Unraveling the functional attributes of the language connectome: crucial subnetworks, flexibility and variability


To cite this version:
hal-03811050

HAL Id: hal-03811050
https://hal.univ-grenoble-alpes.fr/hal-03811050
Submitted on 24 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Unraveling the functional attributes of the language connectome: crucial subnetworks, flexibility and variability

E. Roger a,⁎, L. Rodrigues De Almeida a, H. Loevenbruck a, M. Perrone-Bertolotti a, E. Cousin a, J.L. Schwartz b, P. Perrier b, M. Dohen b, A. Vilain b, P. Baraduc b, S. Achard b, M. Baciu a

a CNRS, LPNC, Université Grenoble Alpes, BSHM, Université Savoie Mont Blanc, UMR 5105, Celex, Grenoble 38000, France
b CNRS, GIPSA-Lab, Grenoble-INP, Université Grenoble Alpes, CNRS, Grenoble 38000, France

ARTICLE INFO

Keywords:
Language
fMRI
Connectome
Functional connectivity
Networks

ABSTRACT

Language processing is a highly integrative function, intertwining linguistic operations (processing the language code intentionally used for communication) and extra-linguistic processes (e.g., attention monitoring, predictive inference, long-term memory). This synergetic cognitive architecture requires a distributed and specialized neural substrate. Brain systems have mainly been examined at rest. However, task-related functional connectivity provides additional and valuable information about how information is processed when various cognitive states are involved. We gathered thirteen language fMRI tasks in a unique database of one hundred and fifty neurotypical adults (Int.Lang [Interactive networks of Language] database), providing the opportunity to assess language features across a wide range of linguistic processes. Using this database, we applied network theory as a computational tool to model the task-related functional connectome of language (LANG atlas). The organization of this data-driven neurocognitive atlas of language was examined at multiple levels, uncovering its major components (or crucial subnetworks), and its anatomical and functional correlates. In addition, we estimated its reconfiguration as a function of linguistic demand (flexibility) or several factors such as age or gender (variability). We observed that several discrete networks could be specifically shaped to promote key functional features of language: coding-decoding (Net1), control-executive (Net2), abstract-knowledge (Net3), and sensorimotor (Net4) functions. The architecture of these systems and the functional connectivity of the pivotal brain regions varied according to the nature of the linguistic process, gender, or age. By accounting for the multifaceted nature of language and modulating factors, this study can contribute to enriching and refining existing neurocognitive models of language. The LANG atlas can also be considered a reference for comparative or clinical studies involving various patients and conditions.

1. Introduction

Language is optimized for human communication. It is an efficient vector of information transmission, shaped under cultural and environmental constraints to meet physical, technological, and social needs (e.g., Kirby et al., 2015; Lupyan and Dale, 2016; Millikan, 2005; Scott-Phillips, 2015; Tamariz and Kirby, 2016). Language is also adapted to thinking and interpretation, playing a scaffolding role in cognition (e.g., Carruthers, 2002; Chomsky, 2014; Clark, 2006; Jackendoff, 1996; Reboul, 2015), and is crucial in meta-cognition including self-evaluation, self-regulation and autoeonic consciousness (Alderson-Day and Fernyhough, 2015; Perrone-Bertolotti et al., 2014).

To combine both effectiveness (Gibson et al., 2019) and utility (Jaeger and Tily, 2011) of language production and comprehension, several essential abilities are required. The first one, is the combinatorial skill (Boer et al., 2012; Friederici et al., 2017; Zuidema and de Boer, 2018). Language is compositional and recursive, implying specialized processing of intra-linguistic aspects (i.e., an aptitude to handle various combinatorics, perceptive, syntactic, or semantic/conceptual; Pylkkänen, 2019). A second important ability relates to multisensory integration, which facilitates spoken communication and enhances speech intelligibility (Chandrasekaran et al., 2009; Ghazanfar and Schroeder, 2006; Luo et al., 2010; Noppeney et al., 2008; Schroeder and Foxe, 2005; Schwartz et al., 2004; Sumby and Pollack, 1954). Beyond the multisensory facilitation (or low-level multimodal integration; Holler and Levinson, 2019), high level cognitive abilities related to top-down multimodal mechanisms have also been emphasized. A shared understanding, a relevant and contextually-adapted discourse, requires aligning the partners’ representations, considering shared knowledge, past experiences, or even making assumptions about the other’s perspectives. Establishing “common ground” between conversational partners (Clark and Marshall, 1981) relies on a wide range of “high-level” cogni-

Received 26 April 2022; Received in revised form 4 October 2022; Accepted 5 October 2022
Available online 6 October 2022.
1053-8119/© 2022 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
tive functions such as working memory (resonance-based theory of common ground, Horton, 2007), long term memory (Brown-Schmidt and Duff, 2016), theory of mind or mentalizing (Vanlangendonck et al., 2018) processes. Language use is therefore adapted online, enabling communication in a range of environmental and social contexts, meeting various cognitive demands and individual metacognitive needs. Given the pressure exerted by communication, cognitive and metacognitive demands, language has evolved as an adaptive and complex system, which requires considering external (context) and internal (individual needs and goals) signals while processing both intra- and extra-linguistic signals (e.g., Holler and Levinson 2019, for a multimodal language-in-situ framework.

How are these various abilities integrated to sustain language functions and how are they implemented in the brain? The exploration of brain networks and their unique lens for understanding cognitive function has become an important part of the cognitive neuroscience landscape (Fornito et al., 2013). Brain network descriptions have revealed that the brain is organized as a “small-world” network (Achard and Bullmore, 2007), favoring optimization of information transfer (Laughlin and Sejnowski, 2003). A balance between segregation and integration characterizes this organization, i.e., by short communication pathways creating specialized subsystems (segregation), whose interconnectivity is coordinated by distant, highly connected brain regions (integration; van den Heuvel and Sporns, 2013a). Functionally, local systems or highly connected “modules” for the processing of information in a given modality (visual, auditory, etc.) are linked together by sparse and specific fiber paths over long distances, according to a connectivity principle of “local richness and long-range sparseness” (Pulvermüller, 2018). This organization allows efficient serial, parallel, and distributed brain activity (Herbet and Duffau, 2020). Integrative areas (sometimes referred to as connector hubs) at the interface between local systems are significant for the multimodal neural integration of information (Cocchi et al., 2013; Fornito et al., 2015; van den Heuvel and Sporns, 2013b). Regarding language circuitry, integration/segregation subsystems and specific connector hubs have been previously identified for both language production and comprehension (e.g., Friederici, 2012; Hagood, 2016; Hertrich et al., 2020; Roger et al., 2022). However, language is multi-faceted, and a comprehensive analysis of the functional properties of the language connectome in a broader framework would contribute to a more accurate description.

To fill the gap, we propose, in this study, to examine the functional attributes of language at the brain level through an integrative perspective, mixing several linguistic tasks explored in the light of graph theory. This research is naturally framed within the substantial legacy of the study of language and its brain foundations, whose growing and diverse observations have been accompanied by the evolution of investigative techniques. In the past decades, many authors have highlighted brain function and structure associated with language through theoretical neurocognitive models (e.g., Duffau et al., 2014; Friederici et al., 2017; Hagood, 2016, 2019; Hickok and Poeppel, 2007; Indefrey, 2011; Levelt, 1989; Price, 2012; Rauschecker and Scott, 2009). However, this study, in direct continuation with past legacy, adds further value by investigating functional cerebral connectivity (FC) based on task data. The anatomic-functional substrates associated with language are indeed highly task-dependent (Hickok and Poeppel, 2000).

The task-based FC analyses presented here rely on an fMRI database compiling a broad spectrum of language-related tasks (Interactive networks of Language database: InLang database). InLang comprises thirteen language tasks, performed cross-sectionally by 150 right-handed neurotypical adults. The database is unique in that it covers a broad spectrum of language features: semantic and conceptual processing, decoding (phonology, sound), lexico-syntactic formulation (production), dialoguality (social aspects of language), monitoring of self and others, and unintentional speech (Fig. 1A; see also the Supplementary Material for more details). Such a database is essential to uncover the functional architecture of the multifaceted language processes in an integrative approach. It makes it possible to model a comprehensive connectomic atlas of language and to explore its fundamental properties in depth.

To this end, we analyzed task connectomes – where interregional FC was estimated by beta-series correlations – using graph metrics applied at multiple scales. It allowed us to expose: (1) the overarching FC profile of different language tasks and latent subprocesses; (2) the architecture of the general language connectome (LANG); (3) the functional roles of crucial language subnetworks and brain regions; (4) the anatomofunctional correlates; and (5) the flexibility and variability exerted on the LANG connectome. In terms of interindividual FC variability, gender (e.g., Filippi et al., 2012; Zhang et al., 2018) or age (e.g., Edde et al., 2021; Jockwitz and Caspers, 2021) are potentially important criteria to consider. Without any a priori hypothesis on the pattern of potential changes related to these factors or their behavioral relevance, we have attempted to identify, in an exploratory way, the main variations in the distribution of key regions belonging to the LANG connectome.

Fig. 1 provides an overview of the InLang database and the methodology used to address these 5 main axes.

2. Materials and methods

2.1. Dataset

The InLang database contains data of 13 different linguistic tasks from 7 previously published fMRI protocols (Baciu et al., 2016; Banjac et al., 2020; Grandchamp et al., 2019; Haldin et al., 2018; Hoyau, 2018; Perrone-Bertolotti et al., 2011; Perrone-Bertolotti et al., 2012; Perrone-Bertolotti et al., 2015, 2017), as well as respective structural MRIs (T1w). In all, 359 functional scans have been acquired between 2010 and 2019 from 150 healthy adults (all right-handed according to the Edinburgh Handedness Inventory (handedness score strictly greater than +80; Oldfield, 1971); 64 females: F/86 males: M; Table 5 contains the subjects’ characteristics, by tasks). The database includes 114 “young/middle-aged adults” (18–59: mean = 36.87; SD = 13.29; 50F/64M) and 36 “seniors” (older adults from 60–85: mean = 70.3, SD = 5.87; 14F/22M) volunteers (Fig. 1). The inclusion and exclusion criteria for participants are described in the Supplementary Material. fMRI and T1w scans of all participants have been formatted in BIDS standard and preprocessed using conventional tools (see Appendix).

2.2. Task-based connectomes

Regions of interest (ROIs) covering both the brain and the cerebellum were defined from 6 mm radius spherical regions built around the 264 coordinates in MNI space proposed by Power et al. (2011). The images used for signal extraction (beta values) were the statistical parametric maps containing the linear contrasts between the HRF parameter estimates for the conditions of interest. Nilearn (https://nilearn.github.io/stable/index.html; Abraham et al., 2014) was used to delineate ROIs and extract the signal. The mean signal for each of these ROIs was extracted by participant and task. Functional connectivity (FC) between brain ROIs was derived from correlating (Pearson correlations) the extracted beta values across subjects, within a given task. The process was repeated for each of the tasks separately, allowing to obtain thirteen 264×264 matrices of interregional FC based on beta-series correlation. Codes and derivatives are available here: https://osf.io/6xm8n/). These matrices of interregional FC were thresholded to obtain the task-based connectomes. We applied a 5% threshold, which defines the 5% of the highest positive correlation values, considered to represent non-spurious intermodal connections. The matrices were binarized: 1 was assigned to internode connections that survived to the given density threshold, and 0 was assigned otherwise. Task-based connectomes were built from these binarized matrices, reduced to a fixed number of edges (top 5%, 3485 edges). Several graph metrics were computed on the task-based connectomes: global (network-wide), intermediate (modularity), and local (nodal), using Networkx
Local efficiency of a node $i$ corresponds to the average global efficiency of a subgraph induced by the neighbors of $i$ (Latora and Marchiori, 2001):

$$E_{loc}(G) = \frac{1}{N} \sum_{i \in G} E_{glob}(G_i)$$

where $N$ is the total number of nodes in the network $G$, the distance $d(i,j)$ corresponds to the number of edges in a shortest path between any two nodes $i$ and $j$. $E_{glob}$ represents the capacity of a given network to efficiently integrate and transmit information between the network components or subnetworks (e.g., Bullmore and Sporns, 2012; Roger et al., 2019b; Stanley et al., 2015). The higher the value, the more likely that information transfer is fast.

The mean geodesic cortical distance ($d$) between functionally interconnected nodes ($d(G)$).

We extracted the relative spatial layout of regions along the cortical surface by using existing scripts (https://github.com/margulies/topography/tree/master/utils), based on an algorithm developed to approximate the exact geodesic distance from triangular meshes (Oligschläger et al., 2017). “Physical” geodesic distances between pairs of nodes, estimated in mm, were quantified from the Power’s nodes coordinates projected onto a template surface mesh (fsaverage5). This resulted in a node-by-node matrix of geodesic cortical distance. From the
geodesic distance matrix, we averaged the distribution of distance-to-connected-areas of the relevant functional connections identified from the thresholded FC matrices. We thus obtained the global geodesic distance of the functional connectivity for each of the task-related connectomes. The higher the mean geodesic distance between functionally connected regions, the further apart the connected regions are on average along the cortical surface.

2.3.2. General LANG connectome and subprocesses

From the main global GT parameters described in the previous section (extracted at the connectome-level, for each task), we assessed the similarity of the FC profiles of the language task connectomes. The similarity was quantified using the Euclidean distance. We applied a data-driven hierarchical clustering approach to the similarity matrix and estimated the partitioning. Thus, we identified categories of tasks with similar FC profiles or global network topology. The internal composition and nature of the tasks assigned to each of the identified groups can reveal putative linguistic subprocesses that may be latent and common to several tasks (see also the Supplementary Material for a rationale and a similar method applied to functional activation maps). Starting from the partitioning, we computed the FC matrices of each of these main task groups — hereafter referred to as subprocess connectomes — from the scans of the respectively involved tasks and using the same procedure described above (Section 2.2.). In addition, and still following the same procedure, we generated the general task-based language connectome — abbreviated LANG — corresponding to the FC matrix derived from all language tasks. The global FC parameters reported in the previous section (Section 2.3.1) were also extracted for the LANG and the respective subprocesses connectomes.

Note that only scans corresponding to the young/middle-aged adults of the InLang cohort (n = 114) were considered for the calculation of the FC matrices (of the tasks, subprocesses, and the general LANG connectome) and the main statistical analyses. Scans of ‘older’ adults were included exclusively for statistics regarding the effect of age on LANG hubs (see Section 2.4.5. for a description of the age-related variability analysis).

2.3.3. Intermediate composition of LANG

We performed modularity analyses to determine community structure of the general LANG connectome. We used the Louvain community detection algorithm (Blondel et al., 2008) implemented in Networkx. Louvain’s method is widely used for community detection in neuroscience and has been previously shown to outperform other community detection methods (Yang et al., 2016). To ensure the stability of the final partition, we repeated the modular partitioning process 100 times (Schedlbauer and Ekstrom, 2019) and we evaluated the best LANG partition (Aynaud, 2018) on the matrix averaging the results of all iterations. Each ROI was assigned to a specific community (i.e., a subunit, here denoted LANG “Net”). To facilitate subsequent analyses and interpretations, the Power’s coordinates of LANG ROIs were mapped to the HCP’s multimodal parcellation (version 1.0: HCP_MMP1.0 proposed by Glasser et al., 2016), which consists of 180 brain parcels. Cerebellum coordinates were mapped to the probabilistic human cerebellum atlas SUIT (Diedrichsen et al., 2009).

Still concerning intermediate configurations between the global level of the network and the nodal ROIs level, we focused on identifying densely interconnected subgraphs of LANG. Complete subgraphs, called cliques, are all-to-all connected sets of brain regions providing architecture that isolates information transmission processes (Giusti et al., 2016) and supports efficient and specialized processing (Sizemore et al., 2018). A maximal clique is one that includes the largest possible number of nodes and to which no more nodes can be added. Using Networkx, we estimated \( \omega(G) \) which is the number of nodes in a maximal clique of G and marked the relevant nodes.

2.3.4. Nodal properties of LANG

We calculated the degree centrality (DC, denoted \( k \)) of each node i as the number of adjacent edges to the node \( (k_i) \), from the reduced and unweighted adjacency LANG matrix. DCs are a convenient metric for highlighting brain regions with a high degree of connectivity or “hubs” and form the basis for other measures of nodal graph theory.

From the DCs \( k \), we computed the within-component degree z-score \( z_i \), that expresses the extent to which node i is connected to other nodes in its respective component and is calculated as follows (Guimerà and Nunes Amaral, 2005):

\[
z_i = \frac{k_{i,S} - k_S}{\sigma_S}
\]

with \( k_{i,S} \) the number of connections of node i to the other nodes in the subgraph component S (i.e., the Net) and \( k_S \) and \( \sigma_S \) respectively the mean and SD of the within-component DC over all nodes in S.

To quantify to what extent a node connects across all components, we measured the participation coefficient (PCI). The following conventional formula (Guimerà and Nunes Amaral, 2005) was applied, with \( m \) the set of components S or Nets (here 4):

\[
PCI = 1 - \sum_{S \in m} \left( \frac{k_{i,S}}{k_i} \right)^2
\]

Following Schedlbauer and Ekstrom (2019) and because of the narrow distribution of the PCs we z-scored the coefficients (zPCI) from each network.

The \( z_i \) and \( z_{PCI} \) values have enabled to assign a specific role to each of the LANG ROIs. The nodes were classified according to their type of functional communication within the connectome as follows: connector (high \( z_{PC_i}/z_{PCI} \); high intra-Net and high inter-Net FC); provincial (high \( z_{i}/\text{low } z_{PCI} \); high intra-Net FC); satellite (high \( z_{i}/\text{low } z_{PCI} \); high inter-Net FC); or peripheral (low \( z_{i}/\text{low } z_{PCI} \); low inter-Net FC). We applied this classification as proposed in previous studies (Bertolero et al., 2015; Cohen and D’Esposito, 2016; van den Heuvel and Sporns, 2013b) and with \( z_{i} \geq 0 \) corresponding to “high \( z_i \)” and \( z_{PCI} > 0 \) corresponding to “high \( P_{CI} \)”.

Finally, we were interested in quantifying the rich club organization of the networks. A rich club reflects a set of nodes in the network of whose level of interconnectivity (i.e., richness) exceeds the level of FC that can be expected by chance. For each degree k, the rich-club coefficient (\( \phi \)) is the ratio of the number of actual to the number of potential edges for nodes with degree greater than k (Colizza et al., 2006):

\[
\phi(k) = \frac{2E_k}{N_k(N_k - 1)}
\]

where \( N_k \) is the number of nodes with degree larger than \( k \), and \( E_k \) is the number of edges among those nodes.

We compared and normalized the rich club coefficient to sets of “equivalent” random networks. An empirical null distribution constituted from the average of 1000 random networks of equal size and degree distribution was generated (\( \phi_{rand}(k) \)).

The difference between \( \phi(k) \) and \( \phi_{rand}(k) \) allowed us to obtain the normalized rich club coefficient \( \phi_{norm}(k) \):

\[
\phi_{norm}(k) = \frac{\phi(k)}{\phi_{rand}(k)}
\]

In line with previous work (Colizza et al., 2006; Grayson et al., 2014; van den Heuvel and Sporns, 2011), a network was considered to have rich club organization when \( \phi_{norm} \) was greater than 1 for a continuous range of increasing k (rich club regime). Rich club nodes were brain regions taking part in these densely connected networks (or rich clubs), forming a functional unit. We considered as rich club hubs the nodes taking part in the club at value k where the strongest rich club effect was observed.
2.4. Statistics

2.4.1. Hemispheric asymmetry
The FC hemispheric asymmetry of the ROIs was estimated with the DCs. We derived a connectivity-based lateralization index (LI), by contrasting the k values of homotopic nodes (comparison of FC between mirror areas), according to the following formula:

\[ LI(k) = \frac{LH(k) - RH(k)}{LH(k) + RH(k)} \]

With LH(k) being the DC for the ROI in left hemisphere, RH(k) the DC for the homolateral ROI in the right hemisphere.

We also calculated global lateralization indexes at the connectome or Net level by averaging the corresponding nodal LIs. LI values can range continuously from -1 to 1 and the following landmarks were considered for interpretation: -1 = complete RH dominance; +1 = complete LH dominance and between -0.2 and +0.2 = no clear dominance (Roger et al., 2019; Rolinski et al., 2020; Seghier, 2008).

2.4.2. Functional and structural matching
We used mappings provided by previously published tools to estimate the spatial concordance between LANG and (1) the neurotransmitter pathways; or (2) the terminations of large white matter (WM) bundles. The “functional” maps were issued from nuclear imaging-derived neurotransmitter maps implemented in the JuSpace toolkit (Dukart et al., 2021), specifically designed to link neuroimaging (MRI data) with underling neurotransmitter information (as revealed by PET and SPECT tracers). The “structural” maps came from the deep-learning algorithm TractSeg (Wassetal al., 2018), which offers the segmentation of the main long-range WM brain bundles. It also allows the generation of grey matter masks that are linked by the bundles (ending masks). These ending masks were used here to define the structural connection maps of each bundle or combination of bundles.

The functional and structural maps were registered to the surface template and binarized. Each parcel of the HCP-MMP1.0 template was coded according to the presence/absence of map coverage, with 1 corresponding to at least 40% coverage of the parcel surface (> 40%); and 0 to less than 40% coverage (< 40%).

We then used the simple matching coefficient (SMC; Boriah et al., 2008; Sokal and Michener, 1958) method to quantify the spatial concordance between LANG and each of the functional and structural binary maps. SMC indicates the coincidence ratio between the mutual presences (and absences) and the length of the binary sequences: 0% means that the labels have nothing in common and 100% that they have identical sequences. Only coefficients exceeding 2/3 of the total agreement (SMC > 0.67) were considered relevant.

2.4.3. Cross-processes flexibility
We computed a flexibility index by using multilayer network model and with a method close to that of Betzel et al. (2017). The layers of the model were constituted from the matrices of the S groups of tasks (i.e., subprocesses) identified with data-driven clustering analyses (see Section 2.3.2). To keep a common reference between layers, the matrices were restricted to the LANG 131 ROIs and re-estimated on this basis. We applied the generalized Louvain package (Jeub et al., 2011), suited to determine community structure in multiplex graphs (Bassett et al., 2011, 2013; Mucha et al., 2010). This method has the advantage of preserving the community labels consistently across layers (here the task groups), avoiding thus the issue of community matching (Yang et al., 2021).

From the communities assigned across layers, we calculated a flexibility score as previously proposed by (Bassett et al., 2011). The flexibility of a node corresponds to the number of times that a node changes its modular assignment between layers, normalized by the total number of possible changes (i.e., the total number of layers minus 1, here 4). In short, the f-score reflects the frequency a brain region changes its community assignment. It ranges from 0 to 1, where 0 corresponds to a region that never changes module whatever the subprocess/task involved (stable on all layers); and 1 corresponds to a region that never belongs to the same module on the 5 layers. We also calculated the mean flexibility (F) over all nodes in the network to examine the global flexibility of the system.

\[ F = \frac{1}{N} - \sum_{i=1}^{N} f_i \]

2.4.4. Inter-individual variability
We assessed inter-subject variability by considering the individual signal values, extracted for each individual and on each of the LANG ROIs. To remove the variability induced by the task, we normalized the beta values, considering the mean and standard deviation of the other subjects who performed the given task (z-score betas). When individuals performed multiple tasks (for subjects enrolled in the same protocol), we averaged the z score betas for these subjects to avoid accounting for additional intra-individual variability. We thus performed the measurement of inter-individual variability by considering the subjects and not the scans. In addition, we considered only the young/middle-aged cohort of participants in which LANG was modeled to (1) prevent a large task effect (older adults primarily performed the NAM object naming task); and (2) not overestimate the variability of LANG by adding older participants (seniors). The average and absolute z scores of each region were then divided into 3 bins of increasing interindividual variability.

2.4.5. Age effect
To highlight LANG ROIs that are the most resilient/vulnerable to the aging effect, we examined the subjects of the whole InLang cohort representative of a wide range of ages (young/middle-aged and older adults) who performed the same task (Object Naming: NAM). Age was considered continuously in our statistical analyses, from age 20 to 85 (n = 82). We computed standard correlation coefficients (Pearson r) between the age and DC (here estimated based on individual connectivity matrices). This allowed us to observe a positive (positive and high r) or negative (negative and high r) age effect on task-based FC.

2.4.6. Gender effect
We estimated the gender effect by generating the FC matrices reduced to the LANG ROIs for males (M) and females (F); self-reported gender) separately. The average DCs obtained for each ROI were then compared between the two groups. We estimated significant differences between the two populations by means of standard t tests. For illustrative purposes, only the top 10% of the largest differences (Males>Females and Females>Males, independently) were retained to define the most diverging ROIs.

3. Results

3.1. Towards a task-based connectomic atlas of language

3.1.1. Global profile of the tasks and latent subprocesses
Data-driven clustering based on the similarity of global measures across conditions/tasks (the global efficiency: \( E_{glob} \), the average geodesic distance: \( d \), and the total number of nodes: \( N \), see Fig. 2A) has revealed an optimal 5-cluster solution. This solution was consistent with that obtained based on BOLD functional activations (Fig. S1, Appendix). We used the internal composition of the five task groups to label them according to the underlying language subprocess that might be primarily involved (Fig. 2A), namely: G1 = MONITORING (MS, MO, DO tasks); G2 = DECODING (PHON, RHYM and PROS tasks); G3 = SEMANTIC (SEM and SP tasks); G4 = PRODUCTION (NAM, FLU, GENE, and REP tasks); G5 = WANDERING (VMW task). Indeed, the monologal and dialogal inner speech with own and other voices (MS, MO, DO) engage MONITORING processes (inner voice control). Phoneme detection (PHON), rhyme judgment (RHYM), and
prosodic detection (PROS) first involve phonology and/or prosody DE-CODING (sound control). Semantic categorization (SEM) and speech perception (SP) respectively engage word and sentence comprehension. They both primarily require SEMANTIC processing (conceptual knowl-edge). Object naming (NAM), categorical fluency (FLU), sentence generation (GENE), and word repetition (REP) rely on lexical/lexico-syntactic formulation or word PRODUCTION (conceptual knowledge). Finally, verbal mind wandering (VMW) involves spontaneous speech production underpinned by introspective WANDERING processes (or unintentional thought). Details of the tasks and the language subprocesses theoretically and primarily targeted according to statistical fMRI contrasts performed are provided in the Supplementary Material (see also the summary Table S1, Appendix).

Interestingly, language tasks and subprocesses can also be grouped according to the physical distance of their functional connectivity within the brain. By contrasting the nodal integration/segregation balance (I:S) with the nodal geodesic distance $d$ of each task and subprocess, we have observed a significant linear relationship between the two parameters ($r = 0.8, p < .001$). However, the best-fit function indicated that the relationship was slightly better described by a nonlinear relationship (i.e., a polynomial curve; Fig. 2B). Furthermore, the unsupervised classification (k-mean method) applied to these data has identified 3 main clusters, denoting a gradual organization of tasks and subprocesses according to 3 canonical profiles of average connectivity that can be interpreted as $C_1 =$ long-range connections; $C_2 =$ middle-range connections; $C_3 =$ short-range connections (Fig. 2B). Overall, the more task- or process-based connectomes were segregated rather than integrated on average (i.e., a negative difference in favor of $E_{loc}$), the shorter the physical intermodal distance (short-distance functional connections, as for the control tasks of language involving DE-CODING and MONITOR-ING subprocesses). Conversely, the more task- or process-based connectomes were integrated rather than segregated on average (i.e., a positive difference in favor of $E_{glob}$), the longer the physical intermodal distance (long-range functional connectivity, as for the WANDERING and SEMANTIC task groups).

3.1.2. Global topology of the general LANG connectome

After excluding irrelevant functional connections (see Material and Methods, Section 2.2.), LANG was composed of 131 non-isolated regions of interest (ROIs; Power’s parcellation: Power et al. 2011), distributed over the two hemispheres (nLH = 80; nRH = 46) and the cerebellum (nCER = 5). Connectivity between LANG ROIs appeared balanced between integration and segregation (I:S = 0.049), associated with a rather long-range connectivity profile ($d = 68.1$). Table S3 (Appendix) summarizes the global network properties of the tasks, the subprocesses, and of the general LANG connectome. Fig. 2C shows the LANG connectome as a graph projected into a reduced two-dimensional space based on principal component analysis (PCA).

3.1.3. LANG partition and hubs (intermediate and local scale)

Community-based detection applied to the LANG connectome has defined 4 distinct components (or functional subnetworks, called Nets; see Fig. 3A, for projection into the reduced PCA space). Fig. 3B shows the mapping of the Nets onto the brain and cerebellum templates. Fig. 3C highlights their internal composition in terms of discrete intrinsic networks (as previously characterized by Ji et al. 2019; Cole-Anticevic Brain-wide Network Partition: CAB-NP).

Considering the composition, Net1 could correspond to the core component of language, engaged in the coding-decoding of linguistic signals of multiple nature: e.g., acoustic, syntactic, conceptual, articularulatory (Coding-Decoding system). Net2 is represented by executive-attentional functional networks (Control-Executive system). Net3 is mainly composed of regions of the default mode network (DMN) known to be involved in high-level cognitive abstraction and can thus be regarded as a “conceptual” knowledge network (Abstract-Knowledge system). Finally, Net4 involves a large majority of perceptual and motor brain areas, suggesting that it is the “Sensori-Motor” system of language. A supported argument and in-depth discussion of the putative functional roles of these LANG Nets is raised in the discussion (Section 4).

The Nets’ composition, coupled with their topological organization in the PCA reduced space, have provided evidence for the possible mean-

---

**Fig. 2.** Global connectomic profiles of the tasks, the subprocesses and the LANG connectome.

(A) Hierarchical clustering of parameters used to define global FC profiles ($E_{glob}$: global efficiency, $E_{loc}$: average local efficiency, $d$: mean geodesic distance; $N$: total number of nodes; top); and of individual tasks, to cluster them into groups of underlying subprocesses (bottom). See also the Supplementary Material for the rationale of the subprocesses labels (Table S1) and for the clustering applied to BOLD functional activations (Fig. S1). Table S3 summarizes the global measures for each task and subprocess. (B) Non-linear relationship between the integration/segregation balance (I:S) and the geodesic distance of the functional connections of the different connectomes. Plain dots correspond to the mean values of each task-related connectomes. Empty dots to the mean values of subprocess-related connectomes. The k-mean clustering revealed 3 distinct types of connectivity according to the I:S/geodesic distance profile (short-range, middle-range, and long-range). The colored lines come from the centroids (crosses) estimated from the observed data (at the level of brain region), in relation to the regression polynomial curve. (C) Global topology of the LANG task-based connectome. The 131 regions of interest (ROIs) of LANG (Power et al., 2011; LH = 61%, RH = 35%, CER = 4%) are projected here in a reduced two-dimensional space (PCA layout) consisting of the first two components: PC1 and PC2. The LANG atlas, node coordinates and properties are described in Table S4. The package including the atlas can be downloaded here: https://osf.io/6xm8n/.
3.2. Properties of the LANG connectomic atlas

3.2.1. Functional correlates

On average, the LANG’s FC laterality index indicated a slight LH predominance (LI(LANG) = +0.23), but hemispheric asymmetry was variable across the Nets. The proportion of nodes that were more strongly connected was higher in LH for Net1 (LI(Net1) = +0.41) than for the other Nets (Fig. 4A). By comparison, the FC of the nodes belonging to the “sensory-motor component” were bilaterally distributed (LI(Net4) = -0.12). Overall, the FC asymmetry of LANG Nets (from bilateral to LH) was arranged along the following gradient: Net4 < Net3 < Net2 < Net1.

In addition, some LANG Nets were spatially congruent with the mapping of neurotransmitter receptor pathways. In particular, the LH nodes of Net2 and Net3 showed a high spatial matching with the serotonin receptors 5HT2a (SMC Net2/5HT2a = 0.68; SMC Net3/5HT2a = 0.74). Those of Net4 overlapped with the noradrenergic transporters NAT_MRB (SMC Net4/NAT_MRB = 0.7). However, based on the SMC relevance criterion (SMC > 0.67), Net 1 did not show significant spatial correspondence with any of the neurotransmitter maps. Fig. 4B shows the distribution of the LANG ROIs that matched (or did not match) with the PET receptors (see also Fig. S5, Appendix).

3.2.2. Structural correlates

The endings of some large white matter (WM) bundles were spatially concordant with the LANG Nets (Figs. 4C, S4, Appendix). The best overlap between the bundles and Net1 (LH ROIs) was obtained by combining the ending masks of the left arcuate fascicle (AF), superior longitudinal fascicle branch III (SLFIII), inferior longitudinal fascicle (ILF), and the thalamo-premotor (T_PREM) projections (SMC Net1/AF-SLFIII-ILF-T_PREM complex = 0.9). The concordance rate increased to 92% when the ending masks of the middle cerebellar peduncle (MCP) and the cerebellar ROIs of Net1 were included. At a more restricted level, the unique contribution of the left AF provided a high spatial concordance with the Net1 lateral LH nodes (SMC Net1/AF = 0.72). Regarding Net2 (LH ROIs), the best matching was reached with the combination of the left superior longitudinal fascicle branch II (SLF-II) and the cingulum (CG) bundle (SMC Net2/SLFII-CG = 0.76). The concordance between the SLFII individually taken and the lateral LH Net2 ROIs was close to 70% agreement (SMC Net2/SLFII = 0.68). Net3 (LH ROIs) had almost complete coverage when considering the combination of CG, the middle longitudinal fascicle (MLF), and the fornix (FX; SMC Net3/CG-MLF-FX = 0.98). Finally, Net4 (LH ROIs) was spatially well covered by the combination of the cortico-spinal (CST) and the striato-precentral (ST_PREC) tracts (SMC Net4/STM-ST_PREC = 0.78).
3.2.3. Flexibility and variability

The module assignments of LANG ROIs varied according to the linguistic subprocesses involved. We have calculated the flexibility coefficients to capture the FC versatility of the ROIs engaged in the different Nets, depending on the subprocess at work. The average flexibility coefficients (F) were low for Net1 (F Net1 = 0.21) and Net4 (F Net2 = 0.32); while those for Net2 and Net3 were twice as high (F Net2 = 0.55; F Net3 = 0.64). Ordering LANG networks according to their functional versatility yielded: Net1 < Net4 < Net2 < Net3; from invariant to highly flexible (Fig. 5A).

Although modest, there was also some inter-individual variability in FC when individuals performed the language tasks (Fig. 5B). We found the highest inter-subject variability on Net3, but the variance remained low on average (mean z score = 0.56). FC variability between participants was more visible at the regional level than at the network scale. In addition, we found a high matching coefficient between the “universal language network” (ULN; as proposed by Malik-Moraleda et al., 2022 and the lateral LH ROIs of Net1 (SMC Net1/ULN = 0.84; Fig. S5, Appendix), suggesting some between-individual and cross-cultural consistency in key language network involvement.

However, the LANG connectome underwent changes with age (Fig. 5D). We observed both positive and negative correlations between age and degree centralities (DCs). Net3 and Net2 were the components showing the most important modulations. More specifically, ROIs of Net2 were negatively correlated with age (mean r = -0.39); while ROIs of Net3 were, on average, positively correlated with age (mean r = 0.31). Thus, the older the individuals, the less functionally connected the Net2 regions were (i.e., a decrease in functional hubs for this network in LANG). By contrast, the Net3 regions tended to be more strongly interconnected in LANG with age.

Finally, gender also modulated LANG connectivity. The strongest LANG ROIs for males compared to females (M > F) in terms of DCs were distributed between Net1 (62.5%) and Net2 (37.5%) in LH. The strongest LANG ROIs for females compared to males (F > M) were practically all located in Net3 (91.67%). Fig. 5C shows the LANG ROIs with the most divergent FC. It is interesting to note that the top 10% were located in LH but see also Table S5 (Appendix) for a complete picture of statistical differences.

4. Discussion

The main objective of this study was to provide an in-depth, multiscale view of the organization of brain function associated with language from a connectomic perspective. We leveraged an extensive fMRI
Fig. 5. Variability of the functional attributes of the LANG connectome.

A. Variability induced by linguistic demand (i.e., the linguistic subprocess involved in the task). Representation of the flexibility score of each of the LANG ROIs as well as their distribution in each of the Nets (see also Fig. S7).

(B) Inter-individual variability. Representation of the betas z-scores estimated on the LANG ROIs and their distribution in each of the Nets (see also Fig. S8).

(C) Variability determined by age. Illustration of the correlation coefficients between age and ROI DCGs, as well as their distribution in each of the Nets. The correlations were performed on the naming task (NAM) that includes healthy participants over a wide age range: 82 subjects aged 18 to 84 years (see also Fig. S9).

(D) Variability induced by gender. Top 10% of the most different LANG ROIs between males (M) and females (F; self-reported gender). In cyan, the top 10% of ROIs where nodal connectivity was significantly higher in females than in males. See Table S5 for a detailed summary of significant differences.

database of multi-paradigm language tasks (InLang [Interactive networks of Language] database), and we applied a state-of-the-art functional connectivity (FC) methodology that provided unique insights on brain networks related to language. The central finding of this research was that the general language connectome (LANG) could be objectively partitioned into four main non-overlapping subnetworks (referred to as “Nets”), possessing distinctive and marked features. Table 1 below provides a complete overview of these Nets.

The most extensive subnetwork, Net1, appears to correspond to a “specialized” language system shaped for the encoding-decoding of auditory-verbal signals. Indeed, Net1 consisted primarily of areas belonging to the intrinsic networks previously designated as “auditory” and “language” (CAB-NP RSNs: Ji et al. 2019, Fig. 3C). It included a set of both primary, secondary, and associative areas, previously noted as specialized for language (e.g., Labache et al., 2019; Price, 2012). In addition, a subset of Net1 composed of crucial brain regions densely interconnected formed a critical set for information integration and communication during language tasks (Fig. 3C). These core network structures were inscribed in the left perisylvian zone, namely: the anterior part of the superior temporal gyrus (STGs), the posterior part of the pars triangularis of the inferior frontal gyrus (pIFG, 45/44), posterior middle frontal gyrus/premotor cortex (pMFG, 55b area), inferior parietal cortex, supramarginal gyrus (PF/PPm), temporo-parieto-occipital junction (STV/TPOJ). Our analyses to determine the functional role of brain regions within networks had identified them all as “connector” hubs, which is consistent with previous observations (e.g., IFG/TPJ/STG: Goucha et al., 2017; Hagoort, 2016; PF/SMG: Braga et al., 2013; pMFG, 55b: Hazem et al., 2021). Connector hubs were located at the contact points of several white matter (WM) fascicles, actively supporting long-distance information transport and processing (e.g., at the AF/SLF convergence areas for the IFG and TPJ; Roger et al., 2022).

Indeed, we observed a robust matching with the AF endpoints (whose involvement in language has been widely and more directly reported; e.g., Forkel et al., 2022, for a meta-analysis) and the lateral perisylvian part of Net1. Nevertheless, more than a one-to-one relationship between structure and function, it was a combination of various WM bundle terminations that underlay the entire network (Fig. 4C). Beyond the clique, Net1 embedded other connector areas. Some of which were in the right hemisphere, others in the basal ganglia (the anterior parts of the thalamus and the putamen) or in the right cerebellum (Crus I) as well (Table S4, Appendix). Large cortico-basal ganglia-cerebellar loops would be involved during language tasks, supporting a substantial role of the subcortical structures in high-level cognition including language (Murphy et al., 2022).

Individually, brain areas have their own anatomical and microstructural properties (cytoarchitectonic features, Zilles and Amunts, 2010)
Table 1

<table>
<thead>
<tr>
<th>LANG Nets</th>
<th>Composition &amp; key features</th>
<th>Functional attributes</th>
<th>Anatomical underpinnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net1 Coding-Decoding System</td>
<td><em>Language/auditory-verbal component (RSNs: LANG/AUD)</em></td>
<td><em>Systematically engaged in language tasks (inflexible)</em></td>
<td><em>Supported by a complex of long-range fascicles (AF-SLFB-ILF-ST_PREM; lateral part: AF +++)</em></td>
</tr>
<tr>
<td></td>
<td>*Strongly interconnected (perisylvian rich-club) *</td>
<td><em>Compatible with the “universal language network”</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>At the crossroads of the other Nets (Core)</em></td>
<td><em>LH lateralized</em></td>
<td></td>
</tr>
<tr>
<td>Net2 Control-Executive System</td>
<td><em>Associative regions involved in high-level cognition (RSNs Net2: CON/FPN; Net3: DMN)</em></td>
<td><em>Flexible according to the linguistic subprocesses involved (versatile)</em></td>
<td><em>Long-range associative WM fascicles (Net2: SLFB-GG; Net3: MFL-CG-FX)</em></td>
</tr>
<tr>
<td>Net3 Abstract-Knowledge System</td>
<td><em>“Grounded” cognition (RSN: SMN)</em></td>
<td><em>Age-sensitive (Net2: DC + with age; Net3: DC + with age)</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Serotonergic pathway</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Globally engaged in language tasks</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>FC bilaterally distributed</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Resistant to aging</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Noradrenergic pathway</em></td>
<td></td>
</tr>
</tbody>
</table>

and may thus be biased—under normal conditions—to respond efficiently and preferentially to certain input types. They can be tuned for functional selectivity to linguistic phonological, syntactic, lexical, or even semantic units (Friederici, 2011). However, the underlying computational processing (i.e., the functional role) of regions belonging to the same Net could be deeply similar. Computational building blocks (called primitives: Poeppel, 2012; elementary linguistic operations: Hagoort, 2019; or neural operations: Buzsáki, 2020) of Net1 could imply here the segmentation-fusion of the linguistic signal, yielding the generation of a verbal information stream of increasing and ordered complexity (Zaccarelli and Friederici, 2015). Multiple combinatorial operations on different linguistic representations have already been reported (e.g., the combinatorial network of language: Pykkänen, 2019). Net1 and its constituents could represent the foundation of these combinatorics in the task. The modularity analysis we applied to multiple language tasks would indeed have captured a common “language combinatorial” computational mechanism for Net1, making this network a cornerstone of a “language-specialized” encoding-decoding system (Hagoort, 2017).

Consistent with a central system, Net1 was topologically situated at the interface of the other components, between an internally and externally oriented cognition (Fig. 3A). Moreover, Net1 was found to be a globally inflexible (unchanged) configuration regardless of task and linguistic demand (Fig. 4A). It also appeared spatially consistent with the “universal language network” proposed by Malik-Moraleda et al., 2022 as an invariant, cross-cultural, functional language network (see Fig. S5, Appendix). Several universals of language (apart from the “universal grammar”; Chomsky, 1995, which is debated) have been reported (Coupé et al., 2019) and concern both semantics (Gibson et al., 2017), syntax (Cutrell et al., 2015) or even pragmatics (Plantadosi et al., 2011). The constraints applied to shape languages seem to follow common rules of optimization of coding and information transfer towards a fundamental principle of efficiency. The functional selectivity of Net1 regions is likely to be inherited from our ancestors and to be part of a language-ready brain (Bocckx and Benitez-Buraco, 2014). They are also supported by a specific brain architecture already present in children (Friederici, 2017) whose functional connectivity is genetically encoded (Mokki et al., 2022, for the genetic regulation specifically involved in the perceptual-motor and semantic pathways of language).

At the boundaries of Net1, we also detected two networks that are integral parts of the general LANG connectome (Nets2-3). Net2 was dominated by intrinsic attentional and executive control networks (cingulo-opercular and frontoparietal networks; Fig. 3C). First, the cingulo-opercular network (CON) is a superordinate system encompassing the salience network (Ji et al., 2019), involved in external-signal-driven attentional control or top-down, “exogenous” redirection of attention (Matthen, 2005). The specialization of such a network in the actively controlled integration of exteroceptive information may lead to the provision of appropriate information in working memory (Parr and Friston, 2017) in order to construct an internal representation of the external world relevant to the individual at a specific time. Second, the frontoparietal network (FPN; close to the Multiple Demand Network (MDN): Smith et al., 2021), or to the Central Executive Network (CEN); Doucet et al. (2019)) is a network involved in all processing requiring controlled attention directed toward internal cues and goals. This network operates for endogenous and top-down attentional redirection (Perrone-Bertolotti et al., 2020) and is engaged in verbal working memory and “fluid” cognition (Assem et al., 2020). Overall, Net2 is a controlled executive language system that captures both endogenous and exogenous attentional aspects. Net3, on the other hand, was almost exclusively composed of DMN (Default Mode Network) regions (Fig. 3C). At rest, the default state is thought to be involved in “random episodic silent thought,” promoting creativity (Andreasen, 2011). Previous studies exploring intrinsic connectivity have shown that some components of the DMN are tightly and specifically coupled with the language network (particularly the anterolateral subnetworks; Gordon et al., 2020). Task-based studies have also shown its involvement in natural language processing (Simoney et al., 2016). As a foundation of the episodic-semantic memory spectrum (and more broadly, language-memory; Roger et al., 2022), the DMN is a multimodal experiential system (Xu et al., 2017) that fosters resonance and binding between environmental features and those derived from similar prior knowledge and states (Binder and Desai, 2011; Constantinetcu et al., 2016). For these reasons, Net3 has been referred to here as the “Abstract-Knowledge system” of language.

Even if their functional role in cognition was distinct, Net2 and Net3 were both involved in high-level cognition. They displayed similar network features in terms of hub properties, with a remarkably high proportion of “satellite” key regions compared to other Nets (Table S4, Appendix). Satellite centers are regions whose functional communication supports dialogue between components (van den Heuvel and Sporns, 2013b). In our case, they favor communication with regions belonging to other Nets, facilitating multimodal integration or information-linking throughout the task. Moreover, we have observed in Net2 and Net3 a clear tendency to reconfigure, according to linguistic demand (i.e., versatile networks with a flexible modular configuration that depends on the language subprocesses involved; Fig. 4A). It is consistent with studies showing that these systems are auxiliary and differentially involved depending on the nature of the task.
(Fedorenko and Thompson-Schill, 2014). For instance, FPN/MDN is functionally active in controlled and challenging semantic tasks but not in less demanding linguistic tasks (Diachek et al., 2020), which is consistent with the role of FPN in attentional processes and fluid cognition (Assem et al., 2020; Perrone-Bertolotti et al., 2020). In the same line, the Net3 configuration was more likely to be engaged in tasks involving the projection of spontaneous and self-oriented thoughts, such as verbal mind wandering (Andreas-Hanna et al., 2014; Binder et al., 2011; Humphreys and Lambon Ralph, 2015; Konishi et al., 2015; Lau et al., 2013; Raichle, 2015; Wang et al., 2020). Interestingly, the brain spatial distribution of Net2 and Net3 corresponded to the mapping of the SHTZ2 receptors (Fig. 4B) involved in serotonergic transmission (Savil et al., 2012; see also Beliveau et al., 2017, for a high resolution and in vivo brain atlas of the serotonergic system), capable of amplifying or sustaining cortical excitation (Puig and Gullede, 2011). These receptors indeed modulate whole-brain connectivity and promote flexibility between brain states and processes (Jancke et al., 2021). They could therefore be a relevant biomarker of the functional flexibility revealed in Nets 2-3.

However, the two networks were distinct, underpinned by specific anatomical connectivity (Fig. 4C). In addition, they were differentially sensitive to gender (Fig. 4D). For example, some hubs in Net3 showed higher FC during tasks in females compared to males, consistent with recent findings on differential DMN connectivity in resting state studies (Liang et al., 2021). Finally, Net2 and 3 were both subject to the pressures of age but here again differently (Fig. 4C). Net2 was negatively impacted by aging. The older the age, the less Net2 regions were functionally connected. This reduction with age in attentional-controlled FC was consistent with the alteration in executive functioning traditionally observed in older subjects (Reuter-Lorenz et al., 2016). Net3, on the contrary, had a high share of hubs that were more densely connected with age, which may reflect a compensatory pathway traditionally observed in aging concerning language (i.e., a semantic strategy: Baci et al., 2021).

The last system, Net4, held bilateral sensorimotor cortico-subcortical brain areas. This fourth component of the language connectome was distinct from the perceptual and motor auditory-verbatim structures included in Net1 (Fig. 2B) but could be an essential part of the action-perception circuits of language. The brain regions involved in Net4 have already been described as engaged in several sensorimotor aspects related to language production, in particular: general action selection (premotor); motor execution (supplementary motor area); orofacial motor activity (precentral and postcentral language areas); or even timing of motor outputs (putamen and cerebellum; Price, 2012 for an exhaustive overview). Besides the primary and secondary sensorimotor regions, Net4 also encompassed a large part of the precuneus. Precuneus supports a prominent level of interconnectivity with other brain regions, which has led to the identification of functional subdivisions (posterior-visual; central-cognitive/associative; anterior-sensorimotor; Margulies et al., 2009) and has indeed been considered a crucial sensorimotor connector hub of the LANG connectome (Table S4, Appendix). Importantly, it is an important site of production-comprehension coupling in natural speech (Silbert et al., 2014). In addition to speech production, Net4 can be engaged in language comprehension. Semantic grounding (i.e., the semantic links between words and their actions, referent objects, and related concepts) appears to depend on semantic circuits that bring together both the circuits related to word form (perisylvian, Net1) and conceptual circuits that underlie, among other aspects, sensory-motor experience (extrasyllian, including Net4). The involvement of the motor system in speech perception and understanding has been observed in various contexts (Fernando et al., 2022; Schomers and Pulvermüller, 2016; Skipper et al., 2017; and see Pulvermüller, 2018 for the hypothesis of neural reuse of action perception circuits in language), which may explain why Net4 was globally engaged regardless of the subprocess involved in the language tasks (Fig. 4A). Finally, several neurotransmitters were involved in regulating the activity of sensorimotor regions.

However, we observed a specific spatial matching between the sensorimotor language system and the noradrenergic receptor mapping (noradrenaline transporter: Hesse et al., 2017; Fig. 4B). Catecholamine noradrenaline has substantial projections to somatosensory and motor areas, including primary cortices, and the modulatory effects of noradrenaline on sensorimotor processing are diverse. While its contribution to modulating arousal states (Holland et al., 2021) and adapting sensory circuits for optimal behavior in animals is well documented (see Jacob and Nienborg, 2018, for a review), its precise function in humans and in language remains to be investigated.

Overall, our observations reinforce and complement past observations about the neurocognitive architecture of language. For example, the concept of multiple language networks (Hagoort, 2019) or the “theoretical” subdivision of the vast language network into a critical system accompanied by several additional systems (or margins; Hertrich et al., 2020) have been discussed. In an opinion article, Fedorenko and Thompson-Schill (2014) have previously reconsidered functional specialization from the perspective of dynamic engagement of interactive networks to support different goals and linguistic complexity. Our observations about the flexibility of the different LANG components are entirely in line with these considerations. For example, some networks may be central (i.e., they are composed of regions that maintain their allegiance through time such as Net1). In contrast, others are composed of peripheral regions that are not necessarily specialized for language but participate in its elaboration (such as Nets2-3). In the same paradigm, Chai et al., (2016) studied the dynamic flexibility of language regions on a comprehension task. Beyond a core set of regions in the left hemisphere – analogous to our perisylvian hubs belonging to the rich-club (Fig. 3C) – they observed flexible reconfigurations of “peripheral” regions, primarily located in the right hemisphere. Within this extended language network, the angular gyrus and the anterior temporal lobe belonging to Net3 of the LANG atlas and assigned to the DMN (e.g., Smallwood et al., 2021), exhibited a singular and highly flexible pattern of activity that may drive the functional inter-hemispheric coordination associated with language (Blank et al., 2016; Chai et al., 2016; Mahowald and Fedorenko, 2016). Although other regions demonstrated weakly lateralized FC in the left hemisphere within the LANG atlas, these regions indeed showed more balanced connectivity between hemispheres (Fig. 4A). They are known to play an active role in the multimodal integration of information in semantic cognition (e.g., Lambon Ralph et al., 2017; Seghier, 2013), which is widely distributed in the brain (Huth et al., 2016).

Other data also support the proposal of a functional architecture of language composed of multiple systems/subsystems that corroborate our observations on LANG of discrete language-related components via the demonstration of specific signatures in terms of intrinsic connectivity (Labache et al., 2019) or of dissociated cortical distribution by intraoperative cortical stimulations allowing to go beyond the traditional correlational framework (Corrivetti et al., 2019). These segregated functional networks appear to dynamically integrate into larger interactive configurations (Roger et al., 2022). This echoes the concept of encapsulation or nested networks (Hilgetag and Goulas, 2020), where domain-specific and domain-general circuits interface to underlying a given mental process (Fedorenko and Thompson-Schill, 2014) (Fedorenko and Thompson-Schill, 2014). Even if the number of proposed networks varies according to methods used or primary theoretical frameworks, the task-based networks or systems defined in the LANG atlas were consistent with previous partitioning proposals.

The great benefit of partitioning emerging directly from data is to pinpoint latent mechanisms that transcend our classical cognitive descriptions (see interesting discussions on the current problem of brain-behavior concordance or the blurriness and ambiguity associated with terminology and definition of psychological constructs: Anderson, 2011; Buzsáki, 2020). Data-driven ontology provides an independent view, here from a neuro-centric perspective (Roger et al., 2022), and can serve as a “lingua franca across disciplines and theoretical gaps” (Eisenberg et al., 2023).
2019). However, since they are derived from the observations, these partitions depend directly on the quantity, quality, sensitivity, and validity of the data used. This study has the advantage of being based on a database including a rich diversity of fMRI protocols, varying in a wide range of language characteristics (Fig. 1). However, task paradigms are less easy to implement than the resting state. They are typically very controlled, require multiple repetitions to obtain a robust signal/noise ratio, are more prone to movement artifacts, and induce higher interindividual variability (Park et al., 2020). This often leads to smaller final samples, which may be a prominent issue for subsequent analyses. Taking into consideration the need to maximize observations to ensure robust results, we have focused most of our analysis on the investigation of the general connectome or on the subprocesses common to several tasks (and not on individual tasks). The compilation of even larger databases will allow for broader and more detailed investigations. For example, it would be important to consider the pragmatic aspects of language (Rasgado-Toledo et al., 2021), which are not specifically valued in the InLang database. Moreover, the current trend is to extend the framework of fMRI paradigms traditionally employed in “laboratory” settings to less controlled and more ecological protocols (Verga and Kotz, 2019). As evidenced by the recent Neuroimage Special Issue (Finn et al., 2022), several initiatives and datasets are steering toward accounts of cognition in more natural settings (e.g., Bhattachai et al., 2020; LeBel et al., 2021; Nastase et al., 2020 for language).

Similarly, (neuroimaging) multimodal initiatives have flourished in recent years (e.g., HCP: Van Essen et al. (2013); UK Biobank: Sudlow et al. (2015); ENIGMA: Thompson et al. (2020); CamCAN: Taylor et al. (2017) data collections). Multimodal datasets would then allow more direct confirmation of the neurofunctional relevance of the spatial correspondences we observed regarding anatomo-functional correlates of LANG. In our study, analysis of the links between LANG and anatomical connectivity and neurotransmitter maps could not be performed in the same individuals. Multimodal data would provide more subtle indicators of the correspondence between structure, function, and neurotransmitter pathways than the simple matching coefficient. It is essential when considering highly interactive complex networks where one-to-one relationships cannot be a viable principle (Suárez et al., 2020).

In addition, considering behavioral and cognitive performance is crucial for a better description of the functional repertoire. We must acknowledge that one of the main limitations of the proposed database concerns the cross-sectional aspect of the cohort and, consequently, the lack of (homogeneous) behavioral/cognitive data (i.e., inconsistent response modalities; some tasks were designed without directly measurable behavioral output, without overt speech production or with inner speech, depending on the targeted linguistic processes or to limit movement artifacts). While databases built from pre-existing data are essential to address current data reuse concerns (see the FAIR principles; Wilkinson et al., 2016), building highly multimodal databases – coupled with improved methods of acquisition, processing, and statistical investigation – is a practical approach to tackle the issue of brain-behavior correspondence more effectively. Such multimodal data acquired systematically across subjects would also limit the inter-subject variability inherent in cross-sectional databases, which may lead to greater stability and accuracy of results. However, the presence of functional (language) tasks and/or behavioral data in the multimodal neurocognitive datasets available to date is still generally limited.

We have attempted to comply with the best practices mentioned for analyzing and sharing FC data following the guidelines of the COBDIS report (Nichols et al., 2017), but some methodological choices may influence the estimation of FC and should be further emphasized. The LANG atlas was selected on independent resting-state and multimodal brain parcellations (Glasser et al., 2016; Power et al., 2011), limiting the circular analyses related to the ROIs selection (Kriegerkorte et al., 2009). Furthermore, the regions belonging to the atlas were identified without a priori, preventing here another bias that is a similar form of circular logic in which assumptions are retrofitted to results (also called HARKing: Button, 2019). However, it is essential to note that we estimated the FC through the correlation of interregional beta-series extracted from the tasks rather than using time-series methods like those applied to resting-state data (Cao et al., 2014). This approach of interregional FC has been implemented, validated (Götlich et al., 2015, 2017), and successfully applied in previous task-based fMRI studies (e.g., for recent studies: Antonucci et al., 2020; Franzmeier et al., 2018; Pang et al., 2022). Nevertheless, as it has also been argued previously that removing inter-event/block variance could help to reduce the chances of circularity in the analysis (Cole et al., 2019), replicating the LANG analyses by removing the average evoked responses from each task would thus be an option to consider to measure the potential bias induced by the FC estimation method.

Finally, the matrix thresholding method is a well-known limitation on resting-state FC. On binary networks, the inclusion of weaker correlations as functional edges leads to the inclusion of noisier connections, which can significantly bias the graph metrics (van den Heuvel et al., 2017). Including false-positive edges has a more deleterious effect than the exclusion of false-negative connections (Zalesky et al., 2016), suggesting that a restrictive threshold should be preferred over a more permissive one. The most widespread practice for choosing an “optimal” threshold is to extract the parameters of interest over a given threshold range and estimate their change (a range between 5 and 25% for most studies, according to a recent methodological review; Hallquist and Hillary, 2019). In our case, the network parameters were robust and stable over a threshold range between 5 and 10% (Fig. 52, Appendix). The chosen threshold of 5% of the highest positive correlations kept the LANG network balanced, neither too sparse nor completely interconnected and composed of relatively high correlation values limiting the inclusion of false-positive connections.

This present study offers a language atlas that relies on a thorough topological (i.e., spatial) analysis of FC. One of the following steps is to identify the causal organization (e.g., effective hierarchies/bertarchies) and precise timeframes in which its components/regions engage depending on subprocesses/mechanisms at work. MEG (such as in the MOUS dataset: Schoffelen et al. 2019) or electroencephalography (EEG; placed on the scalp, the cortex, or intracranial) are valuable tools for the evaluation of such dynamics. To this end, and using intracranial EEG recordings, a recent study examined the dynamic organization of naming (Forseth et al., 2021). They observed that regions were co-activated during extended periods, confirming that complex behaviors such as speech production requires the coordination of discrete network states (defined as a set of reference dynamics that coordinate the generation and transmission of information throughout the cortex; see also the concept of meta-networking of Herbet and Duffau (2020)). They were able to sequence, map, and identify the temporality of the different transient states; globally confirming the seminal model of word production proposed by Indefrey and Levelt (2004). The definition of a comprehensive repertoire of language states and causal relationships (i.e., effective and directed functional connectivity; e.g., Deco et al., 2021) in various tasks may extend our understanding of language functioning.

Although we explored several modulators, such as age and gender, a specific investigation of these factors remains to be done to gain a complete picture of the changes induced by these characteristics. As discussed above, we have observed a change in some hubs with age. However, the effect of age can be refined by attempting to capture critical evolutionary trajectories of the LANG networks. Establishing accurate age-related benchmarks requires more representative databases, both in terms of age and tasks performed, but would mutually benefit models of language and aging. The sample of our cohort of so-called young or middle-aged adults (< 60 years) includes a wide variety of ages (mean = 36.87; SD = 13.29), and it would be interesting to see if there are already significant age-related changes in language networks in this sample (i.e., before age 60). Studies of aging have indeed long shown that brain structure evolves with age and that certain regions – mainly
frontostriatal – experience an early decline (Buckner, 2004; Head et al., 2005; Raz et al., 2005; Raz and Rodrigue, 2006; Ziegler et al., 2011). It could result in subtle but quantifiable neurofunctional changes in the language networks. Exploring the LANG atlas from this angle would allow age categories to be defined not by traditional “chronological age” but by observable differences in functional brain connectivity, which constitutes an exciting perspective.

There are many opportunities for analysis and application of the LANG atlas. It may also involve considering other quantitative metrics or methods for evaluating the properties of LANG Nets/ROIs and their synergies. Given the current view of brain functional organization as nested hierarchies (e.g., Hilgetag and Goulas, 2020), for a review, assessing the “coreness centrality” property of brain networks is timely and of particular interest. Recently, Stanford and colleagues estimated the stability and resilience of brain networks by applying a sophisticated method of decomposing networks into hierarchically ordered subgraphs with progressively increasing connectivity (Stanford et al., 2022), also known as k-core decomposition (Batagelj and Zaversnik, 2003; Seidman, 1983). This method has shown sensitivity, validity, and promise for assessing high-level neurocognitive functioning (see also Arrese Lucini et al., 2019; Lahav et al., 2016; Li et al., 2020; Peng et al., 2019). Using diverse graph-based methods could lead to a comprehensive, multilevel evaluation of the LANG atlas. Exploring a wide range of connectomic properties of LANG, especially in populations where the importance of language is central – such as in multilingual individuals, in pediatric populations for language acquisition, or in pathologies with aphasic manifestations – represents an additional avenue for a deep appreciation of the functioning of language systems.

5. Conclusion

Language is a multi-faceted cognitive function. To account for the multidimensionality of language, we performed functional connectivity analyses on a multi-paradigm fMRI database (InLang [Interactive networks of Language] database), gathering thirteen different language tasks. It allowed us to inspect the language connectome in-depth, particularly its spatial properties and functional attributes. This study reaffirms that high-level cognition, such as language, emanates from synergistic exchanges of external and internal information across specialized systems. Our results highlighted the involvement of essential discrete networks (or components) that are settled around a core “language-related” system. Furthermore, the flexible engagement of some key regions depending on several modulating factors, such as linguistic demand, pointed to the dynamic nature of language. Importantly, and from ontology-based data integration, we have proposed a connectomic atlas of the “language mosaic” (LANG atlas), which can be considered as a reference for investigating additional conditions or pathologies altering language functioning.

Data availability statement

InLang [Interactive networks of Language] Database:

Neuroimaging data (raw and preprocessed T1w/ task-fMRI Nifti files) are available upon request. Data sharing requires approval of the request by the ethics committee of the Grenoble Alps University Hospital (data owner) and the signature of a data use agreement (DUA).

De-identified data related to the results reported in this Article:

Tables/documents, codes, statistical functional maps (group and individual maps), derivatives to perform FC/networks analyses, and the connectivity-based functional atlas of language (LANG atlas) are hosted on the Open Science Framework (OSF) platform and freely available for download [InLang_database; DOI: 10.17605/OSF.IO/6XM8N].

Appendix - Supporting information (SI)

Material and Methods SI:

Description of the InLang database fMRI protocol details
Participants information
Data acquisitions & (pre)processing
Results SI:
Supplementary tables and figures (global and nodal properties)
LANG package files
LANG maps (Nets, correlates, variability)

Data Availability

Data will be made available on request.

Credit authorship contribution statement

E. Roger: Conceptualization, Methodology, Formal analysis, Data curation, Visualization, Validation, Writing – original draft, Writing – review & editing. L. Rodrigues De Almeida: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft. H. Loevenbruck: Conceptualization, Writing – original draft, Writing – review & editing, Resources, Funding acquisition. M. Perrone-Bertolotti: Conceptualization, Writing – original draft, Writing – review & editing, Resources, Funding acquisition. E. Cousin: Conceptualization, Investigation, Resources, Data curation, Methodology. J.L. Schwartz: Conceptualization, Writing – original draft, Writing – review & editing. M. Dohen: Conceptualization, Writing – original draft, Writing – review & editing. P. Perrier: Conceptualization, Writing – original draft, Writing – review & editing. A. Vilain: Conceptualization, Writing – original draft, Writing – review & editing. P. Baraduc: Conceptualization, Writing – original draft, Writing – review & editing. S. Achard: Conceptualization, Methodology, Formal analysis, Visualization, Validation, Writing – original draft. M. Baciu: Conceptualization, Writing – original draft, Writing – review & editing, Resources, Funding acquisition, Supervision, Investigation.

Acknowledgments

We thank Philippe Kahane, Lorella Minotti, Olivier David, Arnaud Attyé, Laurent Lamalle, and Jean Francois Démont for their support and contribution to the NeuroCoG project. We also thank Cédric Pichat, Sonja Banjac, Laurent Torlay, and Eric Guinet for their help in data collection and their useful insights.

Funding

This work was supported by NeuroCoG IDEX UGA, “Investissements d’avenir” program [grant number ANR-15-IDEX-02]; and by the French program “AAP GÉNERIQUE 2017” run by the “Agence Nationale pour la Recherche” [grant number ANR-17-CE28-0015-01].

Ethics statement

The studies and protocols were approved by the local Ethics Committee “Comité de Protection des Personnes” with the following approval codes: NEREC (CPP: 09-CHUG-14, 04/06/2009), NeuroMod (CPP n° 2017-A00834-49), Recovery (CPP-ISIS 07PFR04-N/DCIC/06/25), Plast LANG (CPP: 09-CHUG-14, 04/06/2009), Reorg-GEREC (CPP 09-CHUG-14; MS -14-102), SEMVIE (CPP: 2014-A00569-38). The InnerSpeech project was approved by the local Ethics Committee “Comité de Protection des Personnes” Sud-Est V and by the National Competent Authority France-ANSM (Ref. CPP: 14-CHUG-39, Ref. Promoteur: 38RC14.304, ID-RCB: 2014-A01403-44, Ref. ANSM: 1412000B-31, ClinicalTrials.gov ID: NCT02830100) (CPP: 09-CHUG-14, 04/06/2009). All participants gave informed consent before volunteering.


