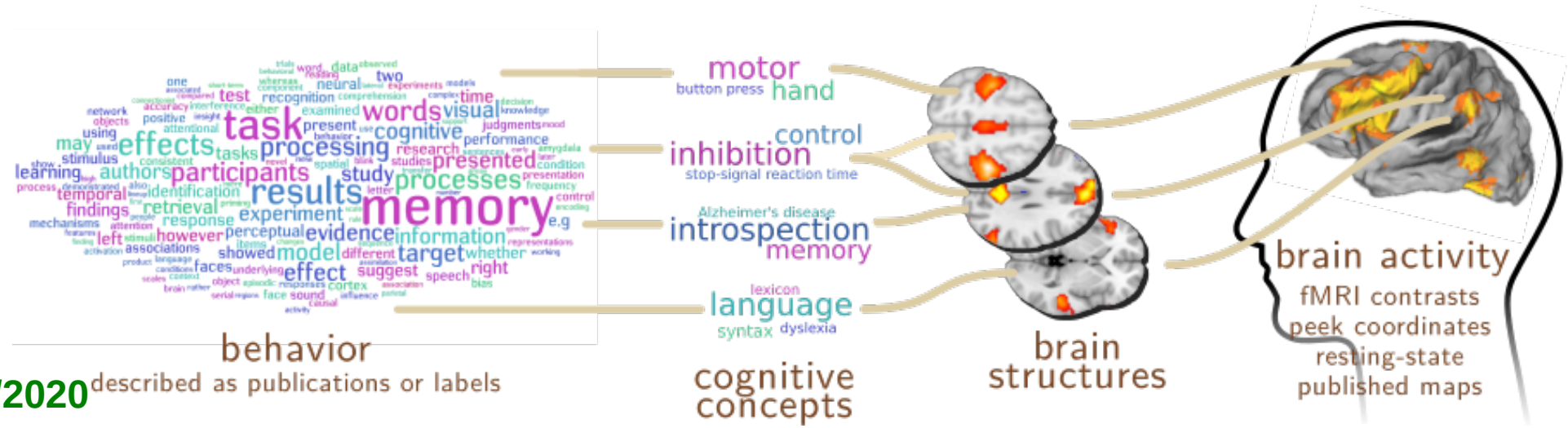
KARAIB

http://project.inria.fr/karaib

Bertrand Thirion, bertrand.thirion@inria.fr

KARAIB
Knowledge And RepresentAtion Integration on the Brain

Bertrand Thirion, bertrand.thirion@inria.fr

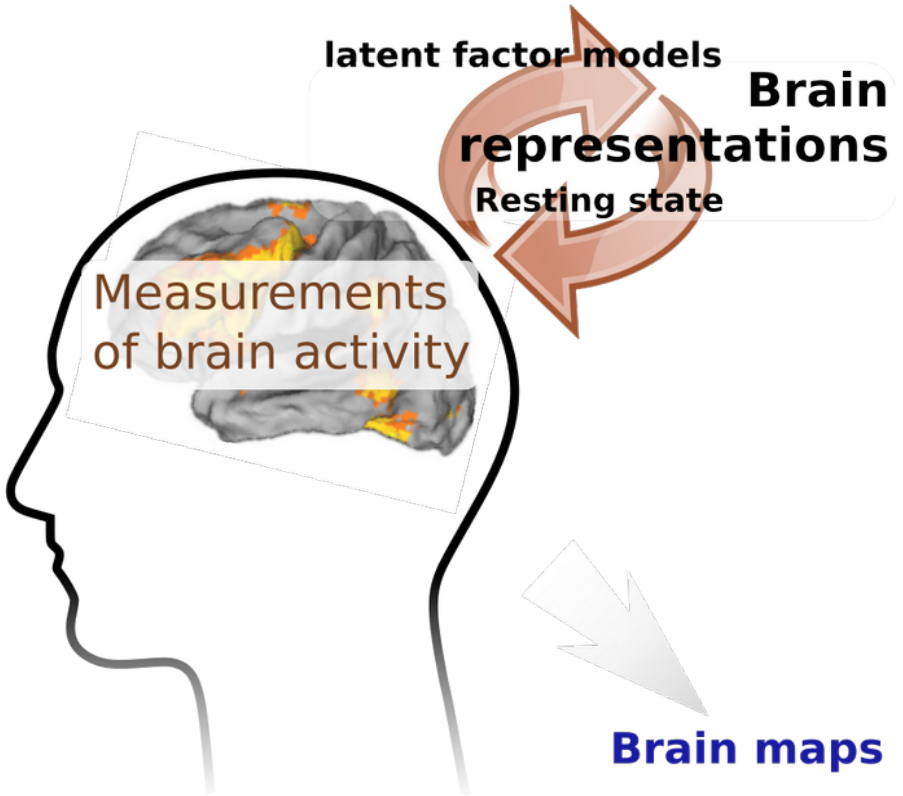


Motivation

- Cognitive concepts and brain pathologies are ill-defined [Yarkoni Poldrack 2016]
  - Psychological constructs
  - Pathologies or symptoms ?
- Leverage existing data w. representation learning
  - Mostly Image repositories & publications
  - Identify latent factors

Combine distributional semantics with brain images

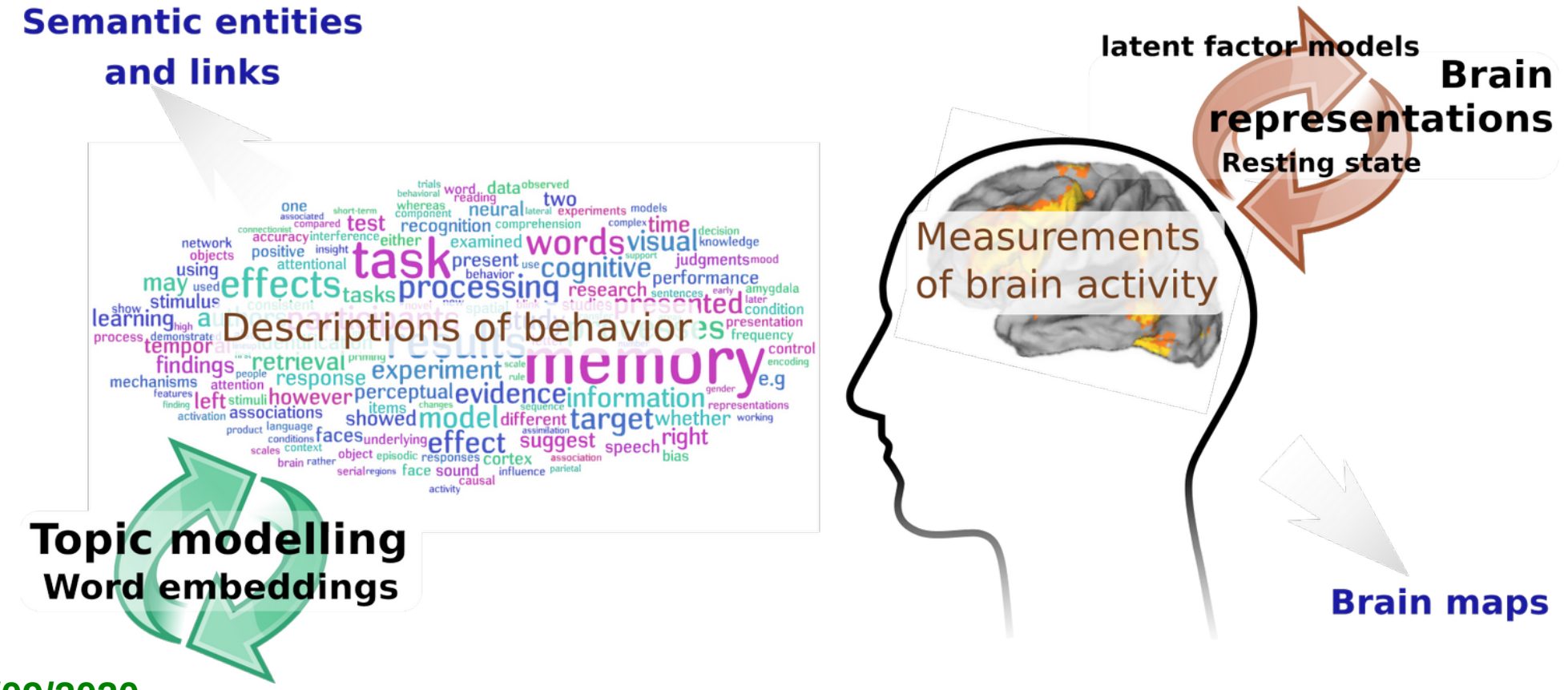
Semantic entities and links



Existing resources

Data source	Data type	Volume	Notes
OpenNeuro	Raw fMRI data, mostly task-related activation	247 studies (as of June 2019), > 6000 subjects	Wide variety of cognitive tasks
NeuroVault	Statistical maps, mostly group-level activation maps	80k maps from 2032 different collections (June 2019)	idem, noisy annotations.
NeuroSynth & Neuroquery	Coordinates from the literature	500 000 from 14 000 publications (June 2019)	Neuroquery developed by PARIETAL
Human Connectome Project	Raw fMRI and diffusion data, mostly rest fMRI	1 200 subjects, with 11 GB per subject	Homogeneous, high-quality data
Individual Brain Charting (IBC)	Task fMRI	12 subjects, 50 acquisitions covering many cognitive domains	Wide cognitive coverage, acquisition and processing by PI team.
UK Biobank Image Data	Rest and task fMRI	100 000 subjects (2021)	A lot of individual behavioral information, aging effects, diseases.

Combine distributional semantics with brain images



Combine distributional semantics with brain images

Semantic entities and links

Encoding
Forward Inference

latent factor models

Brain representations
Resting state

Measurements of brain activity

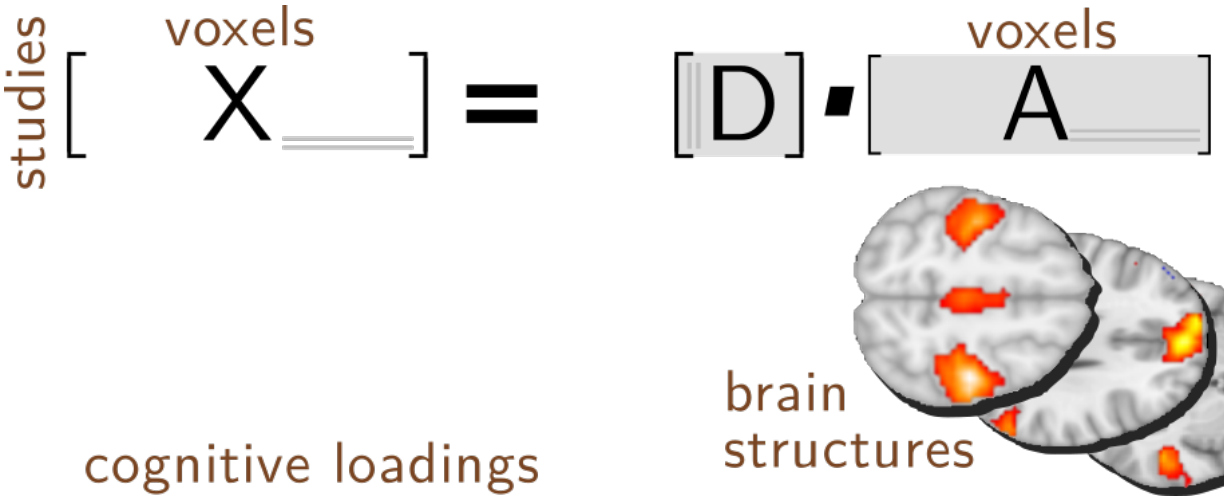
Topic modelling
Word embeddings

Decoding
Reverse Inference

Brain maps

Aim 1: Build coordinated representations bridging psychology and brain maps

Unsupervised learning from large image sets

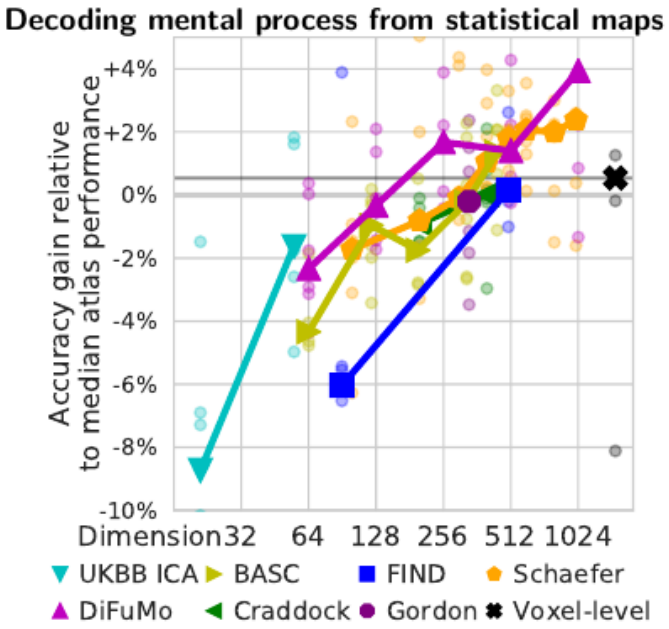


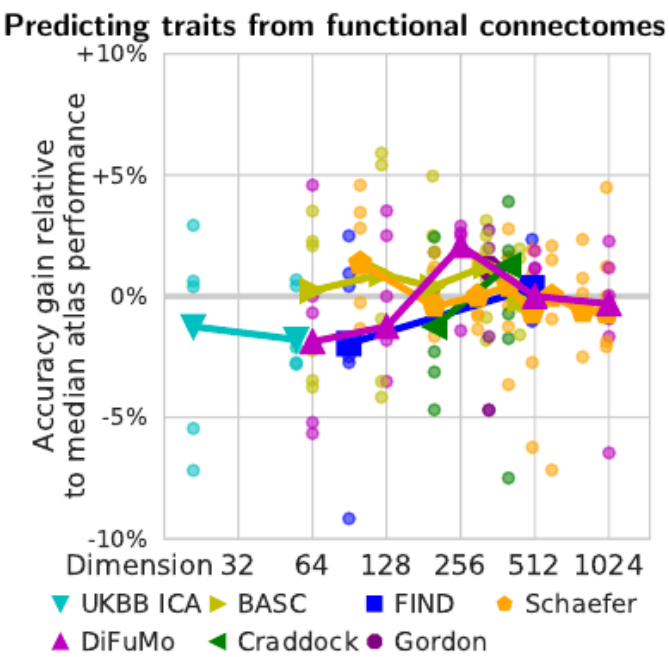
Beyond plain dictionary learning

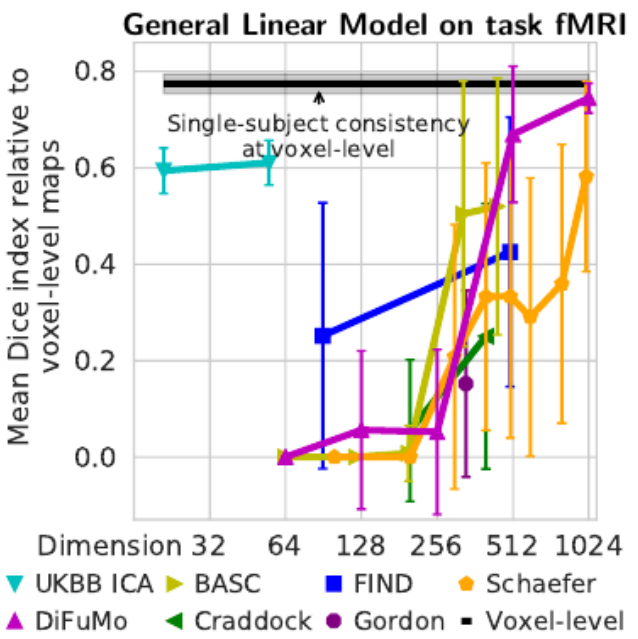
- Nonlinear versions
- Connectivity model
- multi-resolution

Preliminary results: DiFuMo

The most powerful brain atlas for functional data analysis !

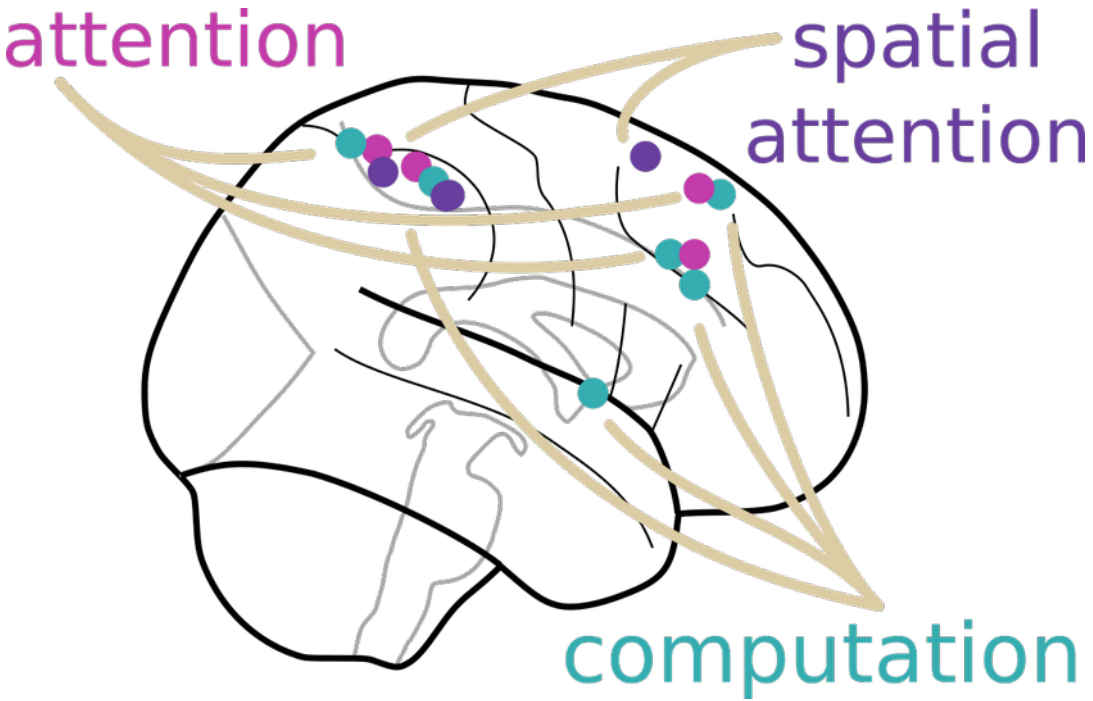






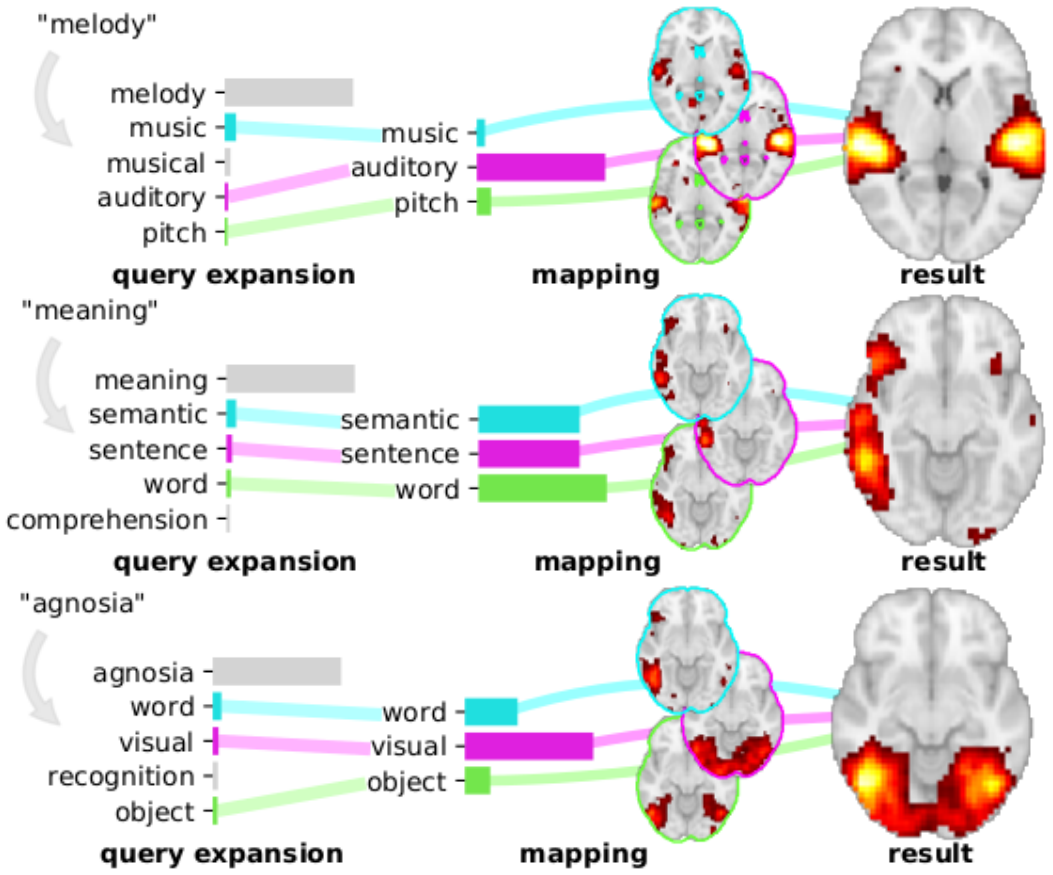
[Dadi et al. Nimg 2020]

Leveraging Modern language models of psychological literature



More powerful (pre-trained) language models in Neuroquery

Preliminary work: Neuroquery



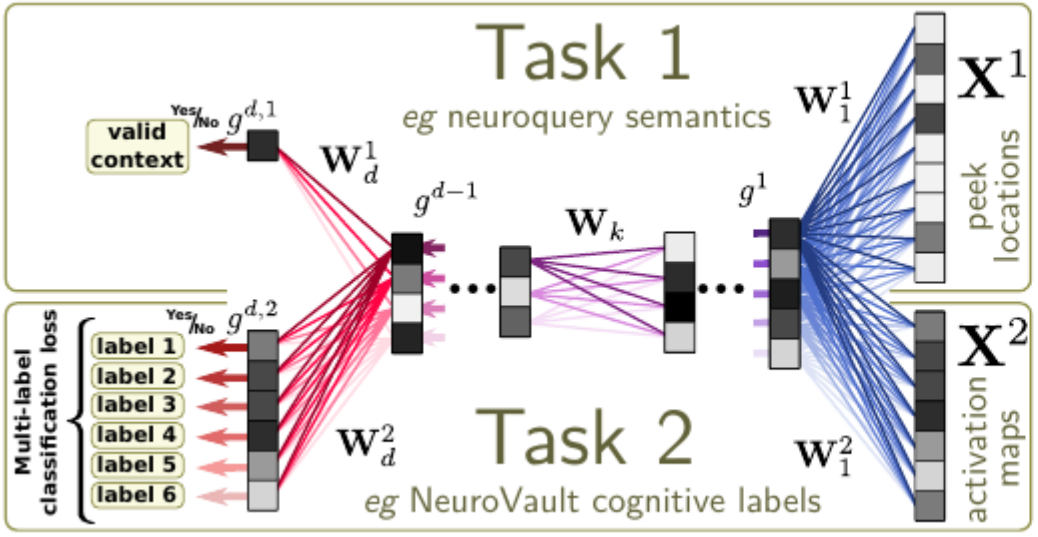
Semantic structure
→ map concepts with few/no data

[Dockes et al. Elife 2020]

Deep learning for coordinated representations

Multi-task architecture:

- models for Tasks 1 and 2 share parameters.
- Joint Minimization → common parameters benefit from both tasks and data sources.



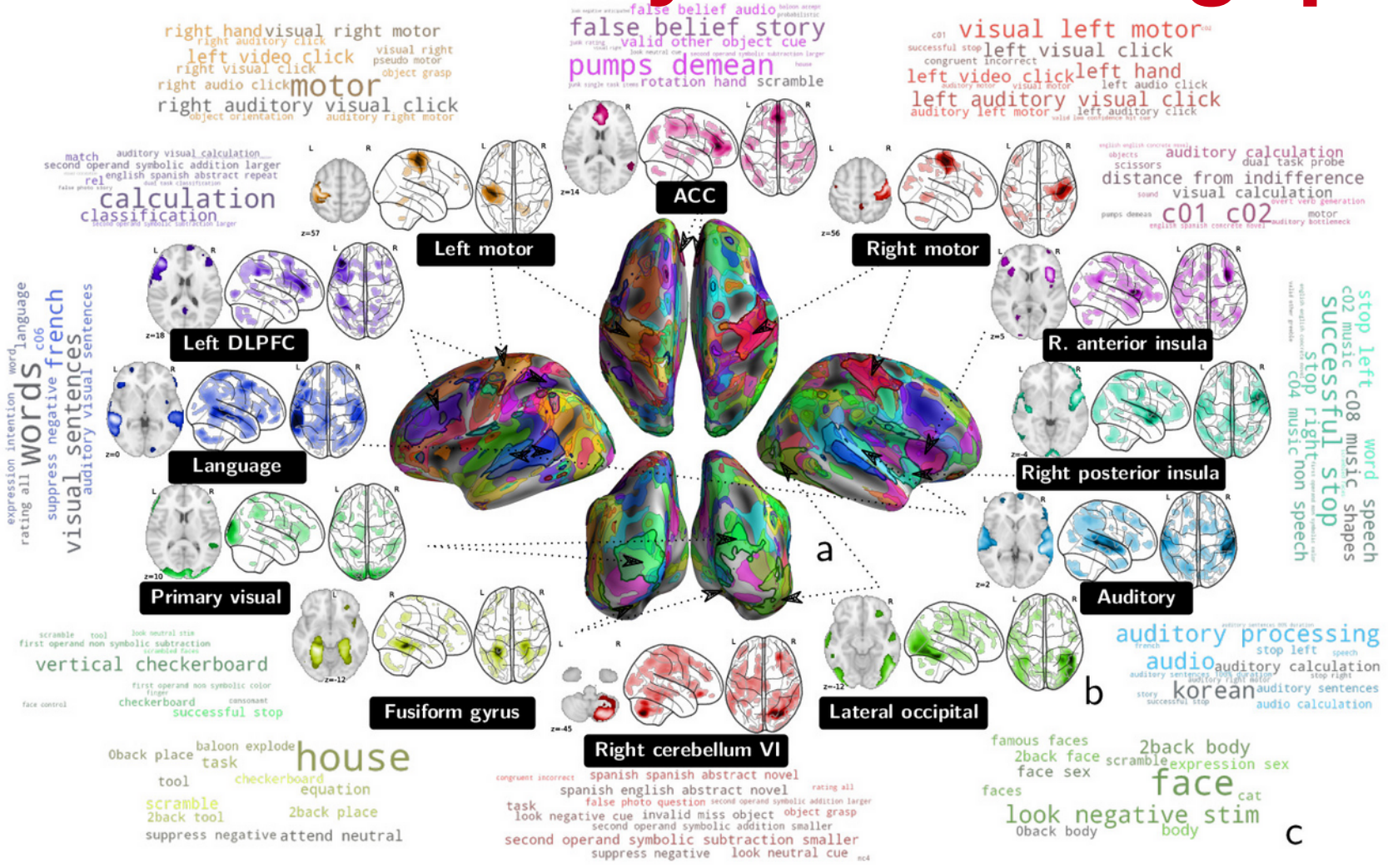
Model for task 1:

$$\mathbf{y}^1 = g^{d,1}\left(\left(\mathbf{W}_d^2 \ \ldots \ g^k\left(\mathbf{W}_k \ \ldots \ g^1\left((\mathbf{W}_1^1\mathbf{x}^1)^\intercal\right)\right)\right)^\intercal\right)$$

Model for task 2:

$$\mathbf{y}^2 = g^{d,2}\left(\left(\mathbf{W}_d^2 \ \ldots \ g^k\left(\mathbf{W}_k \ \ldots \ g^1\left((\mathbf{W}_1^2\mathbf{x}^2)^\intercal\right)\right)\right)^\intercal\right)$$

Preliminary result: Cogspaces



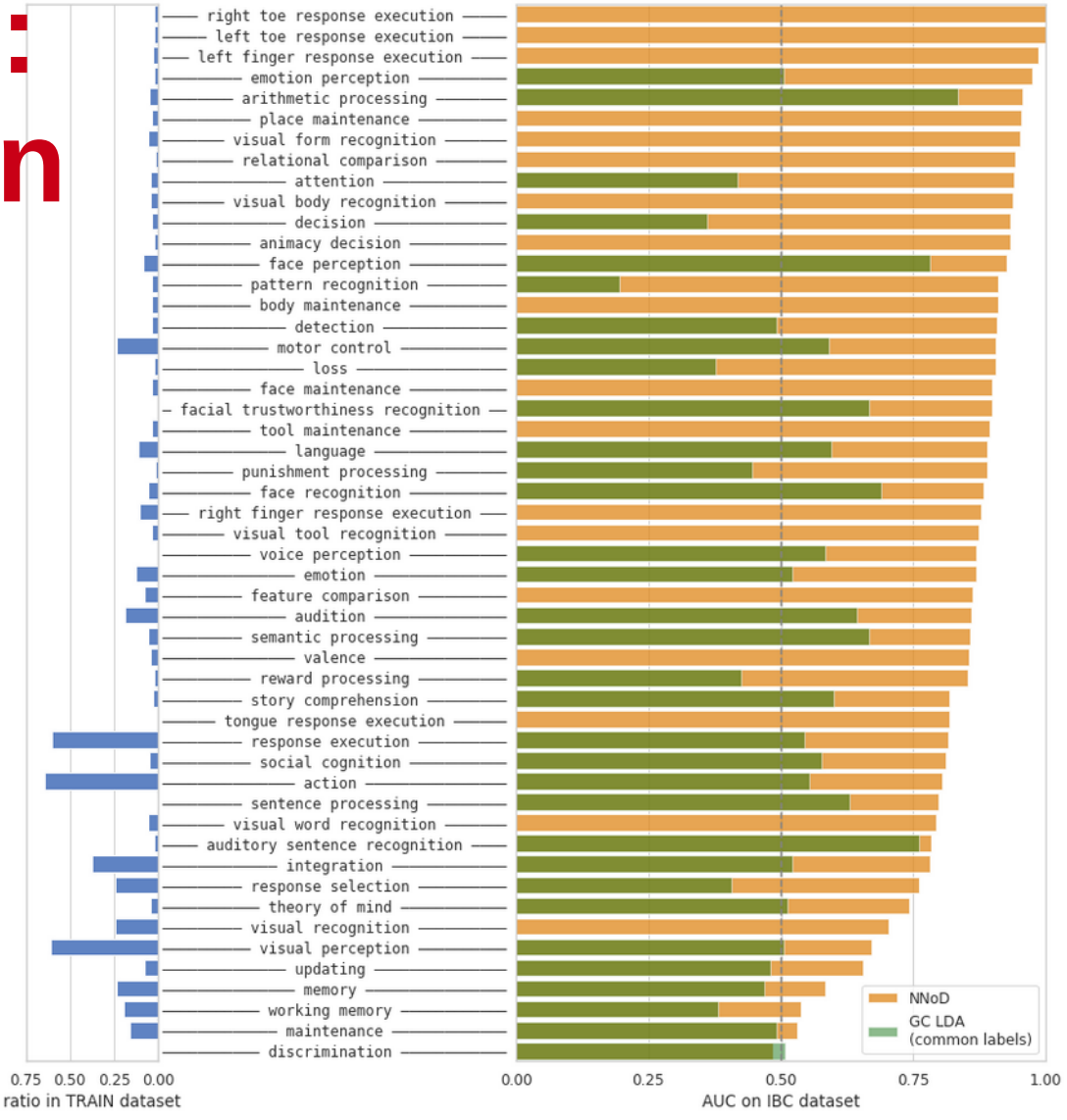
- Decoding model that generalizes across datasets
- Relies on network models

[Mensch et al subm]

Preliminary result: Neural networks on a Dictionary

- 51 terms decoded on NeuroVault

[Menuet et al. Subm.]



09/09/2020

Aim 2: Assessment of coordinated (image/text) representations

- Statistical inference
  - **Aggregation approaches** (across clusterings of brain domain or concepts) [Nguyen ICML 2020]
  - **Generative approaches**:
    - Knockoffs
    - Latent cause models
  - Generalization to **non-linear** models

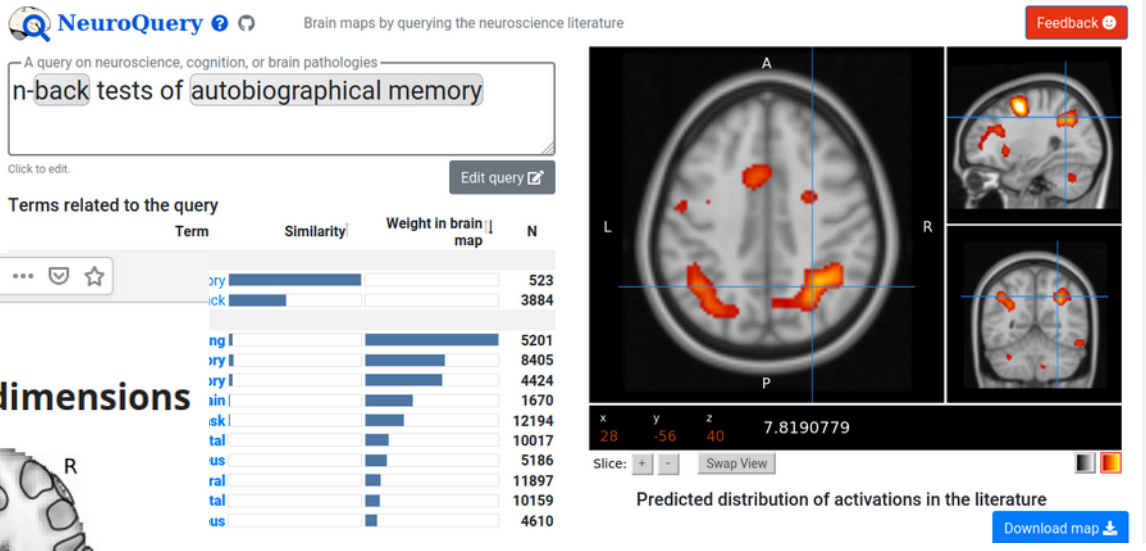
From linked representations to formal reasoning: integration into Neurolang

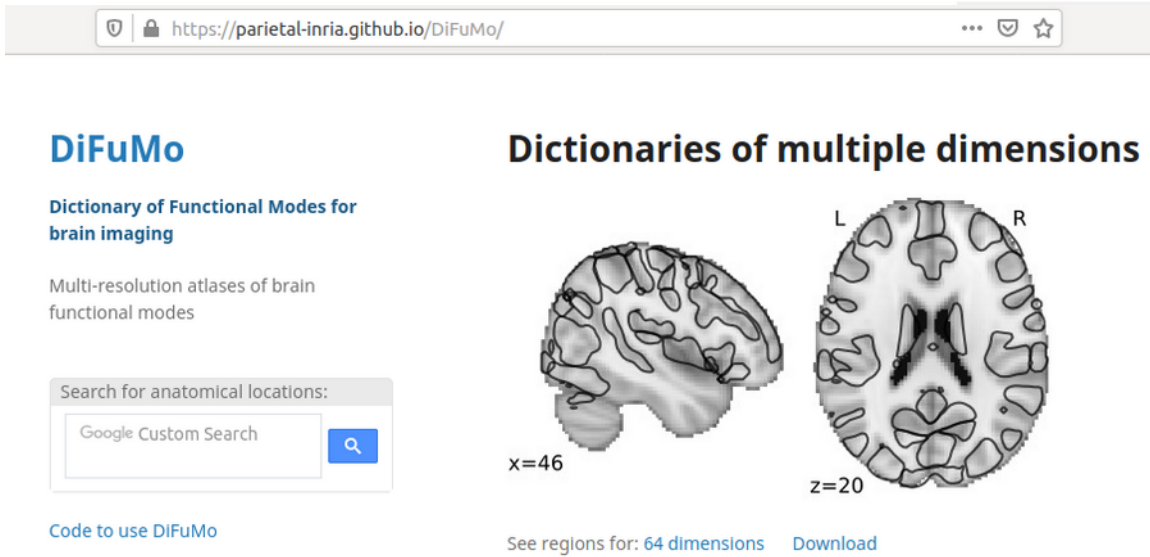
Neurolang → DSL for human neuroscience research
→ combine imaging data, anat. descriptions & ontologies

1. represents information in a syntax close to **natural language**
2. **Querying ontologies** w. same expressive power as current standards (SPARQL, OWL)
3. **Probabilistic** language --- graphical models allowing the implementation of many ML algorithms

Current integration of meta-analytic queries and information in Neurolang

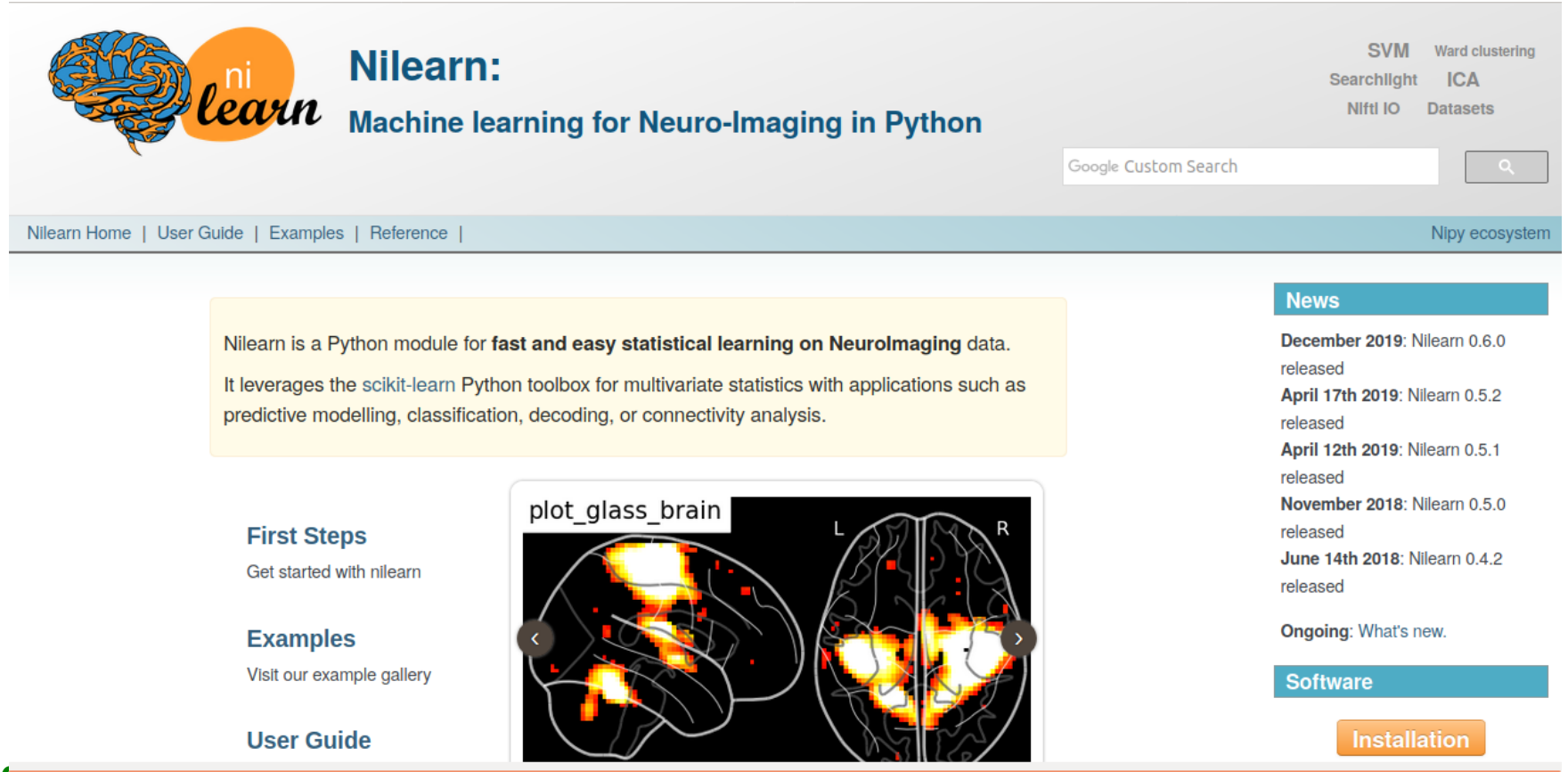
Aim 3: Provide neuroimaging semantic-learning tools and interfaces





http://neuroquery.org

Software development & dissemination



Conclusion

- 5-year project
- A new PhD student
- A starting position
- Integrate & push forward are current assets: MODL, DiFuMo, NnoD, neuroquery, NeuroLang, IBC

Thanks

- Gael Varoquaux
- Demian Wassermann
- Aapo Hyvärinen
- Huges Talbot
- A. Mensch, J. Dockès, R.Menuet, K.Dadi, J.A. Chevalier, Binh Nguyen, H.Richard, T.Bazeille
- All Parietal fellows