MASTERARBEIT | MASTER'S THESIS

Titel | Title

Individual differences in general cognitive measures and functional neuroanatomy related to grammar learning ability.

verfasst von | submitted by

Laurin Daniel Koblicha-Rathausky BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Psychologie
Betreut von | Supervisor:	Assoz. Prof. Narly Golestani BSc PhD

I want to take this opportunity to thank those who have made the writing of this thesis possible and who have supported me either professionally, emotionally or both in that process.

Without the permission to use the data arduously gathered by the people at the Brain and Language Lab there would have been little to write about, so I want to thank Assoz. Prof. Narly Golestani BSc PhD, Olga Kepinska BA MA PhD, Irene Balboni MSc, Alessandra Rampinini BA MA PhD, Giulio Degano BA MA PhD and Sevil Maghsadhagh BSc MSc for the work they have put into gathering, preprocessing and maintaining the data and allowing me to use some of it in this thesis. I also want to thank the same people for providing invaluable feedback and support at different points of this whole process.

I want to especially thank Olga Kepinska for the time she took to help me get this thesis to where it is now by lending me her experience, eye for detail and time.

For tolerating my ramblings, questions and frustrations I want to thank anyone who had to endure them but especially Sarah Bartels and Felix Harreus who sat next to me during a large part of the writing process and who were always kind enough to hear me out.

Finally, thank you Narly Golestani for your supervision and also for giving me the opportunity to get a glimpse into the inner working of a research lab.

Mastering the grammar of a new language involves the acquisition and application of rules that govern both the hierarchical structure of words within sentences (hierarchical syntax) and the modulation of words themselves for the purpose of conveying grammatical relations (morphosyntax). Evidence has suggested that considerable individual differences exist in people's proclivity for acquiring a new grammar. In this thesis these differences are explored on a behavioural level, looking at cognitive variables and language experience, as well as on a functional neuroanatomical level, using resting-state functional connectivity (RSFC) data. The goal was to not only identify variables important for grammar learning generally but also to investigate a putative distinction between the systems responsible for hierarchical syntax and morphosyntax. For this purpose, variables related to performance in a grammar learning task focused on hierarchical syntax (*Brocanto*) and a task focused on morphosyntax (*ArtGram*) were identified in order to find commonalities and differences between the tasks. Better performance in both tasks was related to higher measures of verbal learning but while *ArtGram* performance increased with better rote learning ability, *Brocanto* scores correlated positively with a measure of fluid intelligence. The tasks showed different patterns of RSFC associated with performance. *ArtGram* performance was positively associated with coactivation of the posterior middle temporal gyrus (left and right) and a host of striatal regions purportedly part of a procedural learning system. *Brocanto* performance was predicted by decreased coactivation of the lpMTG and regions immediately surrounding it, a result consistent with theories of neural efficiency.

Die Grammatik einer neuen Sprache zu meistern umfasst das Erlernen und Anwenden von Regeln, die sowohl die hierarchische Struktur von Wörtern innerhalb von Sätzen (hierarchische Syntax) als auch die Modulation der Wörter selbst (Morphosyntax) regeln. In der vorliegenden Arbeit wurden die individuellen Unterschiede in der Fähigkeit sich eine neue Grammatik anzueignen, sowohl auf der Verhaltensebene als auch auf funktionell neuroanatomischer Ebene, unter Verwendung von Daten zur funktionellen Konnektivität im Ruhezustand (RSFC), betrachtet. Das Ziel der Untersuchung bestand darin, Variablen zu identifizieren, die für das Grammatiklernen im Allgemeinen von Relevanz sind und eine mutmaßliche Unterscheidung zwischen den Systemen für hierarchische Syntax und der Morphosyntax zu untersuchen. Zu diesem Zweck wurden Variablen identifiziert, die mit der Leistung in einer hierarchischer Grammatikaufgabe (*Brocanto*) und einer Morphosyntaxaufgabe (*ArtGram*) zusammenhängen, um so Gemeinsamkeiten und Unterschiede zwischen den Systemen zu ermitteln. Die Ergebnisse zeigen einen Zusammenhang zwischen einer besseren Leistung in beiden Aufgaben und höheren Werten für das verbale Lernen. Während die *ArtGram*-Leistung mit einer besseren Auswendiglernfähigkeit korrelierte, korrelierten die *Brocanto*-Werte positiv mit einem Maß für fluide Intelligenz. In Verbindung mit der Leistung in den beiden Tests zeigten sich auch unterschiedliche Muster der RSFC. Die *ArtGram*-Leistung wies eine positive Korrelation mit der Koaktivierung des pMTG (links und rechts) sowie einer Reihe von Striatumregionen auf, die mutmaßlich Teil eines prozeduralen Lernsystems sind. Die beobachtete *Brocanto*-Leistung wurde durch eine verminderte Koaktivierung des lpMTG und der ihn unmittelbar umgebenden Regionen vorhergesagt.

Contents

Contents 5
1 Introduction 7
1.1 Neural Correlates of Syntax 9
1.2 Neural Correlates of Grammar Learning 11
1.3 Measures used in this project 13
1.4 Artificial Grammar Learning 14
1.5 Resting-State Functional Connectivity 15
1.6 Expectations and Goals 16
2 Methods 18
2.1 Sample 18
2.2 Data collection 18
2.2.1 Language experience 19
2.2.2 Artificial grammar learning tasks 19
2.2.3 Independent variable measures 21
2.3 Analysis 24
2.3.1 Behavioural data 24
2.3.2 Neuroimaging data 27
3 Results 29
3.1 Behavioural Results 29
3.1.1 Descriptive statistics 30
3.1.2 Cluster analysis for Brocanto 32
3.1.3 Correlations 35
3.1.4 Brocanto 38
3.1.5 ArtGram 41
3.2 Neuroimaging results 44
3.2.1 Mean connectivity for all ROIs 44

3.2.2 Connectivity predicting task performance 48
4 Discussion 52
4.1 Result interpretation 52
4.1.1 Behavioural 52
4.1.2 RSFC 54
4.2 Big picture integration with Matchin & Hickok model 58
5 References 62
6 Supplementary Materials 71
6.1 Tables 71
6.2 Figures 76

1 Introduction

Considerable individual differences exist within language learning, a faculty that is unique to humans. In addition to the acquisition of new vocabulary and the ability to hear and articulate novel sounds, mastery of the grammatical rules of a new language is one of its fundamental objectives. In order to comprehend the grammatical relations between words in a sentence, it is necessary to parse both sentence-level hierarchical structures and register changes to words themselves. Syntactic processing can therefore be divided into hierarchical lexical-syntactic and morphosyntactic processing. The former describes how grammatical relationships expressed through the positioning of words within a sentence are understood, while morphosyntactic processing refers to the understanding of grammatical relationships indicated through affixation (Matchin & Hickok, 2020). The objective of this thesis is twofold: first, to further explore the variables involved in successful grammar acquisition generally; and second, to determine whether these differ between the two types of syntactic processing. The present study will attempt to differentiate between these approaches on a behavioural level. This will be achieved by examining individual differences in several cognitive variables as a function of actual grammar learning success. Next, an attempt will be made to relate grammar acquisition to brain imaging data.

When it comes to language learning in general, researchers have, since the 1950s, attempted to determine whether a 'talent' for language truly exists, whether it is an unchangeable characteristic or if it can be developed, whether such an ability is measurable, and how it relates to other more general cognitive abilities (Dörnyei & Skehan, 2003; Li, 2016; Singleton, 2017). Within the domain of language acquisition research, this perceived proclivity for linguistic learning is designated as language aptitude. Notwithstanding the considerable debate surrounding the theoretical validity of this construct (Li, 2015), it has been proven to be a useful predictor of learning outcomes (Carroll, 1962; Dörnyei & Skehan, 2003) and varies across individuals.

The predictive power of language aptitude can be traced back to its inception when American psychologist and linguist John B. Carroll, along with his colleague Stanley Sapon, developed the Modern Language Aptitude Battery (MLAT, Caroll & Sapon, 1966) a test battery still in use today (Dale & Sparks, 2023; Dörnyei & Skehan, 2003; Li, 2019). In the process of constructing the MLAT, they identified four categories of abilities that contribute

to successful language acquisition: phonemic coding ability, grammatical sensitivity, inductive language learning ability, and associative memory (Carroll, 1962).

Two components of language aptitude are linked particularly often to second language (L2) grammar learning (Li, 2015). Grammatical sensitivity, as the ability to identify the function of a word in a sentence, and inductive language learning ability, the capacity to formulate new sentences in accordance with established grammatical principles, which are summarized by some language aptitude measures as language analytical ability (e.g. Meara et al. 2002; Meara, 2005). Research has demonstrated the efficacy of measures of language analytical ability in predicting grammar learning in natural languages, as well as the acquisition of artificial grammars (see Kepinska et al., 2017b). Language analytical ability is the component of language aptitude which has been linked most strongly with more general cognitive abilities commonly attributed to general intelligence (Dörnyei & Skehan, 2003).

In addition to the components of language aptitude that have already been established, associations have been found between grammar learning and other cognitive variables. Kidd (2012) discovered a correlation between implicit statistical learning and the acquisition of syntax. The study demonstrated that more robust rebound effects in serial reaction time tasks (a measure of implicit learning) are associated with grammatical learning in children. Furthermore, recent findings have indicated that individual differences in statistical learning serve as a predictor of syntactic comprehension (Kidd & Arciuli, 2016). As demonstrated in the research by Deldar et al. (2020), an association has been identified between working memory and the comprehension and production of syntax. Furthermore, an investigation by Martin and Ellis (2012) has revealed a significant relationship between working memory, in conjunction with short-term verbal memory, and grammar learning success. As posited by Myachykov et al. (2005), attentional processes have been identified as a factor influencing the production of grammatical structures. Further research has demonstrated that these processes are influential, at least in the initial phases of learning (Chen et al., 2022). Investigating several predictors of L2 grammar learning success like statistical learning, executive control, working memory and attention, the study found statistical learning and executive control to be influential over the span of the learning process and some association between learning success and attention, especially in the early stages of learning.

Individual differences in the realm of language learning manifest not only in the dimension of cognitive ability measures. Contemporary brain imaging methodologies have facilitated the identification of cortical regions that demonstrate varying degrees of activation

in accordance with individuals' aptitude for language acquisition (e.g., Antonenko et al., 2012; Kepinska et al., 2017 a, 2017b). One approach to investigating variability between learners has been to analyze resting-state functional connectivity (RSFC) data. Previous studies have demonstrated its capacity to predict learning outcomes in the domain of semantic learning (see Rabini et al., 2023).

One of the objectives of this thesis is to examine the neural correlates of successful grammar learning, and to determine whether these correlates differ between the successful acquisition of hierarchical syntactic and morphosyntactic proficiency. This investigation is informed by a recent model for the cortical organization of syntax, which provides a theoretical foundation for the expectation of such disambiguation.

1.1 Neural Correlates of Syntax

In their article on the cortical organization of syntax, Matchin and Hickok (2020) sought to synthesise diverse findings in the literature and establish a cohesive understanding of the neural underpinnings of syntax. The introduction of a dual-stream model of syntactic processing is proposed, which entails a partial separation of the networks involved in language comprehension and production.

The proposed model is centred around the posterior inferior frontal gyrus (pIFG), commonly referred to as Broca's area, and the posterior middle temporal gyrus (pMTG). The network surrounding the pMTG is linked to hierarchical-lexical syntactic functions, which contribute to both sentence production and comprehension. In contrast, the posterior inferior frontal gyrus (pIFG) network is characterised by a more specialised role, primarily responsible for transforming relational hierarchies into morphosyntactic sequences, the latter being a specific demand for language production.

In accordance with their theoretical framework on language comprehension, the posterior superior temporal gyrus (pSTG) is considered to be predominantly responsible for processing auditory phonological representations. Once perceived, these phonological elements undergo decoding in the pMTG and are subsequently linked to conceptual networks located in the anterior temporal lobe and the angular gyrus. This process facilitates the comprehension of a sentence as an array of auditory phonological elements, representative of meaningful concepts, arranged according to grammatical rules.

The neural substrate changes in accordance with the cognitive operations required for the expression of thoughts and the production of sentences. The process of generation commences with unstructured information that is devoid of hierarchical arrangement

according to grammatical rules, existing instead in a non-sequential form. The initial task is to structure these elements, derive their hierarchical organization (such as subject and object), and subsequently transform them into articulable sequences. Matchin and Hickok (2020) describe the process as beginning in the pMTG, which is responsible for structuring information. The formerly abstract information now transposed into a structure is then transformed into morphological sequences, ready for articulation. This transformation is carried out by the dorsal precentral gyrus, pars triangularis, and sylvian parietal temporal area. The collaboration between the areas involved in this network enables the translation of abstract thoughts into concrete, articulable linguistic forms.

This model of partially separated systems subserving the processes of grammar production and comprehension also provides a basis for an anatomical distinction between more hierarchical syntax, and morphosyntax. Hierarchical syntax refers to the processes of structuring elements according to their grammatical organization on a sentence level and establishing (meaningful) relationships between words based on their position relative to each other (e.g., subject object relations, noun phrases, etc.). Morphosyntax on the other hand describes how changes to words themselves (e.g., conjugation), usually through affixation (the addition of morphemes to word stems), communicate syntactic information. While hierarchical syntax is located primarily in the pMTG but is part of both the production and the comprehension system, morphosyntax, according to the model, is a process specifically involved in syntax production and is located in the pTri.

While Matchin and Hickok (2020) provide a comprehensive overview and a well-argued explanation of the research on the neural correlates of syntax, it is important to note that their account is not an exhaustive integration of all available data (Matchin & Hickok, 2020), but rather a theoretically and conceptually motivated approach. As will be demonstrated in the following discussion, certain studies have identified the presence of further areas implicated in the processing of syntax, which may be of significance.
A meta-analytic study using fMRI (Ardila et al., 2016) investigated coactivation between areas commonly implicated in language processing and the rest of the brain. The study concluded that human language processing is best explained by two distinct networks corresponding to comprehension and production, which is consistent with the conclusions of Matchin and Hickok (2020). However, the authors of the study associated a greater number of areas with grammatical processing by identifying a network of coactivation with Broca's area (IFG) that extended beyond the regions involved in the dual-stream model of syntactic

processing (Matchin & Hickok, 2020). The network in question comprises the left frontal operculum, the supplementary motor area, left parietal and temporal lobes, as well as left secondary visual areas and several subcortical areas, including the left thalamus, basal ganglia, and cerebellum. The study revealed a robust coactivation between Broca's area, the premotor cortex (PMC), and the dorsolateral prefrontal cortex, which is consistent with the findings of Matchin and Hickok (2020). Furthermore, the researchers identified a correlation between the inferior frontal gyrus (IFG) and the insula, a region which they hypothesise plays a regulatory role in the language system.

The broad spectrum of scientific research demonstrates the complexity of the neural correlates of grammar processing. However, variations across these studies may be partly attributable to methodological differences, and there is agreement that the left inferior frontal gyrus, along with the superior temporal gyrus and the posterior superior temporal sulcus/pMTG, appear to be the primary sites of syntax in the human brain (Grodzinsky et al., 2021; Matchin & Hickok, 2020).

1.2 Neural Correlates of Grammar Learning

Typically, the brain regions involved in processing native language grammar are also involved in L2 or artificial grammar learning. A meta-analysis of fMRI studies found bilateral activation in the frontal IFG and surrounding regions, the posterior parietal cortex, as well as the left hemispheric activation in the putamen and caudate nucleus during grammar learning (Tagarelli et al., 2019). However, the acquisition of novel grammar involves additional demands that are reflected in cortical activation.

The premotor cortex (PMC) has been observed to exhibit heightened activation during the acquisition of syntax, as evidenced by the following studies. Some results implicate it in phonological processes like phoneme comparison and segmentation which subserve successful grammar acquisition (Goranskaya et al., 2016). Opitz and Friederici's (2004) findings also indicate the involvement of the left PMC in the acquisition of grammar. The findings of the study suggest that, in the initial stages of training, participants tend to utilise similarity-based learning. This tendency is reflected by increased activation in the left anterior hippocampus, as proficiency increases, learners tend to develop a more rule-based understanding, which activates the left PMC. Another area investigated in connection with grammar acquisition is the IFG. Due to its consistent involvement in the processing of syntax (see Matchin & Hickok, 2020) it seemed a prime candidate for another important region for grammar learning. This role however seems to be more in the realm of rule representation and

rule application, with activity in the IFG increasing over the course of the learning process (Opitz & Friederici, 2003; Opitz and Friederici 2004; Goranskaya et al., 2016). The findings of these studies indicate a shift of neural activity over the course of the learning process, with early, more similarity-based, stages being mostly subserved by the hippocampus and later, more rule-based, stages being accompanied by increased inferior frontal activity.

A subset of studies in the domain of grammar learning seeks to identify the neural networks and specific regions that differentiate between skilled and less skilled learners. These studies account for the complexity of learning success by examining the involvement of different networks across the brain. Thereby, understanding learning as a process comprising and influenced by several contributing factors like working memory (Deldar et al., 2020; Kepinska et al., 2017b), emotional processing and regulation (Kepinska et al. 2017b) and age (Antonenko et al., 2012).

A study that is particularly relevant to this work was conducted by Kepinska et al. (2017b). The researchers identified six functional connectivity networks that present in all participants. They then demonstrated that successful and less successful learners of an artificial grammar exhibited different activation in four of these functional networks. Individuals with high language analytical abilities exhibited increased coactivation of various brain regions, including the left and right middle frontal, precentral and inferior frontal gyrus, as well as the left pars opercularis. Furthermore, heightened coactivation of the central opercular cortex and secondary somatosensory cortex, which were part of the working memory network, and in the amygdala and mamillary body, which were part of the emotional network, was observed. The coactivation patterns exhibited by participants with average language analytical abilities (the study only included high and average language analytical ability participants and excluded those scoring below average) were characterised by increased connectivity in the right hemisphere of the language network, as well as in the default mode network, including the left cingulate gyrus and posterior division, and the right hemisphere paracingulate gyrus, cingulate gyrus, anterior division, frontal pole and middle frontal gyrus.

Overall, high-ability participants showed a pattern of broad recruitment of neural resources in language related areas as well as working memory and emotional processing networks, whereas average-ability learners showed increased connectivity of the default mode network and some areas of the right hemispheric language network.

A study investigating successful grammar learning in older adults (Antonenko et al., 2012) yielded results that seemingly contradict those found in Kepinska et al. (2017b). Antonenko et al. (2012) discovered stronger activation of the areas surrounding Broca's area and its right hemisphere homologue in less proficient learners. In contrast, Kepinska et al. (2017b) found stronger connectivity between left and right language areas to be associated with highly skilled individuals. The apparent incongruity between these two findings is likely to be attributed to the changes in cognitive ability and white matter tract structure that are associated with the process of ageing (Antonenko et al., 2012; Kepinska et al., 2017b).

Although not tied specifically to grammar learning, cortical networks associated with working memory have been shown to interact with language specific networks involved in the comprehension and production of syntactical sequences. In their review of the extant research, Deldar et al. (2020) find evidence for an interaction between working memory involvement and syntactic complexity. As the complexity of a task increases, there is a concomitant activation of the working memory networks in the dorsolateral prefrontal cortex, as well as the basal ganglia and premotor areas.

Having thoroughly explored the individual differences associated with grammar learning and its neural underpinnings, the objective of this thesis is to investigate a potential distinction between morphosyntax and hierarchical syntax. This will be pursued through both behavioural and neuroimaging means. At the behavioural level, the question is whether performance on grammar learning tasks is correlated with different cognitive measures, contingent on whether a task targets the acquisition and application of morphosyntactic or hierarchical syntactic rules. In other words, does high performance in a task targeting morphosyntax come with a different 'profile' of cognitive measures than high performance in a hierarchical-lexical syntactic task? Using neuroimaging, the present thesis will investigate the question of whether the structure of the proposed cortical organization of syntax is reflected in functional connectivity patterns at rest associated with performance on grammar acquisition. More specifically it will examine whether resting-state functional connectivity patterns reflect the proposed cortical organization of syntax, specifically by testing whether performance on tasks engaging different syntactic functions is associated with distinct connectivity profiles.

1.3 Measures used in this project

The following sections will provide an introduction into the measurements and types of data this project will utilize to tackle the questions it poses. The central behavioural variables used

for this thesis are artificial grammar learning tasks. The two tasks used are designed to assess participants' grammar learning abilities on the level of hierarchical organization as well as morphosyntactic manipulation. To investigate the neural correlates of artificial grammar learning (AGL) success, this project will use functional magnetic resonance imaging (fMRI) data collected from participants during rest. The analysis of this type of data, called resting-state functional connectivity (RSFC) analysis, relies on the identification of coactivation between regions of the brain.

The following sections will introduce these paradigms, beginning with artificial grammar learning.

1.4 Artificial Grammar Learning

The AGL paradigm, credited to Reber (1967), has become a significant tool in studying implicit learning of grammatical structures. It emerged from the foundational work of Miller (1958), who observed enhanced recall for letter strings following certain rules, even when participants were unaware of those rules. The AGL paradigm involves exposing participants to sequences generated by a finite state grammar, prompting them to acquire knowledge about the underlying rules without explicit instruction. AGL tests participants' ability to discern grammatical patterns and generalize this knowledge to new, unseen sequences.

Empirical studies have demonstrated that performance in AGL tasks can transfer to natural language grammar learning. Research suggests that individuals who excel in AGL tasks, where they implicitly learn grammatical structures with unfamiliar stimuli, also perform better in tasks involving natural language grammar. This correlation indicates a shared cognitive foundation between AGL and natural language acquisition (Ettlinger et al., 2016). An early finding pointing to such a shared cognitive foundation comes from a study by Friederici and colleagues (2002) who found similarities in the neural correlates of artificial and natural language processing. They recorded the electroencephalographic signatures of people processing natural and artificial language. The artificial language used in this study is called BROCANTO. BROCANTO is akin to natural languages in the sense that it is made up of words which are part of different categories (nouns, verbs, adjectives, etc.) which then must be arranged according to certain rules to form phrases and sentences. Participants in the experimental group learned both words and grammatical rules through a computerized board game, while the control group only received vocabulary training. After the learning phase, participants were presented with sentences containing syntactic violations. Those who were

trained in BROCANTO's grammar showed a pattern of event-related potentials closely resembling the ones commonly observed in native speakers processing natural language.

Apart from a task involving BROCANTO (*Brocanto*), this thesis will also analyze data from another AGL task using a different artificial language. This task, called *ArtGram*, was specifically developed for the research project which gathered the data used in this thesis and is conceptually different from *Brocanto* in two ways. First, instead of examining implicit rule acquisition this task is designed with explicit grammar learning in mind. Second, it was specifically designed to cover the manipulation of morphological units which is necessary to demarcate specific functions of words and their relations (Rampinini et al., 2024).

1.5 Resting-State Functional Connectivity

Resting-state functional connectivity differs from task-based fMRI studies, which involve individuals performing specific activities. RSFC captures spontaneous fluctuations in the blood oxygen level-dependent (BOLD) signal while individuals are at rest. This provides a measure of the baseline connectivity patterns of the brain (Fox & Raichle, 2007; Smith et al., 2009).

One of the features of RSFC networks that makes them valuable for research is their stability both within individuals over time (Chou et al., 2012) as well as the fact that similar networks are identified consistently between subjects (Damoiseaux et al., 2006). This allows researchers to examine the inherent architecture of the brain's networks persisting across different sessions. At the same time, individual differences are reflected in the specific strengths and configurations of these networks, with distinct connectivity patterns linked to differences in cognitive functions (e.g., Mennes et al., 2010; Rabini et al., 2023).

Importantly, research has shown a correlation between RSFC networks and task-positive connectivity networks identified during task processing, supporting the idea that the brain's inherent connectivity is closely tied to its involvement in particular cognitive processes. Smith and colleagues (2009) identified corresponding RSFC networks for various behavioural domains, such as language (e.g., semantics, phonology, and speech), emotion, and perception.

The findings of an RSFC study in the field of cognitive neuroscience of language are particularly relevant for the project at hand. In a recent study, Rabini et al. (2023) demonstrated that variations in RSFC networks predict differences in capacity for semantic knowledge measures, highlighting the plausibility of investigating whether RSFC networks also differ based on grammar acquisition. Together with the functional connectivity networks

associated with the ability to acquire a novel grammar (Antonenko et al., 2012; Kepinska et al., 2017a, 2017b), this provides a framework for locating potential neural differences underlying individual variability in grammatical learning using RSFC.

To explore the possible performance dependent differences in RSFC and investigate the existence of distinct neural systems for distinct grammatical processes, this thesis will use a region of interest (ROI) approach. Connectivity patterns between voxels belonging to predetermined regions of the brain to the rest of the brain will be assessed. Four (two in each hemisphere) ROIs will be defined based on the theoretical model of cortical organization laid out by Matchin and Hickok (2020). The pars triangularis was chosen for its proposed involvement in the application of morphosyntactic rules and the posterior middle temporal gyrus because of its role in encoding and perceiving grammatical information in hierarchical sentence structure.

1.6 Expectations and Goals

To contribute to our understanding of the factors involved in successful second language learning, this study will examine a subcategory of language acquisition, namely grammar learning. The objective of this study is to identify variables that predict successful grammar learning. To this end, individual differences in performance on two AGL tasks will be investigated and associated with functional brain imaging and behavioural measures. Furthermore, the differences between hierarchical syntactic and morphosyntactic processing will be investigated.

In accordance with the research findings outlined in the preceding sections, it can be anticipated that AGL ability will demonstrate variability, accompanied by a discernible pattern of both behavioural measures and RSFC. At the behavioural level, no specific variables are expected to be more strongly associated with success in either of the two AGL tasks, *Brocanto* (hierarchical-lexical syntax) and *ArtGram* (morphosyntax). This prediction is in line with the declarative/procedural model of language (Ullman, 2001) which posits two memory systems that underlie language processing and acquisition. The declarative memory system which, broadly speaking, subserves the learning and representation of facts or events, in the context of language, forms the basis for the mental lexicon. The procedural memory system is classically responsible for the acquisition and control of motor or cognitive skills (such as riding a bicycle) and in the realm of language is thought to be involved in the acquisition and application of grammatical rules (Ullman, 2004). These two systems, however, are not thought to work completely independently from one another and especially

during language learning, the theory predicts interactions and dependencies between them. Early stages of adult second language learning - a process investigated in the present thesis - seem to rely mostly on declarative memory systems which can acquire new information fast and provide a basis for later rule-based understanding (Ullman, 2004). Because of this hypothesized dependence of adult grammatical rule learning on the declarative memory system, both *ArtGram* and *Brocanto* are predicted to be similarly associated with measures of declarative memory.

As previous studies have identified task-activated brain networks that differ based on performance in an AGL task (Kepinska et al., 2017b, 2017a) and the general correspondence of task-activated networks to RSFC networks (Smith et al., 2009), it is anticipated that networks related to task performance will be found. Following the model of syntax organization proposed by Matchin and Hickok (2020), distinct networks are expected in relation to performance on the two different AGL tasks. Since tasks involving hierarchical organization should be predominantly attributed to the pMTG and morphological manipulation should be primarily associated with the pars triangularis, these regions were selected as seed ROIs for the RSFC analysis. The separation of the two systems, however, is only expected to be partial, since morphosyntactic processing also relies in part on the hierarchical-lexical organization performed by the pMTG. Therefore, connectivity of the pMTG could be associated with both Brocanto and ArtGram performance, while connectivity of the pTri should be related exclusively to the latter.

More specifically, it is expected that connectivity between the ROIs and other language-related areas such as the left inferior frontal gyrus (IFG), superior temporal sulcus (STS), and middle temporal gyrus (MTG), as well as with working memory regions like the central opercular cortex and areas linked to emotional processing, will be positively related to AGL performance. In line with prior findings (Kepinska, 2017b), stronger bilateral connectivity between language areas—notably the IFG and pars opercularis—is also likely to be associated with enhanced learning outcomes.

In general, functional connectivity networks between one of the ROIs and the rest of the brain that are associated with performance on one of the tasks should differ from those linked to performance in the other. Conversely, finding that the same coactivation networks support both ArtGram and Brocanto performance could be taken as evidence that the distinction between these two tasks is less pronounced than expected.

2 Methods

2.1 Sample

As part of a broader research project on language aptitude and multilingualism (WP Aptitude), a multitude of tests and scales (for a comprehensive overview see Rampinini et al., 2024) were presented to a sample of participants with a wide range of language experience. The original sample consisted of 152 (102 females, M_{age} = 25, SD = 6.06, range 18-47) healthy adults, recruited mostly in Switzerland (Geneva) as well as France. Included in the original sample were 29 individuals with a previous official diagnosis of dyslexia. This subgroup, along with seven hyperpolyglots (self-reported proficiency in more than 7 Languages), were excluded from the analyses done in this project, resulting in a sample of 116 participants (80 female, M_{age} = 24.1, SD = 5.2).

2.2 Data collection

With the project's aim being an exploration of the factors involved in successful language learning, a wide range of questionnaires and tasks were administered to participants. Included were questionnaires on language experience (LEAP-Q, Marian et al., 2007), musical background (MUSEBAQ, Chin et al., 2018), motivation towards learning foreign languages (MFQ, Ryan, 2008; Thompson & Lee, 2018); behavioural tasks assessing language specific skills like grammar learning, rote learning and lexical memory as well as phonetic perception and production; and lastly tasks assessing more general aspects of cognition like fluid intelligence, working memory, attention and motor skills.

Most of these scales were presented to individuals on their own devices during remote testing sessions supervised by a data collector[1]. In two subsequent in-person sessions, participants completed scales which could not be administered remotely, and some tasks were also presented to participants while inside a magnetic resonance imaging (MRI) scanner. In addition to task-based functional MRI (fMRI), brain oxygenation levels were recorded during rest, and structural imaging as well as diffusion imaging (not included in this thesis) were also acquired.

MRI data was acquired using a Siemens 3T Megnetom-Prisma scanner at Campus Biotech (Geneva), equipped with a 64-channel coil. The resting-state connectivity data used

[1] All of those used in this thesis, apart from the LEAP-Q, which was filled out in an unsupervised session and the CVLT, which was conducted in person.

in this project was acquired in the same session. For the resting-state scan, participants were asked to lie still in the scanner with eyes open and to relax as much as possible. BOLD activation was measured using a whole brain EPI sequence (voxel size: 2 mm isotropic voxel, FOV = 224 mm, TR = 2000 ms, TE 32 ms, Flip Angle: 75°, acceleration factor SMS = 3, 72 slices).

2.2.1 Language experience

The Language Experience and Proficiency Questionnaire (LEAP-Q, Marian et al., 2007) is a broad assessment of multilingual language experience and provides researchers with a range of different subscores. To reduce the results of the questionnaire down to some composite measures of language experience, four continuous multilingualism scores were calculated for this study. One score was based on the current exposure to all reported languages and the other three were measures of self-reported proficiency in speaking, reading and comprehension in all reported languages. The scores for each of these domains were calculated using Shannon's entropy equation (Shannon, 1948). In this project only the entropy score for speaking proficiency was used as a proxy for language experience. Since speaking a language demands a sufficient capability of understanding and applying syntactic rules it was assumed to be the score most relevant to grammar learning ability.

2.2.2 Artificial grammar learning tasks

2.2.2.1 *Brocanto*

To assess implicit syntactic rule acquisition the *Brocanto* task was included in this study. In it, participants must recognize sentences of an artificial language as either grammatical or ungrammatical, without having had direct access to the rules of the artificial language's syntactic system. An array of grammatical and ungrammatical sentences in the artificial language called BROCANTO, originally developed in (Friederici et al., 2002), was constructed to be used as stimulus material for this task.

All sentences were constructed using a small vocabulary of eight words, which could be categorized into different classes like nouns, verbs, adjectives, adverbs and determiners. Each class was uniquely identifiable by a particular vowel (nouns were specified by 'u', verbs by 'e', adjectives by 'ü', adverbs by 'ö' and determiners by 'aa'). The order by which words of different classes could be strung together followed the rules of a syntactic system designed to resemble those also found in natural language. Sentences consisted of a noun phrase and a

verb phrase, with a noun phrase comprising at minimum a determiner followed by a noun and a verb phrase consisting of a verb or a verb followed by an adjective. More complex noun phrases could be constructed by inserting an adjective between the determiner and the noun. To ensure some rule novelty despite these similarities to natural language, BROCANTO has a specific determiner for when a noun phrase contains an adverb and one for those containing just a determiner and a noun.

Similar to a previous study (Kepinska et al., 2017a), the task was split into training and testing blocks, presented in three phases. During training, subjects were instructed to read a set of grammatically correct sentences and to discover the syntactic rules of the BROCANTO language. Each training block comprised 40 sentences from the pool of 80 grammatically correct sentences. During testing, participants had to make grammaticality judgements for 40 sentences by button press. Sentences containing syntactic rule violations were constructed based on the set of correct sentences, by replacing a word from one class with one from one of the other classes. This resulted in the ungrammatical sentences containing either a determiner-noun-agreement violation, word class repetition (two consecutive nouns or verbs) or phrase structure violations (e.g., a noun-phrase following a noun phrase).

Of the 120 sentences presented over all three testing blocks, 50% were grammatically correct and about a third of these had been presented during training.

2.2.2.2 *ArtGram*

The *ArtGram* task, designed for the WP Aptitude and tested in a pilot study before use in data collection, is an extension of the fourth part of the Pimsleur Language Aptitude Battery (PLAB, Pimsleur, 1966). In the original PLAB used to predict student language learning success, the fourth subtest presented subjects with a list of words and sentences in a foreign language as well as their English translations. The students were then prompted to create 15 new sentences in the foreign language, testing their ability to understand and apply grammatical rules (Pimsleur, 1966; Stansfield, 1988). Similarly, *ArtGram* consists of a learning phase in which participants are presented with a small lexicon from an artificial language as well as some fully formed sentences in which the syntactic relationships of words are denoted with inflected word forms. A multiple-choice translation task with 12 novel sentences followed the learning phase.

2.2.3 Independent variable measures

2.2.3.1 Modern language aptitude test V (MLAT5)

A subtest of the MLAT (Carroll & Sapon, 1959), *MLAT5* was used to assess the rote learning abilities and lexical memory of participants. After being presented with a list of 24 words in a foreign language and their French translations for two minutes, participants were asked to choose the correct translation for each word in a multiple-choice test.

2.2.3.2 Ravens advanced abridged progressive matrices (APM)

To test non-verbal fluid intelligence, the *APM* (Raven, 1998) was administered to participants. In this test, one is presented with a set of pictures in a three-by-three matrix, with one spot being empty. To complete the picture set, one chooses the missing block from an array of options. The advanced version of Ravens Matrices was used to attain a higher degree of differentiation on the extremes of the ability spectrum.

2.2.3.3 Corsi blocks

Visual working memory was tested using the block-tapping paradigm introduced by Corsi (1972), adapted for a 2-dimensional, computer-based presentation. Nine black squares were shown on a white background (square layout adopted from Kessels et al., 2000) and sequences were established by squares being briefly shown in yellow. To reproduce the sequence, participants clicked on the squares. After correctly repeating the presented sequence, the length of the next one was increased. Participants completed two versions of this task, one in which the sequence was to be repeated in the order of presentation (*forward*) and one in which it was to be repeated in reverse order (*backward*).

The score used for visual working memory in this study is the summed lengths of the longest sequences reproduced correctly in both the forward and backward trials. For a trial to be counted as accomplished, the presented sequence had to be reproduced completely.

2.2.3.4 Digit span

Working memory was not only tested in the visuo-spatial domain but also in the form of verbal working memory. To this end, the *digit span* task was used. Here participants listened to a sequence of numbers which then had to be repeated immediately after listening. Both a forward (repeating the numbers in the same order as heard) and a backward (repeating in reverse order) version of this task was administered through headphones, and responses recorded via keyboard. As with the *Corsi block task*, performance scores were calculated as

the sum of the lengths of the longest sequences reproduced successfully in both the forward and backward versions.

2.2.3.5 Serial reaction time (SRT)

Procedural memory is a form of learning and memory which is critical for the learning and automation of both cognitive and motor skills. As such, it is hypothesized to be an important component of successful grammar learning which involves the acquisition of rules and their implementation (Arthur et al., 2021, Ullman et al., 2020). To test procedural memory, this study used a Serial Reaction Time task (Nissen & Bullemer, 1987). Participants were shown a series of stimuli (smiley faces) which appeared at fixed locations in a diamond shaped grid. Each time a face appeared at one of the four locations; participants were required to press the arrow key which corresponded to the location of the stimulus. While the location of the stimulus was random in the first and last block, the locations followed a pseudorandom pattern which repeated every tenth trial. The score used to indicate procedural learning in this study is the difference between the median reaction times (in ms) of the last block (random order) and the second to last block (pseudorandom order). Overall, six blocks (2 random, 4 pseudorandom) comprising 80 trials each, were shown.

2.2.3.6 California verbal learning test (CVLT)

Originally designed as an assessment of memory which takes into account and quantifies learning strategies as well as different processes used by examinees (Delis et al., 1987), the *CVLT* is widely used to test episodic verbal learning. During the test, a list of 16 words, consisting of 4 words from 4 different categories, is presented to participants. The first five times this list is presented, subjects are asked to immediately repeat the list before hearing it again. Following the fifth immediate recall trial, a second list (of the same type as the first one) is presented for one round. After recalling this second list, which serves as a short interference, participants are then requested to recall the items from the first list. After a 20-minute interval, in which participants complete some unrelated non-verbal testing, recall of the first list is tested once again. This procedure results in three types of scores for verbal learning: (1) immediate recall from the first 5 trials, (2) short delay recall after interference list and (3) long delay recall after 20 minutes (Delis et al., 1988). For this study, the validated French version of the *CVLT* (Deweer et al., 2008) was used without any 'cued recall' conditions in either the short or long recall. Only the scores of the short and long delays were subjected to analysis.

2.2.3.7 DecLearn

Non-verbal declarative memory was tested using an object learning and recognition memory task. The general paradigm of object recognition is widely used to ascertain explicit long-term memory (Henson, 2005). For this study, the *DecLearn* was adapted from Hedenius (2013). The task consists of an encoding and a learning phase. During the encoding phase, participants were shown pictures of real and made-up objects and were asked to determine which category ("real" or "made-up") they belonged to. The recognition phase started 10 minutes after the encoding phase had ended, and here too participants were presented with a set of objects, some of which they had seen in the encoding phase and others which were novel. In this phase they were asked whether the stimulus had already been presented to them or not.

2.2.3.8 Attention networks (ANT-I)

The Attention Network Test (ANT) (Fan et al., 2002) is a tool developed for the evaluation of three anatomically and functionally distinct attentional networks: (1) *Executive control* comprises attentional processes involved in conflict resolution, novelty and error detection, (2) *alerting* refers generally to the function of maintaining or inducing an adequate state of cognitive activation and (3) *orienting* is responsible for directing attentional and perceptual processing resources to relevant locations (Callejas et al., 2005; Fan et al., 2002). To quantify the efficiency of these attentional networks, the present project made use of a modified version of the ANT, the Attention Network Test-Interaction (ANT-I) (Callejas et al., 2005). This version takes into account the influences among the networks.

The task revolves around participants being shown an arrow pointing either to the left or the right and indicating the orientation of this target arrow as fast as possible by button press. This arrow was shown either above or below the fixation cross located in the middle of the screen and was accompanied by two irrelevant arrows on each side. These flankers were congruent with the target direction on 50% of trials and incongruent on the other 50%. On half of the trials, presentation of the target arrow was preceded by an auditory signal indicating the imminence of the target presentation. Following the audio signal, a spatial cue was shown on two thirds of the trials. Of the trials where a spatial cue was presented, it was valid on one half and invalid on the other.

The scores for the different attention networks were calculated by comparing the reaction times and error rates between the 12 different possible trial conditions. Specifically, the executive control score (from here on referred to as *inhibition* due to its task of inhibiting

the irrelevant information contained in the flankers) was a function of the difference between congruent and incongruent trials, the alerting score reflected the difference between trials with or without an audio cue and the orientation score was arrived at by comparing trials with invalid and valid spatial cues. Overall, 432 trials were presented during this task.

2.3 Analysis

The statistical analysis of the behavioural data was done in R 4.2.2 (latest version used) in the RStudio environment. Brain imaging data was analyzed in FSL (version 6.0.5.) and visualized using MRIcroGL (version 1.2.2.20220720).

2.3.1 Behavioural data

2.3.1.1 Descriptive statistics

To give an overview of the tasks and measures, descriptive statistics will be reported.

2.3.1.2 Identification of *Brocanto* learning patterns

With the BROCANTO task comprising three different points of measurement (i.e. blocks) for each participant, this data was subjected to additional analyses aimed at identifying different learning patterns within the sample. Following Kepinska (2017a), *k*-means clustering was used to identify groups of participants with similar scores at each of the three points of measurement. In *k*-means clustering, the number of clusters (*k*) must be set a priori and is somewhat arbitrary. There are however different statistical instruments which can provide guidance when deciding *k*.

Choosing k. To make an informed decision about the optimal number of clusters, a scree plot was used. The scree plot is probably the most commonly used instrument for this purpose. It visualizes the average distance of data points within a cluster from the centre of their cluster (Within cluster sum of squares, *WCSS*). These distances between a data point and the centroid of its assigned cluster are a measure for how compact a cluster is. The scree plot then shows the *WCSS* for each *k* in a specified range (e.g., 1-10). The *WCSS* by definition decreases with increasing *k,* but by identifying the point at which the rate of decrease in *WCSS* from *k* to *k*+1 clusters drops noticeably, one can pinpoint the number of clusters which maximizes both cluster compactness and model parsimony. This point is described as an elbow in the graph.

After deciding on an appropriate number of clusters to extract from the *Brocanto* data, using the above-described method as well as knowledge from previous research working with the same task, *k*-means clustering was done using the *stats* package in R.

Since most k-means clustering algorithms are sensitive to changes in the first step of their procedure, which is the choice of initial cluster centres, the analysis was run 1000 times to extract the most robust centroids possible. Additionally, for the same reason, two different algorithms (Bradley & Fayyad, 1998; Hartigan & Wong, 1979) with two different methods of choosing the initial centres were used.

2.3.1.3 Correlations

To gain an overview of the direct relationships between all behavioural variables involved, Pearson correlation coefficients were calculated for each variable pair and tested for significance at $\alpha = .05$.

2.3.1.4 Regression

To test the relationship between the predictor variables and performance on both grammar learning tasks, multiple linear regression models were calculated. Due to the exploratory nature of these analyses, no prior assumptions were made regarding relative variable importance, leading to the decision of entering all predictor variables into the model at the same time. One regression model was calculated for *Brocanto* and one for *ArtGram*. Each model consisted of thirteen predictor variables, which were again tested for significance at $\alpha = .05$.

2.3.1.5 Identification of important variables

Due to the complexity of the models resulting from the initial multiple linear regression analyses, the next step was identifying variables of particular importance. For this purpose, two additional statistical approaches were taken. The decision to employ two different methods for the assessment of variable importance was made to compare the results of the two approaches and arrive at a more robust set of important variables.

2.3.1.5.1 Stepwise regression

Stepwise regression procedures can be used to reduce model complexity by including a variable in the model only if its inclusion satisfies some statistical criteria. There are different types of stepwise regression, the main ones being forward selection and backward elimination. The backward elimination procedure used in this project starts with a fully defined model including all predictors, and gradually removes predictors. After removing the weakest predictor, the resulting model is compared to its previous iteration by means of some model quality criterion. This process is repeated until the removal of further predictors no longer improves the quality of the model.

The criterion used in this case was the Akaike Information Criterion (AIC, Akaike, 2011). It expresses model fit corrected for model complexity by adding a penalty for the number of predictors. This enables a relative comparison of model quality and in the case of backwards stepwise regression, serves as a standard for determining if a predictor should be included in the final model. Once the removal of a predictor no longer meaningfully lowers the AIC, the backwards elimination procedure is stopped.

Stepwise regression is however prone to both the selection of noise variables and to the overestimation of model quality measures (Austin & Tu, 2004; Derksen & Keselman, 1992). To increase the robustness of the stepwise regression analysis, it was performed 1000 times on bootstrapped samples. Doing so significantly reduces the risk of selecting irrelevant variables for the final model (Austin & Tu, 2004), since variables that are actually predictive of the outcome would have a higher than chance level of being included in the final model, while noise variables would not be consistently identified as being important.

By inspecting not only a single selection process but instead taking into account the results of many, one is able to compare the number of times a variable was identified as an independent predictor. This frequency can then be expressed as a ratio between the total number of times a variable was selected to the number of samples drawn in the bootstrap procedure.

2.3.1.5.2 Random forest analysis

The second approach, aimed at assessing variable importance for successful artificial grammar learning, was a random forest analysis. Random forest (Breiman, 2001) is an algorithmic prediction method which uses a large number of decision trees that are grown on random subsets of data and with random subsets of candidate variables (Grömping, 2009). One individual decision tree, usually used to predict categorical outcomes, is grown by splitting the sample into more and more homogeneous groups at each ‘node’ (or branching point). Each split is based on measurements in one variable and is made according to some splitting criterion. This splitting process can be repeated until the sample is split based on all variables one wants to consider or until sufficient partitioning of the sample is achieved (Grömping, 2009).

Instead of growing only one decision tree, random forest grows a large number of trees, each with only a subset of candidate variables. This aggregation of trees makes the method applicable both to categorical as well as continuous outcome measures (Grömping, 2009).

Random forest analysis lends itself well to the estimation of variable importance. Multiple different methods within random forest can be used for this purpose (see e.g. (Debeer & Strobl, 2020; Grömping, 2009; Strobl et al., 2008)), but for this project, the measure used to estimate variable importance will be conditional variable importance (CPI), as described in Debeer and Strobl (2020). The general idea of their approach is that when a variable *z* is important for accurately predicting an outcome, introducing random noise into *z* would lead to a drop in overall prediction accuracy of the random forest model. Doing the same to an unimportant variable, however, would not impact the accuracy of the model in the same way. By not only looking at a single variable *z* disconnected from the rest of the set of possible predictors and instead calculating variable importance dependent on change in variables related to *z,* a variable importance measure can be calculated which takes into account inter-predictor relationships.

Random forest was used to identify variables important in both *Brocanto* cluster assignment and *Brocanto* accuracy as well as in *ArtGram* performance. All random forest models comprised 1000 decision trees and 4 candidate variables for each split in the tree (*ntree* = 1000, *mtry* = 4). This analysis was done in R using the *cforest* function of the *party* package (Strobl et al., 2008).

2.3.2 Neuroimaging data

All RSFC analyses as well as extraction of masks for regions of interest were conducted in FSL (version 6.0.5.; FMRIB software library, www.fmrib.ox.ac.uk/fsl; Jenkinson et al., 2012; Smith et al., 2004; Woolrich et al., 2009).

2.3.2.1 ROI masks

Masks for the four regions of interest (left and right pars triangularis and left and right posterior middle temporal gyrus) were defined using the 50% thresholded probability maps from the Harvard Oxford Cortical Atlas implemented in FSL. These masks defined in standard space were then registered to each participant's individual space for subsequent analysis by applying the reversed transformation matrix to the standard space masks.

The type of RSFC analysis conducted in this project examines the coactivation of a-priori defined regions of interest with the rest of the brain. Based on the theoretical considerations of Matchin and Hickok (2020), four (two in each hemisphere) ROIs were defined.

Pars Triangularis. The first region of interest was the pars triangularis (pTri), a cortical structure located in the frontal lobe, more specifically in the inferior frontal gyrus. To minimize the overlap of the ROI with other functional regions in that cortical area, the probabilistic threshold was set at 50%. The ROI mask was defined once for the left and once for the right hemisphere. The left hemispheric pTri ROI (*Figure 5a*) consisted of 124 voxels with a total volume of 992 mm³. The right hemispheric pTri ROI (*Figure 5b)* consisted of 78 voxels with a total volume of 624 mm³.

Posterior middle temporal gyrus. The second region of interest, the posterior middle temporal gyrus (pMTG), is part of the temporal lobe. The probabilistic threshold was again set at 50%, and both left and right ROIs were determined. The left hemispheric pMTG ROI (*Figure 5c)* consisted of 480 voxels with a total volume of 3840 mm³, and the right hemispheric pMTG ROI (*Figure 5d*) consisted of 497 voxels with a total volume of 3976 mm³.

The masks for all four resulting ROIs were then converted from MNI152 standard space into individual space for the subsequent analysis. The centres of gravity for each ROI in MNI152 standard space are summarized in *Table 1*.

Table 1 – Summary of size and location of the four regions of interest. Volume is shown in number of voxels as well as cubic millimetres and the location is indicated by the coordinates of the ROI's centre of gravity.

ROI	Volume in voxels (in mm³)	X	Y	Z
Left pars Triangularis	124 (992)	71.5	77.5	39.5
Right pars Triangularis	78 (624)	17.5	77	39
Left posterior middle temporal gyrus	497 (3976)	76.5	49	30
Right posterior middle temporal gyrus	480 (3840)	13	51.5	30

2.3.2.2 Preprocessing

Functional images were preprocessed using FSL 6.0. Preprocessing was carried out using FEAT (FMRI Expert Analysis Tool, version 6.0) with a standard preprocessing pipeline that included: removal of non-brain tissue using optiBET (Lutkenhoff et al. 2014), B0 unwarping

using fieldmap images and linear registration of functional images to T1 structural images (BBR, FLIRT); nonlinear registration of structural T1 images to MNI template (FNIRT), and spatial smoothing using a Gaussian kernel of 5 mm FWHM. Data were motion corrected using ICA-AROMA (Pruim et al. 2015). Temporal high-pass filtering was set to 90s.

2.3.2.3 Mean RSFC networks

In order to calculate the resting-state functional connectivity between the ROIs and the rest of the brain as a function of task performance, the mean RSFC was first calculated for the whole sample. This was done by first computing the average activation of all voxels within an ROI at each point of measurement. Resting-state activation was recorded at 300 points in time, resulting in 300 average activation scores per participant per ROI. This sequence of activation scores was then used to identify voxels in the rest of the brain that exhibited activation that was temporally correlated with the ROI ($z = 2.3$), a calculation done using FEAT in FSL. The reported coactivation networks represent the group mean coactivation scores with each ROI.

2.3.2.4 RSFC networks predicting task performance

RSFC analyses predicting task performance were conducted using FSL FEAT to examine the relationship between individual task performance and connectivity patterns between the four regions of interest and the rest of the brain. Separate analyses were performed for each ROI in relation to the two tasks, resulting in a total of eight models. RSFC changes were assessed using a threshold of $z = 2.3$, with results obtained both at an uncorrected alpha level (reported in *Supplementary Table 3*) and after Bonferroni correction for multiple comparisons. Each model included age, sex, and handedness as covariates to account for potential confounding effects. Within each analysis, both positive and negative contrasts were defined to identify clusters of coactivation that were either positively or negatively associated with task performance. This approach allowed for the detection of connectivity patterns predictive of better or worse performance while controlling for individual differences in demographic variables.

3 Results

3.1 Behavioural Results

The dataset upon which the following analyses are based contained one missing value in *brocanto_accuracy*. This value was imputed using nearest-neighbour imputation.

3.1.1 Descriptive statistics

3.1.1.1 Independent Variables

Table 1 shows the average scores and standard deviations for all independent variables used in the analysis of the behavioural data. Variables whose names end with 'perc' were converted to percentages for better readability. Percentages are not computed for variables for which such a calculation would be misleading (e.g. visual and verbal working memory span). In this sample, participants solved on average 45.68% (*SD* = 15.78) of the items in the *Ravens Advanced Progressive Matrices* (*APM_perc*). In the long recall condition of the *California Verbal Learning Task* (*CVLT_long*), which is the score that was ultimately used as a measure of verbal learning in this project (see Section 1.1.3. showing the high correlation between the two *CVLT* scores), participants recalled an average of 12.38 items out of 16 (*SD* = 2.68). The average score of correct translations for the novel words presented in the *MLAT5* was 72.84% (*SD* = 17.66). In the test for explicit non-verbal learning, *DecLearn*, the participants recognized 80.85% (*SD* = 8.84) of the objects correctly on average. The scores for working memory are the sum of the longest sequences recalled in both the backward and forward conditions of the two working memory tests (Corsi blocks and digit span). The average visual working memory span (*span_visual)* was 14.12 items (*SD* = 1.89) and the average verbal working memory span (*span*_verbal) was 12.27 items (*SD* = 2.19). The average score for procedural learning as measured by the *Serial Reaction Time task* was 65.35 ms (*SD* = 31.26). The average degree of multilingualism score was 1.3 (*SD* = 0.40).

Table 2 – Group wide means and standard deviations for all independent variables.

Variable	***n***	***Mean***	***SD***
APM_perc	116	45.68	15.78
CVLT_long	116	12.38	2.68
MLAT5_perc	116	72.84	17.66
alerting	116	24.22	14.15
inhibition	116	-61.91	20.56
orienting	116	51.36	18.74
declearn_perc	116	80.85	8.84
span_verbal	116	12.27	2.19

span_visual	116	14.12	1.89
srt	116	65.35	31.26
multilingualism	116	1.31	0.40

3.1.1.2 *ArtGram* and *Brocanto*.

The dependent variables were performance measures on the artificial grammar learning tasks (1) *ArtGram* and (2) *Brocanto*. In *ArtGram*, the average amount of correct translations was 68.03% (*SD* = 18.19), and in *Brocanto*, the average accuracy for identifying sentences as grammatically correct/incorrect was 74.69 % (*SD* = 12.5). The overall accuracy on the *Brocanto* task was measured as a percentage of correct answers over all three trial blocks. Looking at the three blocks individually, there is a noticeable increase in accuracy from block one (*M* = 56.51%, SD = 9.57) to block 2 (*M* = 63.32%, SD = 13.46) to block 3 (*M* = 66.89%, SD = 15.78). Not only does the mean accuracy increase from trial to trial, but so does the standard deviation. The increasing spread of scores across blocks (*Figure 1a*) reflects growing interindividual differences in learning speed and effectiveness (*Figure 1b*).

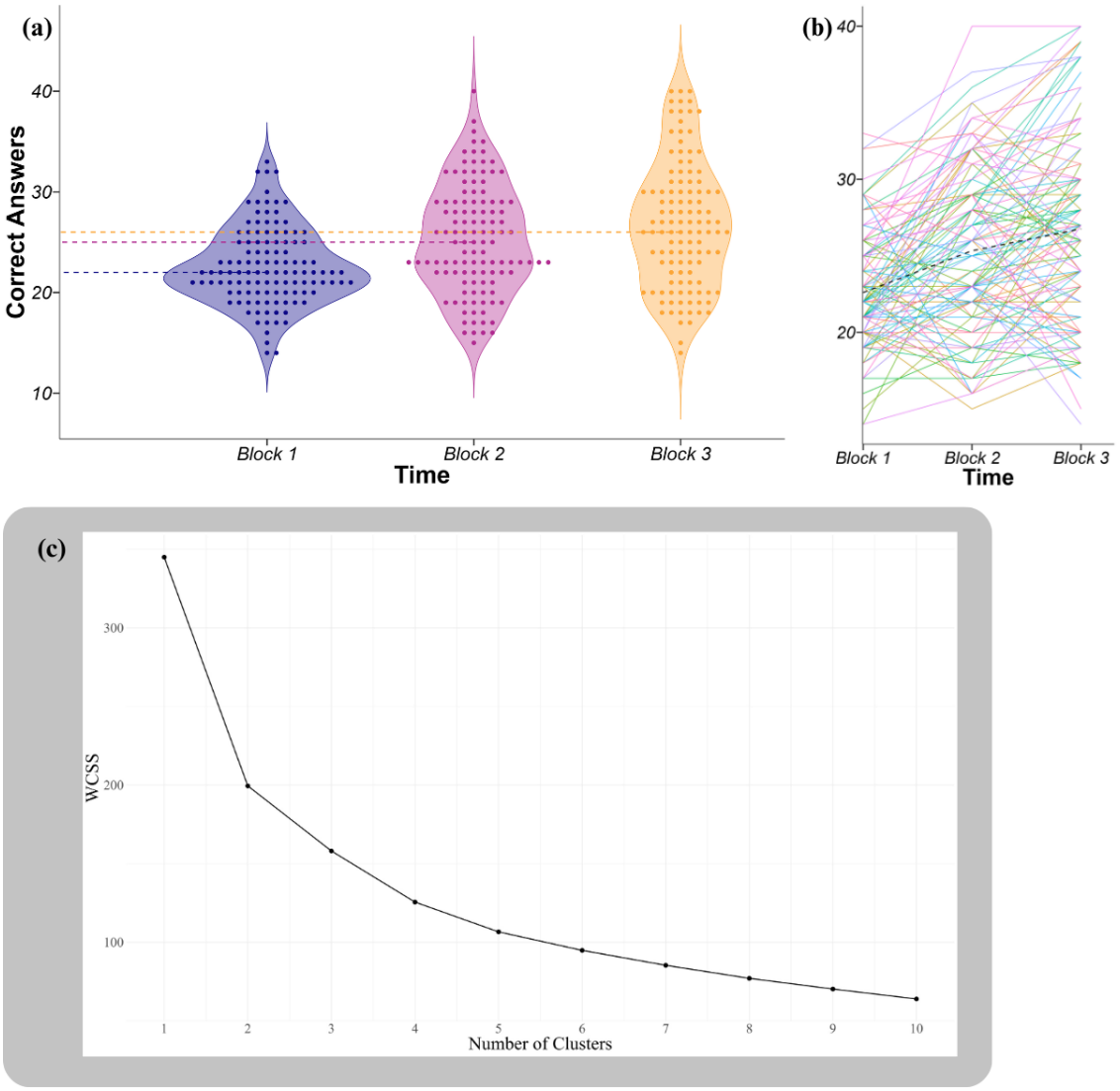


Figure 1 – Panel (a) shows a violin plot with individual data points of the participants Brocanto scores over the three test blocks. The dashed lines represent the group means from Block 1 (blue), Block 2 (pink) and Block 3 (orange). Illustrated in panel (b) is the individual progression of all participants throughout the Brocanto task. The black dashed line again shows the change in group mean score. (c) Scree plot for determining optimal number of clusters (k).

3.1.2 Cluster analysis for Brocanto

The increase in variance throughout the process of the *Brocanto* task, as well as prior studies (Kepinska et al., 2017a) investigating artificial grammar learning using this task, merited the analysis of potential learning patterns emerging from the data. Data from three points in time for each participant enabled such an investigation. This was done by conducting a *k*-means cluster analysis with the aim of identifying groups of participants with similar performance trajectories.

3.1.2.1 Number of Clusters.

The first step in *k*-means cluster analysis is the a-priori choice of how many clusters to extract from the data. Three different statistical instruments providing guidance in finding the optimal number of clusters were used in this project. *WCSS* for 1-10 *k* was calculated and plotted in *Figure 1c*. Identification of an elbow bend in scree plots being subjective in all but the clearest of cases, this plot at most points to a *k* of 2 to maybe 4 clusters being optimal. Ultimately the 4-cluster solution was chosen because it was much more informative in regard to different learning patterns and because it identified a group of participants whose performance seemed to drop over the course of the task. This result seemed plausible as it was congruent with the report from the data collection team indicating that some participants had mentioned exhaustion or tiredness during the *Brocanto* task.

Since most *k*-means clustering algorithms are sensitive to changes in the first step of their procedure, which is the choice of initial cluster centres, the analysis was run 1000 times to extract the most robust centroids possible. Additionally for the same reason, two different algorithms (Bradley & Fayyad, 1998; Hartigan & Wong, 1979) with two different methods of choosing the initial centres were used. As both algorithms converge on very similar solutions, only one solution (Hartigan & Wong, 1979) will be presented here, as it is also the one implemented by default in the R *stats* package.

The results of the *k*-means cluster analysis for this sample are visualized in *Figure 2.*

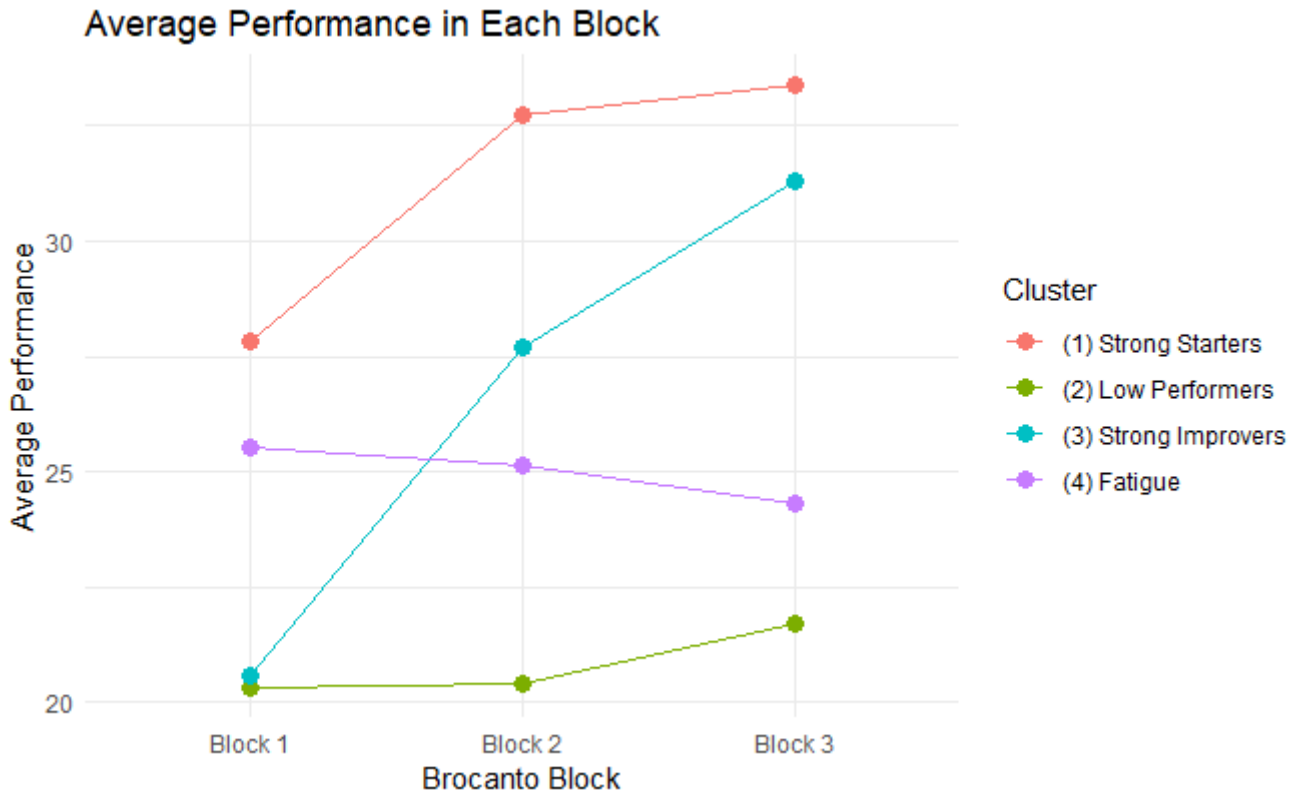


Figure 2 - Average performance (total correct answers) for each trial block of the Brocanto task, with participants separated into four clusters: (1) red line: 'Strong Starters', (2) green line: 'Low Performers', (3) cyan line: 'Strong Improvers', (4) purple line: 'Fatigue'.

The four clusters extracted from the data show distinctly different patterns in their performance trajectories. The first cluster (1) consists of (n = 19) participants who were more likely to already start the first trial block with a comparatively high score. They then improved into block two where they had already almost reached their highest level of performance. The second cluster (2) comprises those participants (n = 43) with comparatively lower *Brocanto* performance. This cluster shows almost no improvement from block 1 to block 2 and only a small amount of it in block three. Cluster number three (3) was the cluster containing participants (n = 32) showing the highest rate of improvement. These participants started out at the bottom of the scale on block one but showed vast improvement over all subsequent blocks. Finally, a fourth cluster (4) (n = 22) was identified which showed a pattern of decreasing performance over time. Given the structure of the task, it is hard to find a reason why with increasing practice, participants would answer fewer questions correctly, so it seems likely that variables other than the ones to be measured by *Brocanto* were driving this effect. During data acquisition it was reported that some participants showed signs of exhaustion and fatigue, particularly during longer tasks, *Brocanto* being one of them. For this reason and

because explanations not tied to fatigue were hard to come by, it was decided that for further analyses, this group of participants would be excluded from the data.

3.1.3 Correlations

The first look at statistical relationships between the variables included in this project was taken using simple correlations. Serving both as a foundation for guiding potential further investigation and as a sanity check to identify any unexpected patterns. The correlations allowed for an assessment of whether tests designed to measure similar constructs were appropriately related; as any substantial divergence from such expected relationships would have warranted closer examination of the data. This very first correlation matrix included scores which ultimately were not part of any further analysis, as well as component scores of measures like the *Serial Reaction Time* task. Reported here is a more condensed version of this correlation matrix, including only those independent variables which were at least considered to be included in further analysis of the behavioural data.

Contained in the upper half of *Figure 3* are the correlation coefficients for all variable pairings. Significant correlations included some expected relations between variables like the long and short recall condition of the *CVLT* ($r = .83$, $p < .001$). This extremely high correlation led to the decision to include only the *CVLT_long* score in further analysis.

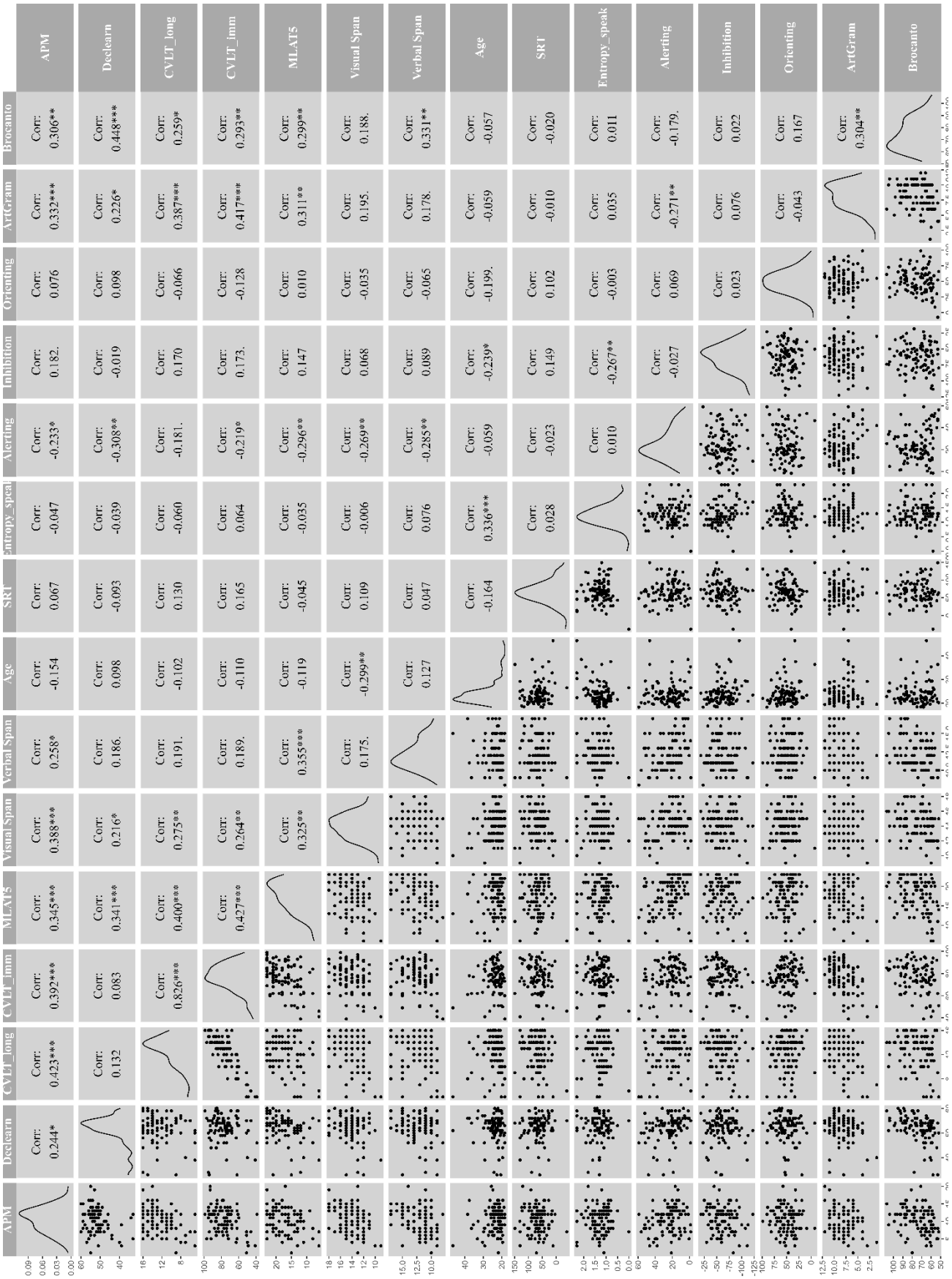
The correlations between these test scores and age fit well with it being a commonly discussed factor, not only in language learning, but learning overall (Cohen et al., 2019). Almost all correlations with performance scores are negative even if only slightly but in most cases under the significance threshold. In this sample, age is significantly correlated with visual working memory ($r = -.27$, $p = .004$), *inhibition* ($r = -.26$, $p = .004$) and *entropy_speak* ($r = .39$, $p < .001$), the last of these correlations being the exception to the pattern of negative correlation with age. Seeing as the *entropy_speak* score is a measure of language experience these results were to be expected.

As for the correlations between dependent and independent variables, *ArtGram* showed significant correlations to *CVLT_long* ($r = .39$, $p < .001$; also to *CVLT_imm*, see *Supplementary Table 1*), *APM* ($r = .25$, $p = .006$), a*lerting* ($r = -.27$, $p = .008$), *DecLearn* ($r = .23$, $p = .028$) and *MLAT5* ($r = .31$, $p = .002$). For an overview of all significant correlations observed in this analysis refer to *Supplementary Table 1*.

Shown on the diagonal in *Figure 3* are the distribution curves for each variable. Visual inspection of these curves shows that most variables are not clearly normally distributed. Age for example noticeably left skews towards the lower age range, while performance measures

like the *CVLT* or the *MLAT5* skew right. Normality was tested using the Shapiro-Wilk test which attributed a normal distribution only to *APM, inhibition* and *SRT*.

Figure 3 - Pair plot based on data after the exclusion of the fatigue cluster. The upper half of the plot shows the correlation coefficients. Depicted on the diagonal is the distribution curve of each variable. The lower half shows a visual representation of the correlation between two variables in the form of a scatter plot.



Looking at the relationship between *Brocanto* and other variables, significant correlations were found between *Brocanto* and verbal working memory span ($r = .3$, $p = .004$), *CVLT_long* ($r = .25$, $p = .014$), *APM* ($r = .31$, $p = .002$), *DecLearn* ($r = .47$, $p < .001$) and *MLAT5* ($r = .28$, $p = .006$). The two AGL tasks, *Brocanto* and *ArtGram*, also showed significant positive correlation ($r = .3$, $p = .003$)

Supplementary Figure 1 shows the same correlation matrix as *Figure* 3 with the only difference being the inclusion of the 'fatigue' cluster. Overall, the pattern of significant correlations is almost congruent between the two correlation analyses, with only a few possibly relevant differences. One of these is the difference in association between *Declearn* and both *Brocanto* ($r_{no\_fatigue} = .47$, $p_{no\_fatigue} < .001$; $r_{fatigue} = .38$, $p_{fatigue} < .001$) and *ArtGram* ($r_{no\_fatigue} = .23$, $p_{no\_fatigue} = .029$; $r_{fatigue} = .19$, $p_{fatigue} = .046$). The only correlation previously not significant, which was significant in this analysis is the one between *verbal* and *visual* working memory span ($r_{no\_fatigue} = .17$, $p_{no\_fatigue} = .1$; $r_{fatigue} = .25$, $p_{fatigue} = .009$)

Supplementary Figure 1 as well as *Supplementary Table 2* show all results from the correlation analysis including the 'fatigue' cluster.

3.1.4 Brocanto

The next step in the process of identifying variables important to successful grammar learning was to look at all our variables at the same time by conducting multiple linear regression analyses. All regressions with the overal *Brocanto* accuracy score as a dependent variable were once again calculated with the slightly smaller sample ($n = 93$), which excludes the 'Fatigue' cluster of the *Brocanto* task.

3.1.4.1 Linear Regression

The first model was a multiple regression model with all predictors entered simultaneously. The independent variables in this first model were almost all the variables introduced before. Only *ArtGram* (because it is one of the dependent variables) and *CVLT_imm* (because of redundancy with *CVLT_long*) were not included as predictors. This resulted in a model with thirteen predictors. Previous analysis had warranted caution regarding the assumption of normality for many of the variables included in this model. However, the Q-Q plot (*Supplementary Figure 2)* shows the distribution of standard residuals deviates only slightly from a theoretical normal distribution.

Out of all variables added to the model, only *DecLearn* ($p < .001$) and verbal working memory span ($p = .04$) emerged as significant predictors for *Brocanto* accuracy. Overall, the model generated with this approach is highly complex and difficult to interpret in its totality.

Table 3 shows all variable parameters, standard errors as well as significance tests for this regression analysis.

In an attempt to simplify the interpretability of the results and possibly identify variables whose importance was masked by the inclusion of other predictors; a stepwise linear regression procedure was conducted.

Table 3 – Results of multiple linear regression analysis with Brocanto as outcome variable and APM, DecLearn, MLAT5, CVLT_long, visual and verbal working memory span, speaking entropy, SRT, alerting, inhibition, orienting and age as predictor variables.

Variable	*B*	*SE*	*t*	*p*
Intercept	2.23	19.41	0.12	.909
APM_correct	0.41	0.4	1.03	.307
Declearn2_correct	0.98	0.25	3.87	.002**
MLAT5	-0.06	0.35	-0.16	.87
CVLT_long	0.69	0.52	1.33	.189
Span_visual	-0.33	0.77	-0.43	.671
Span_verbal	1.33	0.64	2.09	.04*
Entropy_speak	1.77	3.33	0.53	.598
SRT	-0.02	0.04	-0.40	.693
Alerting	0.03	0.1	0.27	.788
Inhibition	-0.03	0.06	-0.40	.689
Orienting	0.07	0.07	1.09	.278
Age	-0.38	0.31	-1.21	.229

*Significance codes: 0.05 * 0.01 **

3.1.4.2 Stepwise Regression

Applying the stepwise regression procedure to the original regression model for *Brocanto* accuracy (see 3.1.4.1), yields a model with four predictors: *DecLearn* ($p < .001$), *CVLT_long* ($p = .053$), *span_verbal* ($p < .001$) and *orienting* ($p = .075$).
shows a summary of the bootstrapped results.

The results of the bootstrapped stepwise regression (summarized in *Table 3*) are broadly consistent with the simple stepwise regression, with the four most frequently identified predictors being *DecLearn* (99.3%), *Span_verbal* (89.2%), *CVLT_long* (58.7%) and O*rienting* (51.4%). Furthermore, the bootstrapped stepwise regression settles on the same final model as the non-bootstrapped version, underlining the consistency of these results.

Table 4 – Summary of bootstrapped stepwise regression for Brocanto accuracy. Shown are the variables and their relative frequency of being identified as a predictor as well as their relative frequency of significance.

Variable	%-Selected	%-Statistically Sign.
DecLearn	99.3	98.7
Span_verbal	89.2	86.7
CVLT_long	58.7	64.7
Orienting	51.4	64.2
Age	38.7	51.9
APM	34.8	56.0
Alerting	29.1	40.6
MLAT5	24.4	52.1
Entropy_speak	24.1	41.1
SRT	23.0	33.0
Span_visual	22.7	37.9
inhibition	20.3	39.4

3.1.4.3 Random forest

3.1.4.3.1 Random forest for Brocanto Cluster Assignment

In order to answer the question of which variables were most closely related to the assignment of participants to the learning pattern cluster identified in section 3.1.2, a random forest was grown on the data (excluding the 'fatigue' cluster).

Figure 4ai shows the results of the CPI calculation for *Brocanto* cluster assignment. The plot displays the variables in their order of importance. CPI values can in some cases be lower than zero, indicating that permuting the values of that variable randomly actually improves the model slightly. These improvements due to randomness are used here to create a positive threshold by taking the absolute value of the lowest observed CPI which is represented by the red dashed line. Variables with values to the left of it have a CPI lower than the absolute value of the lowest conditional permutation importance and can therefore be read as being about as important as randomness. Variables whose values lie to the right of it therefore have a higher CPI than what one might expect from random influence alone (Vanhove & Berthele, 2015).

DecLearn, CVLT and *APM* showed to be the standout variables influencing the predictive quality of the random forest model, further underlining the importance of especially the first two predictors.

3.1.4.3.2 Random forest for Brocanto accuracy

The same analysis was done trying to find variables involved in the prediction of overall accuracy during the *Brocanto* task. *Figure 4aii* again shows the results of this analysis in the form of a Dot-chart, with values displayed in order of importance, and the red dashed line again serving as a relevance threshold by indicating the absolute value of the lowest CPI. Although all variables which were considered relevant in the previous analysis do show up again, the picture is still a slightly different one. *DecLearn* remains the predictor with the highest overall CPI, but the *CVLT* now shows considerably lower importance relative to *DecLearn* than before. Additionally, verbal working memory span emerged as another important predictor and the *APM* are also still above the threshold of importance attributable to randomness alone.

3.1.5 ArtGram

The investigation into variables important for performance on the *ArtGram* task was conducted in the same manner as the one for *Brocanto*. Considering the shared influence of multiple predictor variables, a multiple linear regression analysis, a stepwise regression procedure and a random forest analysis were conducted. The data underlying these statistical analyses now, as mentioned before, included the data from participants assigned to the *Brocanto* 'Fatigue' cluster.

3.1.5.1 Linear Regression

Once again, the first regression model was a large linear model with all previously discussed independent variables being included in a single step as there was no *a priori* reasoning for including certain predictors before others. This resulted in a model with thirteen predictors. The Q-Q plot (*Supplementary Figure 3),* similar to before, showed some signs of non-normally distributed residuals especially at the extremes, but with a majority of data points still on the diagonal.

Out of this regression model only one predictor, namely *CVLT,* emerged as significant ($p = .023$). None of the other predictors crossed even a liberal threshold of $p < .1$, indicating verbal learning ability to be the most important variable involved in performance in *ArtGram*. The full overview of the regression results can be found in *Table 5*.

Table 5 - Results of multiple linear regression analysis with ArtGram as outcome variable and APM, DecLearn, MLAT5, CVLT_long, visual and verbal working memory span, speaking entropy, SRT, alerting, inhibition, orienting and age as predictor variables.

Variable	*B*	*SE*	*t*	*p*
Intercept	3.75	3.44	1.09	.028
APM	0.09	0.07	1.26	.211
DecLearn	0.04	0.04	0.94	.351
CVLT_long	0.21	0.09	2.32	.023*
MLAT5	0.04	0.06	0.64	.524
Span_visual	-0.05	0.14	-0.37	.710
Span_verbal	0.01	0.11	0.10	.920
Age	-0.04	0.05	-0.69	.489
SRT	-0.00	0.01	-0.52	.607
Entropy_speak	0.53	0.60	0.90	.373
Alerting	-0.02	0.02	-1.36	.177
Inhibition	0.00	0.01	0.07	.941
Orienting	-0.01	0.01	-0.49	.626

*Significance codes: 0.05 * 0.01 **

3.1.5.2 Stepwise Regression

The results from the multiple linear regression analysis are further supported by the subsequent analyses including a bootstrapped (1000 subsamples) stepwise-regression analysis, following the procedure from before (3.1.4.2).

The backwards elimination procedure aiming to minimize the AIC resulted in a model with three predictors: *CVLT* (p = .007)*,* the *alerting* (p = .063) component of the attention networks (*ANT-I*) and the *APM* (p = .104).

Table 6 gives an overview of the results of the bootstrapped stepwise regression procedure, underlining the connection of the *CVLT* to *ArtGram* performance. Out of 1000 iterations, *CVLT* remained part of the final model in 900 (90%) of them and was a significant predictor in 88.22% of cases. On the other hand, *Alerting,* as the second predictor, was selected in 54.5% of cases. Importantly, *alerting* is not a positive but a negative predictor (B = -0.02, p = .14 in the simple linear regression model and B = -0.03, p = .06 in the last step of the stepwise regression model). Statistically significant in 65.50% of bootstrap samples it was negative in nearly all models (99.5%) calculated.

Table 6 - Summary of bootstrapped stepwise regression for ArtGram accuracy. Shown are the variables and their relative frequency of being identified as a predictor as well as their relative frequency of significance.

Variable	*%-Selected*	*%-Statistically Sign.*
CVLT	90.0	88.2
Alerting	54.5	65.5
APM	52.6	61.4
DecLearn	39.5	56.2
MLAT5	38.7	58.1
Entropy_speak	34.0	49.4
SRT	28.5	42.3
Orienting	27.0	42.3
Age	26.1	44.8
Span_visual	24.6	43.1
Span_verbal	18.9	34.4
Inhibition	16.8	36.4

3.1.5.3 Random Forest

The random forest analysis for the *ArtGram* task followed the same procedure as the one for *Brocanto* accuracy and produced the results summarized in *Figure 4b.* Once more, *CVLT* emerges as the predictor with the highest importance value. Two other variables, namely *alerting* and *MLAT5,* cross the relevance threshold. Overall, the random forest analysis converges on the same variables as the regressions before, only adding the *MLAT5* (which was only excluded from the regression model in the second to last step of the backwards elimination procedure) to the pool of possibly relevant predictors for *ArtGram* performance.

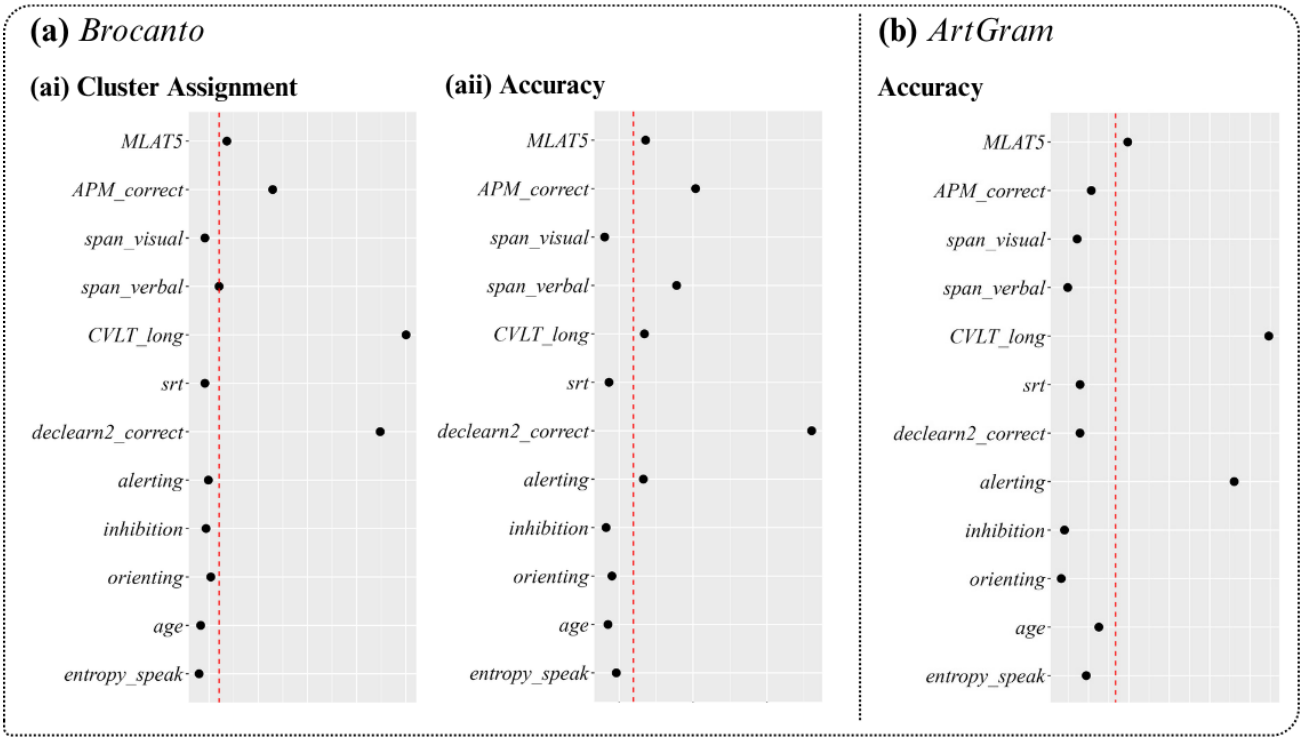


Figure 4 – Dotcharts displaying the results of conditional permutation importance calculation from the random forest analyses. The results for Brocanto (a) include the relative variable importance for cluster assignment (ai) as well as overall score (aii). Panel (b) shows the CPI for ArtGram accuracy.

3.2 Neuroimaging results

The analysis of resting-state functional connectivity as a function of grammar learning performance was carried out on a sample of 93 participants. Of the original 116 participants included in the behavioural analysis, 21 were excluded due to them being identified as part of the 'fatigue' cluster (to ensure comparability between the analysis of *Brocanto* and *ArtGram* analyses) and for another two participants fMRI imaging data could not be obtained.

3.2.1 Mean connectivity for all ROIs

Before delving into the connection between RSFC patterns and artificial grammar learning performance, this section will provide an overview of the mean connectivity pattern between the different ROIs and the rest of the brain. Generally, all ROIs exhibited functional resting-state connectivity with a multitude of areas of the brain but setting a threshold of $z = 5$ allowed us to locate regions with particularly strong coactivations. *Table 7* at the end of this section (3.2.1) shows the clusters of coactivation for each of the four ROIs as well as the region and coordinates of the highest peak of coactivation within that cluster. Due to the size of some of the clusters, particularly the first of each ROI, the coordinates/regions with peak mean connectivity are not adequate summaries of the whole cluster but should rather serve as

points of orientation. The results displayed in *Figure 5* show the mean connectivity maps in MNI152 standard space and were created in *MRIcroGL* (version number 1.2.20220720). They show the same eight axial slices as well as a 3D representation of the corresponding ROI.[2]

3.2.1.1 L pTri mean connectivity

The left pars triangularis exhibited significant resting-state functional coactivation with a host of cortical and subcortical regions as well as the cerebellum (see *Figure 5a)*. The peak of the biggest cluster was found in Crus II of the right cerebellum together with smaller groups of voxels in other parts of the right cerebellum. Although smaller in volume, clusters of high mean connectivity were also found in the left hemisphere of the cerebellum. Several cortical regions showed high levels of coactivation with lpTri, including the left superior frontal gyrus and precuneus, both inferior frontal, inferior middle temporal and occipital gyri. Regions of high coactivations were also located in the left insula, hippocampus, parahippocampus and amygdala as well as other subcortical region regions like pallidum, putamen and cingulum. Although considerable coactivation with lpTri was observed in both hemispheres, left lateralized clusters showed overall stronger measures of coactivation than their right hemisphere homologues.

3.2.1.2 R pTri mean connectivity

Mean resting-state functional connectivity between rpTri and the rest of the brain (see *Figure 5b*), showed a similar pattern as the one between lpTri and the rest of the brain. Particularly large clusters of coactivation were found in regions of the frontal, temporal and occipital lobes. Specifically, both middle occipital, middle temporal, and middle frontal gyri; the right superior temporal, superior frontal and fusiform gyri as well as the left lingual gyrus, temporal pole and inferior temporal gyrus. RSFC was also high with parts of the cerebellum although to a lesser extent than in the case of lpTri. The group of subcortical regions with high connectivity to rpTri comprised the thalamus, pallidum, putamen, caudate, hippocampus and amygdala in both hemispheres.

3.2.1.3 L pMTG mean connectivity

Analysing the group mean connectivity of the left posterior middle temporal gyrus once again revealed a large network of coactivation (see *Figure 5c)*. Particularly strong functional connectivity was once more found in large parts of the cerebellum, specifically the right Crus

[2] MRIcroGL code for reproduction: A L- H -0,1 V -0,1 -55 -40 -25 -15; -8 8 15 25; 35 45 60 S X R 1 R-1

I and II regions. The cortical RSFC network comprised almost the whole MTG in both the left and right hemispheres as well as parts of the left and right superior and inferior temporal gyri; large areas in the frontal lobe, particularly the left and right superior and middle frontal gyri, as well as the pars orbitalis of the left inferior frontal gyrus. The left and right angular gyri as well as parts of the left superior parietal gyrus also contained peaks of coactivation with lpMTG. Peaks were also found in the left and right precuneus and lingual gyri. Subcortical regions exhibiting strong connectivity were found in the left and right hippocampus, posterior and middle cingulum as well as the left caudate nucleus.

3.2.1.4 R pMTG mean connectivity

As was the case with all previous ROIs, rpMTG showed peaks of coactivation within the left and right cerebellum, particularly Crus I and II (see *Figure 5d*). The MTG, inferior temporal gyrus, superior and middle frontal gyri, inferior and superior parietal gyri, precuneus, lingual gyrus and angular gyrus were cortical areas in both the left and right hemisphere, showing strong associations with activity in rpMTG. The left and right hippocampus, middle and posterior cingulum together with the right caudate nucleus formed the set of subcortical structures showing RSFC with rpMTG.

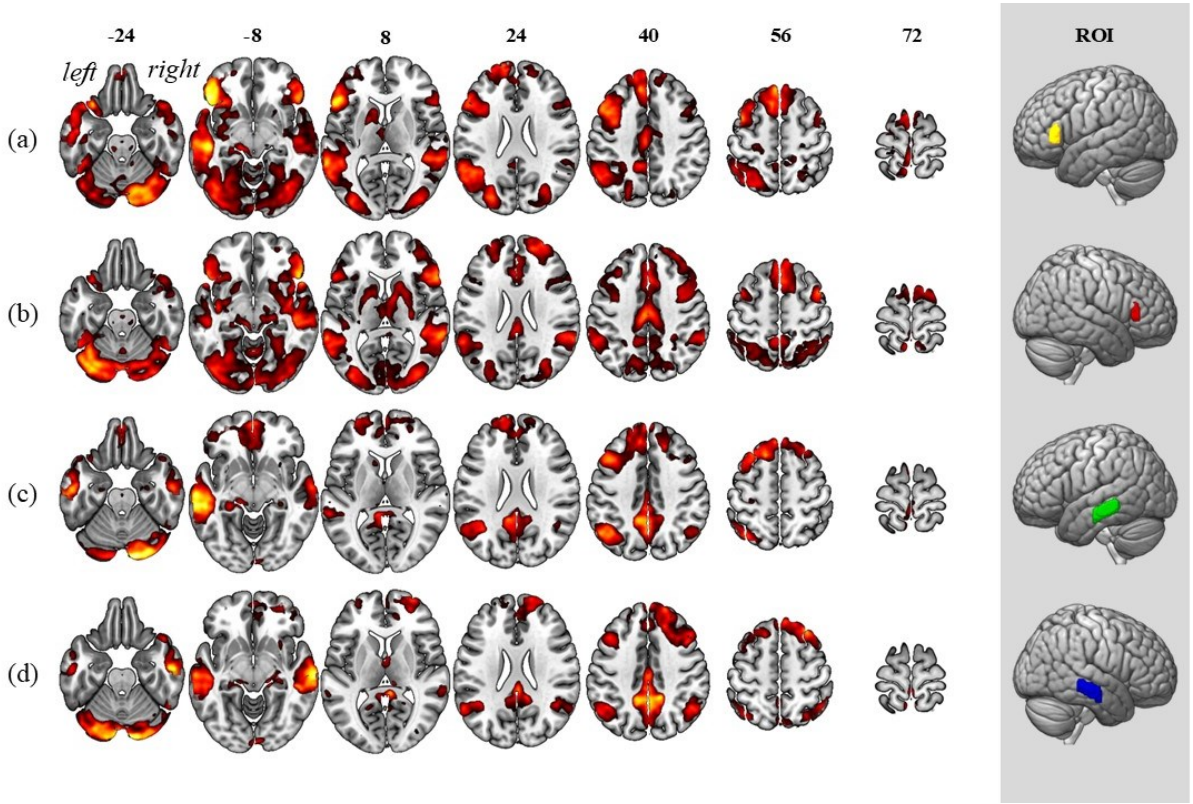


Figure 5 – Mean connectivity networks between all ROIs (shown in the grey column) and the rest of the brain. (a) Shown here is the network of coactivation with lpTri. Row (b) shows coactivation network with rpTri. Row (c) shows coactivation network with lpMTG. Row (d) shows coactivation network with rpMTG.

Table 7 – Peaks of functional connectivity with seed ROIs.

Cortical region (peak)	Size (voxels)	*p*-value	Z_{max}	L/R	Peak location (mm)		
					X	*Y*	*Z*
(a) Seed ROI: left pars triangularis							
Mean connectivity to whole brain (highest peak locations)							
Cerebellum Crus II	62235	< .001	10.4	R	26	-42	-42
Superior parietal lobule; angular gyrus	309	< .001	6.3	R	28	-54	48
Precentral gyrus	220	< .001	5.8	L	-24	-24	56
Cerebellum lobule VIIIa/b	169	< .001	7.2	L	-28	-42	-54
Putamen	96	< .001	5.7	R	20	6	-12
Thalamus	86	< .001	6.3	L	-4	-8	10
Caudate	85	< .001	6	R	16	12	12
Cerebral white matter (precentral sulcus?)	78	< .001	5.4	R	28	-22	52
Insular cortex (opercular cortex)	77	< .001	5.5	R	36	-18	18
Amygdala	36	< .001	5.8	R	24	-12	-12
(b) Seed ROI: right pars triangularis							
Mean connectivity to whole brain							
Inferior frontal gyrus, pars triangularis (large cluster with many peaks, including	69669	<.001	10.4	R	54	26	-2

Cortical region (peak)	Size (voxels)	*p*-value	Z_{max}	L/R	Peak location (mm)		
					X	*Y*	*Z*
Brain stem	271	<.001	6.4	R	56	-58	10
Inferior temporal gyrus	113	<.001	6.3	L	-46	0	-36
Cerebellum lobule V	77	<.001					
(c) Seed ROI: Left posterior middle temporal gyrus							
Mean connectivity to whole brain							
Middle temporal gyrus	18954	<.001	9.5	L	-64	-20	-10
Precuneus/Cingulate cortex	5353	<.001	8.6	L	-12	-48	-36
Cerebellum Crus II	5113	<.001	10	R	26	-86	-28
Middle temporal gyrus	2113	<.001	8.4	R	62	-12	-12
Cerebellum lobule IX	784	<.001	9.5	R	4	-56	-52
Angular gyrus	713	<.001	6.6	R	44	-58	36
Postcentral gyrus/Precuneus	199	<.001	6.3	L	-4	-44	76
Precentral gyrus	41	<.001	5.6	L	-30	-28	56
Caudate	35	<.001	5.6	L	-12	14	10
Frontal pole	28	<.001	5.5	R	34	36	-12
(d) Seed ROI: rpMTG							
Mean connectivity to whole brain							
Middle/superior temporal gyrus	13488	<.001	9.5	R	64	-22	-4
Posterior cingulate gyrus	8232	<.001	8.5	R	14	-44	38
Cerebellum Crus II	6507	<.001	9.6	R	8	-88	-26
Posterior middle temporal gyrus	2278	<.001	8.3	L	-66	-32	-4
Angular gyrus	1973	<.001	7.3	L	-48	-56	50
Cerebellum, lobule IX	867	<.001	8.4	L	-4	-56	-44
Thalamus	115	<.001	6.4	R	4	-2	8
Lateral occipital cortex	27	<.001	5.2	R	34	-80	6
Hippocampus	19	<.001	5.6	L	-11	-38	3
Cerebellum lobule VIIIb	15	<.001	6	L	-26	-42	-56

3.2.2 Connectivity predicting task performance

Next, we assessed whether the patterns of resting-state functional connectivity varied as a function of AGL task performance. For all ROIs both a negative and a positive contrast was calculated for *Brocanto* and *ArtGram* to identify clusters of coactivation that were either positively or negatively associated with task performance. The first analysis was done using the default threshold of $z = 3.1$ and yielded no significant clusters for any ROI nor any task. For this reason, the threshold was lowered to a less conservative $z = 2.3$. The results obtained from this analysis are described in the following sections and are summarized in Table 8 at

the end of this section. Correction for multiple comparisons (4 ROIs X 2 hemispheres) was done using the Bonferroni method resulting in $\alpha = .00625$.

3.2.2.1 RSFC for *Brocanto*

Performance on *Brocanto* was not associated with any significant clusters of RSFC with any of the lpTri, rpTri or rpMTG ROIs. The only RSFC network associated with overall *Brocanto* performance (accuracy) was found using the lpMTG as the seed region.

3.2.2.1.1 L pMTG

There was one cluster of functional connectivity with the lpMTG found to be associated with lower performance scores on *Brocanto*. In other words, stronger measures of RSFC in these regions were found in participants with lower scores, while participants with higher performance exhibited comparatively lower connectivity measures (see *Figure 6a*). The cluster was located predominantly in the left temporal lobe and comprises areas of the superior and middle temporal gyri, postcentral gyrus, central and parietal opercula as well as regions of the insular cortex, planum polare, planum temporale and Heschl's gyrus. Local maxima within this cluster are also located in some cerebral white matter areas adjacent to the aforementioned regions. Coordinates of both the centre of gravity and local maxima as well as the clusters size can be found in Table 8*c*. It includes peaks of connectivity with lpMTG associated with lower *Brocanto* performance ($z_{maxC1} = 3.71$) and is significant at the cluster level ($p < .001$).

3.2.2.2 RSFC for *ArtGram*

The same analysis attempting to connect distinguishable RSFC networks with task performance was done using participants' *ArtGram* scores. Clusters of functional connectivity associated with *ArtGram* performance were identified with two of four ROIs as seed regions: lpMTG and rpMTG.

3.2.2.2.1 L pMTG

Taking the lpMTG as the seed region, one significant cluster ($p = .002$) comprising 660 voxels cantered around the right pallidum, was found to be positively associated with *ArtGram* performance (see *Figure 6b*). Increased connectivity between lpMTG and the regions of this cluster was found in participants with higher task scores. The cluster identified here exhibited local maxima in the pallidum ($z = 4.4$), caudate ($z = 3.7$), thalamus ($z = 3.5$) and areas of cerebral white matter ($z = 3.2$) the exact locations of which can be found in *Table* 8*c*.

3.2.2.2.2 R pMTG

There was one cluster found to show increased functional connectivity with the right pMTG as a function of increased *ArtGram* performance (see *Figure 6b*). This cluster comprised 833 voxels centered around the right pallidum ($z = 4$) and was significant at a level of $p < .001$. Like the RSFC cluster identified with the lpMTG as the seed region, this cluster also showed local maxima in the areas of the caudate ($z = 3.8$), cerebral white matter ($z = 3.8$), thalamus ($z = 3.7$) and putamen ($z = 3.2$). The exact locations of the clusters centre of gravity and local maxima are summarized in *Table* 8*d*.

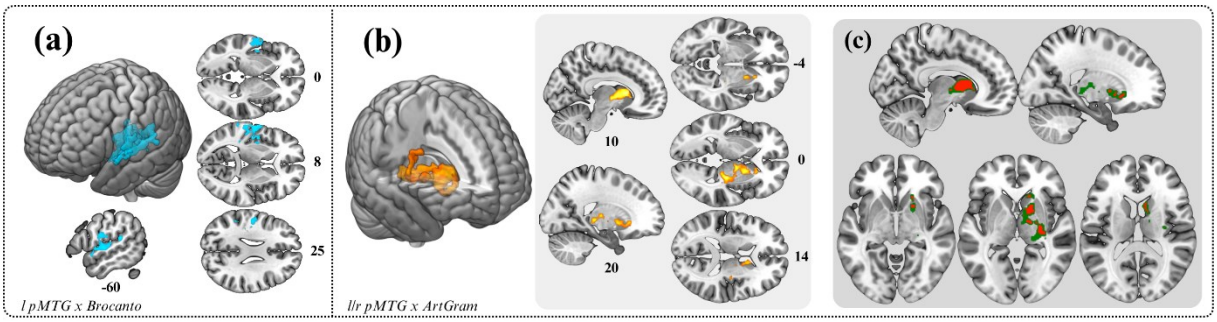


Figure 6 – Summary of connectivity results predicting artificial grammar learning task performance. (a) Displayed here are clusters of increased connectivity using left pMTG as a seed ROI predicting worse performance on Brocanto*. Panel (b) displays brain regions showing increased functional connectivity with either the left or right pMTG that are associated with better performance on the* ArtGram *task. Panel (c) shows the overlap of* ArtGram *performance associated coactivation of left and right pMTG. Green areas indicate regions uniquely connected to either the left or right pMTG (but not both), while red regions reflect overlapping clusters where increased connectivity with both the left and right pMTG was positively associated with* ArtGram *performance.*

Table 8 – Summary of RSFC in relation to task performance. The table shows the results of the seed-based connectivity analysis in relation to Brocanto *and* ArtGram *performance. Within the section for each ROI the results are presented first for RSFC modulated by* Brocanto *and then by* ArtGram *and they are further divided into the positive and negative contrasts. Shown are the sizes and corrected p-values for significant clusters. Image coordinates in the same row as the cluster Index are coordinates for the clusters centre of gravity, while the subsequent coordinates denote locations of within cluster maxima. (Null results are represented by '-'.)*

Cortical region (peak)	**Size (voxels)**	***p*-value**	***Z* max**	**L/R**	**Location (mm)**		
					X	***Y***	***Z***
(a) Seed ROI: left pars triangularis							
Connectivity modulated by *Brocanto* performance							
Positive contrast							
-	-	-	-	-	-	-	-
Negative contrast							
-	-	-	-	-	-	-	-
Connectivity modulated by *ArtGram* performance							
Positive contrast							
-	-	-	-	-	-	-	-
Negative contrast							

Cortical region (peak)	Size (voxels)	*p*-value	Z_{max}	L/R	Location (mm)		
					X	*Y*	*Z*
-	-	-	-	-	-	-	-
(b) Seed ROI: right pars triangularis							
Connectivity modulated by *Brocanto* performance							
Positive contrast							
-	-	-	-	-	-	-	-
Negative contrast							
-	-	-	-	-	-	-	-
Connectivity modulated by *ArtGram* performance							
Positive contrast							
-	-	-	-	-	-	-	-
Negative contrast							
-	-	-	-	-	-	-	-
(c) Seed ROI: Left posterior middle temporal gyrus							
Connectivity modulated by *Brocanto* performance							
Positive contrast							
-	-	-	-	-	-	-	-
Negative contrast							
1 left middle temporal gyrus (CoG)	1489	< .001	3.76	L	-51	-15	-9
Planum polare			3.76	L	-46	-6	-4
Cerebral white matter			3.74	L	-48	-14	24
Central opercular cortex			3.69	L	-62	-8	10
Planum temporale			3.63	L	-48	-40	18
Planum polare/insular cortex			3.5	L	-44	-2	-8
Connectivity modulated by *ArtGram* performance							
Positive contrast							
1 right Pallidum (CoG)	660	.002	4.4	R	18	0	6
Caudate			3.7	R	10	16	8
Thalamus			3.5	R	16	-8	2
Cerebral white matter			3.2	R	24	-10	26
Pallidum			3.1	R	24	-8	2
Negative contrast							
-	-	-	-	-	-	-	-
(d) Seed ROI: rpMTG							
Connectivity modulated by *Brocanto* performance							
Positive contrast							
-	-	-	-	-	-	-	-
Negative contrast							
-	-	-	-	-	-	-	-
Connectivity modulated by *ArtGram* performance							
Positive contrast							
1 Right Pallidum (CoG)	833	< .001	4	R	18	-4	4

Cortical region (peak)	Size (voxels)	*p*-value	Z_{max}	L/R	Location (mm)		
					X	*Y*	*Z*
Pallidum			4	R	14	4	2
Caudate			3.8	R	10	14	8
Cerebral white matter (retrolenticular part of internal capsule)			3.8	R	26	-22	0
Thalamus			3.7	R	14	-6	2
Putamen			3.2	R	32	-22	4
Negative contrast							
-	-	-	-	-	-	-	-

4 Discussion

The objective of this thesis was to identify variables associated with successful grammar learning, both from neurofunctional and behavioural perspectives. The question that was posed was whether a hypothesised distinction in grammatical processing networks (Matchin & Hickok, 2020) would manifest as a difference in variables associated with successful grammar acquisition at either of these levels.

4.1 Result interpretation

4.1.1 Behavioural

Overall, the behavioural results identified verbal learning ability as important for success in both AGL tasks but also revealed differences in association with other variables. *Brocanto* performance was consistently associated with declarative learning ability, verbal working memory span and the orienting component of the ANT-I. In contrast, *ArtGram* demonstrated a consistent correlation with the CVLT. This association of both AGL tasks with verbal learning skills is consistent with the splitting of the language system into a declarative and a procedural memory system, proposed by Ullman (2001). While this declarative/procedural model of language generally assigns the acquisition of grammatical rules to the procedural memory system, it still predicts the earlier stages of second language grammar learning to depend in large part upon the declarative memory system. This is because the gradual proceduralisation of grammar itself is a process which relies on information that is quickly acquired by the declarative memory system (Ullman, 2016).

It is interesting to note that *ArtGram* performance was also associated with an attentional component, the alerting network. However, the direction of this association was reversed. A higher score on alerting, meaning greater effects of an alerting cue, came with lower *ArtGram* performance. The random forest analysis identified additional variables that

had not previously been identified by the regression analyses. For *Brocanto*, it revealed APM, the measure for pattern finding and fluid intelligence, to be predictive of performance, and for *ArtGram*, rote learning ability measured by *MLAT5* was found to be predictive.

In the stepwise elimination procedure for *ArtGram*, MLAT5 was selected in only 35.2% of Bootstrapped samples, yet it demonstrated statistical significance in 54.26% of cases. In the vast majority of cases (98.86%), the relationship with *ArtGram* was positive, thus providing support for the assumption that the dependence is not random but was possibly overlooked in regression analysis due to multicollinearity with *CVLT* ($r = .4$, $p < .001$). It is noteworthy that the two variables identified exclusively by random forests correspond to the distinct cognitive demands of *ArtGram* and *Brocanto*. APM can be conceptualised as a metric of inductive reasoning capability, or, more broadly, fluid intelligence. In contrast, *MLAT5*, in conjunction with CVLT, underscores the significance of word learning and memory in the *ArtGram* task.

The differential involvement of the attentional components in the success of *Brocanto* and *ArtGram* is a phenomenon that is more difficult to explain. Goranskaya et al. (2016) employed a simple AGL task in which four syllables were presented auditorily, following a predefined ordering rule. The successful rule-learning strategies that were reported relied on either the confirmation of a rule-based prediction ("This sound should follow according to the rule I learned") or the memorization of the first part of the structure and its matching to the following part. Non-learners reported experiencing difficulty in ignoring inconsequential information. This finding suggests that directing attention towards specific information plays a crucial part in one's success at this task, which, similar to *Brocanto*, testing implicit grammar learning. The importance of orienting for *Brocanto* can thus be understood as the ability to direct attention to those parts of the sentence that contain the relevant information necessary for making a correct grammaticality judgement.

ArtGram, which used a multiple-choice translation test as its way of measuring explicit morphosyntactic rule learning, might also depend on the ability to direct attention to specific aspects of a presented sentence but even more on the executive attentional network responsible for resolving conflict between competing information. The alerting network has been demonstrated to exert a detrimental effect on the inhibitory function of the executive attention networks (Callejas et al., 2005). Assessments of the interactions between networks found that activation of the alerting network was associated with a higher congruency effect. In other words, if an alerting sound was presented before stimulus onset, resolving the conflict

between relevant and irrelevant information would be harder. When applied to the *ArtGram* findings, it could be argued that higher scores in alerting corresponded to a heightened sensitivity to all the different information present in the multiple-choice translations. Thereby alerting could possibly interfere with the executive attentional network (inhibition) and could disturb the resolution of conflict between the different options.

Despite the established influence of age on language learning outcomes and the neurofunctional correlates of language learning (Antonenko et al., 2012), its role in this particular sample was not significant. This phenomenon is most likely attributable to the distribution of ages being skewed towards the younger (adult) age range.

4.1.2 RSFC

4.1.2.1 Mean connectivity

All ROIs showed an extensive network of coactivation with specific differences being hard to pin down. Generally, the networks are characterized by their overlap in the areas of the middle temporal gyri, large parts of the frontal lobe and parts of the cerebellum. Notable differences between the networks of pTri and pMTG arise mostly in occipital areas, to which pTri seems to be more strongly connected, and medial cortical areas like the precuneus which show stronger coactivation with the pMTG.

4.1.2.2 Predicting performance

As with the analysis of the behavioural data, examination of the neuroimaging data aimed at identifying coactivation networks predicting grammar learning performance. Of special interest was once again the question of whether, given such networks could be identified, these would involve different regions depending on which AGL task's performance (*ArtGram* or *Brocanto*) they predicted.

For both tasks, RSFC networks between at least one ROI and the rest of the brain were identified to be significantly related to task performance. This confirmed the possibility of partially predicting grammar learning ability based on resting-state activation patterns, similar to predicting other language related abilities using RSFC (Rabini et al., 2023).

4.1.2.2.1 Brocanto

Of the four ROIs only the connectivity network of the lpMTG was predictive of *Brocanto* performance. One cluster was found to be significantly negatively related to overall accuracy on the grammar learning task, which demanded the application of implicitly learned hierarchical syntactic rules. The regions of this cluster (*Figure 6c*), centring around the

lpMTG and containing activations in superior and middle temporal gyrus, postcentral gyri, central and parietal opercula as well as regions of the insular cortex, planum polare, planum temporale and Heschl's gyrus, are commonly associated with language or somatosensory related tasks. The superior temporal gyrus (STG), under the model of cortical organization of syntax by Matchin and Hickok (2020), fulfils the role of auditory phonological processing. The posterior portion of the STG is highlighted in particular as a hub for both phonological decoding necessary for speech perception and planning of motor actions involved in speech production by providing targets for sounds to be produced (Matchin & Hickok, 2020). According to the model, pSTG is connected to a sensorimotor network involved in phoneme processing and motor planning. This network is well represented in these results with coactivation in Heschl's gyrus, planum temporale and planum polare. Within this network, Heschl's gyrus, which includes the primary auditory cortex, has been shown to structurally be highly variable among individuals (Dalboni Da Rocha et al., 2020), and related to phoneme learning (Golestani et al., 2006), to individual variability in language learning (Ramoser et al., 2025; Turker et al., 2019) and to multilingualism (Kepinska et al., 2025). Furthermore, the planum temporale, specifically its posterior part, the Sylvian parietal temporal area, has been implicated as an auditory motor interface (Buchsbaum et al., 2005; Matchin & Hickok, 2020). The coactivation of parts of the insula could also be relevant to understand these findings, as the insula has been proposed to serve a coordinating function between the language understanding and production systems surrounding it (Ardila et al., 2016).

The remaining areas, mainly those located in the central and parietal portions of the opercular cortex, are associated with somatosensory perception, with the former being mainly associated with sensations in the oropharynx (Mălîia et al., 2018).

One interesting observation is that the general relationship between RSFC strength and *Brocanto* task performance is negative, indicating better performance in participants who did not exhibit a more widespread coactivation of nearby cortical regions with lpMTG. Neural efficiency is a phenomenon observed in many studies (for an overview, see Neubauer & Fink, 2009b), where individuals with higher cognitive ability seem to recruit fewer neural resources while working on a task than individuals with lower ability. Studies investigating neural efficiency suggest that individuals with higher cognitive ability exhibit both lower activation of fronto-parietal regions during task processing, and that they show stronger network connectivity to closer brain regions (Neubauer & Fink, 2009a), also at rest (Langer et al., 2011). The negative association of a more widespread coactivation network around the

lpMTG with worse performance on *Brocanto*, therefore, could be interpreted as evidence of neural efficiency in the resting brain being a marker for high ability.

Applying the neural efficiency framework to resting-state functional connectivity, however, comes with limitations. Neural efficiency is typically studied in the context of task-related activity, where it is possible to directly measure the relationship between cognitive performance and resource allocation. In contrast, RSFC reflects the brain's intrinsic network architecture, which may not directly correspond to efficiency during task execution. Even in task-based studies neural efficiency has been found to be dependent on factors like task difficulty (Nussbaumer et al., 2015). Neural efficiency effects are observed mostly in the range of moderately difficult tasks, with the relationship between intelligence and neural resource recruitment reversing at high levels of task difficulty (Doppelmayr et al., 2005; Neubauer & Fink, 2009b) and differences not manifesting at very low levels of task difficulty (Neubauer et al., 1999). Learning and training also influence neural efficiency in interaction with intelligence (Neubauer & Fink, 2009b). Brighter individuals have been shown to profit more from training and exhibit a higher gain in efficiency from training, when compared to participants with lower intelligence scores (Haier et al., 1992; Neubauer et al., 2004). Furthermore, in the early stages of learning, individuals with higher cognitive ability recruit more neural resources before efficiency gains emerge (Kepinska et al., 2017a; Neubauer & Fink, 2009).

These interdependencies between neural efficiency and factors like task difficulty or training make it difficult to derive clear, falsifiable predictions for how neural efficiency would manifest in the resting brain. One could argue that neural efficiency at rest should simply be reflected in a less widespread network of coactivation. Conversely, one could also hypothesize that a more widespread network at rest reflects a more flexible and adaptable preparatory state needed during the early learning phase or when facing more difficult tasks, ultimately leading to a more efficient recruitment of neural resources over time.

Despite these limitations, interpreting RSFC findings through the lens of neural efficiency opens interesting avenues for further research. For instance, if neural efficiency is relevant even in the absence of task engagement, one would expect more spatially localized and functionally modular RSFC patterns to predict cognitive ability across domains. Furthermore, investigating whether RSFC in language-related regions is more predictive of grammar learning ability than RSFC in non-language regions could help determine whether

language learning depends on general cognitive efficiency or on a more domain-specific neural architecture.

4.1.2.2.2 ArtGram

Performance on *ArtGram* was positively related to coactivation of a host of striatal regions with both the left and right pMTG. The results obtained using the pMTG as a seed region demonstrate a substantial overlap for both the left and right pMTG (see Figure 6d), and consist of a cluster predominantly involving basal ganglia regions and the thalamus. Increased connectivity between pMTG and the aforementioned regions, which have been described as being part of a procedural learning system (Ullman, 2004), was associated with better performance in *ArtGram*. The caudate, putamen and pallidum, via the thalamus, have been found to possess a multitude of connections to cortical areas. While the putamen is primarily part of a circuit that subserves automatic processing of motor sequences and connects to areas of the motor cortex, the circuit involving the caudate, with its connections to prefrontal areas, is involved in associative cognitive tasks (Middleton & Strick, 2000). Successful *ArtGram* participants show a stronger functional connection between the areas involved in sentence structure processing (pMTG) and a procedural/rule learning hub which has projections into prefrontal areas involved in language and working memory (Ullman, 2004).

Taken together with the fact that in the analysis of the behavioural data, no connection was found between *ArtGram* performance and our chosen measure of procedural learning, this finding is somewhat surprising. If good performance in *ArtGram* is predicted by strong coactivation of the pMTG and areas of a procedural learning hub, it would be at least reasonable to assume some correlation in the behavioural realm as well. Those expectations not coming to fruition in this thesis, might result from the nature of the operationalization of procedural memory, the serial reaction time task (SRT). While this task requires procedural sequence learning, it does not involve language. Procedural memory tested in this way might just not generalize across all domains (movement, language, music, etc.) and the SRT might be a measure suited to assess one aspect of procedural memory, but it might fall short of measuring others.

It can thus be concluded that enhanced performance *ArtGram* is contingent on the robust integration of the hierarchical syntax network in the pMTG with the regions responsible for procedural rule learning in the striatum.

4.2 Big picture integration with Matchin & Hickok model

One of the goals of this project was to assess the model of cortical organization of syntax by Matchin and Hickok in the context of resting-state connectivity and inter-individual variability in grammar learning. Looking at the mean connectivity maps between the ROIs, they align with its predictions. According to the model, the pMTG is integral to both syntactic comprehension and production, while the pTri specifically translates hierarchical representations into sequences for speech. This asymmetry is evident in the connectivity networks presented in this thesis. Both the pMTG and pTri networks show bilateral co-activation in the MTG; however, pMTG activity is not consistently linked to inferofrontal/pTri activation. This pattern reflects the different computational demands of syntactic production versus comprehension.

Importantly, our results stem from resting-state functional connectivity analyses, meaning that they capture the brain's intrinsic network organization rather than task-evoked responses. This intrinsic architecture likely represents a predisposition for how language processing networks are organized: the consistent coupling between pTri and pMTG suggests a stable pattern of coactivation that may prime the system for efficient syntactic production when needed. Furthermore, the autonomous operation of pMTG in the absence of consistent pTri engagement supports the idea that pMTG serves a dual role in both comprehension and production, whereas pTri is more specialized for production-related functions. Thus, even in the resting-state, the distinct connectivity profiles lend empirical support to the production/comprehension asymmetry proposed by Matchin and Hickok, underscoring that these differences are not merely a function of task demands but are embedded in the brain's baseline architecture.

The analysis of RSFC networks predicting task performance once again shows this asymmetry. Even before correcting for multiple comparisons, meaning in less conservative results not reported in this thesis (see *Supplementary Table 3*), performance on the hierarchical syntax task (*Brocanto*) was only associated with one network of coactivation with the lpMTG. Morphosyntactic learning (*ArtGram*) on the other hand was predicted by coactivation with both the pTri (only significant before familywise error correction) and pMTG as seed regions. This lends further credence to the assumption that *Brocanto* and *ArtGram* do indeed target different aspects of syntax processing.

First, successful strategies for the *ArtGram* task might arguably contain some elements of syntax production. Although neither *Brocanto* nor *ArtGram* specifically call for production,

strategies for how participants tackle *ArtGram* might involve more of these processes. This is because *ArtGram* compared to *Brocanto* allows for the construction of meaningful sentences and because it is a translation task, one could generate the correct answer to an item without even needing the multiple-choice options. One strategy for *ArtGram* therefore (see 4.1.1), might be to actively produce the translated sentence in the artificial language and only then compare this prediction to the options provided within the task. From this it could be argued that even though neither *ArtGram* nor *Brocanto* directly involve the production of language, *ArtGram* might at least lend itself better to strategies involving some language production. This in turn could be the reason as to why performance in this task was predicted by connectivity with lpTri, the region involved in production only according to Matchin and Hickok (2020), while performance in *Brocanto* was not.

Second, the tasks are designed to target different levels of syntactic processing. While *Brocanto* measures hierarchical rule acquisition, *ArtGram* targets the correct understanding and use of morphosyntactic rules. Understanding these differences between the tasks, further supports Matchin and Hickok's (2020) model. While pMTG, involved in both production and comprehension and responsible for hierarchical syntax, is the seed region producing networks predictive of *Brocanto, ArtGram* can be predicted with seeds in both pMTG and pTri.

Summary

The aim of this thesis was, first, to contribute to the research investigating the behavioural and neural correlates of successful grammar learning. Second, it was to look for and examine the differences of these correlates between successful acquisition of hierarchical lexical-syntax and morphosyntax. The data analysed herein points to some shared cognitive foundations for these different tasks in the form of significant correlations with declarative and rote learning abilities. Further analysis, however, highlights the different demands placed on the learners, with *Brocanto* being associated more with verbal working memory, declarative learning and pattern identification, and *ArtGram* showing stronger ties to rote learning and verbal learning abilities. Differences on the behavioural level were also found regarding the possible role of attention. Finally, clearly distinguishable RSFC networks were found to be predictive of performance in these tasks.

Limitations and Future Research

Some of the methods used in this thesis to identify behavioural variables relevant to success in *ArtGram* and *Brocanto* have been subject to criticism in past research. The stepwise regression procedure has seen widespread use as a tool for model construction, a task for which it is generally not suitable (see Whittingham et al., 2006). It should therefore be noted that the variables identified to be important in grammar learning in this thesis should not be taken as either an exhaustive list or as a foundation for a solid theoretical model. Instead, these results can be used as a starting point for further research and should be interpreted predominantly as limited evidence for the existence of different cognitive requirements for *Brocanto* and *ArtGram.*

Furthermore, the interpretation of variables identified as important by random forest analysis is limited by the fact that the model functions as a black box designed primarily to maximize predictive power. While random forests can highlight potential candidate variables for models aimed at understanding a process, they do not reveal how these variables interact. Moreover, the notion of variable importance itself is not well defined, as it can refer either to predictive or to explanatory importance (Grömping, 2009).

Both *ArtGram* and *Brocanto* were in some way related to working memory. Although a causal relationship of better working memory to better grammar learning is easy to imagine, it should be noted that not only did this thesis not test any relationship for causation, but there have also been arguments made for the interaction between verbal working memory span and language ability to be more complex (see Kidd et al., 2018).

Grammar learning success in this thesis was measured using two different AGL tasks. Although a pilot study that was performed during the process of developing the *ArtGram* task revealed that it measures some aspect of grammar learning that is different than that measured by *Brocanto*, more should be learned about the ways in which they differ exactly. On the one hand there is the difference in the type of grammatical system, but on the other hand there is also the difference in learning paradigms, since *ArtGram,* unlike *Brocanto,* is designed as an explicit grammar learning task. These differences on separate levels make it more difficult to compare the two tasks and should be kept in mind when looking at the correlated variables found in this thesis (for the difference between explicit and implicit learning see for example: Opitz & Hofmann, 2015).

The limitations outlined above could be remedied by future research. By taking the exploratory findings of this thesis and deriving testable predictions, the differences between morphosyntactic and hierarchic lexical grammar learning could be elucidated further.

Furthermore, instead of focusing on the connectivity of certain regions to the rest of the brain, future studies could also take a broader approach and try to predict grammar learning success using whole brain resting-state networks. Studies taking a similar ROI based approach could also benefit from a more refined definition of the selected seed regions. In their model, Matchin and Hickok (2020), for example, describe the relevant portion of the pMTG as lying between the lateral extent of Heschl's gyrus and the end of the sylvian fissure. With Heschl's gyrus exhibiting considerable individual differences in structure, ROI definition could benefit from tools such as TASH (Dalboni Da Rocha et al., 2020) in order to take this variability into account.

One interesting prospect for future research could also be to dive deeper into the different stages of grammar learning and the strategies used during rule acquisition, also with the use of resting-state fMRI data. This could not only provide even more insight into the inner workings of neurofunctional bases of language processing but also add to the limited understanding of what RSFC means and how to interpret it. For example, investigating whether increased connectivity between lpMTG and the hippocampus can predict how long it takes for people to switch from similarity-based learning to rule-based learning (Opitz & Friederici, 2003), or studying which learning strategy people prefer when faced with an AGL task (Goranskaya et al., 2016), could help in gaining a clearer understanding of how individual differences in RSFC relate to behaviour. It might be the case that increased inherent connectivity between pMTG and the hippocampus goes along with a more readily available language association system. This in turn could increase the time it takes to switch to rule based learning since the path to similarity is easier; or it could mean that similarity learning works more efficiently, enabling people to attain higher scores by relying on it while possibly foregoing the additional performance increase rule-based processing would yield (Opitz & Hofmann, 2015). This question could be answered using a similar approach to this thesis by for example comparing the RSFC networks across the different runs and in the different groups identified in the *Brocanto* cluster analysis.

By not only trying to predict performance but actually predicting behaviour in the sense of which learning strategy will be chosen by people, we could gain broader insight into how thought processes and personal tendencies may be reflected in the functional architecture of the brain.

5 References

Akaike, H. (2011). Akaike's information criterion. *International Encyclopedia of Statistical Science*, 25–25.

Antonenko, D., Meinzer, M., Lindenberg, R., Witte, A. V., & Flöel, A. (2012). Grammar learning in older adults is linked to white matter microstructure and functional connectivity. *NeuroImage*, 62(3), 1667–1674. https://doi.org/10.1016/j.neuroimage.2012.05.074

Ardila, A., Bernal, B., & Rosselli, M. (2016). How Localized are Language Brain Areas? A Review of Brodmann Areas Involvement in Oral Language. *Archives of Clinical Neuropsychology*, *31*(1), 112–122. https://doi.org/10.1093/arclin/acv081

Arthur, D. T., Ullman, M. T., & Earle, F. S. (2021). Declarative Memory Predicts Phonological Processing Abilities in Adulthood. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.658402

Austin, P. C., & Tu, J. V. (2004). Bootstrap Methods for Developing Predictive Models. *The American Statistician*, *58*(2), 131–137.

Bradley, P. S., & Fayyad, U. M. (1998). Refining Initial Points for K-Means Clustering. *Proceedings of the Fifteenth International Conference on Machine Learning*, 91–99.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buchsbaum, B. R., Olsen, R. K., Koch, P. F., Kohn, P., Kippenhan, J. S., & Berman, K. F. (2005). Reading, hearing, and the planum temporale. *NeuroImage*, *24*(2), 444–454. https://doi.org/10.1016/j.neuroimage.2004.08.025

Callejas, A., Lupiàñez, J., Funes, M. J., & Tudela, P. (2005). Modulations among the alerting, orienting and executive control networks. *Experimental Brain Research*, *167*(1), 27–37. https://doi.org/10.1007/s00221-005-2365-z

Carroll, J. B. (1962). The Prediction of Success In Intensive Foreign Language Training. *In Training Research and Education* (pp. 87–136). University of Pittsburg Press.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test* (p. 27). Psychological Corporation.

Chen, Y., Li, L., Wang, M., & Wang, R. (2022). Which Cognitive Factors Predict L2 Grammar Learning: Cognitive Control, Statistical Learning, Working Memory, or Attention? *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.943988

Chin, T.C., Coutinho, E., Sherer, K.R., & Rickard, N.S. (2018). MUSEBAQ: A modular tool for music research to assess musicianship, musical capacity, music preferences, and motivations for music use. *Music Perception*, 35(3), 376-399. https://doi.org/10.1525/MP.2018.35.3.376

Chou, Y. -h., Panych, L. P., Dickey, C. C., Petrella, J. R., & Chen, N. (2012). Investigation of Long-Term Reproducibility of Intrinsic Connectivity Network Mapping: A Resting-State fMRI Study. *American Journal of Neuroradiology*, 33(5), 833–838. https://doi.org/10.3174/ajnr.A2894

Cohen, R. A., Marsiske, M. M., & Smith, G. E. (2019). Neuropsychology of aging. *Handbook of Clinical Neurology*, *167*, 149–180. https://doi.org/10.1016/B978-0-12-804766-8.00010-8

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International*, 34(2-B), 891.

Dalboni Da Rocha, J. L., Schneider, P., Benner, J., Santoro, R., Atanasova, T., Van De Ville, D., & Golestani, N. (2020). TASH: Toolbox for the Automated Segmentation of Heschl's gyrus. *Scientific Reports*, *10*(1), 3887. https://doi.org/10.1038/s41598-020-60609-y

Dale, P. S., & Sparks, R. L. (2023). Mapping the predictive role of MLAT subtests for L2 achievement through regression commonality analysis. *Studies in Second Language Acquisition*, 1–25. https://doi.org/10.1017/S0272263123000530

Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences*, 103(37), 13848–13853. https://doi.org/10.1073/pnas.0601417103

Debeer, D., & Strobl, C. (2020). Conditional permutation importance revisited. *BMC Bioinformatics*, *21*(1), 307. https://doi.org/10.1186/s12859-020-03622-2

Deldar, Z., Gevers-Montoro, C., Khatibi, A., & Ghazi-Saidi, L. (2020). The interaction between language and working memory: A systematic review of fMRI studies in the past two decades. *AIMS Neuroscience*, *8*(1), 1–32. https://doi.org/10.3934/Neuroscience.2021001

Delis, D. C., Freeland, J., Kramer, J. H., & Kaplan, E. (1988). Integrating clinical assessment with cognitive neuroscience: Construct validation of the California Verbal Learning Test. *Journal of Consulting and Clinical Psychology*, *56*(1), 123.

Delis, D., Kramer, J., Kaplan, E., & Ober, B. (1987). The California verbal learning test. *The Psychological Corporation, San Antonio (TX)*.

Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, *45*(2), 265–282. https://doi.org/10.1111/j.2044-8317.1992.tb00992.x

Deweer, B., Poitrenaud, J., Kalafat, M., & Van Der Linden, M. (2008). California Verbal Learning Test. *Editions Du Centre de Psychologie Appliquée (ECPA).*

Doppelmayr, M., Klimesch, W., Hödlmoser, K., Sauseng, P., & Gruber, W. (2005). Intelligence related upper alpha desynchronization in a semantic memory task. *Brain Research Bulletin*, *66*(2), 171–177. https://doi.org/10.1016/j.brainresbull.2005.04.007

Dörnyei, Z., & Skehan, P. (2003). Individual Differences in Second Language Learning. *The handbook of second language learning,* 589-630.

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P. C. M. (2016). The Relationship Between Artificial and Second Language Learning. *Cognitive Science*, 40(4), 822–847. https://doi.org/10.1111/cogs.12257

Fan, J., Mccandliss, B., Sommer, T., Raz, A., & Posner, M. (2002). Testing the Efficiency and Independence of Attentional Networks. *Journal of Cognitive Neuroscience*, *14*, 340–347. https://doi.org/10.1162/089892902317361886

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(1), 529–534. https://doi.org/10.1073/pnas.012611199

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, 8(9), Article 9. https://doi.org/10.1038/nrn2201

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2006). Brain Structure Predicts the Learning of Foreign Speech Sounds. *Cerebral Cortex*, *17*(3), 575–582. https://doi.org/10.1093/cercor/bhk001

Goranskaya, D., Kreitewolf, J., Mueller, J. L., Friederici, A. D., & Hartwigsen, G. (2016). Fronto-Parietal Contributions to Phonological Processes in Successful Artificial Grammar Learning. *Frontiers in Human Neuroscience*, 10. https://www.frontiersin.org/articles/10.3389/fnhum.2016.00551

Grodzinsky, Y., Pieperhoff, P., & Thompson, C. (2021). Stable brain loci for the processing of complex syntax: A review of the current neuroimaging evidence. *Cortex*, 142, 252–271. https://doi.org/10.1016/j.cortex.2021.06.003

Grömping, U. (2009). Variable Importance Assessment in Regression: Linear Regression versus Random Forest. *The American Statistician*, *63*(4), 308–319. https://doi.org/10.1198/tast.2009.08199

Haier, R. J., Siegel, B. V., MacLachlan, A., Soderling, E., Lottenberg, S., & Buchsbaum, M.S. (1992). Regional glucose metabolic changes after learning a complex visuospatial/motor task: A positron emission tomographic study. *Brain Research*, *570*(1–2), 134–143. https://doi.org/10.1016/0006-8993(92)90573-R

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108. https://doi.org/10.2307/2346830

Hedenius, M. (2013). *Procedural and Declarative Memory in Children with Developmental Disorders of Language and Literacy*. https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-204245

Henson, R. (2005). A Mini-Review of fMRI Studies of Human Medial Temporal Lobe Activity Associated with Recognition Memory. *The Quarterly Journal of Experimental Psychology Section B*, *58*(3–4b), 340–360. https://doi.org/10.1080/02724990444000113

Kepinska, O., De Rover, M., Caspers, J., & Schiller, N. O. (2017a). On neural correlates of individual differences in novel grammar learning: An fMRI study. *Neuropsychologia*, *98*, 156–168. https://doi.org/10.1016/j.neuropsychologia.2016.06.014

Kepinska, O., De Rover, M., Caspers, J., & Schiller, N. O. (2017b). Whole-brain functional connectivity during acquisition of novel grammar: Distinct functional networks depend on language learning abilities. *Behavioural Brain Research*, *320*, 333–346. https://doi.org/10.1016/j.bbr.2016.12.015

Kepinska, O., Rocha, J. D. da, Tuerk, C., Hervais-Adelman, A., Bouhali, F., Green, D., Price, C. J., & Golestani, N. (2023). Auditory cortex anatomy reflects multilingual phonological experience. *eLife*, *12*. https://doi.org/10.7554/eLife.90269.1

Kessels, R. P., Van Zandvoort, M. J., Postma, A., Kappelle, L. J., & De Haan, E. H. (2000). The Corsi block-tapping task: Standardization and normative data. *Applied Neuropsychology*, *7*(4), 252–258.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171–184. https://doi.org/10.1037/a0025405

Kidd, E., & Arciuli, J. (2016). Individual Differences in Statistical Learning Predict Children's Comprehension of Syntax. *Child Development*, *87*(1), 184–193. https://doi.org/10.1111/cdev.12461

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Langer, N., Pedroni, A., Gianotti, L. R. R., Hänggi, J., Knoch, D., & Jäncke, L. (2011). Functional brain network efficiency predicts intelligence. *Human Brain Mapping*, *33*(6), 1393–1406. https://doi.org/10.1002/hbm.21297

Li, S. (2015). The Associations Between Language Aptitude and Second Language Grammar Acquisition: A Meta-Analytic Review of Five Decades of Research. *Applied Linguistics*, *36*(3), 385–408. https://doi.org/10.1093/applin/amu054

Li, S. (2019). Six Decades of Language Aptitude Research. In Wen, Z., Skehan, P., Biedroń, A., Li, S., & Sparks, R.L. (Eds.). (2019). *Language Aptitude: Advancing Theory, Testing, Research and Practice* (1st ed.). Routledge. (pp.78-96) https://doi.org/10.4324/9781315122021-5

Mălîia, M.-D., Donos, C., Barborica, A., Popa, I., Ciurea, J., Cinatti, S., & Mîndruţă, I. (2018). Functional mapping and effective connectivity of the human operculum. *Cortex*, *109*, 303–321. https://doi.org/10.1016/j.cortex.2018.08.024

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Martin, K. I., & Ellis, N. C. (2012). The Roles of Phonological Short-Term Memory and Working Memory in L2 Grammar and Vocabulary Learning. *Studies in Second Language Acquisition*, *34*(3), 379–413. https://doi.org/10.1017/S0272263112000125

Matchin, W., & Hickok, G. (2020). The Cortical Organization of Syntax. *Cerebral Cortex (New York, N.Y.: 1991)*, *30*(3), 1481–1498. https://doi.org/10.1093/cercor/bhz180

Meara, P.M. (2005). *LLAMA Language Aptitude Tests*. Swansea: Lognostics.

Meara P.M., Milton J.L. and Lorenzo-Dus N. (2002). *LanguageAptitudeTests*. Newbury: Express.

Mennes, M., Kelly, C., Zuo, X.-N., Di Martino, A., Biswal, B., Castellanos, F. X., & Milham, M. P. (2010). Inter-Individual Differences in Resting State Functional Connectivity Predict Task-Induced BOLD Activity. *NeuroImage*, *50*(4), 1690–1701. https://doi.org/10.1016/j.neuroimage.2010.01.002

Middleton, F. A., & Strick, P. L. (2000). Basal ganglia and cerebellar loops: Motor and cognitive circuits. *Brain Research Reviews*, *31*(2), 236–250. https://doi.org/10.1016/S0165-0173(99)00040-5

Miller, G. A. (1958). Free recall of redundant strings of letters. *Journal of Experimental Psychology*, 56(6), 485–491. https://doi.org/10.1037/h0044933

Myachykov, A., Tomlin, R. S., & Posner, M. I. (2005). *Attention and empirical studies of grammar*. *22*(2–4), 347–364. https://doi.org/10.1515/tlir.2005.22.2-4.347

Neubauer, A., & Fink, A. (2009a). Intelligence and neural efficiency: Measures of brain activation versus measures of functional connectivity in the brain. *Intelligence*, *37*, 223–229. https://doi.org/10.1016/j.intell.2008.10.008

Neubauer, A. C., & Fink, A. (2009b). Intelligence and neural efficiency. *Neuroscience & Biobehavioral Reviews*, *33*(7), 1004–1023. https://doi.org/10.1016/j.neubiorev.2009.04.001

Neubauer, A. C., Grabner, R. H., Freudenthaler, H. H., Beckmann, J. F., & Guthke, J. (2004). Intelligence and individual differences in becoming neurally efficient. *Acta Psychologica*, *116*(1), 55–74. https://doi.org/10.1016/j.actpsy.2003.11.005

Neubauer, A. C., Sange, G., & Pfurtscheller, G. (1999). Psychometric intelligence and event-related desynchronization during performance of a letter matching task. In *Event-Related Desynchronization. Handbook of Electroenceph. And Clin. Neurophysiol.* (pp. 219–231). Elsevier B.V.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Nussbaumer, D., Grabner, R. H., & Stern, E. (2015). Neural efficiency in working memory tasks: The impact of task demand. *Intelligence*, *50*, 196–208. https://doi.org/10.1016/j.intell.2015.04.004

Opitz, B., & Friederici, A. D. (2003). Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *NeuroImage*, *19*(4), 1730–1737. https://doi.org/10.1016/S1053-8119(03)00170-8

Opitz, B., & Friederici, A. D. (2004). Brain Correlates of Language Learning: The Neuronal Dissociation of Rule-Based versus Similarity-Based Learning. *The Journal of Neuroscience*, 24(39), 8436–8440. https://doi.org/10.1523/JNEUROSCI.2220-04.2004

Opitz, B., & Hofmann, J. (2015). Concurrence of rule- and similarity-based mechanisms in artificial grammar learning. *Cognitive Psychology*, *77*, 77–99. https://doi.org/10.1016/j.cogpsych.2015.02.003

Pimsleur, P. (1966). *Pimsleur language aptitude battery (form S)*. Harcourt, Brace and world, Incorporated.

Rabini, G., Ubaldi, S., & Fairhall, S. L. (2023). Task-based activation and resting-state connectivity predict individual differences in semantic capacity for complex semantic knowledge. *Communications Biology*, *6*(1), Article 1. https://doi.org/10.1038/s42003-023-05400-1

Ramoser, C., Fischer, A., Caspers, J., Schiller, N. O., Golestani, N., & Kepinska, O. (2025). Language aptitude is related to the anatomy of the transverse temporal gyri. *Brain Structure and Function*, *230*(1), 1–18. https://doi.org/10.1007/s00429-024-02883-4

Rampinini, A., Balboni, I., Golestani, N., & Berthele, R. (2024). *A Behavioural Exploration of Language Aptitude and Experience, Cognition and More Using Graph Analysis* (SSRN Scholarly Paper No. 4751596). https://doi.org/10.2139/ssrn.4751596

Raven, J. C. (1998). Matrices Progressives de Raven Avancées—Abregées. *Pearson Clinical & Talent Assessment*.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Ryan, S. (2008). The Ideal L2 Selves of Japanese Learners of English [PhD Thesis]. University of Nottingham.

Shannon. C.S. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Singleton, D. (2017). Language aptitude: Desirable trait or acquirable attribute? *Studies in Second Language Learning and Teaching*, 7(1), 89–103. https://doi.org/10.14746/ssllt.2017.7.1.5

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., & Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, 106(31), 13040–13045. https://doi.org/10.1073/pnas.0905267106

Stansfield, C. (1988). *Pimsleur Language Aptitude Battery*. In Keyser D. & Sweetland R. (Eds.), *Test Critiques*, Vol VII (pp. 438-445). Test Corporation of America.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*(1), 307. https://doi.org/10.1186/1471-2105-9-307

Tagarelli, K. M., Shattuck, K. F., Turkeltaub, P. E., & Ullman, M. T. (2019). Language learning in the adult brain: A neuroanatomical meta-analysis of lexical and grammatical learning. *NeuroImage,* 193, 178–200. https://doi.org/10.1016/j.neuroimage.2019.02.061

Thompson, A.S., & Lee, J. (2018). The Motivational Factors Questionnaire in the Korean EFL context: Predicting group membership according to English proficiency and multilingual status. *The Language Learning Journal,* 26(4), 398-414. https://doi.org/10.1080/09571736.2015.1130082

Turker, S., Reiterer, S. M., Schneider, P., & Seither-Preisler, A. (2019). Auditory Cortex Morphology Predicts Language Learning Potential in Children and Teenagers. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00824

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, 2(10), 717–726. https://doi.org/10.1038/35094573

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1–2), 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

Ullman, M. T. (2016). Chapter 76—The Declarative/Procedural Model: A Neurobiological Model of Language Learning, Knowledge, and Use. In G. Hickok & S. L. Small (Eds.), Neurobiology of Language (pp. 953–968). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00076-6

Vanhove, J., & Berthele, R. (2015). *Item-related determinants of cognate guessing in multilinguals*. https://folia.unifr.ch/global/documents/304704

Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, *75*(5), 1182–1189. https://doi.org/10.1111/j.1365-2656.2006.01141.x

6 Supplementary Materials

6.1 Tables

Supplementary Table 1 – Pearson r and p-values for all significant correlations before removal of fatigue group.

Variable 1	*Variable 2*	*r*	*p*
age	span_visual	-.27	.004
age	inhibition	-.26	.004
age	orienting	-.21	.021
age	entropy_speak	.39	< .001
brocanto_accuracy	artgram_corr	.28	.002
brocanto_accuracy	span_verbal	.30	< .001
brocanto_accuracy	CVLT_imm	.29	.002
brocanto_accuracy	CVLT_long	.25	.006
brocanto_accuracy	APM_correct	.29	.002
brocanto_accuracy	declearn2_correct	.39	< .001
brocanto_accuracy	MLAT5	.3	.001
artgram_corr	CVLT_imm	.39	< .001
artgram_corr	CVLT_long	.39	< .001
artgram_corr	APM_correct	.25	.006
artgram_corr	alerting	-.24	.008
artgram_corr	declearn2_correct	.19	.046
artgram_corr	MLAT5	.29	.002
span_visual	span_verbal	.24	.009
span_visual	CVLT_imm	.27	.003
span_visual	CVLT_long	.26	.004
span_visual	APM_correct	.37	< .001
span_visual	alerting	-.31	< .001
span_visual	declearn2_correct	.25	.006
span_visual	MLAT5	.29	.002
span_verbal	CVLT_imm	.23	.01
span_verbal	CVLT_long	.21	.03

Variable 1	Variable 2	r	p
span_verbal	APM_correct	.3	< .001
span_verbal	alerting	-.29	.002
span_verbal	declearn2_correct	.22	.017
span_verbal	MLAT5	.38	< .001
CVLT_imm	CVLT_long	.82	< .001
CVLT_imm	APM_correct	.41	< .001
CVLT_imm	inhibition	.22	.02
CVLT_imm	alerting	-.2	.03
CVLT_imm	MLAT5	.44	< .001
CVLT_long	APM_correct	.4	< .001
CVLT_long	inhibition	.2	.03
CVLT_long	MLAT5	.41	< .001
APM_correct	inhibition	.21	.024
APM_correct	alerting	-.24	.01
APM_correct	declearn2_correct	.29	.001
APM_correct	MLAT5	.35	< .001
inhibition	entropy_speak	-.22	.02
alerting	declearn2_correct	-.32	< .001
alerting	MLAT5	-.28	.002
declearn2_correct	MLAT5	.3	.001

Supplementary Table 2 - Pearson r and p-values for all significant correlations after removal of fatigue group.

Variable 1	Variable 2	r	p
APM_correct	declearn2_correct	.24	.017
APM_correct	CVLT_long	.42	< .001
APM_correct	CVLT_imm	.39	< .001
APM_correct	MLAT5	.35	< .001
APM_correct	span_visual	.39	< .001
APM_correct	span_verbal	.26	.012
APM_correct	alerting	-0.23	.023

Variable 1	*Variable 2*	*r*	*p*
declearn2_correct	MLAT5	.34	< .001
declearn2_correct	span_visual	.22	.036
declearn2_correct	alerting	-0.31	.002
CVLT_long	CVLT_imm	.83	< .001
CVLT_long	MLAT5	.40	< .001
CVLT_long	span_visual	.28	.007
CVLT_imm	MLAT5	.43	< .001
CVLT_imm	span_visual	.26	.010
CVLT_imm	alerting	-0.22	.033
MLAT5	span_visual	.32	.001
MLAT5	span_verbal	.35	< .001
MLAT5	alerting	-0.30	.004
span_visual	age	-0.30	.003
span_visual	alerting	-0.27	.008
span_verbal	alerting	-0.28	.005
age	entropy_speak	.34	< .001
age	inhibition	-0.24	.020
entropy_speak	inhibition	-0.27	.009

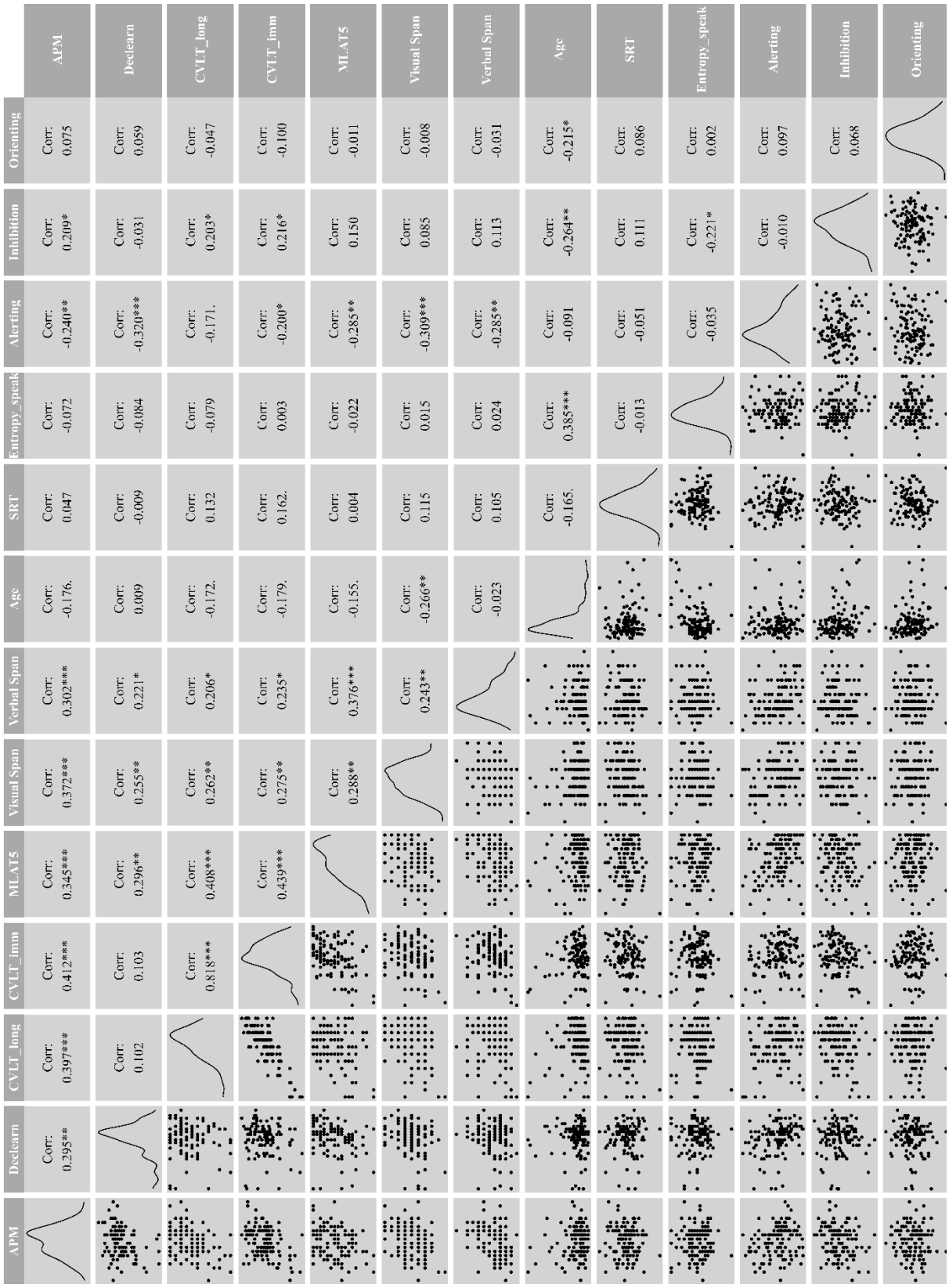
Supplementary Table 3 - Summary of RSFC based on task performance before correction for multiple comparisons. The table shows the results of the seed-based connectivity analysis based on Brocanto and ArtGram performance. Within the section for each ROI the results are presented first for RSFC modulated by Brocanto and then by ArtGram and they are further subsectioned into the positive and negative contrasts. Shown are the sizes and corrected p-values for clusters. Image coordinates in the same row as the cluster index are coordinates for the clusters center of gravity, while following coordinates denote locations of within cluster maxima. (Null results are represented by '-'.)

Cortical region (peak)	**Size (voxels)**	***p*-value**	Z_{max}	**L/R**	**Location (mm)**		
					X	***Y***	***Z***
(a) Seed ROI: left pars triangularis							
Connectivity modulated by *Brocanto* performance							
Positive contrast							
-	-	-	-	-	-	-	-
Negative contrast							
-	-	-	-	-	-	-	-
Connectivity modulated by *ArtGram* performance							
Positive contrast							
-	-	-	-	-	-	-	-
Negative contrast							
-	-	-	-	-	-	-	-
(b) Seed ROI: right pars triangularis							
Connectivity modulated by *Brocanto* performance							
Positive contrast							

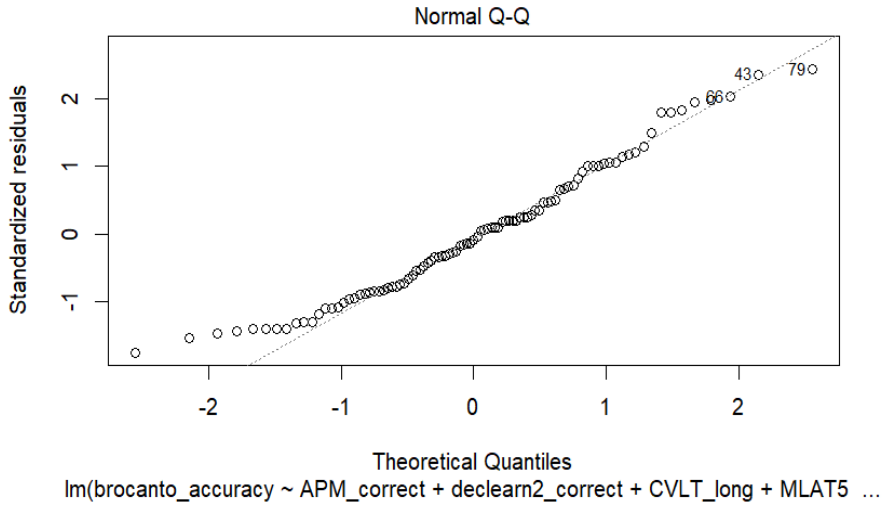
Cortical region (peak)	Size (voxels)	*p*-value	Z_{max}	L/R	Location (mm)		
					X	*Y*	*Z*
-	-	-	-	-	-	-	-
Negative contrast							
-	-	-	-	-	-	-	-
Connectivity modulated by *ArtGram* performance							
Positive contrast							
-	-	-	-	-	-	-	-
Negative contrast							
1 Precuneus (CoG)	439	*.027*	3.4	L/R	0	-58	34
Precuneus			3.4	L	6	-58	32
(c) Seed ROI: Left posterior middle temporal gyrus							
Connectivity modulated by *Brocanto* performance							
Positive contrast							
-	-	-	-	-	-	-	-
Negative contrast							
1 left middle temporal gyrus (CoG)	1489	<.001	3.76	L	-51	-15	-9
Planum polare			3.76	L	-46	-6	-4
Cerebral white matter			3.74	L	-48	-14	24
Central opercular cortex			3.69	L	-62	-8	10
Planum temporale			3.63	L	-48	-40	18
Planum polare/insular cortex			3.5	L	-44	-2	-8
2 right Hippocampus (CoG)	395	*<.001*	3.91	R	27	-4	-20
Amygdala			3.91	R	28	0	-16
Hippocampus			3.76	R	22	-8	-24
Connectivity modulated by *ArtGram* performance							
Positive contrast							
1 right Pallidum (CoG)	660	.002	4.4	R	18	0	6
Caudate			3.7	R	10	16	8
Thalamus			3.5	R	16	-8	2
Cerebral white matter			3.2	R	24	-10	26
Pallidum			3.1	R	24	-8	2
Negative contrast							
-	-	-	-	-	-	-	-
(d) Seed ROI: rpMTG							
Connectivity modulated by *Brocanto* performance							
Positive contrast							
-	-	-	-	-	-	-	-
Negative contrast							
-	-	-	-	-	-	-	-
Connectivity modulated by *ArtGram* performance							
Positive contrast							
1 Right Pallidum (CoG)	833	.001	4	R	18	-4	4

Cortical region (peak)	Size (voxels)	*p*-value	Z_{max}	L/R	Location (mm)		
					X	*Y*	*Z*
Pallidum			4	R	14	4	2
Caudate			3.8	R	10	14	8
Cerebral white matter (retrolenticular part of internal capsule)			3.8	R	26	-22	0
Thalamus			3.7	R	14	-6	2
Putamen			3.2	R	32	-22	4
Negative contrast							
-	-	-	-	-	-	-	-

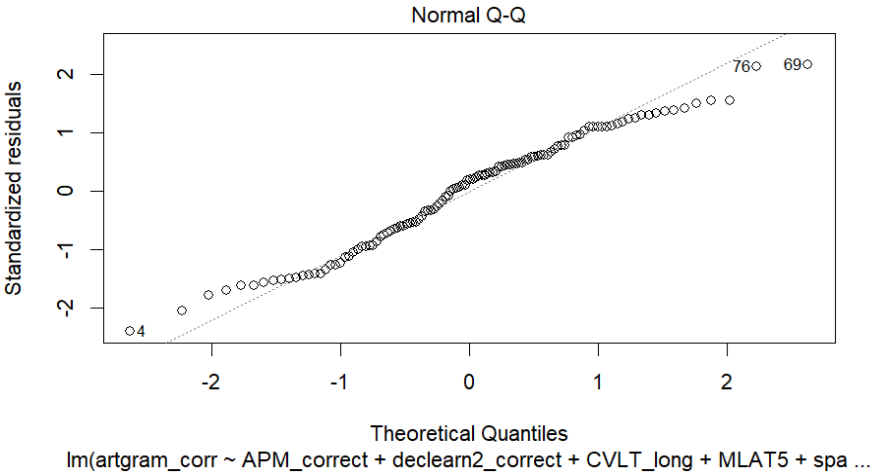
6.2 Figures



Supplementary Figure 1 - Pair plot based on data after the exclusion of the fatigue cluster. The upper half of the plot shows the correlation coefficients. Depicted on the diagonal is the distribution curve of each variable. The lower half shows a visual representation of the correlation between two variables in the form of a scatter plot.



Supplementary Figure 2 - Q-Q plot of multiple linear regression for Brocanto accuracy.



Supplementary Figure 3 - Q-Q plot of multiple linear regression for ArtGram