NEBULA101: an open dataset for the study of language aptitude in behaviour, brain structure and function

Alessandra Rampinini* [a,e], Irene Balboni* [a-d-e], Olga Kepinska[b,c], Raphael Berthele [d,e], Narly Golestani [a,b,c,e]

[a] Department of Psychology, Faculty of Psychology and Education Science, University of Geneva, Geneva, Switzerland

[b] Vienna cognitive Science Hub, University of Vienna, Vienna, Austria.

[c] Department of Behavioural and Cognitive Biology, Faculty of Life Sciences, University of Vienna, Vienna, Austria.

[d] Institute of Multilingualism, University of Fribourg, Fribourg, Switzerland.

[e] National Centre for Competence in Research Evolving Language, Switzerland.

Abstract

This paper introduces the "NEBULA101 - Neuro-behavioural Understanding of Language Aptitude" dataset, which comprises behavioural and brain imaging data from 101 healthy adults to examine individual differences in language and cognition. Human language, a multifaceted behaviour, varies significantly among individuals, at different processing levels. Recent advances in cognitive science have embraced an integrated approach, combining behavioural and brain studies to explore these differences comprehensively. The NEBULA101 dataset offers brain structural, diffusion-weighted, task-based and resting-state MRI data, alongside extensive linguistic and non-linguistic behavioural measures to explore the complex interaction of language and cognition in a highly multilingual sample. By sharing this multimodal dataset, we hope to promote research on the neuroscience of language, cognition and multilingualism, enabling the field to deepen its understanding of the multivariate panorama of individual differences and ultimately contributing to open science.

Background & Summary

Individual differences in language. This paper describes the "NEBULA101 - Neuro-behavioural Understanding of Language Aptitude" dataset. The dataset collects behavioural and brain imaging data of 101 healthy adults for the study of individual differences in language and cognition.

Human language is a complex behaviour, and crucial to understanding its workings is the fact that individuals differ in the way they manifest this and other cognitive skills[1]. The science of individual differences has expanded in recent years, thanks to a more integrated, less modular take on cognition: by combining the study of behaviour and the brain in a deep phenotyping approach, mindful of individual differences, researchers can gain a more comprehensive understanding of complex cognitive functions, language included[2].

Language aptitude. One aspect of language shown to display large individual differences is language *aptitude.* Language aptitude was originally proposed to explain why some people display remarkable abilities when learning additional languages [3–5]. We use the term *additional* language for any language that is not the (or one of the) individual's *first* language(s), i.e., language(s) to which they were exposed from birth. The term "additional" thus includes *second languages* (e.g. in the context of migration), *foreign languages* (e.g. in classroom learning), or *third/additional languages* of multilinguals who use more than two.

Initially, researchers viewed language aptitude as a stable trait, comprising phonetic coding, grammatical sensitivity, inductive learning, and rote learning abilities. These skills were seen as componential, in the climate of the first cognitive revolution, where the brain-mind-behaviour interaction was seen as the execution of algorithms operating within cognitive modules[6]: in this view, phonetic coding was at the core of production and perception of speech sounds, grammatical sensitivity underlay the capacity to identify structure in language (i.e. morphosyntax), inductive learning supported the generalisation of language rules from the input, and rote learning skills the construction of vocabulary via routinary use. Research has since suggested a more parsimonious structure, combining grammatical sensitivity and inductive learning into a global language analytic ability[7].

Nowadays, with advances in neuroscience and experimental psychology methods, the language aptitude construct overall still describes a set of skills operating *across* the hierarchy of all language components, from lower to higher levels of complexity. What has changed is the way we conceptualise language itself, which has also changed the way we view language aptitude: in the second wave of the cognitive revolution, thanks to usage-based linguistics and neural-network psychology, language is seen as a complex adaptive system, rather than a set of rigid structures and fixed operations[8]. Language aptitude is part of this dynamic, as it can vary with age[9–11] or with multilingual experience (likely increasing meta-linguistic awareness[12–14], with complex patterns yet to be clarified[15,16]).
Therefore, despite a certain degree of modularity, cognitive domains are thought to interact in complex ways. Mindful that any quantitative analysis will always require some degree of operationalisation (i.e. to derive scores from tests, we need to identify components of scientific constructs that such tests might tap into), our position is that this view of language as a dynamic system branching out and connecting to more general mechanisms holds promise for better understanding language itself and the human cognitive system more generally.

Aptitude for language(s) and individual differences in other domains of cognition. The construct of language aptitude was developed in the domain of foreign language learning, as explained, but the idea that aptitude only manifests in foreign languages has now been surpassed, since individual differences can be observed in first language skills too, even if it is harder to pinpoint them and isolate them from experiential factors[17]. Moreover, such individual differences might co-exist more globally with individual differences in other

domains of cognition[18,19], giving rise to "neurocognitive profiles" of language aptitude involving the mnemonic domain, fluid reasoning, auditory abilities, and even musicality[20]. It is therefore relevant to ask ourselves whether language, in this modernised and dynamic view, is part of a positive manifold [21,22] originating from the beneficial interactions between cognitive processes, as we proposed in a recent analysis of part of these data[18]. Both the stable and malleable (or *plastic*) features of the human cognitive system are fundamental to such interactions.

Stability and malleability, predisposition and experience. This dataset presents cross-sectional data. Nonetheless, to understand the nature of the language (aptitude) construct, it is important to consider it as the result of a complex interaction between stable traits and malleable states, the first ascribed to genetic predisposition and the latter to environmental influences (what we generally define as "experience")[23–25]. While we provide a *snapshot* of individual profiles at a given moment in time, in order to formulate relevant questions on (and interpretations of) these data, it is important to remember that the observed measures arise from both stable traits and from malleable skills, or states[26]. Cognitive activity[27–30], including language learning and use[31–34], and brain *plasticity* are intrinsically related. Changes in regional activity, network connectivity, or morphology, arising from underlying molecular, neurobiochemical and other changes[27], subtend this state of malleability in brain function and structure[35], both during development and in skill learning. However, brain functional and structural architecture is also highly polygenic (i.e. controlled by the complex interaction of multiple genes, which in turn are expressed in multiple variants across individuals), and in addition, different cortical loci and tissue features (thickness, surface area) are affected to different degrees by genetics[36]. Further, genetic factors also likely influence the degree to which neuroplasticity manifests[37].

The open science of language aptitude. Language aptitude has come a long way since its study was confined to the foreign language classroom, and research restricted to military and government access[38]. The construct now encompasses some of the most promising avenues for research on the human cognitive system more generally: the roles of predisposition and experience, the nature of neuroplasticity and the integrated and multivariate organisation of cognitive domains. Because of its relevance for studying questions on language and cognition more generally (as well as investigating the role of the environment, e.g. the experience of multilingualism), this development in the conceptualisation of language aptitude is of particular interest to the ever-growing world of Findable, Accessible, Interoperable, and Reusable (FAIR) neuroscience data[39], which has faced a substantial growth in the last 15 years[40], and possibly as many challenges[41,42]. FAIR principles are accelerating our comprehension of the human brain[43]. Concurrently, the existence of standardised protocols such as the Brain Imaging Data Structure (BIDS)[44] supports this recent strive towards data and code accessibility, ensuring that data is organized and its analysis reproducible, enabling more efficient and effective use of shared datasets and streamlining data preparation, thus allowing scientists to focus more on discovery.

Mindful that the panorama of shared neuroimaging data is vast and that datasets might be scattered around repositories of the commercial and institutional type, focusing on either a specific (often clinical) population or modality[45], we searched for openly available data similar to ours (MRI and behavioural data from healthy participants focusing on language) on OpenNeuro[46], one of the most recent, easy to access and growing neuroimaging databases. At the time of writing, a search on http://openneuro.org for BIDS-valid MRI raw/derivative datasets with more than 50 healthy adult participants, including the keyword 'language' and accompanied by a published or pre-registered data descriptor, yields the following entries (Table 1), with two of the datasets (ds004215, ds000243) being listed but having no relationship with language (and thus not being included in the below table).

OpenNeuro Accession number	*Dataset name*	*Modality*	*Number of participants*	*Complete dataset (all participants-measures)*	*Associated tasks/measures*	*Main topic*
ds003481	A Dataset to Study Pragmatic Language and Its Underlying Cognitive Processes[47,48]	sMRI, fMRI, behaviour	145	No	• Comprehension of metaphorical phrases and proverbs • Recognition and emotional categorisation of speech acts • Lexical-semantic processing	Language pragmatics
ds004765	Relationship between resting state functional connectivity and reading-related behavioural measures in 69 adults[49,50]	sMRI, fMRI (resting state), DWI	69	Yes	• Word/nonword reading • Spelling • Lexical decision • Spoonerisms • Rapid automatised naming • Non-word repetition • Vocabulary	Reading
ds002382	Age-related differences in auditory cortex activity during spoken word recognition[51]	sMRI, fMRI	61	Yes	• Word listening • Word repetition	Word processing
ds004285	Listening task*	sMRI, fMRI	78	Yes	• Word repetition	Word processing
ds004073	Comparing language lateralisation using fMRI and fTCD[52]	fMRI, functional transcranial doppler sonography (fTCD)	51	No (minimally missing information)	• Word generation • Sentence generation • Phonological decision • Word comprehension • Sentence comprehension • Syntactic decision	Language lateralisation
ds001747	Exploring the Resting State Neural Activity of Monolinguals and Late and Early Bilinguals[53]	sMRI, fMRI (resting state)	92	Yes	• L1 and L2 (where applicable) proficiency • Language background	Bilingualism
ds001796	Bilingualism and the brain**	sMRI, fMRI (task-based), fMRI (resting state), DWI	64	Yes	• Flanker task	Bilingualism
ds002345	Narratives: fMRI data for evaluating models of naturalistic language comprehension[54]	sMRI, fMRI	345	No	• Naturalistic story listening	Language perception
ds003643	Le Petit Prince: A multilingual fMRI corpus using ecological stimuli[55]	sMRI, fMRI	112	Yes	• Naturalistic story listening	Language perception

Table 1. Available neuroimaging datasets on language with more than 50 healthy participants, validated in BIDS. (*)Possibly including or revising ds002382 but no information on dataset version is provided. (**) No data descriptor available.

A similar, broad search on Google Scholar for data descriptors associated with openly available language datasets ("open AND mri dataset AND language") yielded two more datasets: one is MOUS (Mother Of Unification Studies)[56], a dataset comprising sMRI, DWI, resting state and task-based fMRI and MEG in 204 participants assigned to either sentence reading or listening; the other is the Alice dataset[57], where 75 participants listened to the same chapter of Alice in Wonderland during fMRI or during EEG. A third dataset presents a large quantity of data collected in fewer participants with state-of-the-art technology: this is the high-resolution, 7T fMRI Forrest Gump database[58], during which participants watched the 'Forrest Gump' movie. This data can be used to study naturalistic language processing, even though the study did not primarily focus on language. Finally, the search also yielded two speech production datasets of vocal tract MRI[59,60].

At the behavioural level there is more heterogeneity in the type and quantity of shared data[45]. The origins of such heterogeneity have been discussed for quite some time. In a Nature Neuroscience perspective article from ten years ago, the challenges linked to standardisation and accessibility of behavioural data were already discussed[61]. Importantly, behaviour was defined as a complex, highly dimensional, dynamic, and interconnected phenomenon without distinct separable scales, and this was discussed as being one of the leading causes for lack of standardisation and scarce FAIR compliance. The authors insisted on its foundational and unifying nature, and called for improved standards: "Behaviour (…) is the principal function of the brain. (…) Copious, quantitative and open behavioural data has the potential (…) to solidify the foundations of other [disciplines], including neuroscience" [61] (p.1455).

However, even in the cited datasets, which represent timely and relevant steps towards accessibility of neuroscience data, the phenotypical (behavioural) information being shared beyond demographics is still relatively less prominent than the endophenotypical (neural) data: oftentimes, only data from the behavioural tasks being performed during fMRI are available. Moreover, even when behaviour was tested outside the scanner, if the original project did not focus on *both* phenotypical and endophenotypical data and on a specific topic (such as, in the cited examples, pragmatics[48] or reading[50]), the accompanying behavioural information is scarce. Two notable, recent exceptions in the field of individual differences in language are represented by a behaviour-only dataset including 33 measures from 112 adult Dutch speakers[62], and by recent work by Berthele and colleagues in children[63], both representing an important milestone for shared behavioural data.

Given the lack of standardisation in the way we administer behavioural tasks outside the scanner, compared to fMRI task delivery, it seems daunting to force the structure of behavioural paradigms and log files produced by a plethora of software and online platforms to accommodate information within the structure required by the BIDS standard. Some initiatives, such as the Behaverse project[62], are proposing data structures which, they claim, can accommodate phenotypical information better than what BIDS can do. However, it is important to note that any individual differences dataset including both phenotypical and endophenotypical information will have the added strength of multimodality, and to date, BIDS is the only

data structure protocol that can accommodate *both* in a relatively straightforward way, even if it requires extra (and sometimes *post-hoc*) work to prepare the materials intended to be shared. Finally, we must note that when sharing mixed raw and derivative datasets, behavioural data can be easily included as derivatives, these having a more liberal structure (such as in the case of the dataset we present here), lifting from the end-user the load of reprocessing and calculating basic scores starting from item-level data.

Our NEBULA101 dataset aims to promote the study of individual differences in language to better understand a multivariate cognitive system, via the sharing of a truly multimodal dataset in an adequately sized participant sample[64]. We provide sMRI, DWI, task-based and resting state fMRI in 101 individuals, together with broad behavioural data on linguistic but also on non-linguistic, domain general and domain specific tasks, including cognitive and perceptuomotor tasks. This includes measures of all language aptitude components from phonetics to syntax, measures of reading and reading mediators (e.g. phonological awareness), domain-general cognitive skills, numerical processing, musicality and musical experience and rich multilingual language experience measures. By providing these data to the public domain, we hope to contribute new discoveries to the over-arching construct of language (aptitude), embracing the components of individual behavioural and neural phenotypes as widely as possible.

Data types. Behaviour is the phenotype that can be related to, or unify, genetics, neural architecture, neural activity, body structure, physical limitations, and environmental factors[61]. Given the importance of behavioural data in exploring individual differences at the neural level, this dataset includes 28 scores derived from 8 questionnaires, and 74 behavioural measures derived from 21 tasks. Functional neuroimaging data provide information about the neural correlates of cognitive processes, allowing to elucidate how the brain supports specific behaviours and skills. Here, we provide resting-state and task-based functional imaging (fMRI) data, the latter obtained during a language localiser[65]. Finally, NEBULA101 also includes anatomical T1-weighted and diffusion-weighted (DWI) imaging of the brain, which will allow to study brain anatomy and white-matter structural connectivity, to shed light on the brain structural correlates of linguistic behaviours, aptitudes and experiences.

An overview of all measures provided in the NEBULA101 dataset is provided in Table 2.

Test	*Modality*	*Construct*	*Delivery*	*Instructions*	*Derived score*	*Specifications*	*Reference*
Language Experience and Proficiency Qu estionnaire (LE AP-Q)	Questionnaire	Multilingual language experience.	Online, unsupervised	Fill in the provided form.	Entropy	Extended to accommodate up to 50 languages. Added questions on: *Time spent* in contexts such as online communities, fandoms and subcultures where a language is spoken. *Contributors* to language learning: social media, apps and everyday life in the country where a language is spoken. Removed question on cultural identification. French adaptation.	Marian et al., 2007[64]
Code Switching questionnaire	Questionnaire	Code switching habits.	Online, unsupervised	Fill in the provided form.	Voluntary switching Involuntary switching	Selected only voluntary and involuntary switching scores. French adaptation.	Rodriguez-Fornells et al., 2012[66]
Motivational Factors Questionnaire (MFQ)	Questionnaire	Motivation and attitude towards foreign languages (FL).	Online, unsupervised	Fill in the provided form.	Ideal foreign language self Instrumentality International contact Foreign language interest Foreign language anxiety Foreign language confidence Milieu Usage willingness of foreign languages	French adaptation. 5-point Likert scale instead of 6-point. Constructs/items assuming that participants are actively involved in studying foreign languages (academically) were not included. Items that refer specifically to the English language, anglophone world and culture reworded to fit a more global context where possible or excluded. Included constructs are shown in the Derived Score column.	Ryan, 2008[67] Thompson & Lee, 2018[68]
Adult Reading History Questionnaire (AHRQ)	Questionnaire	Reading history.	Online, unsupervised	Fill in the provided form.	Reading history score	French version of the questionnaire.	Lefly & Pennington, 2000[69]
Internal Representations Questionnaire (IRQ)	Questionnaire	Modes of internal reasoning.	Online, unsupervised	Fill in the provided form.	Manipulation Factor Orthographic Factor Verbal Factor Visual Factor	French adaptation.	Roebuck & Lupyan, 2020[70]
Music Use and Background Questionnaire (MUSEBAQ)	Questionnaire	Music training, capacity, preferences, and motivations.	Online, unsupervised	Fill in the provided form.	Index of music training Index of music listening Index of musical instrument playing Cognitive and Emotional Regulation Social Connection Engaged Production Dance Physical Exercise	French adaptation.	Chin et al., 2018[71]
Barratt's Simplified Measure of Socioeconomic	Questionnaire	Socioeconomic status.	Online, unsupervised	Fill in the provided form.	Barratt's Simplified Measure of Socioeconomic Status	French adaptation.	Barratt, 2006[72] Rakesh & Whittle, 2021[73]

Status (BSMSS)							
Edinburgh Handedness Inventory (EHI)	Questionnaire	Handedness.	Online, unsupervised	Fill in the provided form.	Handedness score	10-item version.	Oldfield, 1971[74] Nedjar et al., 1989[75]
Artgram	Behavioural task	Language analytic abilities / Morphosyntax.	Online, supervised	Study the provided dictionary and choose the appropriate adaptation for given sentences by recognising the use of morphological cues.	Accuracy RT	7-word dictionary to study for 3 minutes. 12 sentences to translate, with 4 possible choices, only 1 correct choice. Self-advancement. 15-minute time limit.	Developed in-house.
Modern Language Aptitude Test 5 (MLAT5)	Behavioural task	Rote learning.	Online, supervised	Study the provided dictionary and choose the appropriate adaptation for given words.	Accuracy RT	French adaptation.	Stansfield, 1989[76]
Farsi uvular Production Task	Behavioural task	Foreign sound production.	Online, supervised	Listen to and reproduce words containing a foreign language sound not present in French.	Accuracy	n/a	Golestani & Pallier, 2007[77]
Hindi Dental Retroflex Contrast	Behavioural task	Phonological categorisation/discrimination.	Online, supervised	Categorise a foreign language sound not present in French. Training and testing blocks.	Accuracy	200 trials were presented.	Golestani et al., 2002[78]
Brocanto	Behavioural task	Language analytic abilities / Pattern recognition.	Online, supervised	Recognise grammatical and ungrammatical sentences with different structures and violation types in an artificial language, inductively. Training (reading-only) and testing (judge grammaticality by button press) blocks.	Accuracy RT (For each block 1-3 and overall)	Version from Kepinska et al., 2017 with minor modifications related to timing and condition counterbalancing, as follows: A practice screen was added at the beginning of each testing block to test button press and rule comprehension. Button press instructions: 1 for "grammatical" and 0 for "ungrammatical". We chose a 2-second fixation cross without jittering. We chose a 6-second sentence presentation in both testing and training block types. We maintained 40 sentences per testing block, but we included 8-word sentences and obtained a 1/2 ratio of grammatical and ungrammatical sentences **overall** (N=120, of which the 60 ungrammatical sentences were new and roughly 1/3 of the 60 grammatical ones [N=22] had been presented during training).	Kepinska et al., 2017[79] Opitz and Friederici, 2003[80]
Raven's Advanced Progressive	Behavioural task	Non-verbal intelligence.	Online, supervised	Select the missing block to complete a picture from a set of proposed choices.	Accuracy RT	Advanced and abridged French version. Time limit of 20 minutes for 23 trials. Programmed for computer-based presentation on the Gorilla platform.	Raven, 1998[81]

Matrices (APM)							
Corsi block	Behavioural task	Visuospatial memory	Online, supervised	Watch a sequence of blocks being highlighted, and repeat the same sequence by clicking on them in the same order and then backwards.	Forward span Backward span Total span	Programmed for computer-based presentation on the Gorilla platform in place of in-person. The blocks were presented as 9 2-dimensional black squares arranged on a white background. Blocks were black. Sample block sequences were shown in yellow. Participants clicked on the blocks instead of tapping them with their hand. Selected blocks were shown in green upon clicking. Block placement on the screen followed Kessels et al. 2000[82]	Corsi, 1972[83] Arce & McMullen, 2021[84]
Digit Span	Behavioural task	Auditory working memory	Online, supervised	Recollect the digits immediately after the presentation in the same order, in the 'forward' section of the task, or in the inverse order, in the 'backward' section.	Forward span Backward span Total span	Auditory modality. Response collected through upper numeric keypad.	Wechsler, 1955[85] Ryan et al., 2014[86] Conway et al., 2005[87]
Revised Tempo Test	Behavioural task	Arithmetic abilities	Online, supervised	Solve 60 additions and 60 subtractions in 1 minute each.	Accuracy RT (Additions, subtractions, overall)	Programmed for computer-based presentation on the Gorilla platform in place of paper and pencil, otherwise unchanged. Response collected through upper numeric keypad.	Bellon et al., 2022[88]
Advanced Measures of Music Audiation (AMMA)	Behavioural task	Music audiation, musicality, musical aptitude.	Online, supervised	Judge difference or identity in melody or rhythm between pairs of musical excerpts.	Tonal score Rhythm score Total score	Programmed for computer-based presentation on the Gorilla platform.	Gordon, 1989[89]
ANT-I	Behavioural task	Attention networks: executive control, alerting, orienting.	In person	Detect arrow orientation in presence of flankers and sound cues.	Alerting gain Orienting gain Reorienting gain Inhibition gain	French adaptation (instructions) 18 conditions and 432 trials.	Callejas et al., 2005[90]
CVLT	Behavioural task	Verbal working memory	In person	Remember lists of words short-term and long-term. Recognise previously presented words, long-term.	Accuracy immediate recall Accuracy immediate and short-term recall Accuracy long-term recognition RT long term recognition	Programmed for computer-based presentation on the Gorilla platform. Removed cued recall (short and long-term).	Deweer et al., 2008[91]
Finger tapping Test	Behavioural task	Fine motor skills	In person	Tap with the index finger of the dominant and then non-dominant hand on the spacebar as fast as possible.	Average finger tapping score dominant hand overall Average finger tapping score nondominant hand overall Average finger tapping score of dominant/nondominant overall (Calculated over blocks)	Abbreviated version of the Finger Tapping test as proposed by Ashendorf et al., 2015. 5 trials per hand. 3 dominant hand trials – 1-minute break – 2 dominant hand trials. 3 non-dominant hand trials – 1-minute break – 2 non-dominant hand trials.	Strauss et al., 2006[92] Ashendorf et al., 2015[93]

Purdue Pegboard Test	Behavioural task	Fine motor skills	In person	Insert pegs in holes, first with the dominant and and then with the non-dominant hand. Build assembly of pegs alternating hands as instructed.	Accuracy, dominant hand Accuracy, nondominant hand Accuracy, simultaneous hands Accuracy, assembly task Average accuracy, dominant-nondominant-simultaneous hands (Calculated over blocks)	French adaptation (instructions)	Tiffin & Asher, 1948[94]
Rapid Automatised Naming (RAN)	Behavioural task	Naming automatisation	In person	Rapidly denominating objects, digits, colours	Accuracy RT	French adaptation. Programmed for computer-based presentation on the Gorilla platform.	Frederickson et al., 1997[95]
Phoneme suppression	Behavioural task	Phonological awareness	In person	Repeat words by omitting the first phoneme.	Accuracy RT	Programmed for computer-based presentation on the Gorilla platform. Pre-recorded female voice.	Rutten et al., 2019[96]
Text Reading	Behavioural task	Reading skills	In person	Read two texts of increasing difficulty.	Accuracy RT	Programmed for computer-based presentation on the Gorilla platform.	"Le Pollueur", Gola-Asmussen et al., 2010[97] "L'Alouette", Lefavrais, 1967[98]
Word and Pseudoword Reading	Behavioural task	Reading skills	In person	Reading lists of words and pseudowords.	Accuracy RT (Per stimulus type: regular words, irregular words and pseudowords; overall)	With the target population in mind, this task was made more difficult by merging two standardized dyslexia assessment tests in French, the ECLA16+ and the EVALEC: - There were 56 regular words, 52 irregular words, and 56 pseudowords overall. - For each category 20 words were taken from the Ecla16+ and the remaining from the EVALEC. - The words from the EVALEC were randomized within-list rather than mixing across lists, to be able to time and score the lists per word-type (regular, irregular and pseudowords).	EVALEC: Sprenger-Charolles et al., 2005[99] ECLA16+ : Gola-Asmussen et al., 2010[97]
Spelling task	Behavioural task	Spelling skills	In person	Write down words, pseudowords and irregular words after hearing them.	Accuracy (regular words, irregular words and pseudowords; overall)	Programmed for computer-based presentation on the Gorilla platform.	Gola-Asmussen et al., 2011[97]
Spoonerisms	Behavioural task	Phonological awareness	In person	Swap the first phoneme of word pairs.	Accuracy RT	Programmed for computer-based presentation on the Gorilla platform. Pre-recorded male voice	Szenkovitz & Ramus, 2005[100]
Non-word repetition	Behavioural task	Phonological working memory	In person	Repeat non-word lists of increasing length.	Accuracy of repetition Span, i.e. maximum number of repeated words	Programmed for computer-based presentation on the Gorilla platform.	Majerus, & Van der Linden 2004[101]
Magnetization Prepared -	sMRI	Brain structural anatomy	In person	Lie still in scanner.	n/a	whole-brain coverage 1mm isotropic voxel FOV read = 256mm, FOV phase 93.8% TR = 2300ms, TE = 3.26ms	Work-in-progress compressed sensing sequence © Siemens Healthineers

RApid Gradient Echo (MPRAGE) imaging						Flip Angle: 9° Distance Factor 50 % Orientation: Sagittal Phase Encoding Direction: A >> P	
Diffusion-weighted Imaging (DWI)	dMRI	Fiber tractography	In person	Lie still in scanner.	n/a	Multishell sequence: 1.5mm isotropic voxel FOV read = 225mm, FOV phase 100% whole-brain coverage TR 6700.0 ms, TE 74.00 ms Distance Factor 0 % Acceleration factor SMS = 2 GRAPPA = 3 Orientation: Transversal Phase Encoding Direction: A >> P Diffusion-encoding gradient directions: 117 12 B_0 volumes distributed along the sequence 7 volumes at 700 s/mm2 30 volumes at 1000 s/mm2 68 volumes at 2850 s/mm2	Fedeli et al., 2021[102]
Language network brain functional localiser	fMRI	Functional activation for language	In person	Lie still in scanner, eyes open, fixate cross, listen to story	n/a	Added a condition for degraded second language. Added a fixation at the beginning and at the end of each run.	Malik-Moraleda et al., 2022[65]
Resting-state brain function	fMRI	Resting-state functional activation	In person	Lie still in scanner, eyes open	n/a	2mm isotropic voxel FOV read = 224mm, FOV phase 100% whole-brain coverage 72 slices TR = 2000ms, TE = 32ms Flip Angle: 75° Distance Factor 0 % Acceleration factor SMS = 3 Orientation: Transversal Phase Encoding Direction: A >> P	Developed in-house
Field map	MRI	Intensity of the magnetic field across space	In person	Lie still in scanner.	n/a	Intended to correct B0 distortion in fMRI: 2.4x2.4x2mm^3 voxel FOV read = 225mm, FOV phase 100% whole-brain coverage 72 slices (fMRI), 66 slices (DWI) TR = 700ms, TE1 = 4.92ms, TE2=7.38ms Flip Angle: 60° Distance Factor 0 % Orientation: Transversal Phase Encoding Direction: R >> L	Developed in-house

Table 2. Overview of all tasks and modalities.

Methods

Participants

According to the Organisation for Economic Cooperation and Development (OECD), approximately 40% of Swiss individuals aged 25 to 34 possess upper secondary or post-secondary non-tertiary education, around 50% have attained tertiary education, and about 10% have below-secondary education[103,104]. Additionally, Switzerland is recognized for its linguistic diversity. The Federal Population Census 2022 Structural Survey[105] indicates that French is spoken by 22.8% of the population. In the survey, 70.1% and 23.4% of the population declared that they speak other national (German, Italian, and/or Romansch) or non-national languages (participants could declare more than one language, therefore there is a degree of overlap in these data). Considering that any study examining the interaction between diverse cognitive domains and language skills will always be somewhat culture-dependent[106], we chose our target participants with the aim to test at least 100 healthy, relatively multilingual individuals having French as their first or dominant language, which is the primary language of the Canton where the data collection took place.

We recruited 104 adult participants from the Geneva area, the surrounding French-speaking cantons of Switzerland and neighbouring France through flyers and online advertisements. Participants provided their consent for disclosing their medical history and filled in an online screening survey. To safeguard coherence and avoid confounding factors within the sample, prospective participants were excluded *a priori* if they had musical or simultaneous interpreting professional qualifications (known to interact with language); vision defects that could not be corrected; body implants incompatible with MRI or known claustrophobia; neurological or psychiatric conditions; traumatic head injuries with loss of consciousness; ongoing pregnancies; and past or present serious illnesses that required invasive and/or continued medical treatment (such as cancer, chronic and/or autoimmune diseases). Participants with MRI-conditional implants were evaluated on an individual basis upon providing further documentation. Participants with diagnosed developmental dyslexia, as well as those who reported knowing more than 10 languages with a self-reported proficiency equal to or higher than 4 out of 10 in reading, speaking and listening, are not included in this dataset. Figure 1 reports the number of declared languages for the final sample.

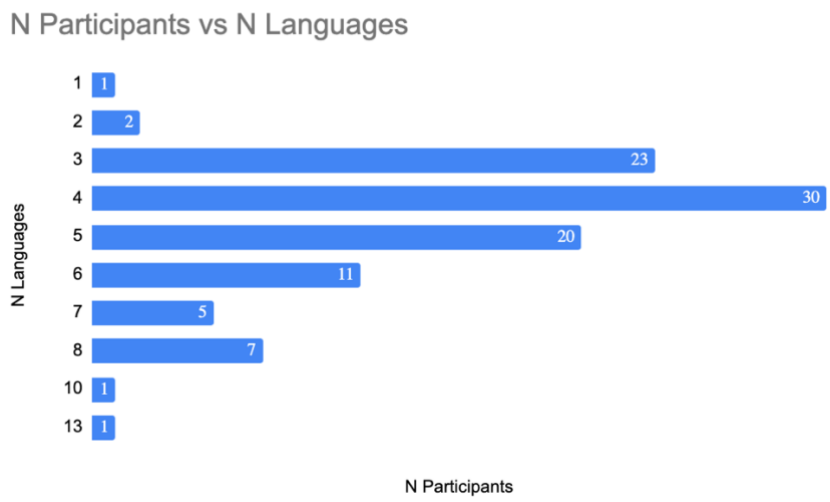

Figure 1. Frequency distribution of the number of reported languages in the final sample (N = 101).

Once eligibility was established, all participants provided a signed informed consent to all subsequent experimental procedures, including anonymised data reuse for open science. One participant subsequently withdrew their consent to data sharing and two were not able to undergo brain imaging due to claustrophobia. Therefore, the final sample includes 101 individuals (M_{age}= 23.35 years *SD* = 4.08, 68F). At study completion, all participants received monetary compensation, an image of their brain and a simplified report on their performance on the behavioural tests. Language background information can be seen in Figure 2.

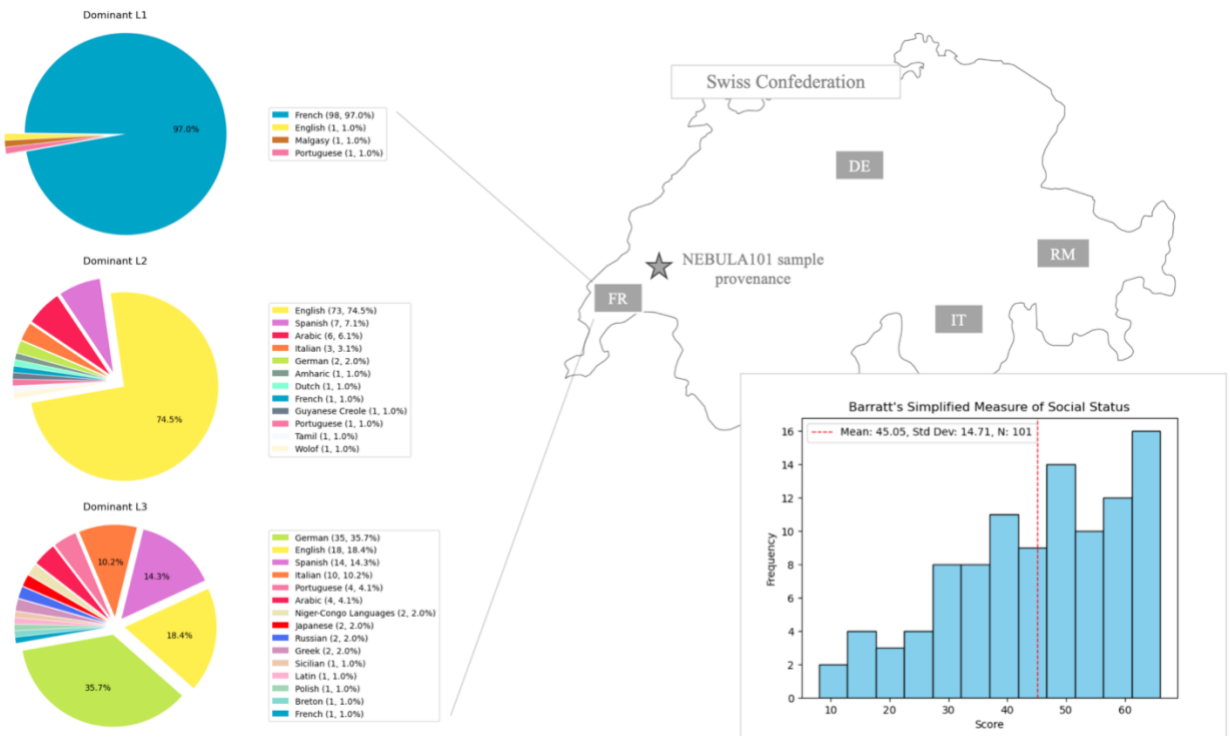

Figure 2. Participants' language background up until their third language (left, languages coded by colour) and Socioeconomic status (bottom right) measured via Barratt's simplified measure of social status[72].

Data Collection

Data collection. All interactions and documents provided to participants were in French. Data were collected by six individuals who were either first-language French speakers, or used French as their second/additional language at an advanced level. All procedures were approved by the Geneva Cantonal Ethical Commission (CCER Protocol N. 2021-01004) prior to the beginning of the study.

Data were collected in two online and two in-person sessions, hereafter named Session 1, 2, 3 and 4, which always occurred on different days but in the same order (1-4) between July 2022 and June 2023. Session 1 was unsupervised, and required participants to fill out online questionnaires. Session 2 involved online behavioural data collection, and was supervised by an experimenter. During Session 3 we collected more behavioural data, this time in person. Finally, in Session 4 we collected neuroimaging (MRI) data.

Behavioural testing. Using established and published tests in cognitive psychology is crucial for addressing the reproducibility crisis and curbing the proliferation of one-off tests, for ensuring that findings are accurate and that they can be compared and replicated by other researchers[107]. Therefore, where possible, we chose existing questionnaires and behavioural tasks, with the exception of an explicit morphosyntax learning test, which we developed and piloted ourselves due to the lack of such a test in the field (see section "Pilot study"). All instructions and tests were delivered in French. Where the French version of a test was not available, it was adapted and checked by at least one first-language French speaker in the team: this process is documented in Table 2 as “French adaptation”. Questionnaire instructions (Session 1) were presented in written form. Behavioural task instructions (Sessions 2 and 3) were read by a commercial natural reader (https://www.naturalreaders.com/commercial/read), using the voice “Renee France” (if not otherwise specified in Table 2), at speed -1. fMRI task instructions were given by the experimenter via a microphone. All technical alterations and task adaptations were related to 1) language of delivery, or 2) adapting a paper & pencil test for computer-based delivery. These are thoroughly documented in Table 2. An overview of the data collection structure is given in Figure 3 and an extensive theoretical explanation of the tests included in sessions 1-3 can be found in our recent exploratory work on the behavioural correlates of language aptitude, which included this sample[18].

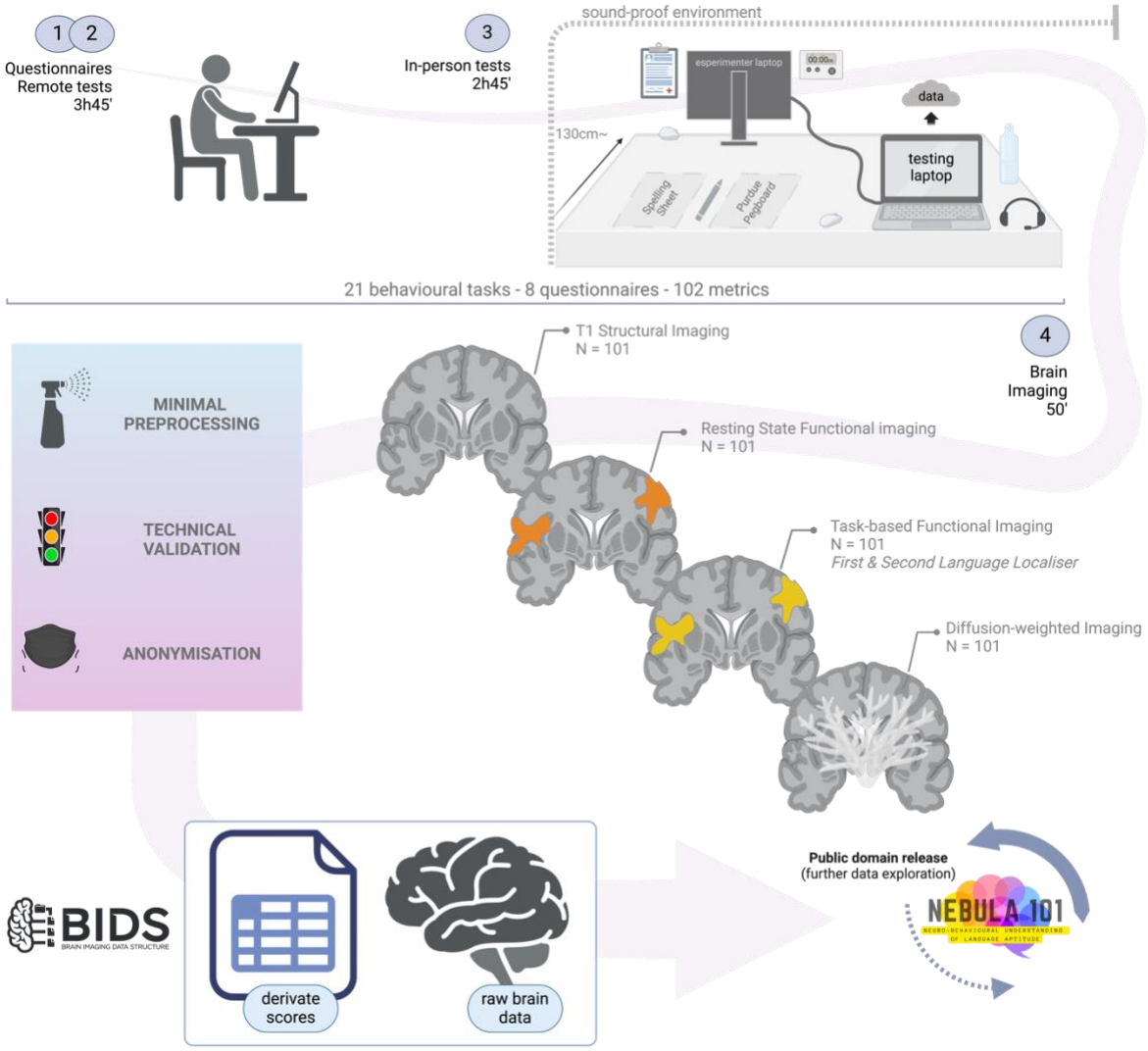

Figure 3. Data acquisition and processing structure (session duration is approximate for behavioural data acquisition due to individual variability in task completion times). [Illustration created with BioRender.com]

Pilot study. Before data collection, an explicit grammar learning task called ArtGram was developed and pilot-tested. ArtGram, designed for adults, extends the PLAB4 task used for language aptitude testing in children and adolescents, for which we could not find an equivalent in adults[108,109]. The test involves learning an artificial, declensional language lexicon with inflected sample sentences, followed by a self-paced multiple-choice, speeded translation task with novel sentences, as described in Table 2 and more extensively in Rampinini et al., (2024)[18]. The pilot study aimed to assess: (1) the task's feasibility, (2) the reliability of the online platform, (3) redundancy with an implicit grammar learning task (Brocanto). Twenty French speakers (11F, M_{age} = 27.65, *SD* = 8.9) without language or reading disorders participated via video conference. Results showed an unsignificant correlation of $r(18) = .44$ between the two grammar tasks.

Session 1: questionnaires. Participants initially filled out a series of questionnaires online using Qualtrics XM©. The questionnaire sequence was fixed: first came the Edinburgh Handedness Test and the BSMSS questionnaire, completed upon recruitment. Then came the demographics, MFQ, IRQ, AHRQ, MUSEBAQ,

Code Switching and LEAPQ (see Table 2). When available from the questionnaire manual, automatic scoring was implemented in Qualtrics via the "Formula Field" and resulted in one column per score in the derivative questionnaire dataset, for each participant.

Session 2: online behavioural testing. During session 2, supervised behavioural data collection was conducted via Zoom©. Participants were guided through a demonstration video on how to share both their sound and entire screen, use wired headphones for reliable online measurements[110,111], and ensure their microphone was functioning properly. Tasks were delivered through the Gorilla web interface[112]. Headphone and microphone tests from the Gorilla open materials section were mandatory before starting the task sequence. Researchers supervised the session, intervening only if technical issues arose. The system prevented participants from logging into the session from mobile phones or tablets, and only allowed the use of Mozilla Firefox© or Google Chrome©.

Participants navigated through tasks autonomously in a predetermined order among 15 possible pseudo-randomizations, one of which was automatically assigned by the system at the start of the testing sequence. Before each task, participants received written instructions in French on screen, and could not proceed until the natural reader finished delivering the same instructions orally. Fixed-length breaks (3 or 5 minutes) were included after the most intensive tasks to optimize concentration and compliance. Participants could end the break early if ready to continue, with a 60-second timer appearing in the last minute of the break if they had not yet proceeded to the next task.

Session 3: in-person behavioural testing. In this session, participants were tested in person for tasks that required closer supervision due to their length (e.g. the ANT-I), or that required manual measurements of reaction times (RT), (e.g. the literacy and literacy mediator tests). Testing was conducted at the Human Neuroscience Platform of "Campus Biotech" in Geneva, in a dedicated, sound-protected room, using the same laptop, mouse, and headset (headphone with microphone) for all participants (see Figure 3). Tasks were organized and delivered via the Gorilla interface using the previously described pseudo-randomization strategy. The same microphone check was included to ensure the safe recording of tasks requiring vocal responses for later assessment.

An experimenter closely supervised the session through a connected screen, mouse, and keyboard while facing the participant, intervening only when necessary due to task requirements or technical issues. For tasks requiring a vocal response, the experimenter manually recorded accuracy and/or RT (assessed via a chronometer) in the session booklet. To prevent data loss, these tasks were also audio-recorded, and the responses were later verified for accuracy by a team member with French as their first language. Task events not requiring manual measurement or verification were recorded directly in Gorilla. During this session, participants also performed three extra tasks that were part of another project[18].

Session 4: Magnetic Resonance Imaging. Following behavioural testing, participants were invited to a brain imaging session on a Siemens 3T Magnetom-Prisma scanner equipped with a 64-channel head coil, again at

Campus Biotech. Prior to scanning, they filled in and signed an MRI safety questionnaire to again verify their eligibility for the procedure. When required, participants were fitted with MRI-compatible goggles to correct their vision. The language localiser task was administered in Matlab r2021b with Psychtoolbox3, through a computer connected to a screen in the back of the scanner room, that participants could see through a mirror placed on top of the RF head coil. During the whole session, we could observe their eyes through an eye-tracking camera, to check that they were awake. During the imaging session, the field map, resting state sequence and the language localiser were administered in random order, but the resting state sequence was always preceded by the T1-weighted anatomical scan to avoid spurious activations due to carrying out an active task just before. After a short break outside the scanner, participants were repositioned and underwent the DWI sequence and its corresponding field map acquisition. During this session, participants also performed three fMRI tasks and one anatomical scan that were part of different projects. An overview of the imaging session parameters is provided in Table 2.

fMRI language localiser. Participants were instructed to keep their eyes open and look at a black fixation cross on a white background while listening to intact and degraded snippets of the story 'Alice in Wonderland', in their first and second most dominant language of choice (L1, L2)[65,113]. This localiser can be used to inspect individual differences in language activation during quasi-naturalistic listening, and their relationship with behavioural measures of language and cognition[114] (Balboni et al., *article in preparation*). The original localiser paradigm is publicly available from the authors' webpage. As described in Table 2 and the Data Records section, we modified this paradigm to include a degraded L2 condition.

Resting-state fMRI sequence. Brain connectivity at rest can be linked to individual differences in language[115,116] , reading[117,118], and other domains of cognition such as executive function[119,120]. To collect information on resting-state brain activity, we asked participants to lie down with their eyes open, instructing them to relax their body and mind as best as they could, while projecting a white fixation cross on a black background.

*BIDS conversion***.** This dataset conforms to BIDS v1.9.0 and was validated using the command line version of `bids-validator v1.14.6` (http://bids-standard.github.io/bids-validator/). A LINUX Terminal print of the output is provided in the Supplementary Materials (sup_mat.pdf, Figure 1s). The BIDS data format conversion occurred in several steps, part of which could be planned before data collection (such as naming of the MRI sequences, folder structure, and participant codes), while others were performed *post hoc* to adapt data generated by environments not optimised for BIDS. In general, the procedure aimed at having a BIDS-coherent data structure and file names, (re)organising the content of tabular files, adding sidecar files to accompany data and customising code to work in a BIDS folder structure. These steps were performed on all data collected during the same imaging session, and the NEBULA101 data were subsequently imported into the /nebula101 data space, at root level. Nonetheless, for clarity, we provide the specific heuristic for the construction of this dataset after DICOM to NiFTI conversion in

`/code/heudiconv/heuristic.py`. Considering this procedure, all code described is specific to the NEBULA101 dataset but is not meant to be rerun, and is given with paths relative to the /nebula101 root folder. Outside of the code performing data import, no further reference is generally made to external/unavailable files in this code. We describe the steps in Table 3.

Step	*Description*	*Language and environment*	*Custom code*	*Code availability*	*Code location*
1	DICOM to NifTI conversion: conversion was handled in heudiconv via dcm2niix.	Python heudiconv	No	No	n/a
2	File structure heuristic: we set up logging, defined constants, and included functions to create keys for file paths and categorise imaging sequences into different modalities based on their metadata.	Python heudiconv	Yes	Yes	`code/heudiconv/heuristic.py`
3	Modifying language localiser log files for BIDS compatibility.	Python	Yes	Yes	`/code/create_nebula/nebula101_bidsify_mri_logs_aliceloc.py`
4	Importing NEBULA101 data from main BIDS data directory, including behavioural scores in /derivatives.	Python	Yes	Yes	`/code/create_nebula/create_nebula.py`
5	Creation of JSON sidecar files for derivative scores.	Python	Yes	Yes	`/code/create_nebula/create_nebula.py` `/code/create_nebula/create_score_id_json.py.`
6	Modifying field map sidecar "intendedFor" field for BIDS compatibility (to remove references to extra scans) and correcting label naming errors in field maps (fid- to acq-).	Python	Yes	Yes	`/code/create_nebula/create_nebula.py`
7	Importing stimuli from fMRI language localiser to store in BIDS dataset and adjusting their folder structure.	Python	Yes	Yes	`/code/create_nebula/create_nebula.py`
8	Further adjustments to language localiser event.tsv files for BIDS compatibility.	Python	Yes	Yes	`/code/create_nebula/create_nebula.py`
9	Defacing T1w MPRAGE scans.	Python PyDeface	Yes*	Yes	`code/1_anat/0_pydeface/nebula101_loop_pydeface.py` * The (minimal) customisation involved creating a script to run the code on the NEBULA101 sample
10	Creation of dataset-level sidecar file dataset_description.json	Python	Yes	Yes	`/code/create_nebula/create_nebula.py`

Table 3. List of the steps taken to make the dataset BIDS-compliant.

Technical Validation

In compliance with the BIDS indications for mixed raw and derivative datasets, to improve user experience and reduce redundancy, we chose to provide pre-processed behavioural data as derivative tables. We also provide quality control (QC) reports for all imaging modalities, together with the code for reproducing them. Below we specify the details of technical validation for each data type.

Questionnaires

Questionnaire source data were retrieved from Qualtrics XM© and preprocessed in Python. Each questionnaire was cleaned with its own script and then merged with the others via another script. The overall process aimed at data cleaning and score calculation (where necessary):

1) Log file data collection: for each questionnaire, from the source tabular file containing all participants, we eliminated irrelevant metadata. In rare cases, when duplicates were identified we eliminated them at source level by keeping the first non-empty instance of the questionnaire, blind to the contents, to avoid selection bias.
2) Data cleaning and handling of missing data: in all questionnaires, string responses were recoded as numbers. Missing data (NaN) was handled as follows: as a general rule, within-questionnaire, all responses were forced (i.e. progression was halted if a response was missing or on-screen warnings appeared to highlight unfilled responses). When information was indeed unavailable for a participant, we set up options to input specific strings, signalling to us that the information was missing. When missing information was nonetheless identified in a questionnaire with *optional* responses, but without the string signalling unavailability, we evaluated on a case-by-case basis whether it was possible to obtain it from the participant. In such rare cases, the questionnaire was retaken by the participant in session 4, assisted by an experimenter (to avoid data alteration), and only missing responses were filled in. In even rarer cases, when an *entirely* missing questionnaire was identified, participants were invited to complete it through an individual link to that specific questionnaire. Any remaining entries with unavailable information were converted to NaN during data preprocessing. With the strategies we put in place, all questionnaires have been completed by all participants in this dataset, except for one monolingual participant not having taken the code-switching questionnaire due to only speaking one language.
3) LEAPQ cleaning: the LEAPQ, due to its length and structure, required additional, *ad-hoc* handling. In Qualtrics, participants reported the number of languages they knew, including extinct ones and dialects, in order of dominance, that is, how much they used each of them over the others, and in order of acquisition, that is, the temporal order in which they learned their languages. They then provided general information for each language in order of dominance (e.g., names, exposure) and specific information (e.g., learning modality, use context). The resulting data file was restructured by language number and question type, filling gaps with NaNs to create a 'staircase-like' structure where participants (rows) and questions (columns) are listed by language number, in ascending order. The resulting file is available in `/derivatives/nebula101_leapq_data.tsv` with a corresponding sidecar file explaining the column contents. The list of languages in order of acquisition was extracted into a separate file to avoid confusion, as it was not always the case that the dominance and acquisition lists coincided.
   Language order information can be accessed at `/derivatives/nebula_101_leapq_langname_order.tsv`.
4) Raw language names as entered by the participants were conformed to ISO 639-3 and Glottolog codes[121] for standardised reference. The conversion map can be found in `/derivatives/nebula_101_leapq_annotation_iso_glottolog.tsv` and its sidecar file.

5) Multilingual language experience entropy. The LEAP-Q is a widely used instrument for assessing multilingual language experience, encompassing contexts of use, learning, use choices, history with the language, and native-likeness. Given that composite measures of multilingualism derived from the LEAP-Q would be beneficial for quantitative analyses, we calculated four different continuous 'multilingualism scores' for each participant, to reflect their multilingual experience cumulatively[64]. Three scores were based on their self-reported proficiency in speaking, reading, and comprehension across all reported languages, while the fourth score was based on current exposure to all reported languages. Following previous research[18,122,123], each participant's multilingualism score combined proficiency or exposure across their different languages using Shannon's entropy equation[124] within the R entropy package[125].

Behavioural tasks

Behavioural tasks were pre-processed in Python for data cleaning, derivate score calculation and BIDS conversion. The process happened in steps:

1) Log file collection: data were downloaded from Gorilla. Given that participants entered 1 out of 15 possible randomised task sequences, automatically assigned to people by Gorilla, the empty log files from the 14 unused randomisations for each participant had to be deleted. Microphone and headphone check files were identified and renamed.
2) Log file cleaning: irrelevant metadata were eliminated. Task log files were renamed to reflect the actual task (as Gorilla provides them in encrypted form) and to contain BIDS-compliant strings by adding the relevant labels, then converted to tab-separated values. Within each task's log file, column names were renamed to more easily interpretable strings where necessary. Audio files from voice-recorded tasks were renamed as well, for easier reference.
3) Derivate score calculation: for each task, one or more derivate scores were calculated, based on the task's protocol. See Table 2**Error! Reference source not found.**.

These actions were performed in subsequent steps by Python scripts acting on source data. Users will not be required to rerun these scripts since we provide cleaned tabular files containing the relevant scores rather than the raw data, but the code can be made available upon request.

Behavioural data: correlations and reliability. Behavioural scores, whether resulting from tasks or questionnaires (here we generally refer to them together as "behavioural data", and we use "scores" or "measures" when referring specifically to their outcomes), derive from item-level data. To validate these measures, which we provide in BIDS at the *derivative* level, we first explored them via correlations, and then ran reliability analyses in the form of internal consistency coefficients, in a conceptually similar approach to our recent exploratory analysis of behavioural data with a larger cohort[18]. Intercorrelation of the scores obtained from behavioural data (tasks and questionnaires) was minimal, and reliability was acceptable to good in most cases. The code for calculating Pearson's correlations on the z-scored data as well as Cronbach Alpha, along with item-level reliability data and summary output (figures, tables) are available in

`/code/validation/beh/` and in the Supplementary Materials (sup_mat.pdf: Figures 2s and 3s, Tables 1s and 2s).

To better understand the correlation patterns and reliability of these data, some considerations must be made about correlations and about reliability measured by internal consistency. Regarding correlations, on the one hand, high intercorrelations can be informative about shared variance between tests, but on the other hand, having an excessively intercorrelated dataset can lead to problems linked to multicollinearity[126–129]. On a conceptual level, the separability of latent variables facilitates the interpretation of the overall construct. Our choices, when planning this data collection and selecting the tests, aimed at isolating the best measures to represent each variable that we wanted to test in the context of this exploratory project on language aptitude. All things considered, a relative degree of independence of our measures was not only expected, but desirable. Here we report all Pearson pairwise linear correlations having a coefficient of at least $|r(100)| > .5$ and significant at $p < .05$. In `/code/validation/beh/` code is provided to generate Table 5 and a matrix (Supplementary Figure 2s) to visualise these data (`correlate_matrix.py`).

Variable 1	***Variable 2***	***Correlation ($r_{Pearson}$)***
purdue_dh_avg	purdue_ndh_avg	0.66
purdue_dh_avg	purdue_both_avg	0.70
purdue_dh_avg	purdue_dh_ndh_both_avg	0.88
purdue_dh_avg	purdue_assembly_avg	0.67
purdue_ndh_avg	purdue_both_avg	0.78
purdue_ndh_avg	purdue_dh_ndh_both_avg	0.90
purdue_ndh_avg	purdue_assembly_avg	0.62
purdue_both_avg	purdue_dh_ndh_both_avg	0.91
purdue_both_avg	purdue_assembly_avg	0.75
purdue_dh_ndh_both_avg	purdue_assembly_avg	0.76
ran_tot_acc	**regular_acc**	**0.62**
ran_tot_acc	**irregular_acc**	**0.64**
ran_tot_acc	**non_word_rep_acc**	**0.52**
ran_tot_acc	**wordreading**	**0.59**
ran_tot_rt	**reading_text_rt**	**0.59**
ran_tot_rt	**regular_rt**	**0.67**
ran_tot_rt	**irregular_rt**	**0.63**
ran_tot_rt	**pseudo_rt**	**0.59**
reading_text_acc	**regular_rt**	**-0.59**
reading_text_acc	**irregular_rt**	**-0.51**
reading_text_acc	**pseudo_rt**	**-0.61**
reading_text_rt	**spelling_irregular_acc**	**-0.52**
reading_text_rt	**spelling_tot_acc**	**-0.51**
reading_text_rt	**regular_rt**	**0.72**
reading_text_rt	**irregular_rt**	**0.69**
reading_text_rt	**pseudo_rt**	**0.68**
cvlt_tot_imm	cvlt_tot_recall	0.99

spelling_regular_acc	spelling_irregular_acc	0.55
spelling_regular_acc	spelling_tot_acc	0.73
spelling_irregular_acc	spelling_tot_acc	0.83
spelling_pseudo_acc	spelling_tot_acc	0.72
spoon_rt_manual	**phon_suppr_rt_manual**	**0.56**
regular_acc	irregular_acc	0.88
regular_acc	pseudo_acc	0.89
regular_acc	wordreading	0.97
irregular_acc	pseudo_acc	0.75
irregular_acc	wordreading	0.91
pseudo_acc	wordreading	0.95
regular_rt	irregular_rt	0.90
regular_rt	pseudo_rt	0.85
irregular_rt	pseudo_rt	0.77
non_word_rep_span	non_word_rep_acc	0.71
rtt_sub_corr	rtt_sub_rt	-0.91
rtt_sub_corr	rtt_sum_corr	0.82
rtt_sub_corr	rtt_sum_rt	-0.78
rtt_sub_corr	arith	0.96
rtt_sub_rt	rtt_sum_corr	-0.81
rtt_sub_rt	rtt_sum_rt	0.88
rtt_sub_rt	arith	-0.90
finger_tapping_dominant	finger_tapping_nondominant	0.83
finger_tapping_dominant	fingertap	0.96
finger_tapping_nondominant	fingertap	0.95
digit_for_span	span_verbal	0.82
digit_back_span	span_verbal	0.87
rtt_sum_corr	rtt_sum_rt	-0.86
rtt_sum_corr	arith	0.95
rtt_sum_rt	arith	-0.86
corsi_back_span	span_visual	0.76
brocanto_corr1	**cvlt_long_corr**	**0.52**
brocanto_corr1	brocanto	0.73
brocanto_rt1	brocanto_rt2	0.72
brocanto_rt1	brocanto_rt3	0.66
brocanto_corr2	brocanto_corr3	0.64
brocanto_corr2	brocanto	0.74
brocanto_rt2	brocanto_rt3	0.85
brocanto_corr3	brocanto	0.74
mlat5_corr	mlat5_rt	-0.55
cvlt_long_corr	**brocanto**	**0.51**
corsi_for_span	span_visual	0.78
amma_tonal	amma_rhythm	0.96
amma_tonal	amma_total	0.99

amma_tonal	**ind_mus_train_imt**	**0.50**
amma_rhythm	amma_total	0.99
amma_total	**ind_mus_train_imt**	**0.50**
ideal_l2_self	instrumentality	0.62
ideal_l2_self	intl_contact	0.57
ideal_l2_self	l2_interest	0.56
l2_anxiety	l2_confidence	-0.72
swt_score_cs	swt_score_us	0.55

Table 4. Pairwise linear Pearson correlations among variables that reached significance and were above |r(100)| > .5. For brevity, between scores having a 'corr' and 'incorr' (correct ,incorrect) columns, only 'corr' was selected. Age and education were not considered. Correlations across tasks are in bold.

Keeping in mind that correlations are just a preliminary way to explore data, here we see that most significant correlations are between metrics within the same test, and between these and their cumulative score, when present. In bold, we highlighted correlations across tasks: the most evident patterns emerging here are the correlations between reading measures and those typically related to or predictive of reading skill or deficit (all sub-scores from the RAN, text reading, word and pseudoword reading, spelling, spoonerisms, phoneme suppression and non-word repetition); between the Brocanto and CVLT long-term recognition score; between AMMA (total and tonal scores) and the musical training score from the MUSEBAQ (time, intensity and level of practice reached). No other measures were highly and significantly correlated.

Regarding reliability, the main advantage of internal consistency measured via Cronbach alpha is that as long as covariance is inspected and there are minimal to no *nonignorable nonresponses* (MNAR data, i.e. missing not at random), it can be applied quite flexibly, unlike other methods[130]. Minimal reliability requirements are a matter of debate, as the α value can be affected by the number of items and improved by the independence of the construct these tap into[131–133]. This being said, it is harder to obtain a reliable measurement in tasks that measure *spans* (e.g. Corsi blocks, Digit span) and tasks where participants stop after reaching a maximum of answers with high variability (e.g. Revised tempo test): these will generate a MNAR pattern, where each row ends with a certain number of NaN because the span has been reached, and thus will be less suited to internal consistency measures, providing hardly interpretable coefficients[134]. This can explain the somewhat unreliable Revised tempo test RT, reflected in a pattern where negative or null covariances outweigh positive ones in a MNAR fashion (Figure 4), (`/code/validation/beh/cronbach_alpha_results.tsv`).

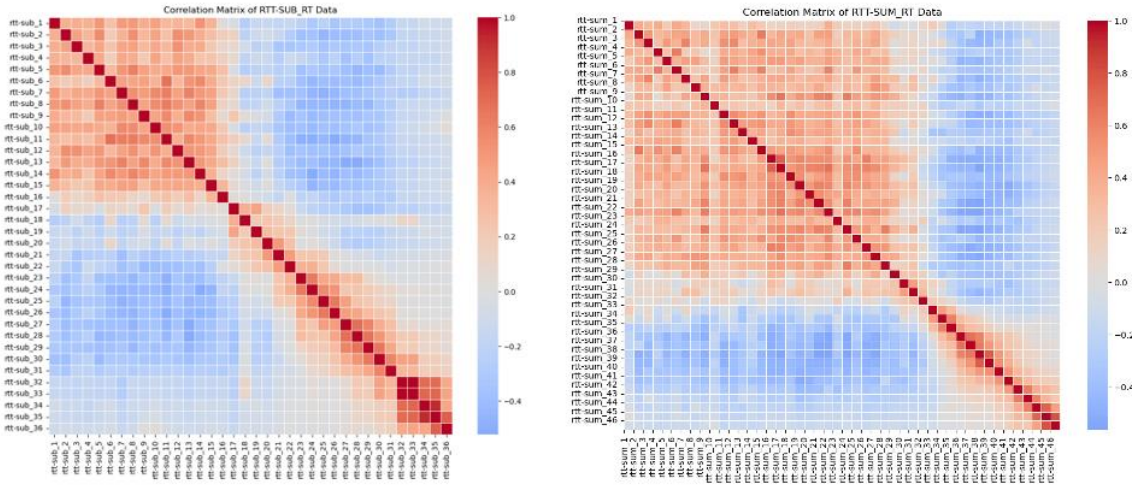

Figure 4 Revised tempo test correlation patterns for subtractions (left) and sums (right) RT data. The different number of items across the 2 figures is due to the fact that 1) participants reached a different number of completed operations across the two conditions and 2) no one reached the end of either condition (60 operations). Knowing that for this task, there is data missing for a reason, this can be interpreted as a MNAR pattern.

Rapid automatised naming accuracy (α = .21) as well as accuracy on the first run of Brocanto (α = .25) also had low reliability. We ascribe the former to the difficulty of the task for the participant, and therefore reaction times may provide more reliable measures than accuracy: RAN measures naming latency[135], whose motor component may show more consistent RTs than accuracy (and this indeed shows in our data, as the two had α coefficients of .93 and .21, respectively). The lower reliability of the first run of Brocanto is conceptually interesting if compared to the subsequent runs and to the RT data: it shows that upon learning an artificial language completely inductively, responses are more variable at the beginning of the learning process, where people tend to guess more frequently, while the time to make a grammaticality judgment is overall always consistent. We provide the correlation matrices for accuracy and RT across the 3 blocks of Brocanto in Figure 5. In a few cases, data were recorded at the item level but digitised at the task level (spoonerisms, non-word repetition and phoneme suppression): item-level data are available upon request. Table 5 shows the internal consistency values for tasks where $\alpha > .5$, and tasks with $\alpha > .6$ are additionally highlighted in bold. Complete reliability data are available in `/code/validation/beh/` and can be regenerated with `calculate_cronbach_alpha.py`. Descriptive statistics and plots are available in the Supplementary Materials (sup_mat.pdf: Table 1s, 2s; questionnaires_violin_plots.pdf; entropy_violin_plots.pdf; tasks_violin_plots.pdf).

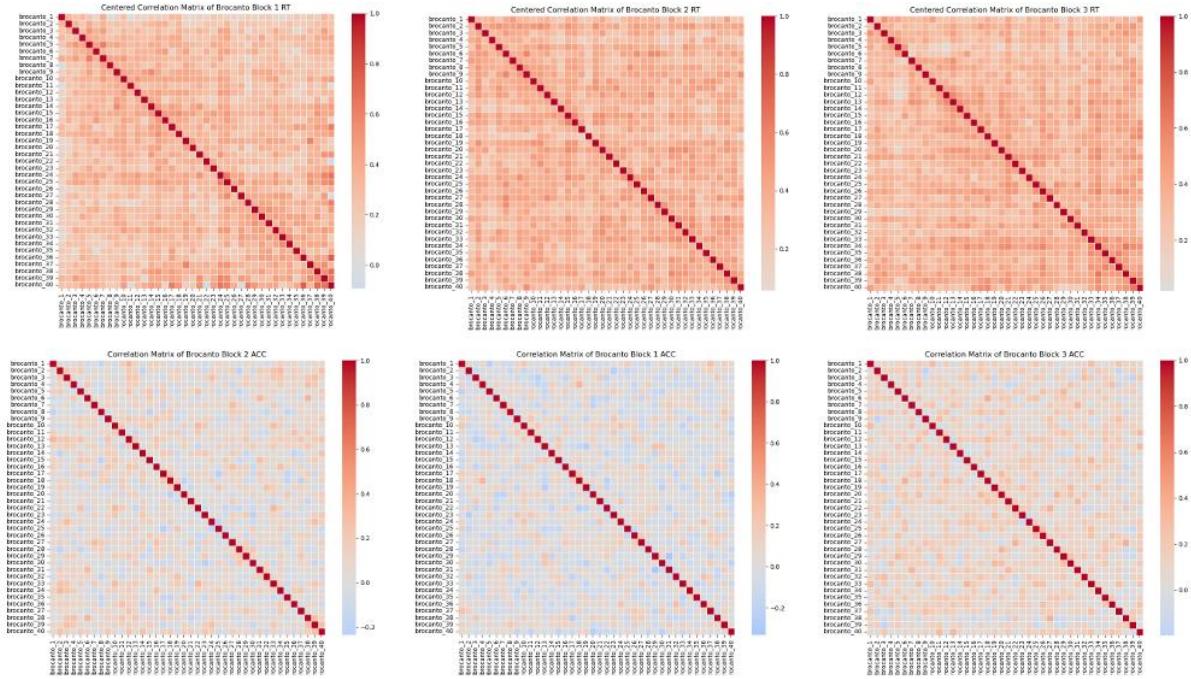

Figure 5 Top row: correlation matrix for the RT data of Brocanto across 3 blocks. Bottom row: accuracy data for Brocanto across 3 blocks. Plots show that overall, RTs were more reliable than was accuracy.

The Farsi uvular sound production task was rated by two first-language Farsi speakers who heard each recorded utterance from participants and gave it a 'native-like production score': therefore, any internal consistency metric would reflect the way the raters scored the task, more than the way the participant performed it (even though these are clearly related). Thus, for this task, we chose to calculate inter-rater reliability between the first and second rater, in a procedure identical to a previous study where this task was used[136]. For the ANT-I task, given its structure (each trial reflecting more than one possible condition from which a score can be derived), we chose to measure the maximal split-half coefficient to assess reliability[137] (Supplementary Materials: ant_split-halfcorr.html and .Rmd).

Test	*Cronbach alpha*	*95% CI lower bound*	*95% CI upper bound*
artgram_corr	0.54	0.39	0.66
corsi-for	0.55	0.42	0.67
ind_mus_train_imt	0.57	0.40	0.70
digit-for	0.58	0.44	0.69
digit-back	**0.61**	0.49	0.71
spelling	**0.62**	0.46	0.73
amma	**0.65**	0.54	0.74
instrumentality	**0.65**	0.54	0.75
irq_manip	**0.65**	0.54	0.75
irq_orto	**0.66**	0.55	0.75
ind_mus_listening_iml	**0.66**	0.50	0.77
physical_exercise	**0.67**	0.53	0.76
l2_interest	**0.67**	0.55	0.76

swt_US	**0.68**	0.56	0.78
l2_anxiety	**0.69**	0.57	0.78
brocanto_corr_2	**0.70**	0.61	0.78
hindi	**0.70**	0.61	0.78
ideal_l2_self	**0.71**	0.61	0.79
cognitive_emotional_regulation	**0.72**	0.63	0.80
apm_corr	**0.72**	0.64	0.79
reading_acc	**0.72**	0.59	0.81
milieu	**0.76**	0.68	0.83
l2_confidence	**0.77**	0.69	0.84
irq_verbal	**0.78**	0.71	0.84
intl_contact	**0.78**	0.70	0.84
brocanto_corr_3	**0.80**	0.74	0.85
irq_visual	**0.81**	0.75	0.86
ahrq	**0.81**	0.75	0.86
cvlt_reco_corr	**0.81**	0.76	0.86
cvlt_reco_rt	**0.82**	0.76	0.87
mlat5_corr	**0.82**	0.77	0.87
mlat5_incorr	**0.82**	0.77	0.87
purdue_NDH	**0.83**	0.77	0.88
artgram_rt	**0.84**	0.79	0.88
social_connection	**0.84**	0.77	0.89
rtt-sum_corr	**0.84**	0.80	0.89
usage_willingness	**0.85**	0.80	0.89
mlat5_rt	**0.85**	0.81	0.89
purdue_DH	**0.85**	0.80	0.90
rtt-sub_corr	**0.86**	0.82	0.90
cvlt_tot-imm	**0.88**	0.84	0.91
apm_rt	**0.88**	0.85	0.91
swt_CS	**0.89**	0.84	0.92
purdue_both	**0.89**	0.84	0.92
dance	**0.90**	0.85	0.93
cvlt_tot-recall	**0.90**	0.87	0.93
engaged_production	**0.92**	0.89	0.94
purdue_assembly	**0.93**	0.90	0.95
reading_rt	**0.93**	0.89	0.95
ran_rt	**0.93**	0.91	0.95
brocanto_rt_1	**0.95**	0.93	0.96
brocanto_rt_3	**0.96**	0.94	0.97
brocanto_rt_2	**0.96**	0.95	0.97
finger-tapping_dominant	**0.97**	0.95	0.98
finger-tapping_nondominant	**0.98**	0.97	0.99
ANT-I alerting	0.51*, CI 95% 0.32, 0.66		
ANT-I orienting	**0.61*, CI 95% 0.46, 0.73**		

ANT-I re-orienting	**0.65*, CI 95% 0.51, 0.76**
ANT-I inhibition	**0.67*, CI 95%, 0.55, 0.77**
Farsi uvular production	***Inter-rater reliability:*** $r_{pearson}$**(95): 0.68** $p < .0001$

Table 5. Internal consistency values measured via Cronbach alpha, where $\alpha >= .5$. Lower and upper bounds are reported at a 95% confidence interval. For clarity, values where $\alpha = > .6$ are in bold.
** Values refer to Spearman-Brown correlation corrected reliability estimates (calculated with split-half correlations, 1000 random splits).*

Brain imaging

Anatomical imaging anonymisation, QC and brain extraction. Facial features in anatomical MRI scans violate the principle of anonymity in open data. Therefore, T1-weighted MPRAGE scans were defaced with PyDeface (https://pypi.org/project/pydeface/) and their quality was assessed in the Computational Anatomy Toolbox for SPM (hereon, CAT12)[138]. Specifically, we used the CAT12 segmentation and sample homogeneity toolboxes, providing easily interpretable quality measures at the participant and group level. The weighted overall image quality (IQR) and the quartic mean Z-score are the two key indicators of image quality. IQR combines noise and spatial resolution measurements before pre-processing, while the mean quartic Z-score assesses the homogeneity of data after pre-processing, with deviations increasing variance and reducing statistical power. The *product* of IQR and quartic mean Z-scores combines these quality measures, with a low number indicating high quality. For each participant, we provide a PDF with the CAT12 report, as well as information on sample homogeneity, in `/code/validation/anat/cat12/`. The CAT12 toolbox can be easily run via a graphical user interface (GUI) in Matlab and requires no custom code. We provide a representative participant report in Figure 6, and the group histogram in Figure 7.

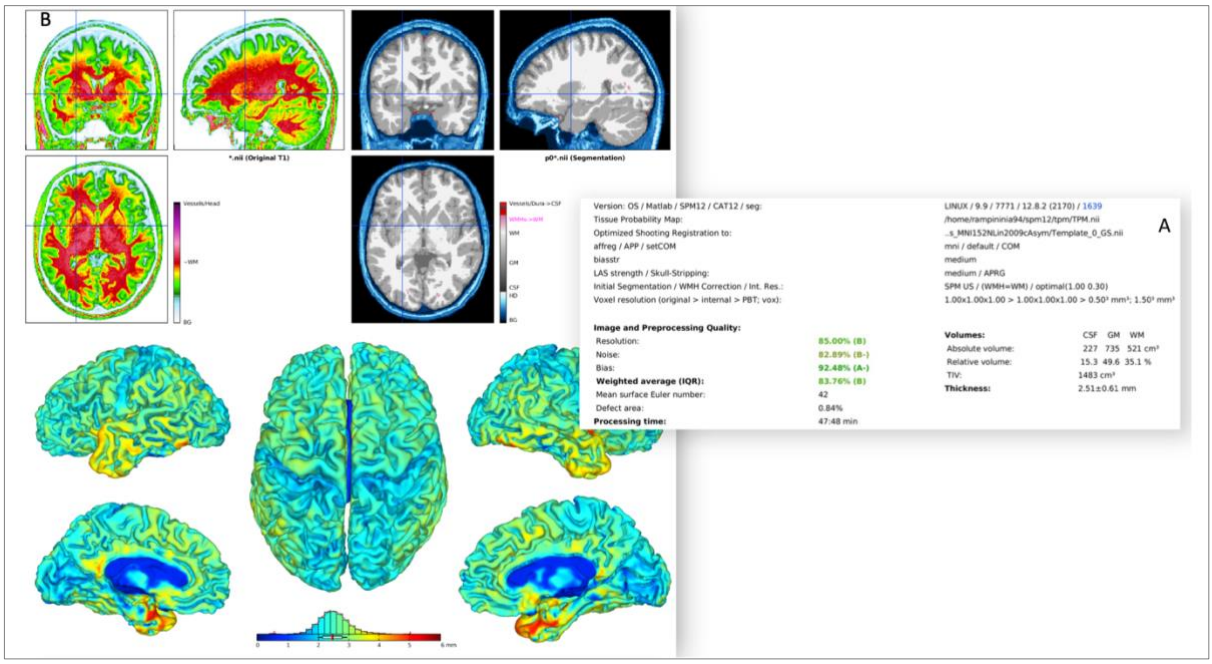

Figure 6. cat12 report for sub-pp171. Panel A: main report, panel B (left) pre-segmentation and (right and bottom) post-segmentation tissue distribution and cortical thickness estimation.

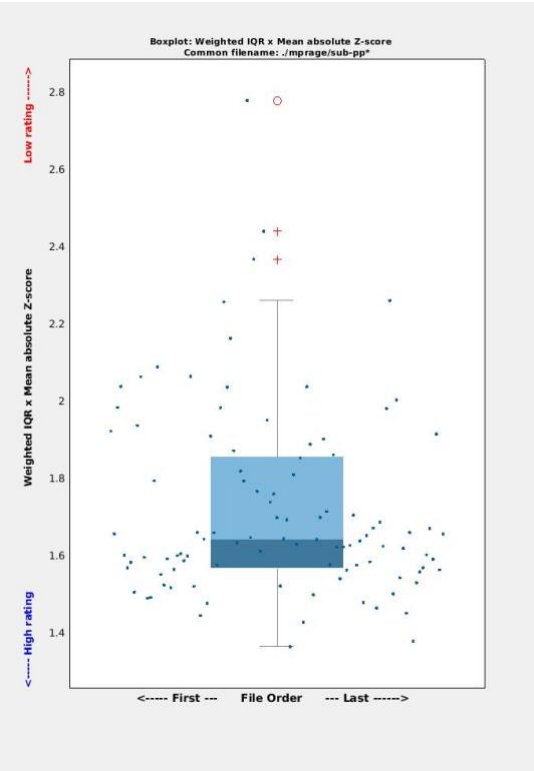

Figure 7. CAT12 output. Boxplot showing the weighted IQR by mean absolute z-score scaled by a factor of 4 (i.e. quartic) to emphasize outliers. As the plot shows, all but 1 participant in this distribution lie in the good to optimal range of the quality measure. Of note, even when outliers are detected, these are not necessarily data to discard on an absolute basis, as the score is calculated relatively to the specific sample being analysed.

To prepare T1 MPRAGE anatomical scans to be fed to FSL[139] (see *Functional neuroimaging* section), we provide code for running an improved skull-stripping as a loop in the NEBULA101 dataset in `/code/preprocessing/1_anat/1_optiBET/nebula101_run_optiBET.sh`. This code will process T1-weighted MRI images stored in a BIDS directory to perform improved brain extraction with optiBET[136] via the following actions:

1) Iterating through all participant directories in the BIDS directory.
2) For each participant, creating an output directory in the derivatives folder if it doesn't exist.
3) Copying the T1-weighted images (`sub-pp*_ses-01_T1w.nii.gz`) to the corresponding output directory.
4) Checking if each T1-weighted image has already been processed.
5) If an image has not been processed, running the `nebula101_optiBET.sh` master script on the image, which performs actual brain extraction.

The script referenced in step 5 is the master optiBET script, which needs to be downloaded from https://montilab.psych.ucla.edu/fmri-wiki/optibet/ and stored in the location where the loop script is stored.

Field maps. We acquired field maps to accompany the DWI and fMRI sequences. Field map preparation consists of a few steps that are partly scanner-specific, and in the case of Siemens Prisma, consisted in recomposing the phase and magnitude images. We provide field map preprocessing code in `/code/preprocessing/2_fmap/fsl/1_nebula101_run_fmap_prep_all.sh`

The script will perform the following steps:

1) Transformation Calculation: calculating the transformation matrix from the anatomical T1 image to the magnitude image and applying the calculated transformation to the T1 brain mask to align it with the magnitude image (`FLIRT`).
2) Magnitude Image Masking: using the transformed brain mask to mask the magnitude image, creating an untrimmed brain image (`fslmaths`).
3) Brain Mask Trimming (`fslmaths`):
    a. Binarising the untrimmed mask.
    b. Smoothing the mask with an 8 mm kernel.
    c. Thresholding the smoothed mask at 75%.
    d. Binarising the thresholded mask.
    e. Trimming the original brain image using the final binary mask.
    f. Removing intermediate untrimmed files.
4) Preparing the final field map (`fsl_prepare_fieldmap`): preparing the final field map using the processed magnitude and phase difference images, with a specified echo spacing of 2.46 ms.

Diffusion-weighted imaging. To validate our diffusion data, we run the `QUAD` (participant) and `SQUAD` (group) programs within the `FSL-FDT`[139,140] suite as part of the diffusion pre-processing pipeline. The DWI data and prepared field map of each participant were fed to `EDDY` to correct for susceptibility, eddy currents, inter- and intra-volume participant movement and signal dropout, using the prepared field map via the `--field` flag. Finally, we ran `eddyQC`[141] for quality assessment. We provide a pdf with QC statistics for each NEBULA101 participant and for the group, as well as the code to reproduce this procedure, in `/code/validation/dwi/fsl/dwi_qc_quad_squad.py`. The code is configured to automatically grab the acquisition parameters and the correct number of shells from the DWI sidecars to generate the `acqparam.txt` and `index.txt` files required by `EDDY`, and will raise flags if the expected numbers do not match with the data. An `eddy_quad_qc_paths.txt` file has also been set up to this aim in the same location, which the user can alter for their needs, should they decide to preprocess DWI data (which we provide in raw form) with the same pipeline (`/code/preprocessing/dwi/fsl/1_dwi_prep.py`). Figure 8 reports the signal- and contrast-to-noise ratios from the group QC metrics generated using `SQUAD`. The complete PDF report is available in `code/validation/dwi/fsl/squad/group_qc.pdf`.

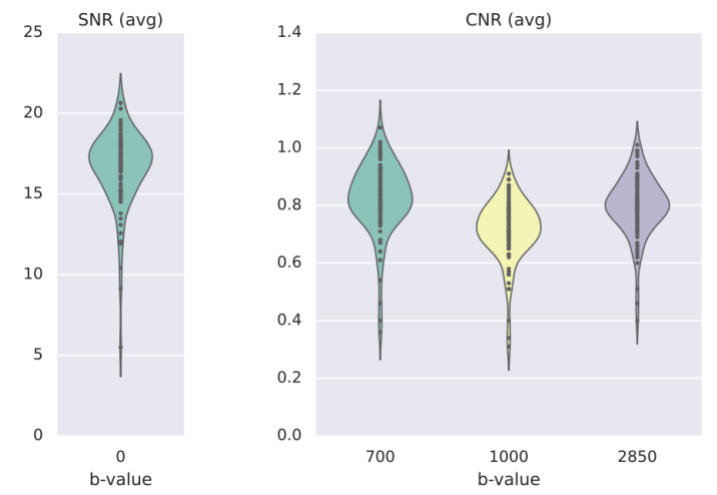

Figure 8. Signal-to-noise and contrast-to-noise ratios from the QC-ed DWI data.

Functional imaging. As concerns fMRI, we provide QC metrics obtained from preprocessing of the whole dataset in FSL, and the code to reproduce the procedure on the NEBULA101 sample.

Figure 9 and Figure 10 report violin plots of the absolute and relative displacement by run for the language localiser and for the resting state fMRI sequence, respectively. Absolute displacement reflects the movement with respect to the reference volume, while relative displacement reports volume-to-volume information. Both can impact SNR but while absolute displacement can be counteracted, relative displacement is usually more harmful, and both must be checked when assessing the quality of fMRI data[142]. Displacement information and plots can be obtained by running `code/validation/func/fsl/1_copy_motion_params.py` and `/code/validation/func/fsl/2_mean_abs_rel_disp_violin_py`. In principle, if the data are not being reprocessed, `1_copy_motion_params` does not need to be rerun as we provide `.rms` output from FSL, containing absolute and relative displacement data for each participant, in `/code/validation/func/fsl/aliceloc/` and `/rest`.

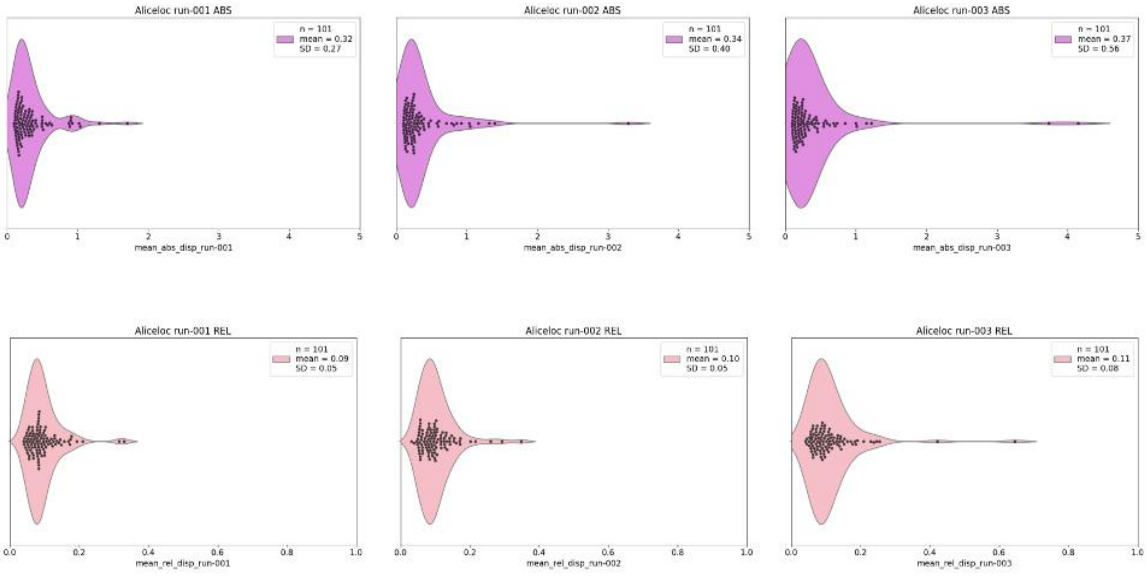

Figure 9. Absolute (ABS) and relative (REL) displacement (mm) across runs of the language localiser.

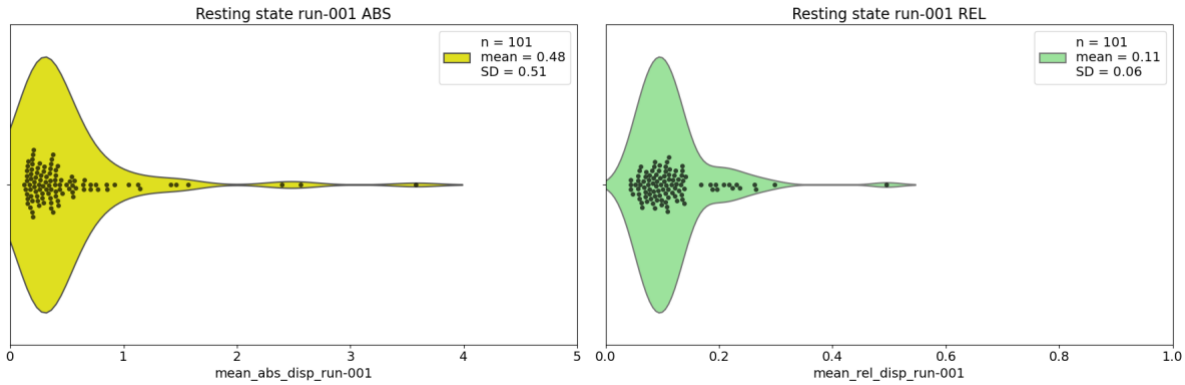

Figure 10. Absolute (ABS) and relative (REL) displacement (mm) of resting state fMRI.

A folder containing the code for performing the basic preprocessing steps of fMRI for the language localiser and resting state sequences has been set up in `/code/preprocessing/3_func/fsl/1_preproc1/1_loop_fsl_preproc1.py`. This code performs the first step of fMRI preprocessing with FSL FEAT, consisting of:

1) Motion correction

2) 4D mean intensity normalization

3) Spatial smoothing (5 mm FWHM)

4) B0 unwarping (includes `BBR reg`)

5) Registration (`BBR, FNIRT`)

This code will run using `/code/preprocessing/3_func/fsl/1_preproc1/design_pp.fsf`, a design file required by FSL which has been set up to contain information on the preprocessing loop, should the user wish to preprocess our data with the same pipeline.

Overall validation of data structure. This dataset is mostly complete, with minimal missing information. However, to provide the user with information that is traceable without necessarily having this data descriptor at hand, we have created a series of scripts within the `/code/data_checks/` folder aimed at providing report logs on the presence of behavioural and brain imaging data.

- `nebula101_bids_report_beh.py`: this script plots behavioural data presence checks and compiles the findings to a report log called `nebula101_beh_report_[timestamp].txt`. It will also create:
    - A tabular file `nebula101_datalist_[date].tsv` listing data as 1 and 0 if present or absent, for easy reference.
    - A heatmap with the above information, called `data_presence_heatmap_[timestamp].png`.
    - A report specific to the LEAPQ data, to have missing and present information at a glance, called `nebula101_leapq_data_report_[timestamp].txt`
- `nebula101_bids_report_mri.py`: this script plots imaging data presence checks and compiles the findings to a report log called `nebula101_mri_report_[timestamp].txt`. This can be used for easy referencing to participants having undergone different conditions (e.g. different field map, see: Usage Notes).

Data Records

In this section we describe the data structure and its contents.

The root-level BIDS folder `/nebula101` contains 101 participant folders with raw imaging data denoted by the code `sub-pp` followed by three digits, code, derivatives, stimulus folders, and the mandatory dataset description and participants file required by BIDS (Figure 11). Inconsistent numbering is due to non-included participants (see Participants section), the participant who denied consent to share their data, and the two participants who could not undergo brain imaging).

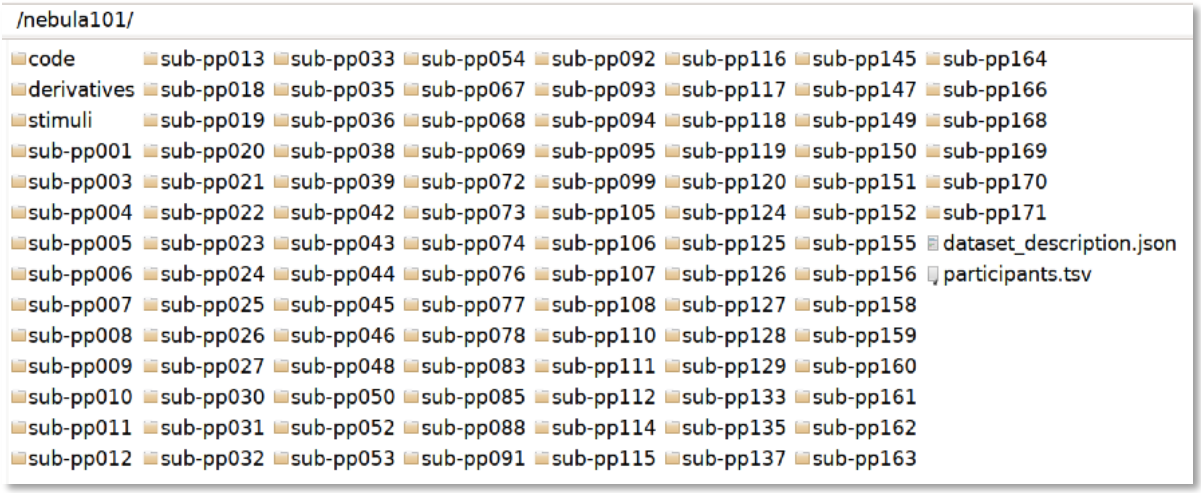

Figure 11. Root dataset folder.

The `/code` folder in Figure 12 contains subfolders with 1) fMRI paradigm data for the language localiser; 2) code to import, modify or generate the BIDS-compliant files as described in Table 3; 3) the BIDS conversion heuristic; 4) the preprocessing information and code folder; 5) the validation materials folder.

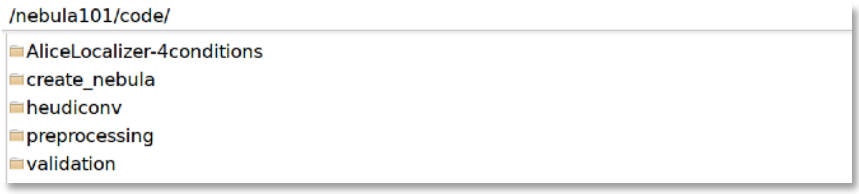

Figure 12. code folder.

The `/derivatives` folder in Figure 13 contains the preprocessed behavioural data and their sidecar files.

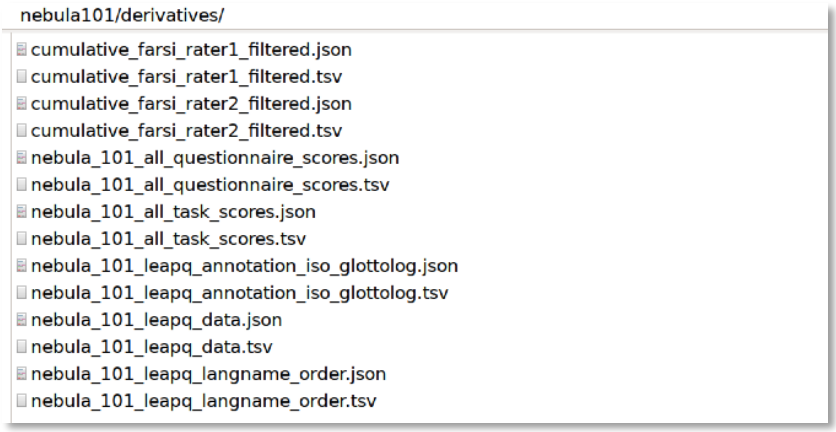

Figure 13. Derivatives folder.

- cumulative_farsi_rater* = derivative scores from the Farsi task ratings and their sidecar files.
- nebula_101_all_questionnaire_scores* = scores from questionnaires and their sidecar file.
- nebula_101_all_task_scores* = scores from tasks and their sidecar file.
- nebula_101_leapq_annotation_iso_glottolog* = mapping of language names to ISO an Glottolog and its sidecar file.
- nebula_101_leapq_data* = LEAPQ data and their sidecar file.

- nebula_101_leapq_langname_order* = LEAPQ languages in order of acquisition and their sidecar file.

The `/stimuli` folder contains a subfolder where the language localiser materials are stored and can be referenced by the BIDS validator.

The participant folders, named `sub-pp*/ses-01/` contain `/anat`, `/dwi`, `/fmap` and `/func` folders. These host the raw imaging data of each participant. An example with `sub-pp001/ses-01/` is shown in Figure 14, Figure 15, Figure 16 and Figure 17.

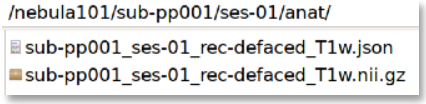

Figure 14. defaced T1 MPRAGE data and their sidecar file.

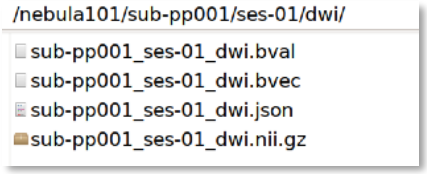

Figure 15. DWI data: beta values, vectors, and sequence data, with its sidecar file.

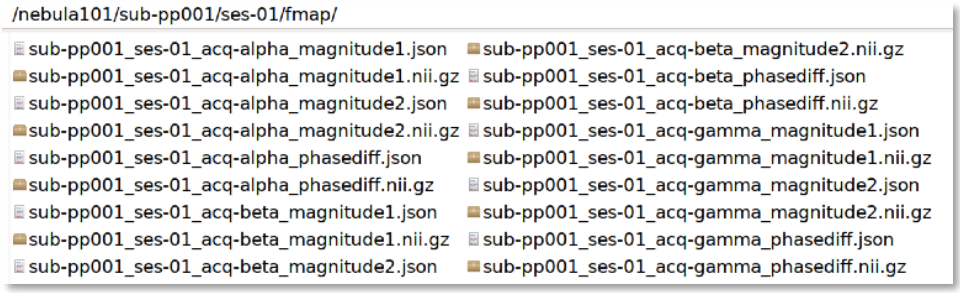

Figure 16. Field maps: phase and magnitude images with their sidecars. This participant has an extra field map (gamma), as explained in the Usage Notes.

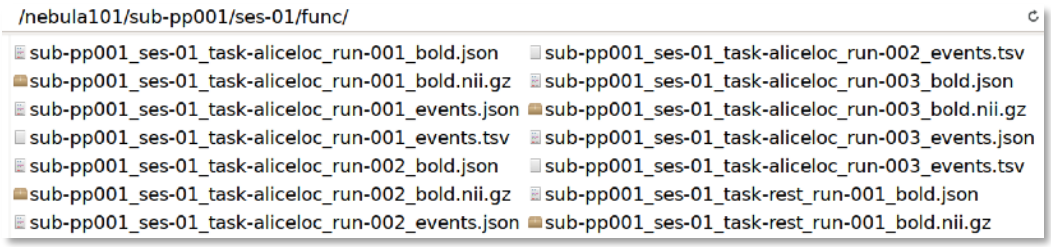

Figure 17. Functional MRI data with their sidecar files and event files. Rest refers to resting state fMRI, aliceloc refers to the language localiser.

The `/code/AliceLocalizer-4conditions/` folder contains our own version of the materials provided in the public domain by the creators of the task[65,113], which we modified to suit our needs, as described in the fMRI language localiser description (see: Data collection). This version of the code can read the stimuli from the structure of the BIDS folder. It will still require manual intervention to create the `events.tsv` files from the Matlab log files, as Matlab currently provides a different tabular structure. We

remind the reader that we altered the structure of the log files to create BIDS events in `/code/create_nebula/ bidsify_mri_logs_aliceloc.py`, and subsequently refined the resulting tabular files in `/code/create_nebula/create_nebula.py` for the `BIDS` validator to be able to reference them.

The `/code/preprocessing` folder includes numerically ordered subfolders containing code to reprocess the anatomical, field map, functional and diffusion data (see Technical Validation). Any steps contained in these subfolders are meant to be run sequentially to recreate the materials needed for technical validation, or if the user decides to preprocess our raw data using the same pipelines.
`1_anat` contains the PyDeface loop and optiBET brain extraction code.
`2_fmap` contains code for field map preparation.
`3_func` contains code for minimal fMRI preprocessing, to get the displacement parameters used in the technical validation.
`4_dwi` contains the preprocessing code for the diffusion data necessary to obtain the technical validation reports via `eddy_qc`.

The `/code/validation` folder contains the following subfolders:
`/anat`: CAT12 segmentation and sample homogeneity reports for each participant.
`/beh`: reliability data for each task at item level, and the code to run reliability analysis:

- `calculate_cronbach_alpha.py`
- Three additional Python notebooks for generating the correlation data presented here and additional tables and matrices contained in the Supplementary Materials (Table 1s, Table 2s and Figure 2s):
    - `correlate_matrix.py`
    - `calculate_descriptives.py`
    - `farsi_inter_rater.py`

`/data_checks`: various Python scripts to create data lists of behavioural and imaging data, and missing data heatmaps, as explained in Technical Validation.
`/dwi`: individual participant `QUAD` pdf reports, the code for generating them, and the group folder from `SQUAD`.
`/func`: two Python scripts, one for importing displacement `.rms` data, and the other for generating the violin plots shown in Figure 9 and Figure 10.

`1_copy_motion_params.py`
`2_mean_abs_rel_disp_violin.py`

This folder also contains `/aliceloc` and `/rest`. Here, the motion parameters imported by the first script are stored for each fMRI modality.
All Python notebooks and executables are heavily commented for user-friendliness.

Usage Notes

Behavioural data. This dataset is mostly complete, but a few technical malfunctions during testing have caused some minimal missing data, as described in Table 6. This information and a heatmap for visualising missing data at a glance can easily be regenerated using the scripts in `/code/validation/data_checks/` called `nebula101_bids_report_beh.py` and `nebula101_missing_data_log.py`.

Participant ID	*Columns with missing data at task level*
sub-pp031	phon_suppr_rt_manual
sub-pp050	brocanto_corr1, brocanto_incorr1, brocanto_rt1, cvlt_long_corr, cvlt_long_incorr, cvlt_long_rt, brocanto
sub-pp116	cvlt_long_rt
sub-pp118	apm_corr, apm_incorr, apm_rt
sub-pp155	non_word_rep_span, non_word_rep_acc

Table 6 Missing behavioural data.

Imaging data. Here we provide some details to facilitate easier navigation of the imaging data. This information can be generated by `/code/validation/data_checks/nebula101_bids_report_mri.py`. JSON sidecar files have already been adjusted to report the below nuances:

- `sub-pp161`: for this participant, the `beta` field map must be used for resting state fMRI processing, and the `delta` field map must be used for DWI processing.
- `sub-pp001, sub-pp010, sub-pp013, sub-pp023, sub-pp043, sub-pp053`: for these participants, the `gamma` field map must be used for resting state fMRI processing.

When running the BIDS validator, this information is reported as *warnings* and saved in the log, unless the `--ignoreWarnings` flag is raised. Warnings to not impede BIDS validation and the dataset will pass it every time. When processing the data, this information will be read automatically by software that can work primarily in BIDS (such as fMRIprep[143] in the case of functional imaging). When using software that lacks this functionality, such as FSL, we recommend processing these participants manually.

Data Availability

Upon publication, the dataset will be publicly accessible and downloadable with no restrictions under a Creative Commons CC0 1.0 (Universal Public Domain Dedication) license from Yareta, the institutional re3-registered repository of the University of Geneva. In all Creative Commons licenses, one essential requirement is attribution, which involves acknowledging the authors by citing the source, and acknowledging the license. Attribution should reasonably encompass all pertinent details provided by the licensor. All participants signed a waiver for anonymised data release in the public domain, and no identifying information was retained, to the best of our knowledge.

Code Availability

All code that can be rerun on these data is provided within the `/code` folder, as described in the Technical Validation section. Given that we do not provide *source* behavioural data, preprocessing code for tasks and questionnaires is not included, but can be shared upon request.

Funding

This work was supported by the Swiss National Science Foundation [Grant #100014_182381], and by the NCCR Evolving Language, Swiss National Science Foundation [Agreement #51NF40_180888].

Author Contributions

CRediT (Contributor Roles Taxonomy) is provided, as follows: Conceptualization: AR IB OK RB NG; Investigation: AR IB; Methodology and Formal Analysis: AR IB OK; Data Curation: AR IB; Technical Validation: AR; Writing - Original Draft AR; Writing - Review & Editing: AR IB OK RB NG; Visualization AR; Resources: RB NG; Supervision: RB NG; Funding acquisition: RB NG.
AR and IB (*) contributed equally to this work.

Competing Interests

The authors declare no competing interests.

Acknowledgments

We would like to express our sincere gratitude to several collaborators who contributed their valuable resources and support throughout this large and intensive project.
We gratefully acknowledge the following people: Roberto Martuzzi, Loan Mattera and Nathalie Philippe from the Human Neuroscience Platform at Campus Biotech for their invaluable and continued assistance with MRI setup and data collection; we thank our student assistants Vanessa Gottofrey, Melody Cascioli, and Aspasia Sfakaki for their contribution to data collection, and Rixa Gruhnert for helping with the LEAPQ annotations.
We are thankful to Michael Dayan for his help with setting up the BIDS workflow; to Priscilla Borges, Jutta Mueller and Hettie Roebuck for discussions on the Modes of Internal Reasoning Questionnaire use and its scoring; to Mark Eckert and Davide Fedeli for their helpful hints on technical validation of anatomical and diffusion-weighted MRI; to Gwendoline Mahé for providing us with the French version of the Adult Reading History Questionnaire; to Peter Schneider for consultations on the musical experience and musicality tests; to Evelina Fedorenko and her team for discussing with us the modifications to the Alice in Wonderland Localiser.
We also thank Isabelle Udry for translating the Motivation Factor Questionnaire to French; Neiloufar Family and our colleague Sevil Maghsadagh for rating the Farsi task.
Finally, we thankfully acknowledge the collaborators who gave us access to and information on materials they have developed themselves: Tobias Kober and Siemens Healthineers© for the WIP-Compressed

Sensing MPRAGE sequence, which crucially shortened MRI time; Will Barratt for sharing his Socioeconomic Status Measure; Franck Ramus and Steve Majerus for sharing their literacy and phonological awareness tasks; Charles Stansfield for granting permission to use the MLAT5; Elsje Van Bergen for providing the Tempo Test Revised.

Lastly, our heartfelt gratitude goes to the 101 participants who participated in this data collection, for their generosity with their time and commitment to complete all procedures.

References

1. Kidd, E., Donnelly, S. & Christiansen, M. H. Individual Differences in Language Acquisition and Processing. *Trends Cogn Sci* **22**, 154–169 (2018).
2. Vogel, E. K. & Awh, E. *How to Exploit Diversity for Scientific Gain Using Individual Differences to Constrain Cognitive Theory*. (2008).
3. Carroll, J. B. & Sapon, S. M. *Modern Language Aptitude Test. Modern language aptitude test.* (Psychological Corporation, San Antonio, TX, US, 1959).
4. Carroll, J. B. Twenty-five years of research on foreign language aptitude. in *Individual differences and universals in language learning aptitude* 867–873 (1981).
5. Stansfield, C. W. & Reed, D. J. The Story Behind the Modern Language Aptitude Test: An Interview With John B. Carroll (1916-2003). *Lang Assess Q* **1**, 43–56 (2004).
6. Miller, G. A. The cognitive revolution: a historical perspective. *Trends Cogn Sci* **7**, 141–144 (2003).
7. Li, S. The Associations between Language Aptitude and Second Language Grammar Acquisition: A Meta-Analytic Review of Five Decades of Research. *Appl Linguist* **36**, 385–408 (2015).
8. Harré, R. The Second Cognitive Revolution. in *After cognitivism* (ed. Leidlmair, K.) 182–187 (Springer, Dodrecht, 2009).
9. Robinson, P. Individual differences, aptitude complexes, SLA processes, and aptitude test development. *Second Language Learning and Teaching* **4**, 57–75 (2012).
10. Robinson, P. Aptitude and second language acquisition. *Annu Rev Appl Linguist* **25**, 46–73 (2005).
11. Wen, Z., Biedroń, A. & Skehan, P. *Foreign Language Aptitude Theory: Yesterday, Today and Tomorrow*. *Language Teaching* vol. 50 (2017).
12. Eisenstein, M. Childhood bilingualism and adult language learning aptitude. *International Review of Applied Psychology* **29**, 159–172 (1980).
13. Grigorenko, E. L., Sternberg, R. J. & Ehrman, M. E. A theory-based approach to the measurement of foreign language learning ability: The canal-F theory and test. *Modern Language Journal* **84**, 390–405 (2000).
14. Sparks, R. L., Ganschow, L. & Patton, J. Prediction of Performance in First-Year Foreign Language Courses: Connections Between Native and Foreign Language Learning. *J Educ Psychol* **87**, 638–655 (1995).
15. Harley, B. & Hart, D. Language aptitude and second language proficiency in classroom learners of different starting ages. *Stud Second Lang Acquis* **19**, 379–400 (1997).

16. Sawyer, M. *Language Aptitude and Language Experience: Are They Related?* https://iuj.repo.nii.ac.jp/records/583 (1992).
17. Dąbrowska, E. Experience, aptitude and individual differences in native language ultimate attainment. *Cognition* **178**, 222–235 (2018).
18. Rampinini, A., Balboni, I., Golestani, N. & Berthele, R. A behavioural exploration of language aptitude and experience, cognition and more using Graph Analysis. *Brain Res* **1842**, 149109 (2024).
19. Udry, I. & Berthele, R. The smart, the motivated and the self-confident: The role of language aptitude, cognition, and affective variables in early instructed foreign language learning. in *Individual Differences in Early Instructed Language Learning: The Role of Language Aptitude, Cognition, and Motivation* (eds. Berthele, R. & Udry, I.) 71–90 (Language Science Press, Berlin, 2021).
20. Turker, S., Seither-Preisler, A. & Reiterer, S. M. Examining Individual Differences in Language Learning: A Neurocognitive Model of Language Aptitude. *Neurobiology of Language* **2**, 389–415 (2021).
21. Van Der Maas, H. L. J. *et al.* A Dynamical Model of General Intelligence: The Positive Manifold of Intelligence by Mutualism. (2006) doi:10.1037/0033-295X.113.4.842.
22. Pluck, G. & Cerone, A. A Demonstration of The Positive Manifold of Cognitive Test Inter-correlations, and how it Relates to General Intelligence, Modularity, and Lexical Knowledge. *UC Merced Proceedings of the Annual Meeting of the Cognitive Science Society* (2021).
23. Kong, X. Z. *et al.* Gene Expression Correlates of the Cortical Network Underlying Sentence Processing. *Neurobiology of Language* **1**, 77–103 (2020).
24. Amelink, J. S. *et al.* Imaging genetics of language network functional connectivity reveals links with language-related abilities, dyslexia and handedness. Preprint at https://doi.org/10.1101/2023.11.22.568256 (2024).
25. Nayak, S. *et al.* The Musical Abilities, Pleiotropy, Language, and Environment (MAPLE) Framework for Understanding Musicality-Language Links Across the Lifespan. *Neurobiology of Language* **3**, 615–664 (2022).
26. Steyer, R., Mayer, A., Geiser, C. & Cole, D. A. A Theory of States and Traits-Revised. (2014) doi:10.1146/annurev-clinpsy-032813-153719.
27. Fuchs, E. & Flügge, G. Adult Neuroplasticity: More Than 40 Years of Research. *Neural Plast* **2014**, 541870 (2014).
28. Dityatev, A. & Schachner, M. Extracellular matrix molecules and synaptic plasticity. *Nature Reviews Neuroscience 2003 4:6* **4**, 456–468 (2003).
29. Ramirez-Amaya, V. Molecular Mechanisms of Synaptic Plasticity Underlying Long-Term Memory Formation. *Neural Plasticity and Memory: From Genes to Brain Imaging* 47–66 (2007) doi:10.1201/9781420008418.ch3.
30. Grafman, J. Conceptualizing functional neuroplasticity. *J Commun Disord* **33**, 345–356 (2000).
31. Wang, R. *et al.* Functional and structural neuroplasticity associated with second language proficiency: An MRI study of Chinese-English bilinguals. *J Neurolinguistics* **56**, 100940 (2020).

32. DeLuca, V., Rothman, J., Bialystok, E. & Pliatsikas, C. Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proc Natl Acad Sci U S A* **116**, 7565–7574 (2019).
33. Pliatsikas, C. Multilingualism and Brain Plasticity. *The Handbook of the Neuroscience of Multilingualism* 230–251 (2019) doi:10.1002/9781119387725.CH11.
34. Li, P., Legault, J. & Litcofsky, K. A. Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex* **58**, 301–324 (2014).
35. DeFelipe, J. Brain plasticity and mental processes: Cajal again. *Nat Rev Neurosci* **7**, 811–817 (2006).
36. Grasby, K. L. et al. The genetic architecture of the human cerebral cortex. *Science (1979)* **367**, 6484 (2020).
37. Navarri, X. *et al.* A biologically informed polygenic score of neuronal plasticity moderates the association between cognitive aptitudes and cortical thickness in adolescents. *Dev Cogn Neurosci* **60**, 101232 (2023).
38. Petersen, C. R. & Al-Haik, A. R. The Development of the Defense Language Aptitude Battery (Dlab. *Educ Psychol Meas* **36**, 369–380 (1976).
39. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data 2016 3:1* **3**, 1–9 (2016).
40. Poldrack, R. A. & Gorgolewski, K. J. Making big data open: Data sharing in neuroimaging. *Nature Neuroscience* vol. 17 1510–1517 Preprint at https://doi.org/10.1038/nn.3818 (2014).
41. Poldrack, R. A. & Poline, J. B. The publication and reproducibility challenges of shared data. *Trends in Cognitive Sciences* vol. 19 59–61 Preprint at https://doi.org/10.1016/j.tics.2014.11.008 (2015).
42. Botvinik-Nezer, R. *et al.* Variability in the analysis of a single neuroimaging dataset by many teams. doi:10.18112/openneuro.
43. Milham, M. P. *et al.* Assessment of the impact of shared brain imaging data on the scientific literature. doi:10.1038/s41467-018-04976-1.
44. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data* **3**, (2016).
45. Eickhoff, S., Nichols, T. E., Van Horn, J. D. & Turner, J. A. Sharing the wealth: Neuroimaging data repositories. *NeuroImage* vol. 124 1065–1068 Preprint at https://doi.org/10.1016/j.neuroimage.2015.10.079 (2016).
46. Markiewicz, C. J. *et al.* The openneuro resource for sharing of neuroscience data. *Elife* **10**, (2021).
47. Rasgado-Toledo, J. *et al.* A Dataset to Study Pragmatic Language and Its Underlying Cognitive Processes. *Front Hum Neurosci* **15**, 666210 (2021).
48. Rasgado-Toledo, J. *et al.* A Dataset to Study Pragmatic Language and Its Underlying Cognitive Processes. *Front Hum Neurosci* **15**, 666210 (2021).
49. Bathelt, J., Rastle, K. & Taylor, J. S. H. Relationship Between Resting State Functional Connectivity and Reading-Related Behavioural Measures in 69 Adults. (2024) doi:10.1162/nol_a_00146.

50. Taylor, J., Bathelt, J. & Rastle, K. Relationship between resting state functional connectivity and reading-related behavioural measures in 69 adults. (2023) doi:10.18112/OPENNEURO.DS004765.V1.0.0.
51. Rogers, C. S. *et al.* Age-Related Differences in Auditory Cortex Activity During Spoken Word Recognition. *Neurobiology of Language* **1**, 452–473 (2020).
52. Bishop, D. V. M., Woodhead, Z. V. J. & Watkins, K. E. Approaches to Measuring Language Lateralisation: An Exploratory Study Comparing Two fMRI Methods and Functional Transcranial Doppler Ultrasound. *Neurobiology of Language* **5**, 409–431 (2024).
53. Gold, C. Exploring the Resting State Neural Activity of Monolinguals and Late and Early Bilinguals. *Theses and Dissertations* (2018).
54. Nastase, S. A. *et al.* The "Narratives" fMRI dataset for evaluating models of naturalistic language comprehension. *Sci Data* **8**, (2021).
55. Li, J. *et al.* Le Petit Prince multilingual naturalistic fMRI corpus. *Scientific Data 2022 9:1* **9**, 1–15 (2022).
56. Schoffelen, J.-M. *et al.* A 204-subject multimodal neuroimaging dataset to study language processing. doi:10.1038/s41597-019-0020-y.
57. Bhattasali, S., Brennan, J. R., Luh, W.-M., Franzluebbers, B. & Hale, J. T. The Alice Datasets: fMRI & EEG Observations of Natural Language Comprehension. 11–16 (2020).
58. Hanke, M. *et al.* A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *Sci Data* **1**, (2014).
59. Isaieva, K. *et al.* Multimodal dataset of real-time 2D and static 3D MRI of healthy French speakers. doi:10.1038/s41597-021-01041-3.
60. Lim, Y. *et al.* a multispeaker dataset of raw and reconstructed speech production real-time MRI video and 3D volumetric images. doi:10.1038/s41597-021-00976-x.
61. Gomez-Marin, A., Paton, J. J., Kampff, A. R., Costa, R. M. & Mainen, Z. F. Big behavioral data: Psychology, ethology and the foundations of neuroscience. *Nature Neuroscience* vol. 17 1455–1462 Preprint at https://doi.org/10.1038/nn.3812 (2014).
62. Hintz, F., Dijkhuis, M., van 't Hoff, V., McQueen, J. M. & Meyer, A. S. A behavioural dataset for studying individual differences in language skills. *Sci Data* **7**, (2020).
63. Berthele, R. & Udry, I. *Individual Differences in Early Instructed Language Learning: The Role of Language Aptitude, Cognition, and Motivation*. (Language science press, 2021).
64. Turner, B. O., Paul, E. J., Miller, M. B. & Barbey, A. K. Small sample sizes reduce the replicability of task-based fMRI studies. doi:10.1038/s42003-018-0073-z.
65. Malik-Moraleda, S. *et al.* An investigation across 45 languages and 12 language families reveals a universal language network. *Nature Neuroscience 2022* 1–6 (2022) doi:10.1038/s41593-022-01114-5.
66. Rodriguez-Fornells, A., Kramer, U., Lorenzo-Seva, U., Festman, J. & Münte, T. Self-Assessment of Individual Differences in Language Switching. *Front Psychol* **2**, (2012).

67. Ryan, S. THE IDEAL L2 SELVES OF JAPANESE LEARNERS OF ENGLISH. (University of Nottingham, 2008).
68. Thompson, A. S. & Lee, J. The Motivational Factors Questionnaire in the Korean EFL context: predicting group membership according to English proficiency and multilingual status. *Language Learning Journal* **46**, 398–414 (2018).
69. Lefly, D. L. & Pennington, B. F. Reliability and validity of the adult reading history questionnaire. *J Learn Disabil* **33**, 286–296 (2000).
70. Roebuck, H. & Lupyan, G. The Internal Representations Questionnaire: Measuring modes of thinking. *Behav Res Methods* **52**, 2053–2070 (2020).
71. Chin, T. C., Coutinho, E., Scherer, K. R. & Rickard, N. S. MUSEBAQ: A modular tool for music research to assess musicianship, musical capacity, music preferences, and motivations for music use. *Music Percept* **35**, 376–399 (2018).
72. Barratt, W. The Barratt simplified measure of social status (BSMSS): Measuring SES. *Unpublished manuscript, Indiana State University* (2006).
73. Rakesh, D. & Whittle, S. Socioeconomic status and the developing brain–A systematic review of neuroimaging findings in youth. *Neurosci Biobehav Rev* **130**, 379–407 (2021).
74. Oldfield, R. C. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
75. Nedjar, T., Touari, M., Mesbah, M., Lellouch, J. & Dellatolas, G. La préférence manuelle dans une population d'étudiants algériens francophones et comparaison avec la population française. *Annee Psychol* **89**, 239–253 (1989).
76. Stansfield, C. W. *Language Aptitude Reconsidered. ERIC Digest*. https://eric.ed.gov/?id=ED318226 (1989).
77. Golestani, N. & Pallier, C. Anatomical correlates of foreign speech sound production. *Cerebral Cortex* **17**, 929–934 (2007).
78. Golestani, N., Paus, T. & Zatorre, R. J. Anatomical Correlates of Learning Novel Speech Sounds. *Neuron* **35**, 997–1010 (2002).
79. Kepinska, O., de Rover, M., Caspers, J. & Schiller, N. O. On neural correlates of individual differences in novel grammar learning: An fMRI study. *Neuropsychologia* **98**, 156–168 (2017).
80. Opitz, B. & Friederici, A. D. Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *Neuroimage* **19**, 1730–1737 (2003).
81. Raven, J. C. Matrices Progressives de Raven Avancées - Abregées. Preprint at (1998).
82. Kessels, R. P. C., Van Zandvoort, M. J. E., Postma, A., Kappelle, L. J. & De Haan, E. H. F. The Corsi Block-Tapping Task: Standardization and normative data. *Appl Neuropsychol* **7**, 252–258 (2000).
83. Corsi, P. M. Human memory and the medial temporal region of the brain. (Mc Gill University, Montréal, Canada, 1972).

84. Arce, T. & McMullen, K. The Corsi Block-Tapping Test: Evaluating methodological practices with an eye towards modern digital frameworks. *Computers in Human Behavior Reports* **4**, 100099 (2021).
85. Wechsler, D. Wechsler Adult Intelligence Scale–Fourth Edition. (2008) doi:10.1037/t15169-000.
86. Ryan, J. J., Townsend, J. M. & Kreiner, D. S. Comparison of Oral, Written, and Pointing Responses to WAIS-IV Digit Span. *Appl Neuropsychol Adult* **21**, 94–97 (2014).
87. Conway, A. R. A., Kane, M. J. & Al, C. E. T. Working memory span tasks: A methodological review and user's guide. *Psychon Bull Rev* **12**, 769–786 (2005).
88. Bellon, E., van Bergen, E. & Dowker, A. D. Is Parental Mathematics Anxiety Associated with Young Children's Arithmetical Performance? *Educ Sci (Basel)* **12**, 812 (2022).
89. Gordon, E. E. *Music Aptitude and Related Tests An Introduction*. (GIA Publications Inc, Chicago, 1989).
90. Callejas, A., Lupiàñez, J., Funes, M. J. & Tudela, P. Modulations among the alerting, orienting and executive control networks. *Exp Brain Res* **167**, 27–37 (2005).
91. Deweer, B., Poitrenaud, J., Kalafat, M. & der Linden, M. CVLT - Test d'apprentissage et de mémoire verbale. Preprint at https://www.pearsonclinical.fr/cvlt-test-dapprentissage-et-de-memoire-verbale (2008).
92. Strauss, E., Sherman, E. M. S. & Spreen, O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary, 3rd Ed*. (Oxford University Press, New York, NY, US, 2006).
93. Ashendorf, L., Horwitz, J. E. & Gavett, B. E. Abbreviating the Finger Tapping Test. *Archives of Clinical Neuropsychology* **30**, 99–104 (2015).
94. Tiffin, J. & Asher, E. J. The Purdue Pegboard: norms and studies of reliability and validity. *Journal of applied psychology* **32**, 234 (1948).
95. Frederickson, N., Frith, U. & Reason, R. *Phonological Assessment Battery (PhAB): Manual and Test Materials*. (NFER-Nelson, Windsor, 1997).
96. Rutten, S., Santoro, R., Hervais-Adelman, A., Formisano, E. & Golestani, N. Cortical encoding of speech enhances task-relevant acoustic information. *Nat Hum Behav* **3**, 974–987 (2019).
97. Gola-Asmussen, C., Lequette, C., Pouget G., Rouyer, C. & Zorman, M. ECLA-16+. Evaluation des compétences en lecture chez l'adulte de plus de 16 ans. 1–63 (2010).
98. Lefavrais, P. Test de l'Alouette: Test d'analyse de la lecture et de la dyslexie. *Paris: Editions du Centre de Psychologie Appliquée* (1967).
99. Sprenger-Charolles, L., Colé, P., Béchennec, D. & Kipffer-Piquard, A. French normative data on reading and related skills from EVALEC, a new computerized battery of tests (end Grade 1, Grade 2, Grade 3, and Grade 4)1. *European Review of Applied Psychology* **55**, 157–186 (2005).
100. Szenkovitz, G. & Ramus, F. Exploring dyslexics' phonological deficit I: Lexical vs sub-lexical and input vs output processes. *Dyslexia* **11**, 253–268 (2005).

101. Majerus, S., Van der Linden, M., Mulder, L., Meulemans, T. & Peters, F. Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm. *J Mem Lang* **51**, 297–306 (2004).
102. Fedeli, D., Del Maschio, N., Sulpizio, S., Rothman, J. & Abutalebi, J. The bilingual structural connectome: Dual-language experiential factors modulate distinct cerebral networks. *Brain Lang* **220**, 104978 (2021).
103. OECD. *Education at a Glance 2023: OECD Indicators*. (OECD, 2023). doi:10.1787/e13bef63-en.
104. OECD. *Education at a Glance 2023 Sources, Methodologies and Technical Notes*. (Organisation for Economic Co-operation and Development, Paris, 2023).
105. Federal Statistical Office Section Demography and Migration. Swiss Federal Population Census Structural Survey - Language. (2022).
106. Milfont, T. L. & Klein, R. A. Replication and Reproducibility in Cross-Cultural Psychology. *Cultural Psychology* (2018).
107. Elson, M., Hussey, I., Alsalti, T. & Arslan, R. C. Psychological measures aren't toothbrushes. *Communications Psychology 2023 1:1* **1**, 1–4 (2023).
108. Pimsleur, P., Reed, D. J. & Stansfield, C. W. *Pimsleur Language Aptitude Battery: PLAB: Manual*. (Second Language Testing Foundation, 2004).
109. Pimsleur, P. *Pimsleur Language Aptitude Battery (Form S)*. (Harcourt, Brace and world, Incorporated, 1966).
110. Woods, K. J. P., Siegel, M. H., Traer, J. & McDermott, J. H. Headphone screening to facilitate web-based auditory experiments. *Atten Percept Psychophys* **79**, 2064–2072 (2017).
111. Zhao, S., Brown, C. A., Holt, L. L. & Dick, F. Robust and Efficient Online Auditory Psychophysics. *Trends Hear* **26**, 23312165221118790 (2022).
112. Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N. & Evershed, J. K. Gorilla in our midst: An online behavioral experiment builder. *Behav Res Methods* **52**, 388–407 (2020).
113. Malik-Moraleda, S. *et al.* The universal language network: A cross-linguistic investigation spanning 45 languages and 12 language families. *bioRxiv* 2021.07.28.454040 (2022) doi:10.1101/2021.07.28.454040.
114. Balboni, I., Rampinini, A., Kepinska, O., Berthele, R. & Golestani, N. Brain activation for language and its relationship to cognitive and linguistic measures: a multimodal exploration. in *Society for the Neurobiology of Language Annual Meeting* (Marseille, France, 2023).
115. Liu, X. & Yang, L. Individual differences in the language task-evoked and resting-state functional networks. *Front Hum Neurosci* **17**, 1283069 (2023).
116. Deng, Z., Chandrasekaran, B., Wang, S. & Wong, P. C. M. Resting-state low-frequency fluctuations reflect individual differences in spoken language learning. *Cortex* **76**, 63–78 (2016).
117. Achal, S., Hoeft, F. & Bray, S. Individual Differences in Adult Reading Are Associated with Left Temporo-parietal to Dorsal Striatal Functional Connectivity. (2015) doi:10.1093/cercor/bhv214.

118. Zhang, G. *et al.* Individual differences in first-pass fixation duration in reading are related to resting-state functional connectivity. *Brain Lang* **213**, (2021).
119. Reineberg, A. E., Gustavson, D. E., Benca, C., Banich, M. T. & Friedman, N. P. The relationship between resting state network connectivity and individual differences in executive functions. *Front Psychol* **9**, 361864 (2018).
120. Reineberg, A. E., Andrews-Hanna, J. R., Depue, B. E., Friedman, N. P. & Banich, M. T. Resting-state networks predict individual differences in common and specific aspects of executive function. *Neuroimage* **104**, 69–78 (2015).
121. Nordhoff, S. & Hammarström, H. Glottolog/Langdoc: Defining dialects, languages, and language families as collections of resources. in *First International Workshop on Linked Science 2011-In conjunction with the International Semantic Web Conference (ISWC 2011)* (2011).
122. Gullifer, J. W. & Titone, D. Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition* **23**, 283–294 (2020).
123. Kepinska, O. *et al.* Language combinations of multilinguals are reflected in their first-language knowledge and processing. *Sci Rep* **13**, 1947 (2023).
124. Shannon, C. E. A mathematical theory of communication. *The Bell system technical journal* **27**, 379–423 (1948).
125. Hausser, J., Strimmer, K. & Strimmer, M. K. Package 'entropy'. *R Foundation for Statistical Computing: Vienna, Austria* (2012).
126. Allen, M. P. The problem of multicollinearity. *Understanding regression analysis* 176–180 (1997).
127. Chan, J. Y.-L. *et al.* Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics* **10**, 1283 (2022).
128. Slinker, B. K. & Glantz, S. A. Multiple regression for physiological data analysis: the problem of multicollinearity. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **249**, R1–R12 (1985).
129. Paul, R. K. Multicollinearity: Causes, effects and remedies. *IASRI, New Delhi* **1**, 58–65 (2006).
130. Green, S. B. & Yang, Y. Evaluation of Dimensionality in the Assessment of Internal Consistency Reliability: Coefficient Alpha and Omega Coefficients. *Educational Measurement: Issues and Practice* **34**, 14–20 (2015).
131. Vaske, J. J., Beaman, J. & Sponarski, C. C. Rethinking Internal Consistency in Cronbach's Alpha. *Leis Sci* **39**, 163–173 (2017).
132. Drennan, J. Quantitative health research: issues and methods. in (ed. Curtis, E.) (McGraw-Hill Education (UK), 2013).
133. Watson, R. Issues and debates in validity and reliability. in *Quantitative Health Research: Issues And Methods* (eds. Curtis, E. & Drennan, J.) (McGraw-Hill Education (UK), 2013).
134. Cronbach, L. J. & Hartmann, W. A note on negative reliabilities. *Educ Psychol Meas* **14**, 342–346 (1954).

135. Wile, T. L. & Borowsky, R. What does rapid automatized naming measure? A new RAN task compared to naming and lexical decision. in *Brain and Language* vol. 90 47–62 (2004).

136. Golestani, N. & Pallier, C. Anatomical Correlates of Foreign Speech Sound Production. *Cerebral Cortex* **17**, 929–934 (2006).

137. Thompson, B. L., Green, S. B. & Yang, Y. Assessment of the maximal split-half coefficient to estimate reliability. *Educ Psychol Meas* **70**, 232–251 (2010).

138. Gaser, C. *et al.* CAT – A Computational Anatomy Toolbox for the Analysis of Structural MRI Data. *bioRxiv* 2022.06.11.495736 (2023) doi:10.1101/2022.06.11.495736.

139. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *Neuroimage* **62**, 782–790 (2012).

140. Jbabdi, S., Sotiropoulos, S. N., Savio, A. M., Graña, M. & Behrens, T. E. J. Model-based analysis of multishell diffusion MR data for tractography: how to get over fitting problems. *Magn Reson Med* **68**, 1846–1855 (2012).

141. Bastiani, M. *et al.* Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. *Neuroimage* **184**, 801–812 (2019).

142. Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J. & Turner, R. Movement-Related effects in fMRI time-series. *Magn Reson Med* **35**, 346–355 (1996).

143. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods* **16**, 111–116 (2019).