# The Earliest Stages of Second Language Learning: A Behavioral Investigation of Long-Term Memory and Age

**Diana Pili-Moss**

A thesis submitted to Lancaster University

for the degree of Doctor in Philosophy (PhD)

**May 2018**

Faculty of Arts and Social Sciences (FASS)

Department of Linguistics and English Language

## Abstract

*The Earliest Stages of Second Language Learning:*

*A Behavioral Investigation of Long-Term Memory and Age*

Diana Pili-Moss        Lancaster University

For the degree of Doctor in Philosophy (PhD)

A study with 40 L1 Italian 8-9 year old children and its replication with 36 L1 Italian adults investigated the role of declarative and procedural learning ability in the early stages of language learning.

The studies investigated: (1) the extent to which memory-related abilities predicted L2 learning of form-meaning mapping between syntax and thematic interpretation, word order and case marking; and (2) the nature of the acquired L2 knowledge in terms of the implicit/explicit distinction.

Deploying a computer game in incidental instruction conditions, the participants were aurally trained in the artificial language BrocantoJ over three sessions. Standardized memory tasks, vocabulary learning ability, and an alternating serial reaction time task provided measures of visual/verbal declarative and procedural learning ability. Language learning was assessed via a measure of comprehension during practice and a grammaticality judgment test.

Generalized mixed-effects models fitted to both experimental datasets revealed that, although adults attained higher accuracy levels and were faster learners compared to children, the two groups did not differ qualitatively in what they learned. However, by the end of the experiment, adults displayed higher explicit knowledge of syntactic and semantic regularities. During practice, declarative learning ability predicted accuracy in both groups, but procedural learning ability significantly increased only in children. The procedural learning ability effect emerged again significantly only in the child grammaticality judgment test dataset. In the practice data declarative learning ability and vocabulary learning ability interacted negatively with procedural learning ability in children, whereas declarative learning ability interacted positively with procedural learning ability in adults. Moreover, the positive interaction in adults only obtained for a subset of practice stimuli, i.e. sentences where the processing of linking between morphosyntax and thematic interpretation was required. Overall, the findings support age-related differences and linguistic target differences in the way abilities related to long-term memory predict language learning.

**Keywords**: procedural memory; declarative memory; L2 learning; age differences; artificial language learning.

# Declaration

I hereby declare that this dissertation is my own work and no portion of it has been submitted in support of an application for another degree or qualification of this or any other university or institute of learning.

May 16th 2018

## Acknowledgements

I am deeply grateful to Jenefer Philp and Patrick Rebuschat, who supervised my work at Lancaster, for their advice and on-going support in the last few years and for the fruitful contacts with the international research community they made possible. I would also like to thank the Department of Linguistics and English Language for awarding me a fee waiver bursary for the duration of my studies.

A large number of other researchers more or less directly contributed to this work in many different ways. I am grateful to Panos Anagnostopoulou, Dimitrinka Atanasova, Silke Brandt, Aina Casaponsa, Rob Davies, Robert DeKeyser, Sarah Grey, Bimali Indrarathne, Motoi Kitamura, Veronika Koller, Judit Kormos, Karen Lichtman, Jarrad Lum, Theo Marinis, Marjie Michel, Padraic Monaghan, Elena Nichele, Michael Ratajczak, Andrea Révész, Anna Samara, Ludovica Serratrice, Michael Ullman, and Elizabeth Wonnacott. I would also like to thank the members of the SLLAT and the PERLL research groups at Lancaster for productive exchange and discussion.

I am especially grateful to Kara Morgan-Short who shared with me her paradigm and helped me develop it, for her advice, the countless Skype sessions, the collaboration that emerged and for organizing a memorable blues night out in Chicago.

This experimental research would have not been possible without the availability of the adult and child participants in Italy and the UK and the collaboration of the school and teachers in Italy. Thank you so much Angela for regularly going the extra mile and making such a difference. I thank Cristina, Don

Enrico, Susanna and Patrizia and my English and Italian families for generously supporting me all along.

But above anybody else I thank you Andrew Moss, for just being yourself and walking beside me on this incredible adventure. I dedicate this work to the loving memory of my mum, who I am sure also had a hand in making sure everything went fine.

# Table of Contents

## List of Tables

## List of Figures

## List of Abbreviations

| | |
|---|---|
| ACC | Accusative |
| Adj | Adjective |
| Adv | Adverb |
| AG | Artificial Grammar |
| AGL | Artificial Grammar Learning |
| AJT | Acceptability Judgment Test |
| AN | Anterior Negativity |
| ASRT | Alternate Serial Reaction Time |
| CA | Case |
| CLIL | Content and Language Integrated Learning |
| CMS | Children Memory Scale |
| CN | Condition Number |
| Conf | Confidence |
| CVMT | Continuous Visual Memory Task |
| D&P | Doors and People (Test) |
| Decl | Declarative Learning Ability |
| DM | Declarative Memory |
| DP | Declarative-Procedural |
| EEG | Electroencephalogram |
| ERP | Event-Related Potential |
| ESL | English Second Language |
| Exec | Central Executive |
| fMRI | functional Magnetic Resonance Imaging |
| GJT | Grammaticality Judgment Test |

| | |
|---|---|
| Gramm | Grammatical |
| L1 | First Language |
| L2 | Second Language |
| LAN | Left-Anterior Negativity |
| LTM | Long-Term Memory |
| LTP | Long-Term Potentiation |
| MLAT | Modern Languages Aptitude Test |
| MTL | Medial-Temporal Lobe |
| NOM | Nominative |
| NP | Noun Phrase |
| O | Object |
| OVa | Object-Verb (asymmetrical) |
| OVs | Object-Verb (symmetrical) |
| PET | Positron Emission Tomography |
| PFC | Pre-Frontal Cortex |
| PhonLoop | Phonological Loop |
| PM | Procedural Memory |
| PMT | Picture-Matching Task |
| Proc | Procedural Learning Ability |
| PT | Processability Theory |
| ROCF | Rey-Osterrieth Complex Figure |
| RR | Representational Redescription |
| RT | Reaction Time |
| S | Subject |
| S1 | Session 1 |

| | |
|---|---|
| S2 | Session 2 |
| S3 | Session 3 |
| SLA | Second Language Acquisition |
| SL | Statistical Learning |
| SLI | Specific Language Impairment |
| SMA | Supplementary Motor Area |
| SOVa | Subject-Object-Verb (asymmetrical) |
| SOVs | Subject-Object-Verb (symmetrical) |
| SpacingS1S2 | Spacing between Session 1 and Session 2 |
| SpacingS2S3 | Spacing between Session 2 and Session 3 |
| SRT | Serial Reaction Time |
| STM | Short-Term Memory |
| SV | Subject-Verb |
| TOL | Tower Of London |
| V | Verb |
| V2 | Verb Second |
| Viola | Type of Violation |
| VocLearn | Vocabulary Learning Ability |
| WO | Word Order |
| WPT | Weather Prediction Task |

## Introduction

In the present thesis I will assume that the mental abilities subsumed under the label 'memory', together with their neural correlates, represents the cognitive basis of all types of human learning. This includes the highly complex skills implicated in the comprehension and production of a second language. Especially relevant to the present thesis is the distinction between declarative long-term memory and nondeclarative long-term memory, and more specifically, between declarative memory and procedural memory (a type of nondeclarative memory). Declarative memory is a fast-learning flexible system suited for the learning of facts, events and semantic information, whilst procedural memory is a system that learns more slowly and incrementally and is suited for probabilistic learning, learning of implicit sequences and skills (Cabeza & Moscovitch, 2013; Squire, 2004; Squire & Wixted, 2011).

Experimental evidence accumulating in neuropsychology and neuroscience has identified specific brain areas associated with the engagement of the two systems. Declarative memory is thought to implicate the medial temporal lobe, in particular the hippocampus, in addition to temporal and frontal neocortical regions, whereas procedural memory engages the basal ganglia and connected areas in the frontal cortex (pre-motor cortex and portions of Broca area). Although capable of processing information independently and in parallel, in certain circumstances the two memory systems may interact competitively or co-operatively in the acquisition/processing of information (Packard & Goodman, 2013), a phenomenon expected to occur in (second) language acquisition, similarly to what is observed for other types of learning (Ullman, 2005, 2015, 2016).

The last 25 years have seen the development of a number of cognitive approaches to language acquisition/learning that explicitly refer to the role of declarative and procedural memory (Paradis, 2009; Ullman, 2005, 2015, 2016) and/or to the role of declarative and procedural knowledge (N. Ellis, 2005; N. Ellis & Wulff, 2015; DeKeyser 2007; 2015). Given that the focus of the present thesis is on learning morphosyntax in a novel miniature language, I will be mostly concerned with the consequences these models have for L2 (second language) learning. As for the terminology, when not further specified, language acquisition refers to the speaker's appropriation of her L1. When referring to the L2, the words 'acquisition' and 'learning' are equivalent and used interchangeably.

An important tenet of the declarative/procedural models (Paradis, 2009; Ullman, 2005, 2015, 2016; cf. 1.6.1) is that certain aspects of language are preferentially learned by the declarative memory system or by the procedural memory system, depending on the specific characteristics of the two systems. For example, aspects of language relating to semantics, and notably vocabulary, intended as a set of arbitrary associations between words and their meanings, are preferentially learned by the declarative memory system. By contrast, rule-based grammatical patterns (e.g., word order) are expected to be learned/processed by the procedural memory system due to its suitability to learn rule-based sequences.

Beside the evidence provided by neuroimaging and ERP studies, declarative and procedural memory have also been investigated behaviorally as individual differences, by deploying standardized measures of declarative and procedural learning ability developed in neuropsychology. This research, conducted mostly, but not exclusively, using miniature language paradigms, has found that the effects of declarative and procedural learning ability are modulated by a number of factors including, for example, amount of practice, type of rule, feedback, and type of learning context (for a review see Buffington & Morgan-Short, 2018; Chapter 4 of this thesis).

Research in neuropsychology and in cognitive psychology has provided evidence that declarative and procedural learning ability undergo developmental changes across the lifespan, with specific memory skills following independent trajectories (Di Giulio, Seidenberg, & O' Leary, 1994; Lum, Kidd, Davis, & Conti-Ramsden, 2010; Ofen et al., 2007). This state of affairs suggests that, if memory has an effect on language learning, we should be able to observe age-related differences in the way language is learnt, processed and used (Morgan-Short & Ullman, 2011; Ullman, 2015).

Declarative and procedural learning abilities have been studied in relation to the L1 in typically developing children and in children with specific language impairment (e.g., Conti-Ramsden, Ullman, & Lum, 2015; Kidd, 2012). For typically developing children, these studies have found that procedural learning ability is a predictor of L1 proficiency. However, no training study to date has examined the relationship between age and long-term memory (LTM) abilities in child L2 learning.

A further point concerns the way L2 learning has been measured in adult behavioral studies. To date these studies have used grammaticality judgment tests (often in conjunction with ERP measurements) or forced-choice tasks, but in general have tended to measure attainment once at the end of the exposure, or twice, if the aim of the study was to compare L2 attainment at low and high levels of proficiency. Beside one study currently under review (Pili-Moss, Brill-Schuetz, Faretta-Stutenberg & Morgan-Short, 2018), no previous research has tracked accuracy during practice to investigate how different LTM abilities shape L2 learning.

Another point of interest is the extent to which other cognitive variables, for example short-term memory (STM) and working memory, determine L2 learning or moderate the effect of declarative and procedural memory on learning, and whether differences can be found between children and adults. For STM and working memory, research on adults has found no effect or mixed results to date (e.g., Antoniou, Ettlinger, & Wong, 2016; Janacsek & Nemeth, 2013).

Overall, the present research project originates from a keen personal interest in the role of memory-related individual differences in (second) language acquisition and from the observation that, whilst a growing body of literature has analyzed the effects of memory in adults, the number of studies focusing on children or age differences is still very limited in comparison. The present thesis compares learning of a novel miniature language in 8-9 year old children and in adults and is the first study to assess the roles played by declarative and procedural learning ability in child L2 learning. Whilst previous comparative studies investigating children in artificial language paradigms have deployed a simplified syntax or single word morphology (e.g., Hudson Kam & Newton, 2009; Kapa & Colombo, 2014), the present thesis examines learning of a fully productive miniature language with a complex phrase structure. Exposure to the language is aural and, after an initial training phase, the language is practiced in the context of a video game in incidental instruction conditions.

The study analyzes the learning of word order, case marking and the relationship between syntactic position and case marking with the thematic interpretation of NPs (linking rules). In the study, linking rules are assessed during the game practice, whilst word order and case marking are assessed via a grammaticality judgment test administered at the end of practice. A pervasive fact in natural languages is that, rather than be classifiable as mainly semantic or syntactic, grammatical phenomena often find themselves at the interface between syntax and semantics. One such example are the rules that determine the linking between the thematic interpretation of an NP (agent, theme, etc.) and its position in the sentence (subject, object, etc.). This type of form-meaning relationship has a clear semantic content (that derives from the meaning and the argument structure of the verb) but is also defined by a rule-based word order. Given the predictions that the declarative/procedural (DP) models make for the learning of different types of linguistic targets, investigating how the two memory systems may be engaged in the learning and processing of linking rules is of particular theoretical interest.

Unlike previous behavioral studies that have investigated declarative and procedural learning abilities as individual differences, the studies in the present thesis include a measure of accuracy that tracks the development of L2 comprehension during L2 practice (i.e., during the game). Finally, the evaluation of the regression models in the data analysis considers the effect on learning of additional covariates, including a measure of short-term memory and of working memory.

By comparing L2 learning in school-aged children and adults, the thesis examines the role of the age factor, a topic that remains debated in the field of second language acquisition (SLA). An established finding in this area of study regards the difference between ultimate attainment and rate of learning. Whilst an early start tends to represent an advantage for the ultimate attainment of L2 learners (i.e., for the highest level of L2 development attainable by a learner), rate of learning studies have consistently showed that, particularly in instructed contexts, there is an overall advantage on L2 attainment for adults, compared to children, for comparable amounts of instruction (DeKeyser, 2013; Muñoz, 2008). This said, not many studies on rate of learning have investigated the child/adult comparison controlling not only amount of instruction, but also confounds introduced by different instructional conditions and language learning measures.

The research involves two studies, one conducted with children (Study 1), and the other conducted with adults (Study 2). The two studies are matched in methodology but are independent, so that the age comparisons are addressed as a part of the discussion (for a recent similarly structured study that compared how executive function modulates L2 learning in children and adults, cf. Kapa & Colombo, 2014). With the exception of minor adaptations, the study adopted the methodology developed by Morgan-Short (2007) and deployed in a number of subsequent studies with adults. Maintaining the same paradigm in a study with children facilitated a direct comparison with previous research, allowing to focus on age effects and to control for confounds that would have arisen from the adoption of different experimental procedures.

In order to examine learning of word order, case marking and linking rules, children and adults (all L1 Italian users) were aurally exposed to an artificial language displaying the word order of Japanese main sentences and including 14 pseudowords. The use of the miniature language ensured that the language was equally novel for all participants. After a short vocabulary training, the participants were exposed to full sentences in the artificial language in incidental condition with no explanation or reference to the language's structural properties. Aural exposure to the language was always associated to corresponding visual stimuli (game configurations), so that the learning task was meaningful, purposeful (aimed at increasing the player's game score), and with a focus on comprehension rather than on linguistic form.

Unlike previous studies deploying the paradigm, the thesis additionally investigated the extent to which the participants were aware or unaware of the L2 knowledge they acquired, i.e. the extent to which their L2 knowledge was explicit or implicit. If the acquired L2 knowledge is explicit, participants will demonstrate it behaviorally and additionally demonstrate some degree of awareness of L2 patterns (some learners may also be able to verbalize them). By contrast, if L2 knowledge is implicit participants will demonstrate it behaviorally but they will not be aware of the L2 patterns they applied or be able to describe what it is that they know. In order to assess language knowledge awareness, the studies deployed verbal reports collected at the end of the experiment as evidence of explicit verbalized knowledge, and subjective ratings relative to the participant's judgments of the sentence stimuli in the grammatical judgment test (GJT) as indirect evidence of implicit and explicit language knowledge (Dienes & Scott, 2005). Specified for children (Chapter 6) and for adults (Chapter 7), the research questions were formulated as follows (in the formulation of the research questions the word 'practice' refers to the language practice the participants had while playing the computer game):

RQ1 To what extent do declarative and procedural learning ability modulate the participants' L2 aural comprehension and learning of the rules linking morphosyntax and thematic interpretation during practice?

RQ2 To what extent do declarative and procedural learning ability modulate the participants' L2 learning of word order and case marking, as measured by a grammaticality judgment test administered at the end of practice?

RQ3 To what extent is the L2 knowledge acquired by the participants implicit/ explicit?

The investigation of the three research questions illustrated above anticipates advances in a number of main research areas relevant to child second language acquisition, adult second language acquisition and to the age comparison perspective. They can be briefly summarized as follows:

(a) Elucidating the role of declarative and procedural learning ability in the earliest stages of child second language acquisition.

(b) Extending the analysis of the role of declarative and procedural learning ability in adult L2 learning to adults whose L1 is not English.

(c) For both children and adults, elucidating the role of potential interactions between declarative and procedural learning ability in the acquisition of different aspects of the L2 in the earliest stages of learning.

(d) A better understanding of the role of declarative and procedural learning ability in the L2 acquisition of constructions where syntactic realization is closely dependent on semantics (such as in thematic linking).

(e) A better understanding of the extent to which children and adults are aware/unaware of their L2 knowledge and of the morphosyntactic properties of a novel L2 in the earliest stages of learning.

(f) Elucidating to what extent the rate of learning of a novel L2 differs in children and adults in the earliest stages of learning.

In order to help the reader navigate the text, I will provide a brief illustration of how its content is structured. The first half of the thesis (Chapters 1-4) provides the theoretical background to the experimental investigation. In particular Chapter 1 outlines the framework in which the study of declarative and procedural learning ability as individual differences is set, and provides the methodological underpinnings to the measurement of these learning abilities in the experiments. Similarly, Chapter 2 introduces the theoretical background and the measures that will be deployed to assess the nature of the language knowledge attained by the participants in terms of the implicit/explicit distinction.

Chapter 3 focuses on issues relative to the current state of the debate around age effects in second language acquisition research, discussing different approaches to the study of the age factor, and a selection of rate of learning studies. The function of Chapter 3 with respect to the experimental studies is to provide a background for the comparison between the results of Study 1 and Study 2 in a SLA perspective (Chapter 8). In the studies, the L2 the participants are exposed to is a miniature artificial language reflecting the morphosyntax of Japanese. Hence, Chapter 4 introduces a detailed review and critical discussion of child and adult miniature language studies and concludes by presenting the variables of interest and the focus of the investigation.

The second part of the thesis (Chapters 5-9) includes the methods of the experimental studies, their results, the discussion and the conclusions. Chapter 5 provides a description of the miniature artificial language deployed in the studies, a rationale for its characteristics, and information relative to the computer game deployed as learning environment in the experiments, including a differentiation between trial types (asymmetric vs. symmetric) that is critical to the experimental design. Chapter 6 and Chapter 7 present the child and adult studies respectively, reporting methodology, data analysis and results. The discussions of both studies and a critical comparison between the results of Study 1 and Study 2 are included in Chapter 8. Finally, Chapter 9 summarizes the main contributions of the studies, their limitations and potential further developments, and concludes with some pedagogical considerations in the light of the thesis findings.

# 1. Language learning and long-term memory

## 1.1 The neurophysiological basis of learning

Neural cells are the fundamental building blocks of the central nervous system and their capacity to engage in complex processes of interaction, information storage and retrieval constitutes the basis of human memory and learning. In the first part of the Twentieth century physiologists studying the interaction among neurons in the nervous system discovered that this mainly consists in the creation and propagation of electrical signals modulated by specific chemical molecules known as neurotransmitters. The transmission of information among neurons (Figure 1.1)

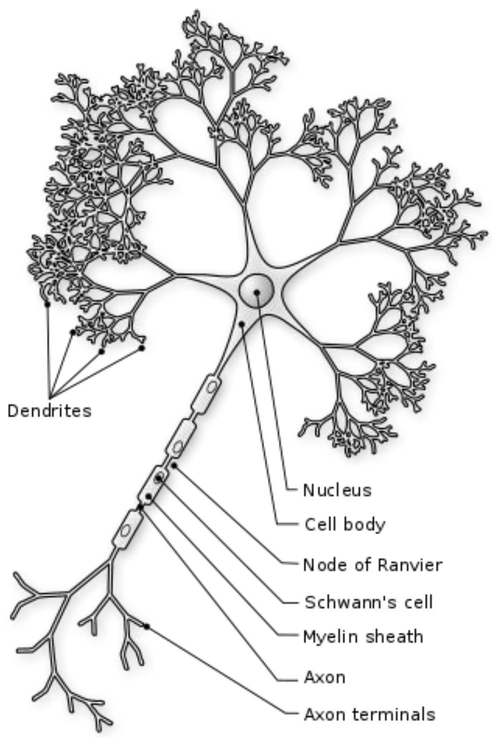


*Figure 1.1*. The Neural Cell[1].

---

[1] Source: Wikimedia Commons. By Nicolas Rougier - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=2192087.

depends on the intensity of the electrical stimulation applied and is determined by a difference in electrical potential originating in pre-synaptic areas located in the final part of a cell's axon (where neurotransmitters are released), and post-synaptic areas located in the dendrites of a neighboring neuron (where specific receptors are located). When the electrical stimulation on a neuron is sufficiently strong and reaches a threshold action potential the neuron activates (i.e., 'fires') and propagates the electric signal further to neighboring neural cells. The study of neuron activation has led to a deeper understanding of how memories may be formed and stored in the brain. Developing this line of research, Bliss and Lømo (1973) discovered *long-term potentiation* (LTP). LTP is the cellular mechanism at the basis of the creation and storage of memories and can be defined as a long-lasting enhancement in signal transmission between two neurons after repeated stimulation.

## 1.2 Types of long-term memory

Defining the difference between the constructs of learning and memory, Squire (1987) indicated that learning is the process by which new information is acquired, whilst memory denotes the persistence of the learned information over time. If learning leads to long-term retention of information, the definitions of the two constructs largely overlap and learning can be defined as a process by which information is acquired, retained in a stable way and available for retrieval for relatively long periods of time. In the latter sense, learning can be understood as the behavioral counterpart of neurophysiological changes arising as a result of long-term modifications in synaptic configurations relative to long-term memory systems (see also Eichenbaum, 2008; 2012).

According to the model proposed in Squire (1987, 2004), and later adopted with minor terminological variations by most authors in the field (Figure 1.2, next page), two main categories of long-term memory are available; declarative and implicit (nondeclarative) memory. Declarative memory is an overarching label comprehending different types of memory that support long-term storage of facts and events. According to Squire's (2004) taxonomy (cf. also Squire & Wixted, 2011) declarative memory, also referred to as cognitive or explicit memory, includes both episodic and semantic memory. Operationally, declarative memory "allows remembered material to be compared and contrasted" (Squire, 2004, p. 173), providing the capacity to single out "what is unique about a single event which by definition occurs at a particular time and space" (Squire, 2004, p. 174). According to Squire's taxonomy, nondeclarative memory (also known as implicit or behavioral memory) includes a number of different forms of long-term memory supporting habituation, perceptual learning (including priming), procedural and sequence learning, conditioning, and nonassociative learning. An important operating principle in nondeclarative memory is "the ability to gradually extract the common elements from a series of separate events" (Squire, 2004, p. 174).

According to this model, different forms of memory can also be classified depending on whether they are conscious or unconscious forms of learning.

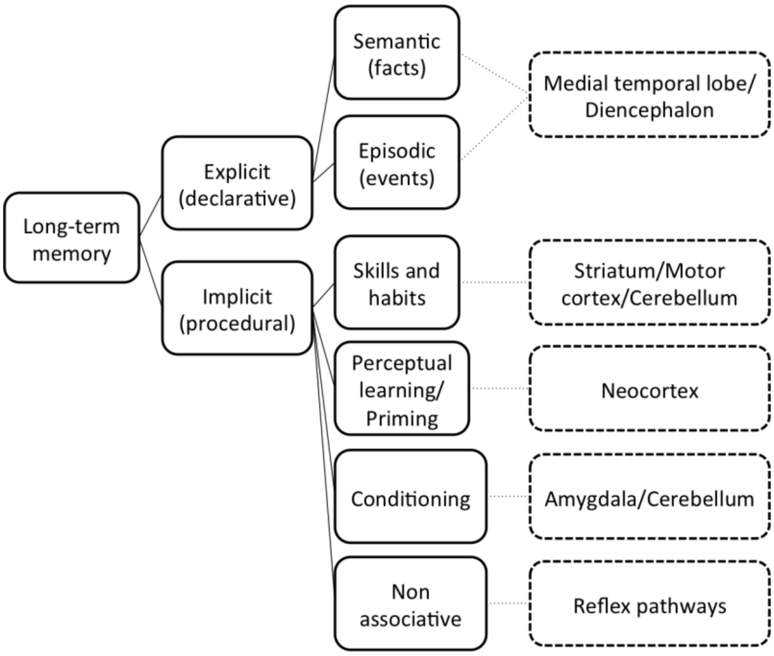


*Figure 1.2*. Taxonomy of Long-Term Memory Systems and Their Neural Correlates Based on Squire (2004).

Unconscious forms of learning include habituation, perceptual learning and procedural learning and they can be classified as forms of behavioral memory involving "the unconscious acquisition and implicit expression of memory through changes in task performance" (Eichenbaum, 2008, p.111). Conscious forms of memory operate under an individual's conscious awareness and include cognitive memory, declarative memory, episodic memory and semantic memory. Recently, models of long-term memory based on differences in processing rather than strict neuroanatomical mapping (e.g., Henke, 2010) have rejected the notion that consciousness should be a variable in describing how different memory systems operate. In particular they maintain that declarative memory is not exclusively associated with consciousness, as indicated by evidence that brain areas traditionally implicated in declarative memory such as the hippocampus appear to support learning of associations even without awareness.

## 1.3 Memory systems and localization of their pathways in the brain.

The two long-term memory systems and their subtypes have also been defined through their association or lack of association with specific neural areas. For declarative memory this mainly involves the hippocampus (Figure 1.3) and neighboring neural structures in the medial temporal lobe (MTL), as well as a two-way flow of information between different parts of the cortex and the hippocampus.

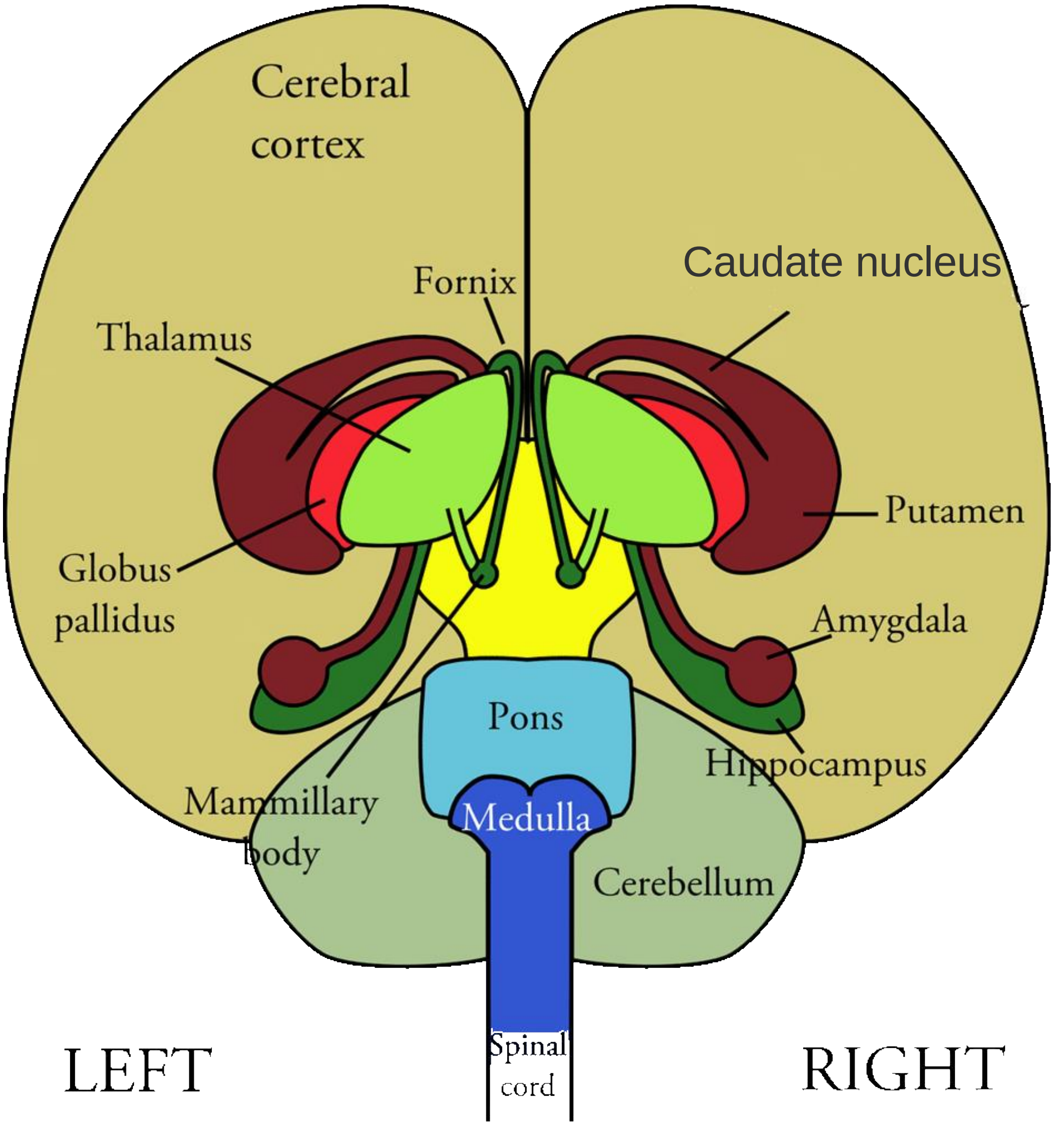


*Figure 1.3*. Location of the Hippocampus and the Basal Ganglia in the Brain[2].

---

2 Source: www.pixabay.com. Creative Commons License CC0.

The localization of long-term nondeclarative memory has not been supported by uncontroversial neurological evidence of the type available for declarative memory. For this reason some authors, for example Reber (2013), suggest that nondeclarative long-term memory functions should be best captured in terms of "an emerging property of general plasticity" (Reber, 2013, p. 2027). This plasticity, i.e. long-term synaptic reconfiguration, would affect different areas of the brain depending on the sensory nature of the stimuli and on how these are experienced in the environment by the individual. In learning involving nondeclarative memory, plasticity has been observed, for example, in the sensory cortex for different types of perceptual learning and in the basal ganglia in relation to category learning, sequence learning and probabilistic classification.

One type of nondeclarative memory, perceptual memory, has a major role in priming, broadly defined as the ability to detect/identify stimuli after recent exposure. While perceptual memory refers to the initial processing of the stimuli by the low level sensory cortex and is related to specific stimuli, perceptual skill learning involves higher level processing leading to the categorization of the stimuli including the evaluation of statistical patterns in the stimuli presentation.

Procedural memory is the type of nondeclarative long-term memory involved in habit learning, sequence and probabilistic learning and learning of complex skills that are performed automatically and typically involve engagement of the brain’s motor areas. The most important brain areas implicated in procedural memory are the basal ganglia (including the neo-striatum - putamen and caudate nucleus - the globus pallidus, part of the thalamus and the substantia nigra; Figure 1.3), areas in the frontal cortex that provide input to and receive input from the basal ganglia (including Broca area - BA 44 and BA 45) and the cerebellum.

### 1.4 Long-term memory vs. working memory

Although the main focus of the thesis will be the role of long-term memory in language learning it is useful to complete this brief review of memory systems and their functions by introducing how short-term and working memory are conceptualized. This will be done using one of the working memory models that have been proposed and will be referred to in the present study, the Multi-Component model (Baddeley, 2000; 2015; Baddeley & Hitch,1974; Baddeley & Logie, 1999). In its initial form this model conceived working memory as a system with three components: a phonological loop, a visual sketchpad and a main attentional component, the central executive. The function of the phonological loop and of the visual sketchpad is to act as storage subcomponents retaining short-term verbal or visual information the central executive operates on/manipulates. Unlike long-term memory, the time storage of information in the working memory subcomponents is very limited (in the order of seconds).

Later conceptualizations of Baddeley's model added an interface with long-term memory and an *episodic buffer*, a further multidimensional storage component with greater capacity compared to the other storage modules and "capable of combining information from the visuospatial and verbal subsystems and linking it" (Baddeley, 2015, p. 25). With regards to its neuroanatomical localization, working memory processes engage the pre-frontal cortex (PFC), although recent evidence shows that sensory cortices may also play an important role in short-term storage.

### 1.5 Developmental aspects of memory systems

The structures that underpin the functioning of memory systems are known to constantly develop throughout childhood and adolescence into young adulthood. However, there is a difference between structures subserving the declarative and the procedural memory system relative to the time at which they reach full anatomical and functional maturity. The procedural memory system relies on neural structures (e.g. the caudate) that are known to develop earlier in life and are believed to reach peak development between age 7 1/2 and 10 (Giedd et al., 1999; Lum, Kidd, Davis, & Conti-Ramsden, 2010). On the other hand neural structures supporting the declarative memory system (medial temporal lobe structures) are known to grow anatomically until age 16 and further develop functionally until up to age 22.

Neuroimaging studies comparing school-aged children and adults confirm that a developmental advantage for older individuals exists in tasks involving the recollection of episodic memories, although a complete longitudinal picture of how declarative memory functions develop is still not available (Ofen et al., 2007). Some studies have indicated a more limited activation of median-temporal structures for younger children compared to older children or adults recalling scenes or short stories (e.g., Chiu, Schmithorst, Brown, Holland, & Dunn, 2006), whilst others have found that more limited activation of prefrontal areas accounted for age differences (Ofen et al., 2007).

Behavioral studies that have investigated the development of procedural memory longitudinally (Lum et al., 2010) or compared children of different age groups cross-sectionally (e.g., Meulemans, Van der Linden, & Perruchet, 1998) generally found no significant between-group differences in procedural memory ability and significant between-group differences for measures of declarative memory. However, at least one study, Thomas et al. (2004), found significantly larger learning effects in adults in a serial reaction time task compared to 7 year old and 11 year old children. The results of these studies suggest that although age-related differences in declarative memory may be comparatively more robust, developmental differences in procedural memory cannot be excluded.

Also in the case of working memory a trajectory of cognitive development and important differences between children and adults are well attested in the literature (Gathercole, Pickering, Ambridge, & Wearing, 2004). In terms of memory span a number of studies have found that the capacity to recall materials in short-term memory increases from three items at 4 years of age to six or seven items in young adulthood. Further differences pertain to the development of the individual components of working memory. Importantly, the efficiency of the phonological loop undergoes substantial development during the primary school years from age 7, due to the availability of more efficient subvocal rehearsal strategies that are supported by the increasing memory span. The efficiency of the phonological loop keeps increasing up to age 12, when it reaches adult levels (Gathercole, Adams, & Hitch, 1994). Increases in the ability to retain short-term information have direct consequences also on the ability to manipulate linguistic information in working memory in language learning processes, and suggest that these abilities should be more developed in older children and adults compared to younger children.

## 1.6 Long-term memory and language learning: a review of theoretical models

### 1.6.1 The declarative-procedural model

***Ullman's Model***. In the declarative-procedural model (DP model) Ullman provides an account of L1 and L2 acquisition rooted in the distinction between declarative and procedural memory (Ullman, 2004, 2005, 2015, 2016). It is important to note that Ullman considers only procedural memory (Ullman, 2004, p. 237), and that the model does not extend to all forms of nondeclarative memory. Although the circuitry involved in the creation and storage of sequential and procedural information is not as well understood as the one supporting the declarative memory system, focusing specifically on procedural memory allows a more precise identification of the network of brain areas that are implicated.

Ullman specifically considers the role of a number of factors modulating how the two memory systems operate. These include molecular and genetic factors, as well as other factors such as sex, age, memory consolidation during sleep, and memory consolidation after prolonged periods of no-exposure (Morgan-Short, Finger, Grey, & Ullman, 2012; Ullman, 2005, p. 161). The two main molecular factors that play a role in modulating long-term memory are the hormone estrogen and the neurotransmitter dopamine. In particular estrogen has been found to improve declarative memory functions in human adults and in rodents, whilst the neurotransmitter dopamine is known to support neural activity in the basal ganglia.

The role of molecular factors, especially the role of estrogen for declarative memory, has been related to sex differences in the reliance on declarative memory processing found in humans (Golomb et al., 1996; Kimura, 1999). Although no experimental studies to date have tested this prediction, Ullman suggested that estrogen could also play a role in developmental changes, as these may arise as a consequence of the higher production of this hormone in both sexes (Ullman, 2004, p. 256).

Due to the different development of declarative and procedural memory across the lifespan, the DP model also predicts an age effect, i.e. a stronger reliance on declarative memory processing in adolescents and adults and a stronger reliance on procedural memory processing in children. Further, external factors also modulate the extent to which one or the other system is relied on. For example, since declarative memory is fast and capable of efficient learning at low level of exposure to a stimulus, it will be more strongly engaged earlier in the learning process compared to procedural memory.

Ullman also suggests that the two memory systems work 'in parallel' and have the potential to acquire the same kind of information and represent it at different levels (in this sense they can give rise to "redundant" representations; Ullman, 2015, p. 139). However they are also capable of interacting, *co-operatively* or *competitively* (Ullman, 2005, p. 161, the italics are mine), depending on the presence vs. absence of the endogenous and external factors modulating their activity. For example, in case of impairment or attenuation of one system, this may give rise to a "seesaw effect" (Ullman, 2015, p. 139), with the second system taking over some of its functions and assuming a predominant role in processing. Evidence from human and animal studies also shows that inhibition of one system by the other during learning and/or retrieval is also possible (Packard & Goodman, 2013).

Ullman claims specifically that the acquisition, learning and use of language can be accounted for by the DP model because 'the brain systems which subserve declarative and procedural memory play analogous roles in language as in their non-language functions' (Ullman, 2004, p. 244; but see also Paradis, 2004, 2009). In doing this, Ullman distinguishes between: (a) lexical, arbitrary and idiosyncratic language knowledge; and (b) mental grammar. Mental grammar is defined as a series of rule-based procedures governing the sequential and hierarchical organization of linguistic units, including syntax, inflectional and derivational morphology, and aspects of sound combination and non-lexical compositional semantics.

Following Ullman, the declarative memory system, which specializes in the storage of discreet, factual pieces of information, is responsible for the acquisition, representation and use of the lexicon, of lexical semantics, irregular morphological forms, and grammatical forms stored as ‘chunks’ or idioms. Additionally, declarative memory in general underlies the acquisition of morphosyntax at low levels of proficiency and exposure in the L2. As a memory system specialized in the representation of sequenced procedures, the procedural memory system presides in general over rule-based grammar in the L1, and in the L2 at increasingly higher levels of exposure and proficiency.

Associations between MTL structures subserving long-term declarative memory and lexical-semantic knowledge in the L1 have been found in healthy subjects among others in Damasio, Grabowski, Tranel, Hichwa, and Damasio (1996), Martin, Unterleider, and Haxby (2000), and Newman, Pancheva, Ozawa, Neville, and Ullman (2001). Associations between the activation of brain areas related to procedural memory and morphosyntactic processing have been found in PET and fMRI studies in Moro et al. (2001) (caudate nucleus), Embick, Marantz, Miyashita, O'Neil, and Sakai (2000), Friederici (2002) (Broca Area - BA 44), Newman et al., (2001), Caplan, Alpert, and Waters (1998) (supplementary motor area - SMA), Friederici and von Cramon (2000) and Ni et al., (2000) (anterior superior temporal gyrus).

A growing body of evidence provided by recent neurocognitive studies also supports a parallelism between how the declarative and procedural memory systems work in the L1 and the L2. These results are important because they provide support for the DP model as well as for the existence of a neurological continuum between rule-based grammar processing in L1 speakers and highly proficient L2 learners. Particularly relevant to this comparison is the tracing of a bi-phasic EEG response (an early anterior negativity after 200-300 milliseconds followed by a P600, see footnote 5), which has been recorded in a number of studies in relation to L1 syntactic ungrammaticality and has been recently observed also for L2 syntactic violations in highly proficient L2 learners (Friederici, Steinhauer, & Pfeifer, 2002; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012).

***Paradis' Model***. Paradis (2004, 2009) proposed a similar version of the declarative-procedural model. However, in relation to how lexis is stored and processed, he argued for a clearer terminological and conceptual distinction between *vocabulary* and *lexicon*. He claimed that vocabulary should be used to denote the sound-meaning pairing between a word form and the entity it refers to in the world (the kind of information a dictionary would provide). On the other hand, the lexicon refers to the grammatical properties of words, including their morphology as well as the syntactic information they encode (for example how many arguments a verb takes, whether they are direct or indirect, etc.). According to this distinction, only vocabulary would be processed by declarative memory, whilst the lexicon would rely on rule-based mechanisms in procedural memory similar to the ones that apply for syntax at phrase or sentence level. In this respect Paradis (2009) further distinguishes between open-class words (nouns, adjectives, verbs, certain classes of adverbs) and function words, whereby the latter but not the former would be predicted to rely on procedural memory for processing.

Further, whilst in Ullman (2015) the declarative memory can be explicit or implicit (cf. Henke, 2010), in Paradis' model declarative and procedural memory map to the explicit/implicit distinction (i.e., declarative memory implies awareness, and procedural memory implies unawareness). A final point concerns the processing of the L2. Whilst for Paradis procedural processing of the L2 is not excluded but "very rare in practice" (Paradis, 2009, p. 16), the results of adult ERP and behavioral studies deploying miniature language systems have indicated that procedural memory can be engaged in L2 processing already after a relatively limited exposure to a novel language (see e.g., Friederici, Steinhauer, & Pfeifer, 2002; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012).

### 1.6.2 Other cognitive models

***Skill Acquisition Theory***. A model for language processing that was formulated independently of the declarative-procedural model but is related to it is the one outlined in DeKeyser's Skill Acquisition Theory (DeKeyser, 1995, 2007, 2015 for a general overview of skill acquisition theory see Anderson, 2007). In this framework DeKeyser does not mention the memory systems implicated but rather refers to declarative and procedural knowledge, although the two types of knowledge can be understood to denote the information processed and stored in the respective memory systems. DeKeyser considers mainly instructed contexts or situations where second languages are learnt explicitly, i.e. relying on the learners' declarative knowledge of rules and language regularities, or the ability to induce rules.

DeKeyser describes the process of attainment of fluent use of the language through three stages. In the first stage learners rely exclusively on declarative knowledge in language processing/use or may even know the rules of a language without attempting to put them into practice. In the second stage, that DeKeyser calls 'proceduralization', learners draw on declarative and, increasingly, on procedural knowledge as rules start to be practiced "*acting on* this knowledge, turning it into a behavior, turning *knowledge that* into *knowledge how*" (DeKeyser, 2007, p. 95[3]). In the third stage, reached after ample opportunity for reiterated practice of the same or a very similar skill is provided, the practiced language skill gradually becomes fully automatized, i.e. tends to rely completely on procedural knowledge.

As DeKeyser points out it is important to clarify that "turning" declarative into procedural knowledge is a label to describe what is observed behaviorally. However, the

---

3 The italics are mine. The words in italics were in quotation marks in the original text.

nature of the process is more correctly captured by the idea that at automatized stages of language use both types of knowledge are available, although a shift from reliance on declarative knowledge to reliance on procedural knowledge has occurred. Note that in this model declarative knowledge not only supports proceduralized knowledge but "plays a causal role" in its development (DeKeyser, 2009, p. 126). Finally, similarly to the DP model, the Skill Acquisition model suggests an initial role of declarative processing followed by a more prominent role of procedural processing for increasing amounts of practice.

***Usage-based approaches.*** Another model adopting a cognitive approach to language acquisition/learning is the one proposed in N. Ellis (1994, 2005) and N. Ellis and Wulff (2015). Similarly to the DP model this approach is underpinned by current neurophysiological evidence of the relationship between brain activation and cognition, but proposes a different account of the way language is processed, as well as of the role attention and declarative knowledge play in language acquisition/learning. According to N. Ellis, an individual's first contact with a novel linguistic object primarily involves perceptual priming. Although this type of priming is an unconscious form of learning, attention plays a crucial role already at this stage, because it organizes and unifies the representation of perceptual stimuli that will emerge as a pattern, after a sufficient number of exemplars of the linguistic object is processed at sensory-cortical level.

Involvement of declarative memory is then required to establish any relationships between the linguistic form and the meaning associated to it. The established form-meaning relationship is further primed in subsequent encounters/uses of the linguistic form and feeds into the implicit representation of the stimuli proceduralizing the construction. With other cognitive models (e.g., the DP models) N. Ellis's approach shares the idea that, in order to process language, learners employ domain-general cognitive mechanisms, i.e., mechanisms that are not exclusive to language learning and processing. A main role for declarative memory in the initial stages of processing is another common point. However, it is important to note that the involvement of procedural memory in this account operates on *constructions,* i.e. sets of form-meaning pairings, whereby a construction can be any linguistic object ranging from lexis, to morphology, phrase structure or more complex syntactic and pragmatic entities.

In comparison, accounts like the DP model seem to place more focus on a separate role of procedural and declarative memory in processing of rule-based components of language vs. processing of meaning or vocabulary. Finally, N. Ellis's account does not emphasize a competitive relationship between the two memory systems. In his model, declarative memory plays a co-operative role in supporting procedural-memory based language acquisition and learning. Indeed, the main function of declarative memory, as emerges from N. Ellis's account, is to functionally enable and support proceduralized language learning.

***Shallow Structure Hypothesis.*** The Shallow Structure Hypothesis (Clahsen & Felser, 2006; Clahsen & Felser, 2017) is mainly a model of language processing in comprehension. It distinguishes between L1 and L2 language processing proposing that whilst L1 processing consistently relies on morphosyntactic computation, L2 processing in comparison relies to various degrees on nongrammatical information, prioritizing semantics (Clahsen & Felser, 2017, p. 2-3). In a way that is reminiscent of the declarative/procedural distinction in the DP model, the Shallow Structure Hypothesis suggests that both processing routes (morphology driven vs. heuristics driven) are available to L2 learners, and reliance on the first depends on the level of language proficiency (although the authors exclude that L2 processing can ever become completely native-like).

Recently, the interaction between age of L2 acquisition (age of onset) and language processing has started to be investigated in this theoretical framework under the hypothesis that the age at which the L2 began to be acquired is a predictor of type of L2 processing in adults (Clahsen & Veríssimo, 2016; Veríssimo, Heyer, Jacob, & Clahsen, 2017). In particular, these studies found that whilst derivational (lexically-mediated) priming was not related to age of onset, age of onset predicted priming of inflected forms that are not mediated by a lexeme (grammatical inflection).

## 1.7 Behavioral measures of long-term memory

In neuropsychological practice a number of memory batteries have been developed to measure aspects of long-term declarative and procedural memory in clinical and research contexts, although in recent years their use has increasingly been extended to research in neighboring disciplines interested in the investigation of the role of memory skills. An important aspect of many of these batteries is that they are subject to rigorous protocols of administration and assessment and include the use of standardized scores and the comparison with normative data based on a number of variables (e.g., gender, number of years of schooling, socio-economic status, etc.).

Criteria followed to select the tests used in the experiments to measure declarative memory include: (a) the availability of recent norms, and (b) norms based on large samples from the populations under study (L1 Italian 8-9 year old children and adults). The selection also considers validation evidence discussed in the literature, the sensitivity of the type of task to measure the construct as documented in the literature, and the previous deployment of the measures in studies with a similar design and/or using similar artificial language stimuli. In the following sections I will briefly review a selection of long-term memory measures normed for UK and US participants and, when available, the corresponding tests normed for the Italian population.

### 1.7.1 Measures of declarative memory

Tasks that assess long-term declarative memory are further classified into tasks that tap visual or verbal aspects of information retention. In these tests participants are often exposed to visual or verbal stimuli and they are asked to discriminate or recall the stimuli immediately afterwards and at a delayed time. Relative to tests originally developed in the US or in the UK, a relatively recent battery that is widely used with adult populations is the Wechsler Memory Scales (WMS), a comprehensive set of tests that measures aspects of visual and verbal short-term memory and long-term declarative memory (Wechsler, 1945, latest revision 2009). It has the advantage to provide updated normed data based on a large sample (16-90 y.o.) in its original version and to be available in additional versions adapted for a number of different languages/countries. In the battery, aspects of verbal declarative memory are measured through a Paired Associates task and through a Logical Memory task. In the Paired Associates task participants are asked to memorize and recall pairs of related and unrelated words, whilst in the Logical Memory task participants listen to a short story and are asked to recall it immediately and after a delay.

The Doors and People Test (D&P; Baddeley, Emslie & Nimmo-Smith, 1994; latest revised version 2006) is a test developed in the UK that originally specifically measured long-term episodic memory in adults, tapping both visual and verbal components (Davis, Bradshow & Szabadi, 1999). In its latest version the range of normative data includes children from age 5 and adults, which makes it a particularly useful tool to investigate memory across age groups.

In a number of studies published in the last ten years (Carpenter, 2009; Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong, 2014, among others) visual declarative memory in adults was also assessed with a computerized version of the Continuous Visual Memory Test (CVMT). The battery was developed by Trahan and Larrabee (1988) in the US, assesses non-verbal visual memory discrimination and recognition and was validated based on the performance of 92 healthy adults and 138 children.

Based on norms for the Italian adult population, a series of tasks along the lines of the ones available in the Wechsler's battery has been developed to assess declarative memory (Bianchi, 2015). Similarly to Wechsler's Logical Memory task, in the Short Story task (*Raccontino "Anna Pesenti..."*) participants memorize and recall a story immediately and after a delay (Novelli, Papagno, Capitani, Laiacona, Cappa, & Vallar, 1986). In its most recent version (Mondini, Mapelli, Vestri, Arcara, & Bisiacchi, 2011) the test was standardized and normed based on a sample of 702 individuals (15-96 years old). Further tests that assess verbal declarative memory include versions of the Paired Associates Test (Zappalà et al., 1995), or tests that require the memorization of lists of words (for a complete review see Dai Pra et al., 2015).

For visual declarative memory Italian tests based on large normative samples include the Rey-Osterrieth Complex Figure Test (Caffarra, Vezzadini, Dieci, Zonato, Venneri, 2002; Carlesimo et al., 2002) and the Modified Taylor Figure Test (Casarotti, Papagno, & Zarino, 2014). In both tests participants are asked to copy a complex abstract figure including a number of elements and to redraw the figure after a delay as precisely as they can recall it.

Materials for the tests of verbal and visual declarative memory used with children are specifically developed for this age group to cater for the age-dependent differences in cognitive development. The Children Memory Scale (CMS; Cohen, 1997) is a comprehensive measure of learning for children aged 5-16, including a total of 9 subtests measuring three domains: Auditory/verbal memory, visual/non-verbal memory and attention/concentration. In particular, recent studies investigating declarative memory in SLI and typically developing child populations (Lum, Gelgic, & Conti-Ramsden, 2010; Conti-Ramsden, Ullman, & Lum, 2015), have deployed two tests from the auditory/verbal component of the CMS. These tests are 'Stories', where children recall events and details of an orally presented story immediately and after a delay, and 'Word Pairs', which tests the ability to recall word pairs over three learning trials immediately and after a delay.

The currently most comprehensive battery to test long-term memory in children for the Italian population is the PROMEA battery (Prove di Memoria e Apprendimento per l'Età Evolutiva [Developmental Memory and Learning Tests]) developed by Vicari (2007). The battery was validated on a sample of 709 Italian children from 5 to 11 years of age and includes tests of verbal and visual aspects of declarative memory, priming and working memory. Verbal declarative memory is assessed with a word retention task and a short story task similar to the one adopted in the CMS battery. Visual/spatial declarative memory is assessed with a picture discrimination task and a spatial learning task, where children memorize and recall the position of pictures of objects on a four-space matrix (see Chapter 6 for details).

### 1.7.2 Measures of nondeclarative memory and procedural learning ability

The Serial Reaction Times Task (SRT) was developed specifically to tap the implicit learning of new associations and sequence learning (Niessen & Bullemer, 1987). In the original version of the paradigm participants are asked to react as fast as they can to a visual stimulus appearing in different locations on a computer screen by pressing corresponding keys on a keyboard. A random block of trials (baseline) is followed by a series of blocks where a fixed sequence of 10 trials is presented, followed by a final random block. A decrease in reaction times (RTs) in fixed-sequence blocks compared to random presentations is expected if learning of the sequence has occurred by the end of training. As the paradigm established itself in the study of implicit sequence learning it became clear that some participants were able to, at least partially, learn the stimuli sequence explicitly (as evidenced by the fact that they were able to recall parts of the sequence after the experiment).

In order to control for this confound new versions of the paradigm were developed where the detection of the implicit sequence through explicit learning strategies was made more difficult, for example by alternating random blocks with sequence blocks (Meulemans, Van der Linden, & Perruchet, 1998) or by presenting sequence patterns that included random trials (Alternating Serial Reaction Task, ASRT; Howard & Howard, 1997).

Compared to similar tasks, the ASRT in particular has provided a paradigm that is more reliable in filtering out learning effects due to declarative strategies (Song, Howard, & Howard, 2007). In this task the repeated sequence is "hidden" and alternates with a series of random events, so that in each sequence of 8 items a fixed item is followed by a random item, for instance following the pattern 1r2r4r3r. According to Hedenius (2013) the ASRT presents at least two additional advantages compared to the original paradigm. First it allows a clearer separation of general motor skill learning from sequence learning in repeating sequences, due to the possibility of comparing RT decreases relative to the complete 8-item sequence with the decreases relative to the hidden 4-item sequence. Secondly, it allows continuous assessment of RTs as they progressively decrease, with an on-going comparison of RTs in random and nonrandom parts of the sequence across blocks (Hedenius, 2013, p. 44).

Variations of the SRT task have been used extensively in the literature to detect implicit learning (Barker, 2012; Jackson, Jackson, Harrison, Henderson, & Kennard, 1995; Knopman & Niessen, 1987; Robertson, 2007; Smith, Siegert, McDowall & Abernethy, 2001), as well as a behavioral measure of procedural memory functioning in both adults and children (Brill-Schuetz & Morgan-Short, 2014; Ferraro, Balota & Connor, 1993; Hedenius, 2013; Janacsek, Fiser, & Nemeth, 2012; Lum et al., 2010, among others). In some studies, measures of motor control (in the case of Lum et. al., 2010 the MOT, from the CANTAB battery) were administered alongside SRT tasks in order to filter out the effect of motor-control in the evaluation of procedural memory scores.

Another family of tests widely used in the investigation of procedural learning ability are probabilistic tasks such as the Weather Prediction Task (WPT; for the use of this task in studies investigating adult L2 learning see for example Carpenter, 2008; Morgan-Short et al., 2014). In this task participants are asked to guess a weather outcome (sun or rain) based on the presentation of a series of cues. After each trial the participants are given feedback (correct/incorrect). Unbeknownst to them, each cue is assigned a fixed probability to give rise to one or the other weather outcome, so that increasing accuracy in the prediction constitutes evidence of learning of the underlying implicit probability pattern (for a probabilistic tasks adapted for use with children, see Mayor-Dubois, Zesiger, Van der Linden, & Roulet-Perez, 2016). Also in the case of probabilistic tasks, studies have found that the use of explicit strategies may represent a potential confound, included in studies with children. Mayor-Dubois and colleagues for example found that performance in the second part of a probabilistic task was related to the deployment of explicit learning strategies, and increasingly so in older children compared to younger children (Mayor-Dubois et al., 2016, p. 729). In Morgan-Short et al. (2014) and other studies, scores obtained assessing parts of the Tower of London task (TOL) have also been used to calculate composite scores of procedural learning ability, although the task itself has been deployed to assess planning and executive function.

In terms of their neuroanatomical correlates, a number of clinical and neuroimaging studies have shown that both motor/cognitive sequence learning and probabilistic tasks implicate the involvement of striatal and cortical areas in the brain. Specifically, sequence motor learning (SRT tasks) has been associated to the activation of the putamen, and cognitive learning (probabilistic tasks) has been associated to the activation of the caudate nuclei, (for a review of studies see Mayor-Dubois et al., 2016).

The first section of this chapter has presented a currently widely adopted dual model of the architecture of long-term memory, as well as our current understanding of the localization of long-term memory functions in the human brain according to this model. This initial neuropsychological background served to introduce a number of (neuro)cognitive models of (second) language acquisition that assume or are compatible with a dual representation of long-term memory (declarative and procedural memory). Finally, in order to set the basis for the methodological choices made in the experimental studies, the third section of the chapter was devoted to a detailed review of normed behavioral batteries and tasks that have been used as measures of declarative and procedural learning ability in SLA studies and in the neuropsychological literature.

# 2. Language learning processes, language instruction and the nature and measure of language knowledge

## 2.1 Explicit vs. implicit learning

In the previous chapter I discussed neuropsychological evidence in support of the idea that two learning systems, respectively based on nondeclarative and declarative memory, implicate different brain areas and neural circuitry. The aim of this chapter is to introduce the notion of the implicit/explicit distinction as it has been applied to L2 processes, to L2 instruction and to the representation of language knowledge, as well as to provide the methodological background for the investigation of the nature of language knowledge in the two experimental studies.

Although the assumption of two different learning modes is not uncontroversial in cognitive psychology (see for example Jimenez, 2003; Shanks, 2003), the distinction between implicit and explicit learning, first introduced in seminal work by Reber (1967, 1976), has long been a topic of research in this discipline. The implicit/explicit dichotomy was subsequently extended to second language acquisition research (N. Ellis, 1994, 2008; DeKeyser, 2003, 2005; Schmidt, 1994; Williams, 2005). In the domain of language learning the implicit/explicit distinction has also been applied to describe different types of linguistic knowledge (see R. Ellis, 2009; Williams, 2005; Rebuschat, 2013, for a review of studies), different instructional treatments (Lichtman, 2013; Norris and Ortega, 2000; Spada and Tomita, 2010), as well as different ways of providing feedback (Long, 2007; R. Ellis et al., 2009).

Some authors have recently suggested that the neurophysiological (nondeclarative/declarative) and the cognitive (implicit/explicit) accounts may in fact represent two strands of evidence pointing at the same underlying neurocognitive difference. For example, in a recent review of the neural basis of implicit learning and memory Reber (2013) indicated a substantial parallelism between the classical cognitive approach to explicit and implicit learning and the existence of two main long-term memory systems underpinning the two processes, declarative and nondeclarative long-term memory. This view however remains debated, and alternative accounts of the organization of memory functions in the brain (e.g., Henke, 2010) maintain that the engagement of neural structures pertaining to the declarative memory system does not imply consciousness/awareness, a construct central to the definition of explicit learning (cf., 2.1.2).

In what follows I will adopt two definitions of learning common to the cognitive psychology and the SLA literature (N. Ellis, 1994; R. Ellis, 2009; Rebuschat, 2013; Schmidt 2001; Williams, 2005). I will use implicit learning to refer to a type of process that is: (a) incidental (there is no intention to learn on the part of the learner); (b) unaware (the learner is not aware of having acquired new knowledge and typically cannot verbalize what he/she learnt); and (c) automatic (the process is not controlled by the learner). I will use explicit learning to refer to a process, which is intentional, aware (learners can typically verbalize what they learnt) and controlled.

In addition to these differences, implicit learning mechanisms are believed to require a longer time of exposure to the target stimulus, whilst explicit strategies provide shortcuts that allow learning (including second language learning) in conditions of restricted access to input and limited time of exposure.

Although it may be convenient to represent the difference between implicit and explicit processes as a dichotomy, most authors agree that 'implicitness' and 'explicitness' are better understood as features belonging to a continuum (e.g., R. Ellis, 2009; Williams, 2005). For learning, Williams (2005) illustrates the explicit and implicit ends of this continuum in the following way: "At one extreme, there is entirely explicit learning involving conscious comparisons between current and previous instances of input and the formation and testing of hypotheses. At the other extreme, there is entirely implicit learning; the learner has no awareness of either the process or product of learning" (Williams, 2005, p. 271).

### 2.1.1 Artificial language learning paradigms

One of the first paradigms used to investigate implicit learning in cognitive psychology is artificial grammar learning (AGL). Initially the artificial grammars (AGs) were semantically neutral finite-state grammars with a set of rules (syntax) generating strings of letters (Reber, 1967, 1976; Reber, Walkenfeld, & Hernstadt, 1991). The relevance of finite-state grammars studies to second language research has been a source of debate in the SLA literature. Some authors argued that the results of experiments with finite-state grammars lacked ecological validity and their results could not be generalized to natural second language learning contexts, due to the fact that artificial languages could not encode meaning and were not suited to communicative use (Van Patten, 1994; for a more recent discussion of the processing differences between AGL and incidental natural language learning see Robinson, 2010).

Later work, especially research interested in the consequences and role of implicit learning for SLA, started to introduce miniature artificial or semi-artificial languages. These resembled more closely the syntactic and morphological characteristics of natural languages and allowed for simplified semantics, whilst also providing the methodological advantages of early AGs, since they lent themselves to experimental manipulation and to a better control of the experimental conditions.

Artificial languages with natural language characteristics vary relative to the degree and type of resemblance to natural languages and have been mainly used to investigate morphosyntactic and semantic learning. Some studies (e.g., Robinson, 2005) have deployed miniaturized versions of real natural languages belonging to linguistic families unrelated to the learners' L1 and have *de facto* treated them as artificial systems.

In other cases, artificial languages are simplified but fully productive linguistic systems mirroring natural language morphosyntax but using made-up words. In some studies the lexis follows the phonotactic characteristics of the learners' L1 in order to control for the effects of phonological variables on the learning of morphosyntactic aspects of the language (Friederici et al., 2002; Morgan-Short, 2007 and related studies). Although using pseudo-words allows full control over phonotactic and morphosyntactic features, the drawback of this type of study, as well as of studies with real foreign words, is that they require a training phase that, depending on the number of trained words, can be lengthy.

In some other studies artificial languages have been built based on the learners' L1 lexis, whilst incorporating morphosyntactic characteristics of other natural languages (e.g., Rebuschat & Williams, 2012; Grey, Williams, & Rebuschat, 2015). Using L1 lexis in artificial language studies constitutes an important practical advantage in experimental design because long vocabulary training phases can be avoided. Additionally, it can be argued that in these studies individual differences in vocabulary learning do not constitute a confounding variable in the assessment of language learning effects. However, a potential drawback could be that the use of L1 lexis in the artificial language paradigm may trigger L1 interference effects that are difficult to control for. For example, it cannot be excluded that the use of L1 lexis in online language processing automatically activates syntactic representations that are highly correlated with the occurrence of the specific lexical entry and would interfere with learning and processing of the alternative syntactic pattern that is the target of learning.

### 2.1.2 Implicit language learning and awareness

Central to the difference between implicit and explicit mental processes is the construct of *awareness*. In cognitive psychology awareness can be defined in a narrow or broad sense depending on its content, i.e., what the individual is aware of. In the narrow sense awareness refers to an individual's capacity to bring already acquired memory representations or new stimuli under focal attention. There is a consensus in the literature that focal attention plays a central role when learning is intentional and explicit. However, the issue of whether implicit and unaware learning is possible independently of focal attention is still debated. Some authors have suggested that forms of attention that do not necessarily require awareness, referred to as *detection*, may be relevant to implicit learning independently of the engagement of focal attention (Tomlin & Villa, 1994; Williams, 2005). On the other hand, others believe that focal attention is conditional to any kind of learning, implicit or explicit (Jiménez & Méndez, 1999), and that, specifically for linguistic stimuli, memory formation is strictly dependent on noticing the stimulus. According to Robinson (1995) "noticing is defined to mean detection plus rehearsal in short-term memory, prior to encoding in long-term memory" (p. 298, for a review of studies and definitions of the construct of noticing see also Philp, 2003).

In a broader use of the term, awareness refers to the conscious identification of the rules and patterns underlying linguistic stimuli (Williams, 2005). Accordingly, in this case implicit learning would be defined as unaware learning of the rules that govern the underlying linguistic generalizations. Based on the two definitions of the construct it is possible to specify the role awareness plays at difference stages of the representation of linguistic stimuli by distinguishing between awareness at the level of *noticing* and awareness at the level of *understanding* (Rosa & Leow, 2004; Rosa & O'Neill, 1999). For example, in Rosa and O'Neill (1999) awareness at the level of noticing was operationalized as "a verbal reference to the target structure without any mention of rules" and awareness at the level of understanding was operationalized as "explicit formulation of the rule underlying the target structure" (Rosa & O'Neill, 1999, p. 530). This is the type of model that will also be used to operationalize awareness in the analysis of the verbal reports in the experimental studies (Chapter 6 and 7).

As linguistic stimuli in natural languages are complex and combine the realization of rule patterns at different levels (phonological, morphological, syntactic, semantic, pragmatic, etc.), awareness at the level of noticing and at the level of understanding must operate in each of these domains as well as engage in identifying regularities at the interface between them. One such example is the learning of form-meaning mappings, which requires the learner to attend to a linguistic form, its meaning and the relationship between the two. The development of awareness of form-meaning relationships could occur at lexical level (e.g., a simple pairing between a noun and its referent), at morphological level (e.g., a pairing between a morpheme and a grammatical property such as gender or case), or at syntactic level (e.g., a pairing between a given word order and a sentence interpretation). A number of studies have shown that learning of form-meaning relationships can occur at least partially implicitly (without awareness at the level of understanding) for morphological endings encoding grammatical relations (Williams, 2005; Grey et al., 2015; Pili-Moss, 2017) and for syntax-semantics pairings (e.g., Williams & Kuribara, 2008).

### 2.1.3 Statistical learning

In the last twenty years a number of studies have investigated evidence of the human ability to learn sequences and identify structure in a stream of input by tracking its statistical regularities. As statistical learning is available from early infancy (Saffran, Aslin, & Newport, 1996) a growing body of research is currently aiming at elucidating its involvement in the development of early learning skills, including language acquisition. The ability for sequential statistical learning shares a number of similarities with implicit learning and some authors have suggested that they may be referring to the same kind of process (e.g., Perruchet & Pacton, 2006). However, although sequential-statistical learning has been shown to occur without conscious awareness of the patterns presented in the exposure, some authors maintain that explicit strategies and awareness may also play a role.

The literature on sequential-statistical learning has focused on two main domains of investigation: sequence learning as evidence of the learning of abstract rules and sequence learning as a product of the tracking of the frequency of occurrence of individual items in the input. Some authors refer to the latter as ‘surface learning’, suggesting that it may constitute an initial level of information structuring on which more abstract forms of sequence encoding could operate. Another dimension along which sequential-statistical learning has been investigated is modality, whereby more abstract types of sequential learning have been shown to be independent of stimuli modality or be transferable across modalities whilst others are more stimulus-specific.

## 2.2 Explicit vs. implicit instruction

A further domain to which the explicit/implicit distinction is applied is second language instruction. The definitions of explicit and implicit instruction adopted in cognitive psychology and second language acquisition research have reflected the respective research methodologies. In cognitive psychology the role of instruction is relevant in specifying the conditions the participants are exposed to in experiments mainly conducted in the laboratory.

A distinction relevant at experimental level is the one between incidental and intentional learning conditions (see e.g., Berry & Dienes, 1993; Leow and Zamora, 2018). In a paradigm where stimuli are presented incidentally, experimental conditions tend to minimize the participants' opportunities to develop awareness of the language regularities during exposure. This includes omitting explanations relative to the content of the object of learning, the use of filler stimuli or dual tasks. In addition, participants are also not informed that they will be tested at the end of the experiment, with the aim to achieve a good approximation to the ideal condition in which the learner has no intention to learn. In contrast, in intentional learning conditions the learners are informed about the fact that they will be tested, and even in the case of rules not explicitly explained, participants may be invited to actively search for patterns and apply problem solving strategies.

In SLA the implicit/explicit dichotomy is mainly applied to instructed contexts and refers to two different strategies to present new linguistic materials as well as to two different ways of providing feedback. According to R. Ellis (2009) implicit instruction is “directed to enable learners to infer rules without awareness” (p. 16) and is best delivered by maintaining the focus of the learner's attention on meaning in a learning environment where the L2 is used in a communicative situation. On the other hand, explicit instruction directly aims at providing explanations or explicit cues relative to the language structure, or invites the learner to create and test hypotheses on the language rules focusing on its forms. Considering R. Ellis' definition it is clear that awareness in this approach is mainly to be intended as the conscious conceptualization of the rules governing a language, i.e. as awareness at the level of understanding.

A similar implicit/explicit dichotomy applies when instruction is delivered through feedback, whereby some feedback techniques are considered more implicit and others more explicit. For instance, the oral feedback technique known as *recasting* is considered implicit because it entails the reformulation of the learner's nontargetlike expression with a targetlike one without disrupting the flow of communication (Long, 2007). That is, it does not explicitly direct attention to the formal aspects of language maintaining the focus of communication on meaning. Instances of explicit oral feedback would include explicit correction, metalinguistic feedback or elicitation, a technique in which the interlocutor would signal that an expression needs 'repair' and the learner is encouraged to provide a reformulation.

In both laboratory and instructed contexts an important point pertains to the relationship between type of language instruction administered, learning processes and type of language knowledge attained as a result (Figure 2.1, next page). As R. Ellis (2009) notes, the fact that instruction is delivered using an implicit strategy can lead to implicit knowledge (via implicit learning, Figure 2.1, part A). However, the provision of implicit instruction does not guarantee that the knowledge will be also implicit. For instance, it is possible that a learner exposed to a certain linguistic feature in implicit instruction conditions may acquire explicit knowledge of it as a result of the application of explicit learning strategies (e.g., language analysis).

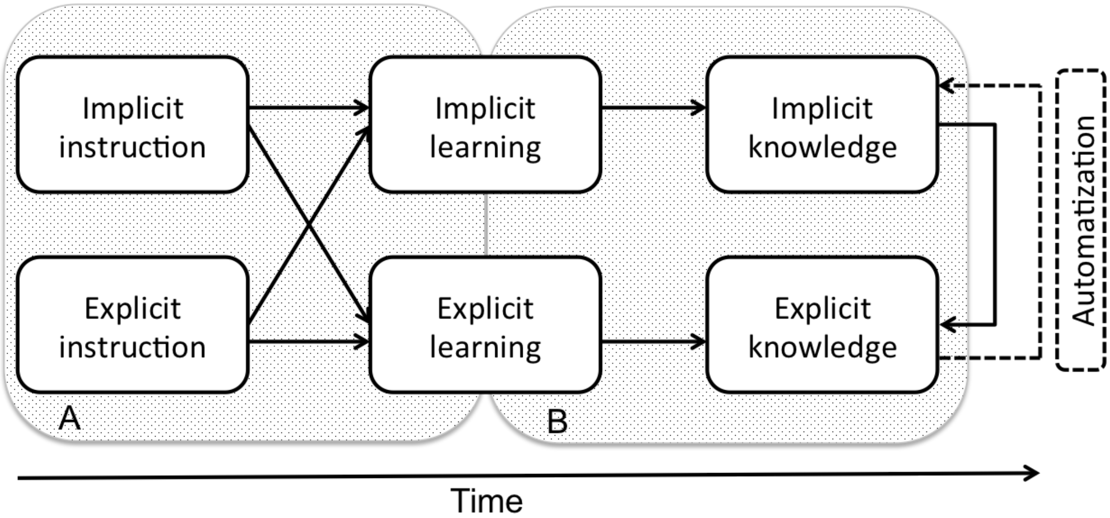


*Figure 2.1*. Interactions Between Instruction, Type of Learning and Type of Language Knowledge in Language Learning Processes.

Conversely, since natural languages are complex systems it is possible that when certain constructions or linguistic features are taught explicitly, the explanation may also provide evidence for other secondary aspects of the language that are related to the explicit object of learning but are learnt and represented implicitly because not directly focused on (from explicit instruction to implicit learning in Figure 2.1). Also at the level of acquired linguistic representations (Figure 2.1, part B), implicit language knowledge can become explicit, as for instance is the case when native speakers become aware of the regularities ruling their own language. In turn, it is possible that, through reiterated use, explicit knowledge is accessed by the speaker in an increasingly automatized way and processed similarly to implicit knowledge. The details of how this 'transformation' can occur are addressed in the next paragraph.

## 2.3 Explicit and implicit language knowledge

After discussing different instruction conditions and different learning processes I now turn to consider in more detail how the language knowledge resulting from them can also be classified according to the implicit/explicit continuum. As discussed in R. Ellis (2005) implicit language knowledge can be defined as the native speaker's ability to process and produce language and judge its grammaticality without conscious introspection of its structural patterns. Implicit knowledge is not represented consciously, hence cannot be verbalized, and is available for automatic and parallel processing with minimal use of working memory resources. On the other hand, explicit knowledge involves awareness of linguistic properties, monitoring and analytic processing. It can typically be verbalized and engages working memory (N. Ellis, 2008; R. Ellis, 2009).

In the SLA literature, the theoretical distinction between implicit and explicit linguistic knowledge has gone hand in hand with a debate regarding how these two types of representations are related (the so-called 'interface issue'; R. Ellis, 2009). According to the *noninterface position* (Hulstijn, 2002; Paradis, 1994) the acquisition of implicit and explicit knowledge hinges on separate and independent neural processes. In particular this hypothesis holds that explicit knowledge cannot become implicit as a result of practice/automatization. In contrast, the *strong interface position* (Sharwood Smith, 1981; DeKeyser, 1998, 2007) maintains that a transformation of explicit knowledge into implicit knowledge (e.g. through automatization) and vice-versa is possible. As suggested by DeKeyser, sufficient amounts of repeated practice can have the effect of speeding up the access to explicit knowledge to the point of full automatization, resembling the effectiveness of language processing and production typical of implicit retrieval. A third theoretical hypothesis is the *weak interface position*. This account allows for the possibility of explicit knowledge transforming into implicit knowledge through practice, provided the learner is developmentally ready (Pienemann, 1989) or by virtue of a mutually supportive interaction between the learning processes responsible for the two types of linguistic knowledge (N. Ellis, 1994, 2005; N.Ellis & Wulff, 2015).

It is important to remember that the notion of knowledge 'transformation' in the definitions of these theoretical models refers to a *behavioral* change in linguistic skills. In the case of DeKeyser's proposal, for example, the 'transformation' would describe the behavioral change produced by the emergence of a parallel and more efficient representation of language knowledge that can be accessed automatically, rather than the replacement of previous linguistic representations (see also 1.7.2).

However, the question concerning which neural structures become engaged in automatized performance and to what extent they are similar to the ones implicated in the retrieval of implicit representations remains open. A clear differentiation between implicit and automatized knowledge may not be easy to determine exclusively on the basis of behavioral evidence. More probably, the teasing out of explicit and implicit knowledge representations will be accomplished by the application of advanced methodologies such as the analysis of neurophysiological responses (ERPs, fMRI, neuroimaging). Further advances in these areas will be able to provide more precise evidence of the type of processing involved in different types of knowledge representation (Morgan-Short, Finger, Grey, & Ullman, 2012; Tagarelli, 2014).

## 2.4 Behavioral measures of language knowledge and language attainment

### 2.4.1 Retrospective verbal reports and subjective measures

In a recent review of the study and measurement of implicit knowledge, Rebuschat (2013) argued for the adoption of methodological standards along the lines of those deployed in cognitive psychology, suggesting the inclusion of a wider range of measures of awareness in the design of SLA studies. Two of these awareness measures are specifically of interest for the thesis; retrospective verbal reports, and subjective measures.

Verbal reports are usually collected at the end of the experiment through structured questionnaires or debriefing interviews. For verbal reports, the assumption in the SLA and cognitive psychology literature is that the ability to verbalize language regularities shows that awareness was involved in the representation of the relevant knowledge. In other words, verbal reports primarily provide direct evidence of which aspects of language knowledge are represented explicitly in the learner's mind.

An above-chance performance on aspects of the linguistic task that subjects could not verbalize in retrospective reports has been also considered indirect evidence that the relevant language knowledge was represented implicitly (Dienes, Broadbent, & Berry, 1991; Reber, 1967). However, since the ability to verbalize conscious knowledge differs among individuals, verbal reports by themselves cannot be considered a reliable diagnostic for the identification of implicit knowledge (see among others Dienes & Berry, 1997).

Similarly to cognitive psychology, the degree of language awareness and the ability to verbalize knowledge in learners' reports are widely used in SLA methodology to assess the nature of linguistic knowledge. Retrospective techniques are probably the most common tool used in SLA research to investigate if subjects noticed or were conscious of explicit patterns in the instructional materials during exposure. In the case of retrospective verbal reports this is accomplished through a structured questionnaire handed out to the subjects at the end of the experiment (Francis, Schmidt, Carr, & Clegg, 2009; Gass & Mackey, 2000; Mackey and Gass, 2005; Rebuschat & Williams, 2012). Other possibilities include a stimulated recall where knowledge of a number of aspects of task and linguistic performance is elicited in a structured interview at the end of treatment. In some cases the recall is aided by showing participants a video of the task performance stopped at specific points, and by asking them to verbalize what they were thinking at that time (Egi, 2004; Gass & Mackey, 2000, 2007).

A further type of awareness measure that to date has been more extensively used in cognitive psychology compared to SLA studies is subjective measures. These can be confidence ratings or measures of source attribution (Dienes, 2004, 2012; Dienes, Altmann, Kwan, & Goode, 1995; Dienes & Scott, 2005). Subjective measures can be built in the design of a grammaticality judgment test (GJT) or other outcome measure of language learning. For example, confidence ratings used in a GJT would assess after each trial how confident participants felt in their judgment (not confident, confident, very confident, etc.; see also Loewen, 2009).

Two techniques, the *guessing criterion* and the *zero-correlation criterion* have been used to assess the explicit/implicit (conscious/unconscious) status of knowledge about trial judgments. According to the guessing criterion, above-chance performance on trials where subjects affirmed to be guessing indicates the implicit status of judgment knowledge (i.e., subjects demonstrate knowledge behaviorally but are not aware that a certain string has the same structure of a training string; Dienes & Scott, 2005). According to the zero-correlation criterion the nature of the judgment knowledge is revealed by the correlations between confidence ratings and trial accuracy. In particular, the absence of a positive correlation between confidence level and accuracy in the response, e.g. low confidence vs. high accuracy provides evidence that the relative judgment knowledge is implicit (Chan, 1992; Dienes et al., 1995). It is important to note that (a) both criteria measure the implicit/explicit status of judgment knowledge, not directly the status of the knowledge of the linguistic regularities (structural knowledge; Dienes & Scott, 2005), and (b) both judgment and structural knowledge can be explicit or implicit. Consider for example the case of an L1 English speaker judging the grammaticality of sentences in her own language. Typically, she will have high confidence and high accuracy in the judgment, indicating explicit knowledge in the judgment, but the relevant structural knowledge could be explicit (she knows the rule of her language explicitly) or implicit (she only has the correct intuitions).

Dienes and Scott (2005) hence conclude that whilst "unconscious structural knowledge can be inferred from unconscious judgment knowledge...conscious judgment knowledge leaves the conscious status of structural knowledge completely open (p. 340)". Finally, measures of source attribution investigate the subjects' beliefs relative to the source of their linguistic knowledge (guess, intuition, rule, memory). As their administration requires participants with a more sophisticated level of metacognition compared to confidence ratings, they are only briefly mentioned in this review but will not be used in the experimental paradigm with primary-school children.

### 2.4.2 Timed vs. untimed GJTs and other measures of language knowledge

SLA studies aiming at investigating the implicit/explicit nature of language knowledge have devised tasks aimed at eliciting one or the other type of knowledge by manipulating a number of dimensions, including for example the time available to perform the task and the focus of attention during the task (for a detailed discussion see R. Ellis, 2009, pp. 38-39). It is important to note that, unlike subjective measures, these tasks are primarily measures of different aspects of morphosyntactic attainment and are not designed to directly assess awareness of linguistic rules. Rather, they are assumed to facilitate/limit awareness of the structures in the linguistic input to different degrees due to their specific design or mode of administration.

For example, the amount of time available to the learner to perform a linguistic task has been a criterion widely deployed in the field to discriminate between tasks that engage implicit or explicit language knowledge. The validity of this criterion is based on the assumption that drawing on explicit knowledge during online language processing requires more time compared to engaging implicit language knowledge, due to the fact that, unlike implicit knowledge, explicit knowledge is not automatically available.

A representative example of one such task is the grammaticality judgment test (GJT). In the GJT learners are asked to give a judgment on the grammatical acceptability of the linguistic stimuli they are exposed to during the test. According to the time-pressure criterion, if the time available to provide a response is sufficiently limited (and the response is correct), it is more likely that participants will draw on implicit knowledge of the linguistic feature. By contrast, if the GJT is untimed, correct responses tend to reflect the result of correct reasoning during task performance based on explicit knowledge (R. Ellis, 2005, 2009; Han & R. Ellis, 1998). Analyzing the eye-tracking scanpaths of sentences in time-pressured and untimed written GJTs, Godfroid, Loewen, Jung, Park, Gass, and Ellis (2015) confirmed the relevance of time pressure as a factor discriminating between less or more controlled types of L2 knowledge and suggested that these "could correspond to implicit and explicit knowledge" (p. 270).

However, as Godfroid et al. admit, time pressure as the sole criterion to detect implicit language knowledge fails to distinguish between fast performance that genuinely reflects implicit knowledge and fast performance that is the result of explicit knowledge automatization. Suzuki (2017a) has recently maintained that, particularly in educated adults that acquired the L2 in instructed contexts, form-focused tasks (including the GJTs) inherently draw on explicit knowledge, suggesting that in the case of time pressure attainment on this task can be at best indicative of automatized explicit knowledge. He suggested that, by contrast, real-time meaning-focused comprehension tasks (visual-world task, word monitoring and self-paced reading) more reliably draw on implicit knowledge because "they indirectly measure grammatical sensitivity without asking for grammaticality judgments" (p. 1233).

Another task that has been argued to tap implicit language knowledge is elicited imitation (Erlam, 2006; 2009). In elicited imitation participants are guided to focus on meaning and asked to repeat aloud grammatical and ungrammatical complex sentences after a delay, in order to avoid verbatim repetition in individuals with sufficiently large working memory spans. In these conditions repetition is thought to rely mainly on syntactic reconstruction of the meaning conveyed by the initial sentence stimuli. In the case of ungrammatical stimuli, correct repair of the nontargetlike structure constitutes evidence of implicit knowledge of the relevant linguistic feature. Although a number of studies have contributed to the validation of elicited imitation as an index of implicit linguistic knowledge so far (see e.g., Spada, Li-Ju Shiu, & Tomita, 2015), some authors (e.g., Suzuki & DeKeyser, 2015) have recently questioned this view, suggesting, again, that rather than being a measure of implicit knowledge, elicited imitation is more likely to assess automatized explicit knowledge.

### 2.4.3 Issues in GJT design

Although the GJT will not be used as a measure of implicit/explicit L2 knowledge in the experimental studies reported in Chapter 6 and 7, it will be used as a measure of morphosyntactic attainment. For this reason, it is useful to review the task in some more detail here, in particular with regard to issues relating to its design and use with adults and children.

In both adult and child L2 learning studies the grammaticality judgment test (GJT) represents one of the most widely deployed test instrument to measure language attainment. Grammaticality judgment tests can be administered auditorily and/or visually, in a computerized environment or 'pen and paper' mode, or presented as a play activity (in the case of studies with young children). They can be manipulated in a number of ways to yield information about the knowledge of specific syntactic domains or grammar features, the type of knowledge attained, the level of knowledge automatization and the subjective confidence or knowledge of the source of judgment.

In their most common format GJTs expose participants to the same number of ungrammatical (ill-formed) and matched grammatical (well-formed) syntactic units to which a certain number of fillers may be also added to take the participant's focus away from the linguistic target of testing. Participants' performance is also compared to chance performance, to make sure that the correct responses are a reflection of genuine learning and are not simply the result of a successful guessing strategy.

As already discussed, timed and untimed GJTs have been considered to be more reliable indexes of implicit and explicit language knowledge respectively (Bialystok, 1979; Han, 2000; Loewen, 2009). For untimed GJTs some authors have suggested that correctness in the GJTs has to be interpreted differently for grammatical and ungrammatical sentences. Correct judgment of ungrammatical sentences would depend more on the engagement of explicit language knowledge, as incorrect sentences invite reflection on the reason of the ill formedness, once the error is detected (R. Ellis, 1991). If the GJT is timed (the time allowed has ranged between 3 and 10 seconds per trial across studies), the correctness is in general expected to draw more on the learner's implicit knowledge, as sufficient time for reflection is not provided.

In addition to a limited time for response, computerized GJTs (such as the ones administered through E-Prime or similar software to program experiments) can also provide a latency measure (the time in milliseconds elapsing between the response request and the response). Further, GJTs can be designed to include additional features, e.g. Likert scales for confidence ratings or source attributions for which latencies can also be measured.

An important point that has been raised with regards to GJT scores validity relates to the widespread use of dichotomous acceptability responses. In general, and specifically for children, it has been argued that the request of a response presenting a binary choice is biased towards a 'yes' or 'correct' response (McDaniel & Cairns, 1996). A common technique used to counterbalance this bias is the computation of so called *A'* (or *d'*) scores, calculated considering the proportion of hits and false alarms in the data instead of simply reporting hits in the two categories (Saxton, Dockrell, Bevan, & van Herwegen, 2008).

Another alternative that has been proposed is the use of graded scales for grammaticality judgments. In adult studies graded grammaticality judgments have deployed magnitude or Likert-type scales with numerical values the participants are asked to select to express a more nuanced judgment compared to the binary option. More recently studies like Ambridge, Pine, Rowland, & Young (2008) have implemented this idea in a more child-friendly format presenting the points of the Likert scale on a test sheet as smiley faces (see also Theakston, 2004). In Ambridge et al. (2008) a five-point scale was used with two green smiling faces on the rightmost side and two red frowning faces on the leftmost side. The face in the middle was half green and half red with a neutral facial expression. The use of the color-coding in addition to the graded scale allowed for a combination of graded and binary acceptability judgment. The study deployed the same GJT paradigm comparing 5-6 year olds, 9-10 year olds and adults. However, since it was not clear whether the younger child group could provide a graded judgment, these children were asked to give a binary judgment first and specify a grade in the scale only subsequently. The authors were confident that both the older children and the adults could easily perform the graded judgment task directly, without the need to provide a binary judgment first.

Metacognition and subjective confidence in children has recently received specific attention in the literature on implicit and explicit learning with the extension/adaptation of assessment instruments used with adult participants to studies with children of primary school age (e.g., Bertels, Boursain, Destrebecqz, & Gaillard; 2015; Fritzsche, Kröner, Dresel, Kopp, & Martsch, 2012; Koriat & Ackermann, 2010). Investigating visual statistical learning in 9 year-old children and adults and in association to a visual triplet-completion task, Bertels et al. (2015) used a binary confidence rating with verbal labels; the participants were instructed to select "guess" if they felt they were answering at random or "remember" if they felt the choice was made on the basis of some form of recall (p. 3). In Koriat and Ackermann (2010) three groups of children of 8, 9 and 11 years of age were presented with a set of age appropriate general knowledge questions and asked to select one of two answers via a computer program. After each item they were asked how confident they were about their choice using the thermometer paradigm, whereby they had to slide a pointer on a scale, which was automatically converted into a percentage confidence score.

## 2.5 The nature of child language learning and language representations

### 2.5.1 Developmental aspects of implicit and explicit language learning

According to widely accepted views about how implicit and explicit learning develop at different ages, implicit learning is a fairly stable ability from early on in childhood and is maintained throughout the lifespan, whilst the ability for explicit learning develops around middle-childhood and parallels the emergence of more complex cognitive skills. However, the literature on implicit/statistical learning has found mixed evidence with regards to the developmental trajectory of these abilities and these issues remain debated.

Some of the developmental evidence relative to implicit learning has already been addressed in the comparison of children and adult cognition earlier (see 1.6). Whilst a number of studies support the view that school-aged children and adults do not differ significantly in their implicit learning abilities, other studies have found child-adult differences, as well as differences among children of different age groups (see for example Thomas et al., 2004). Studies investigating statistical learning (SL) abilities in different age groups also found mixed results. For example, Saffran, Newport, Aslin, Tunick, & Barrueco (1997) found that 6-7 year olds did not significantly differ from young adults in the ability to segment words in a stream of synthetized speech flow, a task indicative of implicit statistical learning abilities.

However, other studies with school-age children found an advantage for adults in implicit conceptual priming (Barry, 2007; Mecklenbräuker, Hupbach, & Wippich, 2003). In a more recent fMRI study McNealy, Mazziotta, & Dapretto (2010) compared adults and 10 year-old children in the parsing of streams of concatenated syllables containing high statistical regularities. Behaviorally the two groups did not differ significantly. However, the study found a larger engagement of dorsal parietal and superior temporal areas (including those subserving attentional networks and working memory) in children, whereas network recruitment in adults was more localized in cortical areas known to be implicated in language processing. McNealy et al. hypothesized that children's greater engagement of working memory and attention during language parsing could constitute a selective advantage because it would facilitate the tracking of transitional probabilities in real acquisition/learning environments.

A possible way to account for the two strands of evidence in the literature on implicit/statistical learning comes from recent theoretical approaches describing statistical learning as a layered set of abilities, with each subset following parallel and (at least partially) independent developmental trajectories. Daltrozzo & Conway (2014) for example distinguish between a *basic* SL mechanism and a parallel, and more refined, *expert* SL mechanism (p. 6). The basic system would be responsible for implicit and automatic bottom-up operations, such as the ones involved in sequence learning depending on the processing of transitional probabilities; the expert system would rely on relatively more explicit top-down mechanisms and on focal attention, and would be responsible for the processing of more abstract sequence patterns.

They propose that the availability of the two systems for sequential/statistical learning varies at different stages of human development, with the basic mechanism posited to be the solely responsible for sequential learning in newborn infants (the authors cited aural learning studies with infants from around 6 months of age). Starting from infancy, and through childhood and adulthood, the expert system would then become increasingly more relevant, until the trend is reversed in older adults, possibly in connection with a weakening of working memory functions.

A number of authors have also questioned the view that explicit knowledge emerges relatively late in development and have suggested that early forms of explicit learning are present in young children even if they may have different characteristics compared with the skills observed in older children and adults (for example more limited verbalization). For instance, in a study investigating children's beliefs about knowledge held by individuals they observed (epistemic beliefs) Matsui, Miura, & McCagg (2006) found that 6 year olds were able to verbalize the source of somebody else's knowledge when asked. Four year olds performing the same task could not verbalize the source of knowledge, but there was behavioral evidence that they could recognize it explicitly.

Theoretical approaches to the structuring of explicit knowledge such as Karmiloff-Smith (1992) or Dienes & Perner (1999) go beyond a simple dichotomous model that distinguishes between implicit and consciously aware knowledge and "claim that explicitness is multifaceted and multi-leveled" (Matsui et al., 2006, p. 1789). Karmiloff-Smith (1992) introduced the representational redescription model (RR), a paradigm that "attempts to account for the way in which children's representations become progressively more flexible, for the emergence of conscious access to knowledge, and for children's theory building (Karmiloff-Smith, 1992, p. 17)".

In Karmiloff-Smith's (1992) developmental model of child knowledge representation four different levels are distinguished: implicit, explicit unconscious, explicit conscious but not verbalizable, explicit conscious and verbalizable. The phases do not correspond to age-related stages in development but apply cyclically and independently for each specific knowledge domain. Furthermore, Karmiloff-Smith did not exclude that the process may extend also to adults for new kinds of learning. In phase one (procedural encoding) knowledge is exclusively data driven and leads to representations that are created and stored in an additive fashion, with no possibility to establish links between them. In phase two operations start to apply to already established internal representations, with an initial temporary tendency to disregard external input (giving rise for instance to phenomena like overgeneralization). In the last two phases knowledge becomes first conscious and then verbalizable, thus reaching the highest level of explicit representation.

In summary, at each level the same knowledge undergoes a process of recoding that creates a higher and more flexible level of information representation maintaining the former one in a system "that indeed may turn out to be a very redundant store of knowledge and processes (Karmiloff-Smith, 1992, p. 23)". Importantly, this model predicts not only that explicit knowledge can be available independently of the child's conscious access to it, but also that explicit knowledge that can be accessed consciously may not yet be verbalizable.

### 2.5.2 Evidence from SLA research

A view commonly held in the SLA literature is that adults and children differ in the learning strategies they deploy in second language learning, with children tending to adopt implicit strategies and adults mainly learning explicitly. However, a number of recent studies that investigated second language learning in instructed contexts found evidence of the important role played by the availability of explicit strategies in child learning.

For example, Lichtman (2016) investigated learning of a miniature artificial language by 5 to 7 year olds and adults under implicit and explicit instruction conditions to test whether age or type of instruction was the most relevant factor in learning. In general, adults were better learners independently of the instruction condition. Investigation of the level of explicit rule knowledge developed by the participants through the training also showed that the only significant difference emerged in the implicit instruction condition, where adults developed a higher level of language awareness compared to children.

In the explicit instruction conditions no significant differences emerged, which supports the view that (relative to the investigated age range) explicit instruction can aid the development of explicit language representations in both children and adults independent of age. In addition to this, the study also found that adding explicit cues to an oral production task enhanced accuracy independent of age. These results confirm the classroom findings in Milton & Alexiou (2008), Suárez & Muñoz (2011), Tellier & Roehr-Brackin (2013), studies that contributed to the evidence of a positive relationship between explicit language instruction and the development of L2 language analytic skills in young school-aged children from age 6 onwards.

This chapter has illustrated the implicit/explicit distinction as it has been applied to language learning, language instruction and the definition and measure of language knowledge in cognitive psychology and SLA studies. Aiming at providing a foundation for the experimental methodology in Chapter 6 and 7, a section of the chapter has been devoted to the discussion of behavioral measures that have been adopted in the literature to probe the implicit/explicit nature of language knowledge. The chapter concluded with a review of theoretical, laboratory and classroom studies exploring the implicit/explicit distinction in the representation of child language knowledge.

# 3. Age effects and theoretical approaches to morphosyntactic learning in SLA

## 3.1 L2 acquisition in children and adults: Ultimate attainment vs. rate of learning

In this chapter I introduce the age variable and its relationship with second language learning/acquisition, an issue that has long been a major topic of debate in SLA. In the light of the experimental studies presented in this thesis (Chapter 6 and 7), the aim of this chapter is to provide a theoretical background for a comparison of the studies along the age dimension (Chapter 8).

In the last forty years the SLA literature on age effects on L2 acquisition and processing has focused around two main issues; the study of ultimate attainment and the study of rate of learning. Ultimate attainment studies have investigated the end state of L2 development in learners exposed to the language (usually in immersion conditions) for a period of time of 10 years or longer (the time after which most researchers agree limited or no further development in the L2 occurs; DeKeyser, 2013). In these studies it is assumed that the L2 acquisition process has been completed at the time of testing and the main predictor of interest is the *age of onset* (AO), i.e. the age at which learners started immersion.

Studies on ultimate attainment have sometimes used the level of proficiency of native speakers as a baseline or deployed native speakers as judges of L2 performance, and have consistently provided evidence of an advantage for early starters. For example, they have found that, in comparable exposure conditions, learners that started immersion in early childhood reached a higher level of ultimate attainment in the language compared to adult starters (e.g., Abrahamsson, 2012; Abrahamsson and Hyltenstam, 2009; DeKeyser, 2000). Some of the authors (e.g., Bley-Vroman, 2009; DeKeyser 2012, 2013; DeKeyser and Larson-Hall, 2005; Granena and Long, 2013a; Gregg, 1996; Hyltenstam and Abrahamsson, 2003; Munninch and Landau, 2010; Newport, 2002; Veríssimo et al., 2017) have also maintained that, although different developmental trajectories are still possible in individual cases, on average ultimate attainment will likely not reach native levels if the onset of L2 exposure occurs beyond certain age boundaries that may vary depending on the specific L2 skill considered.

In contrast with the 'maturational' approach, other authors have minimized the exclusive role of age of onset in predicting proficiency in ultimate attainment and have argued that individual differences other than age, as well as contextual variables, represent the main factors in accounting for what would *prima facie* appear to be age-related variation (Bialystok and Miller, 1999; Hakuta, 2001; Hakuta, Bialystok, and Wiley, 2003; Birdsong 2005, 2006; Herschensohn, 2007). Reviewing the findings on both sides of the debate, some authors have suggested that focusing on how the age variable and the additional individual and contextual variables interact in shaping the observed L2 learning effects across age groups is the methodological approach that would maximally benefit research advancement in this area of investigation (e.g., DeKeyser, 2013). A review of individual and contextual variables that have been found to have an effect in second language acquisition alongside with age is presented in section 3.3.

The second dimension along which age differences in L2 learning have been studied is *rate of learning*, an area of research where studies have been scarcer compared to ultimate attainment. Also, although some studies have investigated rate of learning in naturalistic environments (e.g., Snow & Hoefnagel-Hoehle, 1978), most research in this area has been conducted in instructed contexts. In rate of learning studies it is assumed that the learning/acquisition process is still in progress, and an important variable alongside age of onset is *age at testing*, i.e. the age at which participants' L2 skills are assessed. In this type of study participants matched for amount of L2 instruction but with varying age of onset are trained over periods that go from a few days to a few years. Particularly in longer longitudinal studies, learners are tested at multiple points during the instruction period and their L2 performance is measured on a range of linguistic skills.

In contrast with the picture emerging from ultimate attainment studies, most studies on rate of learning have found that adults (and adolescents) tend to perform better than children, both in terms of L2 learning measures and in terms of the time they require to reach the measured level of attainment. However, a number of studies also found that the advantage seems to be more marked on measures of grammar attainment compared to other L2 skills such as listening comprehension (Muñoz, 2003; 2006a) or oral skills (Cenoz, 2002). One possible line of explanation for this pattern of results may be that, compared to younger children, older children and adults display more developed language analytic and problem solving skills due to their more advanced cognitive development. These skills would support better performance in form-focused tasks of the kind typically deployed to probe grammatical attainment in instructed contexts. By contrast, to the extent that these are focused on meaning and performed in real time, comprehension and oral tasks may offer less opportunity for explicit language analysis, with the consequence that the attainment gap between younger and older learners may be reduced in these areas.

In rate of learning studies conducted in naturalistic conditions, older starters have been also found to display an initial advantage in morphosyntactic development (Snow & Hoefnagel-Hoehle, 1978), although in these contexts the L2 attainment gap between early and late starters tends to close relatively fast compared to instructed contexts. As instructed rate of learning studies are particularly relevant to the investigation pursued in the present dissertation, they will be reviewed in detail in section 3.4. An additional body of research that has looked at age differences in instructed contexts in laboratory conditions will be discussed in Chapter 4.

## 3.2 Individual and contextual variables moderating age effects

### 3.2.1 Cognitive Individual Differences

***Aptitude.*** Aptitude for explicit L2 learning, is one of the most studied individual differences that has been claimed to act as a potential confound in age-effect studies. The interest in the modulating role played by this type of aptitude, broadly defined as the ability to apply analytic skills and hypothesis testing to linguistic input (DeKeyser, 2000; Harley & Hart, 1997, among others), has a long tradition that goes back to the testing work of Carroll and Sapon (1959; MLLAT Test).

A consistent result in this strand of research has been that, in individuals that have started to be exposed to the L2 as adults, high levels of aptitude for explicit learning (see also Chapter 2) are related to better L2 attainment (Abrahamson & Hyltenstam, 2008; DeKeyser, 2000; DeKeyser, Alfi-Shabtay, & Ravid, 2010, Granena & Long, 2013b). Comparing child and adult starters Harley and Hart (1997) found that aptitude for explicit learning correlated with L2 attainment only in the latter group, that is higher analytic ability was related to better L2 attainment in adults, but not in children. However, in another comparative study, Abrahamson and Hyltenstam (2008) found that aptitude was positively correlated with better language attainment also in early immersion bilinguals, and concluded that '[aptitude] plays not only a *crucial* role for adult learners but also a *certain* role for child learners' (p. 499).

More recently, some SLA researchers have set out to broaden the definition of aptitude, either by defining the construct in terms of ability clusters (aptitude complexes; Robinson, 2001) or by validating batteries including a range of tasks designed to tap both aptitude for explicit and implicit language learning in adults (for example the LLAMA battery in Granena, 2013 and the HiLAB battery in Doughty, 2013). Recently, the use of the LLAMA test has been extended to 10-11 years old children (Rogers et al., 2016), whilst the HiLAB test has been validated only for adult learners to date.

***Memory.*** Related to aptitude, the second major type of individual differences potentially interacting with age is cognitive variables measuring different aspects of memory performance (see also Chapter 1). For example, a number of studies have found short-term memory or the central executive to be positively related to L2 learning in children (e.g., Service, 1992, Nicolay & Poncelet, 2013), in adolescents (e.g., Kormos & Sáfár, 2008) and in adults (e.g., Atkins & Baddeley, 1998; Indrarathne & Kormos, 2018; Williams, 1999; Williams and Lovatt, 2003).

Cognitive abilities depending on long-term declarative and procedural memory have also been deployed as individual differences in second language learning (for a review, cf. Buffington & Morgan-Short, 2018). Starting with the work of Carpenter (2008) and Morgan-Short et al. (2014), declarative and procedural learning ability have been found to significantly predict L2 learning depending on L2 proficiency (e.g., Hamrick, 2015; Morgan-Short et al., 2014), type of instruction (e.g., Brill-Schuetz & Morgan-Short, 2014), type of linguistic target (e.g., Antoniou, Ettlinger, & Wong, 2016), context of exposure (e.g., Faretta-Stutenberg & Morgan-Short, 2017) and spacing (Suzuki, 2017). A more detailed review of a selection of these studies is provided in Chapter 4. Although theoretical frameworks such as the DP model make predictions regarding the engagement of different memory systems in processing L2 information at different ages, no child study to date has investigated the relationship between age and declarative and procedural learning ability in L2 learning (cf. Hamrick, Lum, & Ullman, 2018).

### 3.2.2 Affective and personality-related factors

Affective and personality-related factors are believed to predict second language learning as much as cognitive factors (e.g., Otwinowska & De Angelis, 2012, p. 347) or even to a greater extent than cognitive factors (e.g., Kormos, 2013, p. 147). Although the analysis of affective and personality-related factors is beyond the scope of the present thesis, these individual differences are here briefly mentioned as additional potential moderators of age effects. Among affective factors the most studied individual differences are motivation to learn (Dörnyei & Ushioda, 2009), anxiety, and self-confidence. Among the personality-related factors that have been shown to play a significant role in L2 learning are openness to experience, conscientiousness, and extraversion.

### 3.2.3 Type of instruction

Unlike adult L2 teaching, current L2 teaching practices with young children (at least up to middle-childhood) tend to favor instructional techniques where implicit methodologies are used (e.g. songs, stories) and to limit the use of explicit instruction (e.g., Torras, Naves, Celaya & Perez-Vidal, 2006). The beneficial effects of explicit instruction for adult L2 learning are well-attested in the SLA literature (Norris & Ortega, 2000; Spada & Tomita, 2010). However, recent studies conducted in instructed contexts have shown that explicit instruction has a positive effect not only for adult but also for child L2 learning (e.g., Lichtman, 2016). Further, recent research has consistently found that type of instruction is a significant factor in second language learning, that in some cases predicts language learning better than age alone (Llanes & Muñoz, 2013; Pfenninger, 2014; Pfenninger & Singleton, 2017). For example, Pfenninger & Singleton (2017, reviewed in 3.3.1) examined the effects of CLIL (Content and Language Integrated Learning), a classroom-based instructional approach where the L2 is not only the object of learning but is also used as a medium to deliver instruction in other subjects in the school curriculum. Specifically, they compared the effects of type of instruction (CLIL vs. non-CLIL) with the effect of age of onset and found that whilst CLIL predicted accuracy in a number of areas (listening comprehension, vocabulary, written accuracy and complexity), age of onset by itself was not predictive of accuracy in any of the measures.

### 3.2.4 Input effects

Closely related to type of instruction, there is a consensus that type and amount of input are also crucial factors in enhancing L2 processing and L2 proficiency, with some evidence of interaction between amount of input and age. Concerning type of input, Faretta-Stutenberg and Morgan-Short (2017) have recently compared a group of L2 Spanish university students instructed in the US with a group of students from the same university who studied abroad in a Spanish speaking country for a period of 12 to 15 weeks. Deploying a pre-posttest design, they found that learners in both conditions significantly improved in a GJT and were comparable in proficiency at posttest. However, procedural learning ability significantly predicted morphosyntactic gains as well as increases in the magnitude of ERP responses only in the study-abroad group, suggesting that, unlike at-home students, these learners relied on the procedural memory system for morphosyntactic processing as a result of the immersion experience. Although the literature indicates that naturalistic contexts are especially effective for child L2 learners' ultimate attainment, studies that looked at L2 learning in immersion contexts (e.g., Harley, 1986; Genesee, 1987) did not find a proficiency advantage for early starters compared to learners that had started immersion in secondary school. However, to the best of my knowledge studies that investigated age differences in immersion contexts did not look at cognitive individual differences and it is possible that the effects of protracted periods of immersion might more clearly emerge in language processing rather than in measures of proficiency.

Other studies have found that when the effect of the amount of input is partialled out, age differences disappear or become negligible, and that amount of input interacts with age. Larson-Hall (2008, second research question) explored age effects in 200 Japanese university students (*M* = 19 years) that started learning English in instructed contexts before or after age 12. The amount of input the students had been exposed to varied depending on age of onset as well as with the reported amount of additional study. Comparing the two groups and partialling out the effect of amount of input, the study found no significant between-group differences in a GJT and a small-effect advantage for early starters in a phonemic discrimination task.

However, the analysis of GJT scores as a function of amount of input showed a significant advantage of late starters over early starters if input was relatively low (< 800 hours), whilst early starters attained significantly better scores between about 1600 and 2200 hours, after which differences became nonsignificant. Similarly, a significant advantage for early starters in the phonemic discrimination task emerged between 1200 and 2200 hours. Overall the study highlighted the role of input amount in moderating age effects and showed that a substantial number of hours of exposure may be needed for significant advantages for early starters to emerge in instructed contexts.

Support for the role of amount of input in predicting proficiency was provided also by more recent research. Muñoz (2011) tested 162 undergraduates at a university in Catalonia (*M* = 21 years) on their L2 English attainment, using a standardized general proficiency test, two lexical tests and a phonetic identification test. The students' age of onset ranged from 2 to 15 years (*M* = 7.8) and the amount of instruction they had received at testing corresponded to about 2400 hours on average. The study found no significant correlation between age of onset and L2 proficiency on any of the measures and no significant between-group differences with an onset cut-off set at age 11. By contrast, the study found that proficiency outcomes significantly correlated with amount of input in a number of measures including total length of instruction and recent formal and informal contact with the L2.

## 3.3 L2 proficiency and rate of learning in instructed contexts

### 3.3.1 Classroom-based longitudinal studies

As already mentioned, compared to the body of research available for ultimate attainment, a smaller number of studies to date have investigated the role of age for rate of learning in instructed contexts. The first set of studies considered here is the one generated by the BAF (Barcelona Age Factor) project in Catalonia. This longitudinal project, initiated in the mid 1990's, ran over six years and had the aim of tracking the development of L2 proficiency in two cohorts of EFL secondary students who had either started L2 instruction during primary school (at age 8), or at the beginning of secondary school (at age 11). The two cohorts were tested after the same amount of instruction at three times corresponding to 200, 416 and 726 hours of instruction on a range of tasks that probed the four literacy skills in the L2 and required the engagement of different levels of cognitive ability. The tasks included dictation, cloze, listening comprehension, a grammar test, written composition, an oral narrative and an oral interview, phonetic imitation and discrimination tests and a role-play.

As reported in Muñoz (2006a) in one of the project's studies, late starters performed better than early starters in the dictation and cloze tasks at both Time 1 and Time 2, although they were significantly better than early starters in the listening comprehension task only after 416 hours of instruction. Significant group differences also emerged when rate of learning between Time 1 and Time 2 was considered, with late starters showing an advantage particularly in the cloze test. The fact that the advantage of late starters seemed to be reduced in the listening comprehension task is reminiscent of similar findings in previous studies comparing L2 learning in groups with different age of onset (e.g. Cenoz, 2002; 2003; Lapkin, Swain, Kamin, & Hanna, 1980; Muñoz, 2003).

Mora (2006) was another BAF study that looked at differences in L2 attainment on an oral picture task testing a sub-sample of the same participants after 726 hours of instruction (Time 3). He found that whilst late starters on average had a significantly higher speech rate and lower use of L1, early learners had a significantly lower disfluency rate. Overall, the results of the BAF study seem to indicate a clear immediate advantage for late learners compared to early learners in tasks requiring explicit linguistic analysis (e.g. the cloze task), whilst late-learner advantages emerge later for listening comprehension. Age differences in oral skills seem also less marked, as some measures indicate an advantage for early learners and others an advantage for late learners.

The effects of age of onset on L2 proficiency and rate of learning in instructed learners of English were also investigated in another longitudinal project based in Spain (Basque Country) that compared very early starters (AO = 4 years), early starters (AO = 8 years) and late starters (AO = 11 years) after 600 hours of instruction. For example, Cenoz (2003) found significant advantages for an earlier start compared to a later start in most measures of oral proficiency (except pronunciation and fluency). The attainment of late starters was significantly better than the attainment of early starters in written composition in all measures, as well as in listening comprehension, cloze test and reading comprehension. Comparing a subset of the same age-of-onset groups after 6 years of instruction, García Lecumberri and Gallardo (2003) found that the advantages of late starters compared to both early starters and very early starters had extended to sound discrimination and degree of foreign accent (p. 126 -127).

More recently, Pfenninger and Singleton (2017) reported on the results of a seven-year research project in secondary state schools in Switzerland where the interaction of age effects (AO = 8 years and AO = 13 years) and instruction effects (CLIL instruction) was investigated. Four samples of L2 English learners (200 participants in total) were identified, comprising early/late starters without CLIL instruction (50 per group) and early/late starters with CLIL instruction (50 per group). Data were collected for the first time six months after the beginning of secondary EFL instruction (Time 1), and after additional 680 hours of instruction (i.e., five years later, Time 2).

Overall the study found that at Time 1 early starters significantly outperformed late starters in receptive vocabulary and lexical complexity. At that point late starters were already performing better than early starters in morphosyntactic accuracy, whilst no between-group differences were found in written and oral skills and in a grammaticality judgment task. At Time 2 late starters had bridged the gap with early starters. Mixed-effects models fitted on the accuracy data showed that, although age of onset was not a significant predictor of accuracy in any of the measures, type of instruction (and motivation) predicted accuracy in most measures, with the likelihood of accuracy being significantly higher for CLIL learners than non-CLIL learners. As for rate of learning between Time 1 and Time 2, late starters showed significantly faster learning in the areas of productive and receptive vocabulary, grammaticality judgment and accuracy, whilst for other measures no significant between-group differences emerged.

Overall the three longitudinal projects reviewed here (based respectively in Catalonia, the Basque Country and Switzerland) reported partially converging and partially diverging results. Although a direct comparison of the results is complicated by the fact that different tasks and sets of measures were used, all three projects found that late starters displayed a clear advantage over early starters in tasks requiring/probing linguistic analysis (e.g., cloze task) or morphosyntactic accuracy (however Pfenninger & Singleton found no age differences in a grammaticality judgment task). For listening comprehension, a study found that late-starters advantages were reduced in this area (emerged later) while others found they were significant. The results with regard to oral skills are also mixed with studies reporting early starters or late starters advantages (Cenoz, 2003; Mora, 2006), or no significant differences (Pfenninger & Singleton, 2017).

### 3.3.2 Age-effects in L2 morphosyntactic learning

Given the linguistic focus of the present investigation (word order, case marking, and learning of form-meaning relationships) it is useful to review some of the literature that found age effects in the L2 learning of specific morphosyntactic targets. In one of these studies García Mayo (2003) looked at how age of onset interacted with the accurate grammaticality judgment of three syntactic constructions.

The study compared early and late L1 Basque learners of English (AO = 8/9 and AO = 11/12) after 396 hours (Time 1) and 594 hours of exposure (Time 2) on their attainment in a GJT. The GJT probed three constructions whose ungrammaticality in English has been related to the lack of *pro*-drop[4] in this language (sentences with null subjects, V-S inversion and that-trace effects; Rizzi, 1982).

Overall the author reports an advantage for late starters. In particular late starters were better already at Time 1 on the correct identification of the ungrammaticality of null-subject sentences and verb-subject inversion, a difference that became statistically significant at Time 2. Longitudinally, there were significant gains for both groups in all three aspects of the *pro*-drop parameter under investigation.

Comparing the two focal groups of the BAF project and an adult group, Muñoz (2006b) set out to investigate the relationship between the order and rate of acquisition of a set of morphological functors (morphemes, articles, and irregular forms) studied in previous research on developmental sequences (Krashen, Sferlazza, Feldman, & Fathman, 1976). Overall the study found that the order of acquisition strongly correlated with those proposed by Krashen et al. and by Pica (1983) for instructed contexts and that there was no interaction between age and order of acquisition. Also in this case, however, age of onset did predict accuracy with an advantage for later starters, although for the two younger groups the proficiency gap found at Time 1 (200 hours) had closed by Time 3 (726 hours). By the time of the last test battery, child late starters (adults were not tested at Time 3) were more proficient

---

4 In generative linguistics a language is described as *pro*-drop if it allows null pronouns in subject position.

in all functors considered compared to child early starters, except in the correct use of regular past participle endings, where the latter were more accurate.

Looking at a similar selection of morphemes Pfenninger (2011), cited in Pfenninger and Singleton (2017), compared early starters and late starters recruited for her Switzerland-based seven-year longitudinal study. At Time 1 the two groups were tested on morphology production in two writing tasks and on a written GJT. The results showed that the performance of early and late starters was comparable, except for the production and judgment of irregular past forms where late starters had a marked advantage. Pfenninger also observed that, somewhat similarly to the results of Muñoz (2006b), early starters tended to over-regularize past tense forms and suggested distinct patterns of learning for the two age groups, with the early starters relying more on rule-based patterns and the late starters relying more on memory-based associative learning (Pfenninger and Singleton, 2017, p. 73).

Overall these studies indicate an advantage for late starters compared to early starters in the learning of morphosyntax. However, for early starters, two of the studies reported a better attainment on regular morphemes and a tendency to regularize morphological patterns (see 4.3.2 for further evidence of morpheme regularization in 6-year-old children).

### 3.4 Pili-Moss (2017)

Pili-Moss (2017) investigated age-related differences in attainment and rate of L2 learning in laboratory conditions and since it served as a preliminary study for the present thesis it is briefly discussed here. Six 8-9 year olds and eight young adults (all L1 English monolinguals) were trained in the miniature language BrocantoJ for six blocks on three consecutive days (one training block in session 1, two training blocks in session 2, and three training blocks in session 3, cf. 6.3.4). Although the vocabulary was entirely comprised of nonsense words, the structure of the language followed the word order and morphological pattern of Japanese main clauses (cf. Chapter 5). A computer game (Morgan-Short, 2007, Morgan-Short et al., 2010; 2012) provided a meaningful environment for language comprehension and, although participants were aware they were exposed to a new language, instruction on language rules (word order and relationships between NP positions/case markers and thematic roles) was incidental. In order to make points in the game, in each trial participants had to perform correctly the move that was described by an aural sentence stimulus, so that accuracy in the game was a measure of accuracy in language comprehension. Unlike the instructed longitudinal L2 studies reviewed previously, this study focused on the investigation of rate of learning in the very first hours of exposure.

Beside the overall measure of comprehension, two subsets of the game trials provided respectively a measure of the understanding of the linking between the position of syntactic arguments and their thematic interpretation and a measure of the linking between accusative markers and the interpretation of the related NP as patient. Both children and adults scored significantly above chance on all measures, although the proportion of adult correct responses was significantly higher overall, as well as with respect to the linking trials. Specifically, the significant adult advantage was found in session 1 and 2. By the third day, however, the children appeared to have bridged the accuracy gap both in terms of overall performance and in terms of accuracy in the processing of form-meaning relationships. In terms of rate of learning the study found that, overall, adult rate significantly increased earlier during practice compared to children (session 2 vs. session 3).

The overall results of the game task, essentially a listening-comprehension task, are broadly compatible with those of longitudinal studies that found that late learners attained significantly better than children in this linguistic skill (e.g., Cenoz, 2002). However, since children were able to bridge an initial gap with adults in a matter of days, the results also show that in the long run age differences in this linguistic area may level off. Also, significant increases in rate of learning happened in both adults and children, but at different points during exposure. For the set of stimuli measuring form-meaning relationships, it is interesting to note that whilst accuracy in adults increased incrementally, with no significant between-block differences, accuracy in children significantly increased in spurts between sessions. A possibility is that a between-session variable (e.g., sleep) had a significant role in the consolidation of learning in children compared to adults.

This said, it is important to note that a direct comparison between the present study and previous longitudinal research on age differences requires caution. First of all, it is entirely possible that the patterns of proficiency development recorded over an extended time span may differ from or obscure the fine-grained trajectory of attainment variation recorded over a few days of exposure. Also, in longitudinal rate-of-learning studies the tasks used as tests are designed to assess linguistic knowledge that has been trained and consolidated during some time prior to the test. In Pili-Moss (2017) participants had only had a very short exposure to the novel language prior to the game task, so that the game in itself offered a further opportunity for learning. Consequently, what was measured by the game task was to a large extent attainment *during learning* (i.e., language learning ability).

Further, whilst previous studies used natural languages, Pili-Moss (2017) trained participants in an artificial language, which, even if meaningful, is a miniature system with limited syntactic and semantic complexity. Finally, in a training study like the one described here, i.e. one with a very short instruction period, age of onset and age at testing practically coincide. This is not the case in longitudinal studies spanning many years where potential confounds between the two variables inevitably arise.

### 3.5 Two theoretical accounts of L2 morphosyntactic development

With few exceptions (see e.g., García Mayo, 2003) most SLA rate-of-learning studies have compared early and late L2 learners describing their attainment trajectories but have not attempted to frame their results in a more general theory of morphosyntactic representation and/or development. The aim of this section is to introduce two theoretical frameworks that have been developed in SLA to account for the acquisition of L2 morphosyntax. The first theoretical approach is Processability Theory (PT; Pienemann, 1998, 2005; Pienemann & Lenzing, 2015), whilst the second (Goldschneider and DeKeyser, 2005) explores the role of salience as a main driving factor. In view of the experimental studies reported in Chapter 6 and 7, these theories can provide a framework in which to interpret child and adult learning of word order and case marking.

#### 3.5.1 Processability theory

The observation that the acquisition of L2 English morphology in naturalistic conditions appeared to follow a similar sequence independently of type of L1 and learner's age (Bailey, Madden, & Krashen, 1974; Dulay & Burt, 1973; Fathman, 1975; Krashen et al., 1976, Krashen, Long, & Scarcella, 1979) sparked an interest in the study of acquisition sequences and a range of theoretical hypothesis to account for these findings. Specifically it was found that the general order of morphological development (natural sequence) observed for English follows the pattern: -*ing* / plural -*s* / copula >> auxiliary / article >> irregular past >> regular past / third person singular / possessive -*s* (Krashen, 1977).

Advancing a similar line of research, the theoretical framework of Processability Theory (PT; Pienemann, 1998, 2005; Pienemann & Lenzing, 2015), a model of L2 morphosyntactic production, extended the scope of the study of developmental sequences to include languages other than English and a wider range of syntactic phenomena. According to PT, developmental sequences depend on the output of language production procedures that are activated in a given order as the learner's L2 proficiency progresses. In addition to this, a mental grammar component maps word categories to phrase structure and provides a feature unification mechanism (Bresnan, 1982) that "ensures that the different parts that constitute a sentence do actually fit together" (Pienemann, 2005, p. 15).

The sequential availability of procedures has been shown to operate for both adult and child L2 learners (Pienemann, Johnston & Brindley, 1988; Pienemann & Mackey, 1993), to apply cross-linguistically (Al Shatter, 2008; Baten, 2011; Di Biase & Kawaguchi, 2002) and potentially also accounts for L1 developmental acquisition sequences (Pienemann, 2005, p. 40). Although PT is essentially a theory of language production a growing number of studies have tested its predictions for the acquisition of L2 receptive grammar, although with mixed results, in both adults (Spinner, 2013) and children (Buyl & Housen, 2015; Keatinge & Keßler, 2009).

According to the order illustrated in Pienemann (2005, p. 24) the procedure sequence (a) would develop incrementally according to the following levels:

(a)

(1) word/lemma >> (2) category procedure >> (3) phrasal procedure >> (4) VP procedure >> (5) S-procedure >> (6) subord. clause procedure

(b) PRED "eat" (SUBJ, OBJ) (Adapted from Pienemann, 2005, p. 17)

To provide an example of how the model would work let us consider the verb 'eat'. This is a lexical entry, and as such it is stored in the mental lexicon together with a diacritic specification of its linguistic features (e.g., number) and a functional structure specifying the associated thematic roles (b, above). The level of category procedure provides a representation of the verbal phrase and an initial linearization rule producing a basic canonical word order associating thematic roles to syntactic positions (cf. Pienemann, 2005, Table 5, p. 24). A similar category procedure would apply to other word categories as well, for example to nouns, represented as linearized nominal phrases. In both cases no phrasal procedure needs to be initiated at this level, because no feature unification is involved. Importantly for L2 learning, higher procedures become available only when the learner's interlanguage has achieved a sufficient stage of L2 development. If a level of the implicational hierarchy is not yet available "the hierarchy will be cut off at the point of the missing processing device and the rest of the hierarchy will be replaced by a direct mapping of conceptual structure onto surface form" (Pienemann, 2005, p. 13).

Let us now tentatively consider the predictions that Pienemann's model would make for three linguistic domains relevant to the present thesis: linking between nominal elements and thematic function, word order, and case marking. Initially, a correct description of the vocabulary item, including its thematic structure, would be required at word level. Subsequently, both the linking between nominal elements and their thematic interpretation and the production of a basic canonical order would occur at the level of category procedure. Case marking, however, would require a higher level of processing, i.e. a phrasal procedure. Specifically it would require an operation in the nominal phrase to combine/check case features of the noun and of the case particle (feature unification). Hence we can conclude that in the PT framework, case marking, requiring a process of feature unification at phrase structure level, would be more complex than the derivation of a canonical word order for elements inside the verbal phrase.

### 3.5.2 The role of salience

A different approach to determining the factors involved in learning developmental sequences was taken in Goldschneider and DeKeyser (2005). In their meta-analysis of 12 studies focusing on morphological learning in L2 English, they suggested that the combination of five factors connected to a more general construct of 'salience' explained a large portion of the variance in context-correct production of functors. In SLA research on the role of salience word order has often been considered a more salient linguistic target compared to morphology (Long, 2007). However, as the focus of Goldschneider and DeKeyser's paper remains on morphology, this does not allow for predictions to determine/rank the relative salience of syntactic and morphological phenomena in this framework. Also, although the research considered in the meta-analysis indicated a high correlation between order of functor learning in children and adults, the authors did not discuss the extent to which the salience criteria may interact with age of onset or age at testing, or indeed if such interactions are to be expected.

Specifically, the salience criteria discussed in the study included: (a) perceptual salience (depending on number of phones, syllabicity and sonority); (b) semantic complexity (whether more than one meaning was associated to a single form); (c) morphological regularity (depending on whether or not the form changed in different phonological environments); (d) syntactic category (whether the element was functional/lexical and bound/unbound), and (e) frequency in the input.

In the two experimental studies in Chapter 6 and 7 learning of the nominative and accusative markers *ri* and *ru* will be analyzed (see Chapter 5 for a description of the miniature language deployed in the studies). In terms of perceptual salience both particles have two phones, are syllabic and are located in the higher half of the sonority scale adopted by Goldschneider and DeKeyser (in both cases liquid consonant + high vowel, p. 50). Further, their form is semantically unambiguous, they are phonologically independent functional elements and they are morphophonologically regular. In terms of frequency in the input, they are more frequent than any other vocabulary item in the exposure sets but *ru* is about 20% more frequent compared to *ri* (see 6.3.6 and 6.3.7 for the description of the exposure sets). Relative frequency is hence the only salience criterion distinguishing the two markers.

This discussion of two linguistic analyses through which morphosyntactic gains can be evaluated concludes Chapter 3, which has presented issues relative to the debate around age differences in SLA. In doing so, the chapter has provided a SLA theoretical perspective for the comparison of the results of the two experimental studies included in the thesis (Chapter 6 and 7). Chapter 4 consists in a detailed review of child and adult artificial language learning studies paving the way for the introduction of the methodology and the focus of the experimental investigation.

# 4. Learning L2 morphosyntax in a miniature language system

## 4.1 Artificial grammars vs. miniature languages

Since the experimental studies in Chapter 6 and 7 deploy an artificial language paradigm, the aim of the present chapter is to present a literature review of child and adult studies that have deployed miniature languages to investigate learning of morphosyntax in SLA and in cognitive psychology.

Artificial languages are miniature linguistic systems that have been used extensively in experimental research in the last fifty years in a number of disciplines that share an interest in investigating human language learning and processing (Boyd & Goldberg, 2012; Boyd, Gottschalk, & Goldberg, 2009; Braine et al., 1990; Brooks, Kempe, & Sionov, 2006; DeKeyser, 1995; Dienes & Berry, 1997; Ferman & Karni, 2014; Francis, Schmidt, Carr, & Clegg, 2009; Friederici, Steinhauer, & Pfeifer, 2002; Hudson Kam & Newport, 2005; 2009; Kapa & Colombo, 2014; Lichtman, 2012; MacWhinney, 1983; Morgan-Short, 2007; Morgan-Short et al., 2014; Reber, 1967; 1976; Reber et al., 1991; Rebuschat & Williams, 2012; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Williams & Kuribara, 2008; Wonnacott, Boyd, Thomson and Goldberg, 2012; Wonnacott, Newport, & Tanenhaus, 2008). These miniature systems differ with respect to the extent to which they resemble the characteristics of natural languages. In early *artificial grammar* studies finite-state grammars were deployed to generate strings, typically letter sequences (see for example Reber et al., 1991) and investigated the extent to which such strings could be successfully learnt with/without the participants' conscious awareness. Although these sequences can be generated to incorporate regularities that are thought to be typical of natural language syntax (e.g., recursion), they lack semantic reference and for this reason their use in studies that tried to understand how natural languages are learnt/processed has been questioned in the literature (Robinson, 2005; Van Patten, 1994).

On the other hand *artificial languages* (here also referred to as miniature languages) are linguistic systems displaying natural language syntax but whose lexis is constituted by nonwords, i.e. words that are especially devised to be meaningless to the learners prior to training and to follow specific phonotactic rules (for example the rules of the learner's L1). Unlike artificial grammars, artificial languages can be associated to semantics, both at lexical level (meanings associated to lexis during vocabulary training) or at sentence level, when words with different functions are inflected or presented in syntactic constituents and full sentences. In some cases the miniature linguistic system can deploy the lexis of a natural language (typically the learner's L1) but incorporates elements of the morphosyntax of a second natural language (for example word order and/or inflectional morphology).

Research with miniature languages offers a series of advantages compared to natural languages, the first of which is to allow participants to reach high proficiency in the rules that are the object of investigation in a relatively short amount of time. The second main advantage is that they offer a more precise control over a series of experimental variables including for example (a) the phonological/phonotactic differences between the new language and the learner's L1; (b) the possibility of presenting rules that differ or resemble the rules of the L1 (depending on the research design); (c) the age of exposure to the L2 (the stimuli are novel for all participants); (d) the exposure conditions in which the language is learnt.

In this chapter I will review studies that employed miniature languages of different types to investigate adult and child acquisition of morphosyntax and the relationship between a linguistic construction and its semantics under implicit/incidental learning conditions. In what follows I will review child studies and studies that compared children and adults first, and then move on to studies targeting adult populations. Finally, I will summarize the contributions of the two bodies of research to the current understanding of age differences in the initial phases of language learning in the laboratory and the areas that remain open to further investigation.

## 4.2 Child miniature language learning

Child artificial language learning has been studied mostly in developmental and cognitive psychology. Recently some of this research has focused on infants, investigating their ability to identify words and categories tracking the statistical regularities of the auditory input (e.g., Gomez, 2002; Saffran et al., 1996). However, learning studies targeting an older child population are also important because, due to children's higher cognitive development, their design can allow a comparison with an adult group with minimal or no methodological differences. The studies briefly reviewed here include (a) studies that were mainly concerned with child-adult differences in the learning of formal morphosyntactic features, (b) studies that investigated learning of the links between the syntax and the semantics of argument structure, and (c) studies that looked at the role of cognitive individual differences in learning.

### 4.2.1 Child attainment in morphology

MacWhinney (1983) and Braine et al. (1990) investigated inflectional morphology in nominals and were among the first studies to deploy miniature artificial languages with child participants providing evidence for its methodological feasibility and the possibility of age comparisons with adults. In MacWhinney (1983) 16 children between 5 and 7 and 16 adults learned an artificial language consisting of eight nouns and four affixes with a locative meaning (on, in, behind, in front), and were tested on their proficiency in the oral production of accurate noun-affix structures. In order to facilitate the learning of the affixes in the training phase, the children played an interactive comprehension/production game with the researcher. The game consisted in placing toys in different locations according to an aural prompt sentence and in producing sentences to communicate the placement of toys in a new location.

The results of the production tests showed different learning outcomes for the child and the adult group; in particular only children significantly produced more correct suffixes than correct prefixes, significantly overgeneralized the production of suffixes to irregular forms, and were significantly more conservative in producing noun-affix pairings that were presented early in training compared to pairings they learned later on. These results, however, should be interpreted in the light of the study design. Specifically, the author reports that in the presentation and training phases the children were exposed to variable amounts of input (1 to 4 hours) and that, on average, the input adults were exposed to amounted only to half an hour. Secondly, although the study asked whether there was a difference between learning of prefixes and suffixes, the exposure set was not counterbalanced for type of affix, with suffixes being four times more frequent than prefixes. These observations suggest that, instead of depending on the children's preference for one or the other form of affixation, a stronger learning effect for suffixes in children could be related to item frequency and, specifically, to a more marked tendency to adhere to frequent patterns in the input compared to adults.

In another study Braine et al. (1990) also investigated the learning of inflected nominals in 7 to 10 year olds and in adults. Compared to MacWhinney's study, Braine et al. (1990) had a more complex design and focused more on controlling the effects of item frequency and phonological similarity. The artificial language showed a high level of inflectional complexity with 6 noun endings for case in addition to vowel harmony. Vocabulary learning, exposure and testing took place during 3 sessions of between 40 and 60 minutes on 3 consecutive days. The exposure set in this study consisted of a world stimulus of 72 events shown on flashcards and involving Frippy (a monkey) and 24 objects corresponding to nouns in the artificial language. In the presentation the researcher described the event using a sentence in the language, asked the child to repeat it and provided feedback. Nouns were grouped in two classes depending on frequency in the input (high and low), and in three classes with three different case endings encoding three different types of location (movement away from an object, movement towards an object or being close to an object). The testing phase was like the presentation, but this time feedback was not provided.

For the children, the authors found robust learning effects on the matching between affixes and the locative semantics. Compared to children, adults tended to show significantly more robust learning of the correct matching between affixes and locative semantics for low frequency items, better generalization of the correct pattern to novel items irrespective of frequency, and a significantly better performance on items that underwent root vowel harmony when inflected. Similarly to MacWhinney's study, Braine et al. (1990) seemed to point at a greater role of input frequency in supporting learning in children compared to adults.

More recently Ferman and Karni (2010) investigated learning of an artificial morphological rule based on affixation in noun-verb pairs by eight 8-year olds, eight 12-year olds and eight young adults. The morphological rule consisted in changing the ending on the verb depending on whether the preceding noun was animate or inanimate, and training included 10 consecutive sessions (1-3 days apart). In each session, after modeling of the construction, participants engaged in an aural judgment task and in a production task (they had to orally produce a verb after they heard a noun prompt), with both tasks including repeated items from the modeling set as well as new items.

For old items the study found that adult learning gains were superior to both children groups, with 12 year olds performing better than 8 year olds in terms of accuracy (though not speed). As for the generalization to new items, the study found that the performance of adults and 12 year olds was comparable, but 8 year olds did not reach significantly above chance performance. Interestingly the authors report that whilst most adults and older children explicitly reported the semantic distinction on which the rule was based at some point during training, none of 8 year olds did, suggesting that explicit knowledge of the rule may have been pivotal in supporting accuracy in rule learning and in rule generalization (see also Ferman & Karni, 2014).

In sum the three studies provide evidence that children from the age of five can learn to comprehend and produce novel affixes with a semantic content, with an overall advantage for older children and adults. Compared to adults, children appear to be more sensitive to frequency, overgeneralize affixation rules to irregular cases, but are less able to apply rules to novel items and report morphosyntactic rules based on semantics.

### 4.2.2 Form-meaning linking

Research in developmental psychology has shown that children are able to establish a connection between specific word orders and the semantics of the verb from a very early age (Fisher, 1996, 2002; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005, Gertner, Fisher, & Eisengart, 2006). In this section I will review three studies that have specifically investigated the effects of the input structure on language learning and on the ability to generalize syntactic constructions, including the linking rules relating syntactic forms to their semantic interpretation: Casenhiser and Goldberg, 2005; Boyd and Goldberg, 2012; and Wonnacott et al., 2012.

In Casenhiser and Goldberg (2005, Experiment 1) 51 English speaking six year olds were shown 8 short videos, and the corresponding voiceover sentences, displaying 5 made-up verbs and a novel word order consisting of NP1-NP2-V+ present/past tense marker. In order to minimize the influence of the L1 semantics, the novel word order was associated with the presentation of a novel meaning where NP1 appeared in the location NP2 in ways specified by the verb, e.g., sailing, dropping down, rising, rolling, etc. The participants were assigned to three conditions depending on the structure of the input in the exposure set (8 sentences): (a) skewed input, with one verb occurring in 4 of the sentences and the remaining four verbs occurring in one sentence each (4-1-1-1-1); (b) balanced input (2-2-2-1-1), and control, where participants watched the film with the audio turned off.

The testing set consisted of six test items and six distractors and deployed novel verbs that were not used in the exposure. It consisted of a forced-choice task where the children had to match a voiceover target sentence with the video displaying the correct ‘appearance’ semantics. The study found that children in both experimental conditions performed significantly better than controls and that the skewed-input condition significantly outperformed the balanced-input condition.

Boyd and Goldberg (2012) extended the methodology of the previous study to include testing of the linking rules between NP arguments and their syntactic position. In this study a 5-year-old, a 7-year-old and an adult group were compared on the learning of a construction with a syntax of the type NP1-NP2-V, similar to the one adopted by Casenhiser and Goldberg (2005), but matched to a different novel semantics. Here NP1, the agent, approached NP2, the goal, and V encoded the manner of approach. Participants were exposed to the language through 16 videos and corresponding voiceovers. They subsequently performed a forced-choice task where they had to pick the video corresponding to the voiceover out of two presented simultaneously. The task included all-familiar, partially novel and all-novel items and probed the participants' ability to (a) distinguish the correct semantics of the construction from an intransitive construction and to (b) detect the correct linking between semantic roles and NPs across trials.

As for the first measure, significant effects for group were found; older children outperformed smaller children and adults outperformed the child group as a whole. In particular both seven year olds and adults were significantly better than five year olds on all-novel items showing that the latter were less able to generalize the construction to completely novel sets. In the linking trials no significant differences were found between adults and older children, whilst both groups significantly outperformed smaller children, who were at chance. The results of this comparative study provides further support for the idea that input frequency tends to shape children's language learning more than adults' and that children, particularly smaller ones, tend to be less flexible in applying generalizations to novel contexts.

Wonnacott, Boyd, Thomson and Goldberg (2012) represents a further development of the design in previous studies. In this study 42 L1 English five year olds were exposed to an artificial language with a novel syntax (V-ing NP1 NP2) associated to a novel meaning. The language included 14 novel pseudoverbs and 6 English nouns corresponding to the names of animal soft toys. The fourteen verbs all denoted a different manner in which a toy (NP2, the agent) could approach another one (NP1, the goal), e.g. hopping on its head, sliding on its stomach, spinning around, etc. An important difference with previous studies was that in the sentence structure adopted here the agent was not in the initial position, which avoided a bias to associate this thematic role to the first NP appearing in the sentence.

In each trial the researcher pronounced a sentence describing different approach events and after each one enacted the event using two animal toys (exposure set, 16 items). After each presentation the child repeated the sentence aloud, was given feedback and the opportunity to repeat the sentence a second time.

The study manipulated input structure (use of a single verb vs. four different verbs in the exposure input, defining two experimental groups), familiarity (use of verbs already presented in the exposure phase vs. new verbs in the testing set) and day (the children were tested on day two and on day four). Proficiency on three different tasks was used as a measure of learning: (a) an act out activity were children enacted an event with the help of the toys after hearing the corresponding sentence in the language, (b) an oral production task where they had to say a sentence describing the event enacted by the researcher, and (c) a forced choice task where they were asked to pick one of two movies playing simultaneously on a computer after hearing a sentence in the language.

The results showed that the children were able to learn the construction after minimal input, with significant performance effects emerging as a function of exposure both with respect to the global learning of the construction and with respect to the correct linking of the nominal arguments to the correct semantic roles of agent or goal. In particular the learning effect was strong enough to overcome a bias towards the assignment of the role of agent to the first nominal in this construction, which was detected in a control group and found in previous studies with English native speakers (Boyd, Gottschalk, & Goldberg, 2009; Wonnacott, Newport, & Tanenhaus, 2008).

Another important finding of this study was that not only the amount of input but also its structure was a crucial factor affecting learning. Children in the four-verb condition were significantly better than children in the one-verb condition at transferring the construction pattern to novel verbs. This result confirmed previous findings suggesting that evidence of item variation for a given slot in the input sentence supports the identification of abstract categories and pattern learning (see also Gomez and Gerken, 1999; Childers and Tomasello, 2001; Suttle and Goldberg, 2011).

Overall the studies that investigated the learning of linking between syntax and thematic roles found that children older than 5 could learn the syntactic constructions and the form-meaning relationships associated with them after a relatively short exposure. They also found that, although adults and older children were better at generalizing patterns to novel items, a skewed distribution in the input had a an overall significant effect on both learning and generalization.

A final study reviewed in this section, Hudson Kam and Newport (2005), did not investigate linking rules but looked at the conditions under which children's learning adheres to probabilistic rules in the input, and as such it is relevant to the evaluation of which type of rules children are more likely to learn and whether differences have been found with adults. Specifically the authors were interested in replicating in controlled conditions the situation naturalistic learners face when they are exposed to a language deploying optional rules (a characteristic common to pidgin languages).

They hypothesized that children would show a stronger tendency to regularize inconsistent input as a function of input consistency compared to adults, and tested 19 six-year olds and 8 adults on the learning of the distribution of determiners in

Sillyspeak, a miniature artificial language including 17 words (experiment 2). The sentence word order was V-S-O, with an NP structure including a determiner and a noun. For each age group there were two experimental conditions, one in which the input was fully consistent, i.e. where the determiners occurred with specific nouns 100% of the times, and one in which the input was inconsistent, i.e. where the determiners occurred with specific nouns only in 60% of cases (probabilistic distribution). The outcome measures were a sentence completion task and two GJTs, one general and one specifically testing determiners.

The results of the general GJT showed that both groups with no significant differences learnt the overall structure of the syntactic string. However, an analysis of the production data in the sentence completion task showed that the researchers' hypothesis was borne out and that, particularly in the inconsistent condition, children were more likely than adults to systematize their production patterns instead of mirroring the probabilistic distribution of the input.

### 4.2.3 The role of individual differences in child studies

Kapa and Colombo (2014, Study 2) is one of the few child studies considering individual differences in cognitive ability in the learning of a miniature language. After vocabulary training (12 items) 42 5-year-old L1 English children were exposed over two one-and-a-half hour sessions to a modified version of Sillyspeak using short videos (300 videos per session, 60 minutes in total). The individual difference measures included working memory (digit span), attention (Attention Network Test), executive function (visual Simon task), and cognitive flexibility (Card Sorting). The outcome measures of learning were a vocabulary reception and production test, an aural GJT, as well as sentence reception and production tasks (the children watched short videos and chose which one corresponded to an aural sentence stimulus or they were asked to narrate a video sequence using the language).

The study found that although children performed above chance in vocabulary reception and the GJT, they were below chance performance in the receptive sentence task and were better at producing nouns than verbs in the production task. A regression analysis also found that working memory, attention and cognitive flexibility were all significant predictors of language learning gains. However, given the results in the sentence tasks, the extent to which the findings were driven by vocabulary learning (as opposed to sentence learning) remains unclear (see also Stone & Pili-Moss, 2016).

The below-chance performance in the sentence comprehension forced-choice task suggests that, possibly due to the limited amount of exposure received, the children either did not engage in the processing of meaning at this level or their semantic representations were inaccurate or only partially accurate. By contrast, the children performed above chance in an aural GJT, a task where the processing of meaning was not required, suggesting that exposure to a limited amount of language input had been sufficient to develop a sensitivity to the structural properties of the sentence.

## 4.3 Adult miniature language learning

Laboratory studies that have investigated how adults learn L2 morphosyntax using miniature languages have not only analyzed the linguistic gains attained as a result of training, but have also investigated how a number of learners' individual differences (proficiency, aptitude, age, cognitive differences, etc.) and external factors (instruction, input, etc.) modulate learning. In addition to these factors a number of studies have also looked at the type of language knowledge (implicit or explicit) that is gained as a result of L2 training.

### 4.3.1 The role of external factors in adult studies

***Incidental Instruction Conditions***. A number of studies in cognitive psychology and SLA have investigated artificial language learning in incidental instruction conditions (e.g., Altmann, Dienes, & Goode, 1995; Cleary & Langley, 2007; Francis, Schmidt, Carr, & Clegg, 2009; Friederici, Steinhauer, & Pfeifer, 2002; Reber, 1989; Williams, 2010; Williams & Kuribara, 2008) or have compared incidental and explicit instruction conditions (e.g., de Graaf, 1997; DeKeyser, 1994; 1995; Lichtman, 2012; Morgan-Short, 2007; Morgan-Short, Finger, Grey, Ullman, 2012; Morgan-Short, Steinhauer, Sanz, & Ullman, 2010, Robinson, 1996; 2005). Further studies that additionally considered the type of language knowledge obtained as a result of exposure are discussed in subsequent subsections.

For example, Williams and Kuribara (2008) used Japlish, a miniature language with English lexis and Japanese morphosyntax (word order and case) in a single-session design. They tested 41 L1 English university students on the acquisition of syntactic scrambling, after the experimental group (25 participants) were trained in the language in incidental conditions by means of a plausibility judgment task with 194 bimodal (visual and aural) sentence stimuli. In a receptive GJT administered using the same modality they found that the exposure group showed significantly higher accuracy in judging the grammaticality of scrambling compared to the control group both on trained and novel items.

Francis et al. (2009, Experiment 2) also investigated incidental learning of word orders noncanonical for English, but in addition considered how learning was modulated by item frequency. Twenty-nine L1 English speakers were exposed to a mix of English and pseudoword strings of the form V-N-N or N-N-V under implicit conditions (visual input), with one string type more frequent than the other. The study used self-paced reading times as a measure of learning, assuming that a progressive reduction of reading times in noncanonical sentences would provide evidence of learning of the relevant order and that frequency would be a modulating factor. As predicted, in both experiments there was a significant decrease in reading times as the exposure progressed and the statistical analysis yielded an effect for rule with larger gains for the more-practiced noncanonical order over the less-practiced one.

Friederici et al. (2002) was one of the first studies to investigate incidental miniature language learning using electrophysiological measures. It deployed BROCANTO, a miniature language with German phonotactics that was presented in the context of a computer board game where the language syntax was acquired through auditory exposure and gaming practice. The language comprised 14 pseudowords (four nouns, four verbs, two adjectives, two determiners and two adverbs) and had an SVO structure, with determiners and adjectives preceding the nouns and adverbs immediately following the verb.

Twenty-eight L1 German participants learnt the vocabulary of BROCANTO prior to comprehension and production activities, but were not explicitly taught the syntax of the language. A control group of 31 participants only received training in vocabulary with no gaming exposure. In an incidental learning condition pairs of participants were exposed to the language in a meaningful context while playing a computer game against each other. One participant would utter a BROCANTO sentence corresponding to a move (the study report is not very clear about this point but it seems plausible that the player described a move automatically generated by the computer on his/her screen and not a move he/she decided to make, p. 530). Then the other player was instructed to update his/her game constellation according to the first player's utterance. In case of incorrect utterances or misunderstandings the computer program provided the correct utterance auditorily, automatically corrected the move onscreen and provided vocabulary feedback (p. 530). Again, the study report is not clear about which event exactly triggered the feedback provision, whether the incorrect aural instruction given by the first player or the incorrect move performed by the second player. The practice was extensive and lasted for up to 5 hours per session across several sessions until the participants reached 95% accuracy, a level required in order to maximize the possibility to record ERP responses to ungrammatical sentences in subsequent testing.

In the testing phase both groups performed an aurally administered GJT probing the acquisition of BROCANTO word order, during which ERP measures were taken. The GJT consisted of 244 grammatical sentences and 244 matched ungrammatical sentences (each ungrammatical sentence was derived from the corresponding grammatical one by replacing one element with a word from a different lexical category). The results of the behavioral measures revealed a significant effect of practice, with controls performing only slightly above chance (since the difference from chance was not significant, it cannot be excluded that controls' accurate performance was due to accurate guessing). Furthermore, only the ERP results for the trained group (and not the ones relative to controls) revealed the presence of a biphasic pattern of the same type as the one reported in the L1 literature relative to word order violations (anterior negativity followed by centroparietal positivity[1]). According to the authors, the

---

1 The patterns of activation that are most relevant to L1/L2 processing can be summarized as follows (Steinhauer, 2014): (a) N100: A negativity 100ms after the event indicates initial processing of visual or auditory stimuli in the sensory cortex; (b) P200: A positivity at about 200ms after the event is related to pattern recognition; (c) N400: A

finding of this ERP pattern in the experimental group indicates that "a late-learned language, in principle, can be processed in a native speaker-like way" (p. 534).

***Implicit/Incidental vs. Explicit Instruction Conditions***. The studies that have compared the relative benefits of different types of instruction for L2 learning in general have indicated a greater effectiveness of explicit instruction over implicit instruction, an advantage that has also emerged in laboratory and classroom studies investigating natural language learning (for a recent meta-analysis see for example Spada & Tomita, 2010).

DeKeyser (1994, 1995) conducted one of the first studies that used an artificial language to test how implicit and explicit instruction modulate the acquisition of second language morphosyntactic categorical rules and prototypical patterns. To this end he created Implexan, an SVO miniature language with 98 words and a rich morphology, displaying number and gender marking on the verb and number and case marking on the nouns. The morphemes' distribution followed categorical rules but a subset of the morphemes additionally displayed prototypical patterns (allomorphy). Sixty-one L1 English university students participated in the experiment over a series of 20 sessions (25 minutes each) distributed over three weeks. They were exposed to the artificial

---

negativity at about 400ms after the stimulus presentation, that is recorded in central-parietal areas, has been reliably associated to difficulties in lexical and semantic processing; (d) P600: A positivity recorded in the parietal area after between 600 and 900ms is associated with controlled processing and reliably related to the detection of a number of morphosyntactic violations (word order, morphology, local binding, etc.); (e) AN (anterior negativity): Early left-anterior negativities (LANs) emerging after 200-300 milliseconds in left-frontal areas have been identified as signatures of native-like processing of ungrammaticality and automatic processing of sequences.

language using a computer program under two learning conditions; implicit-inductive (no explanation of grammar rules provided) and explicit-deductive (explicit presentation of grammar rules incorporated in the training), and tested with a grammaticality judgment task and a production task administered at the end of training.

No difference in proficiency was found in the GJT where both groups performed at chance. However, the written production test showed that participants in the explicit instruction condition scored significantly better than participants in the implicit instruction condition for both simple categorical rules and categorical rules occurring with allomorphy. Further analysis showed that the significant positive effect of explicit instruction emerged when the forms produced in the test had not been previously encountered in the training, whilst there was no effect of type of instruction on old forms.

The fact that learners trained in the explicit condition did not perform above chance in the GJT, even following a relatively large number of sessions, is not expected considering the findings of most subsequent miniature language studies with a similar design. The fact that participants trained in the explicit condition did not outperform participants trained in the implicit condition with regards to the old forms is also interesting, as it shows that rule learning in the explicit instruction condition did not seem to assist the retention of input stimuli (learning of old items), a result that could have been expected if the explicit learning of a rule supported item learning through input analysis.

In a similar study, de Graaf (1997) tested 56 L1 Dutch undergraduate students under implicit and explicit instruction conditions exposing them to a variation of Esperanto (eXperanto), where the lexis (in this case only about 40 items) was modified to resemble more closely Dutch phonotactics, and both word order and morphology were modeled after Spanish. Language instruction was delivered via a computer based self-study course over 10 sessions (1.5 hours per session), with three test sessions (consisting of speeded ad unspeeded GJTs); one half-way through the training and two posttests.

Instructional conditions were similar to the ones adopted in DeKeyser (1994, 1995). Results showed that, although the explicit group overall outperformed the implicit group at the immediate post-test, there was a clear above-chance learning effect for the implicit group on both GJTs for complex morphological structures (affirmative and negative imperatives in formal and informal contexts) and both simple and complex word order patterns (position of the negation and positions of object clitics respectively).

This study not only confirmed stronger learning effects in the explicit learning condition, but also indicated that a less broad vocabulary may have aided the emergence of more substantial learning effects in both conditions. As for the study design, the testing session half-way through the training had the goal to gain some insight in the learning process. However, this choice presents at least two methodological disadvantages. Not only participants were likely to expect further testing, but they were also made aware of its format, potentially creating a bias towards the use of explicit learning strategies in the second part of training in both conditions.

Another series of studies comparing type of instruction used BROCANTO2, a modification of BROCANTO (Friederici et al., 2002) displaying English-like phonotactics, an SOV order and gender nominal morphology. The full structure of BROCANTO2 is NP-NP-Adv-V at sentence level, and N-Adj-Det at NP level. Morgan-Short (2007) trained 42 L1 English adults in the language BROCANTO2 in the context of a computer game similar to chess over 3 sessions, a maximum of five days apart (for a description of vocabulary training and the gaming environment see 6.3.5 and 5.6).

Participants were assigned to one of two instruction conditions, implicit or explicit. At the beginning of each session participants in the implicit condition listened to a set of sentence exemplars while watching the corresponding game constellations on screen (127 items, 13 minutes). The length of training was the same in the explicit condition but in this case participants listened to 33 exemplars and received metalinguistic information.

After exposure, participants practiced the game with a total of 44 alternating comprehension and production blocks (20 items per block) distributed across the sessions, whereby they listened to a sentence in BROCANTO2 and had to perform the corresponding move (comprehension; cf. 6.3.7), or they had to orally produce a BROCANTO2 sentence to describe a game move they had just watched on screen (production). The behavioral measures of language learning included two aural GJTs probing word order and gender agreement, one administered when participants had reached low proficiency (40% correct trials in two consecutive comprehension blocks), and the other at the end of practice, when participants were also administered a speeded aural GJT, a written GJT, and a free production task.

Accuracy in the GJT at low levels of proficiency showed that participants in the explicit conditions outperformed participants in the implicit condition, but only in the learning of gender agreement structures (noun/article and noun/adjective agreement). At the end of training no significant differences were found between conditions in any of the measures. During the aural GJTs administered at low and high proficiency ERPs were also recorded. They revealed that at high levels of proficiency in the implicit condition the patterns of activation relative to syntactic violations were compatible with the ones observed in L1 processing.

In a behavioral follow-up study Morgan-Short, Sanz, Steinhauer, and Ullman

(2010) investigated specifically the acquisition of gender agreement and trained 30 L1 English adults in BROCANTO2 using a similar methodology. The study found that although participants in both the implicit and the explicit condition showed significant gains between a GJT test administered at low levels of proficiency and a GJT administered at the end of training, only the implicit group improved significantly on noun-adjective agreement (p. 171).

In the same strand of studies, Morgan-Short, Finger et al. (2012) investigated the retention of BROCANTO2 after months of no exposure in learners instructed under incidental or explicit conditions. The period of no exposure ranged from 3 to 6 months (on average around 5 months), and levels of proficiency in the language prior to and after the no-exposure period were comparable across conditions. Analyzing behavioral and ERP data the authors found that not only did the incidentally trained group show more native-like patterns of activation both immediately after training and following the no exposure period, but that the native-like patterns emerged or were re-enforced for all participants independently of the initial training conditions.

Lichtman (2012, Study 2) trained 40 5-7 y.o. children and 40 young adults in the miniature language Sillyspeak and looked at the interaction between age and type of instruction. Half the participants were trained in implicit or explicit instruction conditions in both age groups and were tested on two production tasks and a GJT. The study found no significant effects for age or type of instruction, although adults tended to be more accurate. However, an analysis of the verbal reports indicated that age was relevant in the relationship between type of instruction and language awareness; whilst only explicitly instructed children tended to develop awareness of the rules, adults showed awareness of the rule independently of type of instruction.

In summary, studies that deployed incidental learning conditions found that learning of morphosyntax in these conditions (typically word order) is possible after a relatively limited exposure. When explicit and incidental instruction conditions were compared, learning effects were generally larger in explicit instruction conditions, although some studies found no effect for type of instruction and at least one study (Morgan-Short et al., 2010) found that there was an advantage for morphosyntactic learning in incidental learning conditions. Incidental learning conditions also comparatively benefitted learning when this was measured after a period of no-exposure.

***Input Structure***. The structure of the input exposure set is another external factor that has been found to play a role in the way miniature languages are learnt (see, among others, Boyd, Gottschalk, & Goldberg, 2009; Casenhiser & Goldberg, 2005; Goldberg, Casenhiser, & Sethuraman, 2004; Goldberg, Casenhiser, & White, 2007; Wonnacott, Newport, & Tanenhaus, 2008).

For example Wonnacott et al. (2008) used a miniature artificial language comprising 5 nouns, 12 verbs and one particle and manipulated verb construction and item frequency to investigate learning and generalization of nonnative word orders by a total of 58 L1 English adult participants. The two target structures were of the kind pseudoverb-NP(agent)-NP(patient) and pseudoverb-NP(patient)-NP(agent)-particle. In the first experiment (14 participants) three verbs classes including an equal number of items were presented (4 per class), two of which occurring exclusively with one or the other order and a third one alternating between the two orders, a pattern similar to the distribution of verbs found in the English double-object construction.

The exposure to the language consisted in showing the participants a series of video clips (144 scenes), where animal soft toys performed actions described by a voiceover sentence stimulus. Three tasks were used to assess learning and generalization of the two constructions; a production task, whereby participants were shown a video, were provided the initial verb and prompted to complete a sentence; a forced-choice task to assess comprehension; and a GJT.

The study found significant learning effects on both constructions with the ability to extend nonnative word orders to novel sentences as a function of the statistical structure of the input, with frequency-related entrenchment effects similar to those reported in L1 acquisition studies. In experiment 2 (14 participants) the exposure set was modified to include a larger number of alternating verbs (8 items, compared to 2 items each in the two nonalternating classes). Compared to the first experiment the results showed a stronger tendency to overgeneralize the alternating construction to one-construction verbs with no significant effect of frequency for the latter. In the third experiment (30 participants) the relationship between lexically-based learning and tendency to generalize was explored comparing a language where verbs did not show any alternation (lexicalist language) and a language where all verbs occurred in both constructions (generalist language). The results of the experiment were compatible with a Bayesian account of how learners track input (Perfors et al. 2010) and confirmed the presence of two simultaneous criteria learners took into account when producing new sentences. The first was tracking the occurrences of a given verb in a given construction at the lexical level, and the second consisted in increasing the probability of using a given construction if this occurred extensively in the input.

Boyd et al. (2009) deployed a modified natural language of the form NP1(theme)-NP2(location)-V with pseudoverbs and English NPs, conveying a semantics whereby the theme appeared in a given location in a way specified by the verb. They assigned 16 participants to an experimental condition (16 controls) exposing them to a series of 16 videos for a total of about 3 minutes. The intent was not only to investigate adult learning and generalization of the construction to novel verbs (see also Casenhiser & Goldberg, 2005; Goldberg, Casenhiser, & Sethuraman, 2004; and Goldberg, Casenhiser, & White, 2007) but also to specifically test the learning of the mapping rules between syntactic positions and thematic roles (linking). In order to ascertain the latter point, the testing session included a comprehension task (forced choice) with pairs of videos presented simultaneously depicting scenes where the semantics of the verb (appearance) and the two characters involved were constant but the action was reversible, with each character being the theme in one video and the location in the other. The results of the comprehension and of an additional production task showed significant learning effects after a very short exposure at immediate post-test for the linking rules, although the effects were not maintained at delayed posttest a week later.

#### 4.3.2 Studies focusing on the type of language knowledge attained

A number of studies that investigated adult morphosyntactic learning using a miniature language paradigm also tried to elucidate the nature of the language knowledge gained by the learners as a result of exposure (among others, Francis et al., 2009; Grey, Williams, & Rebuschat, 2015; Rebuschat & Williams, 2012; Robinson, 2002; 2005; Rogers, Révész, & Rebuschat, 2015). For example Rebuschat and Williams (2012) report on two experiments that used a modified natural language with English lexis and German word order (V2 word order in main clauses and OV order in main and subordinate clauses). In experiment 1, 20 L1 English speakers were exposed to 128 aural sentence stimuli under incidental learning conditions and simultaneously asked to judge their semantic plausibility. The stimuli included four word order patterns: V2 in main clauses, with the first element being a phrase or a sentence, V-final in main clauses, and V-final in subordinate clauses. In the testing phase both the experimental and the control group (15 participants) were administered an aural GJT consisting of 64 sentences (half of which ungrammatical).

The test revealed no significant learning effect, nor a significant difference between the experimental group and control. However, the development of conscious knowledge in the learning process, assessed through an analysis of confidence ratings and knowledge source attribution, emerged as a critical modulating factor. In particular, the group who had developed awareness of the language rules showed significant learning effects and there was a significant advantage compared to both the unaware group and the control group, specifically in the case of accuracy in the identification of ungrammatical sentences.

In experiment 2, 30 L1 English speakers were trained in the same modified natural language (15 controls) and a similar design was adopted although fewer syntactic patterns were presented in the exposure. In this second study the experimental group performed significantly above chance and outperformed the control group in the GJT. However, contrary to what was found in experiment 1, this time the advantage emerged in the accurate judgment of grammatical sentences. Again, the analysis of the confidence ratings and the source attribution revealed that although participants could not verbalize the rules of the language, partial awareness of its syntax was related to a positive performance in the experimental group.

Building on the methodology developed in Williams and Kuribara (2008), Grey, Williams and Rebuschat (2015) deployed Japlish to investigate morphosyntax acquisition in a third language under incidental learning conditions. The study tested 36 L1 English low and high proficient learners of Spanish (15 and 21 respectively), immediately after aural exposure to the language (128 aural sentence stimuli) and after a two-week lag. An aural AJT (acceptability judgment test) was deployed as a measure of receptive word order learning. Unlike previous studies with Japlish, the acquisition of the relationship between thematic structure and case marking was investigated using a picture matching task (PMT). The results of this study at immediate posttest revealed that learners performed significantly above chance on word order but not on case marking. However the results of the delayed posttest showed that gains on word order were maintained and accuracy on case marking had improved. A comparison between the scores of low and high proficiency Spanish speakers also revealed that for the advanced group there was a significant correlation between the total number of semesters of Spanish study attended by the students and the AJT performance at delayed posttest.

Unlike Williams and Kuribara (2008), this study included measures of awareness to investigate the extent to which the knowledge acquired during the instructional process was implicit. Analysis of confidence ratings, source attribution and verbal reports showed that confidence in the response and reliance on intuition were the two factors that had a significant influence on accurate performance. Also, the verbal reports revealed that most learners were able to verbalize correct rules at least for noncomplex sentence patterns and some could provide correct exemplifications of how case marking worked, suggesting that incidental instruction had led to a grammatical knowledge representation that was largely explicit.

Finally, Rogers, Révész and Rebuschat (2015) used a modified natural language based on English and including 24 Czech nouns with nominative and accusative endings to test 42 English monolinguals (21 of which controls) on the acquisition of case morphology under incidental learning conditions. The participants were exposed to auditory sentence stimuli (144 items including subject-object canonical and scrambled orders). In each trial they were also asked to perform a force-choice vocabulary task whose purpose was to hinder sentence analysis. The result of a GJT administered immediately after exposure showed that there was a significant learning effect for the experimental group compared to controls, even if learning of the endings was only slightly above chance. The learning effect was driven mainly by the acquisition of one of the two forms (the accusative), a result reminiscent of what Robinson (2002, 2005) found for Samoan (in that case the locative marker was learnt better). There was evidence that participants acquired at least some implicit knowledge of the language morphosyntax during exposure as only GJT responses based on intuition were accurate significantly above chance and none of the participants could verbalize the morphosyntactic rule.

Overall, the picture emerging from adult studies that looked at the nature of morphosyntax knowledge acquired by the participants as a result of exposure suggests that, at least for this age group, awareness of language regularities played a crucial role in supporting L2 attainment. However, in at least one study (Rogers et al., 2015) there was some evidence that above chance learning was associated to knowledge that was at least partially implicit.

### 4.3.3 The role of individual differences in adult studies

Internal factors such as the role of cognitive individual differences (e.g., long-term memory, executive function, phonological short-term memory, working memory) represent a further important variable adult studies on L2 morphosyntax in miniature language paradigms have considered. Studies in this area of investigation include among others Antoniou, Ettlinger, and Wong (2016), Brill-Schuetz and Morgan-Short (2014), Brooks, Kempe and Sionov (2006), Brooks, Kwoka, and Kempe (2017), Carpenter (2008), Ettlinger et al. (2010); Ettlinger, Bradow and Wong (2014), Morgan-Short et al.

(2014), Pili-Moss and Morgan-Short (2018), Tagarelli, Borges Mota, and Rebuschat (2015), Williams and Lovatt (2003).

For example, Brooks, Kempe and Sionov (2006) trained 60 L1 English speakers to comprehend and produce novel locative morphosyntactic patterns providing pictures and auditory input embedded in short dialogues for six one-hour sessions over two weeks. Focusing on a portion of the Russian preposition/noun declension system they investigated how vocabulary size and measures of verbal working memory, nonverbal intelligence and executive function mediated the learning of subject-locative constructions considering both old items and generalization to new items. They found that exposure to a larger vocabulary set (i.e. evidence of a larger number of different contexts where the rule is exemplified) significantly predicted item learning and generalization, but only in participants with higher IQ and executive function.

Adopting a similar design (but with three 2-hour sessions, one week apart) Brooks, Kwoka, and Kempe (2017) exposed 54 L1 English speakers to a miniature natural language consisting in a portion of the Russian case-marking paradigm. They investigated: (1) how recall of trained items and generalization to new items are mediated by the type of input manipulation (skewed vs. balanced distribution of the nouns used to exemplify case marking), and (2) the role of individual differences such as statistical learning ability and nonverbal intelligence. Fitting a mixed-effects model that controlled for the effect of the individual difference variables (statistical learning ability and nonverbal intelligence) they found that only the balanced input condition was significantly associated with accurate generalization of the case marking patterns to novel items. They also found that their measure of nonverbal intelligence significantly predicted morphology learning, whilst statistical learning ability was a predictor of accurate case comprehension and production for old items.

One strand of research in the area of individual differences specifically set out to elucidate the role of learning abilities depending on long-term memory. For example, Morgan-Short, et al. (2014) is a BROCANTO2 study focusing on artificial language learning under incidental conditions in relation to two measures of declarative learning ability (part 5 of the MLAT and the Continuous Visual Memory Task - CVMT) and two measures of procedural learning ability (the Tower of London task and the Weather Prediction task). The study tested Ullman's DP model on the prediction that a positive relationship between procedural memory measures and L2 acquisition should be found at later but not at earlier stages of acquisition, i.e. only when the learner's proficiency in the language is sufficiently high.

Fourteen university students took part in the study over 7 sessions (three to four days apart), with a total of about 10 hours of language practice over 4 nonconsecutive sessions. The measure of learning was the (difference in) performance on two GJTs administered in session 3 (after 2.6 hours of practice) and in session 7 (after 10 hours of practice). The results showed a significant predictive relationship between declarative learning ability and language learning at early stages of practice, and between procedural learning ability and language learning at late stages of practice.

This study followed a similar methodology to Carpenter (2008) and Carpenter, Morgan-Short, and Ullman (2009) where adult performance in BROCANTO2 was also found to significantly relate to measures of declarative and procedural learning ability. In terms of results the main difference between the two studies pertains to the relationship between the procedural learning ability scores and the L2 attainment under incidental exposure. Whereas it appears to be positive and linear in Morgan-Short, Faretta-Stuttenberg, et al. (2014), it followed a parabolic trajectory in Carpenter et al. (2008), with the highest L2 attainment recorded for low and high procedural memory scores and the lowest for mid-range scores.

Consistently with the DP model, which predicts a competitive interaction between procedural and declarative memory, Carpenter (2008) suggested to interpret the high attainment of high procedural memory scorers as a result of procedural memory being strong enough to outcompete declarative memory strategies, and the high attainment of low procedural learning ability scorers as an indication of declarative memory engagement. The data interpretation for mid-range procedural memory scorers is that for those learners procedural memory was strong enough to interfere with learning mechanisms depending on declarative memory but not enough to result in high attainment.

Pili-Moss, Brill-Schuetz, Faretta-Stutenberg & Morgan-Short (2018) investigated the role of declarative and procedural learning ability during practice analyzing accuracy and RT data collected but not discussed in Morgan-Short et al. (2014). They found that, in contrast with the GJT results reported in Morgan-Short et al. (2014), only declarative memory predicted accuracy when it was assessed using a continuous measure, independently of the point in time during practice. Further, they found that procedural learning ability, but not declarative learning ability, was a predictor of automatization in comprehension (measured by the coefficient of variation, Segalowitz, 2010). Finally, the automatization analysis also returned a significant interaction between declarative and procedural learning ability, showing that declarative learning ability enhanced the effect of procedural learning ability leading to increasingly better automatization as practice progressed.

Brill-Schuetz and Morgan-Short (2014) used BROCANTO2 to train 26 L1 English speakers under incidental or explicit learning conditions following the procedure developed in Morgan-Short (2007) and adapted in subsequent studies. Unlike Morgan-Short (2007) the study run over 4 nonconsecutive sessions and included measures of procedural memory. The training period was limited to only 2 sessions (day 1 and day 3), with 10 comprehension and 10 production modules in total.

The measure of learning was accuracy in two aural GJTs administered on day 1 and day 4, respectively. Although the training phase was shorter compared to previous similar experiments, there was still a significant learning effect. However, no significant differences between the two conditions were found, probably due to the fact that the exposure was not long enough (Brill-Schuetz & Morgan-Short, 2014, p. 264). Interestingly, analysis of the procedural memory scores in the implicit condition revealed that participants with high procedural memory performed significantly better than participants with low procedural memory. These findings are consistent with Carpenter (2008), Ettlinger et al. (2014), Morgan-Short, Faretta-Stutenberg, et al. (2014), which also indicated an important role of procedural memory in incidental learning contexts.

Investigating the role of different memory subsystems in language learning, Antoniou et al. (2016) studied how procedural, declarative and working memory modulate the learning of simple and complex morphosyntactic rules in the early stages of training, as well as the role of feedback. In three subsequent experiments they trained a total of 122 participants in an artificial language consisting of 30 noun stems and two affixes (to create diminutive and plural words respectively, 120 words in total). In the simple rule the affixes simply attached to the noun, while in the complex rule affix/stem vowel harmony was added. Rule learning was measured by a forced-choice task, whereby participants had to choose which of two modified spoken words corresponded to a picture they were shown. After responding, feedback was given in terms of a 'correct' vs. 'incorrect' response. Overall the study found that procedural memory significantly predicted simple rule learning and declarative memory was a significant predictor of complex rule learning. They found that neither a measure of working memory, nor the presence vs. absence of feedback during testing, were significantly related to learning in either the simple or complex rule condition.

Investigating phonological short-term memory, Williams and Lovatt (2003) found that it significantly correlated with learning of a morphological rule in L1 English adults exposed to L2 Italian nominal gender agreement. Finally, Tagarelli et al. (2015) found that working memory predicted learning of word order in a semi-artificial language but only in explicit instruction conditions.

Overall, the adult studies that considered the role of cognitive individual differences in miniature language learning show that these cognitive variables have a significant predicting effect on learning that is moderated by a range of factors including e.g. amount of training, target type and complexity, time of no-exposure after training, vocabulary size in training and instruction conditions.

## 4.4 Summary

### 4.4.1 General considerations

The present review highlights a series of aspects in the artificial language literature that deserve a detailed discussion in view of how they could inform the methodology and the research questions of the present study. A clear element emerging is the variety of variables that have been taken into consideration in different strands of research in this area of investigation and the numerous methodological differences among studies.

***Learning Environment/Interaction***. The first consideration concerns the learning environment and the opportunity it provided to deliver language exposure where the language was not the main focus of instruction. In some studies, for example in all the BROCANTO2-based studies, the goal of the activity was gaining more points/doing better in the game, and language comprehension was instrumental in achieving that goal. In other studies the acquisition of new linguistic forms was clearly presented as the main purpose of the learning task, a situation that is likely to support explicit language analysis of the input, especially in adults. Studies also differed with respect to the amount of interaction between researchers and participants, with researchers directly delivering the exposure stimuli at one end of the spectrum and exclusively computer-mediated instruction at the other.

***Type of Language Knowledge***. A further point pertains to the evaluation of learning and its outcomes in terms of the implicit/explicit distinction. As emerges from the review, only some studies have included measures of awareness or additional experimental measures to assess the nature of the knowledge acquired as a result of incidental exposure. We find studies where both verbal reports and multiple subjective measures were deployed (e.g. Rebuschat & Williams, 2012) as well as studies including only retrospective verbal reports or debriefing questionnaires (e.g., Francis et al., 2009; Lichtman, 2012; Morgan-Short, 2007 and subsequent work). To the best of the author's knowledge, measures of awareness have not been used to date in miniature language studies with children.

Recording measures of child language awareness especially in comparative studies is of doubtless interest. If less developed cognitive abilities in children result in a more limited role of explicit learning and explicit knowledge, this would provide an opportunity to observe what type of implicit knowledge emerges as a result of the learning process. At the same time, it could open new research avenues in the study of what type of explicit knowledge is available to children compared to adults.

***Cognitive Individual Differences***. A further point differentiating miniature language studies is the inclusion of measures of individual differences in the design. Among the studies that included individual difference variables, Carpenter (2008), Morgan-Short (2007), and Morgan-Short et al. (2014) and others considered measures of procedural and declarative memory; Grey, Williams, and Rebuschat (2015) took the level of proficiency in another L2 into account, whilst DeKeyser (1994, 1995) and de Graff (1997) assessed the students' aptitude to explicit learning. To the best of my knowledge artificial language studies reporting measures of individual differences have mostly targeted adult populations to date (but see Kapa & Colombo, 2014).

***Type of Miniature Language***. A further aspect bears upon the choice of miniature language deployed, and more specifically on the relationship between the artificial language system adopted and the learners' L1. A number of studies with adult participants used modified natural languages that relied on L1 lexis, and incorporated only certain aspects of a second language (morphology or word order). This methodological choice has undoubtedly the advantage of focusing learning on the formal aspects that are of interest, speeding up the learning process (essentially because less or no vocabulary instruction is needed). On the other hand it is not clear how the use of L1 lexis may affect the activation of native syntactic/semantic representations during processing and how these effects can be controlled. For example, in the case of verbs, interference may occur due to the verb's activation of a representation that links its thematic arguments to specific syntactic positions in the learner's L1 that are different from the ones in the L2 that is the object of learning.

***Length of Exposure***. Another relevant point concerns the length of exposure. Longer exposure to input is arguably relevant to maximize the possibility of detecting a significant learning effect. However it is also key in assisting implicit learning processes, which are thought to emerge more clearly as a result of substantial and reiterated input provision. As a consequence it is to be expected that significant learning emerging as a result of short exposure periods will be related to an increased role of explicit/declarative learning in the learning process. A number of adult studies that included measures of awareness in their design have confirmed this prediction reporting that in short exposure conditions explicit knowledge was the substantial factor driving learning.

***Outcome Measures***. A further point pertains to outcome measures. Methodologically, studies have differed in the type of outcome measures they have adopted, with most studies using behavioral measures (mainly GJTs and language production) and only a few triangulating behavioral and neurophysiological measures (mostly ERPs). Furthermore, in most of the studies reviewed here, outcome measures were taken only once at the end of the exposure/training phase. This means that they provided an indication of the final state of learning but gave no elements to assess intermediate stages in the learning process. The only exception to this has been de Graff (1997) and the series of studies with BROCANTO2, initiated by Morgan-Short (2007), where at least two offline GJT measures were taken.

### 4.4.2 Age comparisons

The most obvious way in which the studies presented in this review differ is with respect to the age group they targeted, with studies including exclusively adults, exclusively children, and only some of them carrying out an analysis of adult and child learning under comparable exposure conditions. The comparison of child and adult learning using artificial languages offers an opportunity to observe how learners of different age groups acquire a new linguistic system in controlled conditions. As such, it constitutes a primary source of evidence relevant to questions of general theoretical interest such as those surrounding the study of age effects. In consideration of this it is interesting to note that only a few of the studies reviewed here have formulated hypotheses/research questions based on specific theoretical predictions in relation to the possible differences in second language learning between the two age groups (see for example MacWhinney, 1983). The comparison of results across age groups seems reasonable also in view of maximizing the generalizability of the research findings in this area of investigation.

It is clear that comparative research in this area faces the challenge of developing a methodology that takes into account specific between-group differences, primarily differences in cognitive development. For example it is known that working memory efficiency improves continually during childhood up to young adulthood and that adults have more efficient working memory and broader attention spans compared to children (Gathercole, Pickering, Ambridge, & Wearing, 2004). Hence the first element to take into account in the design of a comparative study would be the age of the children, selecting and piloting an age range that has potentially the cognitive ability necessary to perform the linguistic task. At the same time, a balanced task that is viable in comparative study design will have to avoid excessive simplification. One element that emerged clearly in recent studies that have compared children and adults, is that the artificial languages deployed tended to display a relatively simple structure. A simple language can arguably guarantee that a learning effect will emerge in a relatively shorter time and within fewer sessions, an advantage when working with age groups whose drop-out rates tend to be high in studies with multiple-session designs (Ambridge & Rowland, 2013). However, comparative studies should also consider that adults, due to their more developed cognitive skills, may resort to learning strategies (such as chunk memorization), leading to ceiling effects early on in training.

A further point concerning a number of comparative studies reviewed here is the importance of the input structure in the experimental design. This methodological point has been considered especially in studies where input type and structure are investigated as a variable modulating learning. A recurring element that emerged in artificial language studies involving children and adults is the different way input structure affects learning in the two age groups. In contexts where input is consistent, children tend to be more conservative than adults and tend not to generalize beyond instances present in the input (e.g., Boyd & Goldberg, 2012). However, if there is evidence that item distribution is probabilistic a different pattern emerges; children (at least in production tasks) show a stronger tendency to regularize item distribution, whilst adults are more likely to re-produce the probabilistic distribution they were exposed to (Hudson Kam & Newport, 2005; 2009). Methodologically, a further finding has been that if the input is consistent, manipulating it so that it presents a skewed distribution seems to be beneficial in facilitating generalization independently of the targeted age group (Casenhiser & Goldberg, 2005). Therefore, independently of a study's specific interest in investigating input's role, considerations of input structure are unavoidable and should inform the creation of exposure sets, particularly when, as it is the case in artificial language studies, input can be more easily manipulated.

## 4.5 Focus of the investigation

Having completed the discussion of the theoretical background relevant to the present study and the literature review, I will now turn to the investigation reported in the present thesis. In this chapter I will start by providing an overview of the piloting phase and of the variables of interest underpinning the research questions. Chapter 5 will then provide a detailed description of the miniature language system and the learning environment that will lead to the introduction of Study 1 and Study 2.

### 4.5.1 Piloting phase

The experimental research undertaken in the present dissertation consists of two studies, a study with 40 L1 Italian 8-9 year old children (Chapter 6) and its replication with 36 L1 Italian young adults (Chapter 7). The experiments investigated comprehension and morphosyntactic attainment in the miniature language BrocantoJ (see Chapter 5) and were conducted in the BROCANTO2 paradigm in incidental instruction conditions controlled across the two age groups.

As the BROCANTO2 paradigm had not been deployed with children before, a piloting phase was necessary to establish whether the game complexity was suitable for typical developing children of the age range of interest, and the miniature language was likely to be learnt at levels of proficiency sufficient to allow meaningful analysis of the data (Pili-Moss, 2017). Another important point was the definition of a viable experimental design with regards to the number and the length of experimental sessions. A preliminary study (Pili-Moss, 2017) provided evidence that the language could be learnt at adequate levels by 8-9 y.o. children, in three sessions delivered on consecutive days, with each session lasting about 1 hour, corresponding to a total length of language exposure of about 2 1/2 hours.

### 4.5.2 Variables of interest

Considering the literature reviewed and the theoretical background outlined in chapters 1-3 of this dissertation the investigation in the two experimental studies focuses on the following research dimensions:

***Age***. The two experimental studies included in the dissertation test children and adults respectively. Overall the paradigm is one of the few in the miniature language literature to offer the possibility of an age comparison in controlled instruction conditions (other examples include Kapa & Colombo, 2014; Lichtman, 2012)

***Linguistic Target***. Language learning is assessed looking at gains in three main linguistic areas: word order, case marking, and learning of the linking rules mapping thematic roles to NP syntactic positions in a sentence. Word order has been extensively investigated in the literature for a range of different age groups, and as such provides ample opportunity to compare results in the dissertation to previous findings. On the other hand the investigation of case marking constitutes an extension of the current literature, since to date it has been studied in a miniature paradigm with adults but not with children. As for the learning of linking rules, previous studies with adults and/or children deployed simplified syntactic structures to model word orders, whilst the artificial language used in this dissertation is more complex both at sentential and NP level.

***Type of Outcome Measure***. The present dissertation adopts two main types of measure; a timed aural GJT administered once at the end of the language practice, and a continuous measure of learning taken online as the participants where practicing with the language while playing the computer game. The GJT has been used extensively in the artificial language literature, but only a few studies to date (DeKeyser, 1995 is one of them) have taken an additional continuous measure of language learning during practice. Since an important focus of the dissertation is to elucidate the role of cognitive individual differences in modulating language learning (see next subsection), this design enables to study the extent to which the predicting value of the cognitive variables varies depending on the type of language learning measure adopted.

***Cognitive Individual Differences***. Following previous studies deploying BROCANTO2 or other paradigms that considered the role of individual differences in long-term memory, this thesis aims at elucidating to what extent declarative and procedural learning ability modulate language comprehension and grammaticality judgments in children and adults. To the best of my knowledge, this is the first training study to examine the relationship between L2 learning and long-term memory in children in any paradigm. As for adults, the contribution of the study will be to investigate to what extent the results that have been found in the BROCANTO2 paradigm for L1 English participants can be extended to participants with a different L1.

Unlike previous BROCANTO2 studies, the ability to retain vocabulary taught in the training sessions will also be considered a type of declarative learning. Vocabulary learning ability measures the long-term ability to correctly *recognize* arbitrary associations between aural and visual stimuli, whereas the tasks used to assess declarative learning ability in the present studies probes the ability to accurately *recollect* visual or verbal stimuli. Although both measures are clearly linked to abilities relating to declarative long-term memory, it would not be methodologically advisable to directly compare the two effects or create a composite measure.

The main reason is that the effect of vocabulary learning ability is likely to be affected by confounding factors such as the re-enforcement of visual/aural associations during training or the fact that, unlike the declarative learning ability measures, it reflects the strength of visual/aural associations that are directly relevant to accurate performance during practice. Furthermore, whilst participants in the declarative learning ability tasks are asked to recollect stimuli in a single modality (visual or verbal), they have to associate stimuli cross-modally in vocabulary learning.

***Type of Language Knowledge***. As a final variable of interest, the present thesis investigates the nature of the language knowledge gained by children and adults along the implicit/explicit continuum. To this end, verbal reports and confidence ratings relative to the GJT trials are analyzed for both groups. In the literature on implicit/explicit language knowledge, child studies, and in particular child studies with confidence ratings, are scarce. To date I am not aware of any study investigating the nature of language knowledge that has compared a child and an adult group trained in the same paradigm and has additionally assessed language knowledge through subjective measures.

# 5. The artificial language BrocantoJ

## 5.1 Introduction

The aim of this short chapter is to present the lexical and syntactic characteristics of BrocantoJ and to introduce the computer game that constitutes the L2 practice environment in the experimental studies. With regards to BrocantoJ morphosyntax, the distinction between symmetric and asymmetric trials (5.7) will be particularly relevant to the experiment design.

## 5.2 Vocabulary items

BrocantoJ is a version of BROCANTO2 (Morgan-Short, 2007) and displays the morphosyntax of Japanese. A previous version of BrocantoJ deployed in the pilot study (Pili-Moss, 2017), was used to instruct L1 English participants and had the same vocabulary items as BROCANTO2. As participants in the present studies are L1 Italian native speakers, vocabulary items were adapted to match the phonotactic characteristics of this natural language as closely as possible (Table 5.1).

In this respect a major difference was that the words in the present version of BrocantoJ are not monosyllabic (except the case markers). This is due to the fact that with the exception of pronouns and prepositions, monosyllabic nouns and verbs are extremely rare in Italian and mostly words of foreign origin. Thus in BrocantoJ all vocabulary items except the case markers are bisyllabic with a syllabic structure either of the form CCVCV / CVVCV or CVCV (C = consonant/V = vowel). Care was also taken that none of the vocabulary items had an independent meaning in Italian. For this reason the original monosyllabic case markers (*li*/*lu*) were modified to *ri*/*ru*, as *lì* is a meaningful word in Italian (meaning 'there'). The change in the first consonant (from *l* to *r*) was motivated by the need to maintain it phonetically as close as possible to the original (the initial consonant in both sets of markers is a liquid). The vowels *a*/*e*/*i*/*o* are used as endings of both nouns and verbs so that they cannot serve as cues to identify nominal categories.

Table 5.1

*Vocabulary Items in the Italian Version of BrocantoJ*

| BrocantoJ (Italian) | Category | Meaning |
|---|---|---|
| blomi | Noun | the 'blomi' token |
| nipo | Noun | the 'nipo' token |
| pleca | Noun | the 'pleca' token |
| vode | Noun | the 'vode' token |
| trose | Adjective | round |
| neimo | Adjective | square |
| klino | Verb | move (intrans.) |
| nima | Verb | capture (trans.) |
| yabe | Verb | release (trans.) |
| prazi | Verb | switch (trans.) |
| noika | Adverb | vertically |
| zeima | Adverb | horizontally |
| ri | Preposition | nominative case |
| ru | Preposition | accusative case |

As illustrated in Table 5.1 the lexicon of BrocantoJ includes 14 items in total; 4 nouns, 2 adjectives, 4 verbs, 2 adverbs, and 2 prepositional case markers. BrocantoJ has the same lexical categories as BROCANTO2 with the difference that in BrocantoJ adjectives do not have gender agreement. In addition to that the monosyllabic particles (used as postnominal determiners in BROCANTO2) are here deployed as the nominative and the accusative prepositional markers respectively. The verbs are all obligatorily transitive, i.e. always occur with an object noun phrase, except *klino,* which is intransitive.

## 5.3 Characteristics of Japanese morphosyntax

In order to understand the morphosyntax of BrocantoJ it is useful to review some properties of Japanese main clauses that are mirrored in this artificial language. The first of these properties is that Japanese is obligatorily verb final in main clauses:

| | | | | | |
|---|---|---|---|---|---|
| (1) | Kodomo | ga | terebi | o | mita |
| | [child | NOM | TV | ACC | watched] |
| | 'The child watched TV' | | | | |

Secondly, Japanese has overt case marking on nominals. However, case marking is not realized through morphological suffixes (as it is for example in German), but deploys prepositions that follow the nouns and are phonologically independent elements. The suffix prepositions *ga*, or *wa* are used for nominative case, and *o* or *wo* for accusative case. There is no person and gender agreement on nominal elements in this language.

Although it has deictic determiners (words corresponding to *this*/*that* in English), Japanese lacks articles. Inside the nominal phrase the relative word order of adjectives and nouns is Adj-N, as shown in (2):

| | | |
|---|---|---|
| (2) | akai | karuma |
| | [red | car] |
| | 'the/a red car' | |

All other arguments and adjuncts of the verbs including –*ly* adverbs (e.g. vertically, horizontally) occur preverbally (3).

| | | | | | | |
|---|---|---|---|---|---|---|
| (3) | Sensei | wa | nihongo | wo | jouzu ni | hanashimasu |
| | [teacher | NOM | Japanese | ACC | skillfully | speaks] |
| | 'The teacher speaks Japanese well' | | | | | |

Finally, Japanese allows null pronouns as subjects of finite sentences if the missing information is specified as *topic* in the context, i.e. under certain circumstances Japanese can be classified as a *pro*-drop language (Rizzi, 1982). Note that in main clauses, the only syntactic context relevant to BrocantoJ, Japanese also allows free word order between subject and direct object (*scrambling*). However, noncanonical word orders in Japanese are optional and it was decided to omit *scrambling* from the BrocantoJ exposure sets in order to limit rule complexity.

### 5.4 Rationale for choosing a language modeled on Japanese

The decision of using Japanese as a model for the artificial language in the two experiments with L1 Italian participants, was based on mainly four considerations. The first is that in terms of word order and morphology Japanese differs significantly from Italian as well as from other Romance languages and English, the languages most widely taught at primary level in Italy. This means that, unlike BROCANTO2, which had elements of Romance morphosyntax, BrocantoJ allows to control for confounding factors due to the similarities of the artificial language with the participants' L1 as well as with the second language they are exposed to in school.

Adopting BrocantoJ also allows to extend the BROCANTO2 paradigm to include the study of the L2 acquisition of case marking and of form-meaning relationships. Unlike gender morphology, which in natural languages is often redundant and normally does not hinder comprehension if not detected accurately, case morphology is directly related to the realization of the verb’s argument structure and indicates the assignment of thematic roles to the arguments. It is important to note that, since the word order in BrocantoJ is fixed, the interpretation of the argument structure relies on both case realization and argument structure in contexts where both subjects and objects are present. However, in sentences with a single object noun phrase (OV sentences) no syntactic cues for thematic interpretation are available, and this relies solely on case marking.

Although the fact that Japanese has postnominal prepositional markers could constitute a potential problem for learnability, due to their limited phonological salience, there is ample evidence coming from previous experiments with BROCANTO2 that postnominal elements of comparable phonological status are learnable.

Finally, BrocantoJ is modeled on the morphosyntax of a natural language that is widely spoken as a first language as well as learnt as a second language.

### 5.5 BrocantoJ sentence types

Consistently with Japanese morphosyntax, the linear structure assumed for a BrocantoJ main clause is as follows:

**(4) (Adj-Noun-NOM marker) - (Adj-Noun-ACC marker) – Adv – Verb**

NP1 NP2

The following are examples of the three BrocantoJ sentence types a participant would be exposed to during the experiment. They are an intransitive SV sentence (5), a transitive SOV sentence (6), and a transitive OV sentence with a null subject (7). These sentences correspond to the move constellations illustrated in Figure 5.1. The example sentences are modeled after Pili-Moss (2017, p. 116):

**SV**

| | | | | |
|---|---|---|---|---|
| (5) | Neimo | blomi ri | noika | klino |
| | Square | blomi NOM | vertically | move |

'The square blomi token moves vertically'

**SOV**

| | | | | | |
|---|---|---|---|---|---|
| (6) | Trose | blomi ri | neimo blomi ru | zeima | nima |
| | Round | blomi NOM | square blomi ACC | horizontally | capture |

‘The round blomi piece captures the square blomi piece horizontally’

**OV**

| | | | | |
|---|---|---|---|---|
| (7) | Neimo blomi ru | zeima | nima |
| | square blomi ACC | horizontally | capture |

‘It/another token captures the square blomi piece horizontally’

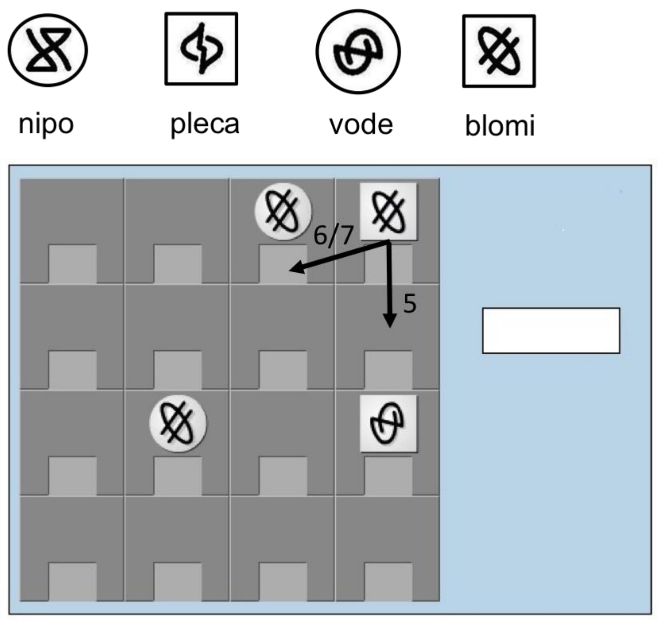


*Figure 5.1*. Game Tokens and Constellations Corresponding to the Sentences (5), (6), and (7) in the Main Text (Pili-Moss, 2017, p. 117).

## 5.6 The computer game

In the experimental paradigm children and adults are aurally exposed to the language BrocantoJ in the context of a computer board game (Morgan-Short, 2007 and subsequent studies; Pili-Moss, 2017). The computer board (depicted in Figure 5.1) is constituted by 16 main token positions (squares), each containing a smaller rectangle. Each main square is referred to as a 'home' position and in the initial rule explanation, a token positioned there is referred to as the 'owner' of the position. However, the participants are told that, under certain circumstances, a token can also find itself inside the smaller rectangle when the 'owner' is 'at home'. Participants are also told that moves are allowed horizontally or vertically (not diagonally).

Once the general game rules have been explained using a physical game board, the 4 move constellations are introduced by showing the participants short animations. The moves are (a) 'move' (*klino*, a simple one-square move, see Figure 5.1, number 5); (b) 'capture' (*nima*, a token ends up in the internal square of an adjacent token's position, Figure 5.2 [a]); (c) 'swap' (*prazi*, two adjacent tokens swap positions with one of the two moving first/ initiating the move, Figure 5.2 (b)); and (d) 'release' (*yabe*, a captured token is released in an adjacent free 'home' position, Figure 5.2 [c]). It is important to note that the move names are never translated and each move is shown on video and then associated to its name with no further cues to the agent-patient semantics involved.

In the computer game participants play individually and gain five points each time they perform the correct move after hearing once a full BrocantoJ sentence that describes it (20 moves per game block, a maximum of 100 points per block). They gain no points if the move is incorrect. After each move is performed, the players receive visual onscreen feedback by means of the words 'Correct' (in green) or Incorrect (in red), appearing in the white rectangle on the side of the game board (Figure 5.1). Immediately after each move, or after one minute if no move is performed, the game presents the next aural stimuli and game constellation (the next move is not sequentially related to the preceding one and a completely new constellation involving different token positions and a different number of tokens is shown on the board). Unlike previous BROCANTO2 studies, the players are not shown the running score, but only a percentage correct final score at the end of each game block.

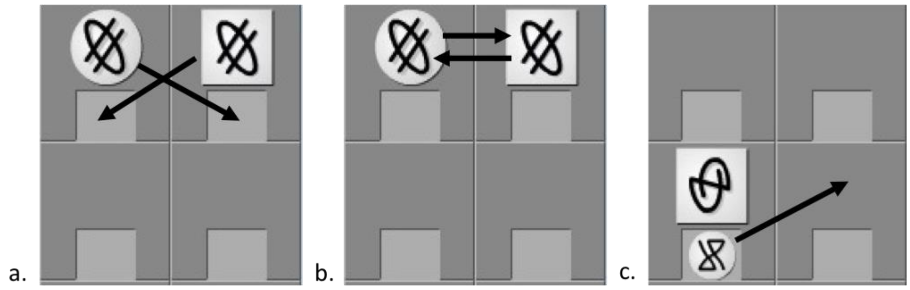


*Figure 5.2*. Symmetric Moves Corresponding to the Verbs *Nima* (a) and *Prazi* (b), and the Asymmetric Move Corresponding to *Yabe* (c).

## 5.7 Symmetric and asymmetric scenarios

It is important to note that, as Figure 5.2 (a) and (b) illustrate, two of the three moves expressed through transitive verbs ('capture' and 'swap') visually correspond to *symmetric* configurations. This means that given an initial game constellation each of the two tokens could capture/be captured or actively swap/be swapped). In this context the visual information offered by the initial constellation does not provide cues as to which of the two tokens will move (or move first), and the performance of the correct move relies completely on the aural comprehension of the mapping between thematic functions and syntactic positions or between thematic functions and case markers.

On the other hand the third move ('release', Figure 5.2 [c]) exemplifies an *asymmetric* configuration (the same can be said of *klino*, Figure 5.1). In the 'release' scenario the visual information provided in the initial constellation unambiguously indicates which of the two tokens will move. This means that, provided that the type of move, the tokens and the direction of movement have been correctly identified in the constellation, an understanding of the morphosyntactic form-meaning relationships in the sentence is not strictly necessary. In this case accurate performance could rely on the application of general heuristic strategies. In conclusion, the possibility to distinguish between symmetric and asymmetric stimuli allows the identification of a subset of stimuli that can be used to specifically assess the participants' comprehension of the relationship between morphosyntax and thematic interpretation during practice.

# 6. Study 1

## 6.1 Introduction

This chapter introduces the first study designed to address the area of investigation broadly identified at the end of Chapter 4. Focusing on a child sample, it will explore the relationship between long-term memory and L2 learning, as well as investigate the type of L2 knowledge developed by the participants as a result of the exposure to the artificial language. The chapter includes Methods and Results and the Discussion is reported in the first part of Chapter 8.

## 6.2 Research questions

In the light of the literature review in the area of artificial L2 learning (Chapter 4) and the theoretical background provided in Chapters 1-3, the following research questions were formulated for Study 1:

RQ1 To what extent do declarative and procedural learning ability modulate child L2 aural comprehension and learning of the rules linking morphosyntax and thematic interpretation during practice?

RQ2 To what extent do declarative and procedural learning ability modulate child L2 learning of word order and case marking as measured by a grammaticality judgment test administered at the end of practice?

RQ3 To what extent is the L2 knowledge acquired by the children implicit/ explicit?

The role of declarative and procedural learning ability in L2 learning has been investigated only in adult studies to date. Although recent theoretical approaches (e.g. DeKeyser, 2012; Ullman, 2005; 2015; 2016) have posited or implied an important role for procedural learning ability in child L2 acquisition, this prediction has not yet been tested and the nature of RQ1 and RQ2 remains highly explorative. In general, based on child artificial language studies conducted with children of 9 years of age or younger (e.g., Boyd & Goldberg, 2012; Braine et al., 1990, Casenhiser & Goldberg, 2005; Ferman & Karni, 2010; Lichtman, 2012; Wonnacott et al., 2012) and on a preliminary study (Pili-Moss, 2017), it can be hypothesized that 9-year-old children should be able to learn the word order and the morphology of an artificial language of the complexity of BrocantoJ in incidental conditions. Further, previous studies also support the prediction that children in the age range will learn form-meaning relationships linking syntactic position (Boyd & Goldberg, 2012; Wonnacott et al., 2012) or morphology (Pili-Moss, 2017) to thematic interpretation.

Note that in Pili-Moss (2017), as well as in the present studies, the instruction conditions were incidental in the sense that there was no explicit instruction on the morphosyntacic rules of the language, the participants were not invited to induce them from the input, they were not aware that their performance was being recorded and that they would be tested. However, they were obviously aware they were being exposed to a new language and that understanding the meaning conveyed by the language would help them perform better in the game.

With regards to RQ3, Lichtman (2012, Study 2), the only other artificial language study that has investigated type of child language knowledge in incidental instruction conditions, found that implicitly instructed children, unlike explicitly instructed children, tended not to display awareness of the linguistic patterns they were exposed to. Since children in Study 1 are exposed to BrocantoJ in incidental instruction conditions, the hypothesis for RQ3 is that they will mainly develop implicit knowledge of BrocantoJ morphosyntax.

## 6.3 Methods

### 6.3.1 School context

The participating primary school was a state school located in Northern Italy in a South-West suburb of Milan with a mixed socio-economic background. The head teacher contacted the researcher after the study was advertised to a number of primary schools in the area. In the Italian school system, pupils are taught for five years in primary school and start the first grade between 5 and 6 years of age. State schools are mixed-gender and at primary level include 20 to 25 children on average.

The current policy at national level is to teach a second language from the start of primary education. The school complied with this policy and additionally belonged to a small group of primary schools in Lombardy offering a CLIL program in which all children from the third grade onwards are taught three curricular subjects in English by L1 Italian teachers of English for a total of 5 hours a week. The majority of pupils came from an Italian monolingual background, with an average of 3/4 pupils per class coming from families where a language different from Italian was spoken. These data are in keeping with the national trend in comparable urban areas that has seen an increase in the number of bilingual children in primary education in the last ten years due to immigration. In total, the data collection period at the school lasted seven weeks in the autumn of 2016.

### 6.3.2 Ethics

The study, including the information and consent materials, was approved by the Ethics Research Committee of Lancaster University (Ref. RS2014/142). In keeping with school-internal requirements, participation in the project was sought and confirmed by a decision of the teachers' council at school level. After that, the forms for written informed consent were sent to the parents of the children in the classes involved in the study. Although the information materials specified that the participation in the study was voluntary and could be withdrawn at any point, the researcher reminded the children of this possibility at the beginning of the study and during the study. At the end of the data collection period all participating children were rewarded with a small gift and a certificate.

### 6.3.3 Participants

The school operated an initial selection of the potential participants based on their attainment in L1 literacy and on general and medical records, to which the researcher had no direct access due to privacy reasons. The school also provided information about the children's handedness. Based on this sample, the participants selected for the study were 53 L1 Italian monolingual children (17 females) from grade 3 (three different classes), 4 (three different classes) and 5 (two different classes). The data from 13 participants were excluded from the final analysis due to a variety of reasons; 4 participants (all female) were excluded because they did not reach the required minimum score in the computer game, 8 participants were excluded because their memory measures were collected but they did not play the computer game, and one participant was excluded because she told the researcher she wanted to stop playing the computer game due to tiredness.

Overall the final sample (Table 6.1) included data from 40 participants (10 females) for which mean age at testing was 9 years and 2 months (in months, $M$ = 109.5; $SD$ = 7.1).

Table 6.1

*Descriptive Statistics Relative to Age and Spacing between Training Sessions*

| | | Age (months) | S1/S2 (days) | S2/S3 (days) |
|---|---|---|---|---|
| Grade | *n* | *M (SD)* | *M (SD)* | *M (SD)* |
| 3 | 15 | 101.6 (4.3) | 1.3 (0.6) | 1.1 (0.5) |
| 4 | 21 | 112.5 (3.6) | 1.2 (0.4) | 1.0 (0.2) |
| 5 | 4 | 119.0 (0.0) | 1.2 (0.4) | 1.0 (0.0) |
| Overall | 40 | 109.5 (7.1) | 1.3 (0.5) | 1.0 (0.3) |

Fifteen participants (4 females) were in grade 3 ($M$ = 101.6; $SD$ = 4.3); 21 participants (3 females) were in grade 4 ($M$ = 112.5; $SD$ = 3.6); 4 participants (3 females) were in grade 5 ($M$ = 119; $SD$ = 0.0). The participants were all typically developing children with no diagnosis of learning differences or hearing impairment, with normal or corrected to normal vision and at least average attainment in L1 literacy.

### 6.3.4 Study design and set-up

The study design included three sessions on separate subsequent days lasting about 40-45 minutes, 50 minutes and 60 minutes respectively (Figure 6.1). Given that testing took place during term time and the children's availability for testing needed to be scheduled around their school commitments, the spacing between sessions ended up varying slightly for each individual child and across children (Table 6.1).

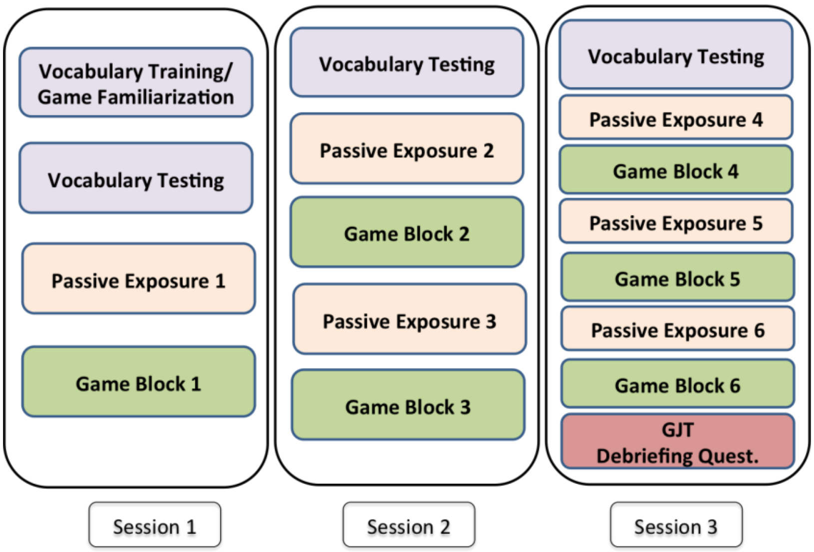


*Figure 6.1*. Design of Study 1.

To limit the duration of the testing session and minimize fatigue, the tasks used to assess memory were administered on different days, with order of days counterbalanced across participants (declarative learning ability and a measure of the phonological loop on one day, lasting about 40 minutes, and procedural learning ability on a separate day, lasting about 30 minutes). Hand motor control data were also collected from children and adults on the same day the measure of procedural learning ability was taken, but they were not analyzed for the purposes of the present studies. Except for the vocabulary training, the children were trained and tested individually, and sat in a quiet room at different laptop computers two at a time. During individually-administered computer tasks the children wore headphones and the volume was adjusted for each child before use.

During the experiment the children were given detailed instructions prior to each task. The opportunity to ask questions was given before the task started, as well as by embedding in-built breaks for this purpose in the task itself. Once the researcher had made sure the instructions were clear, the children were left to proceed by themselves. In the meantime the researcher sat at a different table in the same room ready to intervene if asked, whilst at the same time discreetly controlling the children remained on task.

In what follows the different elements of training and practice and the cognitive tasks will be presented. The description will include materials, procedure, the instructions that were provided to the participants and the operationalization of the outcome measures of L2 learning, L2 awareness and memory. The methods section will conclude with the presentation of the main inferential statistics deployed in the analysis of the results.

### 6.3.5 Vocabulary training and game familiarization

The vocabulary training materials consisted of a video, a section of the original game training materials used in Morgan-Short (2007) and a physical game board with tokens. The training video, initially created in power point and then saved in mp4 video format, lasted 4.38 minutes not considering pauses, and did not have audio. A child-friendly cartoon character (Suzy, Figure 6.2) accompanied the participants through vocabulary and language training introducing different parts of each task. The cartoon character was adapted from a picture freely available on the web (www.pixabay.com) under C0 Creative Commons license.


*Figure 6.2*. The cartoon character Suzy. This slide introduced the second move in BrocantoJ.

The rules of the game were distinct from the rules of the language and were introduced by the researcher using the relevant part of the original BROCANTO2 program (Morgan-Short, 2007 and subsequent studies). The program presented the game board, provided examples of the four move types and after each one gave the participants a chance to practice how to perform them in the program interface using the mouse. The physical game boards were reproductions of the computer game board (24x24cm) and tokens (about 2x2cm each) consisting of a foam base covered with adhesive laminated paper (Figure 6.3a).

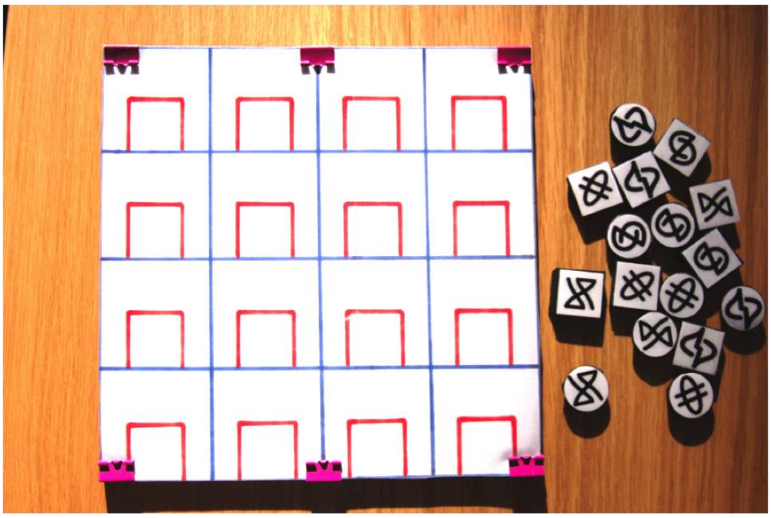

*Figure 6.3a*. Physical Board (24x24cm) and Tokens. Materials: Foam, Paper, Laminated Sheet, Metal.

The participants would initially sit in pairs in front of a laptop at about 50 cm from the screen. The researcher introduced them to the vocabulary of BrocantoJ using the power-point-based video, whilst part of the original BROCANTO2 program was used to practice the moves in the computer environment (game familiarization). Each participant was also given a physical game board and a set of tokens similar to the ones used in the computerized version of the game and was told they were free to manipulate them/try out moves throughout the vocabulary training and game familiarization. All vocabulary items were introduced aurally by the researcher without using translations and in association with a corresponding static picture (game tokens, adjectives and directions) or animation (moves).

In order to support memorization vocabulary items were presented and rehearsed in sets of four (game tokens, moves, and finally shapes/directions). For the same reason, throughout vocabulary presentation and rehearsal the children were encouraged to repeat the item names after the researcher had pronounced them. The postpositional case markers were not presented in the vocabulary training. Vocabulary presentation and rehearsal, game familiarization and vocabulary testing occurred according to the diagram in Figure 6.3b.

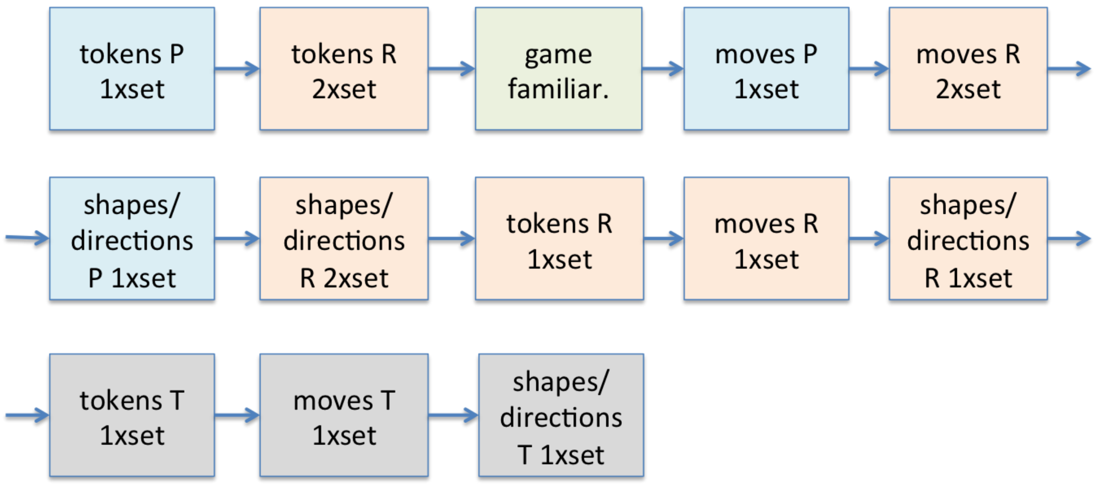


*Figure 6.3b*. Structure of the vocabulary training and game familiarization phases. P = presentation, R = rehearsal and T = testing.

In the rehearsal phase the researcher asked the children in turn to recall vocabulary items/ pictures associations or vocabulary items/animations associations (in the case of moves). For vocabulary items presented in association with static pictures the participants were asked to point at which of four pictures a given vocabulary item would correspond to. For moves, participants were presented with four subsequent animations on the same slide of the power-point video. After each animation the researcher stopped the video and repeated the four move labels in a pseudo-randomized order, asking the participant to repeat the word he/she thought described the move. If children had difficulties pronouncing the BrocantoJ word, responses of the kind of 'the first/second you said' etc. were also counted as correct. Visually, for each set of vocabulary items (tokens, moves or shapes/directions), the position of pictures/animations in the slide was pseudo-randomized with respect to the preceding slide depicting the same set (although it was the same for all participants). During rehearsal the researcher provided corrective feedback (correct/incorrect), additionally indicating the correct response in case of error (cf. Figure 6.3c).

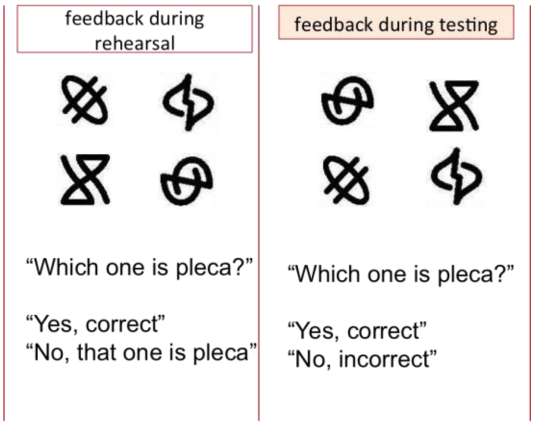


*Figure 6.3c*. Example of feedback given in the rehearsal and in the test phases (tokens).

After the vocabulary training the participants were tested individually. Vocabulary items were tested in sets of four elements (tokens, moves, shapes and directions). In the individual testing phase participants were asked to identify tokens, moves or shapes/directions in the same way as in the rehearsal. However this time the feedback was limited to the information of whether the response was correct or incorrect (without providing the correct response in case of errors, cf. Figure 6.3c). Following Morgan-Short (2007) and subsequent BROCANTO2 studies, the participants had to reach a criterion of 100% correctly identified word/visual associations in order to proceed to the subsequent stage of the experiment. If criterion was not reached, further instruction on the set/s where mistakes were made was provided, followed by a repetition of the vocabulary test (all items). The procedure was repeated for a maximum of four times. If criterion was still not reached at the fourth test the participant was excluded from the experiment. Vocabulary testing was repeated following the same procedure at the beginning of session 2 and 3 to make sure vocabulary knowledge had been retained. The number of errors made prior to reaching criterion was recorded in each session (Table 6.2).

Table 6.2

*Errors in the Vocabulary Test per Session*

| | | S1 | S2 | S3 | Tot |
|---|---|---|---|---|---|
| Grade | *n* | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* |
| 3 | 15 | 2.7 (2.0) | 1.3 (1.8) | 0.7 (1.1) | 4.8 (3.3) |
| 4 | 21 | 2.6 (1.9) | 1.1 (1.7) | 0.5 (0.9) | 4.2 (3.5) |
| 5 | 4 | 1.8 (1.5) | 1.0 (0.7) | 0.0 (0.0) | 2.8 (2.2) |
| Overall | 40 | 2.6 (1.9) | 1.2 (1.7) | 0.5 (1.0) | 4.3 (3.4) |

***Vocabulary Learning Ability Score***. A score of vocabulary learning ability was obtained for each participant standardizing ($z$-scores) the sum of errors in the vocabulary test across the three sessions multiplied by -1.

### 6.3.6 Passive exposure

Here I will refer to passive exposure as the phase of training during which the participants listened to full sentences in BrocantoJ, in association to a corresponding visual animation exemplifying the corresponding move in the game, but did not actively play the game. The passive exposure set consisted of a total of 144 distinct aural stimuli (full BrocantoJ sentences) presented in association with the corresponding game moves in 6 blocks (24 stimuli each; the full set of stimuli is provided in Appendix B). The participants were exposed to Block 1 in Session 1, to Block 2 and 3 in Session 2 and to Block 3, 4 and 5 in Session 3 according to the diagram in Figure 6.1.

In terms of the directions given to the participants, immediately after the vocabulary training and testing the children were told that whilst up to now they had learnt single words in the new language, in what followed they would hear more words spoken together and that these words would describe the game moves shown on screen. They were told that paying attention would help them make more points in the game they would play later on, and that the video would last about 4 minutes with no breaks. After each language training block the children played a game block (cf., Figure 6.1). Unlike previous studies deploying the paradigm (but similarly to Pili-Moss, 2017), the language training was administered in six short blocks instead of a single block in order to minimize the risk of drops in the children's attention.

Item frequency was counter-balanced across elements belonging to the same lexical category, although the set contained a larger number of accusative markers (*ru*) compared to nominative markers (*ri*), given that the number of OV sentences exceeded the number of SV sentences (cf., Table 6.3a).

Table 6.3a

*Frequency of Vocabulary Items in the Passive Exposure Set*

| Vocabulary items | Category | Frequency |
|---|---|---|
| blomi | Noun | 57 |
| nipo | Noun | 57 |
| pleca | Noun | 57 |
| vode | Noun | 57 |
| trose | Adjective | 38 |
| neimo | Adjective | 38 |
| klino | Verb | 12 |
| nima | Verb | 44 |
| yabe | Verb | 44 |
| prazi | Verb | 44 |
| noika | Adverb | 32 |
| zeima | Adverb | 32 |
| ri | Preposition | 96 |
| ru | Preposition | 132 |

The order of presentation of the sentence stimuli was the same for all participants. Overall, sentence length varied between 3 and 6 words in blocks 1 and 2 and between 3 and 7 words from block 3 onwards. The sentence length was kept comparatively shorter in the first part of the passive exposure because it was hypothesized that, initially, a greater amount of attentional and working memory resources would be allocated to familiarization with the mechanics of the game and the establishment/re-enforcement of initial vocabulary correspondences between aural and visual stimuli. In terms of the type of word order, each block included 14 SOV sentences, 8 OV sentences and 2 SV sentences. The frequency of SOV sentences was higher compared to the OV sentences in order to provide a sufficient number of exemplars where the linking rules between thematic interpretation and morphosyntax were expressed both syntactically (subjects preceding objects) and morphologically (subjects marked with nominative case and objects marked with accusative case). Since in OV sentences linking rules are expressed morphologically (through accusative case marking) but not syntactically, it was hypothesized that they would have been harder to interpret. Finally, the two SV sentences were included in the set to mirror the structure of the game set (as explained in 6.3.7 the sequence of game constellations in the game set could not be modified to exclude SV intransitive sentences). Although, overall, the number of transitive verb items was counter-balanced (as in the original BROCANTO2 studies), their distribution relative to the number of sentence stimuli assigned to each verb was skewed in individual blocks (cf., Table 6.3b).

Table 6.3b

*Frequency of Verbs per Sentence Type in the Passive Exposure Set*

| | | symmetric | | asymmetric | | |
|---|---|---|---|---|---|---|
| | | **nima** | **prazi** | **yabe** | **klino** | TOT |
| Block 1 | SOV | 10 | 2 | 2 | - | 14 |
| | OV | 4 | 2 | 2 | - | 8 |
| | SV | - | - | - | 2 | 2 |
| | **Tot** | **14** | **4** | **4** | **2** | **24** |
| Block 2 | SOV | 2 | 2 | 10 | - | 14 |
| | OV | 2 | 2 | 4 | - | 8 |
| | SV | - | - | - | 2 | 2 |
| | **Tot** | **4** | **4** | **14** | **2** | **24** |
| Block 3 | SOV | 2 | 10 | 2 | - | 14 |
| | OV | 2 | 4 | 2 | - | 8 |
| | SV | - | - | - | 2 | 2 |
| | **Tot** | **4** | **14** | **4** | **2** | **24** |
| Block 4 | SOV | 10 | 2 | 2 | - | 14 |
| | OV | 4 | 2 | 2 | - | 8 |
| | SV | - | - | - | 2 | 2 |
| | **Tot** | **14** | **4** | **4** | **2** | **24** |
| Block 5 | SOV | 2 | 2 | 10 | - | 14 |
| | OV | 2 | 2 | 4 | - | 8 |
| | SV | - | - | - | 2 | 2 |
| | **Tot** | **4** | **4** | **14** | **2** | **24** |
| Block 6 | SOV | 2 | 10 | 2 | - | 14 |
| | OV | 2 | 4 | 2 | - | 8 |
| | SV | - | - | - | 2 | 2 |
| | **Tot** | **4** | **14** | **4** | **2** | **24** |
| | TOT | 44 | 44 | 44 | 12 | 144 |

Investigating the effect of input for L2 learning was beyond the scope of the present study. However, the methodological choice of skewing the distribution of verbs in the passive exposure set was taken based on evidence that a skewed distribution in the frequency of lexical verbs in the input has been shown to be beneficial for the learning of linguistic regularities such as novel word order patterns (cf. Goldberg et al., 2004 and references in Chapter 4). Unlike Pili-Moss (2017), where OV sentences were introduced half way through the language training (Block 3), here all word order types were introduced from the start and occurred with constant frequency across blocks.

Each block of stimuli was created using a power-point document subsequently saved in mp4 video format. Each video lasted about 4 minutes. The Suzy character introduced the task, appeared again in a break after 16 stimuli to encourage the participants to pay attention for a little longer, and finally at the end of the video to remind the children that the next activity would be playing with the computer game (a full Suzy's script is provided in Appendix A). At the start of each block, immediately after Suzy's introduction, three consecutive slides were added to provide a countdown (3, 2, 1), with the purpose of focusing the child's attention on the start of the first stimulus slide.

Each move was exemplified by one token (in the case of the intransitive construction with *klino*) or two tokens (all other moves). No distractor tokens were added to the constellations in the passive exposure. Eight sectors of 4 squares each were identified on the game board and the positions the moves appeared on the board were randomized across sectors (same sequence for all participants). Although the aural sentence stimuli and move position varied for each slide, variation in the context was limited by repeating the same move for 4 consecutive slides.

Suzy's instructions and BrocantoJ words were recorded by a female Italian native speaker using monotone intonation and digitized with Audacity version 2.1.0 at a rate of 44100 Hz. Bisyllabic words were standardized in audio files lasting on average 707 ms (*SD* = 80.5 ms), whilst the monosyllabic words lasted on average 538 ms (*SD* = 4.5 ms). Each sentence stimulus was created by concatenating individual word files using AudioJoiner. Fifty milliseconds of silence were also added at the beginning and at the end of each word track for bisyllabic words and at the end of each item for monosyllabic words. Consequently there were 100 ms of silence between any two words in the final concatenated sentence stimulus, except for the transition between nouns and case markers, where the silence amounted to 50 ms.

### 6.3.7 Game practice

The six game blocks used in the experiment were selected from a larger set of 18 blocks used in the comprehension phase of the game in a previous BROCANTO2 study (Grey, 2014; the original blocks were numbered 1, 2, 5, 9, 10 and 13). As the paradigm had never been used with children before, the rationale behind the selection of blocks in the piloting phase was to simplify the game preferring blocks that (a) had a lower number of distractor tokens per game constellation, and (b) had the lowest possible number of moves were explicit use of adverbs was required to univocally identify the move. Using the same game configurations (which could not be modified), new BrocantoJ sentence stimuli were created for the six blocks and uploaded to the program's XML files.

In total the game practice set consisted of 120 BrocantoJ stimuli (20 stimuli per block). Block 1 was administered in Session 1, Block 2 and 3 were administered in Session 2 and Blocks 4, 5 and 6 were administered in Session 3 (cf., Figure 6.1). The order of presentation of the sentence stimuli was the same for all participants.

Overall, sentence length varied between 3 and 8 words. Each block included 12 SOV sentences, 6 OV sentences and 2 SV sentences. To facilitate stimulus retention in short-term memory, the choice was made to keep the aural sentence stimulus as short as possible so that each sentence included only the vocabulary items necessary to univocally identify the correct move, at the expense of perfect item counter-balancing for adjectives, adverbs and token names in the game practice set (Table 6.4a; see Appendix C for the complete set). Again, the practice exposure set contained a larger number of accusative markers (*ru*) compared to nominative markers (*ri*), given that the number of OV sentences in the set exceeded the number of SV sentences.

Table 6.4a

*Frequency of Vocabulary Items in the Game Practice Set*

| Vocabulary items | Category | Frequency |
|---|---|---|
| blomi | Noun | 44 |
| nipo | Noun | 52 |
| pleca | Noun | 45 |
| vode | Noun | 51 |
| trose | Adjective | 58 |
| neimo | Adjective | 51 |
| klino | Verb | 12 |
| nima | Verb | 36 |
| yabe | Verb | 36 |
| prazi | Verb | 36 |
| noika | Adverb | 19 |
| zeima | Adverb | 18 |
| ri | Preposition | 84 |
| ru | Preposition | 108 |

The distribution per block relative to the type and number of sentence stimuli assigned to each verb is reported in Table 6.4b. Unlike the set reported in Table 6.3b, creating a skewed distribution in each of the blocks was not possible in this case due to the fact that the sequence of game constellations in individual game blocks could not be modified.

Table 6.4b

*Frequency of Verbs per Sentence Type in the Game Practice Set*

| | | symmetric | | asymmetric | | |
|---|---|---|---|---|---|---|
| | | **nima** | **prazi** | **yabe** | **klino** | TOT |
| Block 1 | SOV | 4 | 4 | 4 | - | 12 |
| | OV | 2 | 2 | 2 | - | 6 |
| | SV | - | - | - | 2 | 2 |
| | **Tot** | **6** | **6** | **6** | **2** | **20** |
| Block 2 | SOV | 4 | 4 | 4 | - | 12 |
| | OV | 2 | 2 | 2 | - | 6 |
| | SV | - | - | - | 2 | 2 |
| | **Tot** | **6** | **6** | **6** | **2** | **20** |
| Block 3 | SOV | 4 | 4 | 4 | - | 12 |
| | OV | 2 | 2 | 2 | - | 6 |
| | SV | - | - | - | 2 | 2 |
| | **Tot** | **6** | **6** | **6** | **2** | **20** |
| Block 4 | SOV | 3 | 6 | 3 | - | 12 |
| | OV | 3 | - | 3 | - | 6 |
| | SV | - | - | - | 2 | 2 |
| | **Tot** | **6** | **6** | **6** | **2** | **20** |
| Block 5 | SOV | 5 | 4 | 3 | - | 12 |
| | OV | 1 | 2 | 3 | - | 6 |
| | SV | - | - | - | 2 | 2 |
| | **Tot** | **6** | **6** | **6** | **2** | **20** |
| Block 6 | SOV | 4 | 5 | 3 | - | 12 |
| | OV | 2 | 1 | 3 | - | 6 |
| | SV | - | - | - | 2 | 2 |
| | **Tot** | **6** | **6** | **6** | **2** | **20** |
| | TOT | 36 | 36 | 36 | 12 | 120 |

The word tracks for the sentence stimuli were the same as those deployed for the passive exposure set, but an additional 100 ms of silence was added between any two words in the final sentence stimulus, so that the speech rate was slower in the game practice than in the passive exposure. Prior to the start of the first game block the children were given the instruction to listen well to the words in the new language, and then make the move they thought the words were describing as fast as they could. After making their move, they were immediately given feedback on screen in the form of the words 'correct' or 'incorrect' (but were not shown the correct move in case of error). The next stimulus was presented immediately afterwards or after 60 seconds in case of no response. A percentage correct score appeared on screen at the end of each block (corresponding to an increase of 5% for each correctly performed move, with the score starting at 0%). The participants were told that, initially, making many points was not easy, but that if they kept listening carefully they would become better and better at the game.

***Online Measure of Language Learning (Overall Comprehension).*** During the computer game and unbeknownst to the participants the computer program created a by-trial online record of their moves and running score. As accurate performance in the game depended on the comprehension of BrocantoJ sentence stimuli, the running score provided an online overall measure of accuracy of language comprehension at item level.

***Online Measure of Thematic Linking.*** In symmetric contexts (sentences with *nima* and *prazi*) the initial constellation does not provide cues to predict which of the two tokens will move/move first and accuracy exclusively depends on the ability to assign the correct thematic interpretation to the nouns (cf. 5.2.3). For this reason, accuracy in the subset of stimuli relative to symmetric contexts can be used to measure the learning of thematic linking. It is important to recall that although the analysis of SOV sentences in itself can provide evidence of the learning of linking rules, this is not sufficient to allow a differentiation between syntactic and morphological linking. This is because in SOV contexts the information about thematic linking can be conveyed by the word order (which is fixed) and/or by the case markers. On the other hand, OV sentences provide the possibility to isolate morphological linking, because in these sentences correct thematic linking exclusively depends on the accurate interpretation of the accusative case marker.

### 6.3.8 Grammaticality judgment test procedure

The grammaticality judgment test (GJT) was developed in E-Prime (version 2.0.10.356) and administered on an ASUS X553M laptop computer. It comprised a total of 28 novel test sentences (14 ungrammatical) and 4 practice sentences (2 ungrammatical). The test started with the practice block followed by 3 experimental blocks (10, 10 and 8 items respectively), with the possibility for the child to take short self-managed breaks at the end of each block, and ask further questions immediately after the practice items. The practice sentences were trials administered in the same modality as the experimental trials and included 2 grammatical and 2 associated ungrammatical sentences (the complete set is reported in Appendix D). Although detailed instructions were provided to every participant in advance, the aim of the practice trials was to familiarize the participants with the task. This included for example practising the sequence of events in the trial, experiencing the timing of each event and the two different types of judgment that were required together with the associated scales (i.e. GJT judgment and confidence rating). At the end of the four practice trials a screen appeared displaying the text: *Adesso puoi chiedere alla maestra Diana. Altrimenti clicca la barra verde per continuare* [Now you can ask Miss Diana. Otherwise click the green bar to continue]. At this point, if the participant required it, additional clarification information relative to the task instructions was provided. After that, the participants proceeded independently. The order of the practice items was the same for each participant but the order of the experimental blocks, as well as the order of items in each experimental block, was randomized across participants.

The trial started with a fixation cross (3 seconds) after which a sound icon appeared on screen with the associated aural sentence stimulus. Immediately after the aural stimulus, the text *Com'è?* [How is it?] appeared on screen, together with a yellow arrow pointing down at the top-right part of the keyboard where six aligned keys were used to select the judgment response (corresponding to the keys from '7' to '=' on a British Microsoft keyboard). The keys were labelled with yellow stickers depicting six different smileys ranging from very unhappy to very happy and had no further numerical or text indication (Figure 6.4). A six-point scale in the grammaticality judgment was used because it provided the opportunity to code the judgment in a binary way, but at the same time also to assess it as a fine-grained graded judgment (cf. discussion in 2.4.3). Graded judgments deploying multiple-point scales have been considered suitable for use with adults and children aged 4 and upwards (Ambridge & Rowland, 2013; see also Ambridge et al., 2006, who deployed a five-point grammaticality judgment scale with children as young as 5).

After the child had pressed one of the smiley keys, or after 7 seconds, the text *Ti senti sicuro?* [Do you feel sure?] appeared on screen with a picture of a light blue arrow. This time the arrow pointed sideways to a set of four keys on the top-left side of the keyboard (corresponding to the keys from '1' to '4' on a British Microsoft keyboard). The four keys were labelled with light-blue stickers and displayed the writing 'sì molto' [yes very], 'sì' [yes], 'così così' [so so], 'per niente' [not at all]. After pressing one of the blue keys, or after 7 seconds, the next trial started. Trials for which no grammaticality judgments were provided (even if the corresponding confidence ratings were provided) were later excluded from the analysis.

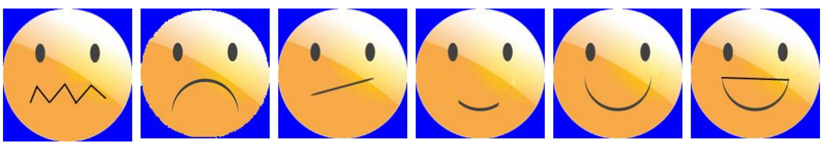

*Figure 6.4*. Keyboard Choices for the GJT.

By allowing a selection of one out of the four options, the confidence rating procedure was designed to rate the participants' confidence in their GJT judgment immediately after this had been given. In practical terms, with the confidence rating the participants confirmed the degree to which they would have picked the same smiley face again, had they been asked to repeat the GJT judgment on the immediately preceding sentence; with very high confidence ratings (corresponding to very sure) they maximally confirmed the judgment they had just given, whilst with very low confidence ratings (corresponding to not at all sure) they signalled a maximally high possibility they would give a different GJT rating given a second chance.

Note that, since it measures confidence, the confidence rating is independent of the specific judgment given in the GJT and applies regardless of it. For example, if the participant judged a GJT sentence to be very good (very happy smiley), the confidence judgment would be about how sure they were the sentence was indeed very good. Similarly, if for the same stimulus they had given a positive judgment but more towards the middle of the scale, the confidence rating would be about how sure they were the sentence was good but not perfect, etc.

In the GJT instructions direct reference to metalinguistic concepts like (un)grammaticality or grammatical acceptability was avoided. The children were told that, having completed all six game blocks, they were now experts in the new language. They were also told that Suzy wanted to create a new game block and had some sentences but needed their expert advice to decide which ones to choose. They could give their ratings of the sentences' suitability using the smiley scale based on how similar the sentences sounded to the ones they heard in the training videos and in the game and pressing the key immediately when they saw the yellow arrow on screen.

It was considered particularly important to ensure that the children understood the difference between sentence rating and confidence rating. In order to clarify the difference between knowledge and higher-order thoughts about knowledge, the children were asked to consider the familiar classroom situation in which a teacher asks a question in class. In such situations, sometimes they would be absolutely sure a certain answer was right. In other cases they would be quite confident but not as much, and sometimes they would not know the answer but try guessing anyway. A very similar situation would happen in the task. After picking a smiley face to help Suzy choose good sentences for the game, they would have to say how sure they felt their choice was correct.

In terms of the composition of the GJT set, vocabulary items, including case markers, were counter-balanced across word categories (Table 6.5a) and the GJT experimental stimuli included 16 SOV sentences, 8 OV sentences and 4 SV sentences (Table 6.5b). The practice stimuli were entirely comprised of SV sentences (two ungrammatical and the two corresponding grammatical). In the game dataset (and as a consequence in the passive exposure and GJT datasets) SV sentences could not be completely excluded due to the impossibility to omit intransitive moves (moves where a single token was involved) from the game constellations. At the same time the number of SV moves was very limited compared to SOV and OV transitive moves. Due to the limited number of GJT items overall, the decision was made to use SV sentences for the practice trials rather than reducing the number of SOV and OV stimuli. Further, since there was an expectation that, overall, word order violations would have shown a greater learning effect compared to case violations, it was decided to include the two SV sentences with word order violations in the GJT set and use the two SV sentences with case violations as practice items.

Table 6.5a

*Frequency of Vocabulary Items in the GJT Set*

| **Vocabulary items** | **Category** | **Frequency** |
|---|---|---|
| blomi | Noun | 12 |
| nipo | Noun | 12 |
| pleca | Noun | 12 |
| vode | Noun | 12 |
| trose | Adjective | 9 |
| neimo | Adjective | 9 |
| klino | Verb | 8 |
| nima | Verb | 8 |
| yabe | Verb | 8 |
| prazi | Verb | 8 |
| noika | Adverb | 8 |
| zeima | Adverb | 8 |
| ri | Preposition | 25 |
| ru | Preposition | 25 |

Table 6.5b

*Frequency of Verbs per Sentence Type in the GJT Set*

| | **nima** | **prazi** | **yabe** | **klino** | TOT |
|---|---|---|---|---|---|
| SOV | 2 | 8 | 6 | - | 16 |
| OV | 6 | - | 2 | - | 8 |
| SV | - | - | - | 8 | 8 |
| **Tot** | **8** | **8** | **8** | **8*** | **32** |

*Note*. *half of these sentences were used as practice

All sentence stimuli (practice and experimental) contained an equal number of words (5; corresponding to 8 or 9 syllables in total). In the ill-formed sentences the ungrammaticality was never triggered by the first word in the sentence. The 14 ungrammatical sentences matched the corresponding grammatical ones and were created inserting violations of case assignment (6 sentences) and word order (8 sentences). Case violations included sentences where the nominative or the accusative case markers were missing (2 sentences) and cases in which the wrong marker was used (4 sentences). Word order violations included ungrammatical order at sentence level (5 sentences) and inside the NP (3 sentences). Appendix D includes the complete GJT set with each sentence labelled for type of word order and, for ungrammatical sentences, type of violation.

***GJT scores.*** In this paradigm the grammaticality judgment test is an offline measure of language learning based on the judgment of aural sentence stimuli as 'good' or 'bad' compared to the ones presented in the language training and practice. Unlike the online measures of learning, the GJT was administered presenting the stimuli outside the game context. Hence it mainly probed morphosyntactic learning independently of its semantic dimension.

Judgments on aural sentence stimuli were given on a six-point-scale (three grades for 'good' and three grades for 'bad', cf. Figure 6.4). For each stimulus this potentially allowed for both binary scoring and graded scoring. As ungrammatical sentences were created to violate specific grammar rules, subsets of the test stimuli could be used to assess learning of word order and case marking.

***Confidence in the Accuracy of the GJT Response.*** The four-graded subjective qualitative rating of confidence in the accuracy of the GJT responses was turned into a four-point numeric scale (with the highest point in the scale corresponding to maximal confidence). Similarly to the GJT this allowed both binary scoring and graded scoring (2 grades for high and low-confidence items respectively, or 4 grades overall).

### 6.3.9 Debriefing questionnaire procedure

The researcher completed a short debriefing questionnaire interviewing the children individually immediately after the GJT was administered. The questionnaire included four open questions (Appendix E) and had the aim of eliciting evidence of the child's explicit knowledge of the artificial language morphosyntax and form-meaning relationships (cf. also 6.3.11).

***Explicit Language Knowledge (Word Order).*** All responses to the debriefing questionnaires were coded by the researcher and by another Italian native speaker trained in linguistics at postgraduate level. The interrater agreement was 90%. The remaining cases were discussed until a consensus was reached on how to code them. Each questionnaire was assigned a language awareness score as follows: (1) reports to have noticed nothing in particular; (2) reports noticing the presence/absence of specific words; (3) reports noticing the presence/absence of a single specific word and refers to its position in the sentence; (4) reports that there is an order involving domains larger than a single word but does not provide examples; (5) reports that there is an order in domains larger than a single word and provides examples; (6) provides a complete example of the sentence word order in the new language (see Appendix E for the questionnaire and a selection of example responses).

The awareness scale contains more categories than a previous three-level model used e.g. in Rosa and Leow (2004) that included no report, awareness at the level of noticing, and awareness at the level of understanding. However it can be reduced to it considering that level (1) would corresponds to no report, level (2) would correspond to awareness at the level of word noticing, and levels (3) to (6) would correspond to awareness at the level of understanding, where scores reflect awareness of word order regularities in increasingly larger syntactic domains.

### 6.3.10 Cognitive tasks materials and procedure

***Visual Declarative Learning Ability (Visual-Spatial Associations).*** The materials for the administration of all declarative memory tasks and the measure of the phonological loop were part of the PROMEA battery (Vicari, 2007). For visual declarative memory they included a set of A4 full-size color pictures of 16 familiar objects, 16 pictures of the same objects occupying a random position in an A4 four-spaced grid, one A4 grid with no pictures, and a booklet to record the child's scores. The task probed the retention of visual-spatial associations immediately and after a delay, and was administered individually according to the battery's manual. It started by showing the children a practice stimulus and explaining that they would have to memorize the picture position in the grid and would be asked to show the position using the empty grid immediately afterwards. After the purpose of the task was clarified with the practice stimulus, a full sequence of 15 A4 grids with pictures was shown, 5 seconds per stimulus.

Subsequently, the children were given the empty grid they would use to indicate the positions and shown the full set of 15 full-size pictures one by one. The children were assigned one point for each picture position in the grid they were able to recollect correctly. No feedback was given on the correctness of single picture-position pairings but the researcher told the children how many matches they got wrong at the end of each series. The whole process was repeated three times. Unbeknownst to the participants, a final delayed recall test was performed 15 minutes after the third immediate test.

***Verbal Declarative Learning Ability (Short Story).*** The materials included a short story of 58 words divided into 28 information units and a form where the version of the story provided by the child in the recall trials could be noted down. The children were asked to listen to a short story and told they would have to repeat it as precisely as they could immediately afterwards. As the children recounted the story the researcher ticked the information units that were remembered and noted down any different words used or changes in the order of the events. Unbeknownst to the participants, they were asked to recall the story a second time after 15 minutes and their performance was assessed in the same way.

***Declarative Learning Ability: Task scoring.*** For declarative learning ability three scores were obtained: (a) visual-spatial declarative memory (from the visual-spatial association task), (b) verbal declarative memory (from the short story task), and (c) a global score of declarative memory. All scores were obtained applying the formulas provided in Vicari (2007) and using the norms provided, i.e. selecting the values corresponding to the raw scores adjusted for the child's age (from 7 years 0 months to 7 years 11 months; from 8 years 0 months to 8 years 11 months; and from 9 years 0 months to 9 years 11 months). The use of the formulas for calculating visual and verbal declarative memory was necessary for standardization in accordance with the manual of the tests.

For visual-spatial declarative memory, given that *TPo* was the sum of positions recalled in the three trials and *Rec1* was the number of position recalled in the delayed trial, the formula applied was (Vicari, 2007, p. 37):

s1 = (adjusted TPo + adjusted Rec1) / 2

visdecl score = (s1 * 5) + 50

For verbal declarative memory the score assigned one point for each information unit accurately recalled in the immediate recall of the story, repeating the procedure for the delayed recall. Accurate recall included when synonymous words were used but excluded recalling accurate information units in the wrong logical order. Given that *Imm* was the number of information units recalled in the first trial and *Rec2* the number of information units recalled in the delayed trial, the formula was:

s2 = adjusted [(Imm + Rec2) / 2]

verbdecl score = (s2 * 5) + 50

Finally, the global measure of declarative memory based on the available measures was calculated as follows:

s3 = (adjusted TPo + adjusted Rec1 + s2) / 3

totdecl = (s3 * 5) + 50

***Phonological Loop (Nonword Repetition).*** The materials included a list of 40 nonwords of different length in a random order: 10 words with two syllables, 10 words with three syllables, 10 words with four syllables, and 10 words with five syllables. Half of each set of words has high phonological similarities and half low phonological similarity with Italian words. For all words the stress falls on the penultimate syllable (the most frequent stress pattern in Italian). The task was administered by the researcher according to the manual, reading the words to the child one at a time whilst holding a light sheet of paper in front of her mouth so that lip movements would not provide visual cues for pronunciation.

***Phonological Loop scores.*** Standardized *z*-scores were also calculated for the nonword repetition task providing a measure of the phonological loop. In the nonword repetition task a raw score was obtained assigning a point for every word repeated correctly, and zero points if any pronunciation errors were made on an item (inaccurate phonemic recall). and the final measure of phonological memory was the standardized raw score thus obtained. According to the administration manual, correct pronunciation was subjectively assessed by the researcher, an Italian native speaker with no diagnosed hearing impairment, who tested all children. The assessment of the accuracy of the non-word repetition task was done immediately.

***Procedural Learning Ability (Alternate Serial Reaction Time Task)***. In a serial reaction time task participants are usually asked to immediately react to the changes in position of an on-screen target by pressing the corresponding buttons on a controller/input box. The alternating serial reaction time task (ASRT) was obtained modifying a version of an E-Prime serial reaction time task (SRT) originally developed in Lum (2010). The main difference between the ARST and the SRT involves the type of sequence the participants are expected to learn as a result of the exposure to the stimuli. In the SRT each item in a sequence of stimuli in nonrandom blocks belongs to the pattern sequence, whilst in the ASRT the pattern sequence is 'concealed' by an alternation of random and pattern positions. The ASRT presents important advantages compared with the SRT. First of all, it provides a continuous measure of learning, whereas learning in the SRT is assessed only once at the end of the task comparing reaction times on a random block to reaction times on a sequence block. Secondly, as the training sequence in the ASRT includes both pattern and random stimuli, it provides the opportunity to differentiate between sequence learning and general motor skill learning without the need to take additional motor skill measures. Finally, as a number of studies have consistently shown, sequences in ASRTs are unlikely to be learnt explicitly, as the pattern sequence is usually not reported and/or identified even after extended practice (Hedenius, 2013; Howard et al., 2004).

The task was administered with an ASUS X553M laptop computer and headphones. The children inputted their responses using a game controller that could be configured for both right-handed and left-handed use (iBUFFALO$_{TM}$ Classic USB

Gamepad BSGP801). The use of the game controller and the task interface were aimed at creating an involving and child-friendly gaming environment (Lum, 2010). The task consisted in pressing one of four buttons in the game controller corresponding to the position a smiley would appear in on screen (Figure 6.5).

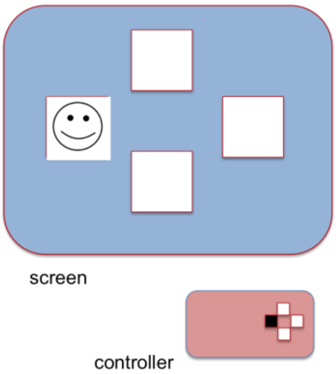


*Figure 6.5.* Alternate Serial Reaction Time Task.

To further discourage the use of explicit learning strategies (and unlike previous version of the ASRT task, e.g., Hedenius, 2013), the task deployed here did not provide indirect feedback on incorrect trials by blocking the transition to the next stimulus until the correct button was pressed. Rather, independently of whether the response to a given stimulus was correct, the next stimulus was immediately presented.

The ASRT task began with on-screen instructions and a series of training trials designed to familiarize the player with the task and the controller. After that there were 8 experimental blocks (80 trials each), with the possibility for the participant of taking brief self-managed breaks between blocks. The blocks and the trials in each block were administered to all participants in the same order, and each block was preceded by 5 warm-up trials. The last screen in each block gave the participants feedback about their accuracy (percentage correct) and about the speed of their performance in the block. In the experimental trials the sequence alternated a fixed position with a random position according to the pattern 1 r 2 r 4 r 3 r (Hedenius, 2013).

The children were told to play the game trying to press the correct controller button as fast as they could. They were also told to check the scores at the end of the block and try to keep an accuracy score around 92% (Howard et al., 2004), whilst at the same time trying to improve their speed score as the game progressed.

***Procedural Learning Ability scores.*** Procedural learning ability was operationalized as learning of the fixed pattern in the alternating stimuli sequence the participants were exposed to. The ASRT task allowed to obtain three measures of procedural learning ability. The first measure was based on the reaction times, the second on (in)accuracy, and the third was a composite of the previous two. For the RT measure, pattern learning was operationalized as the finding of increasingly shorter RTs for pattern trials compared to random trials as training progressed. For the (in)accuracy measure, pattern learning was operationalized as an increasing number of errors on random trials compared to pattern trials as training progressed. The ASRT data were first reduced in Excel excluding practice trials, warm-up trials, incorrect trials, and trials that were the final element in 'trills' (e.g., 212) or in 'repetitions' (e.g., 222), as recommended in Hedenius (2013) and Howard et al. (2004).

For the RT-based measure, the median RT values in milliseconds were calculated separately for sequence trials and nonsequence trials for each participant's block (Bennet at al., 2011, Hedenius, 2013). In both data subsets the scores from Block 1 to Block 4 and the scores from Block5 to Block 8 were averaged obtaining an A and a B score respectively. The difference between A and B (RT Gain) reflected the change in reaction times from the first half to the second half of the training. To obtain a final measure of procedural learning ability based on RTs, RT Gains from nonsequence trials were subtracted from RT Gains from sequence trials, with higher positive differences indicating better sequence learning.

In the case of the (in)accuracy measure the same reduced dataset was used, but this time it also included incorrect items. To obtain a measure of procedural learning ability based on (in)accuracy, the number of inaccurate responses (errors) was calculated for each participant's block and then averaged across blocks (Bennet at al., 2011, Hedenius, 2013). For each participant the difference between the average number of errors in nonsequence trials and the average number of errors in sequence trials provided a measure of sequence learning, with larger positive differences indicating better sequence learning. Finally, a composite measure of procedural learning ability (Proc) was obtained standardizing and then averaging the two components.

### 6.3.11 Statistical analysis

The main inferential statistics deployed were binomial (logit) linear mixed effects models (*glmer* with maximum likelihood, Laplace approximation, from the *lme4* package; Bates, Maechler, & Bolker, 2011) in the *R* environment (R Development Core Team, 2016). When used to model accuracy data, logit mixed effects models do not average across participants or trials and estimate the likelihood that individual trials are accurate measured in the log odds scale. The effect of a predictor is the variation in the log odds[2] that a trial is correct for each unit increase in the predictor. Two further important advantages of mixed effects models are that they

---

[2] When the words 'likelihood' or 'odds' appear in the presentation of the data analysis in the following pages, these expressions should be interpreted to mean 'log likelihood' and 'log odds'.

are robust against missing data and that they allow for the computation of random effects, accounting for portions of variance that are due to random differences (for example random differences between participants or items) and hence returning more precise estimates for the fixed effects compared to techniques where random effects are not controlled.

The procedure adopted to fit the models determined main effects and interactions first, and subsequently investigated random effects. The possibility of inclusion in the equation was explored for all variables that could potentially modulate the effect of the predictors of interest and for their interactions. Fixed effects were added one at a time comparing nested models using the likelihood ratio test and the Akaike Information Criterion (AIC), and effects were included in the model only if they statistically significantly improved the model's fit. Once the structure of fixed effects was determined, random effects were explored starting with random effects on intercepts and subsequently considering random effects on the slopes of the main fixed effects. The analysis tended to fit a maximal random effect structure (Barr, Levy, Scheepers, & Tily, 2013) as long as the model's complexity was supported by the data, i.e. the model converged and the random effect improved the model's fit compared to a less complex nested model (likelihood ratio test and AIC). The condition number (CN) was calculated prior to analysis, and a CN lower than 15 was taken to indicate a low risk of multicollinearity (Davies, 2017). The variables Session and Year were centered and all remaining continuous variables were standardized.

## 6.4 Results

This section reports the results of the descriptive and inferential statistics relative to research questions 1, 2 and 3 in the child study. A detailed discussion of the results including the interpretation of the main effects and interactions in the inferential analysis in the light of the theoretical framework introduced in chapters 1 and 2 can be found in section 8.1.1 (RQ1), 8.1.2 (RQ2) and 8.1.3 (RQ3).

### 6.4.1 RQ 1

RQ1 To what extent do declarative and procedural learning ability modulate child L2 aural comprehension and learning of the rules linking morphosyntax and thematic interpretation during practice?

***Descriptive Statistics.*** In Morgan-Short (2007) the chance score in the computer game was calculated relative to one fourth of all comprehension blocks "as the ratio of all possible correct moves to all possible moves that could be made by a game token that was named by the subject of the sentence" (p. 143). Following Morgan-Short et al. (2007, p.143) and Pili-Moss (2017) the chance level for accuracy during practice in this study was thus set at 14% correct trials per block. Block scores were computed for each participant, and participants' data were further analyzed only if their scores in the game included at least 3 blocks above chance, or at least one block significantly above chance. The analysis included data from 40 participants. Table 6.6a reports overall and by-block accurate performance.

Table 6.6a

*Children Mean Accurate Performance During Practice (Percentage)*

| | *M* (*SD*) | *S.E.* |
|---|---|---|
| Block 1 | 17.2 (10.2) | 1.6 |
| Block 2 | 22.4 (10.4) | 1.6 |
| Block 3 | 24.0 (15.4) | 2.4 |
| Block 4 | 26.6 (15.0) | 2.4 |
| Block 5 | 30.6 (16.3) | 2.6 |
| Block 6 | 36.9 (15.6) | 2.5 |
| Overall | 26.3 (10.3) | 1.6 |

Subsequently, accuracy and inaccuracy counts were aggregated to verify whether learning was significantly above chance at group level for different trial types (Table 6.6b). As the data show, performance was significantly above chance in all subcategories considered, independently of word order and symmetry in the context. However, the differences in effect size (Cramer-V) indicate that learning was more robust in asymmetric trials, where effects are small to medium sized, compared to symmetric trials where effect sizes are negligible to small. At descriptive level no evident differences emerge between word order types, if the symmetric vs. asymmetric distinction is not specified. Means and standard deviations were also calculated for the raw scores of the cognitive individual differences (Table 6.7).

Table 6.6b

*Aggregate Count of Valid Accurate and Inaccurate Trials in the Game (N = 40) and Accuracy Significance Above Chance (14%)*

| | Acc | % | Inacc | Significance above chance | | |
|---|---|---|---|---|---|---|
| | | | | $\chi^2$ | *p* | $\Phi_v$ |
| **Overall** | **1249** | **26.8** | **3418** | **236.33** | **.000** | **.159** |
| **SV** | **128** | **26.9** | **347** | **24.01** | **.000** | **.159** |
| **SOV** | **745** | **26.3** | **2093** | **132.76** | **.000** | **.153** |
| **OV** | **376** | **26.5** | **1041** | **69.22** | **.000** | **.156** |
| **SOVs** | **436** | **21.7** | **1572** | **40.79** | **.000** | **.101** |
| **SOVa** | **309** | **37.2** | **521** | **117.80** | **.000** | **.266** |
| **OVs** | **147** | **17.8** | **679** | **4.35** | **.037** | **.051** |
| **OVa** | **229** | **38.7** | **362** | **92.82** | **.000** | **.280** |
| **symm** | **583** | **20.6** | **2251** | **42.68** | **.000** | **.087** |
| **asymm** | **666** | **35.1** | **1230** | **228.92** | **.000** | **.246** |

*Note*. SOVs/OVs = symmetric SOV/OV trials; SOVa/OVa = asymmetric SOV/OV trials

Table 6.7

*Raw Score Means Relative to the Cognitive Individual Differences*

| | *M* (*SD*) | *SE* |
|---|---|---|
| Decl (visual) | 109.71 (11.9) | 1.94 |
| Decl (verbal) | 114.84 (11.3) | 1.83 |
| Decl | 112.34 (8.5) | 1.38 |
| Proc (Acc) | 2.40 (1.5) | 0.25 |
| Proc (RT) | -5.63 (23.5) | 3.92 |
| Phonological loop | 97.43 (11.6) | 1.90 |

***Model A.*** The fixed effects in the model (Condition Number [CN] = 1.38; Table 6.8) included the two cognitive variables of interest (Decl and Proc), vocabulary learning ability (VocLearn), the categorical variable word order type (WO), year group (Year) and game block (Block). The WO categorical variable was a factor with 5 levels, one for each of the word-order types that could occur in the game trials: (a) SV (which was also the reference category in the model presented in Table 6.8), (b) symmetric SOV, (c) asymmetric SOV, (d) symmetric OV, and (e) asymmetric OV. Interactions that significantly improved the model's fit included a Decl:WO interaction, a Proc:Block interaction, a VocLearn:WO interaction, and a VocLearn:Block interaction. No three-way interactions were found to be significant and/or to significantly improve the fit of the model.

The variables tested but excluded were Sex, PhonLoop (the auditory measure of working memory), the difference in days between Session 1 and Session 2, and the difference in days between Session 2 and Session 3. Two further variables, Age (in months) and Session, individually significantly contributed to the model fit, but not

Table 6.8

*Generalized Mixed Effects Model of the Effects of Declarative Learning Ability, Procedural Learning Ability, Word Order Type, Vocabulary Learning Ability, Year and Game Block on Accuracy During Training (Model A - Children) -* $R^2 = 0.31$

| | | | | 95% CI | | |
|---|---|---|---|---|---|---|
| Fixed effects | β | *SE* | *z* | lower | upper | *p* |
| **(Intercept)** | **-1.23** | **0.28** | **-4.40** | **-1.78** | **-0.68** | **.000*** ** |
| **Decl** | **0.28** | **0.14** | **1.96** | **+0.00** | **0.57** | **.049*** |
| Proc | -0.01 | 0.06 | -0.19 | -0.13 | 0.11 | .851 |
| **Block** | **0.23** | **0.06** | **3.99** | **0.12** | **0.35** | **.000*** ** |
| WO (SOVs) | -0.32 | 0.31 | -1.05 | -0.92 | 0.28 | .292 |
| **WO (SOVa)** | **0.69** | **0.34** | **2.05** | **0.03** | **1.35** | **.040*** |
| WO (OVs) | -0.47 | 0.35 | -1.36 | -1.15 | 0.21 | .174 |
| WO (OVa) | 0.54 | 0.36 | 1.48 | -0.17 | 1.25 | .138 |
| **VocLearn** | **0.52** | **0.16** | **3.26** | **0.21** | **0.83** | **.001**** |
| Year | 0.21 | 0.11 | 1.80 | -0.02 | 0.43 | .072^ |
| **Proc:Block** | **0.04** | **0.02** | **1.98** | **+0.00** | **0.08** | **.047*** |
| **VocLearn:Block** | **0.05** | **0.03** | **1.98** | **+0.00** | **0.11** | **.047*** |
| Decl:WO (SOVs) | -0.19 | 0.14 | -1.30 | -0.47 | 0.09 | .193 |
| Decl:WO (SOVa) | -0.17 | 0.15 | -1.13 | -0.48 | 0.13 | .257 |
| Decl:WO (OVs) | -0.21 | 0.17 | -1.28 | -0.54 | 0.11 | .199 |
| Decl:WO (OVa) | -0.14 | 0.17 | -0.86 | -0.47 | 0.18 | .389 |
| $VL^+$:WO (SOVs) | -0.13 | 0.16 | -0.78 | -0.44 | 0.19 | .432 |
| VL:WO (SOVa) | -0.27 | 0.17 | -1.59 | -0.60 | 0.06 | .111 |
| **VL:WO (OVs)** | **-0.43** | **0.18** | **-2.40** | **-0.78** | **-0.08** | **.016*** |
| VL:WO (OVa) | 0.08 | 0.19 | 0.42 | -0.29 | 0.44 | .674 |

*Note*. $^\wedge p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$; $^{+}$VL = VocLearn

when the, partially overlapping, Year and Block variables were also included in the model. As Year and Block provided a more robust contribution to the model's fit in comparison it was decided to keep these. It is worth noting that the effect of Year includes the effect of age, but may not be limited to it. For example it may encompass an effect of schooling (including the effect to have been exposed to a different number of hours of English instruction). Disentangling the two effects is beyond the scope of the present study and the question remains open to further enquiry.

The random effects included participants and trial items on intercepts as well as random effects of participants and items on the slopes of the Block variable. The analysis also found that including the children’s' class as a random effect would have not improved the model's fit.

***Main Effects (Model A).*** The coefficients reported in Table 6.8 are relative to SV, as this is the initial setting for the WO reference category. Confirming the descriptive data, the intercept's negative coefficient indicates that on average the participants were significantly more likely to be inaccurate than to be accurate in any given SV game trial ($p < .001$). Looking at the cognitive variables of interest, the analysis found a significant positive effect of Decl ($p < .05$).

Three further variables were found to have a significant effect on accuracy in addition to declarative learning ability: Block, VocLearn, and the word order pattern (WO). The positive effect of Block confirms that accuracy significantly increased as training progressed ($p < .001$). Similarly, the ability to retain and recognize picture/aural stimuli associations as measured by the vocabulary test at the beginning of each session was also found to strongly predict accuracy ($p < .001$) .

When interpreting the main-effect coefficients returned by the model it is important to consider whether the relevant variables also appear in higher-order interaction terms, as in this case the coefficient values are *conditional*. For example, in the case of Decl and VocLearn, two variables that also appear in a two-way interaction term with WO, the correct way to interpret the main-effect coefficient is that in both cases it is positive and significant relative to the WO reference category (in this version of the model, SV). In the case of interactions between continuous variables the relevant coefficients are conditional to the value of the other variable being equal to zero, which, for the purposes of this model, corresponds to the mean for standardized variables and to a point half-way through the scale for centered variables (Block and Year).

Turning to the discussion of the categorical variable WO, the model returned a positive significant effect for SOVa sentences ($p < .05$), indicating that they were learnt significantly better than SV sentences (Figure 6.6).

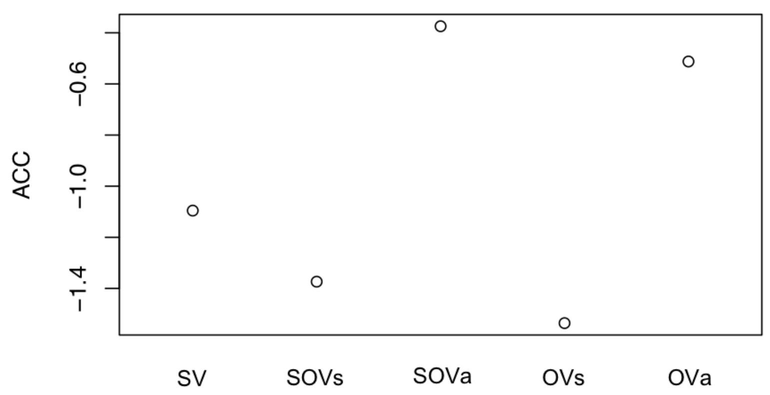


*Figure 6.6*. Effects of the Five Levels of the Word Order Variable (WO) on Accuracy (Intercepts - Log Odds Scale).

***Main Effects: Re-leveling of the Categorical Variable WO.*** In order to obtain a more complete picture, the WO reference category was changed (re-leveling), obtaining 4 further versions of the same model (Appendix F). For convenience, here and in any summary of re-leveled models thereafter, I will discuss only significant effects and omit effects that are not affected by re-leveling, and thus remain constant independently of the reference category.

In the second version of the model (Table 6.9a) SOVs trials were the baseline. This revealed a comparative significant advantage of both SOVa and OVa sentences ($p < .001$ and $p < .01$ respectively). As is to be expected, the new reference category

Table 6.9a

*Summary of WO, Decl and VocLearn Significant Main Effects Depending on WO Reference Category*

| | WO reference category | | | | |
|---|---|---|---|---|---|
| | SV | SOVs | SOVa | OVs | OVa |
| Intercept | -*** | -*** | -** | -*** | -** |
| SV | | | - * | | |
| SOVs | | | - *** | | - ** |
| SOVa | + * | + *** | | + *** | |
| OVs | | | - *** | | - ** |
| OVa | | + ** | | + ** | |
| Decl | + * | | | | |
| VocLearn | + ** | + *** | + * | | + *** |

*Note*. *$p < .05$; **$p < .01$; ***$p < .001$

also affected the conditional coefficients of the fixed effects that appear in interaction terms with WO in the model (Decl and VL). The VocLearn effect was still positive and statistically significant ($p <$ 001). When the baseline was SOVa, the model indicated that all other sentence types were significantly less accurate in comparison; SV ($p < .05$), SOVs ($p < .001$), and OVs ($p < .001$).

When OVs sentences were the baseline the model returned a significant positive learning advantage of SOVa and OVa ($p < .001$ and $p < .01$ respectively). Finally, the model with OVa sentences as baseline returned two significant negative effects indicating that there was a statistically significant lesser chance of accuracy both on SOVs or OVs sentences (both $p < .01$).

Summarizing across all models, it emerges that: (a) there is a statistically significant difference between symmetric and asymmetric contexts, i.e. for both SOV and OV sentences the participants were significantly more likely to be accurate if the stimulus did not require the processing of linking rules; (b) there are no significant differences in accuracy due to type of word order, beside the evidence that SOVa sentences were significantly more accurate than SV sentences; (c) VocLearn predicted accuracy in all sentence types, except in OVs sentences.

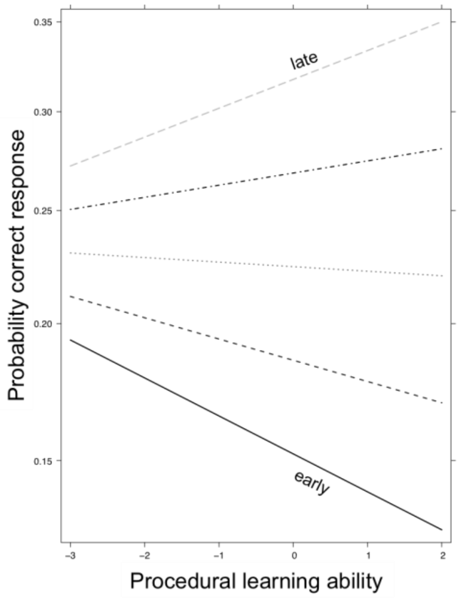


*Figure 6.7*. Development of the Effect of Proc on Accuracy Across Training.

***Two-way Interactions.*** The model in Table 6.8 (the one where SV was the baseline) returned a significant positive effect for the interaction between Proc and Block ($p < .05$; Figure 6.7, above). The positive interaction indicates that, as training progressed, the positive effect of Proc on accuracy increased statistically significantly. Similarly, the positive and significant VocLearn:Block interaction ($p < .05$) also indicates that overall, as training progressed, vocabulary learning ability increasingly supported accuracy. Based on the SV baseline, the VocLearn:OVs effect indicates that VocLearn supported accuracy in OVs trials significantly less than it did in SV trials.

***Interactions: Re-leveling of the Categorical Variable WO.*** The VocLearn:OVs interaction was significantly negative for WO = SOVs, showing that vocabulary learning ability supported accuracy significantly less if the target was OVs

Table 6.9b

*Summary of Significant VocLearn:WO Interaction Effects Depending on WO Reference Category*

| | WO reference category | | | | |
|---|---|---|---|---|---|
| | SV | SOVs | SOVa | OVs | OVa |
| VL$^{+}$:SV | | | | + * | |
| VL:SOVs | | | | + * | |
| VL:SOVa | | | | | - * |
| VL:OVs | - * | - * | | | - ** |
| VL:OVa | | | + * | +** | |

*Note*. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$, $^{+}$VL = VocLearn

compared to SOVs (Table 6.9b). For WO = SOVa, a positive significant effect was found for OVa, indicating that, as far as asymmetric contexts are concerned, accuracy in OV sentences benefitted more from higher vocabulary learning ability compared to SOV sentences (cf. Figure 6.8, next page).

In case WO = OVs, three significant positive effects were found for SV ($p < .05$), SOVs ($p < .05$), and OVa ($p < .01$), indicating that, compared to OVs, accurate responses in these three categories were better supported by vocabulary learning ability. Overall, the data show that asymmetric trials based on OV sentences were the ones with the better odds to be correct for higher levels of vocabulary learning ability (Figure 6.8). Accuracy in SV

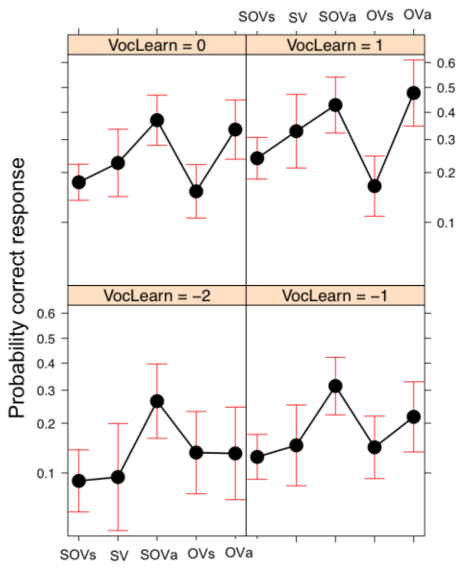


*Figure 6.8*. Interaction Between VocLearn (Measured in SDs) and WO.

and both SOV type sentences was also supported by vocabulary learning ability but to a lesser extent. Finally, OVs sentences were the category that benefitted the least from higher levels of vocabulary learning ability. The latter conclusion is also supported by the observation that the case when the reference category for WO was set to OVs, was the only one for which the main effect of vocabulary learning ability, though positive in sign, was not statistically significant (Table 6.9a; Appendix F).

***Model B: Symmetric Trials.*** In order to further analyze the relationship between cognitive individual differences and the learning of linking rules in the new language, a model that included only symmetric trials was investigated (Table 6.10a). A low condition number (CN = 1.38) indicated that multicollinearity was unlikely.

The most complex model justified by the data included fixed effects of Proc, VocLearn, and a three-way Decl:Year:Session interaction. The random effects included only participants and trial items on intercepts and there was no effect for Class.

Table 6.10a

*Generalized Mixed-Effects Model of the Effects of Declarative Learning Ability, Procedural Learning Ability, Vocabulary Learning, Year and Session on Accuracy During Training in Symmetric (Linking) Contexts (Model B)- $R^2 = .30$*

| | | | | 95% CI | | |
|---|---|---|---|---|---|---|
| Fixed effects | β | *SE* | *z* | lower | upper | *p* |
| **(Intercept)** | **-1.68** | **0.14** | **-11.84** | **-1.96** | **-1.40** | **.000*****|
| Decl | 0.07 | 0.09 | 0.75 | -0.11 | 0.25 | .451 |
| Proc | -0.01 | 0.08 | -0.14 | -0.16 | 0.14 | .890 |
| **Session** | **0.67** | **0.17** | **3.94** | **0.34** | **1.00** | **.000***** |
| **VocLearn** | **0.35** | **0.10** | **3.55** | **0.16** | **0.54** | **.000***** |
| Year | 0.23 | 0.14 | 1.62 | -0.05 | 0.50 | .105 |
| Decl:Year | -0.17 | 0.16 | -1.06 | -0.48 | 0.14 | .288 |
| Decl:Session | -0.01 | 0.08 | -0.10 | -0.17 | 0.15 | .921 |
| Year:Session | 0.08 | 0.12 | 0.68 | -0.16 | 0.33 | .498 |
| **Decl:Year:Session** | **-0.28** | **0.14** | **-2.02** | **-0.55** | **-0.01** | **.043*** |

*Note*. $^{*}p < .05$; $^{***}p < .001$

***Main Effects (Model B).*** The negative intercept indicates that overall, also for this reduced dataset, the participants were significantly more likely ($p < .001$) to be inaccurate than to be accurate in any given game trial. VocLearn was a significant predictor of accuracy ($p < .001$). The effect of Session was also highly significant ($p < .001$), indicating that the odds that a trial was correct significantly increased as training progressed.

***Interactions.*** The model returned a significant negative effect for the three-way Decl:Year:Session interaction ($p < .05$). To visually explore the interaction, the effect of Decl was plotted as a function of Year and Session (Figure 6.9).

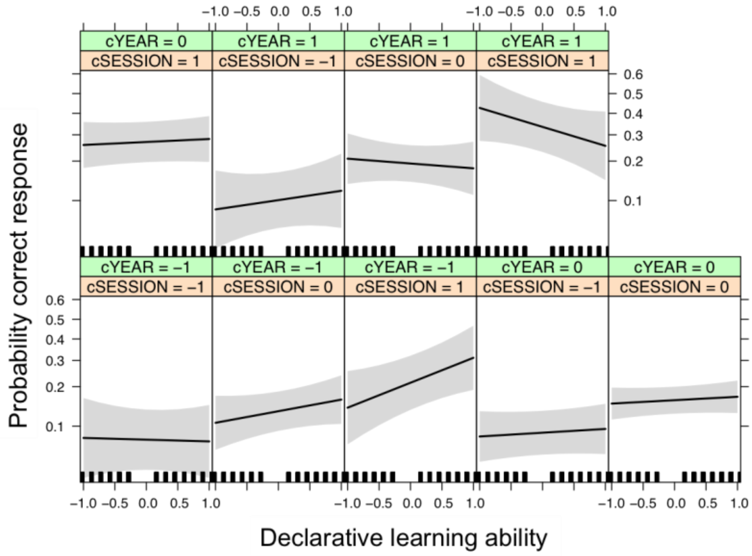


*Figure 6.9*. Three-Way Decl:Year:Session Interaction in Model B.

*Clarification note*: The best way to evaluate the pattern of the interaction is by reading the graph in groups of three panels left to right, starting from the bottom row and proceeding to the top row. The notation cYear/cSession indicates that Year/Session have been centered.

The graph clearly shows a difference between children in lower and higher grades. For younger children (cYear = -1) declarative learning ability has an increasingly positive effect as training progresses. For older children (cYear = 1) instead, the effect of declarative learning ability decreases during training. For intermediate values of Year (cYear = 0, recall that Year is treated as a continuous centered variable) Decl appears to have close to no effect independently of Session.

***Model C: Asymmetric Trials.*** A further model was fitted to a reduced dataset including only the asymmetric trials. The main effects included the two cognitive variables of interest (Decl and Proc), VocLearn, Year, and Block (both centered). The condition number (CN) was 1.38.

Two-way and three-way interactions statistically significantly improved the model's fit. These included a Decl:Year interaction, a Proc:Decl:VocLearn interaction, and a Proc:VocLearn:Block interaction. The only random effects justified by the data were those of participants and items on intercepts. A summary of the model is provided in Table 6.10b (next page).

***Main Effects (Model C).*** Model C shows a more complex structure of main effects and interactions compared with Model B. The negative significant intercept indicates that asymmetric trials were significantly more likely to be inaccurate rather than accurate ($p < .001$). The model returned a significant conditional positive effect of Decl ($p < .05$), Block ($p < .01$), and VocLearn ($p < .001$).

***Two-way Interactions.*** The model returned four significant two-way interactions. There was a significant negative effect of Proc:VocLearn ($p < .01$, accurate for Block = 0), and a significant positive effect of Decl:VocLearn ($p < .01$, accurate for average Proc), Proc:Block ($p < .05$, accurate for average VocLearn), and VocLearn:Block ($p < .001$, accurate for average Proc).

The negative Proc:VocLearn interaction (Figure 6.10) indicates that the effect of procedural learning ability on accuracy significantly decreased for increasing values of vocabulary learning ability (and vice versa). At the same time both procedural learning ability and vocabulary learning ability significantly increased as a function of practice, as indicated by the positive effect of Proc:Block and VocLearn:Block. Taken together these two sets of results indicate the existence of a competitive relationship between procedural learning ability and vocabulary learning ability, with both abilities increasingly supporting accuracy over the course of training.

Table 6.10b

*Generalized Mixed-Effects Model of the Effects of Declarative Learning Ability, Procedural Learning Ability, Vocabulary Learning, Year and Block on Accuracy During Training in Asymmetric Contexts (Model C) -* $R^2 = .28$

| | | | | 95% CI | | |
|---|---|---|---|---|---|---|
| Fixed effects | *β* | *SE* | *z* | lower | upper | *p* |
| **(Intercept)** | **-0.78** | **0.14** | **-5.43** | **-1.06** | **-0.50** | **.000*** ** |
| **Decl** | **0.17** | **0.07** | **2.21** | **0.02** | **0.31** | **.027*** |
| Proc | 0.01 | 0.08 | 0.18 | -0.14 | 0.17 | .860 |
| **Block** | **0.23** | **0.08** | **3.01** | **0.08** | **0.38** | **.003**** |
| **VocLearn** | **0.48** | **0.10** | **4.98** | **0.29** | **0.67** | **.000***** |
| Year | 0.18 | 0.12 | 1.53 | -0.05 | 0.42 | .125 |
| Decl:Year | 0.21 | 0.15 | 1.36 | -0.09 | 0.52 | .172 |
| Decl:Proc | -0.08 | 0.06 | -1.25 | -0.20 | 0.04 | .213 |
| **Proc:VocLearn** | **-0.36** | **0.12** | **-2.98** | **-0.59** | **-0.12** | **.003**** |
| **Decl:VocLearn** | **0.36** | **0.13** | **2.69** | **0.10** | **0.62** | **.007**** |
| **Proc:Block** | **0.07** | **0.03** | **2.24** | **0.01** | **0.14** | **.025*** |
| **VocLearn:Block** | **0.13** | **0.04** | **3.38** | **0.05** | **0.20** | **.001***** |
| **Decl:Proc:VocLearn** | **-0.39** | **0.14** | **-2.78** | **-0.67** | **-0.11** | **.005**** |
| **Proc:VocLearn:Block** | **0.05** | **0.02** | **1.97** | **+0.00** | **0.10** | **.048*** |

*Note*. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

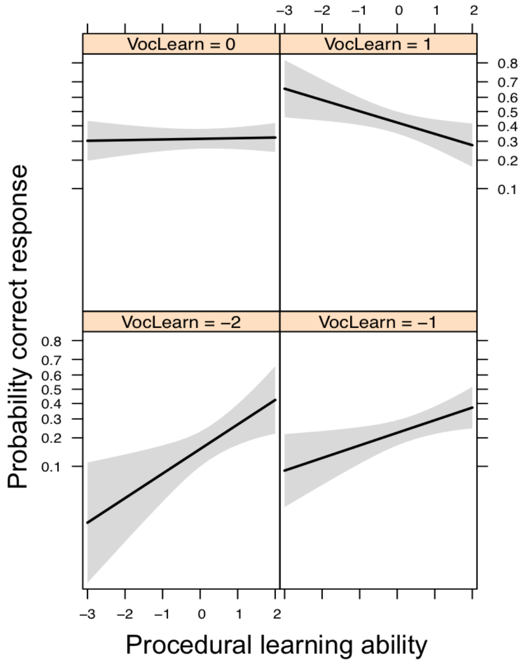


*Figure 6.10*. Effect of the Proc:VocLearn Interaction (Learning Abilities Measured in SDs).

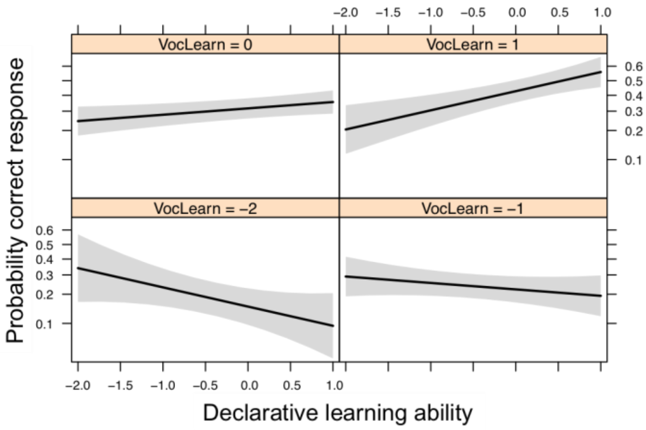


*Figure 6.11*. Effect of the Decl:VocLearn Interaction (Learning Abilities Measured in SDs).

Finally, the positive Decl:VocLearn interaction (Figure 6.11, previous page) indicates that the positive effect of declarative learning ability on accuracy grew for increasing values of vocabulary learning ability (and vice versa).

***Three-way Interactions.*** The model also returned a significant ($p < .01$) negative three-way Decl:Proc:VocLearn interaction. Plotting the data helps clarifying the relationship between the three continuous variables (Figure 6.12).

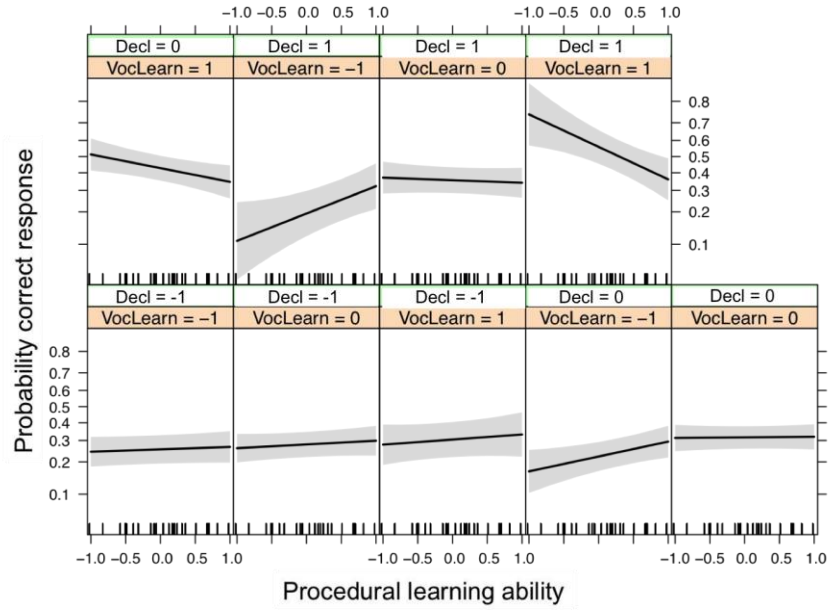


*Figure 6.12*. Effect of the Three-Way Interaction Between Proc, Decl and VocLearn.

For average levels of Decl, the graph in Figure 6.12 corresponds to the graph of the two-way Proc:VocLearn interaction discussed above (Figure 6.10). The additional information the three-way interaction provides is that the way higher levels of VocLearn mitigate the effect of Proc is a function of Decl. In particular, for below-average levels of Decl, the effect of higher VocLearn on Proc is close to zero (i.e., increasing levels of VocLearn do not change the effect of Proc on accuracy). However, for average and above average Decl the mitigating effect of higher VocLearn on the ability of Proc to predict accuracy significantly increases.

The second three-way interaction returned by the model was a significant ($p < .05$) positive interaction between Proc, VocLearn and (centered) Block (Figure 6.13).

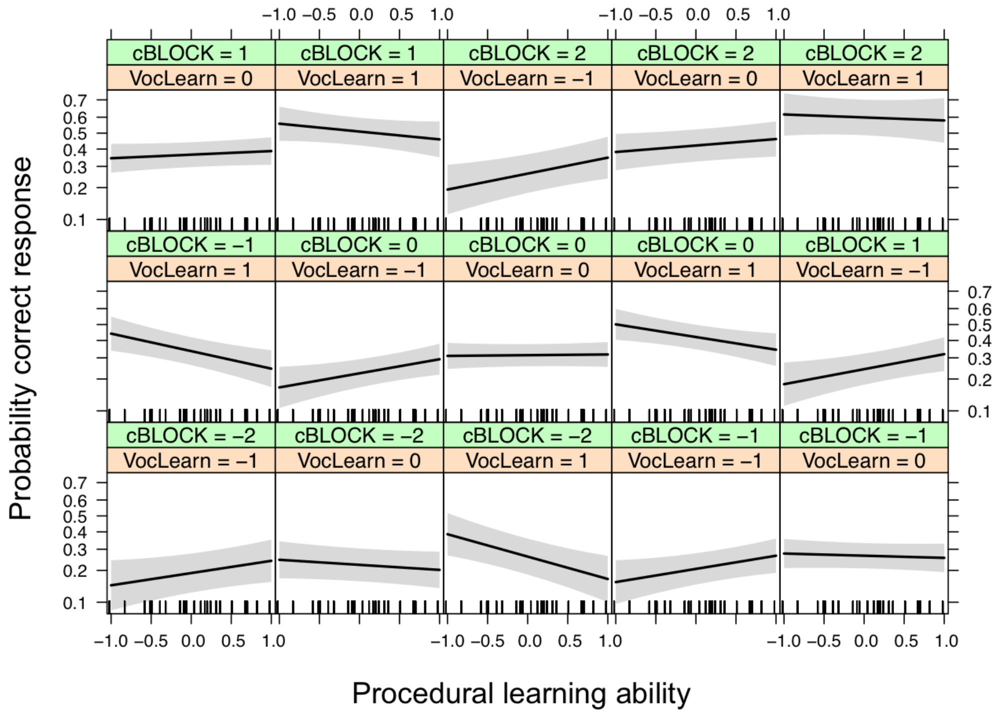


*Figure 6.13*. Effect of the Proc:VocLearn Interaction Across Practice.

At the beginning of training (Block = -2) VocLearn had a distinct mitigating effect on Proc (the odds of Proc of predicting accurate trials sharply declined as VocLearn increased). This mitigating trend weakened over the course of practice, as higher levels of VocLearn limited the effect of procedural learning ability to an increasingly lesser extent.

### 6.4.2 RQ1: Summary of results

The first research question explored the relationship between declarative and procedural learning ability and aural comprehension of a novel L2 during practice, and specifically comprehension of the form-meaning relationships linking morphosyntax and thematic interpretation. The results relative to RQ1 came from the analysis of three generalized mixed-effects models, one fitting the complete practice dataset and the other two fitting the symmetric and the asymmetric subsets respectively.

***Intercepts.*** Although learning was significantly above chance independently of trial type, in all three models the odds that single trials were inaccurate were on average significantly higher than the odds that they were accurate.

***Declarative Learning Ability.*** Decl had a significant positive effect on the accuracy of asymmetric trials (particularly SV trials), but did not seem to play a significant role in symmetric trials. Also, a three-way interaction indicated that, in the case of symmetric trials, the positive effect of Decl on accuracy significantly decreased over the course of training as a function of school Year (i.e., the older the children, the more marked the effect). In asymmetric trials Decl was also found to significantly positively interact with VocLearn. Finally, Decl increased a mitigating effect of VocLearn on the ability of Proc to predict accuracy if VocLearn was above average. However, if VocLearn was below average, Decl enhanced the ability of Proc to predict accuracy.

***Procedural Learning Ability.*** Proc was not a significant main effect in any of the models. However, overall, and in asymmetric contexts in particular, its positive effect on accuracy significantly increased over the course of training. As discussed above, the effect of Proc on accuracy in asymmetric trials was moderated by Decl and VocLearn.

***Vocabulary Learning Ability.*** Vocabulary learning ability was a highly significant positive effect in all three models, although it did not play an important role in the accuracy of OVs sentences. Overall, its effect on accuracy significantly increased over the course of training. As mentioned earlier, VocLearn had a significant mitigating effect on Proc but, for participants with average or above average VocLearn levels, the mitigating effect was significantly reduced as a function of training.

***Word Order.*** In the overall model the word order variable (WO) was a significant effect. It emerged that asymmetric trials were significantly more likely to be accurate than symmetric trials. Model A also found that there were significant differences in the way VocLearn supported the accuracy of different trial categories. In particular, the positive effect of VocLearn on the accuracy of OVs trials was significantly weaker compared to most other trial categories.

***Block/Session.*** The effect of training on accuracy as expressed by the Block or the Session variables was positive and highly significant in all three models. The symmetric subset was the only one where the effect of Session was comparatively superior to the effect of Block.

### 6.4.3 RQ2

In the second research question we set out to investigate the effect of declarative and procedural learning ability on accuracy in a grammaticality judgment test (GJT) administered after the game practice with BrocantoJ was completed.

RQ2 To what extent do declarative and procedural learning ability modulate child L2 learning of word order and case marking as measured by a grammaticality judgment test administered at the end of practice?

***Descriptive Statistics.*** The GJT scores were analyzed as binary choices, with the three 'happy' smileys operationalized as sentence endorsements and the three 'sad' smileys operationalized as no endorsement. Accuracy and inaccuracy were then determined based on the match/mismatch between endorsement and sentence grammaticality. Before analysis, the practice sentences (160 items) and the missing cases (65 items) were excluded from the dataset, leaving a total of 1055 valid cases. Subsequently, counts of accurate and inaccurate valid GJT items were aggregated to verify whether learning was above chance at group level (chance set at 50% correct;

Table 6.11

*Aggregate Count of Valid Accurate and Inaccurate Trials in the GJT (N = 40) and Accuracy Significance Above Chance (50%)*

| | Acc | % | Inacc | Significance above chance | | |
|---|---|---|---|---|---|---|
| | | | | $\chi^2$ | *p* | $\Phi_v$ |
| **Overall** | **663** | **62.8** | **392** | **35.41** | **.000** | **.130** |
| **Gramm** | **355** | **67.5** | **171** | **33.20** | **.000** | **.178** |
| **Ungramm** | **308** | **58.2** | **221** | **7.21** | **.007** | **.083** |
| **SV** | **93** | **61.6** | **58** | **4.12** | **.042** | **.117** |
| **SOV** | **404** | **67.1** | **198** | **36.31** | **.000** | **.174** |
| OV | 166 | 55.0 | 136 | 1.49 | .222 | .050 |
| Case viol. | 112 | 49.3 | 115 | below chance | | |
| **WO viol.** | **196** | **64.9** | **106** | **13.71** | **.000** | **.151** |

*Note*. With reference to the model reported in Table 6.12, the categories SV, SOV and OV correspond to the factors of the categorical variable Word Order (WO), whilst the categories Gramm, Case viol. and WO viol. correspond to the factors of the categorical variable Type of Violation (Viola).

Table 6.11). The analysis was performed for overall scores, grammatical items, ungrammatical items, different word orders, and for ungrammatical items displaying case or word order violations.

On average, performance was significantly above chance for all categories except OV sentences and ungrammatical sentences with case violations (in the latter case performance was slightly below chance). Descriptively, a difference can be noticed between grammatical and ungrammatical sentences, whereby accurate judgment of grammatical sentences was more robust and showed a larger effect size compared to accurate judgment of ungrammatical sentences.

***GJT Model.*** A generalized binomial linear regression model was selected according to the criteria presented in 6.3.12 after ensuring no multicollinearity effects between predictor variables arose (CN = 1.23). The outcome variable was accuracy in the GJT, expressed as the log odds that a given GJT trial was judged accurately for a one-unit increase in the predictor variables. The independent variables included the predictors of interest (declarative and procedural learning ability, both standardized, i.e., Decl and Proc), spacing between the administration of session 1 and 2 (Spacing S1/S2, standardized) and two categorical variables both with three levels each, word order (WO, with levels SV, OV and SOV), and type of violation (Viola, with the levels grammatical, i.e. no violation, violation of word order and violation of case marking).

The interactions included a Proc:WO interaction and a WO:Viola interaction. The only random effects included were those of items on intercepts. Also in this case, Class was found to be irrelevant. Note that, as the GJT set included no valid cases of SV sentences with case violations, the model was rank deficient for coefficients related to this category (for a rationale of why this was the case cf., 6.3.8). A summary is provided in Table 6.12 (next page).

***Main Effects (GJT Model).*** The model in Table 6.12 is relative to the baseline WO = OV and Viola = case violation. The negative intercept confirms that OV sentences, and, in particular, sentences with case violations, tended on average to be judged incorrectly. Looking at the cognitive variables of interest, the analysis found a significant positive effect of Proc on accuracy ($p < .05$).

Table 6.12

*Generalized Mixed-Effects Model of the Effects of Declarative Learning Ability, Procedural Learning Ability, Word Order, Type of Violation and Session Spacing on Accurate Response in the GJT (GJT Model) - $R^2$ = .17*

| | | | | 95% CI | | |
|---|---|---|---|---|---|---|
| Fixed effects | β | *SE* | *z* | lower | upper | *p* |
| **(Intercept)** | **-1.30** | **0.48** | **-2.70** | **-2.25** | **-0.36** | **.007**** |
| Decl | -0.02 | 0.07 | -0.27 | -0.16 | 0.12 | .787 |
| **Proc** | **0.29** | **0.11** | **2.57** | **0.07** | **0.52** | **.010*** |
| **WO (SOV)** | **1.79** | **0.58** | **3.06** | **0.64** | **2.93** | **.002**** |
| WO (SV) | -0.28 | 0.65 | -0.43 | -1.55 | 0.99 | .666 |
| **Gramm** | **2.22** | **0.59** | **3.76** | **1.06** | **3.37** | **.000***** |
| **Viola (WO)** | **1.76** | **0.67** | **2.63** | **0.44** | **3.07** | **.009**** |
| Spacing S1/S2 | -0.14 | 0.07 | -1.88 | -0.29 | 0.01 | .060^ |
| Proc:WO (SOV) | -0.21 | 0.14 | -1.50 | -0.48 | 0.06 | .133 |
| Proc:WO (SV) | -0.25 | 0.19 | -1.32 | -0.62 | 0.12 | .187 |
| **WO(SOV):Gramm** | **-1.91** | **0.71** | **-2.67** | **-3.31** | **-0.51** | **.007**** |
| WO(SV):Gramm | 0.29 | 0.87 | 0.33 | -1.41 | 1.99 | .740 |
| WO(SOV):Viola (WO) | -1.00 | 0.82 | -1.21 | -2.61 | 0.61 | .225 |

*Note*. ^$p$ < .10; $^{*}p$ < .05; $^{**}p$ < .01; $^{***}p$ < .001

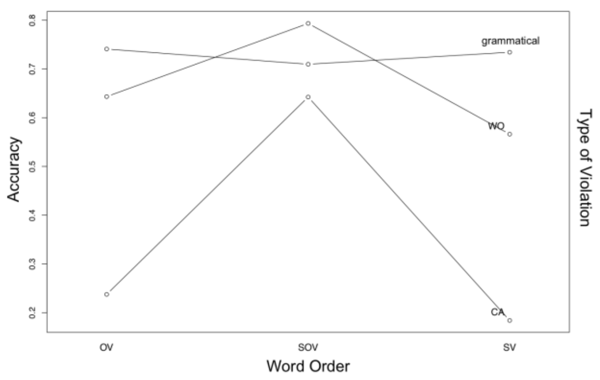


*Figure 6.14*. Effect of Type of Trial and Type of Violation on Accuracy (Probability Correct Response).

Turning to the categorical variables (Figure 6.14), the model returned a positive effect of SOV indicating that (conditionally to Viola = case) participants were on average significantly more accurate on SOV sentences compared to OV sentences ($p < .01$). With regards to type of violation (and conditionally to WO = OV), participants were on average significantly more accurate in grammatical sentences and in sentences containing word order violations compared to sentences containing a case violation ($p < .001$ and $p < .01$ respectively).

***Main Effects: Re-leveling of the Categorical Variables WO and Viola.*** Similarly to the analysis performed on previous models, re-leveling can provide a more complete picture of additional significant effects that may be obscured by specific choices of reference categories for categorical variables. Since both WO and Viola have three levels, this gives rise to a total of nine possible versions of the same model, (see Appendix G). A summary of significant main effects and interactions is provided in Table 6.13a and 6.13b respectively (next pages). Intercepts in different re-leveling scenarios overall indicate significantly positive odds that grammatical sentences were judged accurately. The only positive significant intercept for

Table 6.13a

*Summary of Intercept, Proc, Word Order and Type of Violation Significant Main Effects Depending on Word order and Type of Violation Reference Category*

| | Reference category combination for Word order + Type of Violation | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | OV + case | OV + gramm | OV + word order | SOV + case | SOV + gramm | SOV + word order | SV + case | SV + gramm | SV + word order |
| Intercept | - ** | + ** | | | + *** | + *** | | + ^ | |
| Proc | + * | + * | + * | | | | | | |
| SOV | + ** | | | | | | (+ ^) | | + ^ |
| SV | | | | - ^ | | - ^ | | | |
| OV | | | | - ** | | | | | |
| gramm | + *** | | | | | | (+ ^) | | |
| word order violation | + ** | | | | | | | | |
| case violation | | - *** | - ** | | | | | | |

*Note*. $^{\wedge}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 6.13b

*Summary of Proc:Word Order and Word Order:Type of Violation Significant Interactions Depending on Word Order and Type of Violation Reference Category*

| | Reference category combination for Word order + Type of Violation | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | OV + case | OV + gramm | OV + word order | SOV + case | SOV + gramm | SOV + word order | SV + case | SV + gramm | SV + word order |
| Proc:SOV | | | | | | | | | |
| Proc:SV | | | | | | | | | |
| Proc:OV | | | | | | | | | |
| SOV:gramm | - ** | | | | | | | | |
| SOV:word order | | | | | | | | | |
| SOV:case | | + ** | | | | | | | |
| OV:gramm | | | | + ** | | | | | |
| OV:word order | | | | | | | | | |
| OV:case | | | | | - ** | | | (- **) | |
| SV:gramm | | | | | | | | | |
| SV:word order | | | | | | | | | |

*Note*. $^{**}p < .01$

ungrammatical sentences is for SOV sentences with word order violations. By contrast, in the case of OV sentences with case violations the intercept's coefficient was significantly negative, indicating that on average this type of sentence was significantly more likely to be judged incorrectly in the GJT. The re-leveling revealed that the effect of Proc was positive and significant for all OV sentences independently of grammaticality and type of violation ($p < .05$ in all cases).

***Interactions (GJT Model).*** The only statistically significant interaction in the model (Table 6.12) was between word order type (WO) and type of violation (Viola). In particular the model returned that the difference in accuracy between SOV and OV sentences was significantly less marked when these sentences were grammatical compared to when they contained a case violation ($p < .01$; see Figure 6.14). Alternatively, the negative interaction coefficient can also be interpreted to indicate that the difference between sentences with case violations and grammatical sentences was significantly smaller for SOV sentence, compared to OV sentences.

***Interactions: Re-leveling of the Categorical Variables WO and Viola.*** A further significant interaction effect the model returns can be found for OV and SV sentences in grammatical contexts compared to case violation contexts (Table 6.13b). However, since the interaction involves a comparison with SV sentences in case violation contexts, and this is an empty subset in the GJT dataset, the interaction cannot be meaningfully interpreted and is thus not discussed further.

### 6.4.4 RQ2: Summary of results

***Intercepts.*** The intercepts' β coefficients indicated that all grammatical sentences, as well as SOV ungrammatical sentences with word order violations were significantly more likely to be judged accurately than not. Only for ungrammatical OV sentences with case violations the log odds were significantly negative.

***Procedural Learning Ability.*** Proc had a significant positive effect on the accurate judgment of all OV sentences, independently of grammaticality.

***Word Order and Type of Violation.*** OV grammatical sentences and OV ungrammatical sentences with word order violations were significantly more likely to be accurate compared to OV ungrammatical sentences with case violations. In ungrammatical sentences with case violations SOV sentences were significantly more likely to be judged correctly than OV sentences. An interaction between word order and type of violation also found that the difference in accuracy between SOV and OV grammatical sentences was significantly smaller than the difference between SOV and OV sentences in case violation contexts.

### 6.4.5 RQ3

The third research question investigated whether the participants developed explicit or/and implicit knowledge of BrocantoJ morphosyntax during the experiment.

RQ3 To what extent is the L2 knowledge acquired by the children implicit/ explicit?

In order to answer the question two main sources of evidence were deployed; confidence rating data (for the evaluation of the guessing criterion and the zero-correlation criterion) and verbal reports.

***Descriptive Statistics.*** Similarly to the previous analysis, practice items were removed from the GJT dataset. An aggregate overview of confidence rating categories for valid experimental trials is provided in Table 6.14.

Table 6.14

*Count of Valid Experimental Trials in the GJT According to Level of Confidence*

| Confidence Level | *n* |
|---|---|
| High | 518 |
| Medium High | 309 |
| Medium Low | 156 |
| Low | 104 |
| Total Valid | 1087 |
| Missing | 33 |
| Total | 1120 |

The data show that confidence tended to be rated highly overall. Assigning a decreasing numeric value to confidence levels from higher to lower allowed an additional treatment of the confidence variable (Conf) as a continuous predictor (high confidence = 8; medium high confidence = 7; medium low confidence = 6; low confidence = 5). Overall, the descriptive statistics of the continuous variable confirmed the tendency to high confidence ratings in GJT judgments ($M$ = 7.15; $SD$ = 1.0). The continuous Conf variable was also subsequently standardized for the purposes of model fitting.

The number of total valid trials in this analysis ($N$ = 1025) was lower compared to the previous GJT analysis ($N$ = 1055). This depended on the fact that a different number of errors were made by the participants in selecting the key for the sentence judgment compared to selecting the key for the confidence level judgment. The most common types of errors included pressing a nontarget key, pressing no key, or pressing a valid key too late.

***Guessing Criterion.*** The guessing criterion (see also 2.4.1) is a criterion that can provide an initial indication of whether the knowledge the participants had of their judgments was implicit or explicit and is tested by evaluating the accuracy of low confidence items. According to Dienes et al. (1995), if participants show above chance accuracy on sentences they were not confident about, this reveals that the judgment knowledge involved was largely implicit. Following Dienes and Scott (2005) I assume that implicit structural knowledge can be inferred from implicit judgment knowledge (see also 2.4.1). To test the criterion, the confidence variable was treated like a categorical variable with two levels, high confidence (grouping high and medium high confidence items) and low confidence (grouping low and medium low confidence items).

A Chi-square test was then run to test whether accuracy in the two categories was above chance, with the chance level set at 50% correct (see Table 6.15, next page). The test found that both high confidence and low confidence items were judged correctly above chance, however only high confidence items were accurate significantly above chance. As such, these results did not provide clear enough

Table 6.15

*Count of Valid Experimental Accurate and Inaccurate Trials in the GJT (N = 40) According to Level of Confidence and Assessment of Accuracy Significance Above Chance Performance (50%)*

| | Acc | % | Inacc | Significance above chance | | |
|---|---|---|---|---|---|---|
| | | | | $\chi^2$ | *p* | $\Phi_v$ |
| **High Confidence** | **503** | **64.6** | **276** | **33.81** | **.000** | **.147** |
| Low Confidence | 143 | 58.1 | 103 | 3.27 | .070 | .082 |

evidence of the availability of implicit judgment knowledge in the GJT indicating that further analysis was required. Learning for the low confidence category was not significantly above chance also when the category excluded medium-low confidence cases.

***Zero-correlation Criterion.*** The zero-correlation criterion is a further criterion used in cognitive psychology to assess the implicit vs. explicit nature of judgment knowledge (Dienes & Berry, 1997). The 'zero' correlation refers to the lack of positive correlation between overall confidence and accuracy scores. According to this criterion, if judgment knowledge is explicit, i.e. the participant is applying conscious judgment knowledge in the rating, one would expect her 'to give a higher confidence rating when she actually knows the answer and a lower confidence rating when she is just guessing' (Dienes & Perner, 2004, p. 174). This means that one would expect a positive correlation between level of confidence and accuracy if the judgment knowledge is conscious (i.e., explicit) and no correlation if the judgment knowledge is unconscious (i.e., implicit)[1].

***Confidence Level Model.*** In order to test this criterion a binomial generalized linear mixed-effects model was fitted to the GJT data with accuracy in the GJT as the outcome variable and with the confidence level as the main predictor of interest (CN = 1.18; Table 6.16). The confidence level variable (Conf) was obtained by standardizing the continuous version of the variable.

---

[1] In order to further validate the use of the zero-correlation criterion Dienes and Perner (2004) recommended that the participants' attitude towards the confidence judgment be also assessed. However, as testing attitude would require the participants to rate their confidence for the same items repeatedly, this could not be implemented in the present study design.

Table 6.16

*Generalized Mixed-Effects Model of the Relationship between Judgment Confidence and Accuracy in the GJT Modulated by Declarative Learning Ability, Procedural Learning Ability, and Type of Violation (Confidence Level Model) - $R^2$ = .28*

| | | | | 95% CI | | |
|---|---|---|---|---|---|---|
| Fixed effects | β | *SE* | *z* | lower | upper | *p* |
| (Intercept) | -0.08 | 0.35 | -0.22 | -0.77 | 0.61 | .827 |
| **Conf** | **-0.47** | **0.20** | **-2.39** | **-0.86** | **-0.08** | **.017*** |
| Decl | 0.03 | 0.10 | 0.34 | -0.16 | 0.23 | .733 |
| Proc | -0.14 | 0.15 | -0.94 | -0.42 | 0.15 | .345 |
| **Gramm** | **0.98** | **0.42** | **2.34** | **0.16** | **1.80** | **.019*** |
| Viola (WO) | 0.88 | 0.46 | 1.92 | -0.02 | 1.80 | .055^ |
| **Conf:Gramm** | **1.12** | **0.24** | **4.69** | **0.65** | **1.59** | **.000***** |
| Conf:Viola(WO) | 0.37 | 0.26 | 1.39 | -0.15 | 0.89 | .164 |
| **Conf:Proc** | **-0.21** | **0.07** | **-2.94** | **-0.35** | **-0.07** | **.003**** |
| Proc:Gramm | 0.28 | 0.17 | 1.66 | -0.05 | 0.61 | .096^ |
| **Proc:Viola(WO)** | **0.37** | **0.18** | **2.05** | **0.02** | **0.72** | **.040*** |

*Note*. ^$p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

In order to investigate to what extent the relationship between confidence level and accuracy was modulated by the Type of Violation, the categorical variable Viola was also included in the analysis and kept in the final model as it improved it significantly. In order to understand the extent to which cognitive variables modulated the relationship between confidence and accuracy, Decl and Proc were also included. The following two-way interactions were found to significantly improve the model's fit: Conf:Viola, Conf:Proc, and Proc:Viola. The random effects included effects of subjects and items on intercepts and on the slopes of the Conf variable. In the model output in Table 6.16 the baseline for the Viola variable is Case.

***Main Effects (Confidence Level Model).*** The intercept in the model with Case violation as baseline confirmed that on average sentences with case violations were slightly more likely to be inaccurate than to be accurate (see also Table 6.11 for the relevant descriptive statistics). The model returned a significant ($p < .05$) negative effect of Conf on accuracy. However, the fact that learning was below chance for case violation sentences prevents one to draw meaningful conclusions relative to the consequences of this relationship for the interpretation of judgment knowledge (Dienes & Perner, 2004). The other significant main effect was a positive effect of grammaticality ($p < .05$), confirming that, compared with case violation trials, judgment for grammatical sentences was significantly more accurate.

***Main Effects: Re-leveling of the Categorical Variable Viola.*** The initial model was re-leveled (Table 6.17a, next page; alternative models reported in full in Appendix H). The positive and significant intercept ($p < .01$) for the model with Viola = WO confirmed that trials with ungrammatical word order were significantly likely to be judged accurately in the GJT. In this case however, the relationship between confidence and accuracy was not significant (Appendix H). According to Dienes and Perner (2004) this constellation of results indicates that participants' knowledge of their judgment was implicit. Due to the fact that implicit structural knowledge can be inferred from implicit judgment knowledge, these results also suggest that the participants' structural knowledge relative to the properties of BrocantoJ word order was also largely implicit, i.e. unconscious.

Table 6.17a

*Summary of Significant Main Effects on Accuracy in the GJT Depending on Viola Reference Category (Confidence Level Model)*

| | Type of violation reference category | | |
|---|---|---|---|
| | CA | WO | Gramm |
| Intercept | | + ** | + *** |
| Conf | - * | | + *** |
| Decl | | | |
| Proc | | + ^ | |
| Viola(CA) | | - ^ | - * |
| Viola(WO) | + ^ | | |
| Viola(Gramm) | + * | | |

*Note*. ^$p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

In the third version of the model a statistically significant positive effect for the intercept ($p < .001$) confirms a higher level of correct judgments for grammatical sentences compared to other sentence types. In this case there was also a positive and significant correlation between Conf and accuracy. According to Dienes and Perner (2004), this pattern of results is compatible with the interpretation that judgment knowledge was explicit (conscious). Since explicit judgment knowledge can be related to either implicit or explicit structural knowledge, no conclusions can be drawn in this case relative to the nature of the underlying linguistic (structural) knowledge. A significant negative effect ($p < .05$) for sentences with case violations indicates that (for average levels of procedural learning ability and confidence) these sentences were significantly less likely to be judged correctly compared to grammatical sentences.

***Two-way Interactions.*** The Conf:Gramm interaction indicated that high confidence was significantly more associated to accuracy in grammatical sentences than to accuracy in sentences with case violations ($p < .001$). Further, the model returned a significant negative effect ($p < .01$) for the Conf:Proc interaction, indicating that higher levels of confidence significantly correlated with lower levels of procedural learning ability.

Finally, the Proc:Viola (WO) interaction indicated that, compared to sentences with case violations, sentences with word order violations were significantly more likely to be judged accurately for increasing levels of procedural learning ability. In general, significant results for the Proc:Viola interaction are particularly important because they add additional information relative to the relationship between procedural learning ability and type of violation that could not be extracted from the previous GJT model, where the Viola variable did not appear in an interaction term with Proc.

***Interactions: Re-leveling of the Categorical Variable Viola.*** As reported in Table 6.17b (next page), the version of the model with Viola = WO returned a positive significant ($p < .001$) effect for grammatical sentences. This indicates that, the effect of confidence on accuracy was significantly more positive in grammatical sentences compared with word order violation sentences.

Table 6.17b

*Summary of Significant Interaction Effects on Accuracy in the GJT Depending on Viola Reference Category (Confidence Level Model)*

| | Type of violation reference category | | |
|---|---|---|---|
| | CA | WO | Gramm |
| Conf:Viola(CA) | | | - *** |
| Conf:Viola(Gramm) | +*** | + *** | |
| Conf:Viola(WO) | | | - *** |
| Conf:Proc | - ** | - ** | - ** |
| Proc:Viola(CA) | | - * | - ^ |
| Proc:Viola(Gramm) | + ^ | | |
| Proc:Viola(WO) | + * | | |

*Note*. ^$p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

***Debriefing Questionnaire.*** After coding (see 6.3.11) the debriefing questionnaire scores were analyzed calculating the mean, median and mode (*M* [*SD*] = 2.05 [1.3]; *Mdn* = 2.00; *Mo* = 1.00). The descriptive statistics (Table 6.18) also

Table 6.18

*Raw Score Frequencies in the Debriefing Questionnaire*

| Scores | Frequency | Percent | Cumulative Percent |
|---|---|---|---|
| 1 | 18 | 45.0 | 45.0 |
| 2 | 13 | 32.5 | 77.5 |
| 3 | 3 | 7.5 | 85.0 |
| 4 | 2 | 5.0 | 90.0 |
| 5 | 3 | 7.5 | 97.5 |
| 6 | 1 | 2.5 | 100.0 |
| Tot | 40 | 100.0 | |

revealed that the verbal reports relative to about three quarters of the participants (31 out of 40, corresponding to 77.5%) indicated a very low to low awareness of the structural properties of the new language (these corresponded to scores 1 = 'reports to have noticed nothing in particular' and score 2 = 'reports noticing the presence/absence of specific words'). In the case of score 2, most participants reported having noticed the prepositions 'ri' and/or 'ru', and the fact that occurrence in the input sentences was not consistent.

The remaining participants (9 out of 40, corresponding to 22.5% of the overall number) were assigned scores higher than 2. For scores higher than 2, participants' verbal reports provided explicit reference of words' positions in the sentence and evidence of understanding that word order regularities involved syntactic domains larger than the word. Three participants were assigned score 3, corresponding to 'reports noticing the presence/absence of a single specific word and refers to its position in the sentence', 2 participants were assigned score 4, corresponding to 'reports that there is an order involving domains larger than a single word but does not provide examples', 3 participants were assigned score 5, corresponding to 'reports that there is an order in domains larger than a single word and provides examples'. Only one student was assigned score 6 as he spontaneously reported that the language had a fixed word order at sentence level and provided a complete and correct example of an SOV BrocantoJ sentence. None of the participants reported to have understood the role of the prepositions *ri* and *ru* or attempted to provide hypotheses to account for their function or meaning.

The explicit language knowledge scores were also standardized and correlated to age at testing, to the measures of cognitive ability and the measures of L2 learning. Only two measures, overall game score ($r$ = .604, $p$ < .01) and vocabulary learning ability ($r$ = .455, $p$ < .05) positively and significantly correlated with the explicit language knowledge measure (both sets of estimates reflect bootstrapped and Holm-Bonferroni corrected results).

### 6.4.6 RQ3: Summary of results

***Intercepts and Type of Violation.*** Confirming the results of the previous GJT model, the $\beta$ coefficients in the Confidence Models indicated that grammatical sentences were significantly more likely to be accurate than sentences with case violations, and that sentences with word order violations were marginally more likely to be accurate than sentences with case violations.

***Procedural Learning Ability.*** The Proc:Viola correlation showed that Proc had a significantly larger positive effect in the accuracy of ungrammatical word order trials compared to ungrammatical case trials. Independently of the reference category adopted for Type of Violation, the model returned a significant negative correlation for the interaction between Proc and confidence.

***Confidence.*** The results revealed that judgment confidence was significantly negatively correlated with accuracy in ungrammatical sentences with case violations and significantly positively correlated with accuracy in grammatical sentences. For ungrammatical sentences with word order violations no significant correlation was found between confidence and accuracy. In the light of the zero-correlation criterion, these results can be interpreted as evidence of the role of implicit knowledge in the judgment of ungrammatical word order patterns and, indirectly, as evidence of implicit structural knowledge of word order constraints.

By contrast, judgment knowledge in the case of grammatical sentences was explicit, and no conclusions can be drawn with regards to the nature of structural knowledge in this case. Also, no conclusions can be drawn relative to the nature of judgment or structural knowledge based on ungrammatical sentences with case violations, since learning in this case was not significantly above chance. Differences among sentence types emerged also in the estimates for the interaction between confidence and type of violation, where the difference between grammatical sentences and both types of ungrammatical sentences was positive and highly significant.

***Explicit Language Knowledge.*** A descriptive analysis of the results of the debriefing questionnaire found that, according to the reports of about three quarters of the participants, explicit language knowledge of BrocantoJ at the end of the experiment did not go beyond noticing the occurrence/absence of specific words in the input. Only about one quarter of the participants referred to syntactic properties of the input (word order). None of the participants indicated they understood the function of the case particles *ri* and *ru* or reported hypotheses about their meaning. A correlation analysis revealed that explicit knowledge of BrocantoJ as assessed by the questionnaire significantly directly correlated with vocabulary learning ability and the overall accuracy score in the computer game but not with overall or partial scores in the GJT.

# 7. Study 2

## 7.1 Introduction

This chapter presents a replication of Study 1 with adult participants. Due to the similarities with Study 1 part of the Methods section will refer to Chapter 6. As in the previous study, the results will be presented relative to each research question. The Discussion of Study 2 is also included in Chapter 8.

## 7.2 Research questions

In the light of the literature review in the area of adult artificial L2 learning and memory, the following research questions were formulated for Study 2:

RQ1 To what extent do declarative and procedural learning ability modulate adult L2 aural comprehension and learning of the rules linking morphosyntax and thematic interpretation during practice?

RQ2 To what extent do declarative and procedural learning ability modulate adult L2 learning of word order and case marking as measured by a grammaticality judgment test administered at the end of practice?

RQ3 To what extent is the L2 knowledge acquired by the adults implicit/ explicit?

Based on the adult miniature language studies reviewed in Chapter 4, the prediction is that adults aurally exposed to BrocantoJ for about 2 1/2 hours in incidental conditions should be able to learn formal properties of the language including word order (e.g., Francis et al., 2009; Morgan-Short et al., 2014; Pili-Moss, 2017; Rebuschat & Williams, 2012; Williams & Kuribara, 2008; Wonnacott et al., 2008), morphology (e.g., Brooks et al., 2006; Grey et al., 2015; Morgan-Short et al., 2010; Rogers et al., 2015) and the relationship between morphosyntax and semantic interpretation (e.g., Boyd et al., 2009; Grey et al., 2015; Pili-Moss, 2017). Also, if the study results confirm previous research, case morphemes should be more difficult to learn compared to word order.

With regards to the cognitive abilities, previous BROCANTO2 adult studies with greater amounts of practice (> 5 hours, corresponding to 20+ game blocks) and spanning over a longer period of time (4 - 14 days), have found that declarative learning ability predicts learning at early stages of practice and procedural learning ability has a significant effect later in practice (Brill-Schuetz & Morgan-Short, 2014; Morgan-Short et al., 2014). Based on this evidence, the hypothesis is that, at least initially, declarative learning ability will have a significant effect on L2 learning. However, the amount of practice administered in the present study (6 game blocks) might not be sufficient to observe procedural memory effects in adults. If this is the case, learning of word order and morphosyntactic patterns is expected to rely on declarative memory, even if these structures would eventually be acquired procedurally with sufficient practice.

Based on Ullman's DP model, a further prediction is that accurate comprehension of form-meaning relationships, which requires the processing of semantic constraints, will be likely to engage declarative memory. However, since linking rules can rely on word order, a significant engagement of procedural memory cannot be excluded. Finally, with regards to RQ3, the review in Chapter 4 evidenced that most (if not all) training studies that have investigated artificial language knowledge have found that adults largely acquired explicit knowledge of morphosyntax as a result of incidental exposure. Based on these findings, this will also be my initial hypothesis.

## 7.3 Methods

### 7.3.1 Participants

For the study 36 L1 Italian young adults (17 females) were recruited in Northern Italy (Milan area) and at the University of Lancaster advertising in university libraries and via a Facebook page. Their mean age at testing was 22 years (*SD* = 3.7; range 18-31); 22 years for the female group (*SD* = 3.4; range 18-30), and 23 years for the male group (*SD* = 3.9; range 18-31). The participants reported no history of learning differences or hearing impairment and had normal or corrected to normal vision. Prior to the experimental sessions the participants completed a background questionnaire designed to provide information about their schooling and the languages they had been exposed to (in school for periods of one year or longer and/or during periods of immersion abroad longer than one month). Except for one male and one female (who had completed secondary education and were in work), all participants had received education at university level and were university students at the time of testing. Given the characteristics of the Italian school system it is common for young adults with high levels of education to have been formally instructed in one or more L2 languages.

At the time of testing all participants have had formal instruction in English as a second language for periods >10 years and reported to have had instruction on average in 2.5 languages (range 1-5). Particularly relevant for the design of the present study was to assess whether the participants had been exposed to or taught an SOV language. The questionnaire revealed that 4 participants had received instruction in German, an SOV language (German is SOV in dependent sentences and in main sentences with compound tenses or modals). Beside age, it was decided to include schooling, number of L2 languages known, and whether the participant knew an SOV language as covariates in the analysis to assess whether they improved the model's fit. The descriptive details relative to these variables and further descriptive details relative to the adult sample are provided in Table 7.1.

Table 7.1

*Descriptive Statistics Relative to Years of Schooling, Number of L2 Languages, and Spacing between Training Sessions*

| Participants | *n* | Schooling *M (SD)* | L2 *M (SD)* | S1/S2 (days) *M (SD)* | S2/S3 (days) *M (SD)* |
|---|---|---|---|---|---|
| F | 17 | 15.9 (2.6) | 2.9 (1.0) | 1.1 (0.3) | 1.1 (0.3) |
| M | 19 | 16.9 (3.9) | 2.1 (1.0) | 1.3 (0.5) | 1.2 (0.6) |
| Overall | 36 | 16.5 (3.4) | 2.5 (1.1) | 1.2 (0.4) | 1.2 (0.5) |

The adult study was approved by the Ethics Research Committee of Lancaster University (Ref. RS2014/142). All participants gave their written informed consent and were paid 30 Euros (Italy) or 25 Pounds (UK) for their participation in the three experimental sessions. In total the data collection period for the adult group lasted 4 weeks in the autumn of 2016 (Italy) and 1 week in the autumn of 2017 (UK).

### 7.3.2 Methodological similarities/differences between Study 1 and Study 2

As a replication of Study 1 with participants from a different age group, Study 2 shares important methodological aspects with it. Vocabulary training, language training, language practice (the game), the GJT, the debriefing questionnaire and the ASRT task were administered to adults in the same way as they were to children and using the same materials. In order to maximize the comparability of the results, Study 2 also adopted the same type of statistical analysis. As in the previous study, adults were also tested for motor control, although the analysis of these data is not included in the present dissertation.

The main differences between the two studies, presented in detail in the following paragraphs, included how sessions were structured, where data collection was conducted, and which tasks were used to assess verbal and visual declarative memory and aspects of working memory.

**7.3.3 Study design and set up**

Similarly to Study 1, the design included three sessions where the language learning experimental paradigm was administered (cf. Chapter 6, Figure 6.1). However, in Study 2 the memory tasks were administered within the three sessions, and not on additional days. The three sessions lasted about 40-45 minutes, 65 minutes and 75 minutes respectively.

It has been observed that performance on tasks measuring procedural learning abilities decreases significantly if they are administered immediately after tasks requiring the engagement of declarative memory, although the effects are not observed if the two tasks are administered on different days (e.g., Gagné & Cohen, 2016). For this reason, the memory tasks relative to declarative and procedural memory were administered separately in Session 2 and Session 3, counterbalancing the order of administration across participants. Moreover, the task measuring procedural learning ability (ASRT) was always administered as the first task at the beginning of the relevant session, i.e., prior to vocabulary testing. The declarative memory tasks and the working memory tasks were administered immediately after the game practice (and before the GJT in Session 3).

The data collection for the adult group was conducted in two quiet study rooms available for booking at a local library (Italy) and in two laboratory rooms at Lancaster University (UK)[2].

### 7.3.4 Cognitive tasks materials and procedure

This section exclusively provides a description of the cognitive tasks that differed from those deployed in the child study. For procedures that are not described here see 6.3.

***Visual Declarative Learning Ability (Rey-Osterrieth Complex Figure - ROCF)***. The Rey-Osterrieth Complex Figure (ROCF, Rey, 1941; Osterrieth, 1944) consists in a complex picture widely used in neuropsychological practice to test nonverbal declarative memory. In this study the materials for the administration of the task included an A4 printout of the ROCF, two white A4 paper sheets and a pen. The ROCF (Figure 7.1) was presented to the participant who was given about 5 minutes to

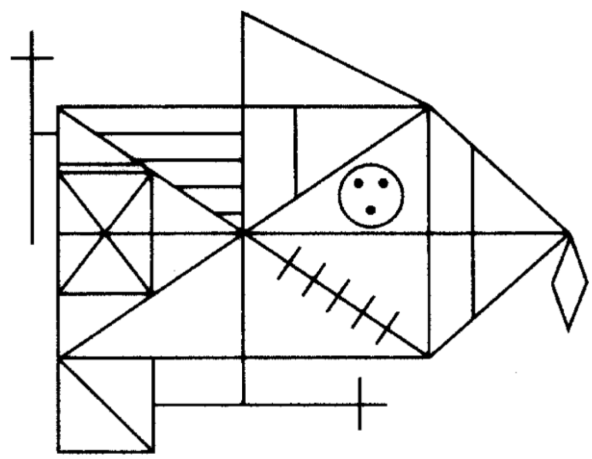

*Figure 7.1*. Rey-Osterrieth Complex Figure (ROCF).

2 Although participants were tested in pairs for computer-based tasks, a second room was necessary to administer the declarative memory tasks and the working memory tasks.

to draw a copy. After 15 minutes, during which the participant was occupied with the first part of a verbal declarative memory task and two working memory tasks, she was asked, with no previous warning, to reproduce the figure again from memory in as much detail as possible. The participant was given about 10 minutes to complete the task.

The ROCF was chosen as a visual declarative memory tasks for this study because fairly recent normative data for the Italian adult population were available (Caffarra, Vezzadini, Dieci, Zonato, Venneri, 2002). In their normative study Caffarra et al. (2002) found that age, sex, and education had significant effects on the performance on the delayed recall.

***Verbal Declarative Learning Ability (Short Story).*** The materials for the administration of this task and its recording protocols are part of a memory and cognitive ability battery recently normed for the Italian adult population (Mondini, Mapelli, Vestri, Arcara, Bisiacchi, 2011). Mapelli et al. (2011) found that age, sex, and education had significant effects on the performance of immediate and delayed recall. Apart from the correction values, the short story (*Raccontino Anna Pesenti*) and the mode of administration and scoring were the same as those described in the child study (see 6.3.10). In the adult study the story and its immediate recall were administered immediately after the ROCF copy, whilst the story's delayed recall followed the ROCF drawing from memory.

***Phonological Loop (Forward Digit Span).*** The materials for the Forward Digit Span this task and its administration and scoring protocols were taken from Monaco, Costa, Caltagirone, & Carlesimo (2013), who also provided recent normative data for the Italian adult population finding that age and education had a significant effect on task performance.

In this task the researcher reads to the participant a series of digit sequences of increasing length (from 3 to 9 items) with monotone intonation and at the rate of about one second per digit. The participant is asked to repeat the full sequence immediately afterwards, and, if she can recall it correctly, is presented a sequence one digit longer. In case of errors, a second sequence of the same length is presented. The test ends when errors on two consecutive lists of the same length occur, or when the participant correctly recalls a nine-digit sequence.

***Central Executive (Backward Digit Span).*** This task is administered in the same way as the previous with the difference that in this case the participant is asked to reproduce the sequence backwards. Providing recent normative data for the Italian adult population also for this task, Monaco et al. (2013) found that age and education had significant effects on performance. Unlike the Forward Digit Span, the Backward Digit Span is considered to be a measure of the Central Executive because it not only requires the participant to retain information in short term memory for immediate repetition, but also to perform an operation (order reversal) before reproduction (Baddeley, 1996).

### 7.3.5 Measures of cognitive ability

For a description of how the measures of language learning, confidence judgment, explicit language knowledge of word order, vocabulary and procedural learning ability were obtained see 6.3.11. For vocabulary learning ability, Table 7.2 shows a descriptive statistics of the participants' errors during vocabulary testing.

Table 7.2

*Errors in the Vocabulary Test per Session*

| | | S1 | S2 | S3 | Tot |
|---|---|---|---|---|---|
| Participants | *n* | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* |
| F | 17 | 1.4 (2.0) | 0.8 (1.1) | 0.3 (0.7) | 2.5 (2.5) |
| M | 19 | 0.8 (1.4) | 1.1 (1.6) | 0.4 (0.9) | 2.3 (3.1) |
| Overall | 36 | 1.1 (1.7) | 1.0 (1.4) | 0.4 (0.8) | 2.4 (2.8) |

***Declarative Learning Ability.*** For declarative learning ability three scores were obtained: (a) a score of visual-spatial declarative memory (relative to the ROCF recall), (b) a score of verbal declarative memory (the score from the short story task), and (c) a composite score of declarative memory. For the ROCF task the recall drawing was scored using the method described in Taylor (1998). In this system 18 elements in the figure are identified and evaluated on a two-point scale (range: 0-36). Two points are given if the item is placed and drawn correctly, one point if it is either well drawn but placed incorrectly or incomplete but placed correctly, half a point if the item is present but incomplete and place incorrectly, and zero points if the element is absent or not recognizable. The raw scores thus obtained were corrected for age, sex and education (Caffarra et al., 2002).

For verbal declarative memory the scoring was performed according to the indications in Mondini et al. (2011), by assigning one point for each information unit accurately recalled in the immediate recall of the story and repeating the procedure for the delayed recall. The raw scores of both recalls were averaged and corrected for age, sex and education. The composite score of declarative learning ability was obtained by standardizing the corrected scores of the visual and verbal tasks and averaging the two scores.

***Phonological Loop.*** To obtain a measure of the phonological loop the Forward Digit Span task was scored assigning the number of points corresponding to the longest correctly recalled sequence. The raw scores thus obtained where corrected for age, sex and education (Mapelli et al., 2011).

***Central Executive.*** To obtain a measure of the central executive the Backward Digit Span task was scored assigning the number of points corresponding to the longest reversed sequence that was recalled correctly. The raw scores thus obtained where corrected for age and education (Mapelli et al., 2011).

***Explicit Language Knowledge (Thematic Linking).*** Unlike children, adults reported different levels of awareness of the linking rules between syntax and thematic interpretation, as well as hypotheses regarding the meaning and function of the case particles. Following a procedure similar to the one adopted to code explicit knowledge of word order regularities, the debriefing questionnaires where assigned to one of the following four categories: (a) 'does not mention *ri*/*ru* or suggest a thematic interpretation for NPs' (score = 0); (b) 'mentions *ri*/*ru* but does not suggest a thematic interpretation for case particles or NPs' (score = 1); (c) 'suggests nonthematic interpretation for *ri*/*ru*' (score = 2); (d) 'suggests thematic interpretation for *ri*/*ru* and/or for the NPs' (score = 3). Also in this case (cf. 6.3.11) the awareness scale can be reduced to the tripartite model deployed in e.g. Rosa and Leow (2004). Level (0) would correspond to 'no report', level (1) would correspond to 'noticing' the particles, and levels (2) and (3) would correspond to increasing levels of understanding of the particles' semantics/of linking rules (cf. Appendix E for rating examples).

## 7.4 Results

Following the same pattern of Study 1, the results will be presented with reference to the relevant research question addressed. For each research question, first an overview of the descriptive statistics will be provided, followed by the discussion of the inferential statistics (mixed-effects models). A detailed discussion of the results including the interpretation of the main effects and interactions in the inferential analysis in the light of the theoretical framework introduced in chapters 1 and 2 can be found in section 8.2.1 (RQ1), 8.2.2 (RQ2) and 8.2.3 (RQ3).

### 7.4.1 RQ 1

RQ1 To what extent do declarative and procedural learning ability modulate adult L2 aural comprehension and learning of the rules linking morphosyntax and thematic interpretation during practice?

***Descriptive Statistics.*** Similarly to the previous study (see also, Morgan-Short, 2007; Pili-Moss, 2017) the chance level for accuracy during game practice was set at 14% correct trials per block (threshold of significance above chance 26.5%) and participants' data were included in the analysis only if they were above chance in at least 3 of the six blocks, or significantly above chance in at least one block. According to this criterion, the data from all tested participants (*N* = 36) were included in the analysis. Table 7.3a reports overall and by-block accurate performance. Accurate and inaccurate trial counts were aggregated to verify whether

Table 7.3a

*Adult Mean Accurate Performance During Practice (Percentage)*

| | *M* (*SD*) | *SE* |
|---|---|---|
| Block 1 | 30.4 (13.7) | 2.3 |
| Block 2 | 42.9 (17.8) | 3.0 |
| Block 3 | 47.6 (22.4) | 3.7 |
| Block 4 | 63.8 (23.3) | 3.9 |
| Block 5 | 63.2 (21.1) | 3.5 |
| Block 6 | 67.9 (23.6) | 3.9 |
| Overall | 52.7 (17.5) | 2.9 |

learning was on average significantly above chance at group level, both overall as well as for different word order types (Table 7.3b, next page).

The data indicate that, on average, learning was significantly above chance overall and in all subcategories, independently of word order type and symmetry in the context. However, differences emerged in the magnitude of the effect size (Cramer *V*), which is medium to large in all categories except in OVs trials, where it is small to medium. Means and standard deviations were also calculated for the raw scores of the cognitive individual differences, including visual and verbal declarative learning ability (Decl), vocabulary learning ability (VocLearn), the measure of the

Table 7.3b

*Aggregate Count of Valid Accurate and Inaccurate Trials in the Game and Accuracy Significance Above Chance (14%)*

| | Acc | % | Inacc | Significance above chance | | |
|---|---|---|---|---|---|---|
| | | | | $\chi^2$ | *p* | $\Phi_v$ |
| **Overall** | **2181** | **52.4** | **1979** | **1385.48** | **.000** | **.408** |
| **SV** | **279** | **67.1** | **137** | **243.60** | **.000** | **.541** |
| **SOV** | **1254** | **50.2** | **1242** | **752.60** | **.000** | **.388** |
| **OV** | **648** | **51.9** | **600** | **405.57** | **.000** | **.403** |
| **SOVs** | **797** | **45.1** | **970** | **411.24** | **.000** | **.341** |
| **SOVa** | **457** | **62.7** | **272** | **365.63** | **.000** | **.501** |
| **OVs** | **274** | **37.6** | **455** | **106.02** | **.000** | **.270** |
| **OVa** | **374** | **72.1** | **145** | **355.99** | **.000** | **.586** |
| **symm** | **1071** | **42.9** | **1425** | **513.04** | **.000** | **.321** |
| **asymm** | **1110** | **66.7** | **554** | **960.16** | **.000** | **.537** |

*Note*. SOVs/OVs = symmetric SOV/OV trials; SOVa/OVa = asymmetric SOV/OV trials

phonological loop (PhonLoop), the measure of the central executive (Exec), and the two measures of procedural learning ability obtained from the analysis of RTs and accuracy in the ASRT task (Proc). The composite measures of declarative and procedural learning ability used in the inferential analysis were obtained from standardized scores and are not reported (Table 7.4).

Table 7.4

*Raw Score Means Relative to the Measures of the Cognitive Individual Differences (N = 36)*

| | *M* (*SD*) | *S.E.* |
|---|---|---|
| Decl (visual) | 17.44 (5.3) | 0.88 |
| Decl (verbal) | 13.85 (4.0) | 0.66 |
| Proc (Acc) | 2.72 (0.5) | 0.09 |
| Proc (RT) | 2.24 (11.9) | 1.97 |
| VocLearn | -2.43 (2.8) | 0.47 |
| Central Executive | 4.56 (1.03) | 0.17 |
| Phonological loop | 5.43 (1.0) | 0.17 |

***Model A.*** After controlling for multicollinearity (CN = 1.98) a generalized binomial linear model was evaluated according to the criteria presented in 6.3.12. All continuous variables were standardized and the Block and Session variables were centered. The fixed effects in the final model included the predictors of interest (Decl and Proc), Session, VocLearn, Exec, the spacing between Session 1 and Session 2, and word order (WO), a categorical variable with 5 levels: (a) SV; (b) symmetric SOV (SOVs); (c) asymmetric SOV (SOVa); (d) symmetric OV (OVs); and (5) asymmetric OV (OVa). The analysis revealed that two- and three-way interactions statistically significantly improved the model's fit. The significant two-way interactions included a positive interaction between Decl and Proc, a positive interaction between Decl and Session, a positive interaction between VocLearn and Session (the only two-way interaction effect that was not conditional), a negative interaction between Proc and Exec, a positive interaction between S1S2 spacing and Session, and an interaction between Proc and WO. Three-way interactions were also found to significantly improve the model's fit and significant effects include a Decl:Proc:Session interaction and an Exec:Spacing:Session interaction.

Additional variables that were not found to improve the model's fit include Block, Sex (whether participants were female or male), PhonLoop, the spacing between Session 2 and Session 3, Age, the number of L2 languages known, whether the participant had been exposed to an SOV language, schooling (years), and whether the participant had ever attended/was attending a university course. The random effects, selected according to the criteria presented in 6.3.12, included effects of participants and trial items on intercepts, as well as effects of participants on the slopes of Decl. A summary of the final output is provided in Table 7.5.

Table 7.5

*Effects of Decl, Proc, Word Order, Vocabulary Learning Ability, Executive Function, Spacing and Session on Accuracy During Practice (Model A - Adults) - $R^2$=.56*

| | | | | 95% CI | | |
|---|---|---|---|---|---|---|
| Fixed effects | β | *SE* | *z* | lower | upper | *p* |
| **(Intercept)** | **1.15** | **0.35** | **3.30** | **0.47** | **1.86** | **.001**** |
| **Decl** | **1.07** | **0.23** | **4.68** | **0.62** | **1.52** | **.000***** |
| Proc | 0.33 | 0.29 | 1.14 | -0.23 | 0.89 | .254 |
| **Session** | **1.05** | **0.14** | **7.47** | **0.77** | **1.32** | **.000***** |
| **WO (SOVs)** | **-1.45** | **0.36** | **-4.05** | **-2.15** | **-0.75** | **.000***** |
| WO (SOVa) | -0.27 | 0.40 | -0.68 | -1.06 | 0.51 | .493 |
| **WO (OVs)** | **-1.71** | **0.40** | **-4.24** | **-2.50** | **-0.92** | **.000***** |
| WO (OVa) | 0.23 | 0.44 | 0.52 | -0.63 | 1.08 | .601 |
| **VocLearn** | **0.39** | **0.09** | **4.59** | **0.23** | **0.56** | **.000***** |
| Exec | 0.05 | 0.16 | 0.32 | -0.25 | 0.36 | .747 |
| **Spacing S1S2** | **0.21** | **0.09** | **2.18** | **0.02** | **0.40** | **.029*** |
| **Decl:Proc** | **0.95** | **0.35** | **2.73** | **0.27** | **1.63** | **.006**** |
| **Decl:Session** | **0.22** | **0.10** | **2.17** | **0.02** | **0.42** | **.030*** |
| Proc:Session | -0.17 | 0.11 | -1.49 | -0.39 | 0.05 | .137 |
| **VocLearn:Session** | **0.15** | **0.06** | **2.29** | **0.02** | **0.28** | **.022*** |
| Exec:Session | 0.03 | 0.06 | 0.41 | -0.10 | 0.15 | .681 |
| Exec:Spac S1S2 | 0.09 | 0.11 | 0.80 | -0.13 | 0.32 | .426 |
| **Spac S1S2:Session** | **0.16** | **0.07** | **2.35** | **0.03** | **0.29** | **.019*** |
| Proc:WO (SOVs) | 0.06 | 0.22 | 0.27 | -0.38 | 0.50 | .784 |
| **Proc:WO (SOVa)** | **-0.50** | **0.25** | **-2.05** | **-0.99** | **-0.02** | **.040*** |
| Proc:WO (OVs) | -0.07 | 0.25 | -0.30 | -0.56 | 0.41 | .766 |
| Proc:WO (OVa) | -0.41 | 0.28 | -1.47 | -0.95 | 0.14 | .143 |
| **Proc:Exec** | **-0.72** | **0.28** | **-2.52** | **-1.28** | **-0.16** | **.012*** |
| **Decl:Proc:Session** | **0.53** | **0.14** | **3.63** | **0.24** | **0.81** | **.000***** |
| **Exec:SpacS1S2:Session** | **0.19** | **0.08** | **2.35** | **0.03** | **0.34** | **.019*** |
| Proc:Exec:Session | -0.18 | 0.15 | -1.20 | -0.48 | 0.11 | .229 |

*Note*. ^$p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

***Main Effects (Model A).*** Since the model included a categorical variable (WO), its results are relative to a specific reference category (in Table 7.5, WO = SV). For a general discussion of the interpretation of effects conditional on the setting of specific values for categorical variables see 6.4.1. The intercept's positive β coefficient ($p < .01$) indicates that on average participants were significantly more likely to be accurate than to be inaccurate in any given SV trial. Relative to the predictors of interest the model returned a positive and significant effect of Decl on accuracy ($p < .001$). Both Session and the spacing between Session 1 and Session 2 had positive and significant effects ($p < .001$ and $p < .05$ respectively) indicating that the likelihood that individual trials were accurate increased over the course of training, but also, interestingly, that on average longer spacing between Session 1 and Session 2 was significantly associated to a higher likelihood that responses were accurate.

VocLearn had also a significant positive effect (p < .001), whilst the effect of Exec was slightly positive but not significant. The β coefficients relative to the WO variable indicate that, compared to SV trials, both types of symmetric trials were significantly less likely to be accurate ($p < .001$; cf. Figure 7.2, next page).

***Main Effects: Re-leveling of the Categorical Variable WO.*** Re-leveling of WO yielded four additional versions of the model (Appendix I). Only significant effects relative to each version of the model will be discussed. Since Proc was the only variable appearing in an interaction with WO in the model, variations in the WO reference category affected the returned coefficient for this main effect.

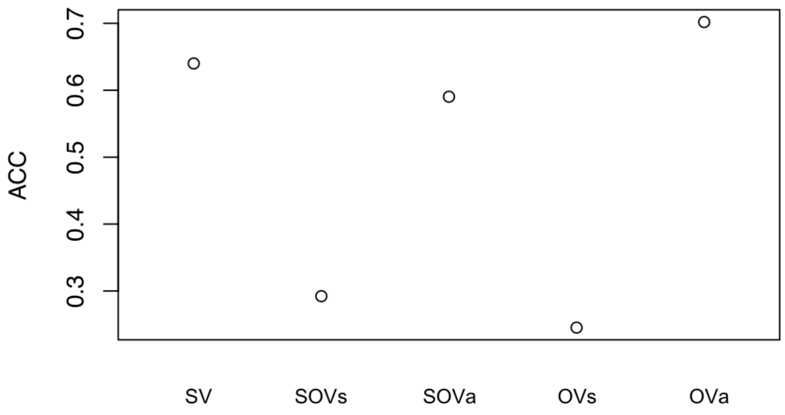


*Figure 7.2*. Effects of the Five Levels of the Word Order Variable (WO) on Accuracy (Probability Correct Response).

For WO = SOVs the model reveals that SV, and both types of asymmetric trials, were on average significantly more likely to be accurate than SOVs trials ($p <$ .001; cf. Figure 7.2).

Setting WO = OVs returned a pattern of results similar to that obtained when the reference category was SOVs, except for the fact that in this case the negative coefficient for the intercept was significant. For WO = OVa the results indicate a pattern similar to that found for versions of the model with reference category SV or SOVa. Summarizing, varying the baseline relative to the categorical variable WO indicates an overall significant difference between symmetric and asymmetric trials, with the former on average statistically significantly less likely to be accurate than the latter (Table 7.6a).

Table 7.6a

*Summary of WO and Proc Significant Main Effects Depending on WO Reference Category*

| | WO reference category | | | | |
|---|---|---|---|---|---|
| | SV | SOVs | SOVa | OVs | OVa |
| Intercept | +*** | | +** | -* | +*** |
| SV | | +*** | | +*** | |
| SOVs | -*** | | -*** | | -*** |
| SOVa | | +*** | | +*** | |
| OVs | -*** | | -*** | | -*** |
| OVa | | +*** | | +*** | |
| Proc | | +^ | | | |

*Note*. ^$p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

***Two-way Interactions.*** The model reported in Table 7.5 returned significant positive effects for the Decl:Proc and the Decl:Session interactions ($p < .01$ and $p < .05$ respectively). These are conditional on a zero value for centered Session and on average procedural learning ability values respectively, and will be discussed in more detail in the section on three-way interactions. A nonconditional VocLearn:Session interaction indicated that the positive effect of VocLearn on accuracy significantly increased over the course of training. A positive interaction between Spacing S1S2 and Session (p < .05), conditional on average values of the Exec variable, also indicated that the positive effect of the spacing between Session 1 and Session 2 significantly increased over the course of the game practice.

A further conditional negative interaction between Proc and Exec ($p < .05$), indicates that higher central executive function had a positive effect on the likelihood that single trials were accurate only for low (below average) levels of Proc and that this effect was mitigated for higher (above average) levels of Proc (Figure 7.3).

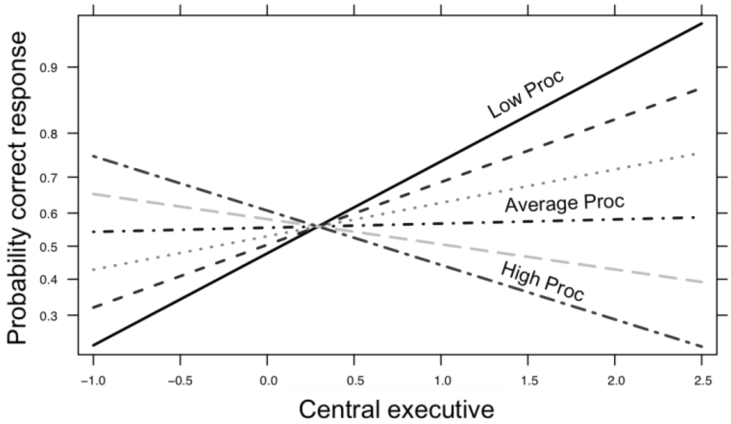


*Figure 7.3*. Effect of the Interaction Between Proc and Exec.

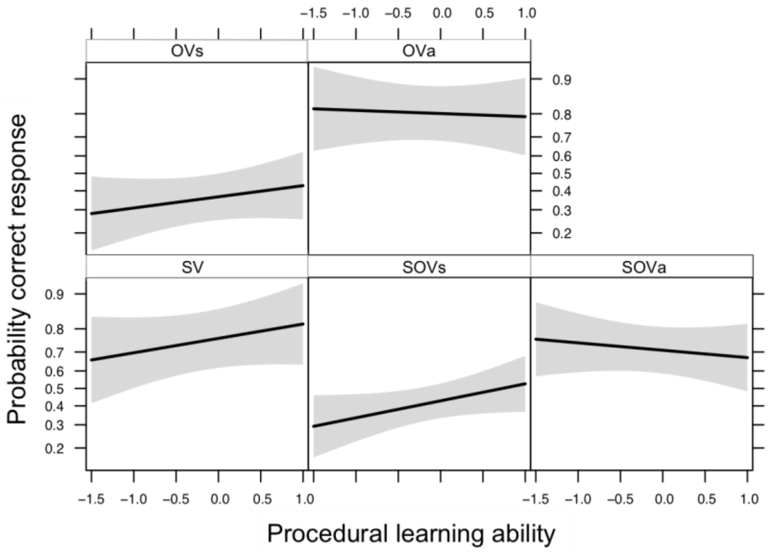


*Figure 7.4*. Effect of Proc on the Accuracy of Different Sentence Types.

Finally, Proc appeared to have a differential effect on the accuracy of different trial types. As Figure 7.4 (above) shows, the general pattern seems to be that whilst the effect of Proc was positive in the case of SV, OVs and SOVs trials, it was negative in the case of SOVa and OVa trials. Specifically, compared to SV, the model returned that the effect of procedural learning ability was significantly worse in SOVa trials ($p < .05$).

***Three-way interactions.*** The model returned a positive Decl:Proc:Session interaction ($p$ <.001) and a positive Exec:Spacing:Session interaction ($p < .05$). Inspection of the plot of the Decl:Proc:Session interaction (Figure 7.5) shows that

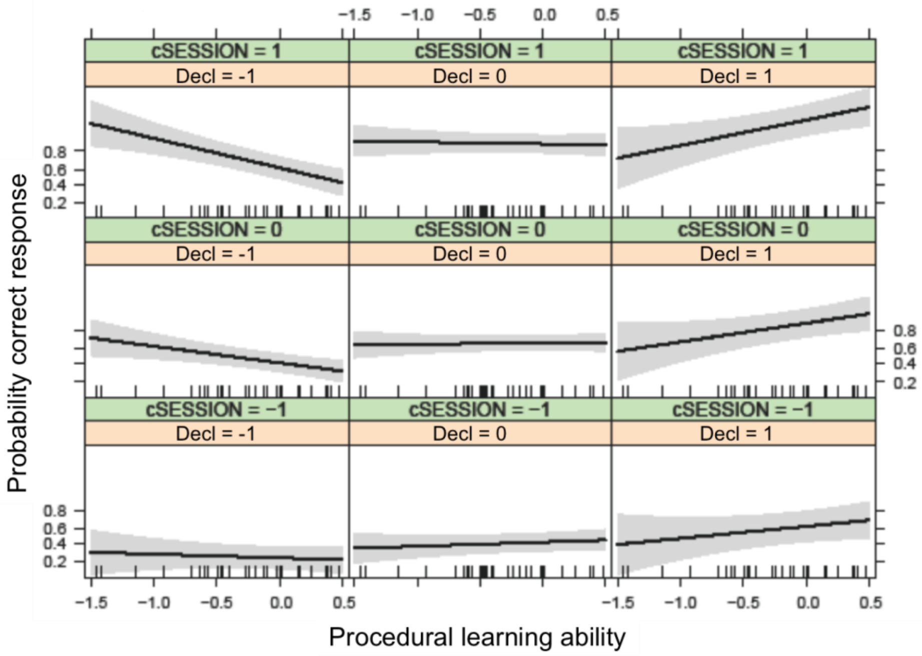


*Figure 7.5*. Effect of the Interaction Between Decl and Proc Across Practice.

in general the relationship between Decl and Proc was co-operative, i.e. resulted in an increase in accuracy for above-average declarative learning ability (Decl > 0). In this case, the positive interaction started at the beginning of training and was maintained over the course of practice. For below-average Decl values, higher Proc was associated to increasingly lower accuracy over the course of practice.

Figure 7.6 is a plot of the Exec:Spacing S1S2:Session interaction and shows that the already mentioned positive effect of spacing between Session 1 and Session 2 in the course of practice was also a function of the central executive. Specifically, higher central executive function enhanced the effect of spacing later in practice. Note that Spacing S1S2 could not have an effect on the accuracy of Block 1 (Session 1).

Hence, caution is needed in the interpretation of the interaction in the first stage of practice (Session = -1).

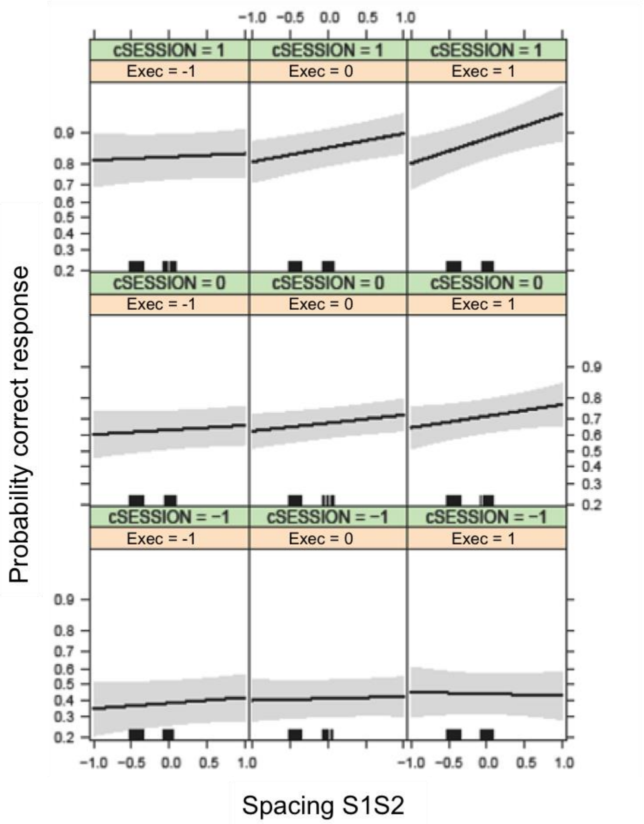


*Figure 7.6*. Effect of the Interaction Between Spacing S1S2 and Exec Across Practice.

***Interactions: Re-leveling of the Categorical Variable WO.*** The only β coefficient affected by the re-leveling of the WO variable was the coefficient of the Proc:WO interaction (Table 7.6b, next page; Figure 7.4). Setting WO to SOVs showed that Proc supported the accuracy of SOVa and OVa significantly less compared to the baseline. By contrast, setting WO to SOVa revealed a significant positive difference in the effects of Proc on accuracy in SV, SOVs and OVs trials.

Table 7.6b

*Summary of Significant Proc:WO Interaction Effects Depending on WO Reference Category*

| | WO reference category | | | | |
|---|---|---|---|---|---|
| | SV | SOVs | SOVa | OVs | OVa |
| Proc:SV | | | +* | | |
| Proc:SOVs | | | +** | | +* |
| Proc:SOVa | -* | -** | | -* | |
| Proc:OVs | | | +* | | |
| Proc:OVa | | -* | | | |

*Note*. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Overall, the re-leveling indicates a significant difference between trial types in the extent to which Proc supported or mitigated learning. With the exception of SV trials, the distinction differentiated between symmetric and asymmetric contexts.

***Model B: Symmetric Trials.*** As in the previous study two follow-up models were fitted. For the symmetric trials the most complex model justified by the data included fixed effects of Decl, Proc, Session, VocLearn, and Exec. The model also included a nonconditional VocLearn: Session interaction, two three-way interactions (Decl:Proc:Session and Proc:Exec:Session) and the two-way interactions conditional on them. The random effects included participants and items on intercepts, as well as effects of participants on the slopes of Decl and Session. Table 7.7 (next page) illustrates the model's output (CN = 1.99).

Table 7.7

*Generalized Mixed-Effects Model of the Effects of Declarative, Procedural, and Vocabulary Learning Ability, Session and Central Executive on Accuracy During Practice in Symmetric (Linking) Contexts (Model B) - $R^2$ = .56*

| | | | | 95% CI | | |
|---|---|---|---|---|---|---|
| Fixed effects | β | *SE* | *z* | lower | upper | *p* |
| (Intercept) | -0.36 | 0.20 | -1.82 | -0.76 | 0.03 | .068^ |
| **Decl** | **0.91** | **0.24** | **3.83** | **0.45** | **1.38** | **.000*** ** |
| Proc | 0.09 | 0.21 | 0.42 | -0.33 | 0.51 | .671 |
| **Session** | **1.12** | **0.20** | **5.57** | **0.73** | **1.52** | **.000*** ** |
| **VocLearn** | **0.56** | **0.12** | **4.73** | **0.33** | **0.79** | **.000*** ** |
| Exec | 0.05 | 0.14 | 0.41 | -0.21 | 0.32 | .645 |
| **Decl:Proc** | **0.83** | **0.37** | **2.24** | **0.10** | **1.55** | **.025*** |
| **Decl:Session** | **0.47** | **0.15** | **3.07** | **0.17** | **0.77** | **.002** ** |
| Proc:Session | -0.26 | 0.15 | -1.75 | -0.56 | 0.03 | .079^ |
| **Proc:Exec** | **-0.73** | **0.24** | **-3.00** | **-1.21** | **-0.25** | **.003** ** |
| Exec:Session | -0.00 | 0.10 | -0.06 | -0.20 | 0.19 | .956 |
| VocLearn:Session | 0.20 | 0.11 | 1.82 | -0.01 | 0.41 | .068^ |
| **Decl:Proc:Session** | **0.86** | **0.21** | **4.01** | **0.44** | **1.28** | **.000*** ** |
| **Proc:Exec:Session** | **-0.52** | **0.19** | **-2.65** | **-0.90** | **-0.13** | **.008** ** |

*Note*. ^$p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

***Main Effects (Model B).*** Decl ($p < .001$), Session and VocLearn ($p < .001$) significantly predicted accuracy (all effects conditional).

***Two-way interactions.*** A significant positive interaction was found between Decl and Proc ($p < .05$), and between Decl and Session ($p < .01$). The Proc:Exec interaction was negative and significant ($p < .01$). These conditional interactions largely confirm the findings in Model A.

***Three-way interactions.*** The model returned two significant three-way interactions, a positive Decl:Proc:Session interaction ($p < .001$) and a negative

Proc:Exec:Session interaction ($p < .01$). Both interactions emerged also in Model A, but Proc:Exec:Session was now returned as a significant predictor of accuracy (Figure 7.7).

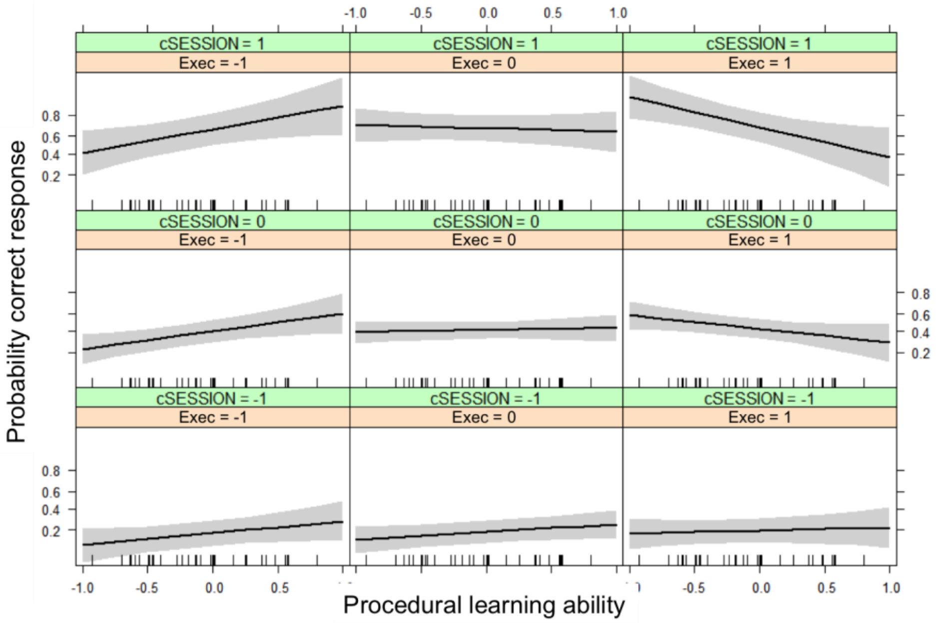


*Figure 7.7*. Effect of the Interaction Between Proc and Exec Across Practice (Symmetric Contexts).

As the plot shows, for average values of Exec (Exec = 0) the effect of the interaction was virtually null over the course of practice. For below-average values of Exec the Proc:Exec interaction contributed to accuracy, with the positive effect increasing over the course of practice. When Exec was above average it mitigated the effect of Proc, going from a lack of effect in the first phases of training to a markedly negative effect later in training. The plot relative to the Decl:Proc:Session interaction replicated the pattern in Figure 7.5 (Model A).

***Model C: Asymmetric Trials.*** A second follow-up model was fitted to the asymmetric dataset (CN = 1.98). The fixed effects in Model C included Decl, Proc, the spacing between Session 1 and Session 2 and (centered) Block.

The model also included two three-way interactions (Proc:VocLearn:Block and VocLearn:Block:S1S2) and all two-way interactions conditional on them. The only random effects included were effects of participants and trial items on intercepts. A summary of the model is provided in Table 7.8.

***Main Effects (Model C).*** The intercept's significant positive β coefficient indicates that trials that did not require learning of the relationships between morphosyntax and thematic interpretation were significantly more likely to be accurate. For Decl, Proc, VocLearn, spacing, and the effect of Block, the model presents a pattern of results that is overall very similar to the one found in the A and B models (all main effects conditional). The main effects of Decl ($p < .001$), Block ($p < .001$) and VocLearn ($p < .05$) were positive and significant. Similarly to Model A, the spacing between Session 1 and Session 2 was also a significant predictor of accuracy ($p < .01$).

Table 7.8

*Generalized Mixed-Effects Model of the Effects of Declarative, Procedural and Vocabulary Learning Ability, Block and Spacing During Practice in Asymmetric Contexts (Model C) -* $R^2 = .46$

| | | | | 95% CI | | |
|---|---|---|---|---|---|---|
| Fixed effects | β | *SE* | *z* | lower | upper | *p* |
| **(Intercept)** | **1.08** | **0.20** | **5.47** | **0.70** | **1.47** | **.000*****|
| **Decl** | **0.83** | **0.24** | **3.44** | **0.36** | **1.30** | **.000***** |
| Proc | 0.20 | 0.25 | 0.78 | -0.30 | 0.69 | .436 |
| **Block** | **0.39** | **0.08** | **4.56** | **0.22** | **0.56** | **.000***** |
| **VocLearn** | **0.40** | **0.19** | **2.14** | **0.03** | **0.76** | **.032*** |
| **Spac S1S2** | **0.55** | **0.18** | **3.12** | **0.20** | **0.90** | **.002**** |
| Proc:VocLearn | -0.10 | 0.34 | -0.29 | -0.76 | 0.56 | .768 |
| **Proc:Block** | **-0.15** | **0.07** | **-2.07** | **-0.30** | **-0.01** | **.038*** |
| **VocLearn:Block** | **0.18** | **0.05** | **3.72** | **0.09** | **0.28** | **.000***** |
| VocLearn:S1S2 | -0.04 | 0.27 | -0.15 | -0.57 | 0.49 | .879 |
| Block:S1S2 | 0.07 | 0.05 | 1.37 | -0.03 | 0.18 | .170 |
| **Proc:VocLearn:Block** | **0.25** | **0.10** | **2.49** | **0.05** | **0.44** | **.013*** |
| **VocLearn:Block:S1S2** | **0.21** | **0.08** | **2.65** | **0.05** | **0.36** | **.008**** |

*Note*. ^$p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

***Two-way Interactions.*** The model returned two significant conditional interactions, a negative interaction between Proc and Block ($p < .05$), and a positive interaction between VocLearn and Block ($p < .001$), respectively indicating an attenuation and an increasing of the effects of these variables on the likelihood that individual trials were correct over the course of practice.

***Three-way interactions.*** The significant positive three-way Proc:VocLearn:Block interaction ($p < .05$), an effect that did not emerge in Model A or B, is plotted in Figure 7.8. The plot shows that, as VocLearn increases, the effect of

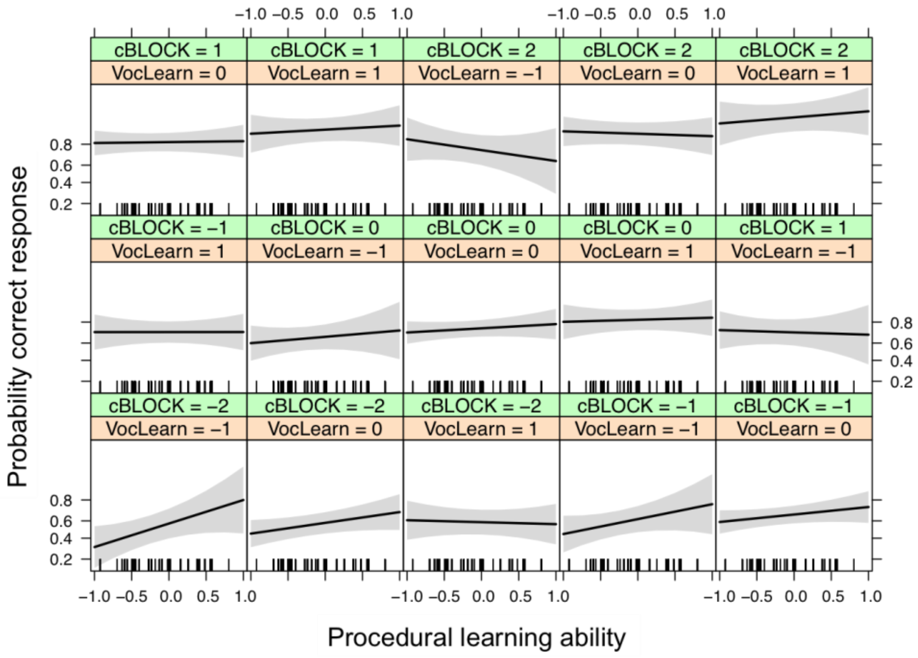


*Figure 7. 8*. Effect of the Proc:VocLearn Interaction Over the Course of Practice (Asymmetric Contexts).

Proc is mitigated early in practice. Note that, differently from the child dataset, higher VocLearn levels do not result in a negative Proc slope (cf. Figure 6.13). The effect of the interaction, that remains virtually null in the central part of practice, changes from negative to positive late in practice.

The plot of the VocLearn:Block:S1S2 interaction (Figure 7.9, next page) shows that the positive effect of spacing between Session 1 and Session 2 was greater for higher levels of VocLearn in the second part of practice. As mentioned above, since S1S2 spacing cannot have had any effect on learning in Block 1, effects in the first part of practice (for centered Block < 0) are difficult to interpret.

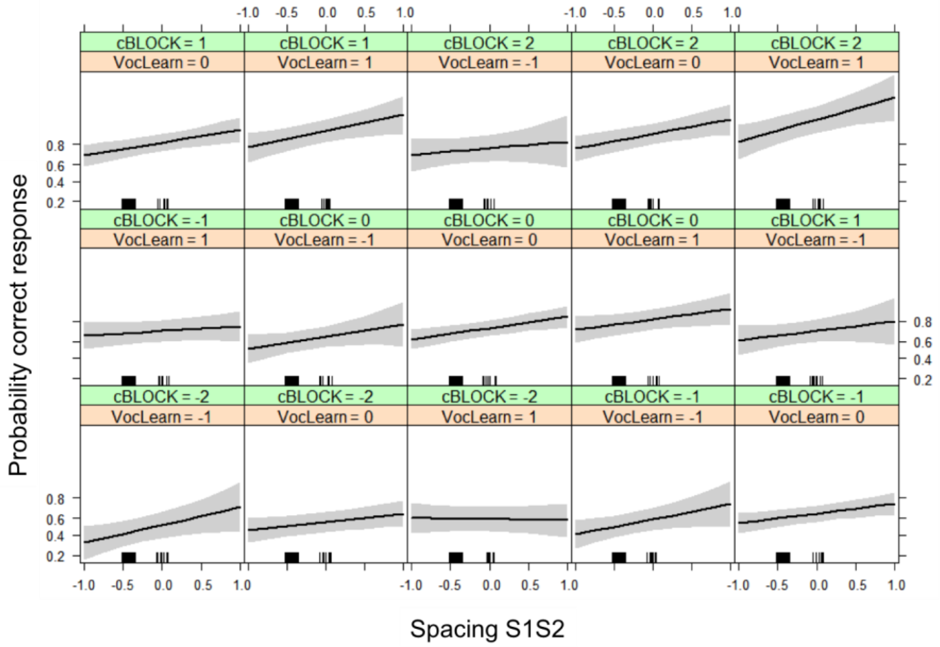


*Figure 7.9*. Effect of the Spacing S1S2:VocLearn Interaction Across Practice.

### 7.4.2 RQ1: Summary of results

The first research question sought to elucidate the extent to which declarative and procedural learning ability modulated adult aural L2 learning during practice in learning conditions that replicated those in the child experiment. In particular it aimed at elucidating the role of these abilities in the learning of rules linking syntax and semantics in the L2.

***Intercepts.*** The β coefficients for the different versions of Model A indicate significant differences in accuracy among trial types, with only SV, SOVa and OVa trials showing on average a significant likelihood of being correct. These findings were confirmed by the difference between the intercept effects in the symmetric dataset (model B, with a marginally significant negative coefficient) and the asymmetric dataset (model C, with a significantly positive coefficient). Taken together, these results indicate a significant advantage in trials where the learning of syntax-semantics relationships was not crucial for accuracy.

***Declarative Learning Ability.*** Decl had a highly significant effect on accuracy in all three models, which also significantly increased over the course of practice in Model A and B. In the overall and symmetric models (but not in the asymmetric model) there was also a significantly positive interaction between Decl and Proc for above average values of Decl, whilst for below-average values of Decl, the interaction was slightly positive early in practice, becoming negative as practice progressed.

***Procedural Learning ability.*** Proc was a term in significant interactions with Decl, VocLearn, Exec and type of sentence trial.

***Vocabulary Learning Ability.*** VocLearn was a highly significant positive effect in all three models. However, Model C was the only one that returned a significant increasing positive effect of VocLearn across practice and a significant positive three-way interaction with Proc and Block.

***Central Executive.*** Exec was returned as a term of a significant negative interaction with Proc in Models A and B. Exec was also found to interact with Spacing S1S2 and Session. The interaction became more positive for longer spacing and later in practice.

***Spacing between S1 and S2.*** Spacing between S1 and S2 was returned as a significant main positive effect in model C, where it also appeared in a significant positive interaction with VocLearn and Block. VocLearn was related to an increasingly positive effect of spacing on accuracy as training progressed, especially for participants with higher levels of VocLearn.

***Word Order.*** The effect of WO revealed more robust learning for asymmetric trials. Inspection of Model A and the relevant re-leveled models indicated that, with the exception of SV trials, the effect of Proc was significantly higher if trials were symmetric.

***Block/Session.*** The effect of practice (likely combined with the effects of sleep and memory consolidation), as reflected by the Block or the Session variables, was positive and highly significant in all three models. The asymmetric subset was the only one where the effect of Block was comparatively superior to the effect of Session.

### 7.4.3 RQ2

The second research question investigated the relationship between declarative/procedural learning ability and accuracy in a grammaticality judgment test (GJT) administered at the end of the game practice that probed learning of word order and case marking.

RQ2 To what extent do declarative and procedural learning ability modulate adult L2 learning of word order and case marking as measured by a grammaticality judgment test administered at the end of practice?

***Descriptive Statistics.*** Prior to analysis the practice sentences (144 items) and the missing cases (41 items) were excluded from the dataset, leaving a total of 1008 valid cases. Out of 28 trials the accuracy was $M(SD) = 19(3.2)$, $SE = 0.5$. Table 7.9 (next page) shows the aggregated counts of accurate and inaccurate trials according to different trial categories, and for each one whether learning was on average significantly above chance. Accurate GJT scores were operationalized in the same way as in Study 1 (see 6.4.3).

On average, performance was significantly above chance for all categories except for ungrammatical sentences with case violations, where performance was below chance. In categories where learning was significantly above chance, small to

Table 7.9

*Aggregate Count of Valid Trials in the GJT and Accuracy Significance Above Chance (50%)*

| | Acc | % | Inacc | Significance above chance | | |
|---|---|---|---|---|---|---|
| | | | | $\chi^2$ | *p* | $\Phi_v$ |
| **Overall** | **665** | **68.8** | **302** | **70.65** | **.000** | **.191** |
| **Gramm** | **378** | **78.1** | **106** | **82.98** | **.000** | **.293** |
| **Ungramm** | **287** | **59.4** | **196** | **8.66** | **.003** | **.095** |
| **SV** | **97** | **70.8** | **40** | **12.43** | **.000** | **.213** |
| **SOV** | **401** | **72.6** | **151** | **59.67** | **.000** | **.232** |
| **OV** | **167** | **60.1** | **111** | **5.70** | **.017** | **.101** |
| Case viol. | 87 | 41.8 | 121 | below chance | | |
| **WO viol.** | **200** | **72.7** | **75** | **30.00** | **.000** | **.233** |

*Note*. With reference to the model reported in Table 7.10, the categories SV, SOV and OV correspond to the factors of the categorical variable Word Order (WO), whilst the categories Gramm, Case viol. and WO viol. correspond to the factors of the categorical variable Type of Violation (Viola).

medium effect sizes (Cramer *V*) were returned for grammatical sentences, SV and SOV sentences, and ungrammatical sentences with word order violations. In the case of ungrammatical sentences and OV sentences the effect size was small (Table 7.9).

***GJT Model.*** After controlling that multicollinearity was not an issue (CN = 1.38), a generalized binomial linear regression model was derived according to the criteria presented in 6.3.12. The dependent variable was the log odds that individual trials were accurate and the independent variables included the two predictors of interest (Decl and Proc), the categorical variable word order (WO) with three levels (SV, SOV, and OV), the categorical variable type of violation (Viola) with three levels (grammatical, case, and word order), and the measure of the phonological loop (PhonLoop). All continuous variables were standardized prior to analysis.

The interactions included Proc:Viola, Proc:PhonLoop, and PhonLoop:Viola. The evaluation found random effects of participants and items on intercepts and of participants on the slopes of type of violation (Table 7.10, next page).

Table 7.10

*Generalized Mixed-Effects Model of the Effects of Declarative and Procedural Learning Ability, Word Order, Type of Violation and Phonological Loop on Accurate Response in the GJT (GJT Model) - $R^2 = .43$*

| | | | | 95% CI | | |
|---|---|---|---|---|---|---|
| Fixed effects | β | *SE* | *z* | lower | upper | *p* |
| **(Intercept)** | **-1.04** | **0.43** | **-2.42** | **-1.87** | **-0.20** | **.015*** |
| **Decl** | **0.35** | **0.16** | **2.21** | **0.04** | **0.66** | **.027*** |
| **Proc** | **-0.67** | **0.28** | **-2.36** | **-1.22** | **-0.11** | **.018*** |
| **WO (SOV)** | **0.87** | **0.37** | **2.35** | **0.14** | **1.59** | **.018*** |
| WO (SV) | 0.05 | 0.53 | 0.09 | -0.99 | 1.09 | .924 |
| **Gramm** | **2.18** | **0.47** | **4.62** | **1.26** | **3.11** | **.000***** |
| **Viola (WO)** | **2.05** | **0.54** | **3.80** | **0.99** | **3.11** | **.000***** |
| PhonLoop | -0.19 | 0.20 | -0.95 | -0.57 | 0.20 | .339 |
| Proc:Gramm | 0.66 | 0.45 | 1.47 | -0.22 | 1.55 | .140 |
| Proc:Viola(WO) | 0.17 | 0.50 | 0.33 | -0.82 | 1.15 | .741 |
| **Proc:PhonLoop** | **-0.34** | **0.17** | **-1.96** | **-0.68** | **-0.00** | **.049*** |
| Gramm:PhonLoop | 0.49 | 0.28 | 1.70 | -0.07 | 1.05 | .088^ |
| Viola(WO): P.Loop | +0.00 | 0.32 | 0.01 | -0.63 | 0.64 | .992 |

*Note*. $^{\wedge}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

***Main Effects (GJT Model).*** The model reported in Table 7.10 indicates coefficients relative to WO = OV and Viola = case violation. The negative intercept ($p < .05$) indicates a significant likelihood that individual baseline trials were incorrect. The effects relative to the other levels of the WO and Viola factors also show that sentences with word order violations and grammatical sentences were significantly more accurate than sentences with case violations ($p < .001$, Figure 7.10) and that SOV trials were significantly more accurate than OV trials ($p < .05$, Figure 7.11).

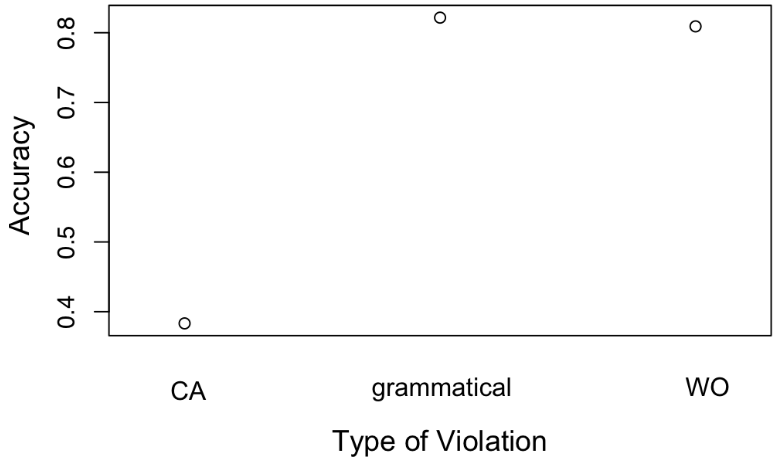


*Figure 7.10*. Effects of the Three Levels of Type of Violation (Viola) on GJT Accuracy.

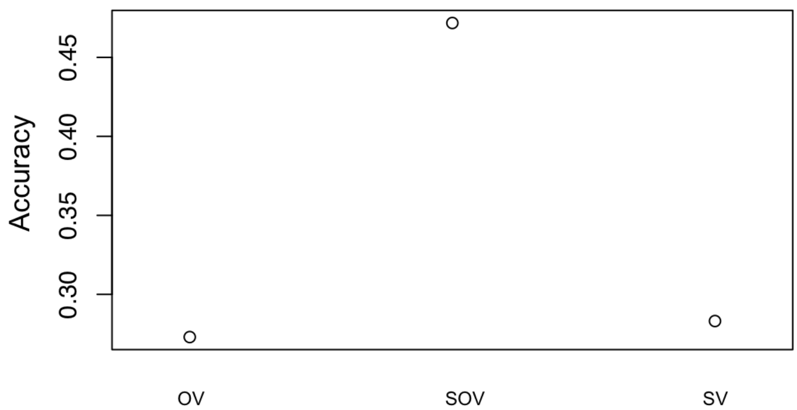


*Figure 7.11*. Effect of Type of Word Order (WO) on GJT Accuracy.

With regards to the predictors of interest, the model showed that Decl was a significant positive predictor of accuracy ($p < .05$), whilst Proc was a significant negative predictor ($p < .05$).

***Main Effects: Re-leveling of the Categorical Variables WO and Viola.*** Re-leveling of WO and Viola produced eight additional versions of Model A (cf. Table 7.11a, end of section 7.4.4; Appendix J). For WO = OV and Viola = Gramm (no violation) the model returned a significant positive intercept and a positive effect for the SOV level, indicating that grammatical SOV sentences were significantly more accurate than grammatical OV sentences. Among OV sentences, case violations where significantly less accurate than grammatical sentences. Overall, the picture that emerged was that, independently of grammaticality and type of violation, SOV sentences were on average significantly more accurate than OV sentences. Finally, independently of type of word order, trials with case violations were significantly less accurate than trials with word order violations or grammatical trials.

***Interactions (GJT Model).*** Going back to the initial model (Table 7.10), the only significant interaction emerging is Proc:PhonLoop. As Figure 7.12 illustrates (next page), the effect of Proc on the likelihood that a response was correct was positive for short phonological loop spans and became increasingly negative for longer spans.

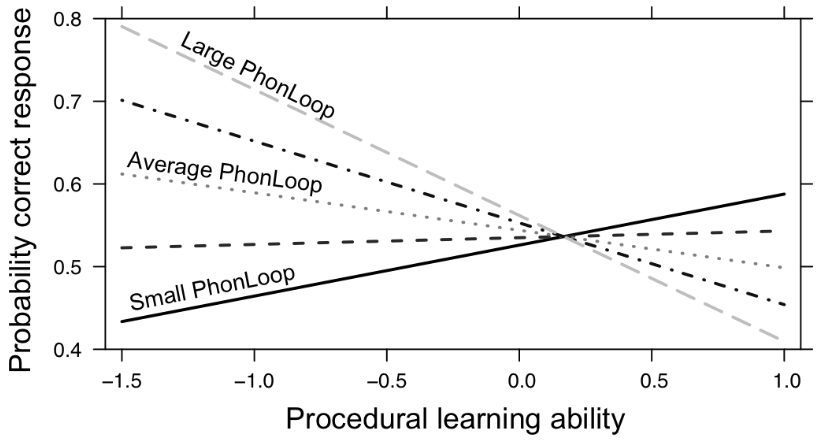


*Figure 7.12*. Effect of the Proc:PhonLoop Interaction on Accuracy in the GJT.

### 7.4.4 RQ2: Summary of results

**Intercepts.** The intercepts' β coefficients in the adult GJT model indicated significantly better accuracy with grammatical trials and ungrammatical word order trials, compared to ungrammatical case trials.

***Declarative Learning Ability.*** There was a significant positive main effect of Decl on accuracy.

***Procedural Learning Ability.*** Proc was a significant negative predictor of accuracy for ungrammatical trials with case violations, but not for grammatical trials or ungrammatical trials with word order violations. Also, the Proc:PhonLoop interaction was a significant negative predictor of accuracy.

***Word Order and Type of Violation.*** The model showed that, independently of grammaticality and type of violation, accuracy was significantly better in SOV compared to OV trials. With regards to Type of Violation, grammatical sentences and ungrammatical sentences with word order violations were highly significantly more accurate than ungrammatical sentences with case violations.

Table 7.11a

*Summary of Intercept, Proc, PhonLoop, Word Order and Type of Violation Significant Main Effects Depending on Word Order and Type of Violation Reference Category (GJT Model)*

| | Reference category combination for Word order + Type of Violation | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | OV + case | OV + gramm | OV + word order | SOV + case | SOV + gramm | SOV + word order | SV + case | SV + gramm | SV + word order |
| Intercept | -* | +** | +* | | +*** | +*** | (-^) | +* | +* |
| Proc | -* | | | -* | | | (-*) | | |
| PhonLoop | | | | | | | | | |
| SOV | +* | +* | +* | | | | (+^) | +^ | +^ |
| SV | | | | (-^) | -^ | -^ | | | |
| OV | | | | -* | -* | -* | | | |
| gramm | +*** | | | +*** | | | (+***) | | |
| word order violation | +*** | | | +*** | | | (+***) | | |
| case violation | | -*** | -*** | | -*** | -*** | | -*** | -*** |

*Note*. ^$p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 7.11b

*Summary of Proc:Type of Violation and PhonLoop:Type of Violation Marginally Significant Interactions Depending on Word Order and Type of Violation Reference Category (GJT Model)*

| | Reference category combination for Word order + Type of Violation | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | OV + case | OV + gramm | OV + word order | SOV + case | SOV + gramm | SOV + word order | SV + case | SV + gramm | SV + word order |
| Proc:gramm | | | | | | | | | |
| Proc:word order | | | | | | | | | |
| Proc:case | | | | | | | | | |
| PLoop:gramm | +^ | | | +^ | | | (+^) | | |
| PLoop:word order | | | | | | | | | |
| PLoop:case | | -^ | | | -^ | | | -^ | |

*Note*. ^$p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

### 7.4.5 RQ3

The third research question aimed at elucidating whether adults developed explicit or/and implicit knowledge of BrocantoJ morphosyntax during the experiment.

RQ3 To what extent is the L2 knowledge acquired by the adults implicit/ explicit?

As in the child study confidence ratings and verbal reports were analyzed to investigate the research question.

***Descriptive Statistics.*** Prior to analysis practice and missing items were removed from the GJT dataset and a count overview relative to the accuracy of valid experimental trials in different confidence categories was computed for descriptive statistics purposes (Table 7.12). However, missing data were included in the subsequent mixed effects model analysis. Confidence was operationalized in the same way as in Study 1 (see 6.4.5). There was a marked prevalence of high confidence accurate trials, confirmed by the values of mean and standard deviation of the Conf variable when treated as continuous ($M$ = 7.31; $SD$ = 0.8, range 5-8).

Table 7.12

*Count of Valid Experimental Trials in the GJT According to Level of Confidence*

| Confidence Level | *n* |
|---|---|
| High | 457 |
| Medium High | 365 |
| Medium Low | 124 |
| Low | 18 |
| Total Valid | 967 |
| Missing | 41 |
| Total | 1008 |

***Guessing and Zero-Correlation Criteria.*** The guessing criterion was evaluated similarly to Study 1 (see 6.4.5 for details). Again, a two-level categorical version of the Conf variable (high and low confidence) was derived and a Chi-square test was run to assess accuracy above chance for each category. The results showed that, although low confidence items were on average accurate above chance, the difference with chance performance was not significant (Table 7.13). As such, the guessing criterion did not provide conclusive evidence of the availability of implicit judgment knowledge in the adult participants. Next, the zero-correlation criterion was investigated by fitting a regression model (cf. 6.4.5).

Table 7.13

*Count of Valid Experimental Accurate and Inaccurate Trials in the GJT According to Level of Confidence and Accuracy Significance Above Chance (50%)*

| | Acc | % | Inacc | Significance above chance | | |
|---|---|---|---|---|---|---|
| | | | | $\chi^2$ | *p* | $\Phi_v$ |
| **High Confidence** | **581** | **70.8** | **239** | **74.56** | **.000** | **.213** |
| Low Confidence | 79 | 56.4 | 61 | 1.16 | .281 | .064 |

***Confidence Level Model.*** The model included the main effects of the judgment confidence (Conf, continuous and standardized) Viola (with three levels, grammatical, word order violation, and case violation). Decl and Proc were not included in the final equation because they did not improve the model statistically significantly as main effects. The model output for Viola = Case is illustrated in Table 7.14 (next page).

***Main Effects (Confidence Level Model).*** Confidence was not a significant predictor of accuracy but lack of a significant positive correlation between confidence and accuracy cannot be taken as evidence of the availability of implicit judgment knowledge, because learning in ungrammatical case trials was not significantly above chance (cf. 7.4.1). The significant positive effects of the grammaticality and word order levels of Viola ($p < .001$) confirm lower accuracy in ungrammatical case trials compared to other trial types.

Table 7.14

*Generalized Mixed-Effects Model of the Effect of Judgment Confidence and Type of Violation on Accuracy in the GJT (Confidence Level Model) - $R^2 = .32$*

| | | | | 95% CI | | |
|---|---|---|---|---|---|---|
| Fixed effects | β | *SE* | *z* | lower | upper | *p* |
| (Intercept) | -0.39 | 0.32 | -1.24 | -1.02 | 0.23 | .214 |
| Conf | -0.24 | 0.16 | -1.53 | -0.56 | 0.07 | .125 |
| **Gramm** | **2.02** | **0.38** | **5.31** | **1.27** | **2.76** | **.000*** ** |
| **Viola (WO)** | **1.57** | **0.42** | **3.77** | **0.76** | **2.39** | **.000*** ** |
| **Conf:Gramm** | **1.07** | **0.20** | **5.33** | **0.68** | **1.47** | **.000*** ** |
| Conf:Viola(WO) | 0.41 | 0.22 | 1.89 | -0.01 | 0.84 | .059^ |

*Note*. ^$p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

***Main Effects: Re-leveling of the Categorical Variable Viola.*** For Viola = WO (Table 7.15a; Appendix K) the model returned a significant likelihood that trials

Table 7.15a

*Summary of Significant Main Effects on Accuracy in the GJT Depending on Viola Reference Category (Confidence Level Model)*

| | Type of violation reference category | | |
|---|---|---|---|
| | CA | WO | Gramm |
| Intercept | | +*** | +*** |
| Conf | | | +*** |
| Viola(CA) | | -*** | -*** |
| Viola(WO) | +*** | | |
| Viola(Gramm) | +*** | | |

*Note*. ^$p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

were accurate ($p < .001$). Importantly, for Viola = WO, confidence did not predict accuracy. Given that ungrammatical word order trials were accurate significantly above chance, the zero-correlation criterion indicates that knowledge of word order judgments was largely implicit in adults.

For Viola = Gramm, both the effects of the intercept and the Conf variable were returned as positive and significant ($p < .001$). According to the zero-correlation criterion, these results indicate the involvement of explicit knowledge in the judgment of grammatical sentences.

***Two-way Interactions and re-leveling.*** The model in Table 7.14 (Viola = case violation) returned a significant positive coefficient for Conf:Viola (Gramm), indicating a significant more positive role of judgment confidence in grammatical sentences compared to the baseline (see also Table 7.15b; Appendix K).

Table 7.15b

*Summary of Significant Interaction Effects on Accuracy in the GJT Depending on Viola Reference Category (Confidence Level Model)*

| | Type of violation reference category | | |
|---|---|---|---|
| | CA | WO | Gramm |
| Conf:Viola(CA) | | -^ | -*** |
| Conf:Viola(Gramm) | +*** | +*** | |
| Conf:Viola(WO) | +^ | | -*** |

*Note*. ^$p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

***Debriefing Questionnaire.*** The same two raters as in the child study coded all questionnaires (agreement on awareness of word order was 100%). The mean, median and mode relative to the awareness of word order as reported in the debriefing questionnaires were computed (*M* [*SD*] = 3.39 [1.5]; *Mdn* = 3.00; *Mo* = 5.00; range 1-

6). The descriptive data (Table 7.16) show that only 10 verbal reports out of 36 (27.8%) revealed a very low to low awareness of the structural properties of the

Table 7.16

*Raw Score Frequencies in the Debriefing Questionnaire for Syntax (N = 36)*

| Scores | Frequency | Percent | Cumulative Percent |
|---|---|---|---|
| 1 | 5 | 13.9 | 13.9 |
| 2 | 5 | 13.9 | 27.8 |
| 3 | 11 | 30.6 | 58.3 |
| 4 | 1 | 2.8 | 61.1 |
| 5 | 14 | 38.8 | 100.0 |
| 6 | 0 | 0 | 100.0 |
| Tot | 36 | 100.0 | |

language (scores 1 and 2). Most reports (26 out of 36) contained explicit reference to regularities in the word order pattern. Eleven participants were assigned score 3, corresponding to 'reports noticing the presence/absence of a single specific word and refers to its position in the sentence'; one participant was assigned score 4, corresponding to 'reports that there is an order involving domains larger than a single word but does not provide examples'; and 14 participants were assigned score 5, corresponding to 'reports that there is an order in domains larger than a single word and provides examples'. None of the reports contained an example of a complete and well-formed BrocantoJ sentence (Appendix E).

The explicit language knowledge scores relative to word order were standardized and correlated to year of schooling, to the measures of cognitive ability and to the measures of language learning. None of the relationships was significant.

Given that, unlike children, adult participants mentioned making hypotheses about the meaning and function of the *ri*/*ru* particles and the thematic interpretations of NPs, this information was also coded and scored according to the criteria presented in the methods section (*M* [*SD*] = 1.91 [1.2]; *Mdn* = 2.00; *Mo* = 3.00; range 0-3; Table 7.17). Raters initially agreed on 94% of the scores, and reached a consensus for the remaining cases.

Table 7.17

*Raw Score Frequencies in the Questionnaire for Ri/Ru Interpretation (N = 36)*

| Scores | Frequency | Percent | Cumulative Percent |
|---|---|---|---|
| 0 | 8 | 22.2 | 22.2 |
| 1 | 3 | 8.3 | 30.6 |
| 2 | 9 | 25.0 | 55.6 |
| 3 | 16 | 44.5 | 100.0 |
| Tot | 36 | 100.0 | |

The verbal reports relative to one quarter of the participants (8 out of 36, corresponding to 22.2%) made no reference to *ri*/*ru* or to the NPs' thematic interpretation (these corresponded to a score of zero = "does not mention *ri*/*ru* or suggest a thematic interpretation for case particles or NPs").

The remaining participants (24 out of 36) mentioned noticing *ri*/*ru*, with a subset of them providing hypotheses about the particles' function. Three participants were assigned score 1, corresponding to 'mentions *ri*/*ru* but does not suggest a thematic interpretation for case particles or NPs', 9 participants were assigned score 2, corresponding to 'suggests nonthematic interpretation for *ri*/*ru*', and 16 participants were assigned score 3, corresponding to 'suggests thematic interpretation for *ri*/*ru* and/or for the NPs' (cf. Appendix E for examples).

Nonthematic interpretations for the particles' functions included that they were connectives or subjunctions, that they were locative adverbs (e.g., up/down), auxiliaries of the verb, or had an ordinal meaning (e.g., first/second). Hypotheses about the function of *ri*/*ru* that explicitly referred to thematic interpretation for the particles or the associated NPs included: (a) that the first NP or *ri* had an agentive meaning, (b) that the second NP or *ru* indicated the 'theme', the element the action was 'done to' or had a passive interpretation, (c) that *ri*/*ru* or the NPs reflected a 'subject/object' or 'agent/patient' pattern. None of the participants provided a complete and correct description of the linking rules of BrocantoJ. The amount and type of metalanguage used to describe grammatical relationships varied across participants but its analysis is beyond the scope of the present dissertation.

The scores relative to the explicit knowledge of thematic relationships were also standardized and correlated to year of schooling, to the measures of cognitive ability and to the measures of language learning attainment. None of the relationships was significant.

### 7.4.6 RQ3: Summary of results

***Intercepts and Type of Violation.*** Reflecting the pattern already found for children in Study 1, the confidence level model indicated that (for average values of the continuous predictors) the likelihood of accuracy in the GJT was positive and significant in grammatical trials and in ungrammatical word order trials and not significantly negative in ungrammatical case trials. Both grammatical and ungrammatical word order trials were significantly more accurate than case violation trials.

***Confidence.*** In this model Conf significantly positively correlated with accuracy only in grammatical trials. There was a negative nonsignificant correlation with accuracy in case violation contexts and positive nonsignificant correlation for ungrammatical word order trials. According to the zero-correlation criterion, these results are compatible with the interpretation that implicit judgment knowledge played a role in the accurate judgment of ungrammatical word order patterns. Under the assumption that implicit structural (linguistic) knowledge can be inferred from implicit judgment knowledge, the result also means that, at the time of the GJT, knowledge of word order was also largely implicit.

The zero-correlation criterion also yields that accurate performance on grammatical trials was mainly related to explicit judgment knowledge. However, in this case no conclusions regarding the nature of structural (linguistic) knowledge can be drawn. Finally, the criterion does not apply to case violations trials because they were not learned above chance.

The difference between grammatical trials and ungrammatical trials with respect to the role of judgment confidence is reflected also in the effects of the Conf:Viola interaction, with a significant positive advantage for grammatical trials compared to ungrammatical trials.

***Explicit Language Knowledge.*** The results of the debriefing questionnaire relative to the explicit knowledge of word order found that most participants made explicit reference to the word order properties of the language. Only 10 participants did not go beyond noticing the occurrence/absence of specific words in the BrocantoJ sentences.

Regarding explicit knowledge of linking rules, slightly less than half the participants assigned the *ri*/*ru* particles or the NPs in BrocantoJ some kind of thematic interpretation. A quarter of the participants reported alternative, nonthematic interpretations.

# 8. Discussion

## 8.1 Discussion of Study 1

### 8.1.1 L2 practice

***L2 Attainment and Main Effects***. This was the first L2 training study to investigate the role of long-term memory (LTM) abilities in children and also one of the first to offer a fine-grained analysis of the role of LTM abilities during L2 practice (see Pili-Moss & Morgan-Short, 2017; Pili-Moss et al., 2018; Suzuki, 2017b). The first research question in the child study asked to what extent learning abilities that depend on long-term memory functions modulate L2 aural learning in children and in particular learning of the rules linking morphosyntax and semantic interpretation. In order to answer this question the study analyzed data relative to a computer game where accurate performance depended on the children's aural comprehension of a novel miniature language and, in a subset of cases, on the understanding of the linking between an NP's syntactic position and its thematic function, or an NP's case marker and its thematic function.

The descriptive accuracy data indicated that children learned the language significantly above chance, independently of the type of sentence stimuli and of whether the stimuli required the understanding of linking rules (including when these depended solely on the case marker, as in OV symmetric sentences). However effect sizes indicated that learning in OVs trials and in symmetric trials in general was less robust compared to other categories (Cramer $V < .100$).

A model of the overall dataset confirmed this and revealed that in trials that did not require the understanding of linking rules (asymmetric) learning was significantly better than in trials that did (symmetric). However, the significantly above-chance learning of symmetric trials indicates that there was some learning of rules linking morphosyntax and thematic interpretation and confirms findings in Casenhiser and Goldberg (2012), Pili-Moss (2017) and Wonnacott et al. (2012). Further, the difference between accuracy in SOVs and OVs trials suggests that linking mainly relied on word order, rather than morphology. Three mixed- effects models were fitted to investigate the first question; an overall model, and two follow-up models analyzing the symmetric and asymmetric subsets respectively.

The present study considered the extent to which visual/aural associations (vocabulary learning ability) were retained as an additional index of declarative learning ability (cf. 4.5.2). In the present discussion I will use the plural form *LTM (long-term memory) declarative (learning) abilities* to collectively refer to declarative learning ability (Decl) and vocabulary learning ability (VocLearn). The findings point to two main phenomena emerging in the early stages of child L2 learning: (a) a significant positive role of abilities linked to LTM declarative abilities (declarative and vocabulary learning ability) accompanied by a significantly increasing positive effect of procedural learning ability across practice; and (b) an initial competitive relationship between LTM declarative abilities and abilities related to procedural long-term memory that decreases as a function of practice.

Analysis of the overall model and of the asymmetric model in particular indicated that declarative learning ability (Decl) had a significant effect on accuracy only in asymmetric trials, whereas no significant main effect of procedural learning ability emerged. However, whilst the effect of declarative learning ability did not increase over time, there was a significant positive increase in the effect of procedural learning ability on accuracy in the overall model and in the asymmetric model over the course of practice.

The analysis found that in the asymmetric model vocabulary learning ability had an overall positive highly significant effect that increased with practice and that it was more robust compared to that of declarative learning ability (Decl). Overall the study confirms the findings of previous behavioral studies with adults (with the exception of Carpenter, 2008) indicating that abilities related to declarative memory are a significant predictor of learning early in practice and at low levels of proficiency (e.g., Brill-Schuetz & Morgan-Short, 2014; Hamrick, 2015; Morgan-Short et al., 2014). This result, now confirmed for child data, is compatible with the predictions of the DP model and other models based on bipartite representations of long-term memory (N. Ellis, 2004; Paradis, 2009; Ullman, 2004, 2005, 2015, 2016) and with the characteristics of declarative memory as a long-term memory system capable of fast learning after minimal exposure. In terms of the DP model, the fact that low amounts of training are related to a bias towards declarative memory constitutes an example of the 'see-saw' effect, a pattern of brain activity for which the engagement of one of the two memory systems occurs in tandem with the attenuation of the other. Not in contradiction with this account, other cognitive models of language learning (e.g., Skill Acquisition Theory) predict a pivotal role for declarative knowledge early in practice (DeKeyser, 2015).

A positive effect of procedural learning ability on learning later in practice is also predicted by the DP model and has been found in previous adult studies where L2 practice was significantly longer compared to Study 1 (e.g., Carpenter, 2008; Morgan-Short et al., 2014). However, Pili-Moss et al. (2018), the only other adult study that has examined LTM cognitive abilities during practice found that declarative learning ability was the only significant predictor in L2 comprehension throughout. In the light of previous/ongoing research, the findings of Study 1 suggest that procedural learning ability may engage earlier in L2 practice in children compared to adults.

A further point common to some cognitive models of L2 learning for which the distinction between declarative and procedural memory/knowledge is relevant (e.g., N. Ellis, 2004; DeKeyser, 2015; Ullman, 2015, 2016), is the explicitly envisaged or implied possibility of some degree of interaction between the two systems (this is not always the case for theoretical models referring to the distinction between implicit and explicit knowledge). Memory systems "are presumed to be activated simultaneously and in parallel" (Packard & Goodman, 2013, p. 1045) but interactions between them during learning, and specifically competitive interactions, have been observed in a number of neuroimaging studies with human participants as well as in many behavioral studies with lower animals, typically rodents (for a review, see Packard & Goodman, 2013). It has also been observed that prominence of one of the two memory systems during learning or performance can be associated to inhibition of the neural pathways of the memory system that is less engaged.

The prevalence of one or the other memory system has been associated to a number of factors, each one contributing to the compound effect shaping the pattern of greater reliance on one or the other system. These can be external (amount of training, environmental factors, spacing, etc.) or endogenous (lesions/attenuation of one of the two memory systems, emotional factors, individual differences etc.). Together with Study 2, the present study is the first L2 learning study that found behavioral evidence of an interaction between LTM learning abilities. Pili-Moss & Morgan-Short (2018) also found a partially co-operative, partially competitive interaction between declarative and procedural learning ability, but in that case the interaction had an effect on L2 automatization, not accuracy.

The asymmetric model returned a significant three-way interaction between declarative learning ability, vocabulary learning ability and procedural learning ability. For ease of exposition let us consider the two components of this interaction separately. As returned by the model, declarative and vocabulary learning ability significantly positively interacted, a synergic effect not unexpected, since different and complementary aspects of declarative memory are likely to be simultaneously engaged and supporting each other in the initial phases of learning.

The second component was a significant negative interaction between vocabulary learning ability and procedural learning ability[1]. To illustrate the effect of the negative interaction in the context of Study 1 let us imagine two learners with comparable levels of procedural learning ability, but with different declarative LTM abilities. Given the significance and magnitude of declarative LTM abilities as main effects in the model, the learning outcome will be mostly affected by how high their level is. However, in terms of accuracy, the negative interaction (a product term) will be comparatively more 'costly' for the learner with higher declarative LTM abilities.

This pattern of results is not incompatible with a situation in which a highly engaged declarative memory system (typical of the early stages of learning) could interfere with/inhibit the procedural memory system resulting in less efficient learning. Note that the interaction does not *per se* provide information about the direction of the effect, and it could as well be that the procedural memory system interferes with declarative memory (for discussion on this point see also Morgan-Short et al., 2014, p. 67).

---

1 Declarative learning ability and procedural learning ability were also inversely related but not significantly.

The data also showed that as training progressed there was a significant increase of the main effect of procedural learning ability together with a significant decrease in the negative value of the interaction between vocabulary learning ability and procedural learning ability. In terms of the previous interaction scenario the mitigation of the negative interaction could be explained with a gradual weakening of the inhibition on the procedural memory system, due to the fact that memory systems tend to be engaged in a more balanced way as practice progresses. Alternatively, a weaker negative interaction could be the result of a diminished interference on the part of the procedural memory system.

Beyond the general role played by separate memory systems in L2 learning, a central idea of the DP model as outlined in Ullman (2004, 2005, 2015, 2016) is that the engagement of declarative and procedural memory in language learning should have consequences for the type of linguistic targets that are acquired (cf., Chapter 1). For example, Antoniou et al. (2016) recently found that procedural learning ability predicted learning of simple rules (affixation), whilst declarative memory predicted learning of complex rules (affixation + vowel harmony). In terms of the extent to which different cognitive abilities supported the learning of specific linguistic targets, the present analysis found evidence of an overall increasing effect of procedural learning ability in contexts where accurate performance did not require learning of the underlying linking rule, i.e. in trials displaying comparatively less semantic complexity.

By contrast, the positive effect of vocabulary learning ability was highly significant in both symmetric and asymmetric trials, whilst declarative learning ability (as measured by recollection tasks) also mainly supported learning of asymmetric sentences. The fact that vocabulary learning ability had a significant effect on the accuracy of trials where semantic interpretation was pivotal, whilst procedural memory did not, would be in line with the view that declarative LTM abilities support the learning of semantics-related aspects of language.

On the other hand, it is not clear why the effect of declarative learning ability was not significant in the symmetric trials, which were arguably the trials where the role of semantics was comparatively more prominent due to form-meaning linking. One possibility is that, in the case of children, the amount of practice with symmetric stimuli was not sufficient for declarative memory to be fully engaged in language processing. In other words, although symmetric trials were learned above chance, they were comparatively more difficult and the amount of exposure was not sufficient for the effects of declarative learning ability to emerge significantly in the child symmetric dataset. Morgan-Short et al. (2014) reached a similar conclusion discussing the lack of declarative learning ability effects found early in practice in Carpenter (2008).

With regards to symmetric trials, it is interesting to note that the effect of vocabulary learning ability was significant for all trial types except OV symmetric trials. In particular there was a significant difference between the effect of vocabulary learning ability in SOVs trials and OVs trials, with a significant advantage for the former. In sum, although the model did not return significant differences in the accuracy of SOVs and OVs trials, it appears that the full benefits of robust vocabulary learning could be reaped during practice only if encoding of linking rules through word order was possible. This indirectly indicates a significantly more robust learning of word-order-based linking compared to morphological linking.

As for additional effects that emerged in the analysis of the first research question, there was (a) a significant effect of practice on accuracy (operationalized as either the effect of Block or the effect of Session), (b) a positive, but not significant, effect of Year, and (c) a negative moderating effect of Year on declarative learning ability that increased later in practice. Whilst the positive effects of practice and Year for proficiency are expected, it is not clear why the effect of declarative learning ability should be more negative as school grade and practice increase. Also, no significant effects for the measure of the phonological loop were found (cf. 8.3.2).

### 8.1.2 GJT evidence

The second research question investigated the role of cognitive abilities in child learning of word order and case patterns in BrocantoJ. The descriptive statistics relative to the GJT data showed that overall, and for grammatical and ungrammatical sentences separately, learning was above chance, although OV sentences were not learnt significantly above chance and sentences with case violations were below chance. The GJT data confirmed the picture returned by the practice data, i.e. that learning of case markers was overall limited. The finding that case markers were not learnt above chance is at odds with previous child studies that looked at L2 case learning with comparable length of exposure (e.g., Braine et al., 1990), but confirms results of other studies (e.g., Ferman & Karni, 2010), that found below-chance learning of affixes whose distribution followed a semantic rule.

With regards to the role of cognitive abilities, the clearest result emerging from the analysis was that procedural learning ability was a significant predictor of judgment accuracy in grammatical and ungrammatical OV sentences, whereas, contrary to the practice dataset, no significant effect of declarative or vocabulary learning ability emerged independently of type of trial. A possible explanation for the diminished effect of declarative LTM abilities lies in the requirements of the task used to measure accuracy in the GJT. Whilst accuracy in the game environment implied the ability to use the language in a meaningful context, the GJT was far less context-related, since exposure to the language did not force semantic processing. Consequently, it has to be expected that abilities related to declarative memory played a more limited role in the GJT compared to the game environment.

When the amount of training and L2 proficiency are taken into consideration, the GJT results in the present study contrast with the findings of adult BROCANTO2 behavioral studies (Brill-Schuetz & Morgan-Short, 2014; Carpenter, 2008; Morgan-Short et al., 2014), pointing at an important difference between children and adults. Similarly to previous research, the present study found that procedural learning ability was a significant predictor of judgment accuracy, but whilst this result was obtained after relatively little input (6 blocks) and at low levels of L2 proficiency in children, it emerged after extensive input provision (up to 72 blocks) and at high proficiency levels in adults. Taken together, the significantly increasing effect of procedural learning ability over the course of L2 practice and the significant effect in the GJT indicate that, compared to adults, children may engage procedural processing earlier in L2 learning, including when they are processing language input in a meaningful context.

These results complement the findings of recent studies that found that procedural learning ability and implicit statistical learning predict proficiency in the L1 in typically developing children of comparable age ranges or younger (Conti-Ramsden et al, 2015; Kidd, 2012). Overall, the picture that emerges is one where procedural memory appears to have a pivotal role in children both in the L1 and in the L2.

The GJT model provided evidence that word order was learnt significantly better than case, in particular OV sentences with case errors were significantly less accurate than sentences with word order errors. Also, whilst there was no significant difference if SOV and OV sentences had word order errors, SOV sentences with case violations were significantly more accurate than OV sentences with case violations.

As children have been found to be particularly sensitive to item frequency in the input (e.g., MacWhinney, 1983; Braine et al., 1990), a possibility is that SOV sentences were learnt better because they were more frequent compared to other word order types.

### 8.1.3 Language knowledge

The third research question sought to investigate the nature of the L2 knowledge acquired by the participants in terms of the implicit/explicit continuum. The present study is among the few that considered the nature of language knowledge in children and, to the best of my knowledge, the first L2 learning study to have applied the zero-correlation criterion to child data.

The analysis found that, whilst judgment of grammatical sentences mainly depended on conscious knowledge, there was evidence that judgment of sentences with word order violations relied on average on implicit knowledge. This was concluded on the basis of the results of the application of the zero-correlation criterion. This found that for word order violation trials, accuracy was significantly above chance and, concurrently, confidence in sentence judgments did not significantly positively correlate with accuracy (in the terms of our model confidence was not a significant positive predictor of accuracy). Since the implicit status of linguistic knowledge (structural knowledge in Dienes' terms) can be inferred from implicit judgment knowledge, it can be concluded that child knowledge of word order restrictions was on average implicit. In the Confidence model, procedural learning ability was a significant predictor of accuracy specifically in sentences with word order violations compared to sentences with case violations (the effect was marginal for grammatical sentences). The role played by procedural learning ability in the learning of word order would thus confirm the DP-model's prediction that learning of syntactic regularities (rules) mainly depends on procedural long-term memory. The model also found an inverse relationship between confidence and procedural learning ability, indicating that an increasing involvement of procedural learning was related to less confidence in the sentence judgment.

The analysis of the debriefing questionnaires confirmed that most participants reported limited knowledge relative to word order and mostly reported that they noticed the occurrence/absence of specific words without reference to where in the sentence the word occurred. About one quarter of the participants mentioned single elements' positions in the sentence or relative word order, providing evidence of explicit knowledge of word order at least for some of the children. It is important to remember that the measure of awareness adopted here is not a comprehensive measure of explicit knowledge, but only an index of explicit knowledge at its most developed level, i.e. verbalizable knowledge (cf., Dienes & Perner, 1999; Karmiloff-Smith, 1992; Matsui et al., 2006). By contrast, confirming results in Ferman and Karni (2010), there was a uniform lack of reports relative to the semantic properties of the particles or the NPs (linking).

A Pearson's correlation found that increased awareness of the syntactic properties of the language was significantly related to vocabulary learning ability and accuracy scores in the computer game but not to the GJT scores. This pattern of correlations may indicate that children with strong vocabulary learning ability may have had sufficient attentional resources to notice other formal properties of the language during practice or devise/try out game strategies to improve their score. A lack of correlation between awareness and the GJT score can be related to the significant effect that procedural learning ability had in determining the accuracy of the GJT task.

Taken together, these results confirm findings in Lichtman (2012, 2016) and indicate that although primary school children exposed to a novel language in incidental conditions, gained both implicit and explicit knowledge of the miniature language in the learning process, awareness of the properties of the language, as emerged from the verbal reports, was limited.

## 8.2 Discussion of Study 2

### 8.2.1 L2 practice

The descriptive data relative to the accuracy in the game practice indicated above-chance learning of the miniature language in adults both overall and in each trial subcategory, including SOVs and OVs trials, with effect sizes from medium to large (Cramer - V). However, the inferential analysis showed that the odds of accurate responses on asymmetric trials were on average significantly better than on symmetric trials, suggesting that high levels of L2 comprehension in the latter case had been more difficult to attain also in adults. That the adults were able to learn linking rules, included when these exclusively depended on the correct interpretation of case particles, confirms findings in e.g. Boyd et al. (2009), Grey et al. (2015), Pili-Moss (2017) and Williams and Kuribara (2008).

As in the child study the first question was investigated fitting three models, one to the overall dataset and two additional ones to the symmetric and asymmetric datasets respectively. In all three models declarative learning ability emerged as a highly significant positive predictor of accuracy. The effect of vocabulary learning ability was also positive and significant in all models, but of smaller magnitude compared to declarative learning ability and stronger in symmetric trials compared to asymmetric trials. There was also evidence that the effect of declarative learning ability significantly increased across the game practice in the symmetric dataset, whilst the same occurred to vocabulary learning ability in the asymmetric dataset. By contrast the effect of procedural learning ability was found to be nonsignificant in all three models, with evidence of the effect significantly decreasing over the course of training in asymmetric contexts.

Summarizing, the overall main effects in the models of the adult practice data discussed to this point indicate: (a) a significant positive role of abilities linked to declarative long-term memory that became stronger over the course of training, with larger effects for declarative learning ability compared to vocabulary learning ability; and (b) a nonsignificant main effect of procedural learning ability. Points (a) and (b) confirm the findings of previous studies that deployed the BROCANTO2 paradigm (Brill-Schuetz & Morgan-Short, 2014; Morgan-Short, 2007; Morgan-Short et al., 2010, 2014; Pili-Moss et al., 2018) or that, more in general, looked at the role of declarative and procedural memory in adult L2 learning (Antoniou Ettlinger and Wong, 2016; Ettlinger, Bradlow, & Wong; 2014; Hamrick, 2015; cf. also Hamrick, Lum, & Ullman, 2018). Relative to the early stages of adult L2 exposure (or at low L2 proficiency) these studies consistently found behavioral and ERP evidence of an asymmetry between declarative memory and procedural memory compatible with the one emerging in the present study.

A series of significant interactions between cognitive variables and between cognitive variables and categorical variables provided further insight. For example, the overall model returned that procedural learning ability interacted with type of trial and was significantly associated to better accuracy outcomes in symmetric SOV and

OV trials compared to asymmetric SOV and OV trials (the main effect of procedural learning ability was marginally positively significant for SOVs trials).

Further, declarative learning ability significantly positively interacted with procedural learning ability, with the effect increasing over the course of training. A positive interaction indicates a positive effect on accuracy over and above the additive contributions of the single variables. Inspection of the follow-up models showed that the significant Decl:Proc interaction specifically emerged in the symmetric dataset, with no interaction in the asymmetric dataset. The possibility of a co-operative interaction between memory systems is predicted by Ullman's DP model but is also not incompatible with other cognitive accounts of language learning (DeKeyser, 2007; N.Ellis, 1994; 2005). As such, the pattern of co-operative interaction between learning abilities found in the data can be interpreted from multiple cognitive perspectives.

The finding that procedural learning ability supports learning of form-meaning relationships would appear to be at odds with the fact that declarative memory and not procedural memory is thought to be preferentially engaged in processing of semantics. However, it can be argued that the learning of form-meaning relationships, especially when they involve the assignment of a specific word order, require both semantic knowledge and rule-based restrictions (e.g. linearization rules) according to which form-meaning relationships are encoded in the L2 syntactic representation. If this is correct, the contribution of both memory systems to the learning of these complex regularities would be expected.

In two recent studies Dominey and Hoen (2006) and Dominey and Inui (2009) discussed evidence that not only sequence learning but also processing of form-meaning relationships involving thematic assignment of the kind explored here, implicate the recruitment of corticostriatal brain areas, compatible with the ones implicated by the procedural memory system. In this light, the finding that procedural learning ability appears to have a significant effect specifically in linking contexts would confirm independent findings in the neuroscience literature at behavioral level and for L2 learning.

A supportive/co-operative role for declarative knowledge has also been hypothesized for L2 automatization, with robust declarative knowledge/processing supporting proceduralization (DeKeyser, 2015; Segalowitz, 2010). Further, re-analyzing reaction time data from Morgan-Short et al. (2014), Pili-Moss et al. (2018) found that an interaction between declarative learning ability and procedural learning ability significantly contributed to L2 automatization in adults. Finally, in a recent neuroimaging study Morgan-Short et al. (2015) found that adults with particularly high levels of declarative learning ability trained in BROCANTO2 in implicit conditions, displayed an earlier switch from engagement of brain areas associated with declarative memory to areas associated with procedural memory, compared to adults with lower declarative learning ability.

The asymmetric dataset also evidenced that an initial negative interaction between vocabulary learning ability and procedural learning ability was mitigated over the course of practice. This effect bears some similarity to an effect found in the child asymmetric dataset, although here the Proc:VocLearn:Block interaction was overall positive.

A further set of results concerns the role of the central executive and of the spacing between sessions in moderating the predictive effect of declarative LTM abilities and procedural learning ability. L2 training studies have found mixed evidence with regards to the role of working memory in L2 learning. Some studies have found that working memory was a predictor of morphosyntactic gains (e.g., Brooks et al., 2006; Kapa & Colombo, 2014; Tagarelli et al., 2015), whereas others did not find that working memory had a significant effect (e.g., Antoniou et al., 2016). The results in the present study confirm that a measure of the central executive was not a significant main effect, but found that it was a variable in a significant interaction.

Specifically, the present study found that in symmetric contexts a measure of the central executive had a significant adverse effect on the ability of procedural learning ability to predict accuracy during game practice (and vice-versa). According to Ashby and Maddox (2005) working memory plays a primary role in supporting hypothesis-testing strategies. Arguably, higher abilities to perform operations on audio-visual input in working memory in the game practice, particularly in symmetric contexts, would directly relate to a better ability to compare different scenarios before selecting a response, hence to a better chance of responding correctly. Considering that each game trial provided a relatively long maximal time to respond (up to 1 minute), it is likely that higher central executive abilities biased adult learners towards explicit hypothesis testing to the detriment of procedural learning strategies.

Finally, unlike previous BROCANTO2 studies, where differences in spacing between sessions were not treated as a covariate, the present study found that variations in the spacing between Session 1 and Session 2 predicted significant differences in accuracy, with longer spacing related to better accuracy. A three-way interaction between vocabulary learning ability, spacing and block also indicated that, this effect was enhanced for higher levels of vocabulary learning ability and was especially evident later in practice. One hypothesis to account for this pattern of results is that the memory consolidation of the visual-aural associations required for vocabulary learning that occurred between Session 1 and Session 2, specifically supported associations that had been established with enough strength in the course of the first vocabulary training session. No effects for the spacing between Session 2 and Session 3 emerged.

### 8.2.2 GJT evidence

The descriptive statistics relative to the GJT data found significant above-chance learning overall, for all word order types and for grammatical sentences and ungrammatical sentences with case violations. Accuracy in ungrammatical sentences with case violations was below chance. In line with the descriptive statistics, modeling of the GJT dataset showed that grammatical trials and ungrammatical word order trials were significantly more accurate than ungrammatical case trials and also that there was significant better learning of SOV trials compared to OV trials.

The present results corroborate the findings of previous adult L2 training studies that reported above-chance learning of word order (e.g., Boyd et al., 2009; de Graaf, 1997; Francis et al., 2009; Friederici et al., 2002; Grey et al., 2015; Lichtman, 2012; Morgan-Short et al., 2014; Rebuschat & Williams, 2012; Rogers et al., 2015; Williams & Kuribara, 2008; Wonnacott, 2008), or better learning of word order compared to gender or case morphemes (e.g., Grey et al., 2015; Morgan-Short et al., 2007; Rogers et al., 2015). Counter to L2 training studies that found evidence that adults learned gender and case affixes (e.g., Antoniou et al. 2016; Brooks et al., 2006, 2017; de Graaf, 1997; Morgan-Short et al., 2010; Robinson, 2005) the present study did not find robust learning of case morphology after about 2 1/2 hours of language exposure.

Relative to the role of cognitive individual differences, the model found a significant positive effect of declarative learning ability and either no effect or a significant negative effect of procedural learning ability. These results replicate the findings of BROCANTO2 studies that have investigated adults and used the GJT as a measures of L2 knowledge in the early stages of learning (e.g., Brill-Schuetz & Morgan-Short, 2014; Hamrick, 2015; Morgan-Short et al., 2014).

This replication is particularly important because it extends the paradigm's methodology to participants with an L1 different from English and deploys partially different measures of declarative learning ability (e.g., the ROCF in the case of visual-spatial declarative learning ability). The fact that results were replicated indicates that the findings in previous experiments are not language- or task-specific and can be generalized to different populations and behavioral tasks that are known to engage brain areas implicated in declarative and procedural memory processing.

The model also returned a significant negative interaction between a measure of the phonological loop and procedural learning ability. Compared to the gaming task, it is plausible to assume that the aural GJT, that only required judging a sentence on the basis of a perceived similarity/difference with previous exemplars, would have recruited central executive resources to a lesser extent. However, it may still have relied significantly on the span of the phonological loop to allow full sentences to be retained in auditory memory for the purposes of a comparative evaluation with sentence stimuli exemplars stored in long-term memory. In this case, one can hypothesize that greater phonological loop capacity may have supported declarative processing of the GJT sentences to the significant detriment of procedural processing in adults.

### 8.2.3 Language knowledge

Finally, in order to address the third research question a further model with confidence in the sentence judgment as a predictor of accuracy was fitted to the GJT data to investigate the outcomes of the zero-correlation criterion in the adult sample.

The results showed that confidence significantly predicted accuracy in the case of grammatical trials, whilst it was a negative predictor of accuracy in ungrammatical case trials. No significant correlation was found for ungrammatical word order trials. In the light of the zero-correlation criterion, these data indicate that, on average, judgment knowledge was explicit in grammatical sentences, whilst judgment knowledge in ungrammatical word order trials was implicit.

Since implicit linguistic (structural) knowledge can be inferred from implicit judgment knowledge, it can be concluded that, at this point in the experiment, adults' linguistic knowledge of word order was on average implicit (unconscious). On the other hand, since explicit judgment knowledge can be related to both implicit and explicit structural knowledge, judgments of grammatical sentences may have been given based on implicit or explicit linguistic knowledge. Accuracy on ungrammatical case trials was below chance in this experiment, so the zero-correlation criterion cannot be applied to this trial subset.

The analysis of explicit language knowledge emerging in the debriefing questionnaires revealed a pattern different from the one observed in the child study. The analysis of the verbal reports provided robust evidence of the availability of explicit language knowledge of word order regularities, as well as evidence of explicit hypothesis about regularities linking syntax to semantic representations. The fact that awareness of word order regularities was observed for most participants is at odds with the result that, according to the zero-correlation criterion, learning of word order regularities was on average implicit. Taken together, the two findings point to the conclusion that adults must have gained awareness of word order rules after (or possibly during) the GJT, mainly as a consequence of the evidence provided by the ungrammatical exemplars they were exposed to in the task (evidence that properties of the test task can support the development of explicit knowledge in adults is discussed in Williams & Lovatt, 2003).

The lack of a significant correlation between the measure of the explicit knowledge of word order and the GJT would be due to the fact that this knowledge either mainly became available after the GJT, or, even if some awareness was developing already during the task, this was not sufficiently robust for the correlation to emerge. The lack of a significant correlation between explicit knowledge of form-meaning relationships and accuracy during the game may be due to the complexity of the rules that had to be mastered in order to perform accurately. It is likely that awareness of form-meaning relationships was gained during the game practice due to the substantial amount of thinking time available. However, even if participants were partially aware during the game, awareness of the existence of some kind of form-meaning rules did not guarantee accuracy in the task because erroneous or incomplete hypotheses on the nature of those rules would have been sufficient to result in inaccurate responses.

Overall, the analysis of the type of language knowledge developed by adults during exposure to BrocantoJ in incidental conditions confirms the results of previous studies that found that adults trained in a novel artificial language developed both implicit and explicit language knowledge (e.g., Grey et al. 2015; Lichtman, 2012; Rebuschat & Williams, 2012; Rogers et al., 2015). Contributing to this body of literature, Study 2 evidenced the implicit nature of word order knowledge of a novel L2 in adults at the time of testing and suggested that the characteristics of the test itself may have determined the development of language knowledge from implicit to explicit during/after testing.

## 8.3 Comparisons between Study 1 and Study 2

### 8.3.1 L2 attainment and rate of learning

Looking at the comparison between children and adults on language attainment during practice (Table 8.1, next page) what emerges clearly is the substantial similarity in the results. Although with differences in effect size, in both cases overall attainment was on average significantly above chance, indicating that in both cases there was evidence of learning of linking rules, including some evidence that the accusative case marker (*ru*) was used as a cue to assign the thematic role of object to the associated NP. This confirmed results already emerged in a preliminary study (Pili-Moss, 2017).

Further, attainment in symmetric trials was for both groups significantly less accurate than in asymmetric trials, indicating that trials whose accuracy depended on form-meaning relationships were more difficult to process independent of age. The attainment of the two groups in the GJT was also comparable, except for the children's worse performance on OV sentences. In the GJT case marking tended to be judged inaccurately in both groups. One possible account for the low levels of accuracy in case trials is the low salience of case markers (Goldschneider & DeKeyser, 2005). However this seems unlikely in this case, since most participants in both groups mentioned noticing the particles in their verbal reports.

Table 8.1

*Summary of Language Attainment*

| | Children | Adults |
|---|---|---|
| **Practice** | Attainment S above chance (small effect sizes)<br><br>S symm/asymm difference | Attainment S above chance (medium effect sizes)<br><br>S symm/asymm difference |
| **GJT** | Overall attainment S above chance (except OV sentences)<br><br>Case violations below chance<br><br>Case marking significantly less learned than word order<br><br>SOV better than OV in case violation trials | Overall attainment S above chance<br><br>Case violations below chance<br><br>Case marking significantly less learned than word order<br><br>SOV better than OV overall |

*Note.* S = significant.

An alternative explanation could be that at such early point in practice the syntactic representation of the new language was in some way still deficient. As discussed introducing Processability Theory in 3.5.1, the processing of case marking would have required a procedure at phrasal level. Developmentally this is a more advanced stage compared to the one at which canonical verb-argument word order is processed (category procedure). It would appear that the limited language exposure was not sufficient for the participants to gain access to the higher level of phrasal procedure independent of age.

Although the content of learning in the two groups was comparable, adults had an advantage on the level of language attainment they reached and on the rate of learning. A *Z*-test comparing the proportion of accurate trials in children and adults shows significant between-group differences overall, as well as for symmetric and asymmetric trials separately (overall, $Z = -24.69$, $p = .000$; symmetric, $Z = -17.59$, $p = .000$; asymmetric, $Z = -18.80$, $p = .000$).

A significant adult advantage was maintained also in the GJT, although here the difference was less substantial (*Z* = -2.80, *p* = .005). Overall, these results confirmed the findings of rate-of-learning studies that indicate significant advantages for adults in L2 learning compared to younger learners (Lichtman, 2012; Muñoz, 2006a, 2006b; Pfenninger & Singleton, 2017) and showed that these emerged already in the first few hours of incidental exposure to a novel language.

Interestingly, the adult advantage in the GJT was somewhat less marked than in the listening comprehension activity during practice. This is apparently at odds with the finding that, compared to GJTs and grammar-related tasks, listening comprehension has been often found to be the area where age differences are less conspicuous (Cenoz 2002, 2003; Muñoz, 2003, 2006a). A possible explanation may be sought in the specific characteristics of the GJT in this study. In particular the GJT was aural and timed, characteristics that may have provided less opportunities for adults to apply their more developed analytic abilities, compared to a GJT administered in the written and untimed modality (Ellis, 2009).

An analysis by block of the practice data indicates that the adult advantage emerged in every block and was at its strongest at the beginning of Session 3. Child rate of learning appeared to pick up faster in Session 3 compared to earlier sessions. Although the timeframe of the data collection did not permit to record the further development of the rate of learning in the two groups, it seems unlikely that children would have surpassed adults before the latter reached ceiling performance (Table 8.2).

Table 8.2

*Comparisons of Percentage Attainment in the Language Practice by Block*

| | Children | Adults | *Z*-test comparison | |
|---|---|---|---|---|
| | | | *Z* | *p* |
| **Block 1** | **17.2** | **30.4** | **-5.97** | **.000** |
| **Block 2** | **22.4** | **42.9** | **-8.41** | **.000** |
| **Block 3** | **24.0** | **47.6** | **-9.46** | **.000** |
| **Block 4** | **26.6** | **63.8** | **-14.39** | **.000** |
| **Block 5** | **30.6** | **63.2** | **-12.53** | **.000** |
| **Block 6** | **36.9** | **67.9** | **-11.85** | **.000** |

A further interesting observation at descriptive level is related to the way rate of learning developed across practice. As Figure 8.1 (next page) clearly shows, rate of learning rose gradually in children, with between-sessions variation more comparable to intrasessional changes, whereas rate of learning substantially increased between sessions in adults. A *Z*-test comparing the transition from Block 1 to Block 2 (Session 1 to Session 2) and from Block 3 to Block 4 (Session 2 to Session 3) in children and adults found consistently significant between-session differences for adults but not for children (adults, $Z = -4.79$, $p = .000$; $Z = -6.11$, $p = .000$; children, $Z = -2.54$, $p = .011$; $Z = -1.17$, $p = .244$). Although assessing the role of sleep is beyond the scope of the

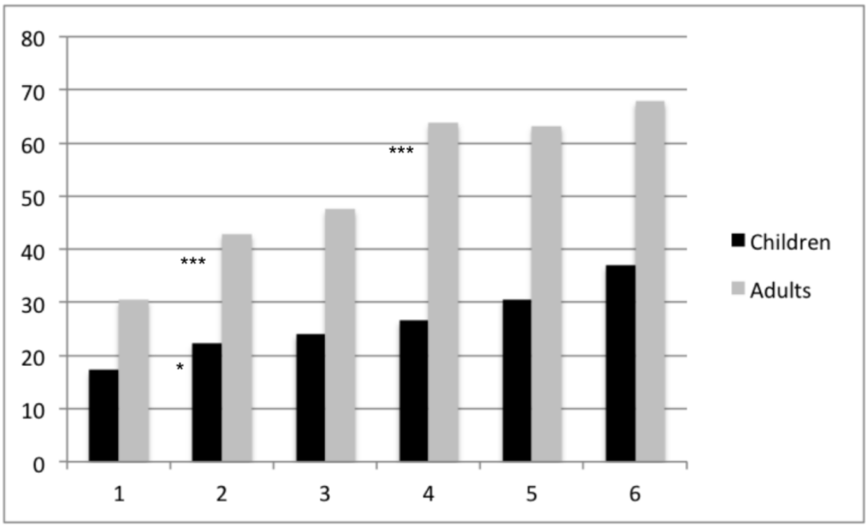


*Figure 8.1*. Percentage of Correct Practice Trials per Block (the Stars Indicate the Level of Significance of the Difference With the Preceding Block).

present thesis, the pattern would be consistent with a stronger effect of memory consolidation during sleep in adults compared to children. However, note that Pili-Moss (2017), a preliminary study run with a smaller number of participants, found the opposite pattern, i.e. significant between-session learning in children and more gradual increases in adults.

### 8.3.2 Cognitive variables and language knowledge

Based on the evidence of the L2 measures taken in this study, and beside the amount of learning, children and adults did not differ substantially relative to the content of L2 learning. However, the analysis of the cognitive abilities involved in L2 learning returns a more complex picture, with similarities between the two groups but also with clear between-group differences. In what follows I will provide an overview of group comparisons based on cognitive main effects and interactions across the domains of the three research questions addressed in Study 1 and Study 2 (also summarized in Table 8.3, Table 8.4, and Table 8.5 in the next pages).

***Practice.*** In both groups LTM declarative abilities were important predictors of accuracy, with evidence that their effect increased over the course of practice in children and adults (vocabulary learning ability only in children). As discussed previously, the reason that declarative learning ability did not emerge as a significant effect in the child symmetric dataset may be due to the comparatively higher difficulty of linking trials. However, even considering only asymmetric trials, it is clear that declarative learning ability was a larger effect than vocabulary learning ability in adults, whereas in children the opposite was the case.

One possibility is that declarative memory (as measured by Decl) and memory for vocabulary (i.e., memory for recognition of picture-sound associations in receptive tasks) are engaged differently in L2 learning in the two age groups due to the specific abilities they require (recall vs. recognition). However, due to the fact that a number of factors in the measurement of vocabulary learning ability were not controlled in these studies, no conclusions with regards to the relevance of the recall vs. recognition distinction can be drawn. Clearly, more research is needed to understand the specific role of the abilities currently collectively labeled as 'declarative' to begin to answer these questions.

More substantial between-group differences were found relative to the effect of procedural learning ability, although during practice it did not emerge as a significant main effect in either group. Whilst there was evidence of a significant increase in the effect of procedural learning ability in children, in adults it significantly decreased in the same context (asymmetric dataset). The evidence appears to suggest that, if the trials were semantically less complex, children's procedural learning ability was a relevant factor early on in language processing, whilst adults tended to rely on declarative memory resources instead.

The model of the symmetric dataset also showed that in adults procedural learning ability was significantly more relevant in determining trial accuracy in symmetric trials compared to asymmetric trials. In sum, the pattern emerging from the data for procedural learning ability is one for which the ability was relevant for children (and significantly more so over the course of training) in asymmetric trials, whilst its role was more marked in symmetric trials in the case of adults.

Table 8.3

*Summary of the Effects of the Cognitive Variables During Practice*

| | Children | Adults |
|---|---|---|
| **Main effects** | | |
| Decl | S in asymm. trials | S overall and sign. increasing across practice (symm. subset) |
| Proc | NS but significantly increasing across practice in asymm. trials | NS and significantly decreasing across practice in asymm. trials<br><br>Significantly more positive in symm. trials |
| VocLearn | S across the board | S across the board |
| SpacingS1S2 | | Positive and S in asymm. trials |
| Decl/VocLearn | VocLearn has a more substantial role | Decl has a more substantial role |
| **Interactions** | | |
| Decl*VocLearn | S positive interaction in asymm. trials | |
| VocLearn*Proc | S negative interaction in asymm. trials that mitigates during practice | Mainly positive interaction in asymm. trials |
| Decl*Proc | | S positive interaction in symmetric trials increasing during practice |
| Proc*Exec | | S negative interaction in symm. trials, becoming more negative with practice |
| VocLearn*Spacing S1S2 | | Longer spacing corresponding to increasingly better accuracy for high VocLearn |

*Note.* S = significant; NS = nonsignificant.

The role of adult procedural learning ability can be better elucidated in the light of the positive significant interaction between declarative and procedural learning ability in the symmetric dataset. As discussed previously, in the case of trials that required the learning of morphosyntactic form-meaning relationships there was a significant positive interaction between declarative and procedural learning ability that increased with practice. In the discussion of Study 2 the suggestion was made that the positive interaction may reflect an underlying dynamics of co-operative engagement between declarative and procedural processing resources in the case of linguistic stimuli where a semantic component is closely related to a rule-based morphosyntactic component (e.g., correct linearization of arguments). No such interaction emerged in the child symmetric dataset. It remains unclear whether it would have been possible to detect this effect in children as well, had learning been more robust, or whether the lack of interaction constituted a genuine difference.

A further significant pattern of interaction, this time negative, was found in the child asymmetric dataset between vocabulary learning ability and procedural learning ability. More precisely, this was an effect conditional to a more complex three way-interaction involving also declarative learning ability. The interaction between the two abilities depending on the declarative memory system significantly positively correlated between each other and together mitigated the effect of procedural learning ability. Moreover, the negative interaction between procedural and vocabulary learning ability significantly decreased over the course of practice. In the discussion of Study 1 the suggestion was made that this negative interaction was consistent with hypotheses relative to the possibility of the declarative and procedural memory system to interact competitively during learning and processing (Packard & Goodman, 2013; Ullman, 2005, 2015, 2016).

A similar (Proc:VocLearn) interaction going from negative to positive during the course of practice was also found in the adult asymmetric dataset. However, in this case the interaction of procedural and vocabulary learning ability across practice was positive on average. Also, importantly, declarative learning ability, a highly significant main effect, was not involved in the interaction in the adult asymmetric dataset. Summarizing, in terms of the relationship between the two cognitive predictors of interest, it appears that children displayed a clearer pattern of competitive interaction and less evidence of co-operative interaction compared to adults. Also, there was evidence of a positive interaction between declarative and vocabulary learning ability in children, which was not found in adults.

As for the remaining cognitive variables that were included as covariates, the measure of the phonological loop had no significant main or moderating effect in any of the practice models, irrespective of age. Hence the present results do not confirm the significant role of the phonological loop evidenced in previous research (e.g., Williams & Lovatt, 2003), but confirm the results of previous studies that found a limited role for the phonological loop in L2 learners at low levels of proficiency (e.g., Kormos & Sáfár, 2008).

By contrast, a significant negative correlation was found between procedural learning ability and a measure of the central executive in the adult symmetric dataset. However, given that it was not possible to take any measures of the central executive in the child group, no comparisons are possible. Finally a significant positive effect of spacing between Session 1 and Session 2 emerged in the adult asymmetric dataset, whilst spacing never predicted accuracy in children. However, only adult learners with higher levels of vocabulary learning ability could take full advantage of the spacing effect. Since spacing was measured in days, the discrepancy between adults and children in the main effect is a further cue pointing to a stronger role of consolidation during sleep in the former compared to the latter.

Summarizing, the main between-group differences that emerged during practice in the comparison of the two studies were: (a) a significant increase of the effect of procedural memory during practice found only in children; (b) a more significant role for declarative learning ability in adults, where the effect significantly increased during practice in some cases; (c) a positive interaction between declarative and procedural learning ability that increased during practice found only in adults; (d) a robust negative interaction between learning abilities relating to the declarative memory system and procedural learning ability in children (mitigated later in practice); (e) a positive relationship between vocabulary learning ability and procedural learning ability later in training in adults; (f) no evidence of the involvement of declarative learning ability in a negative interaction with procedural learning ability in adults vs. evidence in children; (g) in adults, but not in children, evidence that spacing between the first and the second session supported L2 attainment. Similarities between the two groups included: (h) an overall significant effect of declarative and vocabulary learning ability (but no effect of declarative learning ability in the child symmetric dataset); (i) a significant effect of practice on accuracy.

***GJT Task***. The patterns of effects in the GJT task were less complex compared to those found in the game practice. There were clear differences between children and adults on this task. Whilst declarative learning ability was not a significant predictor of accurate sentence judgment in children, it was in adults. Conversely, a significant positive effect was found for procedural learning ability in children, but not in adults. These results confirmed the findings of previous L2 studies with adults tested after limited exposure to a new miniature language (Brill-Schuetz & Morgan-Short, 2014; Hamrick, 2015; Morgan-Short et al., 2014). They also provide further evidence of the more prominent role of procedural learning ability in children compared to adults in the earliest stages of L2 learning.

Table 8.4

*Summary of the Effects of the Cognitive Variables in the GJT*

| | **Children** | **Adults** |
|---|---|---|
| **Main effects** | | |
| Decl | NS effect | S positive effect |
| Proc | S positive effect (OV sentences) | NS or S negative effect |
| **Interactions** | | |
| Proc*PhonLoop | | S negative interaction |

*Note.* S = significant; NS = nonsignificant.

In adults, but not in children, there was a negative interaction between the measure of the phonological loop and procedural learning ability. In the discussion of Study 2 it was suggested that larger PhonLoop spans in adults might have facilitated declarative processing over procedural processing. In this perspective, the lack of the effect may depend on children's limited reliance on declarative memory in this task. Alternatively it may have a developmental explanation, related to the fact that the phonological loop is still undergoing development in children of the age range considered here (Gathercole et al., 1994).

***Language Knowledge.*** The nature of the language knowledge acquired by children and adults was investigated by assessing the participants' confidence in their judgments in the GJT and by analyzing the verbal reports obtained at the end of the experiment. In terms of the evidence provided by the zero-correlation criterion, there were practically no differences between children and adults. In both cases there was evidence that participants were unaware that they were judging word order trials correctly (implicit judgment knowledge) and that they were aware their judgment was correct in grammatical trials (explicit judgment knowledge). Case violation trials were accurate below chance in both groups and the nature of judgment knowledge in this case remained undefined irrespective of age. Based on the findings of the Confidence models it can be inferred that, on average and at the time of testing, participants' structural knowledge of word order restrictions was implicit (unaware) independent of age.

The aim of this model was to assess the correlation between confidence and accuracy in the GJT, but in the case of children the main cognitive variables of interest were also included as moderators. For adults it was found that declarative and procedural learning ability did not improve the fit of the Confidence model, so these variables were not included. As a covariate in the child Confidence model, procedural learning ability significantly predicted accuracy in the case of word order trials. This finding is particularly interesting because it supports the DP model(s) prediction that procedural memory is specifically engaged in the learning of rule-based linguistic sequences. The child Confidence model also evidenced a significant negative correlation between confidence and procedural learning ability.

Comparing children and adults on the quality of their explicit knowledge (Table 8.5, next page), it emerged that, probably unsurprisingly, adult explicit representations of BrocantoJ regularities were more detailed and qualitatively richer. Most children, even if not all, characterized their description of the language at the level of noticing of lexical items (cf. Lichtman, 2012 for similar findings). Adults, on the other hand, not only reported noticing lexical items but reported syntactic information about them (e.g., position in the sentence, relative order). Unlike children, adults also explicitly

Table 8.5

*Summary of the Confidence Effects and Explicit Knowledge*

| | **Children** | **Adults** |
|---|---|---|
| **Main effects** | | |
| Conf | S positive effect for gramm. trials | S positive effect for gramm. trials |
| | NS for WO trials | NS for WO trials |
| | S negative effect for Case trials | NS for Case trials |
| Proc | S effect for WO trials | |
| **Interactions** | | |
| Conf*Proc | S negative interaction | |
| **Zero-correl. crit.** | Word order implicit<br>Gramm. trials explicit | Word order implicit<br>Gramm. trials explicit |
| **Explicit knowledge** | Reports of noticing specific words in the input | Reports of word order regularities/hypotheses on particle semantic functions |
| VocLearn | S correlation with awareness of morphosyntax | |
| Accuracy (Practice) | S correlation with awareness of morphosyntax | |

*Note.* S = significant; NS = nonsignificant.

expressed hypotheses with regards to the semantic functions of the case particles, whereas none of the children did (cf., Ferman & Karni, 2010 for similar results in an 8-year-old group).

Overall, this pattern of results may suggest that for most children conscious knowledge was largely limited to the vocabulary, which they were taught explicitly (but note that in the case of *ri*/*ru*, they noticed they had *not* been given any explicit instruction about them, in some cases reporting this spontaneously during the experiment). By contrast, adults reported more information relative to aspects of the language they were not explicitly taught.

Finally, for children awareness of morphosyntax significantly correlated with vocabulary learning ability and accuracy during practice. In 8.1.3 it was suggested that efficient vocabulary learning in children might have freed attentional or/and processing resources to notice distributional properties of the input. Maintaining the same line of argument, the lack of a significant correlation between vocabulary learning ability and awareness in adults may suggest two possibilities. Adults could benefit from a more robust pool of attentional resources during practice (re-directing resources from vocabulary processing did not make a significant difference). Alternatively, development of language awareness was less dependent on the availability of attentional resources during practice (for example this would be the case if it mainly happened offline/ at a later stage).

# 9. Conclusions

## 9.1 Contributions of the studies

The present thesis investigated the role of cognitive learning abilities rooted in long-term memory in the very early stages of L2 learning and elucidated the extent to which cognitive abilities are differentially engaged depending on age. A strength of the methodology was that a miniature language paradigm and a learning environment, previously used only in adult studies, were adapted to be deployed with children. This not only allowed a comparison with previous studies, but also provided matched instructional conditions for both child and adult participants.

One of the main findings of the thesis concerns the role of procedural learning ability in L2 learning. In particular, Study 1 found that, for children, procedural learning ability is a predictor of L2 learning and engages early on in practice. During practice it significantly supported the learning of sentence stimuli that were comparatively less complex semantically, and was a significant effect driving accuracy in a grammaticality judgment test.

Taken together, these results contrast with the findings of comparable adult studies (including Study 2) where procedural learning ability was not found to be a predictor of L2 learning at early stages of practice. This points to an important difference between children and adults. The fact that procedural learning may have a more prominent role in child L2 learning compared to adult learning has been envisaged theoretically (e.g., DeKeyser, 2012; Ullman, 2015, 2016) but had not been directly shown experimentally before. Declarative and vocabulary learning ability were significant predictors of accuracy in adults and children, but in general the findings suggested a more prominent role of vocabulary learning ability in children and a more prominent role of declarative learning ability in adults.

More generally, the results of Study 1 relative to the role of procedural learning ability in child L2 learning, complement the findings of recent research that has investigated procedural learning ability in child L1 (Conti-Ramsden et al., 2015; Kidd, 2012). Overall, these studies constitute convergent evidence for a prominent role of procedural memory in child language processing and learning.

Although procedural learning ability did not emerge as a significant main effect in adults, it significantly interacted with declarative learning ability supporting accuracy in constructions encoding linking between syntax (word order) and semantics (thematic interpretation). A further significant interaction, this time negative, was observed in children between LTM declarative abilities (in particular vocabulary learning) and procedural learning ability. Overall, what emerges is a pattern of co-operative interaction in adults, and of competitive interaction in children. More research is needed to investigate the extent to which these differences can be generalized. The findings relative to the interactions are compatible with the predictions of Ullman's DP model (2005, 2015, 2016) and supported by evidence of the interaction between memory systems in neuroscience. However, they are also relevant to other cognitive models of L2 learning including Skill Acquisition Theory (DeKeyser, 2015) and Usage-Based approaches (N. Ellis, 1994, 2004; N. Ellis & Wulff, 2015).

Relative to the relationship between LTM abilities and type of linguistic target, the findings of Study 1 and Study 2 are compatible with cognitive models of language learning that predict that procedural learning ability should have a role in the learning of word order. Compatibly, the studies found that procedural learning ability plays a pivotal role in the learning of word order (children) and in the learning of word-order based form-meaning linking (adults).

Given the important differences in processing found between children and adults, the similarity of their L2 attainment is remarkable. Beside differences in rate of learning, both children and adults attained above-chance learning in the game practice and in the GJT, learned word order and failed to show significant learning of the syntactic patterns of case marking. This indicates that conclusions relative to age differences solely based on language attainment would not have been sufficiently informative. To the extent that such results are replicated in future studies, they indicate that, beside attainment, research interested in age differences may benefit from an increased focus on language processing.

The analysis of language knowledge returned a less fairly balanced picture, although with between-group similarities. It emerged that, at the time of the GJT, both children and adults displayed implicit knowledge of word order. However, most adults reported explicit knowledge of word order in the verbal reports, whilst most child reports confirmed limited awareness of syntax. This led to the conclusion that adults' explicit language knowledge must have been gained during or after the GJT exposure. By contrast, there was some indirect evidence that children's explicit knowledge of syntax might have derived from exposure to the aural stimuli during the game. Also, only adults reported hypotheses about the semantic content of case markers, whilst no children did.

Study 2 also found that measures of the central executive and of the phonological loop significantly negatively interacted with procedural learning ability in adults. In both cases it was suggested that more robust central executive function and auditory short-term memory would support declarative processing to the detriment of procedural processing. Finally, the spacing between Session 1 and Session 2 had a positive effect on L2 learning in adults but not in children. Although the present study was not designed to test the effects of sleep on L2 learning, the evidence from spacing, matched to the observation that rate of learning consistently significantly increased between sessions in adults but not in children, led to the hypothesis that memory consolidation during sleep might have played a more substantial role for adults.

Methodologically, generalized mixed-effects models, used as the statistical analysis of choice, revealed a structure of main effects and interactions between cognitive predictors and moderators that had not emerged in previous studies with similar designs and represents a promising approach that should be adopted more widely in future research on individual differences (cf., Pfenninger and Singleton, 2017).

## 9.2 Limitations of the studies

The studies included in the present dissertation have a number of limitations, a selection of which is discussed in what follows.

***Learning Environment***. One of the main reasons for adopting the BROCANTO2 paradigm was that it was an established learning environment for the study of the effects of memory-related individual differences in adults. As no previous similar studies with children existed, adopting the paradigm was ideal from a comparative perspective. However, although the BROCANTO2 game was fully adaptable syntactically (aural stimuli), the order and type of game constellations inside each block was fixed, which limited its semantic flexibility.

***Participants***. Although fairly balanced, the number of participants in the child and in the adult group was not identical. Furthermore, although participants were almost perfectly balanced for sex in the adult group, they were not in the child group, due to the higher number of female participants who were not included in the final sample.

***Amount of Instruction***. In the child study, data were collected in school during term time and children's availability was subject to time limitations due to classroom commitments. This restricted the amount of time allocated for instruction in the study design of both experiments (children and adults had to be matched for instruction). As a consequence, the rate of learning was more limited than it could have been with longer exposure or/and a larger number of sessions.

***Ecological Validity***. Since the studies were experimental and conducted in controlled laboratory-like conditions, caution needs to be used in generalizing their results. In particular, this is true for direct comparisons with the classroom environment or other environments where social or interactional variables play an important role in shaping the learning experience.

At the same time the learning environment adopted in the present studies is more directly comparable with computer-assisted language learning. Considering the increasing role and the integration of computer-assisted education with traditional classroom-based teaching both in school and outside, evidence from language learning studies adopting this medium is undoubtedly relevant and of interest.

It can be also argued that exposure to a miniature artificial language put the learner in a somewhat 'unnatural' learning situation. However, since the miniature language adopted had the characteristics and productivity of a natural language, the 'unnaturalness' of the experience would depend more on the fact that the applicability of the linguistic system the learners were exposed to was limited to the computer game.

***Cognitive Measures***. In a more extended experimental design the set of memory-related tasks used as measures of learning ability could have been larger, and in particular would have ideally included an additional measure of procedural learning ability. Also, the studies did not include a measure of declarative learning ability based on recognition that was independent on the game. This would have allowed a controlled comparison between the recall and recognition components of declarative learning ability in both age groups. Finally, due to the fact that measures of the central executive were not taken in the child group, the role of working memory in child language learning could not be fully appreciated and a comparison with adults along that dimension was not possible.

***Task effects***. The present studies were not designed to test task effects but the results showed that, keeping other variables constant, the engagement of the same type of cognitive ability varied depending on the type of task that was deployed to measure it (game vs. GJT). Also, it cannot be excluded that the type of task interacted with age. Future studies will seek to control these effects for different age groups.

***Inferential Comparisons***. In this dissertation, for reasons of complexity and space, a decision had to be made as to whether to proceed with an in-depth analysis of the role of cognitive variables in L2 learning for each group separately, or prioritize the between-group age dimension. Since the memory dimension of the study was highly exploratory it was decided to privilege a detailed analysis of this aspect. Consequently, group comparisons along the age dimension remained mainly qualitative at this stage.

## 9.3 Further developments

In this section I will briefly review some of the potential further developments that can be envisaged in the light of the results of the present studies.

***Replication***. Analyzing the role of cognitive abilities in child L2 learning in a comparative dimension, the dissertation had a high exploratory value. For this reason it is particularly important that studies with a similar design/with comparable populations are conducted to test the extent to which the reported findings can be confirmed and generalized.

***Role of Instruction***. In both studies, exposure to morphosyntactic regularities was incidental. Further studies with a similar design could manipulate instruction by adding explicitly instructed groups and could investigate how the engagement of different cognitive abilities as well as the representation of language knowledge varies depending on type of instruction.

***Role of Input***. A number of studies have produced converging evidence showing that the structure of the exposure sets has a significant effect in miniature language learning. Future research could manipulate the exposure sets and investigate how the effects of cognitive abilities in L2 learning vary according to different distributions/statistical properties of the input. To the best of my knowledge, this area of statistical learning remains largely unexplored to date.

***Cognitive Measures***. This dissertation found some evidence that two learning abilities relating to the declarative memory system may be engaged differently during L2 learning depending on age or linguistic target. A promising avenue of investigation in this area is a more detailed approach to the effects of different measures tapping individual aspects of declarative and procedural memory. For declarative memory, a relevant distinction could be the difference between measures tapping recall of verbal/visual information vs. measures assessing verbal/visual recognition. For procedural memory, probabilistic learning could be compared to implicit sequence learning.

***Automatization***. In this dissertation, accuracy was used as the main outcome measure of language learning. However, other methods can be used to assess how L2 attainment progresses. One of these is the analysis of reaction times. Unlike accuracy, reaction times provide a measure of how the efficiency of language processing develops as a result of practice. In previous studies with adults (included BROCANTO2 studies) RTs recorded during language practice have been deployed to assess cognitive fluency (automatization in comprehension). A study following up on the ones presented here could extend the study of automatization to children by investigating how cognitive abilities engage in automatization across age groups.

***Language knowledge***. Further studies could aim at devising methods to better understand to what extent the representation of language knowledge, in particular explicit language knowledge, differs in children and adults, as well as gain more insight into the factors that support the crystallization of explicit knowledge in incidental instruction in the two groups. As verbal reports remain the major source of evidence of explicit knowledge in language studies, it would be particularly important to further develop research methods aimed at elucidating to what extent lack of evidence of explicit knowledge in verbal reports genuinely reflects lack of knowledge or lack of verbalization.

**9.4 Pedagogical considerations**

If confirmed by further research, the finding that procedural memory has a prominent role in child language learning, and in particular supports certain linguistic targets over others, has potentially very important implications for foreign language pedagogy. Recent studies comparing L2 learning in child early starters and late starters (e.g., Pfenninger & Singleton, 2017) have provided evidence that, overall, an early start in traditional school settings may not automatically translate to an advantage in language attainment. Part of the reason for this is that much of the teaching provided in the early years/primary school relies on communicative activities administered in a context (the classroom) where input is scarce and unlikely to support efficient implicit learning leading to substantial and lasting language gains.

Recent L2 instructed research has evidenced that children, including children of primary school age, benefit from explicit instruction (e.g., Lichtman, 2012; Suarez & Muñoz, 2011; Tellier & Roehr-Brackin, 2013). However, children's explicit representations of language most certainly differ from adult ones (Karmiloff-Smith, 1992). By investigating the specificity of child explicit language representations, current SLA research can significantly inform the design of explicit L2 instruction catering for this age range. This said, if children do have a specific advantage in procedural learning that is age related, this should also be fostered through appropriate instructional intervention. Computer mediated instruction (e.g., videogames) used as a component of a comprehensive instructional strategy could be particularly suited to assist language proceduralization and automatization. As our understanding of which linguistic features are ideally acquired procedurally increases, these instructional interventions could discriminate between aspects of the language that can benefit more from implicit instruction (one very good candidate could be word order) and aspects of the language that are best taught explicitly in an age-appropriate way.

This tailored 'mixed' approach is not limited to child language learning but extends more generally to learners of any age who for different reasons may not be characterized by robust aptitude for explicit language learning or may be particularly efficient procedural learners. Understanding the role of long-term memory variables represents an important further step in the study of how individual differences shape language learning. This line of research has had and will continue to have a great potential in supporting policy makers and educators in their efforts to provide a more tailored and effective learning experience for all learners and a more inclusive approach to language education.

## References

Abrahamsson, N. (2012). Age of onset and nativelike L2 ultimate attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition*, *34*(2), 187–214. https://doi.org/10.1017/S0272263112000022

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*(4), 481–509.

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*, 249-306.

Al Shatter, G. (2008). The development of verbal structures in L2 Arabic. In: J-U. Keßler (Ed.), *A brief introduction to Processability Theory*, (pp. 267-302). Newcastle upon Tyne: Cambridge Scholars.

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 899-912.

Ambridge, B., Pine, J. M., Rowland, C. F., & Young, C. R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition, 106*(1), 87–129. doi:10.1016/j.cognition.2006.12.015

Ambridge, B., & Rowland, C. F. (2013). Experimental methods in studying child language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(2), 149–168.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.

Antoniou, M., Ettlinger, M., & Wong, P. C. M. (2016). Complexity, Training Paradigm Design, and the Contribution of Memory Subsystems to Grammar Learning. *PLOS ONE*, *11*(7), e0158812. doi:10.1371/journal.pone.0158812

Ashby, F. G., & Maddox, W. T. (2005). Human Category Learning. *Annual Review of Psychology*, *56*(1), 149–178. doi:10.1146/annurev.psych.56.091103.070217

Atkins, P.W.B., & Baddeley, A.D. (1998) Working memory and distributed vocabulary learning. *Applied Psycholinguistics*, *19*, 537–552.

Baddeley, A. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences*, *93*(24), 13468-13472. doi:10.1073/pnas.93.24.13468

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417-423.

Baddeley, A. D. (2015). Working memory in second language learning. In Z. Wen & M. Borges Mota (Eds.), *Working memory in second language acquisition and processing* (pp. 17–28). Bristol, UK: Multilingual Matters.

Baddeley, A. D., Emslie, H., & Nimmo-Smith, I. (1994). *The Doors and People Test: A Test of Visual and Verbal Recall and Recognition*. Bury St. Edmunds: Thames Valley Test Company.

Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G.A. Bower (Ed.), *Recent advances in learning and motivation* (pp. 47–89). New York: Academic Press.

Baddeley, A.D., & Logie, R.H. (1999). Working memory: The multiple component model. In A. Miyake and P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). New York: Cambridge University Press.

Bailey, N., Madden, C., & Krashen, S. (1974). Is there a "natural sequence" in adult second language learning? *Language Learning*, *21*, 235-243.

Barker, L. (2012). Defining the Parameters of Incidental Learning on a Serial Reaction Time (SRT) Task: Do Conscious Rules Apply? *Brain Sciences*, *2*(4), 769–789.

Barr, D.J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effect structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.

Barry, E. S. (2007). Does Conceptual Implicit Memory Develop? The Role of Processing Demands. *The Journal of Genetic Psychology*, *168*(1), 19–36. doi:10.3200/GNTP.168.1.19-36

Baten, K. (2011). Processability Theory and German case acquisition. *Language Learning*, *61*, 455–505.

Bates, D., Machler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using s4 classes. http://cran.R-project.org/package=lme4.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Hove: Lawrence Erlbaum Associates.

Bertels, J., Boursain, E., Destrebecqz, A., & Gaillard, V. (2015). Visual statistical learning in children and young adults: how implicit? *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.01541

Bialystok, E. (1979), Explicit and implicit judgments of L2 grammaticality. *Language Learning*, 29: 81–103. doi:10.1111/j.1467-1770.1979.tb01053.x

Bialystok, E., & Miller, B. (1999). The problem of age in second-language acquisition: Influences from language, structure, and task. *Bilingualism: Language and Cognition*, *2*(2), 127-145.

Bianchi, A. (2015). *L'esame neuropsicologico dell'adulto*. Firenze, Italia: Giunti O.S.

Birdsong, D. (2005). Nativelikeness and non-nativelikeness in L2A research. *IRAL*, *43*, 319–328.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*, 9–49.

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition*, *31* (2), 175-198.

Bliss, T.V.P., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of Phisiology*, *232*, 331-356.

Boyd, J. K., & Goldberg, A. E. (2012). Young children fail to fully generalize a novel argument structure construction when exposed to the same input as older learners. *Journal of Child Language*, *39*(3), 457–81.

Boyd, J. K., Gottschalk, E. A., & Goldberg, A. E. (2009). Linking Rule Acquisition in Novel Phrasal Constructions. *Language Learning*, *59*, 64–89.

Braine, M. D., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., & Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language*, *29*(5), 591–610.

Bresnan, J. (1982). *The mental representation of grammatical relations*. Cambridge, Mass.: MIT Press.

Brill-Schuetz, K.A., & Morgan-Short, K. (2014) The role of procedural memory in adult second language acquisition. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 260–265). Austin, Texas: Cognitive Science Society.

Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics*, *27*(2). doi:10.1017/S0142716406060243

Brooks, P. J., Kwoka, N., & Kempe, V. (2017). Distributional Effects and Individual Differences in L2 Morphology Learning: Determinants of L2 Morphology Learning. *Language Learning*, *67*(1), 171–207. doi:10.1111/lang.12204

Buffington, J., & Morgan-Short, K. (2018). Declarative and procedural memory as individual differences in second language aptitude. Manuscript submitted for publication.

Buyl, A., & Housen, A. (2015). Developmental stages in receptive grammar acquisition: A Processability Theory account. *Second Language Research*, *31*(4), 523–550.

Cabeza, R., & Moscovitch, M. (2013). Memory systems, processing modes, and components: Functional neuroimaging evidence. *Perspectives on Psychological Science*, *8*(1), 49-55. doi:10.1177/1745691612469033

Caffarra, P., Vezzadini, G., Dieci, F., Zonato, F., & Venneri, A. (2002). Rey-Osterrieth complex figure: normative values in an Italian population sample. *Neurological Sciences*, *22*(6), 443–447.

Caplan, D., Alpert, N., & Waters, G. (1998). Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *Journal of Cognitive Neuroscience*, *10*(4), 541-552.

Carlesimo, G.A., Buccione, I., Fadda, I., Graceffa, A., Mauri, M., & Lorusso, S. (2002). Standardizzazione di due test di memoria per uso clinico: Breve Racconto e Figura di Rey. *Nuova Rivista di Neurologia*, *12,* 1-13.

Carpenter, H. S. (2008). *A behavioral and electrophysiological investigation of* *different aptitudes for L2 grammar in learners equated for proficiency level* (Unpublished doctoral dissertation). Georgetown University, US.

Carpenter, H., Morgan-Short, K., & Ullman, M. T. (2009). Predicting L2 using declarative and procedural memory assessments: A behavioral and ERP investigation. Presented at the Georgetown University Round Table, Washington, DC.

Carroll, J. B., & Sapon, S. (1959). *Modern Language Aptitude Test* (1959 ed.). New York: The Psychological Corporation.

Casarotti, A., Papagno, C., & Zarino, B. (2014). Modified Taylor Complex Figure: Normative data from 290 adults. *Journal of Neuropsychology*, *8*(2), 186-198.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science, 8*(6), 500–508.

Cenoz, J. (2002). Age differences in foreign language learning. *ITL Review of Applied Linguistics*, *135/136*, 125-142.

Cenoz, J. (2003). The influence of age on the acquisition of English: General proficiency, attitudes and code mixing. In M. Garcia Mayo, & M.L. Garcia Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 77-93). Clevedon: Multilingual Matters.

Chan, C. (1992). *Implicit cognitive processes: Theoretical issues and applications in computer systems design* (Unpublished doctoral dissertation). University of Oxford, UK.

Chiu, C.Y.P., Schmithorst, V.J., Brown, R.D., Holland, S.K. & Dunn, S. (2006). Making memories: a cross-sectional investigation of episodic memory encoding in childhood using fMRI. *Developmental Neuropsychology*, *29*, 321–340.

Clahsen, H., & Felser, C. (2006). Continuity and shallow structures in language processing. *Applied Psycholinguistics, 27*, 107–126.

Clahsen, H., & Felser, C. (2017). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*. https://doi.org/10.1017/S0272263117000250

Clahsen, H., & Veríssimo, J. (2016). Investigating grammatical processing in bilinguals: The case of morphological priming. *Linguistic Approaches to Bilingualism*, *6*, 685–698.

Cleary, A. M., & Langley, M. M. (2007). Retention of the structure underlying sentences. *Language and Cognitive Processes, 22*, 614–628.

Cohen, M. J. (1997). *Children's Memory Scale*. San Antonio, TX: Psychological Corporation.

Conti-Ramsden, G., Ullman, M. T., & Lum, J. A. G. (2015). The relation between receptive grammar and procedural, declarative, and working memory in specific language impairment. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01090

Dai Prà, M., Pagliuca, D., Barletta Rodolfi, C., Gasparini, F., Ghidoni, E., et al. (2015). Una rassegna dei test neuropsicologici italiani. In A. Bianchi (Ed.), *L'esame neuropsicologico dell'adulto* (pp. 158-209). Firenze, Italia: Giunti O.S.

Daltrozzo, J., & Conway, C. M. (2014). Neurocognitive mechanisms of statistical-sequential learning: what do event-related potentials tell us? *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00437

Damasio, H., Grabowski, T., Tranel, D., Hichwa, R., & Damasio, A. (1996). A neural basis for lexical retrieval. *Nature*, *380*(6574), 499-505.

Davies, R. (2017). *PSYC402: Advanced Statistics* [Class materials]. Department of Psychology, Lancaster University, Lancaster, UK.

Davis, C., Bradshaw, C. M., & Szabadi, E. (1999). The Doors and People Memory Test: Validation of norms and some new correction formulae. *British Journal of Clinical Psychology*, *38*, 305-314.

de Graaf, R. (1997). The eXperanto Experiment. *Studies in Second Language Acquisition*, *19*, 249–276.

DeKeyser, R. M. (1994). Implicit and Explicit Learning of L2 Grammar: A Pilot Study. *Tesol Quarterly*, *28*(1), 188–194.

DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, *17*(3), 379–410.

DeKeyser, R. M. (1997). Beyond explicit rule learning. Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, *19*(2), 195–221.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–533.

DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty and J. Williams (eds.) *Handbook of Second Language Acquisition* (pp. 313-349). Malden, MA: Blackwell.

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97–113). Mahwah, NJ: Lawrence Erlbaum.

DeKeyser, R. M. (2009). Cognitive-psychological processes in second language learning. In M. Long & C. Doughty (Eds.), *Handbook of second language* *teaching* (pp. 119–138). Oxford, England: Wiley-Blackwell.

DeKeyser, R. M. (2012). Interactions Between Individual Differences, Treatments, and Structures in SLA. *Language Learning, 62*(2), 189-200.

DeKeyser, R. M. (2013). Age Effects in Second Language Learning: Stepping Stones Toward Better Understanding. *Language Learning, 63*, 52-67.

DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 94-112). Mahwah, NJ: Lawrence Erlbaum Associates.

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics, 31*, 413-438.

DeKeyser, R., & Larson-Hall, J. (2005). What does the Critical Period really mean? In J.F. Kroll, & A.M.B. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic perspectives* (pp. 88-108). Oxford: OUP.

Di Biase, B., & Kawaguchi, S. (2002). Exploring the typological plausibility of Processability Theory: language development in Italian second language and Japanese second language. *Second Language Research*, *18*(3), 274–302. doi:10.1191/0267658302sr204oa

Dienes, Z. (2004). Assumptions of subjective measures of unconscious mental states: Higher order thoughts and bias. *Journal of Consciousness Studies, 11*(9), 25–45.

Dienes, Z. (2012). Conscious versus unconscious learning of structure. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 337–364). Berlin, Germany: Mouton de Gruyter.

Dienes, Z., Altmann, G., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology:*

*Learning, Memory, and Cognition*, *21*, 1322–1338.

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bullettin & Review*, *4*(1), 3-23.

Dienes, Z., Broadbent, D. E., & Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 875–882.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, 22(5), 735-755.

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research, 69,* 338–351.

Dominey, P. F., & Hoen, M. (2006). Structure mapping and semantic integration in a construction-based neurolinguistic model of sentence processing. *Cortex*, *42*, 476–479. doi:10.1016/S0010-9452 (08)70381-2

Dominey, P. F., & Inui, T. (2009). Cortico-striatal function in sentence comprehension: Insights from neurophysiology and modeling. *Cortex*, *45*, 1012–1018. doi:10.1016/j.cortex.2009.03.007

Dörnyei, Z., & Ushioda, M. (2009). Motivation, language identities, and the L2 self: A theoretical overview. In Z. Dörnyei & M. Ushioda (Eds.), *Motivation, language identity, and the L2 self* (pp. 1-8). Bristol: Multilingual Matters.

Doughty, C. J. (2013). Optimizing post-critical-period language learning. In G. Granena, & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 153-175). Amsterdam, The Netherlands: John Benjamins.

Dulay, H., & Burt, M. (1974). Natural sequences in child second language acquisition. *Language Learning*, *24*, 37-53.

Egi, T. (2004). Verbal reports, noticing, and SLA research. *Language Awareness*, *13*, 243-264.

Eichenbaum, H. (2008). *Learning & Memory*. New York: W.W. Norton & Company.

Eichenbaum, H. (2012). *The cognitive neuroscience of memory: An introduction* (2nd ed.). Oxford, England: Oxford University Press.

Ellis, N. C. (1994). *Implicit and Explicit Learning of Languages*. London: Academic Press.

Ellis, N. C. (2005). At the interface: dynamic interaction of explicit and implicit knowledge. *Studies in Second Language Acquisition*, *27*, 305–352.

Ellis, N. C. & Wulff, S. (2015). Usage-based approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 75-93). New York, NY: Routledge.

Ellis, R. (1991). Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition*, 13(2), 161-186.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*, 141–172.

Ellis, R. (2009). Implicit and Explicit Learning, Knowledge and Instruction. In R. Ellis et al. (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 3-25). Bristol: Multilingual Matters.

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J. & Reinders, H. (2009). *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*. Multilingual Matters.

Embick, D., Marantz, A., Miyashita, Y., O'Neill, W., & Sakai, K. L. (2000). A syntactic specialization for Broca's area. *Proceedings of the National Academy of Sciences USA*, *97*, 6150-6154.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27* (3), 464-491.

Erlam, R. (2009). The Elicited Imitation Test as a Measure of Implicit Knowledge. In R. Ellis et al. (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 65-93). Bristol: Multilingual Matters.

Ettlinger, M., Bradlow, A., & Wong, P. C. M. (2014). Variability in the learning of complex morphophonology. *Applied Psycholinguistics*, *35*(4), 807-831.

Faretta-Stutenberg, M., & Morgan-Short, K. (2017). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research*, *34*(1), 67-101.

Fathman, A. (1975). The relationship between age and second language productive ability. *Language Learning*, *25*, 245-266.

Ferman, S., & Karni, A. (2010). No Childhood Advantage in the Acquisition of Skill in Using an Artificial Language Rule. *PLoS ONE*, *5*(10), e13648. doi:10.1371/journal.pone.0013648

Ferman, S., & Karni, A. (2014). Explicit versus Implicit Instruction: Which Is Preferable for Learning an Artificial Morphological Rule in Children? *Folia Phoniatrica et Logopaedica*, *66*(1–2), 77–87.

Ferraro, F. R., Balota, D. A., & Connor, L. T. (1993). Implicit memory and the formation of new associations in nondemented Parkinson’s disease individuals with senile dementia of the Alzheimer type: a serial reaction time (SRT) investigation. *Brain and Cognition*, *21*, 163-180.

Fisher, C. (2002). Structural limits on verb mapping: The role of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science*, *5*, 56–65.

Francis, A.P., Schmidt, G.L., Carr, T. H., & Clegg, B.A. (2009). Incidental learning of abstract rules for non-dominant word orders. *Psychological Research, 73*(1), 60-74.

Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84.

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, 99*, 529–534.

Friederici, A. D. & von Cramon, D. Y. (2000). Syntax in the brain: Linguistic versus neuroanatomical specificity. *Behavioral and Brain Sciences*, *23*(1), 32-33.

Fritzsche, E. S., Kröner, S., Dresel, M., Kopp, B., & Martschinke, S. (2012). Confidence scores as measures of metacognitive monitoring in primary students? (Limited) validity in predicting academic achievement and the mediating role of self-concept. *Journal for Educational Research Online*, *4*(2), 120.

García Mayo, M. (2003). Age, length of exposure and grammaticality judgements in the acquisition of English as a foreign language. In M. García Mayo, & M.L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 94-114). Clevedon: Multilingual Matters.

García Mayo, M., & García Lecumberri, M. L. (2003). English FL sounds in school learners od different ages. In M. García Mayo, & M.L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 115-135). Clevedon: Multilingual Matters.

Gagné, H.M., & Cohen, H. (2016). Interference effects between memory systems in the acquisition of a skill. *Experimental Brain Research*, *234*(10), 2883-2891.

Gass, S.M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S. & Mackey, A. (2007). *Data Elicitation for Second and Foreign Language Research*. Mahwah, NJ: Lawrence Erlbaum.

Gathercole, S.E., Adams, A.M. & Hitch, G.J. (1994). Do young children rehearse? An individual-differences analysis. *Memory & Cognition*, *22*, 201-207. https://doi.org/10.3758/BF03208891

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory From 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, *17*, 648-691.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., et al. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience*, *2*(10), 861-862.

Godfroid, A., Loewen, S., Jung, S., Park, J. H., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition*, *37*(2), 269-297.

Goldberg A. E., Casenhiser, D., & Sethuraman, N. (2004). Learning Argument Structure Generalizations. *Cognitive Linguistics*, 15, 289-316.

Goldberg A. E., Casenhiser, D., & White, T.R. (2007). Constructions as Categories of Language. *New Ideas in Psychology*, *25*(2), 70-86.

Goldschneider, J. M., & DeKeyser, R. M. (2005), Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, *55*: 27–77. doi:10.1111/j.0023-8333.2005.00295.x

Golomb, J., Kluger, A., de Leon, M. J., Ferris, S. H., Mittelman, M., Cohen, J., & George, A. E. (1996). Hippocampal formation size predicts declining memory performance in normal aging. *Neurology*, *47*(3), 810-813.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431-436.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109-135.

Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. In G. Granena, & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105-129). Amsterdam, The Netherlands: John Benjamins.

Granena, G., & Long, M. (2013a). Age of onset, length of residence, language aptitude and ultimate attainment in three linguistic domains. *Second Language Research, 29*(1).

Granena, G., & Long, M. (2013b). *Sensitive periods, language aptitude, and ultimate L2 attainment*. Amsterdam: John Benjamins.

Gregg, K. R. (1996). The logical and developmental problems of second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 49–81). San Diego: Academic Press.

Grey, S. (2014). *Natural patterns of language learning in adulthood: bilingual vs. monolingual* (Unpublished doctoral dissertation). Georgetown University, US.

Grey, S., Williams, J., & Rebuschat, P. (2015). Incidental Exposure and L3 Learning of Morphosyntax. *Studies in Second Language Acquisition*, 1-34.

Hakuta, K. (2001). A critical period for second language acquisition? In D.B. Bailey, J.T.Bruer, F.J. Symons, & J.W. Lichtman (eds.), *Critical thinking about critical periods* (pp. 193–205). Baltimore: Paul Brookes.

Hakuta, K., Bialystok, E., & Wiley (2003). Critical Evidence: A Test of the Critical-Period Hypothesis for Second-Language Acquisition. *Psychological Science 14*(1): 31-38.

Hamrick, P. (2015). Declarative and procedural memory abilities as individual differences in incidental language learning. *Learning and Individual Differences*, *44*, 9–15. doi:10.1016/j.lindif.2015.10.003

Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences*, *115*(7), 1487–1492. doi:10.1073/pnas.1713975115

Han, Y. (2000). Grammaticality judgment tests: How reliable and valid are they? *Applied Language Learning*, 11(1), 177-204.

Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, *2*(1), 1-23.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, *19*(3), 379-400.

Hebb, D. (1949). *The Organization of Behavior*. New York: Wiley & Sons.

Hedenius, M. (2013). Procedural and Declarative Memory in Children with Developmental Disorders of Language and Literacy. *Acta Universitatis Upsaliensis. Digital Comprehensive Summaries of Uppsala Dissertations from* *the Faculty of Medicine, 917*. Uppsala. ISBN 978-91-554-8707-2.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews. Neuroscience*, *11*(7), 523-532.

Herschensohn, J. (2007). *Language development and age*. Cambridge: CUP.

Howard, J. H., Jr., & Howard, D. V. (1997). Age differences in implicit learning of higher-order dependencies in serial patterns. *Psychology and Aging*, *12*(4), 634-656.

Howard, D. V., Howard, J. H., Japikse, K., DiYanni, C., Thompson, A., & Somberg, R. (2004). Implicit Sequence Learning: Effects of Level of Structure, Adult Age, and Extended Practice. *Psychology and Aging*, *19*(1), 79–92. doi:10.1037/0882-7974.19.1.79

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development*, *1*(2), 151–195.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30–66.

Hulstijn, J.H. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, *18*, 193-223.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. Doughty, & M.H. Long (Eds.), *The handbook of second language acquisition* (pp. 539- 588). Oxford: Blackwell.

Jackson, G. M., Jackson, S. R., Harrison, J., Henderson, L., & Kennard C. (1995). Serial reaction time learning and Parkinson's disease: evidence for a procedural learning deficit. *Neuropsychologia*, *33*, 577-593.

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science*, *15*(4), 496–505. doi:10.1111/desc.2012.15.issue-4

Janacsek, K., & Nemeth, D. (2013). Implicit sequence learning and working memory: Correlated or complicated? *Cortex*, *49*(8), 2001–2006. doi:10.1016/j.cortex.2013.02.012

Jiménez, L. (2003). *Attention and Implicit Learning*. Amsterdam: John Benjamins.

Jiménez, L., & Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 236-259.

Kapa, L. L., & Colombo, J. (2014). Executive function predicts artificial language learning. *Journal of Memory and Language*, *76*, 237–252.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.

Keatinge, D., & Keßler, J-U. (2009). The Acquisition of the passive voice in English as a foreign language: Production and perception. In J-U. Keßler & D. Keatinge (Eds.), *Research in second language acquisition: Empirical evidence across languages* (pp. 41-68). Newcastle upon Tyne: Cambridge Scholars.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171-184.

Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: MIT Press.

Knopman, D. S., & Nissen (1987). Implicit learning in patients with Alzheimer's disease. *Neurology*, *37*, 784-788.

Koriat, A., & Ackerman, R. (2010). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science*, *13*(3), 441–453. doi:10.1111/j.1467-7687.2009.00907.x

Kormos, J. (2013). New conceptualizations of language aptitude in second language attainment. In G. Granena, & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 131-152). Amsterdam, The Netherlands: John Benjamins.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, *11*(2). doi:10.1017/S1366728908003416

Krashen, S. (1977). Some issues relating to the Monitor Model. In H. D. Brown and R. Crymes (Eds.). *On TESOL '77. Teaching and learning English as a second language: Trends in research and practice* (pp. 144-158). Washington: TESOL.

Krashen, S., Long, M., & Scarcella, R. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL Quarterly*, *13*, 573-582.

Krashen, S., Sferlazza, V., Feldman, V., & Fathman, A. (1976). Adult performance on the SLOPE test: More evidence for a natural sequence in adult second language acquisition. *Language Learning*, *26*, 145-151.

Lapkin, S., Swain, M., Kamin, J., & Hanna, G. (1980). *Report of the 1979 evaluation of the Peel County late French immersion program, grades 8, 10, 11 and 12.* Unpublished report, University of Toronto, OISE.

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research*, *24*(1), 35-63.

Leow, P. L., & Zamora, C. C. (2017). Intentional and Incidental L2 Learning. In S. Loewen and M. Sato (Eds.). *The Routledge Handbook of Instructed Second Language Acquisition*. Routledge Handbooks Online (accessed 12 Sept 2018).

Lichtman, K. M. (2012). *Child-adult differences in implicit and explicit second language learning*. Unpublished PhD dissertation, University of Illinois at Urbana-Champaign.

Lichtman, K. M. (2013). Developmental Comparisons of Implicit and Explicit Language Learning. *Language Acquisition, 20*(2), 93-108.

Lichtman, K. M. (2016). Age and learning environment: Are children implicit second language learners? *Journal of Child Language*, *43*(3), 707–730. doi:10.1017/S0305000915000598

Llanes, À., & Muñoz, C. (2013). Age Effects in a Study Abroad Context: Children and Adults Studying Abroad and at Home. *Language Learning*, *63*(1), 63–90. doi:10.1111/j.1467-9922.2012.00731.x

Loewen, S. (2009). Grammaticality Judgment Tests and the Measurement of Implicit and Explicit Knowledge. In R. Ellis et al. (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 94-113). Bristol: Multilingual Matters.

Long, M. H. (2007). *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lum, J. A. G., Kidd, E., Davis, S., & Conti-Ramsden, G. (2010). Longitudinal study of declarative and procedural memory in primary school-aged children. *Australian Journal of Psychology*, *62*(3), 139e148.

Lum, J. A. G., Gelgic, C., & Conti-Ramsden, G. (2010). Procedural and declarative memory in children with and without specific language impairment. *International Journal of Language & Communication Disorders*, *45*(1), 96–107. https://doi.org/10.3109/13682820902752285

Mackey, A., & Gass, S. M. (2005). *Second language research. Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (1983). Miniature linguistic systems as tests of the use of universal operating principles in second-language learning by children and adults. *Journal of Psycholinguistic Research*, *12*(5), 467–478.

Martin, A., Ungerleider, L. G., & Haxby, J. V. (2000). Category specificity and the brain: the sensory/motor model of semantic representation of objects. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1023-1036). Cambridge, MA: MIT Press.

Matsui, T., Miura, Y., & McCagg, P. (2006). Young children’s implicit and explicit understanding of speaker knowledge. *Proceedings of The 28th Annual Conference of The Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Mayor-Dubois, C., Zesinger, P., Van der Linden, M., & Roulet-Perez, E. (2016). Procedural learning: A developmental study of motor sequence learning and probabilistic classification learning in school-aged children. *Child Neuropsychology*, *22*(6), 718-734.

McDaniel, D., & Cairns, H.S. (1996). Eliciting grammaticality judgments from children . In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methodology in assessing children's syntax* (pp. 233—244). Cambridge, MA: MIT Press.

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2010). The neural basis of speech parsing in children and adults. *Developmental Science*, *13*(2), 385–406. doi:10.1111/j.1467-7687.2009.00895.x

Mecklenbräuker, S., Hupbach, A., & Wippich, W. (2003). Age-related improvements in a conceptual implicit memory test. *Memory & Cognition*, *31*(8), 1208–1217.

Meulemans, T., Linden, M. V. der, & Perruchet, P. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology*, *69*, 199-221.

Milton, J., & Alexiou, T. (2006). Language aptitude development in young learners. In C. Abello-Contesse, R. Chacón-Beltrán, M. D. López-Chiménez & M. M. Torreblanca-López (Eds.), *Age in L2 acquisition and teaching* (pp. 177-192). Oxford: Peter Lang.

Monaco, M., Costa, A., Caltagirone, C., & Carlesimo, G. A. (2013). Forward and backward span for verbal and visuo-spatial data: Standardization and normative data from an Italian adult population. *Neurological Sciences*, *34*(5), 749–754. doi:10.1007/s10072-012-1130-x

Mondini, S., Mapelli, D., Vestri, A., Arcara G., & Bisiacchi, P. S. (2011). *Esame neuropsicologico breve 2*. Milano, Italia: Raffaello Cortina Editore.

Mora, J. C. (2006). Age Effects on Oral Fluency Development. In C. Muñoz (Ed.), *Age and Rate of Foreign Language Learning* (pp. 65-88). Clevedon: Multilingual Matters.

Morgan-Short, K. (2007). *A neurolinguistic investigation of late-learned second language knowledge: The effects of explicit and implicit conditions* (Doctoral dissertation, Georgetown University). Retrieved from https://uic.academia.edu/KaraMorganShort

Morgan-Short, K., Deng, Z., Brill-Schuetz, K. A., Faretta-Stutenberg, M., Wong, P., & Wong, F. C. K. (2015). A view of the neural representation of second language syntax through artificial language learning under implicit contexts of exposure. *Studies in Second Language Acquisition*, *37*(2), 383–419.

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K.A., Carpenter, H. & Wong P.C.M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition, 17*, 56-72.

Morgan-Short, K., Finger, I., Grey, S., & Ullman, M. T. (2012). Second language processing shows increased native-like neural responses after months of no exposure. *PLoS ONE, 7*(3), e32974.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second Language Acquisition of Gender Agreement in Explicit and Implicit Training Conditions: An Event-Related Potential Study. *Language Learning*, *60*(1), 154–193.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24*, 933–947.

Morgan-Short, K., & Ullman, M. T. (2011). The neurocognition of second language. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 282-299). Oxford: Routledge. doi:10.4324/9780203808184.ch17

Moro, A., Tettamanti, M., Perani, D., Donati, C., Cappa, S. F., & Fazio, F. (2001). Syntax and the brain: disentangling grammar by selective anomalies. *Neuroimage*, *4*, 257-267.

Munnich, E., & Landau, B. (2010). Developmental decline in the acquisition of spatial language. *Language Learning and Development, 6*(1), 32-59.

Muñoz, C. (2003). Variation in oral skills development and age of onset. In M. García Mayo, & M.L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 161-181). Clevedon: Multilingual Matters.

Muñoz, C. (2006a). The Effects of Age on Foreign Language Learning: The BAF Project. In C. Munoz (Ed.), *Age and Rate of Foreign Language Learning* (pp. 1-40). Clevedon: Multilingual Matters**.**

Muñoz, C. (2006b). Accuracy orders, rate of learning and age in morphological acquisition. In C. Munoz (Ed.), *Age and Rate of Foreign Language Learning*. Clevedon: Multilingual Matters**,** 107-126.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, *24*(4), 578–596.

Muñoz, C. (2011). Is input more significant than starting age in foreign language acquisition? *International Review of Applied Linguistics*, *49*, 113-133.

Newman, A. J., Pancheva, R., Ozawa, K., Neville, H. J., & Ullman, M. T. (2001). An event-related fMRI study of syntactic and semantic violations. *Journal of Psycholinguistic Research*, *30*(3), 339-364.

Newport, E. (2002). Critical periods in language development. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 737-739). London: Macmillian/Nature Publishing Group.

Ni, W., Constable, R. T., Menci, W. E., Pugh, K. R., Fullbright, R. K., Shaywitz, S. E., Gore, J. C., & Shankweiler, D. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *Journal of Cognitive Neuroscience*, 12(1), 120-133.

Nicolay, A.C., & Poncelet, M. (2013) Cognitive abilities underlying second-language vocabulary acquisition in an early second-language immersion education context: A longitudinal study. *Journal of Experimental Child Psychology, 115*, 655–671. doi:10.1016/jecp.2013.04.002

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, 19(1), 1-32.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-analysis. *Language Learning, 50*(3), 417-528.

Novelli, G., Papagno, C., Capitani, E., Laiacona, M., Cappa, S., & Vallar, G. (1986). Tre test clinici di memoria verbale a lungo termine. Taratura su soggetti normali. *Archivio di Psicologia, Neurologia e Psichiatria*, 47(2), 278-296.

Ofen, N., Kao, Y.-C., Sokol-Hessner, P., Kim, H., Whitfield-Gabrieli, S., & Gabrieli, J. D. E. (2007). Development of the declarative memory system in the human brain. *Nature Neuroscience*, *10*(9), 1198–1205. https://doi.org/10.1038/nn1950

Osterrieth, P. A. (1944). Le test de copie d'une figure complexe: Contribution al'etude de la perception et de la memoire [Copying a complex figure: Contributions to the study of perception and memory]. *Archives de Psychologie*, *30*, 206–356.

Otwinowska, A., & De Angelis, G. (2012). Introduction: social and affective factors in multilingualism research. *International Journal of Multilingualism*, *9*(4), 347–351. doi:10.1080/14790718.2012.714379

Packard, M. G., & Goodman, J. (2013). Factors that influence the relative use of multiple memory systems. *Hippocampus*, *23*(11), 1044–1052. doi:10.1002/hipo.22178

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. C. Ellis (Ed.), *Implicit and* *explicit learning of languages* (pp. 393–419). London: Academic Press.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam, Netherlands: John Benjamins.

Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, *37*, 607–642.

Perruchet, P., & Pachton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, 10(5), 233-238.

Pfenninger, S. E. (2011). Age effects on the acquisition of nominal and verbal inflections in an instructed setting. *Studies in Second Language Learning and Teaching* , *1*(3), 401–420. doi:10.14746/ssllt.2011.1.3.5

Pfenninger, S. E. (2014). The misunderstood variable: Age effects as a function of type of instruction. *Studies in Second Language Learning and Teaching*, *3*, 529–556. doi:10.14746/ssllt.2014.4.3.8

Pfenninger, S. E. and D. Singleton. (2017). *Beyond age effects in Instructional L2 learning: Revisiting the age factor*. Bristol: Multilingual Matters.

Philp, J. (2003). Constrains on noticing the gap: Nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition*, *25*, 99-126.

Pica, T. (1983). Adult acquisition of English as a second language under different conditions of exposure. *Language Learning*, *33*, 465–497. doi:10.1111/j.1467-1770.1983.tb00945.x

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics, 10*, 52-79.

Pienemann, M. (1998). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins.

Pienemann, M. (2005). *Cross-linguistic aspects of Processability Theory*. Amsterdam: John Benjamins.

Pienemann, M., Johnston, M., & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition*, *10*, 217-224.

Pienemann, M., & Lenzing, A. (2015). Processability Theory. In B. VanPatten, & J. Williams (Eds.), *Theories of second language acquisition: An introduction* (2nd ed., pp. 159-179). Mahwah: NJ: Lawrence Erlbaum Associates.

Pienemann, M., & Mackey, A. (1993). An empirical study of children's ESL development and rapid profile.  In P. McKay (Ed.), *ESL development. Language and literacy in schools*, (pp. 115-259). Commonwealth of Australia and National Languages and Literacy Institute of Australia.

Pili-Moss, D. (2017). Tracking the early stages of child and adult comprehension of L2 morphosyntax: A pilot study. *Journal of the European Second Language Association*, *1*(1), 113–125. doi:10.22599/jesla.25

Pili-Moss, D. & Morgan-Short, K. (2017). Contributions of declarative and procedural memory to second language development. Paper delivered at the Sixth Implicit Learning Seminar, 25-26 May, ELTE University, Budapest, Hungary.

Pili-Moss, D., Brill-Schuetz, K., Faretta-Stutenberg, M., & Morgan-Short, K. (2018). Contributions of Declarative and Procedural Memory to Accuracy and Automatization During L2 Practice. Manuscript under review.

R Development Core Team (2016) R: A Language and Environment for Statistical

Computing. R Foundation for Statistical Computing, Vienna.

Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*(6), 855-863.

Reber, A. (1976). Implicit learning of synthetic learners: The role of instructional set. *Journal of Experimental Psychology, Human Learning and Memory, 2*, 88-94.

Reber, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219-235.

Reber, A., Walkenfeld, F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11*, 888-896.

Reber, P. J. (2013). The neural basis of implicit learning and memory: A review of neurophychological and neuroimaging research. *Neuropsychologia*, *51*, 2026-2042.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research: A review. *Language Learning*, *63,* 595–626.

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics, 33,* 829–856.

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique [Psychological examination of cases of traumatic encephalopathy]. *Archives de Psychologie*, *28*, 215–285

Rizzi, L. (1982). *Issues in Italian syntax*. Dordrecht: Foris.

Robertson, E. M. (2007). The Serial Reaction Time Task: Implicit Motor Skill Learning? *Journal of Neuroscience*, *27*(38), 10073–10075.

Robinson, P. (1995). Attention, memory, and the "Noticing" hypothesis. *Language Learning*, *45*(2), 283-331.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition*, *18*, 27–67.

Robinson, P. (2001). Individual differences. Cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, *17,* 368-392.

Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA: A replication and extension of Reber, Walkenfeld and Hernstadt, 1991. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211-266). Amsterdam: John Benjamins.

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 leaning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition*, *27*, 235–268.

Robinson, P. (2010). Implicit Artificial Grammar and Incidental Natural Second Language Learning: How Comparable Are They? *Language Learning, 60*(2), 245-263.

Rogers, J., Révész, A., & Rebuschat, P. (2015). Implicit and explicit knowledge of inflectional morphology. *Applied Psycholinguistics*, 1–32.

Rogers, V. E., Meara, P., Aspinall, R., Fallon, L., Goss, T., Keey, E., & Thomas, R. (2016). Testing aptitude: Investigating Meara's (2005) LLAMA tests. *EUROSLA Yearbook*, *16*, 179–210. https://doi.org/10.1075/eurosla.16.07rog

Rosa, E. M., & Leow, R. P. (2004). Awareness, different learning conditions, and second language development. *Applied Psycholinguistics*, *25*(2), 269–292.

Rosa, E. M., & O'Neill, M. D. (1999). Explicitness, intake, and the issue of awareness: Another piece to the puzzle. *Studies in Second Language Acquisition, 21*(4), 511–556.

Sáfár, A., & Kormos, J. (2008). Revisiting problems with foreign language aptitude. *International Review of Applied Linguistics in Language Teaching, 46*, 113-136.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*(2), 101-105.

Saxton, M., Dockrell, J., Bevan, E., & Van Herwegen, J. (2008). Grammaticality judgements of children with and without language delay. In T. Marinis, A. Papangeli, & V. Stojanovik (Eds.), *Proceedings of the Child Language Seminar 2007* (pp. 136-150). Reading, UK: University of Reading.

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammar and SLA. In N. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp. 165-209). London: Academic Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction*. Cambridge: CUP.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge: New York/London.

Service, E. (1992) Phonology, working memory and foreign-language learning. *Quarterly Journal of Experimental Psychology, 45A*(1), 21–50.

Shanks, D. R. (2003). Attention and awareness in implicit sequence learning. In L. Jimenez (Ed.) *Attention and Implicit Learning* (pp. 11-42). Philadelphia, PA:

John Benjamins.

Sharwood Smith, M. (1981). Consciousness-raising and the second language learner. *Applied Linguistics, 2*, 159-169.

Smith, J., Siegert, R. J., McDowall, J., & Abernethy, D. (2001). Preserved Implicit Learning on Both the Serial Reaction Time Task and Artificial Grammar in Patients with Parkinson's Disease. *Brain and Cognition*, *45*(3), 378–391.

Snow, C., & Hoefnagel-Hoehle, M. (1978). Age differences in second language acquisition. In E. Hatch (Ed.), *Second Language Acquisition: A Book of Readings* (p. 333-344). Rowley, MA: Newbury Press.

Song, S., Howard, J. H., & Howard, D. V. (2007). Implicit probabilistic sequence learning is independent of explicit awareness. *Learning and Memory*, *14*, 167–176.

Spada, N., & Tomita, Y. (2010). Interactions Between Type of Instruction and Type of Language Feature: A Meta-Analysis. *Language Learning*, *60*(2), 263-308.

Spada, N., Shiu, J. L.-J., & Tomita, Y. (2015). Validating an Elicited Imitation Task as a Measure of Implicit Knowledge: Comparisons With Other Validation Studies. *Language Learning*, *65*, 723–751. doi:10.1111/lang.12129

Spinner, P. (2013). Language Production and Reception: A Processability Theory Study. *Language Learning*, *63*(4), 704–739. doi:10.1111/lang.12022

Squire, L. R. (1987). *Memory and brain*. New York: Oxford University Press.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171-177.

Squire, L. R., & Wixted, J. T. (2011). The cognitive neuroscience of human memory since HM. *Annual Review of Neuroscience*, *34*. doi:10.1146/annurev-neuro-061010-113720

Steinhauer, K. (2014). Event-related Potentials (ERPs) in second language research: A brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2. *Applied Linguistics*, *35*(4), 393-417. doi:10.1093/applin/amu028

Stone, H. A., & Pili-Moss, D. (2016). Executive Function and Language Learning: Differentiating Vocabulary and Morpho-Syntax. In E. Nichele, D. Pili-Moss, & C. Sitthirak (Eds.), *Papers from the 10th Lancaster Postgraduate Conference in Linguistics and Language Teaching* (pp. 53-74). Lancaster: Lancaster University.

Suzuki, Y. (2017a). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics*, 38(5), 1229-1261.

Suzuki, Y. (2017b). The role of procedural learning ability in automatization of morphology under different learning schedules. *Studies in Second Language Acquisition*, 1–15. doi:10.1017/S0272263117000249

Suzuki, Y., & DeKeyser, R. (2015), Comparing Elicited Imitation and Word Monitoring as Measures of Implicit Knowledge. *Language Learning*, *65*, 860–895. doi:10.1111/lang.12138

Tagarelli, K. (2014). The Neurocognition of Adult Second Language Learning: An fMRI Study. Unpublished PhD dissertation. Georgetown University.

Tagarelli, K. M., Borges Mota, M., & Rebuschat, P. (2015). Working Memory, Learning Conditions and the Acquisition of L2 Syntax. In Z. Wen (Ed.), *Working Memory in Second Language Acquisition and Processing* (pp. 125-137). Bristol: Multilingual Matters.

Taylor, L. B., (1998). Scoring criteria for the Rey-Osterrieth Complex Figure Test. In

O. Spreen, & E. Strauss (Eds.), *A compendium of neuropsychological tests. Administration, norms, and commentary* (pp. 350-351). New York: Oxford University Press.

Tellier, A., & Roehr-Brackin, K. (2013). Awareness in primary-school children: A classroom study. *Essex Research Reports in Linguistics*, *62*(1). Colchester, UK: University of Essex.

Theakston, A. L. (2004). The role of entrenchment in children's and adults' performance on grammaticality-judgement tasks. *Cognitive Development*, *19*(1), 15–34.

Thomas, K. M., Hunt, R. H., Vizueta, N., Sommer, T., Durston, S., Yang, Y., & Worden, M. S. (2004). Evidence of developmental differences in implicit sequence learning: An fMRI study of children and adults. *Journal of Cognitive Neuroscience*, *16*(8), 1339–1351.

Tomlin, R., & Villa, V. (1994). Attention in Cognitive Science and Second Language Acquisition. *Studies in Second Language Acquisition*, *16*(2), 183-203. doi:10.1017/S0272263100012870

Torras, M. R., Navés, T., Celaya, M. L., Pérez-Vidal, C. (2006). Age and interlanguage development in writing. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 156-182). Bristol: Multilingual Matters.

Trahan, D. E., & Larrabee, G. J. (1988). *Continuous Visual Memory Task: Professional Manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*, 105– 122.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*, 231–270.

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Processing approaches to adult SLA: Theory and practice* (pp. 141–178). Washington, DC: Georgetown University Press.

Ullman, M. T. (2015). The declarative/procedural model: A neurobiologically-motivated theory of first and second language. In B. VanPatten, & J. Williams (Eds.), *Theories of second language acquisition: An introduction* (2nd ed., pp. 135-158). Mahwah: NJ: Lawrence Erlbaum Associates.

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge and use. In G. Hickok, & S. A. Small (Eds.), *The neurobiology of language* (pp. 953-968) Elsevier. doi:10.1016/B978-0-12-407794-2.00092-4

VanPatten, B. (1994). Evaluating the role of consciousness in second language acquisition: Terms, linguistic features, and research methodology. *AILA Review*, *11*, 27–36.

Veríssimo, J., Heyer, V., Jacob, G., & Clahsen, H. (2017). Selective effects of age of acquisition on morphological priming: Evidence for a sensitive period. *Language Acquisition*. doi:10.1080/10489223.2017.1346104

Vicari, S. (2007). *Prove di memoria e apprendimento per l'età evolutiva (PROMEA)*. Firenze, Italia: Giunti O. S.

Wechsler, D. (1945). A standard memory scale for clinical use. *Journal of Psychology*, *19*, 87-95.

Williams, J. N. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition*, *21*(1), 1–48.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*, 269–304.

Williams, J. N. (2010). Initial Incidental Acquisition of Word Order Regularities: Is It Just Sequence Learning? *Language Learning*, *60*(s2), 221–244.

Williams, J. N., & Lovatt, P. (2003). Phonological Memory and Rule Learning. *Language Learning*, *53*(1), 67–121.

Williams, J. N., & Kuribara, C. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua*, *118*, 522–553.

Wonnacott, E., Boyd, J. K., Thomson, J., & Goldberg, A. E. (2012). Input effects on the acquisition of a novel phrasal construction in 5 year olds. *Journal of Memory and Language*, *66*(3), 458 – 478.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, 56(3), 165-209.

Zappalà, G., Measso, G., Cavarzeran, F., Grigoletto, F., Lebowitz, B., Pirozzolo, F., et al. (1995). Aging and memory: corrections for age, sex and education for three widely used memory tests. *The Italian Journal of Neurological Sciences*, *16*(3), 177–184.

*This page is intentionally left blank*

**Appendix A**

**Suzy's script (Language Training)**

<u>Block 1</u>

*Ciao, mi chiamo Suzy. Adesso che hai imparato le pedine e le mosse di Brocanto, osserva alcune mosse del gioco e ascolta come si dicono. Ascolta con attenzione! Se lo fai avrai più possibilità di ottenere un punteggio alto nel gioco. Pronti.... via!*

[Hi, my name is Suzy. Now that you have learnt the tokens and moves of Brocanto, watch some Brocanto moves and listen to how to say them. Ready.... go!]

----------------------------------

*Perfetto! Adesso farai la tua prima vera partita. In bocca al lupo!*

[Well done! Now you will play your first Brocanto game. Good luck!]

<u>Block 2</u>

*Eccomi di nuovo! La prima volta che hai giocato a Brocanto forse è stato un pò difficile. Ma vedrai che andrà sempre meglio con un pò di pratica! Guarda ancora qualche mossa e ascolta con attenzione prima di fare un'altra partita. Pronti.... via!*

[Here I am again! The first time you played with Brocanto maybe it was a bit hard. But you will get better and better at it with a bit of practice! Watch some more Brocanto moves and listen carefully before you play another game. Ready.... go!]

----------------------------------

*Fantastico! Adesso puoi fare un'altra partita! In bocca al lupo!*

[Great! Now you can play another game! Good luck!]

<u>Block 3</u>

*Ciao! Sono di nuovo io! Osserva ancora qualche mossa di Brocanto e ascolta con attenzione prima della tua prossima partita. Pronti.... via!*

[Hi! It's me again! Watch some more moves and listen carefully before you start your next game. Ready.... go!]

----------------------------------

*Fantastico! Adesso puoi fare un'altra partita! In bocca al lupo!*

[Great! Now you can play another game! Good luck!]

<u>Block 4</u>

*Eccomi qui! Sono sicura che adesso il tuo punteggio è molto più alto di quando abbiamo iniziato. Ascolta con attenzione e farai ancora meglio. Pronti.... via!*

[Here I am! I am sure that now your score is much better than when we started. Listen carefully and you will do even better. Ready.... go!]

----------------------------------

*Fantastico! Adesso puoi fare un'altra partita! In bocca al lupo!*

[Great! Now you can play another game! Good luck!]

<u>Block 5</u>

*Ciao! Guarda e ascolta qualche altra mossa di Brocanto per migliorare ancora di più il tuo punteggio. Pronti.... via!*

[Hi! Watch and listen to some more Brocanto moves to earn even more points. Ready.... go!]

----------------------------------

*Fantastico! Adesso puoi fare un'altra partita! In bocca al lupo!*

[Great! Now you can play another game! Good luck!]

<u>Block 6</u>

*Ciao! Questa è l'ultima opportunità che hai di guardare e ascoltare le mosse di Brocanto per migliorare il tuo punteggio. Pronti.... via!*

[Hi! This is your last chance to watch and listen to some more Brocanto moves to improve your score. Ready.... go!]

-----------------------------------

*Perfetto! Adesso giocherai a Brocanto per l'ultima volta. In bocca al lupo e grazie per la tua attenzione. A presto!*

[Well done! Now you will play Brocanto for the last time. Good luck and thank you for watching. See you soon!]

Break after 16 moves in each block

*Ci siamo quasi! Ce ne sono ancora un paio.*

[We have almost finished. Just a few more]

**Appendix B**

**Passive exposure set**

<u>Block 1</u>

| | |
|---|---|
| **Nouns:** | **pleca; vode; blomi, nipo** |
| **Case particles:** | **ri (NOM); ru (ACC)** |
| **Adjectives:** | **neimo; troise** |
| **Verbs:** | **nima; prazi; yabe; klino** |
| **Adverbs:** | **noika; zeima** |

**SOV = 14 OV = 8 SV = 2**
**symmetric = 18 asymmetric = 6**

| | | | | |
|---|---|---|---|---|
| pleca ri | vode ru | nima | | SOV |
| pleca ri | vode ru | noika | nima | SOV |
| pleca ri | trose vode ru | nima | | SOV |
| neimo pleca ri | vode ru | nima | | SOV |
| vode ri | blomi ru | nima | | SOV |
| vode ri | blomi ru | zeima | nima | SOV |
| vode ri | trose blomi ru | nima | | SOV |
| neimo vode ri | blomi ru | nima | | SOV |
| blomi ru | nima | | | OV |
| blomi ru | noika | nima | | OV |
| neimo blomi ru | nima | | | OV |
| neimo blomi ru | noika | nima | | OV |
| vode ri | vode ru | zeima | yabe | SOV |
| vode ri | blomi ru | zeima | yabe | SOV |
| pleca ru | yabe | | | OV |

| | | | | |
|---|---|---|---|---|
| pleca ru | zeima | yabe | | OV |
| blomi ri | pleca ru | prazi | | SOV |
| blomi ri | pleca ru | noika | prazi | SOV |
| pleca ru | prazi | | | OV |
| pleca ru | noika | prazi | | OV |
| vode ri | klino | | | SV |
| vode ri | noika | klino | | SV |
| nipo ri | blomi ru | nima | | SOV |
| nipo ri | blomi ru | noika | nima | SOV |

Block 2

**Nouns:** **pleca; vode; blomi, nipo**
**Case particles:** **ri (NOM); ru (ACC)**
**Adjectives:** **neimo; troise**
**Verbs:** **nima; prazi; yabe; klino**
**Adverbs:** **noika; zeima**

**SOV = 14 OV = 8 SV = 2**
**symmetric = 8 asymmetric = 16**

| | | | | |
|---|---|---|---|---|
| vode ri | pleca ru | yabe | | SOV |
| vode ri | pleca ru | zeima | yabe | SOV |
| vode ri | trose pleca ru | yabe | | SOV |
| neimo vode ri | pleca ru | yabe | | SOV |
| blomi ri | vode ru | yabe | | SOV |
| blomi ri | vode ru | noika | yabe | SOV |
| blomi ri | neimo vode ru | yabe | | SOV |

| | | | | |
|---|---|---|---|---|
| trose blomi ri | vode ru | yabe | | SOV |
| | | | | |
| nipo ru | yabe | | | OV |
| nipo ru | noika | yabe | | OV |
| trose nipo ru | yabe | | | OV |
| trose nipo ru | noika | yabe | | OV |
| | | | | |
| vode ri | pleca ru | nima | | SOV |
| vode ri | pleca ru | noika | nima | SOV |
| nipo ru | nima | | | OV |
| nipo ru | zeima | nima | | OV |
| | | | | |
| nipo ri | nipo ru | zeima | prazi | SOV |
| blomi ri | nipo ru | zeima | prazi | SOV |
| vode ru | prazi | | | OV |
| vode ru | noika | prazi | | OV |
| | | | | |
| nipo ri | klino | | | SV |
| nipo ri | zeima | klino | | SV |
| pleca ri | nipo ru | yabe | | SOV |
| pleca ri | nipo ru | noika | yabe | SOV |

Block 3

**Nouns:** **pleca; vode; blomi, nipo**
**Case particles:** **ri (NOM); ru (ACC)**
**Adjectives:** **neimo; troise**
**Verbs:** **nima; prazi; yabe; klino**
**Adverbs:** **noika; zeima**

**SOV = 14 OV = 8 SV = 2**
**symmetric = 18 asymmetric = 6**

| | | | | |
|---|---|---|---|---|
| nipo ri | blomi ru | prazi | | SOV |
| nipo ri | blomi ru | noika | prazi | SOV |
| neimo nipo ri | blomi ru | prazi | | SOV |
| nipo ri | neimo blomi ru | prazi | | SOV |
| | | | | |
| vode ri | pleca ru | prazi | | SOV |
| vode ri | pleca ru | noika | prazi | SOV |
| trose vode ri | pleca ru | prazi | | SOV |
| trose vode ri | trose pleca ru | prazi | | SOV |
| | | | | |
| nipo ru | prazi | | | OV |
| nipo ru | noika | prazi | | OV |
| neimo nipo ru | prazi | | | OV |
| neimo nipo ru | noika | prazi | | OV |
| | | | | |
| blomi ri | nipo ru | nima | | SOV |
| blomi ri | nipo ru | zeima | nima | SOV |
| pleca ru | nima | | | OV |
| pleca ru | noika | nima | | OV |
| vode ri | blomi ru | yabe | | SOV |
| vode ri | blomi ru | zeima | yabe | SOV |
| vode ru | yabe | | | OV |
| vode ru | zeima | yabe | | OV |
| | | | | |
| pleca ri | klino | | | SV |
| pleca ri | noika | klino | | SV |
| vode ri | nipo ru | prazi | | SOV |

| | | | | |
|---|---|---|---|---|
| vode ri | nipo ru | noika | prazi | SOV |

Block 4

**Nouns:** **pleca; vode; blomi, nipo**
**Case particles:** **ri (NOM); ru (ACC)**
**Adjectives:** **neimo; troise**
**Verbs:** **nima; prazi; yabe; klino**
**Adverbs:** **noika; zeima**

**SOV = 14** **OV = 8** **SV = 2**
**symmetric = 18** **asymmetric = 6**

| | | | | |
|---|---|---|---|---|
| nipo ri | vode ru | nima | | SOV |
| vode ru | nima | | | OV |
| nipo ri | vode ru | zeima | nima | SOV |
| vode ru | zeima nima | | | OV |
| nipo ri | trose vode ru | nima | | SOV |
| trose vode ru | nima | | | OV |
| nipo ri | trose vode ru | zeima | nima | SOV |
| trose vode ru | zeima | nima | | OV |
| neimo vode ri | klino | | | SV |
| neimo vode ri | noika | klino | | SV |
| nipo ri | neimo blomi ru | nima | | SOV |
| nipo ri | neimo blomi ru | noika | nima | SOV |
| nipo ri | pleca ru | zeima | nima | SOV |
| nipo ri | pleca ru | noika | nima | SOV |
| nipo ri | trose pleca ru | nima | | SOV |

| | | | | |
|---|---|---|---|---|
| neimo nipo ri | pleca ru | nima | | SOV |
| vode ri | neimo blomi ru | yabe | | SOV |
| trose vode ri | blomi ru | yabe | | SOV |
| trose pleca ru | yabe | | | OV |
| trose pleca ru | zeima | yabe | | OV |
| blomi ri | neimo pleca ru | prazi | | SOV |
| trose blomi ri | pleca ru | prazi | | SOV |
| neimo pleca ru | prazi | | | OV |
| neimo pleca ru | noika | prazi | | OV |

Block 5

**Nouns:** **pleca; vode; blomi, nipo**
**Case particles:** **ri (NOM); ru (ACC)**
**Adjectives:** **neimo; troise**
**Verbs:** **nima; prazi; yabe; klino**
**Adverbs:** **noika; zeima**

**SOV = 14 OV = 8 SV = 2**
**symmetric = 8 asymmetric = 16**

| | | | | |
|---|---|---|---|---|
| pleca ri | blomi ru | yabe | | SOV |
| blomi ru | yabe | | | OV |
| pleca ri | blomi ru | zeima | yabe | SOV |
| blomi ru | zeima | yabe | | OV |
| pleca ri | neimo blomi ru | yabe | | SOV |

| | | | | |
|---|---|---|---|---|
| neimo blomi ru | yabe | | | OV |
| pleca ri | neimo blomi ru | zeima | yabe | SOV |
| neimo pleca ru | zeima | yabe | | OV |
| | | | | |
| trose nipo ri | klino | | | SV |
| trose nipo ri | noika | klino | | SV |
| pleca ri | trose nipo ru | yabe | | SOV |
| pleca ri | trose nipo ru | noika | yabe | SOV |
| | | | | |
| nipo ri | blomi ru | yabe | | SOV |
| nipo ri | blomi ru | noika | yabe | SOV |
| neimo nipo ri | blomi ru | yabe | | SOV |
| nipo ri | trose blomi ru | yabe | | SOV |
| | | | | |
| neimo vode ri | pleca ru | nima | | SOV |
| vode ri | trose pleca ru | nima | | SOV |
| neimo nipo ru | nima | | | OV |
| neimo nipo ru | zeima | nima | | OV |
| | | | | |
| neimo blomi ri | nipo ru | prazi | | SOV |
| blomi ri | trose nipo ru | prazi | | SOV |
| trose vode ru | prazi | | | OV |
| trose vode ru | noika | prazi | | OV |

Block 6

**Nouns:** **pleca; vode; blomi, nipo**
**Case particles:** **ri (NOM); ru (ACC)**
**Adjectives:** **neimo; troise**
**Verbs:** **nima; prazi; yabe; klino**

**Adverbs:** **noika; zeima**

**SOV = 14** **OV = 8** **SV = 2**
**symmetric = 18** **asymmetric = 6**

| | | | | |
|---|---|---|---|---|
| pleca ri | blomi ru | prazi | | SOV |
| blomi ru | prazi | | | OV |
| pleca ri | blomi ru | zeima | prazi | SOV |
| blomi ru | zeima | prazi | | OV |
| pleca ri | neimo blomi ru | prazi | | SOV |
| neimo blomi ru | prazi | | | OV |
| pleca ri | neimo blomi ru | zeima | prazi | SOV |
| neimo blomi ru | zeima | prazi | | OV |
| trose pleca ri | klino | | | SV |
| trose pleca ri | noika | klino | | SV |
| vode ri | neimo nipo ru | prazi | | SOV |
| vode ri | neimo nipo ru | noika | prazi | SOV |
| pleca ri | nipo ru | noika | prazi | SOV |
| pleca ri | nipo ru | zeima | prazi | SOV |
| pleca ri | trose nipo ru | zeima | prazi | SOV |
| neimo pleca ri | trose nipo ru | prazi | | SOV |
| trose blomi ri | nipo ru | zeima | nima | SOV |
| trose blomi ri | trose nipo ru | nima | | SOV |
| trose pleca ru | nima | | | OV |

| | | | |
|---|---|---|---|
| trose pleca ru | noika | nima | OV |
| vode ri | trose blomi ru | zeima | yabe SOV |
| neimo vode ri | trose blomi ru | yabe | SOV |
| neimo vode ru | yabe | | OV |
| neimo vode ru | zeima | yabe | OV |

**Appendix C**

**Game practice set**

Block 1

| | |
|---|---|
| **Nouns:** | **pleca; vode; blomi, nipo** |
| **Case particles:** | **ri (NOM); ru (ACC)** |
| **Adjectives:** | **neimo; troise** |
| **Verbs:** | **nima; prazi; yabe; klino** |
| **Adverbs:** | **noika; zeima** |

**SOV = 12 OV = 6 SV = 2**
**symmetric = 12 asymmetric = 8**

| | |
|---|---|
| trose nipo ri klino | SV |
| neimo nipo ri neimo nipo ru zeima yabe | SOV |
| neimo blomi ri trose pleca ru nima | SOV |
| neimo pleca ru zeima nima | OV |
| pleca ri nipo ru nima | SOV |
| trose vode ri pleca ru nima | SOV |
| neimo pleca ru nima | OV |
| trose nipo ri nipo ru noika nima | SOV |
| neimo blomi ri pleca ru yabe | SOV |
| neimo vode ri nipo ru zeima yabe | SOV |
| trose nipo ru yabe | OV |
| blomi ri klino | SV |
| trose blomi ru yabe | OV |
| vode ri blomi ru noika yabe | SOV |
| trose pleca ri blomi ru prazi | SOV |
| vode ri trose blomi ru prazi | SOV |
| trose blomi ri neimo vode ru prazi | SOV |

| | |
|---|---|
| neimo vode ri trose vode ru zeima prazi | SOV |
| vode ru noika prazi | OV |
| vode ru prazi | OV |

<u>Block 2</u>

**Nouns:** **pleca; vode; blomi, nipo**
**Case particles:** **ri (NOM); ru (ACC)**
**Adjectives:** **neimo; troise**
**Verbs:** **nima; prazi; yabe; klino**
**Adverbs:** **noika; zeima**

**SOV = 12 OV = 6 SV = 2**
**symmetric = 12 asymmetric = 8**

| | |
|---|---|
| pleca ri noika klino | SV |
| trose nipo ri vode ru prazi | SOV |
| neimo vode ru noika yabe | OV |
| trose vode ru nima | OV |
| trose blomi ri neimo pleca ru nima | SOV |
| neimo vode ri pleca ru nima | SOV |
| pleca ri trose pleca ru nima | SOV |
| pleca ru zeima nima | OV |
| blomi ri nipo ru noika yabe | SOV |
| trose vode ri trose nipo ru yabe | SOV |
| neimo blomi ri neimo nipo ru yabe | SOV |
| nipo ru noyka yabe | OV |
| trose pleca ri nipo ru nima | SOV |
| vode ri blomi ru noika yabe | SOV |
| neimo pleca ri neimo blomi ru prazi | SOV |

| | |
|---|---|
| nipo ri blomi ru prazi | SOV |
| trose blomi ru prazi | OV |
| neimo vode ru prazi | OV |
| pleca ri vode ru prazi | SOV |
| trose nipo ri noika klino | SV |

<u>Block 3</u>

**Nouns:** **pleca; vode; blomi, nipo**
**Case particles:** **ri (NOM); ru (ACC)**
**Adjectives:** **neimo; troise**
**Verbs:** **nima; prazi; yabe; klino**
**Adverbs:** **noika; zeima**

**SOV = 12** **OV = 6** **SV = 2**
**symmetric = 12** **asymmetric = 8**

| | |
|---|---|
| neimo blomi ri klino | SV |
| trose pleca ri trose pleca ru yabe | SOV |
| nipo ri blomi ru prazi | SOV |
| trose vode ru nima | OV |
| trose vode ru nima | OV |
| nipo ru yabe | OV |
| trose pleca ri vode ru nima | SOV |
| nipo ri neimo vode ru zeima nima | SOV |
| trose vode ri neimo vode ru nima | SOV |
| neimo pleca ru noika yabe | OV |
| neimo nipo ri noika klino | SV |
| neimo nipo ri pleca ru zeima yabe | SOV |
| blomi ri neimo nipo ru yabe | SOV |

| | |
|---|---|
| vode ri nipo ru noika yabe | SOV |
| trose pleca ri nipo ru prazi | SOV |
| trose pleca ru zeima prazi | OV |
| trose blomi ri blomi ru prazi | SOV |
| trose vode ri trose blomi ru prazi | SOV |
| neimo blomi ru prazi | OV |
| neimo pleca ri blomi ru nima | SOV |

<u>Block 4</u>

| | |
|---|---|
| **Nouns:** | **pleca; vode; blomi, nipo** |
| **Case particles:** | **ri (NOM); ru (ACC)** |
| **Adjectives:** | **neimo; troise** |
| **Verbs:** | **nima; prazi; yabe; klino** |
| **Adverbs:** | **noika; zeima** |

**SOV = 12 OV = 6 SV = 2**
**symmetric = 12 asymmetric = 8**

| | |
|---|---|
| trose vode ri zeima klino | SV |
| trose pleca ri pleca ru prazi | SOV |
| trose blomi ru nima | OV |
| neimo nipo ri blomi ru noika nima | SOV |
| nipo ru nima | OV |
| trose vode ri blomi ru nima | SOV |
| trose pleca ri blomi ru nima | SOV |
| vode ru nima | OV |
| neimo blomi ri nipo ru zeima yabe | SOV |
| neimo vode ri vode ru zeima yabe | SOV |
| trose vode ru yabe | OV |

| | |
|---|---|
| neimo vode ru noika yabe | OV |
| neimo blomi ri pleca ru yabe | SOV |
| pleca ru noika yabe | OV |
| blomi ri zeima klino | SV |
| nipo ri trose pleca ru prazi | SOV |
| trose blomi ri neimo nipo ru prazi | SOV |
| neimo vode ri trose nipo ru prazi | SOV |
| pleca ri neimo nipo ru prazi | SOV |
| trose nipo ri trose nipo ru prazi | SOV |

Block 5

| | |
|---|---|
| **Nouns:** | **pleca; vode; blomi, nipo** |
| **Case particles:** | **ri (NOM); ru (ACC)** |
| **Adjectives:** | **neimo; troise** |
| **Verbs:** | **nima; prazi; yabe; klino** |
| **Adverbs:** | **noika; zeima** |

**SOV = 12 OV = 6 SV = 2**
**symmetric = 12 asymmetric = 8**

| | |
|---|---|
| vode ri blomi ru nima | SOV |
| nipo ri nipo ru prazi | SOV |
| trose pleca ri neimo nipo ru noika prazi | SOV |
| trose nipo ru nima | OV |
| trose pleca ri klino | SV |
| pleca ri neimo blomi ru nima | SOV |
| trose nipo ri neimo vode ru nima | SOV |
| blomi ri blomi ru zeima nima | SOV |
| vode ru noika yabe | OV |

| | |
|---|---|
| trose pleca ri trose vode ru zeima yabe | SOV |
| neimo vode ru yabe | OV |
| trose vode ri nipo ru yabe | SOV |
| vode ri trose vode ru yabe | SOV |
| pleca ru zeima yabe | OV |
| neimo nipo ri pleca ru prazi | SOV |
| neimo pleca ru prazi | OV |
| neimo pleca ru prazi | OV |
| trose blomi ri klino | SV |
| neimo nipo ri vode ru nima | SOV |
| blomi ri nipo ru zeima prazi | SOV |

Block 6

| | |
|---|---|
| **Nouns:** | **pleca; vode; blomi, nipo** |
| **Case particles:** | **ri (NOM); ru (ACC)** |
| **Adjectives:** | **neimo; troise** |
| **Verbs:** | **nima; prazi; yabe; klino** |
| **Adverbs:** | **noika; zeima** |

**SOV = 12 OV = 6 SV = 2**
**symmetric = 12 asymmetric = 8**

| | |
|---|---|
| nipo ri klino | SV |
| neimo vode ri pleca ru nima | SOV |
| trose vode ri neimo pleca ru prazi | SOV |
| neimo nipo ru nima | OV |
| neimo blomi ri noika klino | SV |
| nipo ru nima | OV |
| nipo ri neimo nipo ru nima | SOV |

| | |
|---|---|
| blomi ri vode ru nima | SOV |
| nipo ru yabe | OV |
| neimo blomi ru zeima yabe | OV |
| trose nipo ri trose blomi ru yabe | SOV |
| neimo blomi ri blomi ru zeima yabe | SOV |
| trose vode ri neimo vode ru yabe | SOV |
| vode ru noika yabe | OV |
| trose nipo ri vode ru prazi | SOV |
| blomi ri trose vode ru prazi | SOV |
| trose nipo ri vode ru nima | SOV |
| neimo pleca ri trose pleca ru prazi | SOV |
| nipo ri pleca ru noika prazi | SOV |
| pleca ru prazi | OV |

**Appendix D**

**GJT set**

| **Grammatical sentences** | **Ungrammatical sentences** | **WO Type** | **Violation type** |
|---|---|---|---|
| *Practice (does not count towards total score)* | | | |
| neimo pleca ri noika klino | neimo pleca **noika ru** klino | SV | sent WO/Case |
| trose nipo ri zeima klino | trose nipo **ru** zeima klino | SV | Case |
| *Experimental stimuli* | | | |
| trose vode ru noika nima | trose vode ru **nima noika** | OV | sent WO (V-Adv instead of Adv-V) |
| vode ri blomi ru yabe | **blomi ru vode ri** yabe | SOV | sent WO (OSV instead of SOV) |
| neimo blomi ru noika yabe | **blomi ru neimo** noika yabe | OV | NP WO (N-Marker-Adj instead of Adj-N-Marker) |
| nipo ri pleca ru nima | nipo ri **ru pleca** nima | SOV | NP WO (Marker-N instead of N-Marker) |
| pleca ri nipo ru prazi | pleca **ru** nipo ru prazi | SOV | Case (ru instead of ri) |
| vode ri vode ru prazi | vode **ru** vode ru prazi | SOV | Case (ru instead of ri) |
| blomi ri blomi ru prazi | blomi ri **prazi blomi ru** | SOV | sent WO (SVO instead of SOV) |
| nipo ri pleca ru prazi | nipo ri **prazi pleca ru** | SOV | sent WO (SVO instead of SOV) |

| | | | |
|---|---|---|---|
| trose pleca ru zeima nima | trose pleca **ri** zeima nima | OV | Case (ri instead of ru) |
| vode ri nipo ru yabe | vode ri trose nipo // yabe | SOV | Case (ru missing) |
| neimo pleca ru noika nima | neimo pleca **ri** noika nima | OV | Case (ri instead of ru) |
| blomi ri nipo ru yabe | trose blomi // nipo ru yabe | SOV | Case (ri missing) |
| neimo vode ri zeima klino | **vode neimo ri** zeima klino | SV | NP WO (N-Adj-Marker instead of Adj-N-Marker) |
| neimo blomi ri zeima klino | trose blomi ri **klino zeima** | SV | sent WO (V-Adv instead of Adv-V) |

# Appendix E

## (a) Debriefing interview

1) Pensi che alcune delle parole fossero speciali? Perché?

[Do you think any of the words of the new language were special? Why?]

2) Pensi che la nuova lingua che hai ascoltato avesse delle regole speciali? Per esempio?

[Do you think the new language you heard had any special rules? For example?]

3) Immagina che il/la tuo/a migliore amico/a voglia giocare a questo gioco. Adesso che sei un/a esperto/a, cosa potresti dirgli per aiutarlo/a a fare molti punti più velocemente?

[Suppose your best friend wanted to play this game. Now that you are an expert, what would you tell him/her to help them make a lot of points quickly?]

4) C'è nient'altro che vorresti dirmi sulla nuova lingua o il gioco?

[Is there anything else you would like to tell me about the language or the gaming experience?]

**(b) Examples of language awareness rating (children)**

| Rating (WO) | Example | Question | Participant | Grade |
|---|---|---|---|---|
| 1 | 'No. I haven't noticed anything special'<br>'You have to pay attention' | 1/2<br><br>3 | C17 | 4 |
| 2 | 'Sometimes I heard a word you didn't teach me, *ri* or *ru*' | 1 | C32 | 3 |
| 3 | 'It mixed up the words. The last word it said was the move'<br>'You have to pay attention' | 2<br><br>3 | C19 | 4 |
| 4 | 'The words are not random, there is an order'<br>'You need to pay attention and repeat in your head' | 2<br><br>3 | C1 | 5 |
| 5 | 'Yes, *klino* first, then *zeima*. There is a certain order' | 2 | C14 | 4 |
| 6 | 'For example if it said *Trose blomi ri nipo ru zeima prazi*, the first was the shape, then the token..'<br>'At the beginning it was confused, then it became less confused' | 2<br><br>4 | C26 | 3 |

*Figure E.1*. Examples of Rating of Awareness of Word Order for the Child Group (Translated from Italian).

**(c) Examples of language awareness rating (adults)**

| Rating (WO) | Rating (SEM) | Example | Question | Participant |
|---|---|---|---|---|
| 1 | 0 | I haven't noticed anything special | 1/2 | A1 |
| 2 | 3 | 'At some point I thought that *ri* and *ru* meant active and passive, especially with *prazi*' | 2 | A34 |
| 3 | 1 | 'There is an order between *ri* and *ru*. *Ri* comes first' | 2 | A19 |
| 4 | 0 | 'There is a linguistic structure. I was expecting a structure with subject, verb, preposition, complement' | 2 | A26 |
| | | 'The important thing is memorize, but not get too fixed on an idea, also have a go' | 3 | |
| 5 | 2 | 'There is a word order. Maybe *ri* and *ru* mean first and second' | 1 | A35 |
| | | '*Ri* comes before *ru*. The move comes at the end' | 2 | |
| | | 'Most of all you have to pay attention to *ri* and *ru* with *prazi* and *nima*' | 3 | |
| 6 | - | n/a | n/a | n/a |

*Figure E.2*. Examples of Rating of Awareness of Word Order and *Ri*/*Ru*/NPs Thematic Interpretation (SEM) for the Adult Group (Translated from Italian).

**Appendix F**

## Re-leveling: Model A (children)

### WO = SOVs

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | -1.55336 | 0.15038 | -10.330 | < 2e-16 | *** |
| DECL | 0.09485 | 0.09215 | 1.029 | 0.30331 | |
| WO(SV) | 0.32323 | 0.30682 | 1.054 | 0.29211 | |
| WO(SOVa) | 1.01231 | 0.24429 | 4.144 | 3.41e-05 | *** |
| WO(OVs) | -0.14958 | 0.25316 | -0.591 | 0.55462 | |
| WO(OVa) | 0.86082 | 0.27557 | 3.124 | 0.00179 | ** |
| PROC | -0.01160 | 0.06199 | -0.187 | 0.85162 | |
| cBLOCK | 0.23575 | 0.05908 | 3.990 | 6.61e-05 | *** |
| VocLearn | 0.39040 | 0.09902 | 3.943 | 8.06e-05 | *** |
| cYEAR | 0.20803 | 0.11577 | 1.797 | 0.07235 | . |
| DECL:WO(SV) | 0.18920 | 0.14526 | 1.302 | 0.19276 | |
| DECL:WO(SOVa) | 0.01335 | 0.10769 | 0.124 | 0.90133 | |
| DECL:WO(OVs) | -0.02530 | 0.12401 | -0.204 | 0.83836 | |
| DECL:WO(OVa) | 0.04552 | 0.12359 | 0.368 | 0.71265 | |
| PROC:cBLOCK | 0.04248 | 0.02147 | 1.979 | 0.04783 | * |
| WO(SV):VocLearn | 0.12633 | 0.16090 | 0.785 | 0.43235 | |
| WO(SOVa):VocLearn | -0.14384 | 0.11615 | -1.238 | 0.21556 | |
| WO(OVs):VocLearn | -0.30481 | 0.13017 | -2.342 | 0.01920 | * |
| WO(OVa):VocLearn | 0.20549 | 0.14128 | 1.454 | 0.14583 | |
| cBLOCK:VocLearn | 0.05350 | 0.02695 | 1.985 | 0.04714 | * |

### WO = SOVa

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | -0.54105 | 0.20875 | -2.592 | 0.00955 | ** |
| DECL | 0.10821 | 0.10519 | 1.029 | 0.30362 | |
| WO(SV) | -0.68908 | 0.33642 | -2.048 | 0.04053 | * |
| WO(SOVs) | -1.01231 | 0.24428 | -4.144 | 3.41e-05 | *** |
| WO(OVs) | -1.16190 | 0.29308 | -3.964 | 7.36e-05 | *** |

| | | | | | |
|---|---|---|---|---|---|
| WO(OVa) | -0.15150 | 0.31218 | -0.485 | 0.62746 | |
| PROC | -0.01160 | 0.06199 | -0.187 | 0.85162 | |
| cBLOCK | 0.23575 | 0.05908 | 3.990 | 6.61e-05 | *** |
| VocLearn | 0.24655 | 0.11089 | 2.223 | 0.02619 | * |
| cYEAR | 0.20803 | 0.11577 | 1.797 | 0.07235 | . |
| DECL:WO(SV) | 0.17584 | 0.15531 | 1.132 | 0.25756 | |
| DECL:WO(SOVs) | -0.01335 | 0.10769 | -0.124 | 0.90131 | |
| DECL:WO(OVs) | -0.03865 | 0.13548 | -0.285 | 0.77543 | |
| DECL:WO(OVa) | 0.03216 | 0.13515 | 0.238 | 0.81189 | |
| PROC:cBLOCK | 0.04248 | 0.02147 | 1.979 | 0.04783 | * |
| WO(SV):VocLearn | 0.27018 | 0.16950 | 1.594 | 0.11095 | |
| WO(SOVs):VocLearn | 0.14384 | 0.11615 | 1.238 | 0.21555 | |
| WO(OVs):VocLearn | -0.16096 | 0.13948 | -1.154 | 0.24849 | |
| WO(OVa):VocLearn | 0.34933 | 0.15129 | 2.309 | 0.02094 | * |
| cBLOCK:VocLearn | 0.05350 | 0.02695 | 1.985 | 0.04714 | * |

**WO = OVs**

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | -1.70295 | 0.22210 | -7.668 | 1.75e-14 | *** |
| DECL | 0.06956 | 0.12183 | 0.571 | 0.56804 | |
| WO(SV) | 0.47281 | 0.34803 | 1.359 | 0.17430 | |
| WO(SOVs) | 0.14958 | 0.25315 | 0.591 | 0.55459 | |
| WO(SOVa) | 1.16190 | 0.29308 | 3.964 | 7.36e-05 | *** |
| WO(OVa) | 1.01039 | 0.31946 | 3.163 | 0.00156 | ** |
| PROC | -0.01160 | 0.06199 | -0.187 | 0.85162 | |
| cBLOCK | 0.23575 | 0.05908 | 3.990 | 6.61e-05 | *** |
| VocLearn | 0.08559 | 0.12565 | 0.681 | 0.49578 | |
| cYEAR | 0.20803 | 0.11577 | 1.797 | 0.07235 | . |
| DECL:WO(SV) | 0.21449 | 0.16707 | 1.284 | 0.19919 | |
| DECL:WO(SOVs) | 0.02529 | 0.12401 | 0.204 | 0.83838 | |
| DECL:WO(SOVa) | 0.03865 | 0.13548 | 0.285 | 0.77545 | |
| DECL:WO(OVa) | 0.07081 | 0.14859 | 0.477 | 0.63367 | |
| PROC:cBLOCK | 0.04248 | 0.02147 | 1.979 | 0.04783 | * |
| WO(SV):VocLearn | 0.43114 | 0.17943 | 2.403 | 0.01627 | * |

```
WO(SOVs):VocLearn  0.30481    0.13016   2.342  0.01919 *
WO(SOVa):VocLearn  0.16097    0.13948   1.154  0.24848
WO(OVa):VocLearn   0.51030    0.16239   3.143  0.00168 **
cBLOCK:VocLearn    0.05350    0.02695   1.985  0.04714 *
```

**WO = OVa**

```
Fixed effects:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       -0.69255    0.24706  -2.803  0.00506 **
DECL               0.14037    0.12289   1.142  0.25334
WO(SV)            -0.53758    0.36277  -1.482  0.13837
WO(SOVs)          -0.86081    0.27559  -3.124  0.00179 **
WO(SOVa)           0.15150    0.31220   0.485  0.62749
WO(OVs)           -1.01040    0.31950  -3.162  0.00156 **
PROC              -0.01160    0.06199  -0.187  0.85162
cBLOCK             0.23575    0.05908   3.990 6.61e-05 ***
VocLearn           0.59589    0.13869   4.297 1.73e-05 ***
cYEAR              0.20803    0.11577   1.797  0.07235 .
DECL:WO(SV)        0.14368    0.16677   0.862  0.38894
DECL:WO(SOVs)     -0.04552    0.12359  -0.368  0.71264
DECL:WO(SOVa)     -0.03217    0.13515  -0.238  0.81188
DECL:WO(OVs)      -0.07082    0.14860  -0.477  0.63367
PROC:cBLOCK        0.04248    0.02147   1.979  0.04783 *
WO(SV):VocLearn   -0.07916    0.18799  -0.421  0.67370
WOs(SOVs):VocLearn-0.20549    0.14129  -1.454  0.14583
WO(SOVa):VocLearn -0.34933    0.15129  -2.309  0.02094 *
WO(OVs):VocLearn  -0.51030    0.16239  -3.142  0.00168 **
cBLOCK:VocLearn    0.05350    0.02695   1.985  0.04714 *

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

**Appendix G**

# Re-leveling: GJT Model (children)

## WO = OV; Viola = grammatical (no violation)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 0.912023 | 0.336610 | 2.709 | 0.006740 | ** |
| DECLtot | -0.019625 | 0.072681 | -0.270 | 0.787148 | |
| PROC | 0.295751 | 0.115158 | 2.568 | 0.010222 | * |
| WO(SOV) | -0.121730 | 0.410712 | -0.296 | 0.766934 | |
| WO(SV) | 0.008557 | 0.577880 | 0.015 | 0.988186 | |
| Viola(CA) | -2.217158 | 0.590091 | -3.757 | 0.000172 | *** |
| Viola(WO) | -0.461141 | 0.570711 | -0.808 | 0.419084 | |
| S1S2 | -0.141626 | 0.075322 | -1.880 | 0.060073 | . |
| PROC:WO(SOV) | -0.209182 | 0.139417 | -1.500 | 0.133509 | |
| PROC:WO(SV) | -0.251297 | 0.190448 | -1.320 | 0.186999 | |
| WO(SOV):Viola(CA) | 1.909564 | 0.714673 | 2.672 | 0.007541 | ** |
| WO(SOV):Viola(WO) | 0.913284 | 0.709862 | 1.287 | 0.198246 | |
| WO(SV):Viola(WO) | -0.288888 | 0.869240 | -0.332 | 0.739628 | |

## WO = OV; Viola = WO (word order)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 0.45088 | 0.46127 | 0.978 | 0.32833 | |
| DECL | -0.01963 | 0.07268 | -0.270 | 0.78715 | |
| PROC | 0.29575 | 0.11516 | 2.568 | 0.01022 | * |
| WO(SOV) | 0.79155 | 0.57920 | 1.367 | 0.17174 | |
| WO(SV) | -0.28033 | 0.64949 | -0.432 | 0.66602 | |
| Viola(CA) | -1.75602 | 0.66869 | -2.626 | 0.00864 | ** |
| Viola(gramm) | 0.46114 | 0.57070 | 0.808 | 0.41908 | |
| S1S2 | -0.14163 | 0.07532 | -1.880 | 0.06007 | . |
| PROC:WO(SOV) | -0.20918 | 0.13942 | -1.500 | 0.13351 | |
| PROC:WO(SV) | -0.25130 | 0.19045 | -1.319 | 0.18700 | |
| WO(SOV):Viola(CA) | 0.99628 | 0.82246 | 1.211 | 0.22576 | |
| WO(SOV):Viola(gramm) | -0.91328 | 0.70986 | -1.287 | 0.19824 | |

```
WO(SV):Viola(gramm)  0.28889    0.86922   0.332  0.73962
```

**WO = SOV; Viola = CA (case)**

```
Fixed effects:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)            0.48270    0.32785   1.472  0.14094
DECL                  -0.01962    0.07268  -0.270  0.78715
PROC                   0.08657    0.08078   1.072  0.28385
WO(OV)                -1.78783    0.58435  -3.059  0.00222 **
WO(SV)                -1.07188    0.57620  -1.860  0.06285 .
Viola(gramm)           0.30759    0.40377   0.762  0.44617
Viola(WO)              0.75974    0.47960   1.584  0.11317
S1S2                  -0.14163    0.07532  -1.880  0.06007 .
PROC:WO(OV)            0.20918    0.13942   1.500  0.13351
PROC:WO(SV)           -0.04212    0.17197  -0.245  0.80653
WO(OV):Viola(gramm)    1.90956    0.71468   2.672  0.00754 **
WO(SV):Viola(gramm)    1.20216    0.77963   1.542  0.12308
WO(OV):Viola(WO)       0.99628    0.82245   1.211  0.22576
```

**WO = SOV; Viola = grammatical (no violation)**

```
Fixed effects:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)            0.79029    0.23621   3.346 0.000821 ***
DECL                  -0.01962    0.07268  -0.270 0.787149
PROC                   0.08657    0.08078   1.072 0.283847
WO(OV)                 0.12173    0.41072   0.296 0.766928
WO(SV)                 0.13029    0.52575   0.248 0.804278
Viola(CA)             -0.30759    0.40377  -0.762 0.446186
Viola(WO)              0.45215    0.42147   1.073 0.283367
S1S2                  -0.14163    0.07532  -1.880 0.060073 .
PROC:WO(OV)            0.20918    0.13942   1.500 0.133509
PROC:WO(SV)           -0.04212    0.17197  -0.245 0.806534
WO(OV):Viola(CA)      -1.90957    0.71469  -2.672 0.007542 **
WO(OV):Viola(WO)      -0.91330    0.70986  -1.287 0.198237
```

| | | | | | |
|---|---|---|---|---|---|
| WO(SV):Viola(WO) | -1.20218 | 0.77964 | -1.542 | 0.123081 | |

## WO = SOV; Viola = WO (word order)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 1.24244 | 0.35060 | 3.544 | 0.000395 | *** |
| DECL | -0.01963 | 0.07268 | -0.270 | 0.787148 | |
| PROC | 0.08657 | 0.08078 | 1.072 | 0.283847 | |
| WO(OV) | -0.79156 | 0.57922 | -1.367 | 0.171751 | |
| WO(SV) | -1.07189 | 0.57621 | -1.860 | 0.062849 | . |
| Viola(CA) | -0.75974 | 0.47961 | -1.584 | 0.113176 | |
| Viola(gramm) | -0.45215 | 0.42148 | -1.073 | 0.283376 | |
| S1S2 | -0.14163 | 0.07532 | -1.880 | 0.060073 | . |
| PROC:WO(OV) | 0.20918 | 0.13942 | 1.500 | 0.133509 | |
| PROC:WO(SV) | -0.04212 | 0.17197 | -0.245 | 0.806534 | |
| WO(OV):Viola(CA) | -0.99628 | 0.82247 | -1.211 | 0.225773 | |
| WO(OV):Viola(gramm) | 0.91329 | 0.70987 | 1.287 | 0.198249 | |
| WO(SV):Viola(gramm) | 1.20218 | 0.77964 | 1.542 | 0.123080 | |

## WO = SV; Viola = CA (case)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | -0.58917 | 0.66280 | -0.889 | 0.3741 | |
| DECL | -0.01962 | 0.07268 | -0.270 | 0.7871 | |
| PROC | 0.04445 | 0.15269 | 0.291 | 0.7709 | |
| WO(OV) | -0.71596 | 0.82008 | -0.873 | 0.3826 | |
| WO(SOV) | 1.07187 | 0.57621 | 1.860 | 0.0629 | . |
| Viola(gramm) | 1.50975 | 0.81261 | 1.858 | 0.0632 | . |
| Viola(WO) | 0.75973 | 0.47960 | 1.584 | 0.1132 | |
| S1S2 | -0.14163 | 0.07532 | -1.880 | 0.0601 | . |
| PROC:WO(OV) | 0.25130 | 0.19045 | 1.319 | 0.1870 | |
| PROC:WO(SOV) | 0.04211 | 0.17197 | 0.245 | 0.8065 | |
| WO(OV):Viola(gramm) | 0.70741 | 1.00336 | 0.705 | 0.4808 | |
| WO(SOV):Viola(gramm) | -1.20216 | 0.77963 | -1.542 | 0.1231 | |

| | | | | | |
|---|---|---|---|---|---|
| WO(OV):Viola(WO) | 0.99628 | 0.82245 | 1.211 | 0.2258 | |

## WO = SV; Viola = grammatical (no violation)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 0.920573 | 0.469916 | 1.959 | 0.05011 | . |
| DECL | -0.019625 | 0.072681 | -0.270 | 0.78715 | |
| PROC | 0.044454 | 0.152686 | 0.291 | 0.77094 | |
| WO(OV) | -0.008548 | 0.577885 | -0.015 | 0.98820 | |
| WO(SOV) | -0.130280 | 0.525756 | -0.248 | 0.80429 | |
| Viola(CA) | -0.307594 | 0.403771 | -0.762 | 0.44618 | |
| Viola(WO) | -0.750025 | 0.655748 | -1.144 | 0.25272 | |
| S1S2 | -0.141626 | 0.075322 | -1.880 | 0.06007 | . |
| PROC:WO(OV) | 0.251297 | 0.190448 | 1.319 | 0.18700 | |
| PROC:WO(SOV) | 0.042114 | 0.171970 | 0.245 | 0.80654 | |
| WO(OV):Viola(CA) | -1.909565 | 0.714682 | -2.672 | 0.00754 | ** |
| WO(OV):Viola(WO) | 0.288882 | 0.869243 | 0.332 | 0.73963 | |
| WO(SOV):Viola(WO) | 1.202168 | 0.779645 | 1.542 | 0.12309 | |

## WO = SV; Viola = WO (word order)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 0.17055 | 0.45743 | 0.373 | 0.7093 | |
| DECL | -0.01963 | 0.07268 | -0.270 | 0.7871 | |
| PROCTOT5 | 0.04445 | 0.15269 | 0.291 | 0.7709 | |
| WO(OV) | 0.28033 | 0.64949 | 0.432 | 0.6660 | |
| WO(SOV) | 1.07188 | 0.57619 | 1.860 | 0.0628 | . |
| Viola(CA) | -0.75974 | 0.47960 | -1.584 | 0.1132 | |
| Viola(gramm) | 0.75002 | 0.65572 | 1.144 | 0.2527 | |
| S1S2 | -0.14163 | 0.07532 | -1.880 | 0.0601 | . |
| PROC:WO(OV) | 0.25130 | 0.19045 | 1.319 | 0.1870 | |
| PROC:WO(OV) | 0.04212 | 0.17197 | 0.245 | 0.8065 | |
| WO(OV):Viola(CA) | -0.99628 | 0.82244 | -1.211 | 0.2258 | |
| WO(OV):Viola(gramm) | -0.28888 | 0.86922 | -0.332 | 0.7396 | |

```
WO(SOV):Viola(gramm)   -1.20217    0.77961  -1.542   0.1231

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

**Appendix H**

# Re-leveling: Confidence Level Model (children)

## Viola = WO (word order)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 0.80854 | 0.31143 | 2.596 | 0.009426 | ** |
| CONF | -0.10367 | 0.18081 | -0.573 | 0.566419 | |
| Viola(CA) | -0.88554 | 0.46148 | -1.919 | 0.054995 | . |
| Viola(gramm) | 0.09249 | 0.38184 | 0.242 | 0.808615 | |
| DECL | 0.03477 | 0.10181 | 0.342 | 0.732686 | |
| PROC | 0.23149 | 0.13428 | 1.724 | 0.084716 | . |
| CONF:Viola(CA) | -0.36768 | 0.26449 | -1.390 | 0.164488 | |
| CONF:Viola(gramm) | 0.75131 | 0.22540 | 3.333 | 0.000858 | *** |
| CONF:PROC | -0.21225 | 0.07226 | -2.937 | 0.003310 | ** |
| Viola(CA):PROC | -0.36985 | 0.18052 | -2.049 | 0.040484 | * |
| Viola(gramm):PROC | -0.08816 | 0.15902 | -0.554 | 0.579286 | |

## Viola = grammatical (no violation)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 0.90103 | 0.24060 | 3.745 | 0.000180 | *** |
| CONF | 0.64764 | 0.13594 | 4.764 | 1.90e-06 | *** |
| Viola(WO) | -0.09248 | 0.38185 | -0.242 | 0.808624 | |
| Viola(CA) | -0.97802 | 0.41775 | -2.341 | 0.019224 | * |
| DECL | 0.03477 | 0.10181 | 0.342 | 0.732688 | |
| PROC | 0.14333 | 0.11513 | 1.245 | 0.213132 | |
| CONF:Viola(WO) | -0.75131 | 0.22540 | -3.333 | 0.000859 | *** |
| CONF:Viola(CA) | -1.11898 | 0.23876 | -4.687 | 2.78e-06 | *** |
| CONF:PROC | -0.21225 | 0.07226 | -2.937 | 0.003310 | ** |
| Viola(WO):PROC | 0.08816 | 0.15902 | 0.554 | 0.579286 | |
| Viola(CA):PROC | -0.28168 | 0.16914 | -1.665 | 0.095841 | . |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Appendix I**

# Re-leveling: Model A (adults)

## WO = SOVs

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | -0.29957 | 0.20500 | -1.461 | 0.143928 | |
| DECL | 1.07009 | 0.22860 | 4.681 | 2.85e-06 | *** |
| PROC | 0.38821 | 0.22761 | 1.706 | 0.088077 | . |
| SESSION | 1.05002 | 0.14054 | 7.471 | 7.94e-14 | *** |
| VocLearn | 0.39557 | 0.08622 | 4.588 | 4.47e-06 | *** |
| Exec | 0.05039 | 0.15632 | 0.322 | 0.747168 | |
| S1S2 | 0.20871 | 0.09594 | 2.175 | 0.029593 | * |
| WO(SV) | 1.45195 | 0.35902 | 4.044 | 5.25e-05 | *** |
| WO(SOVa) | 1.17644 | 0.28905 | 4.070 | 4.70e-05 | *** |
| WO(OVs) | -0.25701 | 0.28752 | -0.894 | 0.371376 | |
| WO(OVa) | 1.68040 | 0.33452 | 5.023 | 5.08e-07 | *** |
| DECL:PROC | 0.95206 | 0.34833 | 2.733 | 0.006272 | ** |
| DECL:SESSION | 0.21941 | 0.10090 | 2.174 | 0.029668 | * |
| PROC:SESSION | -0.16670 | 0.11208 | -1.487 | 0.136915 | |
| SESSION:VocLearn | 0.14985 | 0.06535 | 2.293 | 0.021844 | * |
| SESSION:zExec | 0.02694 | 0.06559 | 0.411 | 0.681321 | |
| Exec:S1S2 | 0.09223 | 0.11593 | 0.796 | 0.426299 | |
| SESSION:S1S2 | 0.15904 | 0.06770 | 2.349 | 0.018823 | * |
| PROC:WO(SV) | -0.06121 | 0.22374 | -0.274 | 0.784402 | |
| PROC:WO(SOVa) | -0.56694 | 0.18030 | -3.144 | 0.001664 | ** |
| PROC:WO(OVs) | -0.13535 | 0.17815 | -0.760 | 0.447412 | |
| PROC:WO(OVa) | -0.46838 | 0.21987 | -2.130 | 0.033154 | * |
| PROC:Exec | -0.71955 | 0.28519 | -2.523 | 0.011635 | * |
| DECL:PROC:SESSION | 0.52746 | 0.14509 | 3.635 | 0.000278 | *** |
| SESSION:Exec:S1S2 | 0.18744 | 0.07990 | 2.346 | 0.018986 | * |
| PROC:SESSION:Exec | -0.18111 | 0.15047 | -1.204 | 0.228749 | |

**WO = SOVa**

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 0.87686 | 0.27617 | 3.175 | 0.001498 | ** |
| DECL | 1.07009 | 0.22858 | 4.681 | 2.85e-06 | *** |
| PROC | -0.17873 | 0.25348 | -0.705 | 0.480740 | |
| SESSION | 1.05002 | 0.14053 | 7.472 | 7.92e-14 | *** |
| VocLearn | 0.39557 | 0.08621 | 4.588 | 4.47e-06 | *** |
| Exec | 0.05039 | 0.15631 | 0.322 | 0.747153 | |
| S1S2 | 0.20871 | 0.09593 | 2.176 | 0.029584 | * |
| WO(SOVs) | -1.17644 | 0.28899 | -4.071 | 4.68e-05 | *** |
| WO(SV) | 0.27551 | 0.40274 | 0.684 | 0.493922 | |
| WO(OVs) | -1.43345 | 0.34212 | -4.190 | 2.79e-05 | *** |
| WO(OVa) | 0.50395 | 0.38188 | 1.320 | 0.186952 | |
| DECL:PROC | 0.95206 | 0.34826 | 2.734 | 0.006261 | ** |
| DECL:SESSION | 0.21941 | 0.10090 | 2.175 | 0.029667 | * |
| PROC:SESSION | -0.16670 | 0.11207 | -1.487 | 0.136904 | |
| SESSION:VocLearn | 0.14985 | 0.06535 | 2.293 | 0.021843 | * |
| SESSION:Exec | 0.02694 | 0.06559 | 0.411 | 0.681315 | |
| Exec:S1S2 | 0.09223 | 0.11593 | 0.796 | 0.426282 | |
| SESSION:S1S2 | 0.15904 | 0.06770 | 2.349 | 0.018821 | * |
| PROC:WO(SOVs) | 0.56694 | 0.18028 | 3.145 | 0.001662 | ** |
| PROC:WO(SV) | 0.50573 | 0.24699 | 2.048 | 0.040604 | * |
| PROC:WO(OVs) | 0.43160 | 0.21127 | 2.043 | 0.041063 | * |
| PROC:WO(OVa) | 0.09856 | 0.24345 | 0.405 | 0.685586 | |
| PROC:Exec | -0.71955 | 0.28517 | -2.523 | 0.011627 | * |
| DECL:PROC:SESSION | 0.52746 | 0.14509 | 3.635 | 0.000278 | *** |
| SESSION:Exec:S1S2 | 0.18744 | 0.07990 | 2.346 | 0.018984 | * |
| PROC:SESSION:Exec | -0.18111 | 0.15047 | -1.204 | 0.228738 | |

**WO = OVs**

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | -0.55659 | 0.27586 | -2.018 | 0.043628 | * |
| DECL | 1.07009 | 0.22860 | 4.681 | 2.85e-06 | *** |
| PROC | 0.25286 | 0.25371 | 0.997 | 0.318944 | |
| SESSION | 1.05002 | 0.14054 | 7.471 | 7.93e-14 | *** |
| VocLearn | 0.39556 | 0.08622 | 4.588 | 4.47e-06 | *** |
| Exec | 0.05040 | 0.15632 | 0.322 | 0.747140 | |
| S1S2 | 0.20871 | 0.09594 | 2.176 | 0.029591 | * |
| WO(SOVa) | 1.43345 | 0.34217 | 4.189 | 2.80e-05 | *** |
| WO(SOVs) | 0.25701 | 0.28750 | 0.894 | 0.371349 | |
| WO(SV) | 1.70896 | 0.40351 | 4.235 | 2.28e-05 | *** |
| WO(OVa) | 1.93741 | 0.38268 | 5.063 | 4.13e-07 | *** |
| DECLtot:PROC | 0.95205 | 0.34830 | 2.733 | 0.006268 | ** |
| DECLtot:SESSION | 0.21941 | 0.10090 | 2.174 | 0.029669 | * |
| PROC:SESSION | -0.16670 | 0.11208 | -1.487 | 0.136912 | |
| SESSION:VocLearn | 0.14985 | 0.06535 | 2.293 | 0.021844 | * |
| SESSION:Exec | 0.02694 | 0.06559 | 0.411 | 0.681312 | |
| Exec:S1S2 | 0.09223 | 0.11593 | 0.796 | 0.426290 | |
| SESSION:S1S2 | 0.15904 | 0.06770 | 2.349 | 0.018823 | * |
| PROC:WO(SOVa) | -0.43159 | 0.21128 | -2.043 | 0.041081 | * |
| PROC:WO(SOVs) | 0.13535 | 0.17815 | 0.760 | 0.447401 | |
| PROC:WO(SV) | 0.07414 | 0.24970 | 0.297 | 0.766540 | |
| PROC:WO(OVa) | -0.33303 | 0.24655 | -1.351 | 0.176769 | |
| PROC:Exec | -0.71954 | 0.28519 | -2.523 | 0.011634 | * |
| DECL:PROC:SESSION | 0.52746 | 0.14509 | 3.635 | 0.000278 | *** |
| SESSION:Exec:S1S2 | 0.18744 | 0.07990 | 2.346 | 0.018985 | * |
| PROC:SESSION:Exec | -0.18110 | 0.15047 | -1.204 | 0.228747 | |

**WO = OVa**

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 1.38083 | 0.32403 | 4.261 | 2.03e-05 | *** |
| DECL | 1.07009 | 0.22859 | 4.681 | 2.85e-06 | *** |
| PROC | -0.08018 | 0.28318 | -0.283 | 0.777076 | |
| SESSION | 1.05002 | 0.14054 | 7.472 | 7.93e-14 | *** |
| VocLearn | 0.39557 | 0.08621 | 4.588 | 4.47e-06 | *** |
| Exec | 0.05039 | 0.15632 | 0.322 | 0.747172 | |
| S1S2 | 0.20871 | 0.09593 | 2.176 | 0.029589 | * |
| WO(OVs) | -1.93742 | 0.38265 | -5.063 | 4.12e-07 | *** |
| WO(SOVa) | -0.50397 | 0.38186 | -1.320 | 0.186909 | |
| WO(SOVs) | -1.68041 | 0.33444 | -5.025 | 5.05e-07 | *** |
| WO(SV) | -0.22847 | 0.43675 | -0.523 | 0.600893 | |
| DECL:PROC | 0.95205 | 0.34829 | 2.733 | 0.006267 | ** |
| DECL:SESSION | 0.21941 | 0.10090 | 2.175 | 0.029667 | * |
| PROC:SESSION | -0.16670 | 0.11208 | -1.487 | 0.136910 | |
| SESSION:VocLearn | 0.14985 | 0.06535 | 2.293 | 0.021844 | * |
| SESSION:Exec | 0.02694 | 0.06559 | 0.411 | 0.681317 | |
| Exec:S1S2 | 0.09223 | 0.11593 | 0.796 | 0.426290 | |
| SESSION:S1S2 | 0.15904 | 0.06770 | 2.349 | 0.018822 | * |
| PROC:WO(OVs) | 0.33304 | 0.24654 | 1.351 | 0.176752 | |
| PROC:WO(SOVa) | -0.09856 | 0.24346 | -0.405 | 0.685597 | |
| PROC:WO(SOVs) | 0.46839 | 0.21986 | 2.130 | 0.033142 | * |
| PROC:WO(SV) | 0.40717 | 0.27782 | 1.466 | 0.142753 | |
| PROC:Exec | -0.71955 | 0.28518 | -2.523 | 0.011632 | * |
| DECL:PROC:SESSION | 0.52746 | 0.14509 | 3.635 | 0.000278 | *** |
| SESSION:Exec:S1S2 | 0.18744 | 0.07990 | 2.346 | 0.018984 | * |
| PROC:SESSION:Exec | -0.18111 | 0.15047 | -1.204 | 0.228737 | |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Appendix J**

# Re-leveling: GJT Model (adults)

## WO = SOV; Viola = CA (case)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | -0.170321 | 0.368120 | -0.463 | 0.643595 | |
| DECL | 0.348067 | 0.157570 | 2.209 | 0.027177 | * |
| PROC | -0.667800 | 0.282576 | -2.363 | 0.018115 | * |
| Viola(gramm) | 2.183300 | 0.472434 | 4.621 | 3.81e-06 | *** |
| Viola(WO) | 2.050733 | 0.539794 | 3.799 | 0.000145 | *** |
| WO(OV) | -0.866392 | 0.368132 | -2.353 | 0.018599 | * |
| WO(SV) | -0.815932 | 0.494086 | -1.651 | 0.098658 | . |
| PhonLoop | -0.188113 | 0.197111 | -0.954 | 0.339906 | |
| PROC:Viola(gramm) | 0.663809 | 0.450403 | 1.474 | 0.140532 | |
| PROC:Viola(WO) | 0.166083 | 0.503106 | 0.330 | 0.741313 | |
| PROC:PhonLoop | -0.340177 | 0.173223 | -1.964 | 0.049553 | * |
| Viola(gramm):PhonLoop | 0.488016 | 0.285817 | 1.707 | 0.087740 | . |
| Viola(WO):PhonLoop | 0.003102 | 0.323100 | 0.010 | 0.992339 | |

## WO = SV; Viola = CA (case)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | -0.986251 | 0.589410 | -1.673 | 0.094271 | . |
| DECL | 0.348068 | 0.157570 | 2.209 | 0.027176 | * |
| PROC | -0.667800 | 0.282570 | -2.363 | 0.018113 | * |
| Viola(gramm) | 2.183300 | 0.472418 | 4.622 | 3.81e-06 | *** |
| Viola(WO) | 2.050733 | 0.539776 | 3.799 | 0.000145 | *** |
| WO(SOV) | 0.815929 | 0.494074 | 1.651 | 0.098651 | . |
| WO(OV) | -0.050463 | 0.529242 | -0.095 | 0.924038 | |
| PhonLoop | -0.188113 | 0.197108 | -0.954 | 0.339899 | |
| PROC:Viola(gramm) | 0.663808 | 0.450393 | 1.474 | 0.140524 | |
| PROC:Viola(WO) | 0.166084 | 0.503094 | 0.330 | 0.741306 | |
| PROC:PhonLoop | -0.340177 | 0.173223 | -1.964 | 0.049552 | * |
| Viola(gramm):PhonLoop | 0.488016 | 0.285811 | 1.707 | 0.087734 | . |

| | | | | | |
|---|---|---|---|---|---|
| Viola(WO):PhonLoop | 0.003102 | 0.323093 | 0.010 | 0.992340 | |

**WO = OV; Viola = WO (word order)**

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 1.014016 | 0.478752 | 2.118 | 0.034171 | * |
| DECL | 0.348067 | 0.157569 | 2.209 | 0.027176 | * |
| PROC | -0.501718 | 0.447341 | -1.122 | 0.262052 | |
| Viola(CA) | -2.050730 | 0.539771 | -3.799 | 0.000145 | *** |
| Viola(gramm) | 0.132564 | 0.476785 | 0.278 | 0.780984 | |
| WO(SV) | 0.050462 | 0.529235 | 0.095 | 0.924037 | |
| WO(SOV) | 0.866390 | 0.368123 | 2.354 | 0.018596 | * |
| PhonLoop | -0.185009 | 0.302989 | -0.611 | 0.541455 | |
| PROC:Viola(CA) | -0.166084 | 0.503090 | -0.330 | 0.741303 | |
| PROC:Viola(gramm) | 0.497723 | 0.502417 | 0.991 | 0.321853 | |
| PROC:PhonLoop | -0.340177 | 0.173222 | -1.964 | 0.049551 | * |
| Viola(CA):PhonLoop | -0.003103 | 0.323094 | -0.010 | 0.992337 | |
| Viola(gramm):PhonLoop | 0.484912 | 0.327445 | 1.481 | 0.138634 | |

**WO = SOV; Viola = WO (word order)**

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 1.880403 | 0.443261 | 4.242 | 2.21e-05 | *** |
| DECL | 0.348067 | 0.157569 | 2.209 | 0.027176 | * |
| PROC | -0.501722 | 0.447339 | -1.122 | 0.262046 | |
| Viola(CA) | -2.050726 | 0.539784 | -3.799 | 0.000145 | *** |
| Viola(gramm) | 0.132567 | 0.476797 | 0.278 | 0.780985 | |
| WO(OV) | -0.866388 | 0.368126 | -2.354 | 0.018597 | * |
| WO(SV) | -0.815926 | 0.494073 | -1.651 | 0.098651 | . |
| PhonLoop | -0.185008 | 0.302992 | -0.611 | 0.541462 | |
| PROC:Viola(CA) | -0.166080 | 0.503079 | -0.330 | 0.741304 | |
| PROC:Viola(gramm) | 0.497727 | 0.502420 | 0.991 | 0.321852 | |
| PROC:PhonLoop | -0.340177 | 0.173222 | -1.964 | 0.049551 | * |
| Viola(CA):PhonLoop | -0.003104 | 0.323096 | -0.010 | 0.992335 | |

| | | | | |
|---|---|---|---|---|
| Viola(gramm):PhonLoop | 0.484911 | 0.327449 | 1.481 | 0.138640 |

**WO = SV; Viola = WO (word order)**

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 1.064484 | 0.526015 | 2.024 | 0.043003 | * |
| DECL | 0.348067 | 0.157569 | 2.209 | 0.027176 | * |
| PROC | -0.501719 | 0.447348 | -1.122 | 0.262058 | |
| Viola(CA) | -2.050729 | 0.539778 | -3.799 | 0.000145 | *** |
| Viola(gramm) | 0.132564 | 0.476791 | 0.278 | 0.780987 | |
| WO(SOV) | 0.815922 | 0.494071 | 1.651 | 0.098652 | . |
| WO(OV) | -0.050469 | 0.529233 | -0.095 | 0.924027 | |
| PhonLoop | -0.185008 | 0.302989 | -0.611 | 0.541458 | |
| PROC:Viola(CA) | -0.166083 | 0.503095 | -0.330 | 0.741307 | |
| PROC:Viola(gramm) | 0.497724 | 0.502424 | 0.991 | 0.321859 | |
| PROC:PhonLoop | -0.340177 | 0.173222 | -1.964 | 0.049551 | * |
| Viola(CA):PhonLoop | -0.003105 | 0.323095 | -0.010 | 0.992333 | |
| Viola(gramm):PhonLoop | 0.484912 | 0.327445 | 1.481 | 0.138635 | |

**WO = OV; Viola = grammatical (no violation)**

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 1.146580 | 0.367636 | 3.119 | 0.00182 | ** |
| DECL | 0.348067 | 0.157569 | 2.209 | 0.02718 | * |
| PROC | -0.003994 | 0.295269 | -0.014 | 0.98921 | |
| Viola(WO) | -0.132565 | 0.476810 | -0.278 | 0.78099 | |
| Viola(CA) | -2.183295 | 0.472433 | -4.621 | 3.81e-06 | *** |
| WO(SV) | 0.050463 | 0.529251 | 0.095 | 0.92404 | |
| WO(SOV) | 0.866391 | 0.368134 | 2.353 | 0.01860 | * |
| PhonLoop | 0.299903 | 0.208991 | 1.435 | 0.15129 | |
| PROC:Viola(WO) | -0.497722 | 0.502434 | -0.991 | 0.32187 | |
| PROC:Viola(CA) | -0.663808 | 0.450401 | -1.474 | 0.14053 | |
| PROC:PhonLoop | -0.340177 | 0.173223 | -1.964 | 0.04955 | * |
| Viola(WO):PhonLoop | -0.484912 | 0.327450 | -1.481 | 0.13864 | |
| Viola(CA):PhonLoop | -0.488015 | 0.285815 | -1.707 | 0.08774 | . |

## WO = SOV; Viola = grammatical (no violation)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 2.012970 | 0.316216 | 6.366 | 1.94e-10 | *** |
| DECL | 0.348067 | 0.157570 | 2.209 | 0.0272 | * |
| PROC | -0.003994 | 0.295263 | -0.014 | 0.9892 | |
| Viola(WO) | -0.132563 | 0.476802 | -0.278 | 0.7810 | |
| Viola(CA) | -2.183293 | 0.472423 | -4.621 | 3.81e-06 | *** |
| WO(OV) | -0.866391 | 0.368126 | -2.354 | 0.0186 | * |
| WO(SV) | -0.815928 | 0.494073 | -1.651 | 0.0987 | . |
| PhonLoop | 0.299903 | 0.208990 | 1.435 | 0.1513 | |
| PROC:Viola(WO) | -0.497724 | 0.502421 | -0.991 | 0.3219 | |
| PROC:Viola(CA) | -0.663808 | 0.450390 | -1.474 | 0.1405 | |
| PROC:PhonLoop | -0.340177 | 0.173222 | -1.964 | 0.0496 | * |
| Viola(WO):PhonLoop | -0.484912 | 0.327447 | -1.481 | 0.1386 | |
| Viola(CA):PhonLoop | -0.488015 | 0.285813 | -1.707 | 0.0877 | . |

## WO = SV; Viola = grammatical (no violation)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 1.197044 | 0.488268 | 2.452 | 0.0142 | * |
| DECL | 0.348067 | 0.157569 | 2.209 | 0.0272 | * |
| PROC | -0.003995 | 0.295255 | -0.014 | 0.9892 | |
| Viola(WO) | -0.132565 | 0.476773 | -0.278 | 0.7810 | |
| Viola(CA) | -2.183296 | 0.472408 | -4.622 | 3.81e-06 | *** |
| WO(SOV) | 0.815928 | 0.494049 | 1.652 | 0.0986 | . |
| WO(OV) | -0.050462 | 0.529211 | -0.095 | 0.9240 | |
| PhonLoop | 0.299904 | 0.208988 | 1.435 | 0.1513 | |
| PROC:Viola(WO) | -0.497722 | 0.502405 | -0.991 | 0.3218 | |
| PROC:Viola(CA) | -0.663807 | 0.450376 | -1.474 | 0.1405 | |
| PROC:PhonLoop | -0.340177 | 0.173221 | -1.964 | 0.0496 | * |
| Viola(WO):PhonLoop | -0.484914 | 0.327443 | -1.481 | 0.1386 | |
| Viola(CA):PhonLoop | -0.488016 | 0.285810 | -1.707 | 0.0877 | . |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Appendix K**

# Re-leveling: Confidence Level Model (adults)

## Viola = WO (word order)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 1.1807 | 0.2933 | 4.026 | 5.68e-05 | *** |
| CONF | 0.1667 | 0.1537 | 1.085 | 0.278026 | |
| Viola(CA) | -1.5753 | 0.4179 | -3.770 | 0.000163 | *** |
| Viola(gramm) | 0.4450 | 0.3560 | 1.250 | 0.211308 | |
| CONF:Viola(CA) | -0.4110 | 0.2177 | -1.888 | 0.058983 | . |
| CONF:Viola(gramm) | 0.6616 | 0.1977 | 3.348 | 0.000815 | *** |

## Viola = grammatical (no violation)

Fixed effects:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 1.6257 | 0.2364 | 6.877 | 6.10e-12 | *** |
| CONF | 0.8283 | 0.1289 | 6.428 | 1.29e-10 | *** |
| Viola(WO) | -0.4450 | 0.3560 | -1.250 | 0.211304 | |
| Viola(CA) | -2.0203 | 0.3801 | -5.315 | 1.06e-07 | *** |
| CONF:Viola(WO) | -0.6616 | 0.1976 | -3.348 | 0.000815 | *** |
| CONF:Viola(CA) | -1.0727 | 0.2011 | -5.334 | 9.63e-08 | *** |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1