Conversational production and comprehension: fMRI-evidence reminiscent of but deviant from the classical Broca–Wernicke model

Caroline Arvidsson 1, Ekaterina Torubarova 2, André Pereira 3, Julia Uddén 1,3,4

1 Department of Linguistics, Stockholm University, Universitetsvägen 10 C, 114 18 Stockholm, Sweden
2 Division of Speech, Music, and Hearing, KTH Royal Institute of Technology, Lindstedtsvägen 24, 114 28 Stockholm, Sweden
3 Department of Psychology, Stockholm University, Albanovägen 12, 114 19 Stockholm, Sweden
4 Corresponding author: julia.udden@psychology.su.se

A key question in research on the neurobiology of language is to which extent the language production and comprehension systems share neural infrastructure, but this question has not been addressed in the context of conversation. We utilized a public fMRI dataset where 24 participants engaged in unscripted conversations with a confederate outside the scanner, via an audio-video link. We provide evidence indicating that the two systems share neural infrastructure in the left-lateralized perisylvian language network, but diverge regarding the level of activation in regions within the network. Activity in the left inferior frontal gyrus was stronger in production compared to comprehension, while comprehension showed stronger recruitment of the left anterior middle temporal gyrus and superior temporal sulcus, compared to production. Although our results are reminiscent of the classical Broca–Wernicke model, the anterior (rather than posterior) temporal activation is a notable difference from that model. This is one of the findings that may be a consequence of the conversational setting, another being that conversational production activated what we interpret as higher-level socio-pragmatic processes. In conclusion, we present evidence for partial overlap and functional asymmetry of the neural infrastructure of production and comprehension, in the above-mentioned frontal vs temporal regions during conversation.

Key words: interaction; contextual language processing; LIFG; LMTG; functional asymmetry.

Introduction

Conversation is integral to the everyday experience of almost every human. It is thus not surprising that language processing, as occurring during conversation, is the dominant explanation in psycho- and neurolinguistics. Notwithstanding, isolation paradigms, where participants produce or listen to linguistic signals in a non-interactive setting, are standard in behavioral and in particular neuroimaging experiments. Conversation entails flexibly managing the roles of speaker and listener while simultaneously considering linguistic, social, and other contextual factors to encode and decode meaning (Austin 1973; Grice et al. 1975; Clark and Murphy 1982). As all of that is missing in the isolation paradigm, one can question the validity of typical psycho- or neurolinguistic experiments. The goal of this fMRI study was to address this issue by investigating the underlying processes of speech production and comprehension during conversation.

A long-standing yet ongoing debate is to what extent production and comprehension systems (Pickering and Garrod 2007) diverge concerning the recruitment of regions in the perisylvian language network (Okada and Hickok 2006; Matchin and Hickok 2020; Hickok 2022; Hu et al. 2022; Matchin et al. 2022; Rutten 2022; Giglio et al. 2022a, b). The perisylvian language network (Caplan 1987) is a left hemisphere-dominant network of cortical areas including the left inferior frontal gyrus (LIFG), the left middle/superior temporal gyri (LMTG/STG) and the posterior inferior parietal cortex. This network is known to be crucial to higher-level linguistic processing, e.g. syntactic and semantic processing at the sentence level (Tyler et al. 2010; Vlooswijk et al. 2010; Fedorenko et al. 2011; Mollica et al. 2020; Hickok 2022; Malik-Moraleda et al. 2022; Stockert et al. 2023). In the classical model of the neurobiology of language, originating from the pioneering work of Carl Wernicke in the late 19th century, “Broca’s area” (the posterior LIFG) was described as a motor speech center, and “Wernicke’s area” (the posterior LSTG) a sensory speech center (Wernicke and Eggert, 1885/1977; Geschwind 1970; Rutten 2022). At variance with the classical model were later lesion and neuroimaging studies, providing compelling evidence that frontal and temporal regions both subserve aspects of production and comprehension (Caramazza and Zurif 1976; Seg aer et al. 2012; Fridriksson et al. 2015; Hagoort 2016). Today, the classical model’s assumption of modularity is considered outdated (see e.g. Tremblay and Dick 2016), as the field has moved toward an understanding of language as a complex system of processes subserved by multiple regions. Still, questions on the relative contribution of the perisylvian language regions to production and comprehension remain unresolved.

Hu et al. (2022) argue for a model in which production and comprehension rely on the same knowledge representations [see also Pickering and Garrod 2004; Chomsky 2014] and, by extension, the same neural structures. In an experimentally controlled fMRI
study, these authors found no evidence of any brain regions, within or outside the language network, that selectively supported processing in one of the systems (production or comprehension) but not the other. Hu et al. (2022) also found that all language regions (localized by contrasting reading sentences vs lists of nonwords) were more engaged during production than during comprehension, which they explained by arguing that production is overall more demanding than comprehension. However, it is reasonable to assume that the language processes tapped by the tasks in Hu et al. (2022) differ from conversational language processes. For example, their comprehension tasks involved reading or listening to context-independent sentences containing single clauses (e.g., “the girl is smelling a flower”) and did not require the listener to integrate both previous linguistic and other contextual information to understand the message. The comprehension tasks in their study were therefore likely less demanding than conversational comprehension.

In contrast to Hu et al. (2022), models such as Hickok (2022) and Matchin and Hickok (2020) argue for a functional asymmetry of inferior frontal and temporal regions. According to these models, the posterior LIFG supports processes primarily tied to production (e.g., the transformation of abstract morphosyntactic representations into linear sequences of morphemes). Moreover, these models suggest that the temporal lobes support processes crucial to both production and comprehension. For instance, the left posterior middle temporal gyrus (LpMTG) is proposed to link morphosyntactic representations to conceptual-semantic systems in the anterior temporal lobes (ATL) (Matchin and Hickok 2020; Hickok 2022).

A functional asymmetry of inferior frontal and temporal regions has also been observed in recent fMRI investigations (Matchin and Wood 2020; Giglio et al. 2022b). In Giglio et al. (2022b), participants listened to and produced word sequences of fixed lengths. Following the classical paradigm in Pallier et al. (2011), the words comprised phrases of varying sizes depending on condition. Giglio et al. (2022b) found an effect of constituent size in frontal and temporal regions for both production and comprehension. However, when contrasting production and comprehension regardless of the constituent size, activation of inferior frontal regions was stronger for production, while activation of middle temporal regions was stronger for comprehension. Results from an ROI analysis in their study also indicated that production entails stronger LIFG activity than comprehension, while the opposite, i.e., more activity for comprehension than production, was the case in the LMTG. Notably, Giglio et al. (2022b) used large masks that covered regions outside of the LIFG, which is problematic because areas adjacent to the LIFG (e.g., the left middle frontal gyrus (LMFG)/inferior frontal sulcus) are part of the multiple demand (MD) network (Stiers et al. 2010; Duncan 2013; Camilleri et al. 2018; Wehbe et al. 2021; MacGregor et al. 2022). The MD network plays a crucial role in a set of domain-general processes often denoted by the umbrella term cognitive control or executive functions (e.g., working memory and inhibition; Miller and Cohen 2001) that are recruited while speaking but do not primarily support linguistic processing (Diachek et al. 2020).

The current study had two main objectives. The first was to address the question of the relative contribution of frontal and temporal regions in the left-lateralized perisylvian language network, now in the more ecologically valid conversational setting. We expect the conversational setting to have implications for the division of labor in the language network. For instance, the conversational setting may incur a greater cost on conceptual regions (e.g., ATL), considering the need to integrate context to decode utterance meaning (as discussed in e.g., Grice et al. 1975; Wilson and Sperber 2002). Another possibility is that comprehension employs mechanisms that favor processing speed and undermine the need for a full analysis of the incoming utterance by using lexical information, simple heuristics connecting syntax and semantics (e.g., the first incoming noun phrase is the agent of an action), and world knowledge (see “good enough processing”: Christianson et al. 2001; Ferreira et al. 2001; Townsend and Bever 2001; Ferreira et al. 2002; Ferreira 2003; Ferreira and Patson 2007). This tendency could be due to the demands of timing in conversational turn-taking (Stivers et al. 2009; Kendrick and Torreira 2015; Levinson and Torreira 2015). This is in contrast with the situation in production, where a detailed representation of the utterance always needs to be generated (Bock 1982; Garrett 1988; although, see “good enough production”, Goldberg and Ferreira 2022). Under this explanation, a stronger relative activation is expected for conversational production vs comprehension in regions supporting combinatorial processing on different levels of language (e.g., syntactic, lexical, contextual/socio-pragmatic; potentially a widely distributed network, involving the LIFG, as suggested in Giglio et al. 2022b and Hagoort 2016). Another consideration is that conversational production is qualitatively different from production in isolation, as resources available during conversational comprehension are allocated for production processes. For instance, speakers begin planning their upcoming utterance before their interlocutor has finished their turn (Bogels et al. 2018; 2020). Investigations based on conversational data are, in other words, crucial to understanding the neurocognition of everyday language use.

The second objective of this study was to investigate the relative contribution of regions outside of the perisylvian language network. We were particularly interested in the potential division of labor in regions previously implicated in socio-pragmatic processing (above and beyond word- and utterance/sentence-level phonology, syntax, and semantics). For instance, the medial prefrontal cortex has been linked to communicative production planning (Vanlangendonck et al. 2018) and inferring the indirect meanings of utterances (Bavsnáková et al. 2014; Bavsnáková et al. 2015; Bendtz et al. 2022). It is also reasonable to believe that regions related to the reasoning about others’ mental states [i.e., theory of mind regions, e.g., the temporo-parietal junction; the superior/middle frontal gyrus; Schurz et al. 2014] are recruited during conversational production and comprehension. However, these regions have not been localized for production and comprehension in a conversational context (see, however, Rauchbauer et al. 2019, Hogenhuis and Hortensius 2022; who both used fMRI to investigate conversational processes on human–robot interaction, as further discussed in the supplement) and the relative contribution of these regions across systems is unknown.

In summary, we aimed to investigate the neural architecture of conversational production and comprehension. We were particularly interested in eventual similarities and asymmetries in the recruitment of (1) inferior frontal and superior/middle temporal regions within the perisylvian language network, and (2) regions outside the perisylvian language network, linked to socio-pragmatic processing. Moreover, in modern research on the neurobiology of language, divisions are made between the frontal and temporal nodes (e.g., Matchin and Hickok 2020; Hickok 2022; Matchin et al. 2022; Giglio et al. 2022a, b). As we were interested in these differences, we chose an ROI approach (also following Hu et al. 2022; Giglio et al. 2022b) to complement a whole-brain analysis.
Materials and methods

Data

Raw MRI images and TextGrid-formatted orthographic transcriptions were retrieved from a publicly available data set provided by Rauchbauer et al. (2020). The MRI data and transcriptions were retrieved from OpenNeuro and Ortolang (https://hal.archives-ouvertes.fr/hal-02612820, https://www.ortolang.fr/market/corpora/convers/v2). The 25 participants in Rauchbauer et al. (2020) reported normal or corrected-to-normal vision and had no prior history of psychiatric or neurological conditions. One of these participants was excluded from the present study because of excessive head movement (max movement > 4 mm). Included in the main analysis were 24 participants (18 female, 6 male, M age = 28.8, SD = 12).

In the Rauchbauer et al. (2020) corpus, participants held conversations in their L1 (French) with a confederate in the control room. The confederate was either an experimenter or a robot (controlled by the experimenter through a Wizard of Oz paradigm), but in the present study, we were only interested in human–human conversation. We, therefore, modeled the images acquired during the human–robot conversations in the same way (using the same categories, see below) but separately from the images acquired during the human–human conversations. In the present study, only data from human–human conversations (12 min/participant in total) were used in the first-level contrasts and the second-level analysis. Henceforth, all the mentioned events refer to those of human–human conversations.

Interlocutors were connected via bidirectional audio (using active noise-cancellation), and unidirectional video transmission (the participant saw the confederate’s face on a video monitor, but not vice versa). To provide a framework for naturalistic conversation, participants were told that they would discuss images from an advertising campaign with another participant. These images portrayed anthropomorphized fruits (i.e., fruits with faces). Rauchbauer et al. (2020) reported that all participants confirmed that they believed the cover story after participation. There were four runs per participant, each consisting of six blocks with the following block structure: 8-s presentation of the image of the fruit, 4-s fixation cross, 1-min conversation with the confederate, 4-s fixation cross.

Rauchbauer et al. (2020) collected MRI data with a 3T Siemens Prisma and a 20-channel head coil. Functional images were acquired using an EPI sequence with the following parameters: echo time (TE): 30 ms, repetition time (TR): 1205 ms, matrix size: 84 × 84, field of view (FOV): 210 mm × 210 mm, voxel size (VS): 2.5 × 2.5 × 2.5 mm3, 54 slices co-planar to the anterior/posterior commissure plane (axial), flip angle: 65 deg. Functional images were acquired with multiband acquisition factor 3. Parameters for the acquisition of structural images were: TE: 0.002 ms, TR: 2.4 ms, FOV: 204.8 × 256 × 256 mm, VS: 0.8 × 0.8 × 0.8 mm3, 320 slices (sagittal).

Rauchbauer et al. (2020) automatically segmented audio files of speech from individual speakers into inter-pausal units (blocks of speech surrounded by silences ≥ 200 ms) that were visually inspected and manually transcribed. In the present study, we extracted onsets and offsets of three events: production (when the participant spoke), comprehension (when the confederate spoke), and silence (when both interlocutors were silent), from the transcribed data in Rauchbauer et al. (2020). This extraction was performed using a Python script (https://github.com/carolinearvidsson/RobotfMRI). To avoid extremely short events in the analysis, utterances were merged into a single utterance if they were surrounded by silences < 300 ms within the same speaker. Utterances shorter than 300 ms were removed.

Ethical statement

The acquisition and availability of the Rauchbauer et al. (2020) corpus were approved by the ethics committee “Comité des Protection des Personnes Sud Mediterranéen I”. No sensitive data have been used or reported in the current study. The study has been conducted following the EU’s data protection law.

fMRI Preprocessing

Preprocessing was performed using fMRIPrep (v21.0.1; Esteban et al. 2019). The T1-weighted (T1w) image was skull-stripped and corrected for intensity non-uniformity, and furthermore used as reference throughout the workflow. Brain tissue segmentation of cerebrospinal fluid, white matter, and gray matter was performed on the brain-extracted T1w. Volume-based spatial normalization to two standard spaces (MNI152NLin2009cAsym, MNI152NLin6Asym) was performed through nonlinear registration with ‘antsRegistration’ (Avants et al. 2009), using brain-extracted versions of both T1w reference and the T1w template.

For each of the BOLD runs per subject, the following preprocessing was performed. Head-motion parameters (transformation matrices, and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using mcflirt (part of the FSL package; Jenkinson et al. 2012). The BOLD time-series were resampled onto their original, native space by applying the transforms to correct for head motion. The BOLD reference was then co-registered to the T1w reference. Co-registration was configured with 6 degrees of freedom. Additionally, a set of physiological regressors was extracted to allow for component-based noise correction. Automatic removal of motion artifacts using independent component analysis was performed after the removal of non-steady state volumes and spatial smoothing with an isotropic, Gaussian kernel of 6-mm FWHM (full-width half-maximum). Head movements in coordinates x, y, and z were inspected independently. As previously mentioned, one participant had > 4 mm in max head movement and was therefore excluded from the following analyses. The BOLD time-series were resampled to MNI152NLin2009cAsym standard space. The preprocessed AROMA-cleaned images were used in further analyses.

Whole-brain analysis

For the first-level single-subject analysis, production, comprehension, and silences were modeled as three separate regressors. Images acquired during the presentation of the fixation cross (fixation) and the presentation of the advertising image (advertisement) were modeled as two separate regressors. Head movements were modeled as six motion parameters. The events were convolved with a canonical hemodynamic response function. The three regressors used in the contrasts were production, comprehension, and fixation. Production and comprehension were contrasted against fixation and each other: production > fixation, comprehension > fixation, production > comprehension, comprehension > production. The number and duration of the production and comprehension events are available in Table 1. The second level analysis was conducted with one-sample t-tests on the contrast images defined at the first level. A cluster-forming threshold of an uncorrected P-value was set to 0.001 (no extent-level threshold, k = 0). Family-wise error, as implemented in SPM12, was used as the multiple comparison correction.
**Table 1.** The mean number of production and comprehension events per participant.

<table>
<thead>
<tr>
<th></th>
<th>Production</th>
<th>Comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td>N events</td>
<td>133</td>
<td>129</td>
</tr>
<tr>
<td>Mean duration (s)</td>
<td>2.07</td>
<td>1.77</td>
</tr>
<tr>
<td>SD duration (s)</td>
<td>1.68</td>
<td>1.39</td>
</tr>
<tr>
<td>Range duration (s)</td>
<td>0.3–10.43</td>
<td>0.3–9.96</td>
</tr>
</tbody>
</table>

The mean, SD, and range of the event durations are given in seconds. Events shorter than 0.3 s were removed from the analysis.

method (cluster and peak level). Only clusters with \( p_{\text{FWE}} \) < .05 at cluster level were reported in the current investigation. The test statistic of each cluster’s highest peak (voxel) is also reported. No additional voxels were reported, even if they were significant at \( p_{\text{FWE}} \) < .05 at the voxel level. Cluster labeling was performed using the Automated anatomical labeling atlas toolbox for SPM (Rolfs et al. 2020).

**ROI analysis**

Publicly available functional masks or "language parcels" (downloadable at https://evlab-mit-edu.ezp.sub.su.se/funcloc/), were intersected with subject-specific gray-matter probability maps to account for individual variation. The functional masks originally defined in Fedorenko et al. (2010) were since then updated by increasing the number of participants to N=220. These masks were derived from a probabilistic activation overlap map using the contrast listening to sentences > nonwords and designed to (1) include voxels that selectively show activation for higher-order linguistic processing, (2) exclude lower level phonetic-articulatory activation, and (3) decrease the influence of higher-level, domain-general (e.g. MD) activation. Moreover, our subject-specific gray matter probability maps (from fMRIprep) were thresholded \( P \geq 0.2 \) (this threshold is common when using structural probability maps; Johnson et al. 2005; Taki et al. 2011; Callaert et al. 2014; Dukart and Bertolino 2014; Zhang et al. 2021). The four ROIs were labeled anterior and posterior LIFG, anterior LMTG/STS, and posterior LMTG/STS (antLIFG, postLIFG, antLMTG/STS, antLMTG/STS; see Fig. 3). We note that the LpostIFG did not include any voxels from the frontal operculum, while the LIFG did have a small overlap with this area.

We extracted mean beta values per participant in each of these ROIs for production and comprehension relative to baseline using MarsBar (Brett et al. 2002) in SPM12. Using beta value as dependent variable, we ran a linear mixed effects model to investigate the interaction effect of LOBE (frontal, temporal) and SYSTEM (production, comprehension), with by-participant random intercepts. The linear mixed models were conducted using the lme4 package (Vazquez et al. 2010) in R (R Core Team 2021), with an alpha level of \( \alpha = 0.05 \). P-values were retrieved using R package afex (Singmann et al. 2018). One sample t-tests were conducted to investigate whether the mean beta weights differed from zero. Paired sample t-tests were conducted to compare ROI recruitment in production with comprehension. As a multiple comparisons method, we used Bonferroni correction for four comparisons (one comparison per ROI).

**Results**

**Whole-brain analysis**

We investigated the main effect of production and comprehension against the baseline (looking at a fixation cross). The contrast production > fixation yielded clusters in the STG, MTG, and the superior temporal pole (STP) bilaterally. Significant activations were also observed in the LIFG, encompassing pars orbitalis, triangularis, and opercularis, as well as in the pre- and postcentral gyrus. Other activated areas in this contrast were the left and right occipital gyri, right insula, bilateral cerebellum, bilateral supplementary motor area, pre/postcentral gyrus, and the right cuneus. The contrast comprehension > fixation generated large bilateral clusters spanning from the MTG/STG to the STP and the middle temporal pole. Comprehension activation was also observed in the occipital gyri bilaterally and the right cerebellum (see Fig. 1 and Supplementary Table S3 for cluster labels, local maxima coordinates, and test statistics from all whole-brain contrasts). To evaluate the extent of the production-comprehension overlap, we performed a conjunction analysis, using a conservative method (Nichols et al. 2005) showing activity where both systems are significantly active, individually. This analysis revealed clusters in the bilateral MTG/STS, LIFG (pars triangularis, opercularis, and orbitalis), the pre- and postcentral gyrus, and the occipital gyri. We also contrasted the systems against each other to understand if any areas were overall more active in production or comprehension (see Fig. 2). The contrast production > comprehension generated stronger activity in lateral and medial frontal areas, including the LIFG (pars orbitalis, opercularis, and triangularis), frontal operculum and anterior insula, the bilateral middle and superior frontal gyrus (SFG), the bilateral anterior cingulate cortex

![Fig. 1. Whole-brain results for production vs fixation (left; blue) and comprehension vs fixation (right; yellow). The red outline shows the results of the conjunction analysis, i.e. where production and comprehension activation overlapped. The figure shows clusters with a cluster-forming threshold of \( p_{\text{uncorr}} = .001 \). Only clusters with \( p_{\text{FWE}} < .05 \) are reported in text and figures.](https://example.com/f1.png)

![Fig. 2. Whole-brain results from contrasting systems against each other. Blue: areas more active in production than comprehension. Yellow: areas more active in comprehension than production. Saggital slice at MNI coordinate x: −38 confirms stronger activation for production, compared to comprehension, in the LIFG/frontal operculum, not visible in the surface rendering. See Fig. 1 for details of the visualization settings.](https://example.com/f2.png)
Fig. 3. Top line plot: Interaction of LOBE (frontal, temporal) and SYSTEM (production, comprehension). Lines (blue: production; red: comprehension) show the mean and line ribbons show the standard error of the mean. Bottom violin plots: ROI-wise differences in BOLD response between conversational production and conversational comprehension. The functionally defined language ROIs included the anterior and posterior inferior frontal gyrus (LanIFG, LpIFG) and anterior and posterior left middle temporal gyrus/sulcus (LanMTG, LaMTG). Violin plots show the density distribution of the participants’ mean beta weights from the contrasts production/comprehension vs baseline (white: production, gray: comprehension) and ROI. The lines across violinss how the difference in means across tasks within each ROI. The color bar shows the number of overlapping subject-specific masks in the displayed voxels. Star notation: * P < .05, *** P < .001. P-values Bonferroni-corrected for multiple comparisons (one test per ROI).

and the bilateral supplementary motor area. Activation stronger in production compared to comprehension was observed in the precuneus, the pre- and postcentral gyri, the cuneus and the occipital gyri, bilaterally. The contrast comprehension > production generated two clusters: a left-lateralized cluster spanning from the MTG/STG, reaching to the superior ATL, and a right-lateralized cluster in the STG/MTG.

ROI analysis
All eight distributions of beta weights (one distribution per system in each ROI) individually passed a Shapiro–Wilks normality test (P-values were larger than 0.9). The mixed effects model with LOBE and SYSTEM as factors and participants as random effects showed that the effect of LOBE was significant, so that temporal lobe activation was stronger than frontal lobe activation (β = 1.09, SE = 0.37, t(165) = 2.93, P < .01). Activation overall was stronger in production than comprehension (β = 1.14, SE = 0.37, t(165) = 3.06, P < .01). We report these results for completeness, although we set up the model mainly to test the interaction. Crucially, there was a significant interaction between LOBE and SYSTEM (see Fig. 3), showing opposite signs of LOBE activity differences (LanIFG vs LpMTG, LpMTG). Comprehension activation was also significantly different from zero (LanIFG: t(23) = 5.78, P < .001; LpIFG: t(23) = 6.20, P < .001; LanMTG: t(23) = 2.87, P < .01; LpMTG: t(23) = 4.75, P < .001). Comprehension activation was also significantly different from zero (LanIFG: t(23) = 3.46, P < .01; LpIFG: t(23) = 4.04, P < .001; LanMTG: t(23) = 7.58, P < .001; LpMTG: t(23) = 6.71, P < .001). The mean beta values in production and comprehension differed significantly from each other in the LpIFG and the LaMTG, but not in the LaIFG or the LpMTG (LanIFG: t(23) = 1.95, P < .06; LpMTG: t(23) = −1.58, P < .12). All significant P-values reported in the ROI analysis survived Bonferroni correction for four comparisons (one test per ROI).
Discussion

Using a neuroimaging paradigm where participants engaged in unscripted conversations, we provide evidence that aspects of both inferior frontal and superior/middle temporal regions in the perisylvian language network subserve conversational production and comprehension, but to different extents. Our conjunction analysis shows that the conversational production and comprehension systems largely share neural infrastructure in the perisylvian language network. However, there was activation not seen in the conjunction, including (1) comprehension activation in the anterior aspects of the bilateral temporal lobes, and (2) production activation in the anterior and posterior LIFG, indicating some system-specific processing. Contrasting the two systems further revealed a functional asymmetry within this shared network: the recruitment of the posterior LIFG was stronger in production than in comprehension, while the recruitment of the anterior aspects of the temporal lobe (the left anterior superior temporal gyrus/sulcus, LaSTG/STS; the left anterior middle temporal gyrus, LaMTG) was stronger in comprehension than in production.

The asymmetric recruitment pattern found further support in our ROI analysis, showing significantly greater activation of the posterior LIFG in production than comprehension. In turn, LaMTG/STS was more strongly recruited during comprehension than production. While these results will certainly remind the reader of the classical model (Wernicke and Eggert, 1885/1977; Geschwind 1970; Tremblay and Dick 2016; Rutten 2022), this more anterior location of the temporal activation for conversational comprehension and the functional overlap of the two systems are crucial differences relative to that model (further discussed below).

As expected, we show that conversational production and conversational comprehension engage regions outside of the perisylvian language network. Producing language, in a conversational context, involves the recruitment of the SFG, motor regions, and the medial frontal cortex. In the more crucial contrast of production vs comprehension systems, production recruited regions involved in higher-level sociocognitive processing, such as the bilateral medial prefrontal cortex (mPFC), which has been implicated in communicative perspective-taking during utterance planning (Vanlangendonck et al. 2018), along with the superior/middle frontal gyrus. Interestingly, no regions outside of the perisylvian language network are more activated for conversational comprehension than conversational production.

Production-comprehension overlap

Our results support the notion of partial overlap of conversational production and comprehension systems in the frontal and temporal lobes. Aspects of this overlap are predicted by a recent version of the Dual-stream model (Hickok 2022). In this model, a network subserving both comprehension and production is proposed, reaching from the posterior to the ATL. Compared to other studies investigating the production-comprehension overlap using whole-brain analysis (Okada and Hickok 2006; Menenti et al. 2011; Segaert et al. 2012; 2013; Giglio et al. 2022b), our results indicate the largest production-comprehension overlap in the bilateral temporal lobes to date. A unique contribution of our results is thus that the right pSTG/STS/MTG subserves not only comprehension but also production. Our results may raise a call for further specification of the Dual Stream model, e.g. by highlighting the involvement of both hemispheres. The involvement of the rMTG in production may reflect the processing of communicative actions (Stolk et al. 2013), while rSTG/STS involvement possibly reflects prosody production (Sammier et al. 2015) as well as the self-monitoring of speech (Indefrey and Levelt 2004), all of which might be enhanced in the conversational setting.

Notably, two of our results are not predicted by the updated Dual Stream model. First, the model describes the LIFG as mainly linked to production, but our conjunction analysis identifies the LIFG as shared across systems (albeit the posterior LIFG shows greater activation for production in our ROI analysis). In addition, the model describes the ATL as shared, but our whole-brain results suggest comprehension-selective activation of the anterior aspects of the ATL (we elaborate on these LIFG and ATL findings in the second and third paragraphs in the next section).

Functional asymmetry in the perisylvian language network

The results extend the functional asymmetry of frontal and temporal regions in the perisylvian language network, previously observed in the controlled production and comprehension experiment of Giglio et al. (2022b); to the crucial conversational setting. Compared to the ROI analysis in Giglio et al. (2022b); our ROIs were based on functional masks specifically designed to tap higher-order linguistic processing and account for individual variation in gray matter structure. Despite differences with their study, we still observed similar results, which, contrary to recent accounts (Hu et al. 2022), favor the notion of a functional asymmetry of the discussed frontal vs temporal regions.

A plausible explanation for the observed functional asymmetry between inferior frontal and temporal regions is that the LIFG may facilitate supramodal syntactic parsing (or even building a full syntactic representation, available at some point during processing; Uddén et al. 2022). Under this explanation, this parsing process is essentially shared across production and comprehension, but can however be circumvented in comprehension and not in production. Our results point in the direction that these “good enough” comprehension processes, previously mentioned in the introduction, are supported by the LaMTG/STS. This explanation is consistent with (1) stronger recruitment of the LIFG during production than during comprehension in tasks where participants are required to build syntactic representations (e.g. produce sentences while describing events depicted in images, as in Giglio et al. (2022b) and Hu et al. (2022); or produce utterances while engaging in conversation, as in our study) and (2) absence of system differences in LIFG recruitment in tasks where participants repeat sentences that they hear (e.g. Matchin and Wood 2020).

We observe a functional asymmetry of production and comprehension processes in the ATL, unlike Hu et al. (2022). The anterior temporal cortex is considered a semantic hub that organizes semantic features of mental concepts (Patterson et al. 2007; Lambon Ralph and Patterson 2008; Hickok 2009). Conversational comprehension involves analyzing utterances by integrating contextual information to resolve semantic ambiguities and decipher indirect meanings (Grice et al. 1975). Tasks that do not require the integration of e.g. prior linguistic information or other contextual cues, may not strongly engage these complex semantic processes in comprehension (e.g. Hu et al. 2022). On the other hand, the whole-brain results in Giglio et al. (2022b) did indicate a stronger activation of the LaMTG in comprehension than production. This observed asymmetry could be ascribed to their use of embedded clauses (e.g. “The woman saw that the man clapped”), which not only requires more syntactic, but also semantic, combinatorial processes (for an account of the ATL in semantic composition, see Pylkkänen 2020).
Separable view of production and comprehension

Our results provide neurobiological support for the so-called separable view of production and comprehension systems (Kittredge and Dell 2016), often endorsed in psycholinguistics (Gahl and Strand 2016; Meyer et al. 2016). According to the separable view, speaking and listening “cannot be understood as the same processes running in opposite directions” (Meyer et al. 2016), because processes of one system may be specific to or more important to that system. This claim is supported by studies, e.g. suggesting (1) that brain damage can selectively impair one of the systems (e.g. Martin et al. 1999), and (2) that pre-school-aged speakers of Turkish correctly choose specific morphological forms before correctly interpreting these forms in utterances (Ünal and Papafragou 2016). Importantly, the separable view holds that the systems are not completely independent, as some processes used in one system are also used by the other system (see e.g. Kittredge and Dell 2016; Hickok 2022). Therefore, this view is compatible with previous studies that have found evidence of a shared infrastructure of individual subprocesses (e.g. syntactic, lexical, phonological) within the two larger systems (e.g. Menenti et al. 2011; Segaert et al. 2012; 2013).

New perspectives from conversation

In this study, we investigated the roles of speaker vs listener in the conversational setting. Our results draw attention to the possibility that the speaker, to a greater extent than the listener, may draw on higher-level sociopragmatic processing in the bilateral mPFC, the LMFG, and the bilateral SFG. These regions are part of the theory of mind network (Astonington et al. 1988; Schurz et al. 2014) and have been implicated in communicative vs non-communicative production planning, where LMFG involvement was specific to when the speaker’s visual perspective differed from the addressee’s (Vanlangendonck et al. 2018).

Our approach of investigating production and comprehension as whole systems diverges from focusing on individual subprocesses (e.g. syntactic parsing in Segaert et al. 2012; 2013; Matchin and Wood 2020; Hu et al. 2022 or lexical access in Hu et al. 2022). We adopt this approach for several reasons. Crucially, the granularity of our investigation matches the overarching intentional structure of conversation (for a discussion of granularity, see Poeppel and Embick 2017). This intentional structure is sometimes referred to as “the speaker-listener gap” (Brown 1995, p. 25) and builds on key differences between the speaker and the listener. For the informative function of language (Jakobson 1976), the speaker strives to be sufficiently informative and the listener strives to obtain a sufficient level of information, while both strive to minimize cognitive effort. A multitude of processes have evolved to implement these two functions and may e.g. compensate each other, with the goal of functionality at the role level of speaker or listener.

Moreover, we gladly note that we are not the first to pose questions on this functionally appropriate and broader level. For instance, Giglio et al. (2022b) investigated the neural infrastructure of production and comprehension in non-conversational contexts, and their analysis included the contrasting of entire systems. Menenti et al. (2011) also formulated their question on this level, but focused on three individual processes (lexical, semantic, and syntactic), also in a non-conversational context. It is important to complement the investigation of single processes, even in combination as in Menenti et al. (2011), with an investigation of the systems in their entirety. Our study’s unique contribution is this broader focus on production and comprehension, but now in the conversational context.

Limitations

Two limitations of the study relate to core features of conversation. First, in conversation, it is not uncommon for speech from two individuals to overlap, e.g. during the transition from one speaker’s turn to another speaker’s turn (Sacks et al. 1978; Heldner and Edlund 2010). In our study, overlapping speech leads to overlapping production and comprehension events. We do not regard this as a crucial issue, partly because conversational production and comprehension processes naturally overlap, even when there is no overlapping speech (Levinson and Torreira 2015; Bögels et al. 2018); the listener plans and encodes their upcoming response when the incoming turn from their interlocutor is still unfinished and needs to be monitored. This production and comprehension-overlap is a necessary condition for studying these processes in conversational settings and thus part of our unique contribution. Furthermore, our main result is the asymmetric recruitment pattern of production and comprehension. Such an asymmetry could not be driven but obscured by overlapping production and comprehension processes. Second, conversational turns are often short in time (the mean duration in the current study was 1.67 for production and 1.96 for comprehension, see Table 1). Low-duration events (i.e. trials) are not standard in fMRI data analysis, because most MRI studies are based on controlled experiments where the timing of events is predetermined. However, methodological research has shown that responses to stimuli with durations as short as 5 ms can be reliably detected with fMRI (Yeşilyurt et al. 2008).

Summary

We have addressed the long-lived yet ongoing debate of whether there is a functional division of production and comprehension systems in frontal vs temporal perisylvian language regions, finally including the conversational context. Our whole-brain results can be understood in terms of partial overlap of production and comprehension systems. Specifically, the whole-brain results indicate (1) shared neural infrastructure including inferior frontal and superior/middle temporal perisylvian regions, (2) functional asymmetry of these shared loci, in which production more strongly recruits the LIFG, while comprehension more strongly recruits the left ATL (extending the results in Giglio et al. 2022b), (3) comprehension-selective activation (i.e. comprehension activation in voxels not covered by the conjunction activation map) in the anterior aspects of the bilateral temporal lobes, and (4) production-selective activation in the anterior and posterior LIFG and midline structures of the frontal cortex. Our ROI analysis further supports (1)–(2), suggesting that all the functionally defined language ROIs were activated for both systems but modulated differently.

Conclusions

We have provided evidence favoring partial overlap of the production and comprehension systems in the inferior frontal and temporal perisylvian language regions. Within this overlap, we observed functional asymmetry in the expected direction following the classical Broca–Wernicke model. Apart from the functional overlap, our results depart from the classical model in the greater engagement of anterior, rather than posterior, temporal regions during comprehension compared to production. Our results also depart from accounts in which production incurs a greater cost for
all regions in the language network (e.g. Hu et al. 2022), or where comprehension the comprehension system does not involve the LIFG (Hickok 2022). This is an example of how investigations of conversational production and comprehension will alter current models built on isolation paradigms. These results, together with the results on system asymmetries outside the classical language regions, indicate that the production and comprehension systems share processes, although some processes may be more important, or even dedicated, to one of the systems. In conclusion, while each system includes unique regions and is not a subset of the other, conversational production and comprehension share crucial neural infrastructure.

Acknowledgments

We would like to thank Dr. Rita Almeida at Stockholm University Brain Imaging Centre for support with fMRI data preprocessing and analysis.

CRediT statement

Caroline Arvidsson (Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing—original draft, Writing—review and editing), Ekaterina Torubarova (Software, Writing—review and editing), André Pereira (Funding acquisition, Project administration, Supervision, Writing—review and editing), Julia Uddén (Conceptualization, Data curation, Funding acquisition, Project administration, Software, Writing—review and editing).

Supplementary material

Supplementary material is available at Cerebral Cortex online.

Funding

This work was supported by Digital Futures project ”Using Neuroimaging Data for Exploring Conversational Engagement in Human-Robot Interaction”. JU received additional support from Bank of Sweden Tercentenary Foundation (http://dx.doi.org/10.13039/501100004472) and the Swedish Collegium of Advanced Studies and Stiftelsen Marcus och Amalia Wallenbergs Minnesfond (2022.0034).

References


Bock JK. Toward a cognitive psychology of syntax: information processing contributions to sentence formulation. Psychol Rev. 1982:89(1):1–47.


Fridriksson J, Fillmore P, Guo D, Rorden C. Chronic Broca’s aphasia is caused by damage to Broca’s and Wernicke’s areas. Cereb Cortex. 2015:25(12):4689–4696.


