion-Related Correction

To prevent excessive motion artifacts due to speaking out loud, participants' heads were secured in a pillow and a tape was attached across their foreheads to provide them with feedback in case of movement, which was shown to reduce motion (Krause et al. 2019). In addition, subjects with FD values above the voxel size were excluded. ICA-AROMA was used to estimate components related to motion that were later added as nuisance regressors together with motion parameters, FD, DVARS and aCompCor in the first-level design matrix.

#### fMRI Analysis

##### Whole-Brain Analysis

We used the non-denoised preprocessed BOLD images in MNI152NLin2009cAsym standard space for first-level single-subject analysis. We applied spatial smoothing with an isotropic Gaussian kernel of 4 mm FWHM in SPM12 in Matlab2019a. For the production runs, we computed a general linear model (GLM)

in SPM12 with the following condition regressors: correct trials for each of the four conditions, all incorrect trials, temporal derivative, and parametric modulations of speech onset times. For the comprehension runs, the GLM was identical except for the absence of an incorrect trial regressor and parametric modulations. The onset of each trial was set as the picture onset time, and trial duration was set as time until speech offset, hence accounting for differences in duration between individual stimuli and conditions. In addition, we added confound regressors that were computed in fMRIPrep. We included regressors for DVARS, Framewise Displacement, 6 aCompCor parameters and 6 motion parameters. Finally, we added the AROMA noise components computed in fMRIPrep as additional nuisance regressors, to perform non-aggressive denoising. Contrast images for the main effect of constituent size (with weights  $[-4 -1 5]$  based on constituent size of C1, C2 and C4, respectively), main effect of modality (production vs. comprehension) and interaction between constituent size and modality were computed for each participant. For the main effect of constituent size we selected a numerical linear contrast based on Pallier et al. (2011) that reflects activation with a linear increase according to the number of words integrated in a constituent: C1=one content word per constituent, C2=2 content words, C4=4 content words. This led to a contrast with weights  $[-4 -1 5]$  after mean-centering. By contrasting the three conditions together, the results were less sensitive to other types of differences between individual conditions (e.g., the contrast C4 vs. C2 might be sensitive to verb argument structure differences). The contrast images were tested with a one-sample T-test at the group level following Henson (2015). We thresholded brain responses at the voxel-level at  $p=0.001$  uncorrected, and then used  $p=0.05$  Family-Wise Error corrected as the cluster threshold. We also ran a conjunction analysis to specifically look at the overlap between production and comprehension in the response to constituent size. To run the conjunction analysis, we created contrast images for the constituent size effect separately in production and comprehension, and then we entered them into a one-way ANOVA in SPM, with each as a separate cell. By defining separate contrasts for each, we could then run the conjunction analysis for the group-level contrast image of constituent structure in production and comprehension.

#### ROI Analysis

We took functional regions of interest (ROIs) based on the keyword “syntactic” in Neurosynth (<https://www.neurosynth.org/>, accessed on 08/01/2020, Yarkoni et al. 2011). This allowed us to select voxels that are reported to be active in multiple studies related to a key search word, here “syntactic”. We downloaded the active voxels with a z-score threshold of 9. This revealed two clusters, one in left IFG and one in left anterior and posterior middle temporal lobe. We extracted mean beta values per participant in each of these ROIs per condition (C1, C2, C4, in production and comprehension) relative to baseline using MarsBar (Brett et al. 2002) in SPM12. We then compared the beta weights in a mixed-effects model in R (version 3.6) using *lme4* (Bates et al. 2015), with constituent size (C1, C2, C4), modality (Production vs. Comprehension) and ROI (LIFG vs. LMTG) as factors. Deviation coding was used for factors modality and ROI, while a linear contrast with weights  $[-4 -1 5]$ , as in the whole-brain analysis, was used for constituent size. We added by-participant random slopes for the interaction of ROI and modality and for the main effect of constituent size. We computed the contribution of

factors using Type-III Wald tests in *car* (version 3.0–7, Fox et al. 2020) and pairwise comparisons for significant effects with the package *emmeans* (version 1.4.6, Lenth et al. 2020).

#### Exploratory Analysis: BOLD Peak Latency

As an additional exploratory analysis, we extracted BOLD time courses to determine whether the time to peak was influenced by region, modality and constituent size. To capture a delay in peak times, we used a finite impulse response (FIR) basis set as implemented in Marsbar in SPM12. This allowed us to get estimates of BOLD activity at each TR in the two ROIs for each participant. We then extracted BOLD peak times as the timepoint with highest amplitude between 1.5 and 9 s post stimulus onset for each participant. We ran a linear mixed-effect model with constituent size, modality and ROI as fixed effects, and by-participant random slopes for ROI. We used a linear contrast with weights  $[-4 -1 5]$  for constituent size, and deviation coding for modality and ROI. We computed the contribution of factors using *car* (version 3.0–7, Fox et al. 2020) and pairwise comparisons for significant effects with the package *emmeans* (version 1.4.6, Lenth et al. 2020).

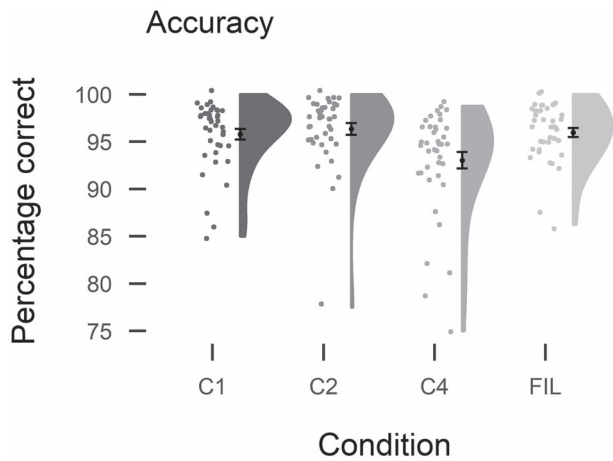
## Results

### Behavioral Results

Accuracy was generally high across participants and conditions (mean percentage correct: C1: 95.4, C2: 96.2, C4: 92.9, Fillers: 95.9; Fig. 2). There were slightly more errors in the C4 condition than in the C1 ( $\beta=0.55$ ,  $SE=0.12$ ,  $Z=4.7$ ,  $p<0.001$ ) and in the C2 conditions ( $\beta=0.71$ ,  $SE=0.14$ ,  $Z=5.2$ ,  $p<0.001$ ). Types of errors included using the wrong determiner (in Dutch, *het* is used with *meisje*-girl, and *de* with boy, man and woman; across all sentences for all participants,  $n=117$ ), the wrong actor ( $n=170$ ), a wrong verb or the correct verb in the wrong inflection/pronunciation ( $n=105$ ), the wrong condition ( $n=119$ ), not finishing within 5 seconds ( $n=77$ ), or other types of errors ( $n=62$ ). Unsurprisingly, onset times varied between conditions due to the characteristics of the conditions (mean onset times (in seconds): C1: 1.25, C2: 1.33, C4: 1.39, Fillers: 1.38, Fig. 3A). In particular, C1 elicited shorter reaction times than C2 ( $\beta=0.06$ ,  $SE=0.01$ ,  $t=5.7$ ,  $p<0.001$ ) and C4 ( $\beta=0.11$ ,  $SE=0.01$ ,  $t=8.2$ ,  $p<0.001$ ), as only the first verb had to be planned to initiate speech output. The other conditions, instead, required sentence planning, including subject (determiner and noun) as well as verb planning. The C4 condition elicited longer onset times than C2, too ( $\beta=0.05$ ,  $SE=0.01$ ,  $t=5.1$ ,  $p<0.001$ ). Similarly, duration times varied by condition (mean durations (in seconds): C1: 2.46, C2: 1.86, C4: 1.90, Fillers: 1.46; Fig. 3B). C1 production was characterized by the separate production of each lexical item, introducing pauses between words, and was thus characterized by the longest durations (vs. C2:  $\beta=0.28$ ,  $SE=0.02$ ,  $t=16.6$ ,  $p<0.001$ ; vs. C4:  $\beta=0.27$ ,  $SE=0.02$ ,  $t=15.4$ ,  $p<0.001$ ), while C2 and C4 did not differ in duration.

### Whole-Brain Analysis

We focused on the main effects of constituent size in production and comprehension, and on the interaction between modalities (production vs. comprehension) and constituent size. For the main effect of constituent size, a large bilateral network centered around areas of the language network and the corresponding right hemisphere areas, with cerebellar and occipital



**Figure 2.** Individual and mean accuracy per condition. Black dots indicate mean with standard error of the mean. Gray dots represent individual participants' mean.

activity, was found (Fig. 4A; Supplementary Table 1). We found a large left lateralized cluster including peaks in the left IFG, STG, MTG, temporal pole, precentral gyrus, postcentral gyrus, fusiform gyrus and superior parietal lobule. Similar right lateralized activity was found in a cluster in the temporal pole and IFG (*pars orbitalis*), a cluster in postcentral and precentral gyrus and a cluster in superior and middle temporal gyri, a cluster in superior parietal lobule and a cluster in the more posterior parts of the IFG (*pars triangularis* and *pars opercularis*). Additionally, we found clusters in the left and right supplementary motor area, in the left thalamus, left putamen, and right cerebellum.

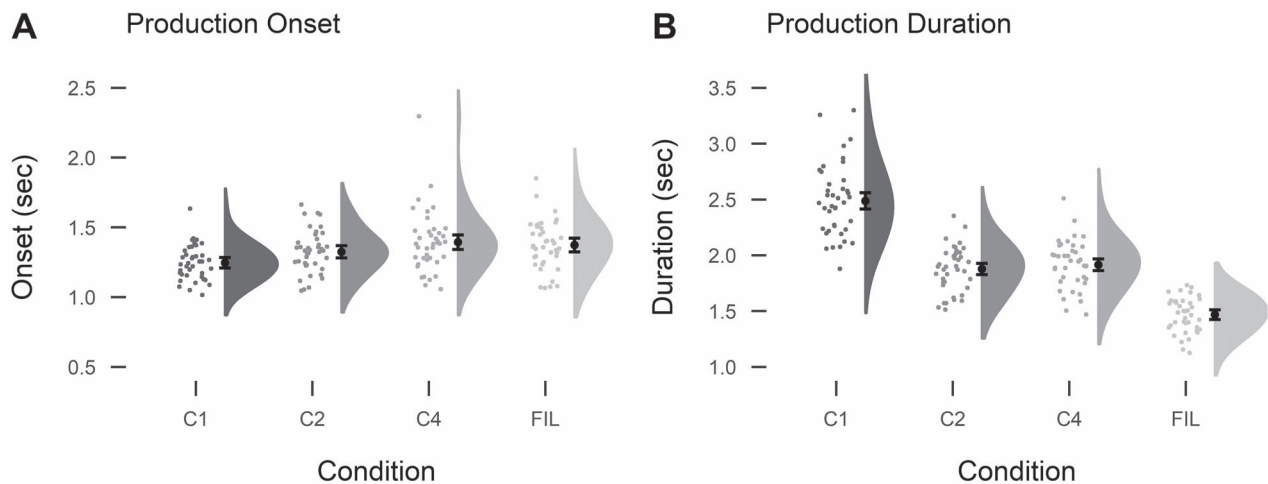
To evaluate to what extent the activated network was overlapping between comprehension and production, we performed a conjunction analysis of the separate constituent size contrast for production and comprehension. This analysis revealed that in part the constituent size effect was reliably active in both modalities, with clusters in anterior and posterior MTG, LIFG, left precentral gyrus, left fusiform gyrus and right cerebellum (see Fig. 4A, Supplementary Table 1).

We also looked at the main effect of modality to understand if any areas were overall more active in production or comprehension (see Fig. 4B, Supplementary Table 2). We found bilateral frontal areas and parietal areas, as well as subcortical and cerebellar regions, to have larger activity in production than comprehension, partly reflecting articulatory requirements in production. Bilateral superior temporal areas were more engaged in comprehension, which was likely due to auditory processing. Bilateral angular gyrus, precuneus and superior frontal regions were also more engaged in comprehension.

A few areas responded differently to constituent size in production and comprehension (Fig. 4C, Supplementary Table 3). Areas that were more left lateralized and included the LIFG (*pars triangularis*, *pars orbitalis* and *pars opercularis*), middle frontal gyrus, precentral gyrus, supplementary motor area, inferior and superior parietal lobule, supramarginal gyrus, angular gyrus, and posterior sections of the LMTG. Regions in the right hemisphere included precentral gyrus, postcentral gyrus, superior parietal lobule, supplementary motor area and cerebellum. A complementary network was more active in the comprehension of larger constituents, with peaks in bilateral Heschl's gyrus, STG and temporal pole, and right hemisphere areas, including angular gyrus, precuneus, frontal pole and superior and middle frontal gyri.

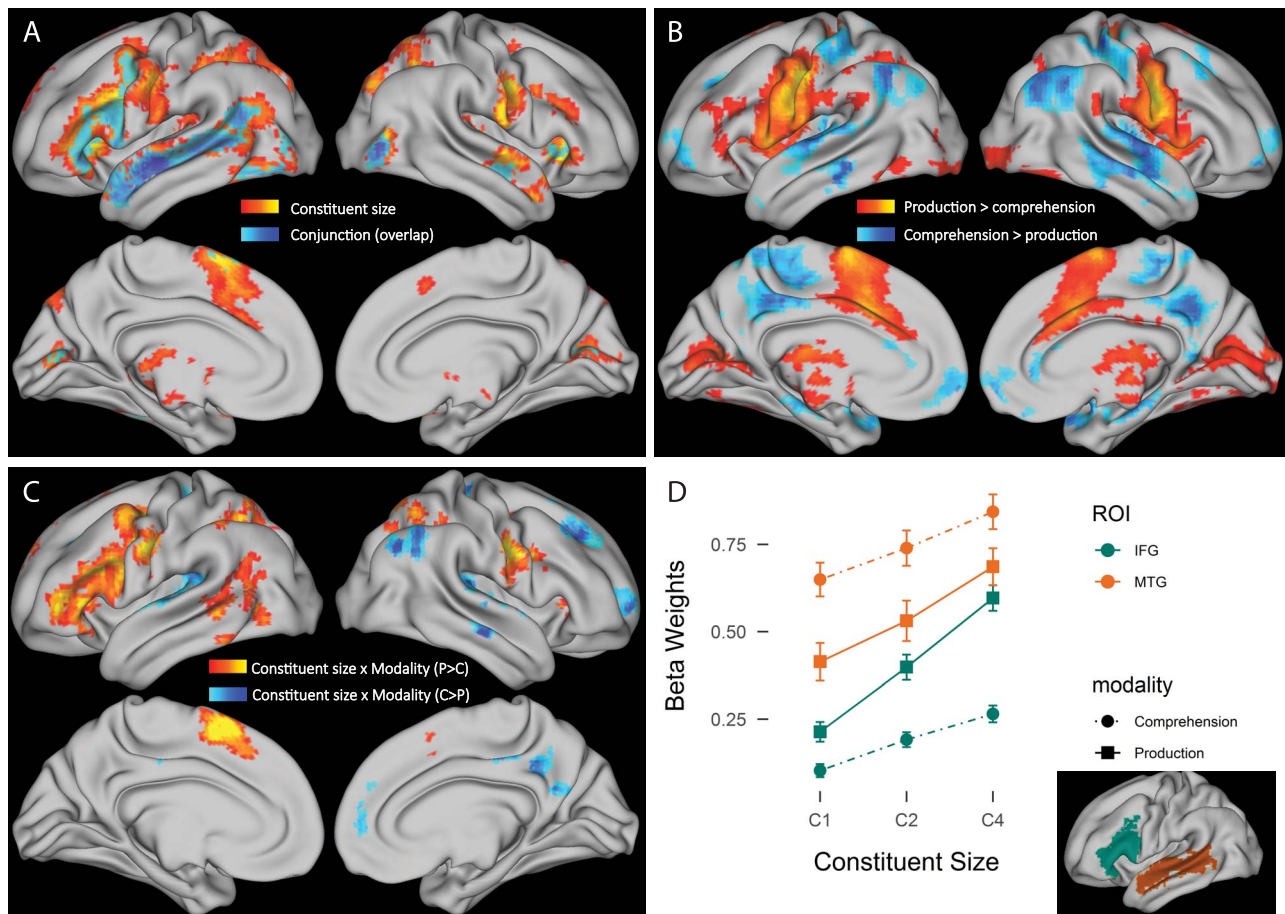
#### ROI Analysis: LIFG and LMTG

We extracted beta weights for the average activity in regions previously associated with syntactic effects to inspect patterns of activation for each condition in production and comprehension (Fig. 4D). We ran a linear mixed-effects model on the beta estimates that we extracted per condition per region. We found a main effect of constituent size ( $\beta = 0.027$ ,  $SE = 0.001$ ,  $t = 20.03$ ,  $\chi^2 = 401.04$ ,  $p < 0.0001$ ), indicating that beta weights increased with larger constituent sizes. Pairwise comparisons indicated that beta weights for C4 were significantly larger than C2 and C1 in all modalities and ROIs (estimates  $> 0.10$ ,  $ts > 8.2$ ,  $ps < 0.0001$ ). We also found a main effect of ROI ( $\beta = 0.18$ ,  $SE = 0.02$ ,  $t = 8.62$ ,  $\chi^2 = 74.4$ ,  $p < 0.0001$ ), with larger beta estimates in LMTG than LIFG (estimate = 0.35,  $SE = 0.04$ ,  $t = 8.52$ ,



**Figure 3.** Onset (A) and duration (B) times per condition (of correct trials only). Black dots indicate mean with standard error of the mean. Gray dots represent individual participants' mean.





**Figure 4.** Whole-brain and ROI results. **A:** orange: main effect of constituent size with a linear contrast for the three constituent sizes. Blue: conjunction analysis of production and comprehension constituent size effects representing areas active in both production and comprehension following the conjunction of null hypotheses (Friston et al. 2005). The blue area is superimposed on the corresponding cluster found as main effect of constituent size. **B:** whole-brain results for the main effect of modality. Orange: areas more active in production than comprehension. Blue: areas more active in comprehension than production. **C:** whole-brain results for the interaction between constituent size and modality. Orange: areas with larger response to constituent size in production than comprehension. Blue: areas with larger response to constituent size in comprehension than production. **D:** mean beta weights extracted from the predefined ROIs (depicted in figure), error bars represent standard error of the mean.

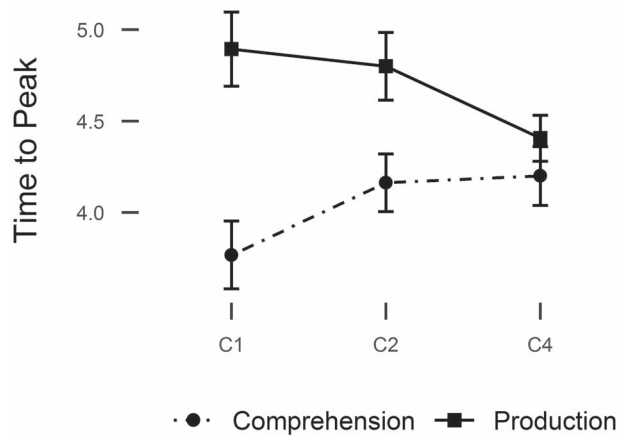
$p < 0.0001$ ). The effect of ROI interacted with modality ( $\beta = 0.11$ ,  $SE = 0.009$ ,  $t = 11.45$ ,  $\chi^2 = 131.03$ ,  $p < 0.0001$ ), since there was a larger difference in activity between ROIs in comprehension than in production (MTG—IFG, Production: estimate = 0.14,  $SE = 0.04$ ,  $t = 2.97$ ,  $p < 0.025$ ; Comprehension: estimate = 0.56,  $SE = 0.04$ ,  $t = 13.16$ ,  $p < 0.0001$ ). Importantly, there was a three-way interaction between constituent size, modality and ROI ( $\beta = 0.004$ ,  $SE = 0.0009$ ,  $t = 4.24$ ,  $\chi^2 = 18.0$ ,  $p < 0.0001$ ). Inspection of the slopes for constituent size in each modality and ROI indicated that production elicited the steepest slope in the response to constituent size in the IFG: there was a larger slope difference between modalities in the IFG (Production—Comprehension: estimate = 0.023,  $SE = 0.003$ ,  $t = 9.34$ ,  $p < 0.0001$ ) than in the MTG (Production—Comprehension: estimate = 0.009,  $SE = 0.003$ ,  $t = 3.37$ ,  $p = 0.005$ ), and there was a slope difference between ROIs in production (IFG—MTG: estimate = 0.012,  $SE = 0.003$ ,  $t = 4.5$ ,  $p = 0.0001$ ), but not in comprehension (MTG—IFG: estimate = 0.003,  $SE = 0.003$ ,  $t = 1.47$ ,  $p = 0.46$ ). These results, therefore, show that: (i) larger constituent structures elicit higher activity in both regions and modalities, (ii) there is a stronger effect of constituent size in production than in

comprehension, especially in LIFG, (iii) there is a higher response in LMTG than LIFG overall, and (iv) production elicits stronger activity than comprehension in the LIFG, while the opposite is the case in LMTG: more activity for comprehension than production.

#### Exploratory Analysis: BOLD Peak Latency

We extracted BOLD times-to-peak for each condition to understand whether the regional and modality-specific effects highlighted by the ROI analysis were also characterized by BOLD time course differences. Pallier et al. (2011) had found that larger constituent sizes were associated with later peak times in the superior temporal sulcus and IFG, in line with the idea that activation is stronger towards the end of a constituent. A model with ROI, modality and constituent size as predictors for time-to-peak showed a main effect of modality ( $\beta = 0.33$ ,  $SE = 0.06$ ,  $t = 5.39$ ,  $\chi^2 = 29.01$ ,  $p < 0.0001$ ), with comprehension peaking earlier than production (estimate = 0.66,  $SE = 0.12$ ,  $t = 5.34$ ,  $p < 0.0001$ ). In addition, we found an interaction between modality and constituent size ( $\beta = 0.049$ ,  $SE = 0.016$ ,





**Figure 5.** BOLD peak times averaged across participants and ROIs for each constituent size in production and comprehension. Error bars represent standard error of the mean.

$t = 3.002$ ,  $\chi^2 = 9.01$ ,  $p = 0.0027$ )<sup>3</sup>. Inspection of the slopes in the response to constituent size showed that comprehension elicited a positive slope, with larger constituent structures peaking later, while production elicited a negative slope, with larger constituent structures peaking earlier (Comprehension—Production: estimate = 0.098, SE = 0.033,  $t = 2.98$ ,  $p = 0.0031$ ) (Fig. 5). Therefore, the constituent size effect on peak latency that was found before (Pallier et al. 2011) seems to be dependent on modality, since in production an opposite pattern was found relative to comprehension.

## Discussion

We examined neural responses to the production and comprehension of utterances with increasing constituent size to clarify the neural correlates of sentence production and comprehension. We found that larger constituent sizes engaged areas traditionally part of the language network. These included inferior frontal regions, temporal and inferior parietal regions, mainly in the left hemisphere. Through a conjunction analysis, we confirmed that the LIFG and LMTG responded to constituent size in both comprehension and production. Increased syntactic complexity resulted in stronger activation in these areas. Moreover, we found a modality-specific dissociation, with production recruiting the LIFG more strongly than comprehension, and comprehension recruiting the LMTG more strongly than production. At the same time, the network was found to be differentially responsive to constituent size across modalities. While comprehension elicited similar responses to constituent size in LIFG and LMTG, in production the LIFG was more sensitive to constituent size than the LMTG. Finally, constituent size had opposite effects on BOLD peak latencies in comprehension and production: increasing constituent size elicited later peaks in comprehension but earlier peaks in production.

3 Since peak time extraction provided values in 1.5 sec resolution, we also ran a model with time (as well as constituent size, modality and ROI) as a predictor for percent signal change to ensure that extracted peak times were consistent with the time courses. The model returned a significant interaction between modality and time, and a three-way interaction between modality, constituent size and time, confirming the results of the peak time analysis.

By demonstrating that the response to constituent size is largely shared between comprehension and production, we extend Pallier et al.'s constituent size effect (2011) to sentence production. Our results are in line with evidence associating sentence-level processes with left inferior frontal and temporal activation (Bornkessel-Schlesewsky et al. 2009; Shetreet et al. 2009; Segaert et al. 2012; Hagoort and Indefrey 2014; Blank et al. 2016; Henderson et al. 2016; Walenski et al. 2019; Indefrey 2018). Pallier et al. (2011) found that the LIFG and the posterior superior temporal sulcus were responsive to constituent size also with jaberwocky stimuli, while the ATL and the temporoparietal junction only responded to stimuli with real words. All of these areas were responsive to constituent size also in the present study. Whether the IFG and the posterior temporal sulcus are sensitive to constituent size also with jaberwocky stimuli in production will have to be determined in future studies specifically designed to address the distinction between syntactic and semantic compositional processes. Finally, the activation delay for larger constituent structures was replicated here, but critically only in comprehension.

Our results, therefore, implicating both LIFG and LpMTG, as well as other areas of the language network, suggest that the inconsistent evidence for sentence production was due to low power in the single studies and in the meta-analyses (Walenski et al. 2019; Indefrey 2018). With 40 participants and a large number of trials per condition, we had enough power to detect effects in areas previously linked with sentence processing in comprehension. It is unlikely that the effects we found are reducible to the type of paradigm used to elicit sentence production, since other studies using picture descriptions or sentence reorganization paradigms also found activations in LIFG and/or LpMTG, but, critically, in an inconsistent way (e.g., pictures descriptions, Indefrey et al. 2001; Grande et al. 2012; Menenti et al. 2012; Segaert et al. 2012; sentence generation from words, Haller et al. 2005; Golestani et al. 2006; Collina et al. 2014). Moreover, although our paradigm was partly artificial in eliciting sentence production, it allowed us to cleanly manipulate constituent structure, ensuring consistent behavioral responses across participants. Previous studies used similar types of constrained elicitation paradigms or more constraining ones when more control over the production was required (cf. Matchin and Hickok 2016; Takashima et al. 2020; Matchin and Wood 2020).

Crucially, the conjunction analysis showed that production and comprehension engage largely overlapping areas in constituent structure building. An extensive network is engaged in sentence production that does not diverge from the one observed for comprehension in previous studies. The activation pattern, including left anterior and posterior MTG and LIFG is similar to the syntactic adaptation effects found across modalities in fMRI studies with repetition suppression (Menenti et al. 2011; Segaert et al. 2012). Our results do not provide information on whether verb-specific processing is also shared between production and comprehension, since the linear-contrast analysis avoided sensitivity to verb argument structure differences between sentences. Thus, these results confirm shared resources in sentence-level processes across modalities and provide no support for spatial segregation as a basis for distinct processes or representations. These findings, therefore, reconcile the previous inconsistent findings between sentence production and comprehension networks, as shown by meta-analyses (Walenski et al. 2019; Indefrey 2018), with the adaptation effects across modalities (Menenti et al. 2011; Segaert et al. 2012). Common neural resources provide a neural basis

for views of shared linguistic representations and processes, such as retrieval and unification, between production and comprehension (Kempen 2000; Kempen et al. 2012; Dell and Chang 2014; Momma and Phillips 2018).

While the networks overlapped, there were differences in the degree to which each modality recruited core areas. In particular, we found that comprehension engaged the LMTG more than production, and production engaged the LIFG more than comprehension. This finding was consistent with the modality differences in the whole-brain results. Larger activity in the LMTG in comprehension was also found by Humphreys and Gennari (2014), and is likely due to the fact that the auditory input is processed in superior temporal areas with activity spreading in the temporal lobe, whereby the LpMTG might be involved in retrieval and integration of lexical, syntactic and semantic information, given its extensive connectivity patterns (Baggio and Hagoort 2011; Turken and Dronkers 2011; Binder 2017). The clusters showing more activity in production included not only the LIFG, but also more dorsal areas, extending to the precentral gyrus and the supplementary motor area. Together with the cerebellar activation, these latter areas are involved in articulation and motor planning (Price 2012; Basilakos et al. 2018).

The greater involvement of inferior frontal regions in production than comprehension is likely attributable to stronger sentence planning requirements, also reflected in the stronger effect of constituent size in the LIFG and to a smaller extent in the LpMTG in production than comprehension. In production, the syntactic structure of sentences needs to be fully and correctly computed in order to produce a well-formed utterance (Indefrey 2018; Garrett 1982; Garrett 1980). In comprehension, instead, inferring sentence meaning can often be done by retrieving word meanings and world knowledge, bypassing the need for a full syntactic analysis of the input (cf. good-enough processing, Ferreira et al. 2002). For instance, it has been shown that passive or object-relative sentences are sometimes interpreted in line with world knowledge but not necessarily in agreement with the syntactic structure (Ferreira 2003; Ferreira and Lowder 2016). Therefore, reduced sensitivity to constituent structure in comprehension may signal reduced syntactic processing in reaching the conceptual interpretation for these sentences. This fundamental difference between production and comprehension on the importance of “getting it right” may also explain the larger engagement of the default mode network in comprehension (in particular, right angular gyrus, right pre-cuneus, right superior frontal gyrus and right frontal pole). We speculate that production disengaged the default mode network more than comprehension in responding to constituent size, due to the stronger requirements for accurate sentence planning (Raichle et al. 2001; Raichle and Snyder 2007).

The interaction effects between constituent size and modality cannot be reduced to task differences between modalities and in particular to the absence of an explicit task in comprehension. On the one hand, the constituent size effect in comprehension and the finding of larger comprehension activity in the LMTG confirm that participants processed the input even in the absence of a task (see Fig. 4D). On the other hand, the task requirements in production were very similar across levels of constituent size: what varied was the linguistic complexity of the output. Differences between modalities may instead show task effects, including cognitive control differences. However, as mentioned above, production is inherently a “task” as opposed to comprehension being more passive also in naturalistic situations. Task effects thus need not reflect spurious task

differences due to the current design, but could be related to inherent differences in cognitive control between production and comprehension. Studies of spontaneous production may be able to address to what extent cognitive control is needed during naturalistic production as opposed to comprehension.

An additional dissociation in the response pattern for production and comprehension was found in the BOLD time courses. Production and comprehension elicited opposite profiles of response latencies in relation with constituent size. Larger structures were characterized by later peaks in comprehension, confirming previous evidence suggesting that larger structures take longer to be computed (Pallier et al. 2011). In contrast, larger structures elicited earlier peaks than smaller structures in production. This was likely due to planning differences between conditions. Reaction time analyses showed that onset times increased with constituent size, with C2 taking longer than C1, and C4 taking longer than C2. Since high-level processing can be initiated for the whole clause before speaking (Smith and Wheeldon 1999), it is likely that more extensive planning at the message or structural level took place in early stages for the more complex structures, inducing early peaks in BOLD activity. In contrast, in the conditions with smaller constituent size the structures to be computed were smaller and planning may have been in a word-by-word fashion interleaved with articulation, hence inducing sustained activity with later peaks. Since this was an exploratory analysis for which the stimuli and the design were not optimized *a priori*, future studies will need to clarify whether BOLD peak latencies in production are indeed influenced by planning scope and if the inverse relationship between onset times and production peak latencies holds with different stimuli and paradigms.

Overall, the current results are striking in showing how production and comprehension share resources but modulate them differently. Spatially, frontal and temporal regions are engaged in both modalities, but to different extents. Temporally, constituent size affects BOLD peak latencies in both modalities but in opposite directions. Rather than providing support for a distinction of core processes and representations between modalities (Meyer et al. 2016), this unbalanced sharing of resources reveals a “computational asymmetry” (Matchin and Hickok 2020) or “directional” differences (Pickering and Garrod 2013; Gambi and Pickering 2017). In production, linguistic processes map from higher to lower linguistic levels, i.e., meaning to phonology, and in comprehension from lower to higher linguistic levels, i.e., phonology to meaning (Pickering and Garrod 2013). This directional difference implies that the inputs and outputs of each modality are opposite in production and comprehension, which results in differences in recruitment patterns within the shared language network (Momma and Phillips 2018; Indefrey 2018), reflected not only in different regional levels of activity, but also in timing patterns.

In conclusion, the current results extend the constituent structure effect found in comprehension (Pallier et al. 2011) to production, and robustly show the involvement of both LpMTG and LIFG in constituent structure building in production, helping to clarify the inconsistencies in the previous studies on the neurobiology of language production. Additionally, the results confirm that the neural resources for sentence production and comprehension are largely overlapping, supporting accounts of shared representations between modalities. Finally, our results highlight modality-specific differences in regional and time course patterns that underline inevitable differences in the requirements of speaking and listening.

## Supplementary Material

Supplementary material can be found at *Cerebral Cortex* online.

## Notes

We would like to thank Eva Poort for help with stimulus recording and coding of production recordings, Maarten van den Heuvel for help with presentation scripts and Marcel Zwiers for support with data preprocessing.

## Funding

This work was supported by the Max Planck Society. P.H. was supported by the NWO Grant Language in Interaction, grant number 024.001.006.

## Data Availability

The data will be made available on the Donders Repository at <https://data.donders.ru.nl/>.

## References

- Baggio G, Hagoort P. 2011. The balance between memory and unification in semantics: a dynamic account of the N400. *Language and Cognitive Processes*. 26(9):1338–1367. [10.1080/01690965.2010.542671](https://doi.org/10.1080/01690965.2010.542671).
- Barr DJ, Levy R, Scheepers C, Tily HJ. 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*. 68(3):255–278. [10.1016/j.jml.2012.11.001](https://doi.org/10.1016/j.jml.2012.11.001).
- Basilakos A, Smith KG, Fillmore P, Fridriksson J, Fedorenko E. 2018. Functional characterization of the human speech articulation network. *Cereb Cortex*. 28(5):1816–1830. [10.1093/cercor/bhx100](https://doi.org/10.1093/cercor/bhx100).
- Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. 67(1):1–48. [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01).
- Behzadi Y, Restom K, Liu J, Liu TT. 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*. 37(1):90–101. [10.1016/j.neuroimage.2007.04.042](https://doi.org/10.1016/j.neuroimage.2007.04.042).
- Bemis DK, Pylkkänen L. 2013. Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cereb Cortex*. 23(8):1859–1873. [10.1093/cercor/bhs170](https://doi.org/10.1093/cercor/bhs170).
- Bhattachali S, Fabre M, Luh W-M, Saied HA, Constant M, Pallier C, Brennan JR, Spreng RN, Hale J. 2019. Localising memory retrieval and syntactic composition: an fMRI study of naturalistic language comprehension. *Language, Cognition and Neuroscience*. 34(4):491–510. [10.1080/23273798.2018.1518533](https://doi.org/10.1080/23273798.2018.1518533).
- Binder JR. 2017. Current controversies on Wernicke's area and its role in language. *Curr Neurol Neurosci Rep*. 17(8):58. [10.1007/s11910-017-0764-8](https://doi.org/10.1007/s11910-017-0764-8).
- Blank I, Balewski Z, Mahowald K, Fedorenko E. 2016. Syntactic processing is distributed across the language system. *NeuroImage*. 127:307–323. [10.1016/j.neuroimage.2015.11.069](https://doi.org/10.1016/j.neuroimage.2015.11.069).
- Bock K. 1996. Language production: methods and methodologies. *Psychonomic Bulletin & Review*. 3(4):395–421. [10.3758/BF03214545](https://doi.org/10.3758/BF03214545).
- Bornkessel-Schlesewsky I, Schlesewsky M, von Cramon DY. 2009. Word order and Broca's region: evidence for a supra-syntactic perspective. *Brain and Language*. 111(3):125–139. [10.1016/j.bandl.2009.09.004](https://doi.org/10.1016/j.bandl.2009.09.004).
- Boylan C, Trueswell JC, Thompson-Schill SL. 2017. Relational vs. attributive interpretation of nominal compounds differentially engages angular gyrus and anterior temporal lobe. *Brain and Language*. 169:8–21. [10.1016/j.bandl.2017.01.008](https://doi.org/10.1016/j.bandl.2017.01.008).
- Brett M, Anton J-L, Valabregue R, Poline J-B. 2002. Region of interest analysis using an SPM toolbox. Presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan.
- Brysbaert M, Stevens M, De Deyne S, Voorspoels W, Storms G. 2014. Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*. 150:80–84. [10.1016/j.actpsy.2014.04.010](https://doi.org/10.1016/j.actpsy.2014.04.010).
- Chang CHC, Dehaene S, Wu DH, Kuo W-J, Pallier C. 2020. Cortical encoding of linguistic constituent with and without morphosyntactic cues. *Cortex*. 129:281–295. [10.1016/j.cortex.2020.04.024](https://doi.org/10.1016/j.cortex.2020.04.024).
- Clark EV, Hecht BF. 1983. Comprehension, production, and language acquisition. *Annual Review of Psychology*. 34(1):325–349. [10.1146/annurev.ps.34.020183.001545](https://doi.org/10.1146/annurev.ps.34.020183.001545).
- Collina S, Seurinck R, Hartsuiker RJ. 2014. Inside the syntactic box: the neural correlates of the functional and positional level in covert sentence production. *PLOS ONE*. 9(9):e106122. [10.1371/journal.pone.0106122](https://doi.org/10.1371/journal.pone.0106122).
- Dale AM. 1999. Optimal experimental design for event-related fMRI. *Hum Brain Mapp*. 8(2–3):109–114.
- Dale AM, Fischl B, Sereno MI. 1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*. 9(2):179–194. [10.1006/nimg.1998.0395](https://doi.org/10.1006/nimg.1998.0395).
- Dell GS, Chang F. 2014. The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 369(1634):20120394. [10.1098/rstb.2012.0394](https://doi.org/10.1098/rstb.2012.0394).
- Esteban O, Blair R, Markiewicz CJ, Berleant SL, Moodie C, Ma F, Isik AI, Erramuzpe A, Kent M, James D, et al. 2018. fMRIPrep. *Zenodo*. [10.5281/zenodo.852659](https://doi.org/10.5281/zenodo.852659).
- Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Snyder M, et al. 2018b. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*. 16(1):111–116. [10.1038/s41592-018-0235-4](https://doi.org/10.1038/s41592-018-0235-4).
- Europa E, Gitelman DR, Kiran S, Thompson CK. 2019. Neural connectivity in syntactic movement processing. *Front Hum Neurosci*. 13:27. [10.3389/fnhum.2019.00027](https://doi.org/10.3389/fnhum.2019.00027).
- Fedorenko E, Nieto-Castañón A, Kanwisher N. 2012. Lexical and syntactic representations in the brain: an fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia*. 50(4):499–513. [10.1016/j.neuropsychologia.2011.09.014](https://doi.org/10.1016/j.neuropsychologia.2011.09.014).
- Ferreira F. 1991. Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*. 30(2):210–233. [10.1016/0749-596X\(91\)90004-4](https://doi.org/10.1016/0749-596X(91)90004-4).
- Ferreira F. 2003. The misinterpretation of noncanonical sentences. *Cognitive Psychology*. 47(2):164–203. [10.1016/S0010-0285\(03\)00005-7](https://doi.org/10.1016/S0010-0285(03)00005-7).
- Ferreira F, Bailey KGD, Ferraro V. 2002. Good-enough representations in language comprehension. *Curr Dir Psychol Sci*. 11(1):11–15. [10.1111/1467-8721.00158](https://doi.org/10.1111/1467-8721.00158).
- Ferreira F, Lowder MW. 2016. Chapter Six - Prediction, Information Structure, and Good-Enough Language Processing. In:



- Ross BH, editor. *Psychology of Learning and Motivation*. Vol 65. Academic Press, pp. 217–247.
- Flick G, Pylkkänen L. 2020. Isolating syntax in natural language: MEG evidence for an early contribution of left posterior temporal cortex. *Cortex*. 127:42–57. [10.1016/j.cortex.2020.01.025](https://doi.org/10.1016/j.cortex.2020.01.025).
- Fonov V, Evans A, McKinstry R, Almlí C, Collins D. 2009. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*. 47(Supplement 1):S102. [10.1016/S1053-8119\(09\)70884-5](https://doi.org/10.1016/S1053-8119(09)70884-5).
- Fox J, Weisberg S, Price B, Adler D, Bates D, Baud-Bovy G, Bolker B, Ellison S, Firth D, Friendly M, et al. 2020. Car: companion to applied regression. [accessed 2020 Jun 30]. <https://CRAN.R-project.org/package=car>.
- Friederici AD. 2012. The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences*. 16(5):262–268. [10.1016/j.tics.2012.04.001](https://doi.org/10.1016/j.tics.2012.04.001).
- Friederici AD, Gierhan SM. 2013. The language network. *Current Opinion in Neurobiology*. 23(2):250–254. [10.1016/j.conb.2012.10.002](https://doi.org/10.1016/j.conb.2012.10.002).
- Friston KJ, Penny WD, Glaser DE. 2005. Conjunction revisited. *NeuroImage*. 25(3):661–667. [10.1016/j.neuroimage.2005.01.013](https://doi.org/10.1016/j.neuroimage.2005.01.013).
- Gambi C, Pickering MJ. 2017. Models linking production and comprehension. In: *The handbook of psycholinguistics*. Wiley Online Library. p. 157–181.
- Garrett M. 1980. Levels of processing in sentence production. In: *Language production Vol. 1: Speech and talk*. Academic Press. p. 177–220.
- Garrett MF. 1982. Remarks on the relation between language production and language comprehension systems. In: *Neural models of language processes*. Elsevier, pp. 209–224.
- Golestani N, Alario F-X, Meriaux S, Le Bihan D, Dehaene S, Pallier C. 2006. Syntax production in bilinguals. *Neuropsychologia*. 44(7):1029–1040. [10.1016/j.neuropsychologia.2005.11.009](https://doi.org/10.1016/j.neuropsychologia.2005.11.009).
- Grande M, Meffert E, Schoenberger E, Jung S, Frauenrath T, Huber W, Hussmann K, Moormann M, Heim S. 2012. From a concept to a word in a syntactically complete sentence: an fMRI study on spontaneous language production in an overt picture description task. *NeuroImage*. 61(3):702–714. [10.1016/j.neuroimage.2012.03.087](https://doi.org/10.1016/j.neuroimage.2012.03.087).
- Grill-Spector K, Henson R, Martin A. 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*. 10(1):14–23. [10.1016/j.tics.2005.11.006](https://doi.org/10.1016/j.tics.2005.11.006).
- Grodzinsky Y. 2000. The neurology of syntax: language use without Broca's area. *Behavioral and Brain Sciences*. 23(1):1–21. [10.1017/S0140525X00002399](https://doi.org/10.1017/S0140525X00002399).
- Hagoort P. 2005. On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*. 9(9):416–423. [10.1016/j.tics.2005.07.004](https://doi.org/10.1016/j.tics.2005.07.004).
- Hagoort P. 2013. MUC (memory, unification, control) and beyond. *Front Psychol*. 4:416. [10.3389/fpsyg.2013.00416](https://doi.org/10.3389/fpsyg.2013.00416).
- Hagoort P. 2019. The neurobiology of language beyond single-word processing. *Science*. 366(6461):55–58. [10.1126/science.aax0289](https://doi.org/10.1126/science.aax0289).
- Hagoort P, Indefrey P. 2014. The neurobiology of language beyond single words. *Annual Review of Neuroscience*. 37(1):347–362. [10.1146/annurev-neuro-071013-013847](https://doi.org/10.1146/annurev-neuro-071013-013847).
- Haller S, Radue EW, Erb M, Grodd W, Kircher T. 2005. Overt sentence production in event-related fMRI. *Neuropsychologia*. 43(5):807–814. [10.1016/j.neuropsychologia.2004.09.007](https://doi.org/10.1016/j.neuropsychologia.2004.09.007).
- Henderson JM, Choi W, Lowder MW, Ferreira F. 2016. Language structure in the brain: a fixation-related fMRI study of syntactic surprisal in reading. *NeuroImage*. 132:293–300. [10.1016/j.neuroimage.2016.02.050](https://doi.org/10.1016/j.neuroimage.2016.02.050).
- Hendriks P, Koster C. 2010. Production/comprehension asymmetries in language acquisition. *Lingua*. 120(8):1887–1897. [10.1016/j.lingua.2010.02.002](https://doi.org/10.1016/j.lingua.2010.02.002).
- Henson RN. 2015. Analysis of variance (ANOVA). In: *Brain Mapping: an encyclopedic reference*. Elsevier, pp. 477–481.
- Hickok G, Poeppel D. 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience*. 8(5):393–402. [10.1038/nrn2113](https://doi.org/10.1038/nrn2113).
- Hirotsani M, Makuuchi M, Rüschemeyer S-A, Friederici AD. 2011. Who was the agent? The neural correlates of reanalysis processes during sentence comprehension. *Human Brain Mapping*. 32(11):1775–1787. [10.1002/hbm.21146](https://doi.org/10.1002/hbm.21146).
- Humphreys GF, Gennari SP. 2014. Competitive mechanisms in sentence processing: common and distinct production and reading comprehension networks linked to the prefrontal cortex. *NeuroImage*. 84:354–366. [10.1016/j.neuroimage.2013.08.059](https://doi.org/10.1016/j.neuroimage.2013.08.059).
- Indefrey P. 2018. *The Relationship Between Syntactic Production and Comprehension*. The Oxford Handbook of Psycholinguistics. doi:10.1093/oxfordhb/9780198786825.013.20. p. 486–505.
- Indefrey P, Brown CM, Hellwig F, Amunts K, Herzog H, Seitz RJ, Hagoort P. 2001. A neural correlate of syntactic encoding during speech production. *PNAS*. 98(10):5933–5936. [10.1073/pnas.101118098](https://doi.org/10.1073/pnas.101118098).
- Indefrey P, Hellwig F, Herzog H, Seitz RJ, Hagoort P. 2004. Neural responses to the production and comprehension of syntax in identical utterances. *Brain and Language*. 89(2):312–319. [10.1016/S0093-934X\(03\)00352-3](https://doi.org/10.1016/S0093-934X(03)00352-3).
- Jackendoff R, Audring J. 2020. *The texture of the lexicon: relational morphology and the parallel architecture*. USA: Oxford University Press.
- Jaeger TF. 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*. 59(4):434–446. [10.1016/j.jml.2007.11.007](https://doi.org/10.1016/j.jml.2007.11.007).
- Jenkinson M, Bannister P, Brady M, Smith S. 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*. 17(2):825–841. [10.1006/nimg.2002.1132](https://doi.org/10.1006/nimg.2002.1132).
- Kempen G. 2000. Could grammatical encoding and grammatical decoding be subserved by the same processing module? *Behavioral and Brain Sciences*. 23(1):38–39. [10.1017/S0140525X00402396](https://doi.org/10.1017/S0140525X00402396).
- Kempen G, Olsthoorn N, Sprenger S. 2012. Grammatical workspace sharing during language production and language comprehension: evidence from grammatical multitasking. *Language and Cognitive Processes*. 27(3):345–380. [10.1080/01690965.2010.544583](https://doi.org/10.1080/01690965.2010.544583).
- Keuleers E, Brysbaert M, New B. 2010. SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*. 42(3):643–650. [10.3758/BRM.42.3.643](https://doi.org/10.3758/BRM.42.3.643).
- Kircher TTJ, Oh TM, Brammer MJ, McGuire PK. 2005. Neural correlates of syntax production in schizophrenia. *The British Journal of Psychiatry*. 186(3):209–214. [10.1192/bjp.186.3.209](https://doi.org/10.1192/bjp.186.3.209).
- Krause F, Benjamins C, Eck J, Lühns M, van Hoof R, Goebel R. 2019. Active head motion reduction in magnetic resonance imaging using tactile feedback. *Human Brain Mapping*. 40(14):4026–4037. <https://doi.org/10.1002/hbm.24683>.
- Lambon Ralph MA, Jefferies E, Patterson K, Rogers TT. 2017. The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*. 18(1):42–55. [10.1038/nrn.2016.150](https://doi.org/10.1038/nrn.2016.150).



- Lee DK, Fedorenko E, Simon MV, Curry WT, Nahed BV, Cahill DP, Williams ZM. 2018. Neural encoding and production of functional morphemes in the posterior temporal lobe. *Nature Communications*. 9(1):1877. [10.1038/s41467-018-04235-3](https://doi.org/10.1038/s41467-018-04235-3).
- Lenth R, Singmann H, Love J, Buerkner P, Herve M. 2020. Emmeans: estimated marginal means, aka least-squares means. [accessed 2020 May 8]. <https://CRAN.R-project.org/package=emmeans>.
- Levelt WJ. 2013. *A history of psycholinguistics: The pre-Chomskyan era*. Oxford University Press.
- Mack JE, Meltzer-Asscher A, Barbieri E, Thompson CK. 2013. Neural correlates of processing passive sentences. *Brain Sciences*. 3(3):1198–1214. [10.3390/brainsci3031198](https://doi.org/10.3390/brainsci3031198).
- Matchin W. 2018. A neuronal retuning hypothesis of sentence-specificity in Broca's area. *Psychon Bull Rev*. 25(5):1682–1694. [10.3758/s13423-017-1377-6](https://doi.org/10.3758/s13423-017-1377-6).
- Matchin W, Hammerly C, Lau E. 2017. The role of the IFG and pSTS in syntactic prediction: evidence from a parametric study of hierarchical structure in fMRI. *Cortex*. 88:106–123. [10.1016/j.cortex.2016.12.010](https://doi.org/10.1016/j.cortex.2016.12.010).
- Matchin W, Hickok G. 2016. 'Syntactic perturbation' during production activates the right IFG, but not Broca's area or the ATL. *Front Psychol*. 7:241. [10.3389/fpsyg.2016.00241](https://doi.org/10.3389/fpsyg.2016.00241).
- Matchin W, Hickok G. 2020. The cortical Organization of Syntax. *Cereb Cortex*. 30(3):1481–1498. [10.1093/cercor/bhz180](https://doi.org/10.1093/cercor/bhz180).
- Matchin W, Wood E. 2020. Syntax-sensitive regions of the posterior inferior frontal gyrus and the posterior temporal lobe are differentially recruited by production and perception. *Cereb Cortex Comm*. 1:1, tga029. [10.1093/texcom/tgaa029](https://doi.org/10.1093/texcom/tgaa029).
- Menenti L, Gierhan SME, Segaert K, Hagoort P. 2011. Shared language: overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychol Sci*. 22(9):1173–1182. [10.1177/0956797611418347](https://doi.org/10.1177/0956797611418347).
- Menenti L, Segaert K, Hagoort P. 2012. The neuronal infrastructure of speaking. *Brain and Language*. 122(2):71–80. [10.1016/j.bandl.2012.04.012](https://doi.org/10.1016/j.bandl.2012.04.012).
- Mesulam M-M, Rogalski EJ, Wieneke C, Hurley RS, Geula C, Bigio EH, Thompson CK, Weintraub S. 2014. Primary progressive aphasia and the evolving neurology of the language network. *Nature Reviews Neurology*. 10(10):554–569. [10.1038/nrneurol.2014.159](https://doi.org/10.1038/nrneurol.2014.159).
- Meyer AS, Huettig F, Levelt WJ. 2016. Same, different, or closely related: what is the relationship between language production and comprehension? *Journal of Memory and Language*. 89:1–7.
- Momma S, Phillips C. 2018. The relationship between parsing and generation. *Annual Review of Linguistics*. 4(1):233–254. [10.1146/annurev-linguistics-011817-045719](https://doi.org/10.1146/annurev-linguistics-011817-045719).
- Mumford JA, Nichols TE. 2008. Power calculation for group fMRI studies accounting for arbitrary design and temporal autocorrelation. *NeuroImage*. 39(1):261–268. [10.1016/j.neuroimage.2007.07.061](https://doi.org/10.1016/j.neuroimage.2007.07.061).
- den Ouden D-B, Hoogduin H, Stowe LA, Bastiaanse R. 2008. Neural correlates of Dutch verb second in speech production. *Brain and Language*. 104(2):122–131. [10.1016/j.bandl.2007.05.001](https://doi.org/10.1016/j.bandl.2007.05.001).
- Pallier C, Devauchelle A-D, Dehaene S. 2011. Cortical representation of the constituent structure of sentences. *PNAS*. 108(6):2522–2527. [10.1073/pnas.1018711108](https://doi.org/10.1073/pnas.1018711108).
- Phillips C. 2013. Parser-grammar relations: We don't understand everything twice. In: *Language Down the Garden Path: The Cognitive and Biological Basis for Linguistic Structures*. Oxford University Press. p. 294–315.
- Pickering MJ, Garrod S. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*. 36(4):329–347. [10.1017/S0140525X12001495](https://doi.org/10.1017/S0140525X12001495).
- Pinheiro J, Bates D. 2000. *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag (Statistics and Computing).
- Posse S, Wiese S, Gembris D, Mathiak K, Kessler C, Grosse-Ruyken M-L, Elghahwagi B, Richards T, Dager SR, Kiselev VG. 1999. Enhancement of BOLD-contrast sensitivity by single-shot multi-echo functional MR imaging. *Magnetic Resonance in Medicine*. 42(1):87–97. [10.1002/\(SICI\)1522-2594\(199907\)42:1<87::AID-MRM13>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1522-2594(199907)42:1<87::AID-MRM13>3.0.CO;2-O).
- Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*. 84(Supplement C):320–341. [10.1016/j.neuroimage.2013.08.048](https://doi.org/10.1016/j.neuroimage.2013.08.048).
- Price CJ. 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*. 62(2):816–847. [10.1016/j.neuroimage.2012.04.062](https://doi.org/10.1016/j.neuroimage.2012.04.062).
- Pruim RHR, Mennes M, van Rooij D, Llera A, Buitelaar JK, Beckmann CF. 2015. ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*. 112(Supplement C):267–277. [10.1016/j.neuroimage.2015.02.064](https://doi.org/10.1016/j.neuroimage.2015.02.064).
- Pylkkänen L, Bemis DK, Blanco EE. 2014. Building phrases in language production: an MEG study of simple composition. *Cognition*. 133(2):371–384. [10.1016/j.cognition.2014.07.001](https://doi.org/10.1016/j.cognition.2014.07.001).
- Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. 2001. A default mode of brain function. *Proceedings of the National Academy of Sciences*. 98(2):676–682.
- Raichle ME, Snyder AZ. 2007. A default mode of brain function: a brief history of an evolving idea. *NeuroImage*. 37(4):1083–1090. [10.1016/j.neuroimage.2007.02.041](https://doi.org/10.1016/j.neuroimage.2007.02.041).
- Rogalsky C, Matchin W, Hickok G. 2008. Broca's area, sentence comprehension, and working memory: an fMRI study. *Front Hum Neurosci*. 2:14. [10.3389/fnhum.2008.00014](https://doi.org/10.3389/fnhum.2008.00014).
- Santi A, Grodzinsky Y. 2010. fMRI adaptation dissociates syntactic complexity dimensions. *NeuroImage*. 51(4):1285–1293. [10.1016/j.neuroimage.2010.03.034](https://doi.org/10.1016/j.neuroimage.2010.03.034).
- Schell M, Zaccarella E, Friederici AD. 2017. Differential cortical contribution of syntax and semantics: an fMRI study on two-word phrasal processing. *Cortex*. 96:105–120. [10.1016/j.cortex.2017.09.002](https://doi.org/10.1016/j.cortex.2017.09.002).
- Segaert K, Kempen G, Petersson KM, Hagoort P. 2013. Syntactic priming and the lexical boost effect during sentence production and sentence comprehension: an fMRI study. *Brain and Language*. 124(2):174–183. [10.1016/j.bandl.2012.12.003](https://doi.org/10.1016/j.bandl.2012.12.003).
- Segaert K, Menenti L, Weber K, Petersson KM, Hagoort P. 2012. Shared syntax in language production and language comprehension—an fMRI study. *Cereb Cortex*. 22(7):1662–1670. [10.1093/cercor/bhr249](https://doi.org/10.1093/cercor/bhr249).
- Shetreet E, Friedmann N, Hadar U. 2009. An fMRI study of syntactic layers: sentential and lexical aspects of embedding. *NeuroImage*. 48(4):707–716. [10.1016/j.neuroimage.2009.07.001](https://doi.org/10.1016/j.neuroimage.2009.07.001).
- Smith M, Wheeldon L. 1999. High level processing scope in spoken sentence production. *Cognition*. 73(3):205–246. [10.1016/S0010-0277\(99\)00053-0](https://doi.org/10.1016/S0010-0277(99)00053-0).
- Snijders TM, Vosse T, Kempen G, Van Berkum JJA, Petersson KM, Hagoort P. 2009. Retrieval and unification of syntactic

- structure in sentence comprehension: an fMRI study using word-category ambiguity. *Cereb Cortex*. 19(7):1493–1503. [10.1093/cercor/bhn187](https://doi.org/10.1093/cercor/bhn187).
- Takashima A, Konopka A, Meyer A, Hagoort P, Weber K. 2020. Speaking in the brain: the interaction between words and syntax in sentence production. *Journal of Cognitive Neuroscience*. 32(8):1466–1483. [10.1162/jocn\\_a\\_01563](https://doi.org/10.1162/jocn_a_01563).
- Thothathiri M. 2018. Statistical experience and individual cognitive differences modulate neural activity during sentence production. *Brain and Language*. 183:47–53. [10.1016/j.bandl.2018.06.005](https://doi.org/10.1016/j.bandl.2018.06.005).
- Thothathiri M, Rattinger M. 2015. Ventral and dorsal streams for choosing word order during sentence production. *PNAS*. 112(50):15456–15461. [10.1073/pnas.1514711112](https://doi.org/10.1073/pnas.1514711112).
- Tremblay P, Dick AS. 2016. Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language*. 162:60–71. [10.1016/j.bandl.2016.08.004](https://doi.org/10.1016/j.bandl.2016.08.004).
- Turken AU, Dronkers NF. 2011. The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. *Front Syst Neurosci*. 5:1. [10.3389/fnsys.2011.00001](https://doi.org/10.3389/fnsys.2011.00001).
- Walenski M, Europa E, Caplan D, Thompson CK. 2019. Neural networks for sentence comprehension and production: an ALE-based meta-analysis of neuroimaging studies. *Human Brain Mapping*. 40(8):2275–2304. [10.1002/hbm.24523](https://doi.org/10.1002/hbm.24523).
- Wernicke C. 1892. *Gesammelte Aufsätze und kritische Referate zur Pathologie des Nervensystems*. Saarbrücken: VDM Verlag.
- Willems RM, van Gerven MAJ. 2018. New FMRI methods for the study of language. *The Oxford Handbook of Psycholinguistics*. 2:975–992. [10.1093/oxfordhb/9780198786825.013.42](https://doi.org/10.1093/oxfordhb/9780198786825.013.42).
- Wilson SM. 2017. Lesion-symptom mapping in the study of spoken language understanding. *Language, Cognition and Neuroscience*. 32(7):891–899. [10.1080/23273798.2016.1248984](https://doi.org/10.1080/23273798.2016.1248984).
- Wilson SM, DeMarco AT, Henry ML, Gesierich B, Babiak M, Mandelli ML, Miller BL, Gorno-Tempini ML. 2013. What role does the anterior temporal lobe play in sentence-level processing? Neural correlates of syntactic processing in semantic variant primary progressive aphasia. *Journal of Cognitive Neuroscience*. 26(5):970–985. [10.1162/jocn\\_a\\_00550](https://doi.org/10.1162/jocn_a_00550).
- Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. 2011. Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*. 8(8):665–670. [10.1038/nmeth.1635](https://doi.org/10.1038/nmeth.1635).
- Zaccarella E, Meyer L, Makuuchi M, Friederici AD. 2017a. Building by syntax: the neural basis of minimal linguistic structures. *Cereb Cortex*. 27(1):411–421. [10.1093/cercor/bhv234](https://doi.org/10.1093/cercor/bhv234).
- Zaccarella E, Schell M, Friederici AD. 2017b. Reviewing the functional basis of the syntactic merge mechanism for language: a coordinate-based activation likelihood estimation meta-analysis. *Neuroscience & Biobehavioral Reviews*. 80:646–656. [10.1016/j.neubiorev.2017.06.011](https://doi.org/10.1016/j.neubiorev.2017.06.011).
- Zhang Y, Brady M, Smith S. 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*. 20(1):45–57. [10.1109/42.906424](https://doi.org/10.1109/42.906424).