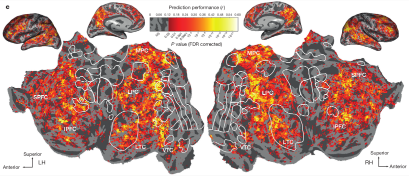
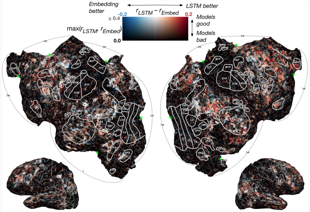
fMRI predictors based on language models of increasing complexity recover brain left lateralization

Laurent Bonnasse-Gahot[1] & Christophe Pallier[2]

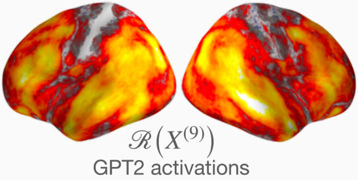
[1] Centre d'Analyse et de Mathématique Sociales, EHESS, CNRS, Paris
[2] Cognitive Neuroimaging Unit, CNRS, INSERM, CEA, Neurospin Center, Gif-sur-Yvette



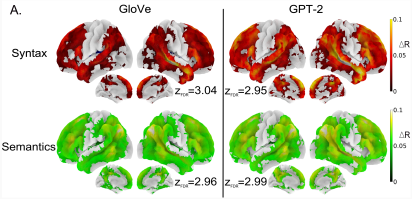
Huth et al. (2016)



Jain and Huth (2018)



Caucheteux et al. (2021)



Pasquiou et al. (2023)

Evidence for left hemisphere dominance of language

- Aphasia associated mostly to lesions in the left hemisphere
  (Dax, 1836; Broca, 1865)
- Electrocortical stimulation
  (Penfield and Roberts, 1959)
- Hemispheric anaesthesia with intracarotid amobarbital injections
  (Wada and Rasmussen, 1960)
- fMRI (language – control stimuli) generally yields more/stronger activations in the LH than in the RH
  (Binder et al., 1996; Just et al., 1996; Malik-Moraleda et al., 2022)
- . . .

90% of the healthy population has left hemispheric dominance, more or less correlating with handedness

Methods: Encoding models

Large language models

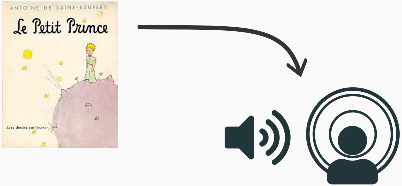
- 28 pretrained models, from 124M to 14.2B parameters
- 8 different families
- mainly Transformers based, but also State Space models architecture

Baselines

- random vectors
- random embeddings
- GloVe embeddings

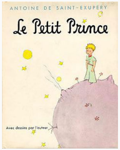
model name	$n_{parameters}$	n_{layers}	$n_{neurons}$
gpt2	124M	12	768
gpt2-medium	355M	24	1024
gpt2-large	774M	36	1280
gpt2-xl	1.56B	48	1600
opt-125m	125M	12	768
opt-350m	331M	24	1024
opt-1.3b	1.32B	24	2048
opt-2.7b	2.65B	32	2560
opt-6.7b	6.66B	32	4096
opt-13b	12.9B	40	5120
Llama-2-7b-hf	6.74B	32	4096
Llama-2-13b-hf	13.02B	40	5120
Qwen1.5-0.5B	464M	24	1024
Qwen1.5-1.8B	1.84B	24	2048
Qwen1.5-4B	3.95B	40	2560
Qwen1.5-7B	7.72B	32	4096
Qwen1.5-14B	14.17B	40	5120
gemma-2b	2.51B	18	2048
gemma-7b	8.54B	28	3072
stablelm-2-1_6b	1.64B	24	2048
stablelm-3b-4e1t	2.80B	32	2560
stablelm-2-12b	12.14B	40	5120
Mistral-7B-v0.1	7.24B	32	4096
mamba-130m-hf	129M	24	768
mamba-370m-hf	372M	48	1024
mamba-790m-hf	793M	48	1536
mamba-1.4b-hf	1.37B	48	2048
mamba-2.8b-hf	2.77B	64	2560

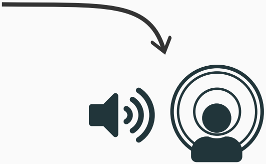
Methods: fMRI data & Encoding models



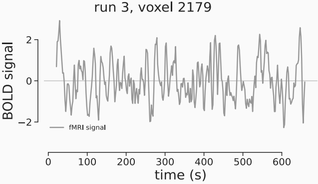
Li et al. (2022) *Le Petit Prince* multilingual naturalistic fMRI corpus. **Scientific Data, 9(1):530.**

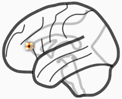
Methods: fMRI data & Encoding models



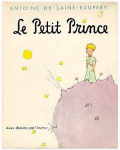


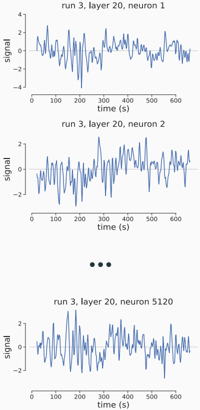
- English data (n = 49 participants)
- Functional data resampled to 4 mm isotropic voxels
- **Average subject**



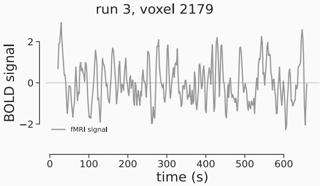


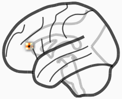
Methods: fMRI data & Encoding models



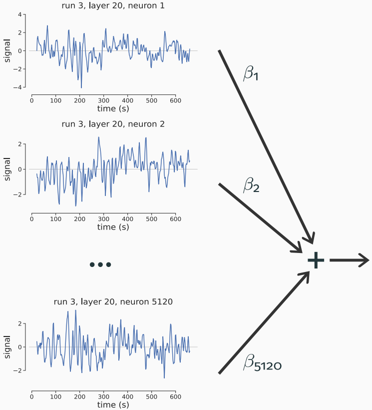


- English data (n = 49 participants)
- Functional data resampled to 4 mm isotropic voxels
- **Average subject**

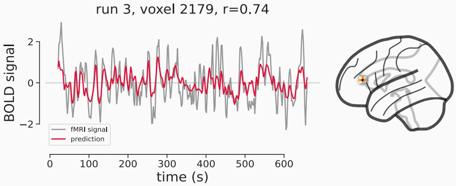




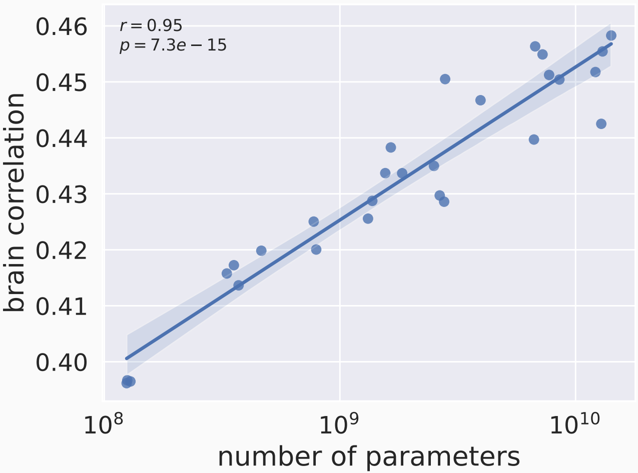
Methods: fMRI data & Encoding models

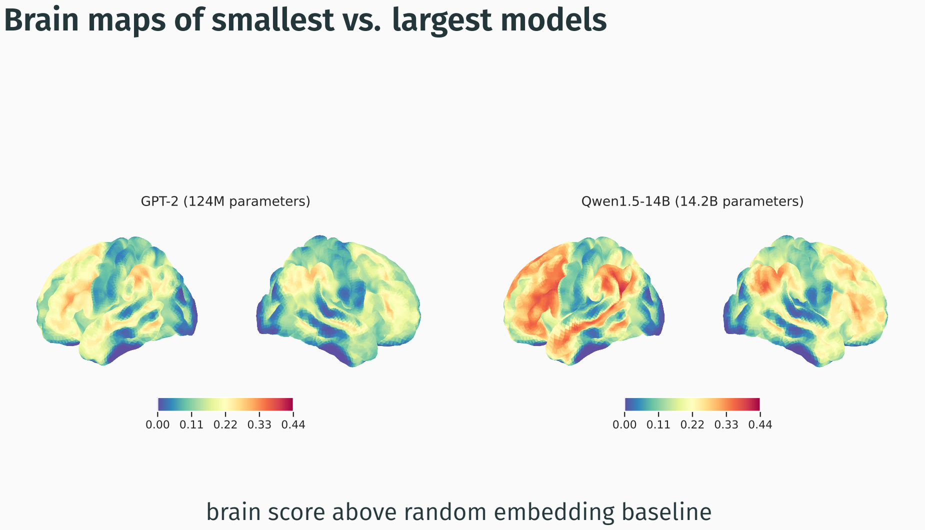


- English data (n = 49 participants)
- Functional data resampled to 4 mm isotropic voxels
- **Average subject**

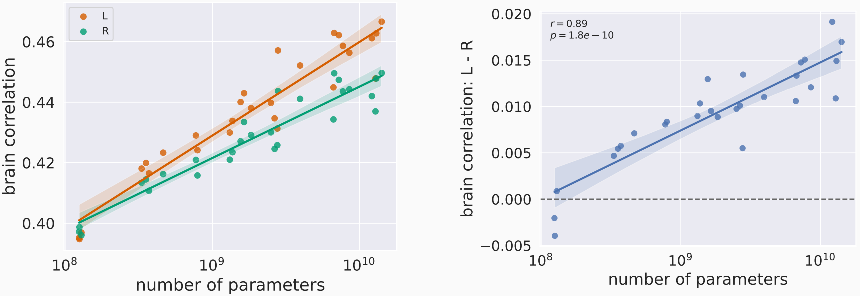


Brain correlation & Scaling laws




Brain maps of smallest vs. largest models
GPT-2 (124M parameters)
Qwen1.5-14B (14.2B parameters)
0.00
0.11
0.22
0.33
0.44
0.00
0.11
0.22
0.33
0.44
brain score above random embedding baseline

Emergence of left lateralization with the size of the models



Additional analyses, Discussion & Conclusion

We found this left-right growth in asymmetry:

- in other languages (namely Chinese and French)
- during the training process of a large language model
- at the level of the individual
  (future work: compare with independent assessments of individual language hemispheric dominance)

Why this emergence of asymmetry?

- Language models may better capture linguistic information (e.g., have more precise semantics)
- Future work: for which tasks are larger models better than smaller ones? How and where does this translate into better fMRI prediction?

Thank you for your attention!

arXiv preprint

code