The Interaction Between Children’s Working Memory And Long-Distance Subject-Verb Agreement: An Experimental And Corpus Investigation

Thesis

How to cite:
Roque Gutierrez, Ernesto (2022). The Interaction Between Children’s Working Memory And Long-Distance Subject-Verb Agreement: An Experimental And Corpus Investigation. PhD thesis The Open University.

© 2021 Ernesto Roque Gutierrez

https://creativecommons.org/licenses/by-nc-nd/4.0/

Version: Version of Record

Link(s) to article on publisher’s website:
http://dx.doi.org/doi:10.21954/ou.ro.00014b21

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
THE INTERACTION BETWEEN CHILDREN’S WORKING MEMORY AND LONG-DISTANCE SUBJECT-VERB AGREEMENT: AN EXPERIMENTAL AND CORPUS INVESTIGATION

Ernesto Roque Gutierrez

Faculty of Wellbeing, Education and Language Studies

Thesis Submitted to the Open University for the Degree of Doctor of Philosophy

September 2021
Abstract

The status of human language as either a recycler of general cognition (Tomasello, 2003; Bybee, 2010; Goldberg, 2006) or an encapsulated module with processes unique to itself (Chomsky, 1980; Fodor, 1983), has been intensely contested. The debate is important because it speaks to what language is made of and how it is acquired. Although evidence is building of a significant interaction between language and cognition (e.g., Dąbrowska & Divjak, 2015, 2019), the extent of this interaction, the development periods in which it occurs, and the domains to which it corresponds, remain unclear. This thesis contributes to the understanding of these issues with four studies. Study 1 explores the directionality of cognitive transfer between working memory (WM) and syntax among L1 and L2-learning 7-year-old children in a randomised controlled trial. Study 2 analyses individual variation in WM performance among these same participants by examining speed/accuracy trade-offs in the context of gender. Using corpus data, study 3 correlates the development of English and Spanish long-distance subject-verb agreement (SVA) with that of non-verbal WM in children between 2 and 12 years of age. Study 4 examines the extent to which WM is implicated in processing long-distance SVA given the level of formulaic language used. The studies revealed that: [1] WM training improved L1 and L2 syntactic performance, but not vice versa; [2] the reaction times to the WM task were bimodally distributed, and boys tended to respond quicker but girls were more accurate; [3] the development of long-distance SVA was predicted by the development of non-verbal WM; [4] the degree of formulaicity observed in sentences with long-distance agreement was consistent with some form of WM involvement. Interpreted together, these results present a compelling and comprehensive case in favour of a deep integration between WM and syntax in the process of language acquisition.
Acknowledgments

From the very beginning of this doctoral journey, I have received immense support and guidance from my supervisors, Dr Paul Ibbotson, whose passion and dedication to research I continue to envy to this day and whose scientific ingenuity has deeply influenced the way I do research; and Dr Lina Adinolfi, whose invaluable advice and patience helped to steer this PhD ship in the right direction.

I want to thank my mother, because she is prouder of me than anyone else, my father, who has missed me more than anyone else, and my brother, who I have missed more than anyone else.

I am always indebted to my friends Javier, Leo and Casuso, who I know will argue about the order in which I have introduced their names. They have historically enjoyed partaking in all my endeavours and have encouraged me to do better, and even more so on this occasion. To every other friend who wanted to but did not stop me from palavering about working memory and language learning.

And finally, I want to thank my partner, Taimy, for enduring every single day of these last three years and for looking after my sanity throughout it all, especially when I lacked the strength to do so. We made it, Ona!
Declaration of Authorship

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own.

References to Relevant Works:


# Table of Contents

**CHAPTER 1 - INTRODUCTION** ................................................................. 16

1.1 CHAPTER OVERVIEW ........................................................................... 16
1.2 MODULARITY AND LANGUAGE .......................................................... 16
1.3 MEMORY AND LANGUAGE ................................................................. 23
1.4 WM AND Sdv .................................................................................. 26
1.5 RESEARCH QUESTIONS ....................................................................... 28
   RESEARCH QUESTION 1 (RQ1) .......................................................... 28
   RESEARCH QUESTION 2 (RQ2) .......................................................... 30
   RESEARCH QUESTION 3 (RQ3) .......................................................... 32
   RESEARCH QUESTION 4 (RQ4) .......................................................... 34
1.6 SUMMARY ....................................................................................... 36

**CHAPTER 2 – LITERATURE REVIEW** ..................................................... 39

2.1 CHAPTER OVERVIEW ........................................................................... 39
2.2 WORKING MEMORY ........................................................................... 39
2.3 WM DEVELOPMENT ........................................................................... 41
2.4 MEASURING WM ............................................................................ 49
2.5 TRAINING WM ................................................................................. 53
2.6 AGREEMENT .................................................................................... 57
2.7 WM AND SVA ................................................................................ 62
2.8 SVA IN SPANISH AND ENGLISH ..................................................... 66
2.9 SVA IN L1 AND L2 .......................................................................... 67
2.10 SUMMARY ..................................................................................... 70

**CHAPTER 3 – STUDY 1: EXPERIMENTAL STUDY** ................................. 72

3.1 CHAPTER OVERVIEW ........................................................................... 72
3.2 INTRODUCTION ................................................................................ 72
3.3 METHODS ....................................................................................... 77
   PARTICIPANTS ................................................................................ 77
   DESIGN ........................................................................................... 78
   CONTEXT ........................................................................................ 79
   MATERIALS .................................................................................. 81
   PROCEDURE .................................................................................. 86
   ETHICS ........................................................................................... 87
3.4 RESULTS ........................................................................................... 87
   L1 PRE-TEST ................................................................................ 88
   L1 POST-TEST ............................................................................ 88
   L2 PRE-TEST ................................................................................ 89
   L2 POST-TEST ............................................................................ 89
   WM PRE-TEST ........................................................................... 90
   WM POST-TEST ........................................................................... 90
3.5 DISCUSSION .................................................................................... 91

**CHAPTER 4 – STUDY 2: REACTION TIMES** .......................................... 96
List of figures

Figure 1. Starling’s recognition of recursive patterns ................................................................. 21
Figure 2. Recursion in visual grouping ......................................................................................... 22
Figure 3. Tree diagrams of centrally embedded and right branching patterns ......................... 27
Figure 4. Four hypotheses tested in Study 1 ................................................................................. 30
Figure 5. Visual conceptualisation of the dissertation ................................................................. 37
Figure 6. Baddeley’s multi-component working memory system ................................................. 40
Figure 7. Memory development from pre-birth to 7 years of age .............................................. 49
Figure 8. N-back stimuli .............................................................................................................. 50
Figure 9. Four hypotheses tested in Study 1 ................................................................................. 73
Figure 10. Study 1 design ............................................................................................................ 79
Figure 11. N-back task instructions .............................................................................................. 82
Figure 12. Variables explored in Study 1 ..................................................................................... 86
Figure 13. L1 SVA by training group ........................................................................................... 89
Figure 14. L2 SVA by training group ........................................................................................... 90
Figure 15. WM performance by training group ............................................................................ 91
Figure 16. Hypothesis A .............................................................................................................. 93
Figure 17. Reference-back task .................................................................................................. 99
Figure 18. N-back procedure ..................................................................................................... 103
Figure 19. Reaction times of hits and false alarms .................................................................. 104
Figure 20. Reaction times of stimuli 1 – 4 .................................................................................. 105
Figure 21. Reaction times of stimuli 5 – 8 .................................................................................. 106
Figure 22. Mean reaction times of all participants .................................................................... 107
Figure 23. Reaction times by gender ........................................................................................ 108
Figure 24. Composition of early and late responses ................................................................. 100
Figure 25. Reaction times to the 1-back by gender ................................................................. 110
Figure 26. Reaction times to the 2-back by gender ................................................................. 110
Figure 27. Reaction times to the 1-back by accuracy .............................................................. 111
Figure 28. Reaction times to the 2-back by accuracy .............................................................. 111
Figure 29. Reaction times to the 1-back and 2-back ............................................................... 112
Figure 30. Median reaction times by gender and modality .................................................... 112
Figure 31. Tree diagrams of centrally embedded and right branching patterns ................... 117
Figure 32. Scatter graph of the English SdV and non-SdV sentences ...................................... 126
Figure 33. Distribution of syllable length between subject and verb in English ...................... 126
Figure 34. Possible routes (Rs) to establish SVA in Spanish ..................................................... 132
Figure 35. Boxplots of Spanish SdV length between 39 and 149 months of age .................... 133
Figure 36. Scatter of the Spanish sentences extracted with clear SV relationships ................ 134
Figure 37. Distribution of syllable length between subject and verb in Spanish ..................... 134
Figure 38. Maximum syllable length between subject and verb ............................................ 135
Figure 39. Trajectories of maximum syllable length of utterance and SdV distance in English ......................................................................................................................... 137
Figure 40. Trajectories of maximum syllable length of utterance and SdV distance in Spanish .......................................................................................................................... 138
Figure 41. Trajectories of English SdV, maximum syllable length and WM ........................... 140
Figure 42. Trajectories of Spanish SdV, maximum syllable length and WM .......................... 141
Figure 43. Sankey representation of five English CS sentences ............................................... 152
Figure 44. Measuring the connectivity of two sentences ......................................................... 153
Figure 45. Measuring the connectivity of three-four sentences ............................................. 153
Figure 46. Sankey distribution of English SdV sentences ........................................155
Figure 47. Sankey distribution of Spanish SdV sentences ........................................156
Figure 48. Linear representation of repetitiveness of English SdV and non-SdV sentences 157
Figure 49. Linear representation of repetitiveness of Spanish SdV and non-SdV sentences 158
Figure 50. Visual summary of the four studies of this dissertation .................................162
List of tables

Table 1. Contextual information of the schools intervened in Study 1 ................................................. 80
Table 2. L1 and L2 frequency analyses in Spanish and English .......................................................... 85
Table 3. Chi-square table representing accuracy by early/late response ........................................ 104
Table 4. Chi-square table examining the relationship between gender and early/late responses .......................................................... 107
Table 5. Chi-square table (gender by accuracy). .............................................................................. 108
Table 6. Characterisation of the three English corpora that were sampled .................................. 120
Table 7. Working taxonomy of SdV sentences ................................................................................. 123
Table 8. Characterisation of the two Spanish corpora that were sampled .................................. 127
Table 9. Analysis of variance of syllable distance in English SdV sentences .............................. 142
Table 10. Analysis of variance of syllable distance in Spanish SdV sentences ............................ 142
List of acronyms

ANCOVA – analysis of covariance
ANOVA – analysis of variance
CDS – child-directed speech
CHILDES – Child Language Data Exchange System
CLAN – Child Language Analysis Program
CS – child speech
DCL – developmental cognitive linguistics
EF – executive function
ERP – event-related brain potential
ESOL – English for Speakers of Other Languages
FA – false alarm
FLB – faculty of language (in the) broad (sense)
FLN – faculty of language (in the) narrow (sense)
fMRI – functional magnetic resonance imaging
Gf – fluid intelligence
GJT – grammaticality judgment task
H – hits
JSON – JavaScript Object Notation
L1 – first language
L2 – second language
LDD – long-distance dependency
LTM – long-term memory
MaxSV – maximum syllable distance between subject and verb
PM – picture memory task
PS – picture span task
RQ – research question
RT – reaction time
SoV – Sentences with 0 syllables between subject and verb
SdV – Sentences with at least one syllable (distance) between subject and verb
SPRT – self-paced reading task
STM – short-term memory
SVA – subject-verb agreement
TD – typical development
WISC-V – Wechsler Intelligence Scale for Children – Fifth Edition
WM – working memory
WMT – word-monitoring task
WPPSI-IV – Wechsler Pre-School and Primary Scale of Intelligence – Fourth Edition
Chapter 1 - Introduction

1.1 Chapter overview

This chapter introduces the general theories of language and cognition that underpin the theoretical framework of this dissertation. Section 1.2 provides a brief recount of the position shared by the theories that see language deeply integrated with, and somewhat contingent on, domain-general cognition (Tomasello, 2003; Bybee, 2010; Goldberg, 2006; Ibbotson, 2020) in contrast to that derived from the mind modularity theory (Chomsky, 1980; Fodor, 1983). This contributes to defining the raison d’être of the developmental cognitive linguistics (DCL) approach within the integrationist group, and to declaring the stance adopted by this research project. Section 1.3 proceeds to describe in general terms some of the well-researched connections between working memory (WM) and language, including ties with phonological, semantic, and syntactic development. This is followed by section 1.4, which zooms in on the previously established links between WM and the processing of subject-verb agreement (SVA), which is one of the objects of the research. The brief characterisation of the field sets the tone for the introduction of the four research questions that motivate each study (section 1.5) and the discussion of the rationale behind each one. Lastly, section 1.6 provides a conceptual map of the entire dissertation highlighting the internal relationships among all studies.

1.2 Modularity and language

‘... a genetic inheritance, a mathematical system, a social fact, the expression of individual identity, the expression of cultural identity, the outcome of a dialogic interaction, a social semiotic, the intuitions of native speakers, the sum of attested data, a collection of memorized chunks, a rule-governed discrete combinatory system, or electrical activation in a distributed network... We do not have to choose. Language can be all of these things at once.’ (Cook & Seidlhofer, 1995)

This dissertation sets out to explore the connections between WM and SVA and their implications for a more comprehensive understanding of the relationship between general cognition and language. There are multiple accounts of both WM and SVA in the literature so perhaps it is convenient to clarify what both constructs mean in the context of this dissertation from the very start. In simple terms, WM is a multi-component system responsible for organising, storing, and updating sensory information needed to perform a cognitive task. This processing and manipulation of
information is also mediated by other executive functions (EFs), specifically attention and inhibition — a full definition and a critical review of the sources that inform it is provided in section 2.2. SVA, on the other hand, refers here to the realisation of morphological features through memory retrieval from verb to subject, whilst being subject to interference — this rationale behind this definition is discussed in section 2.6.

The study of language and cognition, was, for many years defined by two opposing positions. The first, originating in the middle of the 20th century and gaining more traction in the 1980s, bases itself on the ‘mind modularity’ paradigm (Chomsky, 1980; Fodor, 1983), a metaphorical representation of the human brain as a ‘cognitive mosaic’ composed of largely independent units or modules, one of which is language, that interact minimally with each other. Proponents of the second stance favour a more integrated view of human cognition, supported by findings of developmental psychology research on how children’s learning mechanisms, including those related to language, develop over childhood (Bybee, 2010; Goldberg, 2009; Ibbotson, 2020; Tomasello, 2003). This opposition in the field was perhaps more noticeable some decades ago between nativists (i.e., Fodor) and constructivists (i.e., Piaget) but the dichotomy between modular and interactive theories has eroded over the last years in light of new discoveries reporting important interactions between language and cognition (e.g., in Dąbrowska & Divjak, 2015, 2019; Ibbotson, 2020). However, much remains unknown about the extent of this interaction, the development periods in which it occurs, and the domains to which it corresponds. After more than twenty years of empirical research, the lessons learned from many usage-based studies is that language learning, like all human learning, is exemplar-based. Young learners first understand individual instances of language occurring in local, meaningful communication with others, and then proceed to create more abstract, interactive schematic representations. However, this network of representations of a mental grammar still remains metaphorical, to a large extent, as more empirical studies interested in testing specific interactions between domain-general cognition and language are needed. This chapter aims to transit from one group of theories to the other whilst underlining key differences in how they define the relationship between domain-general cognition and language and more specifically between WM and syntax.
The field of cognitive linguistics also places cognition at the centre of language learning, but thus far the main focus remains on studying mostly adult language acquisition, in both their first (L1) and second (L2), and considerably less attention has been allotted to child language learning. One of the aims is to situate this research in this overlapping space, by adopting a position that amalgamates key theoretical and methodological aspects of both developmental psychology and cognitive linguistics, aided by both cross-sectional and longitudinal approximations to children’s acquisition and processing of syntax through a cognitive-tinted lens. ‘Developmental cognitive linguistics’ (henceforth DCL) (Ibbotson, 2020) is the term that will be used to refer to this ‘marriage’ between developmental psychology and cognitive linguistics, and the one that best summarises the theoretical stance of this investigation. The research is ‘developmental’ because of the longitudinal focus on children’s cognitive and linguistic development. The internal use of the ‘cognitive’ tag relates to children’s WM, while the ‘linguistics’ tag refers to sentences with long-distance SVA. These terms will be explained in full later in this chapter, following a brief overlook of how language and cognition have been conceptualised by proponents of both group of theories, highlighting some of the theoretical discrepancies between them with regards to language acquisition and processing. This will serve to provide the theoretical context to frame this research.

According to The modularity of mind: an essay on faculty psychology (Fodor, 1983), one of the distinctive characteristics of a modular system, namely the human brain, is that input systems (i.e. five senses plus language) are informationally encapsulated. That is, the input that is available to one module is not available to others, so visual-motor information, for example, is only handled by a dedicated processor that is specific to the visuospatial domain. Fodor has used visual illusions to explain his view of a fragmented brain. Visual illusions, even if they are explained prior to exposure, cannot be deciphered. The information containing what the illusion is and why it happens does not change how the visual input system processes the stimulus. The fact that our eyes cannot escape being deceived by the illusion is used as support for the claim that visual information is solely processed by a domain-specific apparatus, impervious to the rest of cognition. Modules are isolated from each other and are incapable of accessing information in a shared repository or
central monitoring system so input systems are bound to process the information they receive autonomously and automatically, in the same way that reflexes are triggered by specific stimuli and occur independently from conscious thought (Fodor, 1983).

There is a caveat to this analogy, however, one that Fodor uses to dissociate his theory of mind modularity from Gall's (1835) 19th century, pseudo-scientific practice of phrenology. Gall believed in a more radical form of faculty autonomy, where no cognitive mechanisms (i.e., attention, memory) were shared at all, so the processes involved in remembering the lyrics to a song would be exclusive and distinct to those involved in remembering a picture. On the other hand, Fodor, a reader of Gall, acknowledges that systemic relations must be operating across modules, especially because memory and attentional resources are limited in usage. That is, the mental effort required to remember the song lyrics cannot be too dissimilar from that needed to remember the picture, given that allocating too much or too little mental energy on one, affects the other. In this way, access to the computational resources employed in remembering is not only limited but shared. Therefore, Fodor's mental units are not autonomous in the sense they are self-reliant in operation, but autonomous mostly in an ‘informational’ sense.

For language, that means that the dedicated linguistic input processor cannot access information that is available to each of the sensorial systems and vice versa, but they all compete for the computation resources needed to decode and store their own information. Both strict and loose interpretations of this relationship between language and general cognition have been abundant in the field, which has inevitably led to confusion (Jackendoff & Pinker, 2005). In an attempt to address specifically the extent to which language was separated from the rest of cognition, in 2002 Chomsky revised the original language-as-module construct and made a distinction between the Faculty of Language in a Narrow sense (FLN) and the Faculty of Language in a Broad sense (FLB) (Hauser et al., 2002). FLN was conceived as the linguistic module containing all the information that is special and unique to language, the language organ per se, the nucleus. FLB, on the other hand, would represent the interface or ‘soft boundary’ with general cognition that enables each
module, including language, to recruit attentional, memory and other resources for processing.

It is perhaps worth highlighting at this stage that the discussion around ‘modularisation’ has shifted from the original dichotomy between the type of nativism described by Fodor (1983) and Piaget (1950)’s constructivism. One very well-documented attempt of reconciling the best of both epistemological stances, specifically the arguments around domain-specificity and domain-generality in developmental change, comes via Karmiloff-Smith’s representational redescriptions model (1992). For Karmiloff-Smith, cognition cannot be understood nor explained if developmental changes are not thoroughly accounted for. In that regard, she believes in ‘gradual modularisation’, where adult brains become more modular over time as different brain circuits, including those associated with language processing, become domain-specific (Lorandi & Karmiloff-Smith, 2012). Children’s brains show higher levels of plasticity and rely more heavily on domain-general mechanisms. In other words, Karmiloff-Smith’s model rejects the ‘nativeness’ of the brain modularity paradigm (Fodor, 1983; van der Lely, 2005; Spelke & Kinzler, 2007) and envisages domain-specificity as a developmental process, which is to say that when it comes to language processing, children, unlike adults, start out by recruiting domain-general processes that become increasingly specialised in their content as they move towards adulthood.

However, in spite of Karmiloff-Smith’s efforts to discourage people from re-engaging into the nativist versus empiricist debate, the conversation is still ongoing as the boundaries between domain-specificity and domain-generality in language continue to be mapped. Attempting to unpack the contents of FLN and FLB, for example, and the nature of intra- and intermodular interactions has been an elusive task, as Hauser et al. provided very limited testable examples of one or the other. For instance, when describing the uniquely human ‘endemicity’ of FLN, they exclusively relied on recursion, what Jackendoff & Pinker (2005) have called the ‘recursion-only claim’. In linguistics, recursion is the syntactic ability to embed (and parse) grammatical phrases in other phrases, across different hierarchies and effectively ad infinitum. Among the reasons provided to support the candidature of recursion as FLN are its observed lack of presence in animal communication systems (Fitch &
Hauser, 2004) and in other human cognitive domains (Fitch et al., 2005). Both of these claims, which suggest that recursion is a human language-specific principle, have been partially challenged.

In the case of animal communication, European starlings have been observed to recognise different self-embedding recursive sounds that were part of a centrally embedded pattern $AB$, where $A$ were rattle sounds and $B$ warble sounds (Gentner et al., 2006). Column b in Figure 1 shows the recursive pattern that starlings were able to identify whenever a new $AB$ pattern was inserted centrally. For example, the pattern $ab$, in red and green, respectively, would move from level 1 to level 4 differently depending on the grammar. For grammar A, it would retain its initial position, while in grammar B, $a$ and $b$ would cascade towards the ends as new patterns were introduced. However, although these findings support the claim that starlings are able to identify central embedded patterns after being trained to do so, positive recognition of true centre-embedded recursion would also imply understanding each constituent part in the system, not just recognising the sequential pattern, hence the partial challenge.

![Figure 1](image)

In human language terms, this means that being able to signal that sentences 2 and 3 (below) are centrally embedded, as opposed to sentence 1, means not only responding to a defined *that*-pattern but also understanding the connection between
the constituting parts, that is that woman is linked to running, dog to biting and mouse to scaring.

1- The woman ran
2- The woman that the dog bit ran
3- The woman that the dog that the mouse scared bit ran

On the other hand, the second supportive claim for recursion, namely its apparent absence in other human cognitive domains, has been disputed by Jackendoff & Pinker (2005). Using evidence from human visual cognition (Figure 2), they have shown that visual perception is sensitive to sequential patterns governed by the same principles that are imperative to recursion, namely the endless repetition of gestalt-like hierarchical structures. In Figure 2, the repetition of discrete pairs leads to larger clusters that follow the same hierarchical structure (i.e., pairs lead to quartets, then octets and so on).

Additionally, similar grouping principles have been exposed by the generative theory of tonal music (Lerdahl & Jackendoff, 1983), where Western tonal music has been observed to be built off tension-relaxation patterns that can be repeated recursively. Further evidence challenging the recursion-only hypothesis can be found in other linguistic studies of a typological nature, especially those focusing on the Amazonian language Pirahã (Everett, 2005; Futrell et al., 2016). In a corpus analysis of naturalistic Pirahã originally collected by Everett (2005), Futrell and colleagues found no compelling evidence of syntactic centre embeddedness, which is the paragon of recursion. Therefore, the assertion that recursion is what makes language special is flawed if at least one language does not exhibit recursive patterns.
Over time, FLN has continued to shrink in size, being reduced to the point of exhaustion, where all that is left is ‘merge’, the basic property of merging two syntactic objects to form a new one, responsible for recursive operations at the syntactic level (Chomsky, 2015). The narrow faculty that allegedly makes language distinct from other mental modules has encountered multiple obstacles from its very inception. Even before delineating the contents of FLB, it is not unreasonable to expect its involvement in language to be considerable, given how little FLN seems to contribute to using language. Notwithstanding the actual credibility of the existence of FLN, recursion does not seem to answer the question of what makes language unique to humans.

1.3 Memory and language

While the interconnection between language and short-term (STM) and long-term memory (LTM) has been extensively documented in the past (Caplan & Waters, 1990; Daneman & Merikle, 1996), this dissertation is particularly concerned with the one between language and WM (Baddeley & Hitch, 1974), a multi-component cognitive system that has replaced the traditional view of STM (Atkinson & Shiffrin, 1968; Broadbent, 1958) as a unitary entity acting independently from the rest of cognition. The WM construct has been updated and amended throughout the years. Initially, efforts were primarily concerned with describing the architecture of the newly proposed system and the content of its components (Baddeley, 1986, 1998, 2000, 2003). Once this was somewhat settled, the main focus shifted towards characterising the functionalities of the system, highlighting its properties of storage, maintenance and the updating of information, including in situations with interfering conditions (Conway et al., 2005; Daneman & Carpenter, 1980; Miyake & Shah, 1999). Generally speaking, the consensus is that WM is a system that orders, stores and processes immediate sensory details which are subsequently integrated into general cognition. The amount of data that can be stored temporarily is limited and the speed of activating it varies. In other words, WM allows humans to temporarily store and update specific information while they simultaneously perform other related or unrelated cognitive tasks.
Following these principles, many have theorised that WM capacity is responsible for individual differences between humans in their language comprehension ability (Hitch et al., 2001; Just & Carpenter, 1980, 1992; Miyake, 2001), as reading and listening often involve holding some information momentarily while old information is updated or new input is processed. In the case of anaphoric references, for example, any time a pronoun or substitute is used (i.e., she, they), readers/listeners need to establish the correct match with the corresponding source for coherent comprehension, as in sentence 4 below.

4- *I know that Laura gets really uncomfortable when she is being stared at.*

This is not only true of syntactic relations but also applies to the on-line processing of polysemous words, such as the noun *drill* in example 5, a well-known example taken from Daneman & Merikle (1996). In this case the newly encountered information about *drill* challenges the semantic choice previously determined by what was initially the context.

5- *The dentist used the drill. He found it useful for mastering the Latin and Greek verbs.*

The body of psycholinguistic evidence arguing for clear connections between WM and children’s language acquisition and processing is extensive and well-established. For instance, Gathercole and colleagues have observed positive correlations between variation in phonological STM and the vocabulary development of children with typical development (TD) over a 4-year span, from 4 to 8 years of age (Gathercole et al., 1992). Their findings for the first two years were particularly compelling in favour of a deep integrationist view of language and cognition, and phonological memory skills were thought to play a causal role in the development of children’s vocabulary.

Unambiguous links between children’s WM and syntax have been also thoroughly documented. Usually, the evidence is of a correlational nature and between TD, school-aged children’s verbal WM, and on-line syntactic processing (Gaulin & Campbell, 1994). After surveying both children and adults, McDonald (2008), for example, found that WM capacity, along with phonological ability, accounted for individual variance in grammaticality judgement tasks (GJTs), after controlling for age effects. The GJTs in question included sentences whose ungrammaticality consisted of syntactic features such as word order (6) and third person agreement.
(7), alongside cases involving word morphology (8). Participants with larger WM capacity found it easier to identify those instances of ungrammaticality than those with more reduced capacity.

6- *The teacher the tests graded
7- *The boy jump whenever he is startled
8- *Several of the mans decided not to go to the football game

(Taken from McDonald, 2008b)

Some studies have, however, gone beyond the correlational and pushed forward the understanding on the possible causal relationship between children’s WM and syntax acquisition, following Gathercole and colleagues’ blueprint. A study by Swanson discovered significant intercorrelations between verbal and visuospatial WM measures when both were tested against standardised intelligence and aptitude measures (Swanson, 1996b). Among the measures with strong correlational effects were verbal processing tasks that required remembering event sequences in a short story and numbers in short sentences, and visuospatial tasks where children had to memorise pictures in a sequence and the spatial organisation and distribution of objects. Some task pairs showed a poor correlation, for example those where phonological WM involvement (a rhyming sequence task) was matched with non-verbal sequencing. One possible explanation for the lack of correlation, from the DCL standpoint, concerns the intervening role of other executive functions (i.e., attention, inhibition) in facilitating the compatibility of two tasks. This is always true for DCL. Just because a task has been meticulously designed and trialled to measure WM performance does not entail that it will solely measure WM. That said, the results by Swanson did show that, generally speaking, children who performed well in verbal sequencing tasks also excelled at non-verbal sequencing tasks. This finding suggests that verbal WM and non-linguistic WM are more interconnected than what the language-as-module constraints permit, given the mutual exclusion between the vision and language domains. Furthermore, it represents an important step towards a more accurate delineation of the boundaries between children’s WM and syntactic processing, and a preview of the type of interaction between cognition and language proposed by DCL.
Many of the findings of these developmental studies on children have been replicated in comparable studies on adult participants (Caplan & Waters, 1999), and positive correlational effects between WM and syntax have been observed in mixed-age participant samples, regardless of age differences (McDonald, 2008b; Swanson, 1996b).

1.4 WM and SdV

As the collection of studies above have shown, indisputable links have been established between WM and language, and more specifically syntax, and it is precisely to this area of the field that this research intends to cater to, by focusing on the WM processing demands of grammatical concord in sentences with distal subject-verb interactions or long-distance agreement (SdV). This type of sentences are ‘loosely’ related to long-distance dependencies (LDD), also called filler-gap or unbounded dependencies, which include resolutions in questions such as What do the parents of the children think?, previously studied by Dabrowska et al. (2009), Dąbrowska & Lieven (2005), and more recently by Bentea & Marinis (2021), which are not the focus of this dissertation.

Revisiting the sample sentences 1-3 showing recursion (below) shows that the iteration could be performed endlessly by moving from sentence 9 to 10, 10 to 11, etc. Yet there are obvious limitations to the number of embedded structures that humans can process, given their WM capacity (Gibson, 1998). There is a memory cost to be paid. It could be argued that sentence 11 is more difficult to parse than sentence 10 because [1] it is longer, as a result of [2] having an extra noun-verb pair. However, [1] and [2] are also true for sentence 12, but its decentralised structure, which follows a right-branching pattern, makes the processing of its constituents arguably easier than those of sentence 11 (see Figure 3 below). One structural component of sentence 11 that adds to its complexity is the longer distance in long-distance SVA. The memory load of the ‘distance’ in SdV sentences, namely that the dog that the mouse scared bit, competes for the same resources needed to establish concord. The speaker has to mentally hold eight words before establishing concord, which unsurprisingly places an additional burden on their working memory.
9- The woman ran
10- The woman that the dog bit ran
11- The woman that the dog that the mouse scared bit ran
12- The woman ran when the dog bit her after the mouse scared him

Figure 3. Tree diagrams of [A] centrally embedded and [B] right branching patterns. Words in red highlight the head-argument (subject-verb) relationship for grammatical agreement. The long SV distance in A is marked by the intervening embedded structure.

Therefore, centre-embedding does not seem to be limited solely by one’s ability to use recursion, but also by one’s general cognitive abilities, specifically in this instance WM capacity. The processing of centrally embedded sentences and the resulting load on WM, as illustrated by sentence 3, has been documented in the past (Caplan & Waters, 1999). However, because the recursive patterns cannot be processed in isolation as constituents, they compete for the same memory resources required to effectively process the sentence’s subject-verb concord. Therefore, finding a systematic method for weighing the memory burden of what is embedded when establishing SVA becomes a viable way of exploring the interaction between WM and recursive mechanisms not only in sentences that are centrally embedded but also in sentences with any type of interrupted agreement (i.e., SdV). If WM is observed to play a subsidiary role to more significant language-specific processes when parsing sentences with distal subject-verb relationships, then this would support the FLN/FLB account. But if, in fact, there are strong transfer effects from working memory to establishing agreement in sentences with distal subject-verb relationships, this would back a more unified view of cognition and language. This is because, under the modular view, domain-general cognition plays no serious role in constructing language itself, but rather, developing cognition applies a filter through which the encapsulated language system emerges.
1.5 Research questions

After having discussed the theoretical backdrop of the interactions between WM and language, the following research questions that circumscribe the studies to be discussed in Chapters 3, 4, 5, and 6 are posited below.

Research question 1 (RQ1)

(RQ1) Can non-linguistic WM training lead 7-year-olds to perform more accurately SVA in L1 and L2 SdV sentences in a GJT transfer task and vice versa?

RQ1 motivates the experimental study discussed in full detail in Chapter 3. The question merits unpacking as it informs the other RQs. Accordingly, this section will first, dissect the question by parts and explain the constructs of WM training and cognitive transfer; second, address its scope and what the predictions of this study are; and finally, comment on some of the limitations of these predictions and how they lead to research questions 2, 3 and 4.

First, the question as a whole summarises the intent to move beyond correlational data and into possible causal associations between WM and syntactic processing. Methodologically speaking, the experimental approach of using a transfer-of-training design facilitates the exploration of this causality, as discussed hereon. Extensive practice in several WM tasks has been found to lead to better performance in these same tasks (Simons et al., 2016). However, more often than not, researchers have found that the benefits of that practice do not improve performance on a different WM task, leading many to believe that cognitive transfer is only possible when the knowledge and skills learned are identical to those applied later and others to doubt its very existence (Detterman, 1993). Whenever training in a task is assessed in an untrained task that shares many elements of its content and abilities, any positive effects are referred to as near transfer, whereas if the pair of training and transfer task are significantly different, far transfer is alluded to. Near transfer refers to situations where the training task and the untrained task are closely related and the observed effects from one to the other are short-lived. Conversely, far transfer applies to tasks that are more dissimilar and the effects have a longer-term impact.
In an exhaustive review of transfer studies, including commercial cognitive training packages aimed at children, such as Cogmed and Cognifit, Simons et al. (2016) concluded that cognitive training leads to better performance in trained tasks, but these effects are less likely to transfer to untrained tasks that are distantly related.

There are WM training studies with child participants that have reported cognitive gains, although these are almost exclusively near-transfer effects (cf. Barnett & Ceci, 2002). Because these similarities are not always clear-cut for each field, in the case of WM training studies, near transfer has been documented between tasks of the same paradigm (recall, span) (Gathercole et al., 2019b) and far transfer between WM and other cognitive abilities (Jaeggi et al., 2008). A 2013 meta-analytic review of 23 WM training studies corroborated that positive near-transfer effects had been observed for verbal WM and visuospatial WM and cited limited evidence suggesting that the effects could be maintained long-term in the case of the latter. However, they failed to find sufficient data to support transfer between WM and more general cognitive skills, including linguistic processing (Melby-Lervåg & Hulme, 2013a). A more recent meta-analysis of 24 WM training studies where children and adults were tested on 113 trained and untrained task pairs found that only specific WM tasks that shared the same principles of processing (e.g. backward span) showed transfer effects across domains (verbal and visuospatial) (Gathercole et al., 2019b).

The transfer of training proposed here is twofold, involving a non-linguistic visual WM task and a syntactic processing task. That is, the transfer effects of training children's visual WM will be assessed in an untrained grammatical exercise, and conversely, the effects of training children on SVA will be tested in an untrained visual WM task. This allows for the exploration of the potential bidirectionality of the effects, which have not been accounted for in previous studies in the field, and yet exemplify the view of integrated cognition that DCL advocates. In other words, if only one direction were to be tested, for instance the effects of WM training to syntactic processing, the results would be insufficient to make any claims of the interaction between general cognition and language, at least from a developmental cognitive perspective. The principal benefit of following an experimental approach of transfer is that it equips researchers with the type of long-focus methodological lens
that zooms in on the interactions between WM and syntax that would otherwise remain undetected.

Second, if children who undergo non-linguistic WM training are able to improve their processing skills in SVA in the transfer task (GJT) in their L1 and L2, then positive cognitive transfer will be established. This would be suggestive of a deep integrationist account in which domain-general WM abilities exert a significant influence on what were thought to be language-specific processing skills. There are four possible outcomes for each language being tested (Figure 4).

![Diagram showing four hypotheses](image)

**Figure 4.** These are the four hypotheses tested: [A] Improved WM performance will transfer into improved syntactic ability, but not the other way around; [B] improved syntactic ability will transfer into working memory performance, but not the other way around; [C] interactions between working memory and syntax will run in both directions; [D] interactions will run in neither direction.

The implications of obtaining similar or divergent results for the L1 and L2 are reviewed in chapter 4. The predictions are that A-C are possibilities that are aligned with the type of integration proposed by DCL and D is the one that most accurately depicts the interdependence between language and cognition of the modular brain. Scenario C, on the other hand, is perhaps the most accurate representation of the type of interdependency between domain-general cognition and language. Cases of A are scarce, or even more generally from WM to language processing, but do exist in the literature (for a detailed review see section 2.5 in the following chapter) while no records supporting B have been uncovered in the literature reviewed.

**Research Question 2 (RQ2)**

The experimental study designed to answer RQ1 provided considerable data offering insights into the processing of n-back tasks well beyond what was required for the
purposes of RQ1. The answer to RQ1 is based on measurements of accuracy, in this case indicated by the number of true responses as a proportion (%) of the total stimuli. However, another set of data that is to some extent tangential to the connections between language and cognition theorised here, but still relevant to this research is that concerned with reaction times (RTs). Because variation in children’s RTs to the n-back tasks has direct implications for any claims about the possible effects of WM training, RQ2 was formulated to characterise in more detail the cognitive side of the processing discussed in RQ1. Accordingly, RQ2 was framed as follows:

(RQ2) What motivates individual variation, if any, in the reaction times to the n-back task?

Variability in response latencies performance in different cognitive tasks has been documented extensively for populations of different ages (Woods et al., 2015) and with neurological differences (Epstein et al., 2011). More specifically, studies testing n-backs in the field have used response latencies in tandem with values for accuracy (i.e. % of correctness) as dependent variables to offer more nuance of the overall performance of a task (for a list of studies see Meule, 2017). Traditionally, increasing the difficulty of the task (n+1) leads to lower accuracy and longer RTs, and longer RTs have also been linked to more mistakes. However, recent studies have also observed less obvious dissociations between accuracy and RTs. Jaeggi et al., (2010), for example, reported positive correlations between accuracy, and not RTs, and fluid intelligence, in more complex versions of the n-back (3 factorial) using visuospatial, auditory, and dual stimuli. RTs in visuospatial n-backs did correlate well with other span tasks such as reading and digit forward. Another study by Hur et al. (2017) offers a different perspective on this matter of accuracy versus reaction times by focusing on the influence of emotions on perception and WM. They did so by comparing how participants would perform in a modified o-back task showing emotional words or pictures of emotional scenes versus a traditional 2-back test. Unsurprisingly to them, participants showed longer response times to emotional stimulus and there was less variability in RTs in the 2-back test. This is because in general, participants seem to be more concerned with doing the task well than doing it fast, especially when affective information is involved (see review in Meule, 2017).
Additionally, patterns in variation in RTs to n-backs have been linked to gender differences in children (Vuontela et al., 2003; Pelegrina et al. 2015). Vuontela et al. (2003), found that school-aged boys were quicker to respond to visual and auditory stimuli than girls. In general, girls aged 6 to 8 took longer to respond and were more accurate than boys, although this difference was less significant with older groups (aged 8 to 10 and 10 to 12). Pelegrina et al. (2015), on the other hand, provided normative data for the n-back task from a sample of 3722 Spanish children and adolescents between 7 and 13 years of age. Likewise, they observed significant effects of gender, with girls outperforming boys whilst taking longer to respond.

These findings suggest that accuracy and RTs should not be interpreted as interchangeable values of performance in n-backs and that reporting both sets of results, when possible, bolsters any claims to be made about the processing demands of the task. The study of individual differences in WM is especially relevant when taking into consideration that the understanding of the construct and others associated to it (i.e., executive control) has benefited greatly from important investigations arising from this subfield. More importantly, much of this research has established strong connections between individuals’ performance in WM tasks and cognitive skills employed regularly in real life activities (Ackerman et al., 2005; Kane et al., 2005).

**Research question 3 (RQ3)**

Even when the construction of the SdV sentences used in the GJT can be informed by the real frequency of occurrence of the lexical items that integrate them in naturalistic settings, these were assembled to test children’s processing prowess, but are not samples of real-life language. With special interest on language production rather than language processing, RQ3 was devised specifically to trace the emergence of SdV structures in L1 naturalistic child speech (CS) as follows:

(RQ3) Can non-linguistic WM act as a developmental predictor of children’s ability to establish grammatical agreement in SdV sentences?
While the answer to RQ1 could provide empirical evidence suggesting a causal link between general WM mechanisms and the processing of grammatical agreement in SdV sentences, thus addressing the C and L in DCL, it cannot answer the developmental (D) implications of these findings. As such, RQ3, discussed fully in chapter 5, attempts to trace the emergence of SdV sentences in naturalistic L1 CS over a 10-year span, starting from 2 years of age, using English and Spanish corpus evidence. More specifically, what is measured is the maximum distance between subject and verb, using the number of syllables as the unit of measurement.

The primary reason for using a syllable count as opposed to a word count is that memory span has been revealed to be more sensitive to the syllable length of phrases than word length (Baddeley et al., 1975). In concert with this observation by Baddeley and colleagues, more recent empirical studies have detected that Arabic- and Finnish-speaking children struggle more to recall derived and inflected words than uninfl cted or base words, implying that morphological complexity seems to add to the load of verbal WM, even when the languages are morphologically distinct (Cohen-Mimran & Adwan-Mansour, 2013; Service & Tujulin, 2002). In Service & Tujulin’s study, for example, a complex span task called Last Word Span required that children listened to a sentence in Finnish (L1) that might or not match an image shown on a screen, while simultaneously memorising the last word of the sentence (see 5-8, taken from Service & Tujulin, 2002). Their findings revealed that children found it easier to remember conscious and teach in sentences 13 and 15 than unconscious and teacher in sentences 14 and 16.

13- The man lying on the ground is conscious
14- The man lying on the ground is unconscious
15- Next to the sick man stands a teacher
16- Next to the sick man stands a teacher

The maximum syllable distance between subject and verb (MaxSV) is therefore used as a more accurate measurement of the language load taxing WM in SdV sentences, and, as such, serves as the dependent variable [0] in the developmental comparison. Unsurprisingly, the first predictor [1] proposed is a systemic measurement of

---

1 The original sentences are in Finnish. These are English equivalents provided by the authors in the original study.
children’s WM development for the same age period. For this, normative scores of WM subtests of the Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition (WPPSI-IV) and the Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) intelligence tests were used. The specific subtests, Picture Memory and Picture Span, are visual WM sequential tasks, especially selected given that no language processing is required in their completion. A second predictor [2], conceived as a representation of the incremental length of right-branching sentences, was also measured in syllables and extracted from the same corpus of [0]. Right-branching structures, as opposed to centrally embedded ones, do not burden WM resources more than any other chunks of syntax (refer back to sentences Figure 3), and are therefore designated to illustrate the somewhat language-specific, memory-independent expansion of CS syntax.

In essence, RQ3 is designed to project these three measurements over the same longitudinal period to explore whether the developmental trajectory of the maximum distance between subject and verb in SdV sentences is to some extent contingent on WM. If, for example, the development of MaxSV correlates well with that of WM, this would add to the suggestion that processing SdV sentences is contingent on WM. That is to say that children can only succeed at establishing grammatical concord in sentences with distal subject-verb relationships if their WM allows for it. This longitudinal observation thus serves not only to address the qualitative concerns of the nature of the English and Spanish SdV sentences in the training/transfer tasks but also to complement the cross-sectional results of the experimental study.

Research question 4 (RQ4)

It is reasonable to assume that even if transfer is established in Study 1 and positive correlations are made between the development of WM and that of SdV sentences, other intervening factors could be facilitating improved WM performance, one of which concerns the morphosyntactic formulaicity of the SdV sentences. The attention in Study 4 was directed towards a qualitative appraisal of the English and Spanish SdV sentences extracted from the corpora in Study 3. As such, the following research question was posited:
(RQ4) What are the implications of the morphosyntactic formulaicity of SdV sentences for WM?

Earlier it was mentioned that by DCL accounts, cognitive functions are inextricably interwoven, so attentional and inhibitory mechanisms, for example, could play a decisive role in both training and the transfer task. There is at least one measurable linguistic variable in the transfer study which could assist participants in the grammatical concord task and thus have confounding effects on the results, and that concerns the syntactic composition of the subject-verb distance. As discussed above, in SdV sentences the grammatical concord between subject and verb is characterised by their distal relationship, and successful agreement is theorised here to be dependent on the WM resources required to hold the elements between subject and verb active. If the syntactic composition of the intervening elements is flexible, meaning, if the structures to be hold mentally between subject and verb are not unusually fixed or repetitive, then WM will be taxed on every occasion. But if, however, there is some ‘formulaicity’ to the syntactic structures enclosed by subject and verb, then it could be argued that the WM burden is somehow alleviated by the syntactic iteration.

This is to more generally grounded on usage-based models of language learning, where children learn in a piecemeal, item-based type of fashion, moving from fixed formulas to patterns to more abstract schemas (Abbot-Smith & Tomasello, 2006; Dabrowska et al., 2009; Dąbrowska & Lieven, 2005; Tomasello, 2003). More specifically, studies such as Dabrowska et al. (2009), for example, have observed that children as old as six continue to rely on formulaic patterns to process questions with long-distance dependencies, albeit less than they do when they are two and three years old (Dąbrowska & Lieven, 2005). Studies using response time data have observed that the time it takes children and adults to process familiar 4-word chunks is significantly shorter than the time they take to process unfamiliar 4-word chunks. Familiarity in this case is tested by continuously exposing participants to prefabricated syntactic frames where the first three words are similar in all instances but the fourth word changes. In a meta-analysis by Arnon & Snider (2010), the continuous frequency or repetitiveness of the chunks were shown to be better predictors of participants reaction times.
Chunking has been conceptualised as an information-processing mechanism summoned by auditory and visual perceptual systems, learning and cognition operating in humans and animals (Gobet et al., 2001). Essentially, it is an online processing skill that is often recruited by one’s senses and other general cognitive domains, such as WM and attention. As a cognitive tool, it assists child learners from a very early age in chunking single items of information into multicomponent sequences for faster processing and recall. In the case in point, if chunking processing skills are believed to be less proficient due to the lack of repetitiveness of the target language, which complicate efficient parsing, then WM skills could be more heavily involved in the SVA, narrowing the margin for confounding effects. To this end, measures of the levels of syntactic formulaicity of the structures intervening between subject and verb could be used to infer whether chunking strategies were deployed.

Study 4 is therefore concerned with measuring the degree of morphosyntactic formulaicity of the English and Spanish SdV sentences extracted from the corpora of Study 3. Two steps were taken towards providing a reliable measurement of formulaicity. The first one involved representing these sentences visually through a Sankey diagram, as these highlight the flow and frequency of use of the components of a network. The rationale was that higher sentence connectivity would suggest higher levels of formulaicity, which in turn would imply that WM is less burdened because recall from LTM is possible. The second involved designing a novel measure of network connectivity that could be used internally as a reliable index of the formulaicity of the SdV sentences. This was then used in contrast to comparable non-SdV sentences to determine if SdV sentences exhibited more or less formulaic patterns and whether these could influence the involvement of WM in SVA.

1.6 Summary

This chapter has introduced some the complexities that permeate language learning predictions on the accounts of language being either a mental module barricaded from the rest of cognition, or an interconnected system reliant on domain-general mechanisms. Yet, many critical issues remain unanswered. For example, [a] are the
links between WM and language processing similarly observed in L1 and L2 learning? [b] are the interactions between L1 and WM considerably different between children and adults? [c] is WM training really training WM? Chapter 2 will elaborate on aspects of WM and syntax processing that are relevant to the three studies that steer this dissertation. While the approaches are methodologically diverse (i.e., experimental, corpus-based) in the array of materials used (n-back, GJTs, parser) and the cultural intricacies and technical challenges faced, there are theoretical commonalities among them. These are the focus of chapter 2. This means that [a] more attention can be devoted in chapters 3-6 to detailing the methodology and analysis of the results without having to revisit the same theoretical paradigms, and [b] the studies are not interpreted as independent, stand-alone research pieces but as part of the same enterprise.

As indicated above, chapters 3 to 6 focus on each of the research questions. A conceptual flowchart was devised (Figure 5) to offer a visual summary of the aim of each of the studies as a roadmap to the dissertation.

**Figure 5.** Visual conceptualisation of the dissertation, labelling each study and providing the driving question for each one. The arrows from 1 to 2 and 3 to 4 represent how the studies on reaction times and formulaicity derive from the cognitive transfer and the corpus study, respectively. The arrow from 1 to 3 emphasises the move from cross-sectional (1 and 2) to longitudinal (3 and 4) data.
Chapter 3 tests RQ1 through a randomised controlled study of the bidirectionality of ‘transfer of training’ skills between visuospatial WM and the syntactic processing of agreement of SdV sentences. Chapter 4 examines RQ2 by measuring variability in the reaction times to the n-back stimuli and in doing so provides a clearer depiction of children’s individual WM performance. In the original conception of the research project, RQ1 and RQ2 were meant to be followed up by a second intervention 6 months later, and this would have facilitated the longitudinal observation of the same speakers and the assessment of the longevity of the training effects. With the outbreak of the COVID-19 pandemic the second intervention was cancelled and a new route towards the procurement of additional, complementary data was devised, factoring in the travel and physical distancing restrictions. The new plan brought about RQ3 and RQ4. Chapter 5 addresses RQ3 through a longitudinal, corpus-based study that explores possible associations between the emergence of SdV sentences and WM development across English and Spanish. Chapter 6 seeks to answer RQ4 by creating a computational representation of the linguistic material that constitutes the distance in CS SdV sentences, accounting for chunking mechanisms. Finally, chapter 7 summarises and integrates the cross-sectional and longitudinal evidence before moving on to a general discussion of their implications for domain-general cognition and language processing.
Chapter 2 – Literature review

2.1 Chapter overview

Chapter 1 introduced the general theoretical context underpinning the research questions of this research project and introduced some of the critical connections between WM and morphosyntax. This chapter will start by introducing WM, and briefly address its architecture and functional properties (2.2). These structures and operations will be explored in more depth in 2.3, where some of the key milestones in the development of WM from birth into adolescence will be highlighted. This section also discusses theoretical issues concerning WM, especially with regards to arguments in favour of its domain specificity versus its domain generality (Miyake, 2001). The direct implications carried by a pro-generality versus a pro-specificity account for memory in morphosyntactic processing are at the heart of this dissertation and as such will be addressed here. Next (2.4), a critical review of the methodological apparatus used to assess WM will serve to emphasise the challenges involved in measuring the construct. Section 2.5 will reflect the potential and limitations of using some of these WM measures in training studies to explore cognitive transfer to untrained abilities. The proposed untrained skill, syntactic agreement, will be presented in section 2.6 and a case for its compatibility with a WM task will be made. Section 2.7 will address the differences in SVA processing in Spanish and English while section 2.8 will do the same for L1 and L2 SVA. Finally, the last section (2.9) will summarise the crucial theoretical and methodological issues reviewed in the preceding sections that will preface the studies that compose this research project.

2.2 Working memory

The original representation of working memory (WM) as a complex, multifunctional system only came into existence in the 1970s, when STM was viewed, in contrast to LTM, as a unitary system for the temporary storage of information (Baddeley & Hitch, 1974). WM embodied the immediacy of processing and storing information in situ that had been overlooked by the relationship between LTM and STM. Baddeley & Hitch’s main purpose was to delineate the underlying architecture of WM to show its structural distinctiveness from STM, which they did by describing a central
executive system ruling over two dependent subsystems, the visuospatial sketchpad and the phonological loop (Baddeley & Hitch, 1974). The central executive system is purported to be the overall, domain-general controller that manages attention, updates information, and taps into long-term memory. As the one domain-general component, its management of resources affects processes involved in higher level cognition. On the other hand, as part of the slave, domain-specific system, phonological working memory is associated with storing and processing all speech-based information, while the visuospatial sketchpad holds and manipulates visual and spatial information. The last addition to the slave system came about some decades later in the form of an episodic buffer; a necessary interface that allows information to be shared among these systems and bound together, as illustrated in Figure 6 (Baddeley, 2000). This interface applies both to the ‘fluid’ slave systems as well as to the ‘crystalised’ knowledge that integrates LTM. As the depiction in Figure 6 shows, the phonological loop is associated with the acquisition of language, more specifically vocabulary, in L1 and L2, while the visuospatial sketchpad is associated with visuospatial semantics. Through the buffer, the central executive is theorised to bind information from both sources into episodes or chunks which are more easily retrievable either immediately or in the longer term.

Figure 6. The multi-component working memory system by Baddeley (2000). The central executive at the top controls and binds the information from the slave system into episodes. The episodic buffer acts as an interface between the systems and LTM.

Once the ‘architecture’ of the model had been proposed, the main research focus shifted towards the functionality of WM which is often described as a complex
system where information is stored and processed, despite interference and
distraction (Miyake, 2001; Conway et al., 2005; Diamond, 2013). Cowan (1988) also
proposed a multi-component view but without a precise definition of the
components of such system. In the model, the basic subdivisions of working memory
were said to be activated memory with the focus of attention within it, and central
executive processes that manipulate stored information. Baddeley’s
(1986) phonological loop and visuospatial sketchpad would be subsumed by
activated memory, with both being susceptible to interference. Similarly, Baddeley’s
(2000) posterior addition of an episodic buffer is likened to Cowan’s focus of
attention within the system (2008). Even when some of the characterisations of WM
in the field may vary to some degree, all models agree on a system that orders, stores,
and revises immediate sensory details which are later to be integrated into general
cognition. The amount of data that can be stored temporarily is limited and the speed
by which it is activated varies.

It is important to note that although the focus here is solely on WM, because WM is
responsible for maintaining and updating the relevant information to accomplish a
specific cognitive task, its ‘performability’ is influenced, at least partially, by other
forms of executive functions (EF) such as attention and inhibition. The domain-
generality of the governing central executive system makes this possible. As observed
in the previous chapter, the efficiency in which high volumes of information are
orderly chunked prior to being stored relies heavily on the attentional abilities of the
individual. Analogously, if the WM system is capable of updating stored information
continuously based on incoming sensorial input, it must do so in a way that filters
out irrelevant environmental information, hence the essential role of inhibitory
control.

2.3 WM development

Children exhibit memory abilities from a very early age. In fact, evidence from pre-
natal studies shows that, even at 33 weeks of gestational age, foetuses seem to show
sensitivity to properties of maternal speech and native-language speech, which
suggests memory is functioning in some primitive way at least prior to children’s
birth (Kisilevsky et al., 2009). Experience with and sensitisation to language input
starts prenatally and continues at birth (e.g., Abboub et al., 2016). In addition, over the first year of life, before children are fully productive with language or can fully comprehend the communicative intentions of others, there are a range of categorization, memory, inhibition, analogy, attention, and social cognition skills also emerging in development, and these resources are available to the child at the time or before they are beginning to use language more productively. In their first months of age, infants have shown they can hold representations of objects mentally for a considerable length of time, for example, while an important line of evidence supporting early recognition memory (children aged 3 to 6 months) originates from child gaze studies (Caron et al., 1977; Fagan, 1974, 1990), Diamond (1995), went a step further in providing more compelling data demonstrating that 6-month-old children are able to remember previously recognised stimuli. In her study, children were habituated to a sample object and after delays as long as 10 minutes, they would not prefer to look at the sample object over a distractor, but they would reach for it. The results from both response modalities were highly comparable but, because in the delayed non-matching to sample (stimulus = reward) children are required to display more evident behaviour (reaching), presumably to a greater cognitive cost, this dispels the inherent doubts of interpreting infants’ gazes, especially because parents are often present and act as distractors. Infants between 5 and 10 months old have also shown they are capable of updating old stored information with new information in tasks such as A-not-B (Bell & Cuevas, 2015). In the A-not-B task, children observe how a desirable object is hidden in front of them, and after their gaze is interrupted with a distractor, they try to remember the location of the object. The hiding locations and the length of the delay of the response, marked by the distractor, are controlled. Therefore, longer delays suggest heavier WM loads, and greater inhibitory control, as children have to process more interfering information.

Over the first years of life, children’s memory duration increases substantially. This development is not, however, always linear and plenty of u-shaped trajectories have been reported (van Geert, 2008) during infancy. Past characterisations of the 0-3-year period used to portray infants’ memory as ‘fleeting’ and ‘primitive’, and usually use ‘infantile amnesia’, the inability of adults to remember episodic experiences from 0-3 years of age, and other extrapolations from findings in adults, as support for this claim (Nelson, 1994). While the apparent inability of older children and adults to
communicate their early life experiences has been documented extensively (Hayne, 2004; Rovee-Collier & Cuevas, 2009), the theories that propose that this forgetfulness is rooted in neurophysiological immaturity (Newcombe et al., 2007) have often underestimated the empirical evidence in support of associative learning in infants (Alberini & Travaglia, 2017; Mullally & Maguire, 2014). The latter line of evidence is problematic for a number of reasons, but fundamentally because it presumes that aging adults with amnesia and children share the same limited memory resources. This is meant to reflect an imperfect underlying structure that differs from the superior system shared by older children and healthy adults, which is fundamentally biased because it conceptualises children’s memory solely from adult experiences (Karmiloff-Smith, 2002).

From 4 years of age until adolescence, WM is always increasing. In the case of the phonological loop, rehearsal has been a well-documented source for WM gains in younger children (Gathercole et al., 1994, 2004; Hulme et al., 1984). Rehearsal is a form of maintenance of phonological information through articulation that is often subvocal, and children as young as 4 years of age have shown sensitivity to this, with strong correlations found between articulation rate and memory span (Gathercole et al., 1994). During the pre-school and early school years, spontaneous rehearsal has been recorded sporadically, which suggests the phonological storage of children of that age is limited to the speech rate they listen to, and the phonological loop acts mostly as storage. This changes from the age of 6-7 onwards, as children become more used to relying on rehearsal strategies for also recoding visual input into phonological information, which, in turn, leads to an improved recall ability. Reaching this developmental milestone means that providing reliable measures of the visuospatial sketchpad in isolation has proven to be problematic from the experimental point of view. Even when input is presented visually, and especially when it consists of everyday objects, children are able to recode it in their WM, aided by their phonological form and/or their semantic representation. ’Pure’ visuospatial WM tasks do exist and have been employed in the past (see Logie & Pearson, 1997, for a study contrasting two of them). Yet, while these tests have corroborated previous findings of increasingly consistent visuospatial WM gains in the 5-12 age period, they explain little about why this improvement occurs or whether these gains are a reflection of an improved capacity or improved strategies. Less ‘pure’
visuospatial tests, including visual recall tasks, on the other hand, have offered
detailed insights into what types of developmental changes occur during childhood
(see review in Pickering, 2001). Hitch et al. (1988), for example, tested visual
similarity effects and word length effects on an object recall task to compare the use
of both WM subsystems on children aged 5 to 10 years old. They reported three key
findings. First, younger children’s recall benefited from labelling the objects prior to
the task, as opposed to relying solely on visual information. Second, younger children
(5-year-olds) were affected by the visual similarities among the visual stimuli. Third,
older children were affected by the length of the name of the visual stimuli. The
consensus is that after the age of approximately 7 years, children resort to
phonological coding of visual information in their WMs, whenever possible, in recall
tasks (Gathercole et al., 2004). In other words, children under 7 tend to depend more
on visuospatial information when remembering objects, but an important
developmental shift that occurs around this point in time unlocks the power of
rehearsal to act across domains and in turn improve their WM recall abilities
considerably.

Lastly, as discussed in chapter 1, the DCL account advocates for an integrated view of
cognition which sees the development of WM not as an isolated system but
intertwined with other forms of EF. Therefore, understanding how memory evolves
presupposes understanding how other executive functions evolve. The
developmental changes in the central executive have been mostly measured via
memory span tasks with dual demands in terms of storage and processing. A well-
known example of central executive assessment in adults is the seminal study by
Daneman & Carpenter (1980) using the reading span paradigm. In their reading
span task, college students were asked to answer questions about several sentences
they had just read (processing demand) while simultaneously remembering the last
word of each sentence (storage demand). Because the functional properties of the
central executive are a reflection of how intertwined EF operate in domain-general
cognition, it is not surprising that there are fewer documented developmental
approaches using reliable WM measures than those conducted on adult participants.
As previously noted, in the case of research focused on WM’s subsystems, for many
years the convention was to describe children’s performance in central executive
tasks only when contrasted to adults’ performance, with little attention to the
developmental trajectory of the master system during childhood. However, more recent studies have pushed the boundaries of our understanding of the changes that the central executive experiences during childhood and their connection to neurophysiology (Luciana & Nelson, 1998), the effects of gender and age in individual variation (Pelegrina et al., 2015; Vuontela et al., 2003), and the developmental role of interference control (Schleepen & Jonkman, 2009). These different lines of evidence converge on the same conclusion that at approximately 7-8 years of age, children can cope with high attentional, inhibitory and memory demands, notwithstanding the lack of functional integrity among their EFs which is only achieved later, in adolescence.

Multiple assessments of the central executive, the phonological loop, and the visuospatial sketchpad (Alloway et al., 2004; Gathercole et al., 2004) have confirmed similar findings to the ones presented thus far, namely that, by the time children are 7 years old, the key structural components of the WM system and subsystems are already set in place.

Whilst impressive, this early developmental start and linear increase does not equal rapid WM attainment. Even when children can make associations to remember things that adults cannot learn, age differences in WM capacity and processing skills have been observed in children above 13 years of age up until early adulthood (Cowan et al., 2011; Luciana et al., 2005). For example, Cowan et al. observed continuous developmental gains in visual WM capacity in three different age groups, children 6-8, 11-13 years old, and college students. This contradicted their own previous findings where the differences between older children’ and adults’ WM capacity had been less significant (Cowan et al., 2010). While there is compelling evidence suggesting that in general children’s WM capacity expands during their childhood (see review in Cowan & Alloway, 2009), less is known about why and how it occurs. The challenges for Cowan et al. (2011) were first, examining whether adults outperforming children in visual recall tasks was indicative of a superior WM capacity or a result of their use of strategic rehearsal and other mnemonic techniques, and second, tracing how these age differences changed, if they did, over time. After presenting visual items at a slower rate and one at a time, which were two key methodological differences between this study and its predecessor, Cowan et al.
concluded that there is a substantial increase in visual WM capacity during school years, and that this gap is narrower, but still noteworthy, between older children and young adults.

The second study cited above, by Luciana et al. (2005), provided insightful data suggesting that the development of non-verbal WM processing skills in adolescents (9 to 20 years old) could not be interpreted monolithically, given the marked age differences found in different non-verbal WM tests. Their experiment consisted of four recall tasks, namely [1] nonverbal face recognition, [2] spatial delayed response, [3] spatial memory span, both forward and backward, and [4] spatial self-ordered search. These were hierarchically ordered in terms of multitasking demands, with [1] being the least complex, and [4] the most. The results for [1] showed no age differences among the age groups, suggesting that the ability to remember previously recognised human faces (visual stimuli) appears to have reached adult-like maturity by the time children are 9 years old. In the case of [2] and [3], the strictly spatial tasks, no developmental changes were reported after 11-12 years of age. As for [4], where spatial recall was used in a situation requiring problem-solving, there were significant changes up until the age of 16, after which participants’ performance stabilised. Altogether, the results from Luciana et al. (2005) suggest that children’s visuospatial WM develops in an orderly progression based on the task demands they face, with significant age effects being observed as late as 16 years of age.

This is somewhat consistent with other EF findings showing a ‘late’ critical period for age-related improvements in inhibitory control. Hale et al. (1997), for example, tested school-age children’s and young adults’ visuospatial and verbal WM using domain-specific and cross-modal interference and found that the latter tended to affect only children under 10 years old. That is, whilst engaged in a primary memory task, both children’s and young adults’ processing skills were disturbed by interference from the same domain (e.g., spatial task and spatial interference), but only children under 10 were highly susceptible to a secondary task from a different domain (e.g., spatial task versus verbal task and vice versa). This led Hale et al. to conclude that the central executive system, responsible for cross-modal interference-control, only reaches maturity later in childhood, perhaps somewhere between 8 and 10 years of age. However, differences between unimodal and cross-modal
interference-control have been established elsewhere for longer periods of time. For example, a recent study by Hirst et al. (2019) found age effects on children’s (6-11 years old) and adults’ processing of auditory distractors while they were focused on visual tasks and vice versa (Stroop test). More specifically, they reported that [a] children experience more stimulus-interference than adults and [b] children are more susceptible to cross-modal interference. The former finding adds to [an] existing literature reporting the same using other psycholinguistic paradigms (flanker test in Cragg, 2016). Evidence from neuropsychology and neuroimaging confirms that, because cross-modal interference when updating and maintaining information is controlled by the central executive, and this activity occurs primarily in the frontal lobes, not until children are late into their teenage years are they able to suppress interference like adults, as these are the last brain regions to fully form (Dempster, 1992; Fletcher & Henson, 2001).

It should be noted that none of the studies that integrate this research project are interested in characterising children’s WM processing abilities by contrasting them to those of adult populations; nor are they strictly concerned with the development of other forms of EF such as attention and inhibition. However, and as mentioned in the previous chapter, because the theoretical positioning of DCL is inextricably linked to cognitive linguistics, which often relies on adult data, and because the integrated view of cognition adopted here entails that WM does not operate independently from other domain-general functions, it is necessary to discuss WM development in relationship to the development of other closely related cognitive faculties and their adult end state. The expectation is that doing so, albeit somewhat tangentially, contributes to the realisation that the ontogenetic changes in children’s WM development, including individual variation, respond to a multitude of influences, from the maturation of the neurological system and the changes that ensue, especially the interaction of WM with other forms of EF, to the superficial characteristics and underlying apparatus of the measuring tasks. A recent review of theoretical and empirical evidence of WM development in infants by Cuevas & Sheya (2019) adopted a broader biopsychosocial perspective on the matter. They posit that because WM is essentially a developing biological system, and children move through different sociocultural contexts, the changes observed in their memory capacity and ability are paced by the ecological demands of the niche they occupy.
There is a terminological issue as well. The words ‘memory’ and ‘attention’ entered the lexicon centuries before scientists were able to understand what they entail within human cognition. Russ Poldrack, a psychologist and neuroscientist at Stanford University, highlighted this mismatch between terminology and psychological constructs in a recent interview to Quartz: “People can talk past another. If one person says I’m studying ‘working memory’ and the other people says ‘attention’, they can be finding things that are potentially highly relevant to one another but they’re talking past one another” (Goldhill, 2018). This echoes past concerns in the field (Duncan, 2010) and warns of how essentialist category labels could misrepresent brain functions (see comprehensive analysis across psychology and neuroscience in Brick et al., 2021).

The lessons hopefully learned thus far are that first, the developmental onset of children’s WM occurs exceptionally early, and while the system does not reach maturity or adult-like capacity until early adolescence, by the time children are 6-7, the fundamental underlying structure and function is there (see Figure 7 below for a graphic summary). Even when significant age effects have been documented beyond this point, and even when some respond to undisputed age-related gains, many of the perceived gaps are reflections of the differences in the ecological niches occupied by children, young adults, and older adults, and their demands, and not due to underlying deficits or a developmental discontinuance. Second, the on-line processing, storing, updating and retrieval of information is subject to effects from other forms of EF such as attention and inhibition, and this dictates how WM can be measured. Domain-specific tasks that assess phonological and visuospatial memory, as discussed above, can be tightened to control for minimal cross-modal interference. Tasks devised to measure the dominant central executive (i.e., complex span), however, need to be cautiously implemented because of cross-modal interference and the inherent cognitive demands on the child. It is this latter group that provides a gateway towards the potential interactions between domain-general cognition and syntactic processing theorised here.
2.4 Measuring WM

The previous section cautioned about the dangers of assuming that WM tests provide exclusive measurements of WM processes (Rovee-Collier & Cuevas, 2009). This is not surprising given that children’s WMs do not develop independently from the rest of cognition and are in fact conditioned by other forms of EF, and that ‘spillage’ across domains has been observed in domain-specific tasks. Add to this the methodological challenges intrinsic to measurements of online processing in situ and the overall message becomes clear, interpretations of the underlying processes in WM are constricted by the existing psycholinguistic measurements of WM products.

Online psycholinguistic methods to measure WM have the considerable advantage over traditional pen and paper tasks in that they attempt to measure the processing of tasks in real time, or at least closer to real time. Among many of the classic WM measurement paradigms is the span task, a complex test with storage and control demands which can be conducted forward or backward, using digits, words, locations, etc. There are many ways in which WM tasks are complex. For example, cross-modal binding tasks require participants to respond to visual and auditory stimuli. While the argument used to be that this information is stored and processed separately, more complex tasks have informed a more functional emphasis in the resources used by the WM system. Normally, complex span tasks require participants to perform an additional task in between the items in a serial recall, while simple span tasks only involve performing the serial recall, for example.
Until the 1990s, WM was typically and almost exclusively measured using dual tasks, that is tasks with processing and storing demands, as these were believed to be more accurate reflections of the construct, especially when contrasted to short-term memory. Some of the most important dual WM tasks of the last century include a combination of recall plus reading (Daneman & Carpenter, 1980), counting (Case et al., 1982), or arithmetic operation (Turner & Engle, 1989). A closer re-examination of span tasks, however, has shown that they are, in fact, reliable measures mostly of WM storage, an important predictor of individual variance. However, these tasks would be inappropriate for the transfer study discussed in chapter 3, as the cognitive mechanisms theorised to be involved in SVA relate to the processing and manipulation of continuous input, not just the storage of it. This implies that other forms of WM measures are needed to gain a better understanding of the system beyond its capacity, and, conveniently, other forms of complexity or duality in WM assessment beyond recall have gained similar status over the past three decades, one of which is the n-back task.

The n-back task was originally devised by Kirchner (1958) as a visuospatial task with four load factors (0-back to 3-back), but a further understanding of its cognitive demands has shown how the continuous recognition of new information while simultaneously storing it makes it a dual task. To borrow Fletcher & Henson’s (2001) definition, ‘n-back tasks require the manipulation and maintenance of the contents of WM’, and this makes it, at least at face value, an accurate assessment of the central executive. Procedurally, participants have to respond if the stimulus they see on a screen matches the stimulus back in the sequence. For example, in a 1-back task, 100% accuracy would mean pressing the button in P8 (−P7), using the forms depicted in Figure 8 as stimuli. In a 2-back task, true positives are instances where the child presses a button if the image was presented two positions before, so that would mean P3 (−P1), P6 (−P4) and P7 (−P5). If the same sequence were used in a 3-back task P8 (−P5) would be correct.

![Figure 8](image.png)

*Figure 8. Shape-based stimuli used to represent the sequential presentation of an n-back.*
The recurrent use of n-backs in neuroimaging studies throughout the 1990s certainly contributed to their current paradigmatical status as an effective measure of WM (Owen et al., 2005). This is not surprising given that n-backs require simple button-pressing responses, provide records of accuracy and reaction times, and their difficulty setting is easily adjustable in terms of memory load (0 to 3 factors) and stimulus latency (e.g., 500ms to 3000ms). Because of this flexibility, n-backs are well suited for studying children’s WM, not only from a neurological perspective (functional magnetic resonance imaging, fMRI, in Casey et al., 1995) but also behaviourally (Vuontela et al., 2003).

However, the concurrent validity of n-back tests with other traditional measures of WM has been called into question in the past, especially in studies with adult populations. There is neuropsychological evidence showing a lack of correlation between a modality of the n-back test and a digit span backward test in an experiment with adults with Parkinson’s disease (Miller et al., 2009). In their study, Miller et al. (2009) did acknowledge that one possible explanation why they failed to observe correlations between the two tasks may lie in the differences in presentation of stimuli, as the n-back used images of consonants and the span test digits presented orally. What is to Miller and colleagues a conjecture about within-domain versus between-domain differences has gained momentum in neuroimaging research. A quantitative meta-analysis by Owen et al. (2005) of 24 primary studies using multiple variants of n-back tasks on healthy subjects provided detailed neuroimaging evidence for consistent activation of frontal and parietal cortical regions, which have been traditionally linked to WM. But perhaps one of their most critical findings was that there were noticeable differences in subregional and laterised activation depending on the content and processing demands of the n-back. Therefore, this content-specific activation in WM registered in n-backs suggests that, even within the same domain, say using images of human faces and images of letters, there is discernible variation. And in the case of different domains, one of the prominent differences in regional activation reported by Owen et al. (2005) was in fact between n-backs using verbal and non-verbal stimuli.
Similarly, Kane et al. (2007) have scrutinised the construct validity of n-backs in a combined experimental-correlational study comparing them to WM span tasks. Because of the weak correlations found, they concluded that both tasks cannot be valid measures of the same construct, or in their view, that n-backs fail to demonstrate convergent validity with an established WM measure (i.e., WM span). There is one critical methodological decision in the study, that may explain, to some extent, this lack of correlation. It concerns the procedure followed in the implementation of the n-backs, specifically the reduced time to remember the stimulus (500ms). This becomes even more arduous with two or three factors, which were the only instances of n-backs used in the experiment. A caveat to their findings and this decision, partly influenced by the preliminary data of a study by Shelton et al. (2007) published shortly after, is that the differences between the n-back and complex span task were only observed in conditions that involved speeded recognition. Shelton et al. (2007) did establish positive correlations between an operation word span task and a lagged version of a recall n-back task that used words presented on a screen for 1000ms, which implies that Kane et al.’s (2007) experiment could potentially yield different results if the time allotted for each stimulus is adjusted.

The irregularity in the correlations established between n-backs and other measures of WM seems to be a direct reflection of the poor compatibility between the contents of the tasks. Therefore, it would be scientifically prudent to avoid equating interpretations based on different types of WM measures. In addition, the consistently high correlations between n-back tasks and other measures of domain-general cognition, such as fluid intelligence (Jaeggi, Buschkuehl, et al., 2010; Kane et al., 2007), may well suggest not that n-back tasks do not measure WM but that n-back tasks measure more than WM (cf. Conway et al., 2005). In other words, the n-back task captures processes which are crucial to WM performance, and, as discussed thus far, these include those that are ‘purely’ WM, such as storage and maintenance, as well as those that involve other forms of general cognition, such as interference resolution and inhibition (Kane et al., 2007; Oberauer, 2005). In the n-back, irrelevant items must be inhibited and abandoned from WM, and a counting and matching process between the upcoming and stored stimulus in WM is necessary to make a decision whether the stimuli are the same to initiate a correct
response (Rac-Lubashevsky & Kessler, 2016). The underlying cognitive process is one that enables active maintenance and regulation of a limited amount of task-relevant information. Participants are asked to hold some information in mind (the sequence of the stimuli) while simultaneously performing an additional task (matching the stimuli).

This raises the question whether one should conceive of the transfer proposed here as one of working memory or broader inhibitory control. The answer to that question very much depends on the theoretical position one takes of what working memory is and how domain specific it is. For example, the theoretical framework of Cowan (1988) places greater emphasis on the possibility of domain interference within working memory than does Baddeley (1986) and more generally there has been a continuing controversy about the extent to which modes (verbal, nonverbal) and components of the executive function (attention, working memory) interfere with one another (e.g., Cocchini et al., 2002; Fougnie & Marois, 2011). The approach adopted here was inspired by modern frameworks of sentence comprehension (often using sentences with distal subject-verb relationships such as the ones used here) that claim that language processing, while it might operate on specialised representations, is nevertheless subject to general processing principles and constraints that govern other domains of memory (Marcus, 2006; Lewis, Vasishth & Van Dyke 2006).

Because this dissertation aims to explore possible causal links between WM and morphosyntactic processing, which will be covered later, the next section explores the potential and limits of the n-back task beyond correlational studies, not as a WM measuring tool, but as a WM training one.

2.5 Training WM

WM training has received enormous attention within and outside the scientific community over the past two decades, with arguably equal amounts of approval and criticism. Cognitive training packages that advertise different WM training plans aimed at school-age children (+6 years old) and sometimes adults have been developed by companies such as Cogmed (for a review see Shinaver et al., 2014) and
Cognifit (for a review see Shah et al., 2017) and achieved considerable commercial success. This is not surprising considering many of the claims as to what WM training can do, including a reduction of symptoms of neurodevelopmental disorders (Gathercole & Alloway, 2006), improved academic performance (Bergman Nutley & Söderqvist, 2017), and gains in general intelligence (von Bastian & Oberauer, 2014) in children and adults.

The popularity of the n-back as a WM measure has led to it being one of the most commonly used computerized experimental WM training paradigms, and many modalities of the n-back have been used in the commercial cognitive training products cited above. Its paradigmatical status has not escaped criticism and to this day still attracts plenty of attention from the community. This is best exemplified perhaps by the fact that there are numerous well-known critical meta-analyses of n-back training efficacy conducted in this last decade, some of which include those by Melby-Lervåg & Hulme (2013), Au et al. (2015), Melby-Lervåg et al. (2016), Soveri et al. (2017) and Gathercole et al. (2019).

The study by Au et al. (2015) focused on the training effects of n-backs on fluid intelligence (Gf). The sample contained 20 studies training adults (18-50 years of age), using active and passive control groups, single and dual n-back tasks, and verbal and visuospatial transfer tasks, following a pre-test – post-test design. Upon review of all 98 Gf outcome measures, the authors reported a small but significant positive effect on Gf gains after just weeks of n-back training. It is worth noting that, even if the effect sizes recorded by Au et al. (2015) were not large, these were still Gf gains, equivalent to 2-3 points in an intelligence scale, among adults at their cognitive peak, which is encouraging for demographics with more room for improvements in cognitive performance. A methodological observation made was that, while most of the studies sampled had a passive control group, meaning the participants did not partake in any activity, the size of the transfer effect was not different in studies with active control groups, where participants carried out an unrelated task during the training period.

Melby-Lervåg & Hulme, (2016) scrutinised this finding and argued that, even if there was no statistical significance between the types of control listed by Au et al. (2015),
only studies with active controls can provide convincing evidence of cognitive training effects. This is because studies have shown that participants in active control groups benefit from increased motivation and beliefs that they have improved after training (Mohr et al., 2009) and these effects are overlooked and conflated with training effects in passive control trials. In their own previous meta-analysis, Melby-Lervåg & Hulme (2013) noted the same, that short-term WM training effects on children’s and adults’ performance on verbal and non-verbal WM tasks decreased considerably when they adjusted the inclusion criteria for randomisation and active controls. Three years later, in a new meta-analysis with 145 experimental comparisons Melby-Lervåg et al. (2016:524) reiterated that ‘only studies with treated control groups can provide convincing evidence of specific benefits from working memory training’. However, they acknowledged that, apart from the characteristics of the methodological design, the type of training was also a possible moderator of the results, and in fact highlighted a significant transfer of n-back training to non-verbal ability in their review.

The fourth study is a multi-level meta-analysis that subsumes the samples cited by the other three. In their research, Soveri et al. (2017) investigated the transfer to untrained tasks in four variables [1] n-back task-specific, [2] other WM measures, [3] Gf, and [4] cognitive control. While concluding that all three previous meta-analyses had overestimated the transfer effects of n-back tests on WM by including untrained variants, their review of 203 effect sizes from 33 n-back training studies showed that a significant part of transfer post-n-back-training was specific to the task, and that only small effects were observed in other WM tasks, Gf, and cognitive control. Moderators such as age, training dose and training contents showed no significant effects. What these four meta-analyses show is that, while n-back training has shown positive effects of transfer to other tasks, these tend to be stronger if the untrained task is related in content, and that overestimations of transfer effects are common when the experimental design is not stringent enough.

This summarises the views of a process-specific theory of transfer, namely that transfer occurs when the underlying cognitive processes and the contents of the trained-untrained task pair are shared. Other theories of transfer (Klingberg, 2010; Takeuchi & Kawashima, 2012) have attributed training gains to the cortical plasticity
of the underlying neural system linked to WM but have failed to clarify the inconsistencies in transfer reported by many of the studies cited above.

Following a skill-based approach to transfer, Gathercole et al. (2019), in contrast, has argued that sharing processes is not enough for transfer to be established between a WM trained task and an unrelated untrained task with similar WM demands. According to their framework, for training-induced transfer to occur, a new cognitive routine has to be created between a trained and an untrained task that share a common structure. A cognitive routine is understood as ‘a structured specification of the coordinated sequence of processes that must be implemented to accomplish a mental activity’ (Gathercole et al., 2019:21). Because Gathercole’s model of transfer calls for a new sequence of existing cognitive processes, if the cognitive routine triggered by a specific WM task has already been established, there is little chance for training products to transfer to an untrained task even if they share features. To assess these two assumptions (routine and shared structure), they conducted a meta-analysis of randomised controlled trials reporting transfer between WM tasks. Situations testing for stimulus input modality, stimulus domain, stimulus category, recall modality, and recall paradigm were matched and unmatched for task pairs and later compared. The results show that transfer effects were more prominent if the trained and untrained tasks were part of the same paradigm and the same domain, and that input and output modalities had no significant impact on transfer. The WM paradigms that were shown to require new routines between trained and untrained tasks and therefore lead to transfer were visuospatial serial recall, complex span, backward span, n-back and running span.

The conclusion that can be drawn from this is that cognitive transfer requires similar WM demands from unfamiliar pairs of trained-untrained tasks. To guarantee a good pair match, the novel routines and subroutines created to perform both tasks need to be common to the systems involved. If the assumption is that routines are hierarchically organised, a subroutine, responsible for encoding, maintaining, or retrieving information, for example, is a good place to start. For instance, chunking is a form of information codification in WM that facilitates retrieval commonly used in visuospatial processing and language. Therefore, if two unfamiliar tasks are believed to recruit chunking and other existing strategies in a novel sequence, the prediction
is that transfer of training will occur. This may explain the irregularities in the findings of the meta-analyses reported above as the criteria established for strong/weak transfer were unable to account for all effects reported.

It seems to be the case that exploring n-back training effects on an unrelated and untrained language task that also demands maintaining information while simultaneously processing distractors could be a viable way to explore the types of connections between WM and language processing that are the subject of this research. Training participants in one WM task followed by testing in an untrained task, say, a linguistic one, enables researchers to explore causal links between the two domains, thereby pushing the boundaries beyond the type of evidence that can be obtained from correlational studies.

### 2.6 Agreement

A particular aspect of language that is thought to especially tax working memory, is agreement. Broadly speaking, agreement refers to a grammatical system where words or morphemes must interrelate their form with one another, for example, in person (e.g., ‘I am’ vs ‘he am’), number (e.g., ‘they is happy’ vs ‘they are happy’) gender (el\_masculine gato\_masculine vs. la\_feminine gato\_masculine) or case (e.g., ‘He is kissing her’ vs ‘him is kissing she’). This type of grammatical relationship is an obligatory part of many of the world’s languages and thus successful acquisition represents an important developmental milestone.

In strictly morphological terms, agreement traditionally referred to the realisation of morphological features on different lexical tokens. In their comprehensive work of English grammar *A Grammar of Contemporary English and A Comprehensive Grammar of the English Language*, Quirk and colleagues (1972) provided clear definitions for ‘concord’, the grammatical relationships referred to here as agreement, as the link between two units that resides in their compatible features (such as singularity/plurality in English, gender and number in Spanish). In English, this grammatical concord most relevantly applies to singular subjects requiring singular verbs and plural subjects requiring plural verbs. However, this agreement is often affected by interference of ‘proximal’ units. The principle of proximity, also
termed ‘attraction’, leads to overriding the agreement between the head of the subject and the main verb by one between the verb and a closely preceding noun or noun phrase as in [17], where *are* agrees with the noun group *English*... but not with the subject *knowledge*.

17- *A good knowledge of English, Russian and French are required for this position*

(Taken from Quirk et al., 1972, p. 757).

Quirk et al. (1972) noticed that the longer the distance between the head of the subject phrase and the verb the more likely grammatical concord is subject to interference from nouns that are ‘local’ to the verb. This suggests that, in cases where the subject and verb are discontinuous, such as when the subject is followed by a modifying expression, ‘mental energy’ is required to keep track of the source until the target becomes available (Franck et al., 2002). Franck and colleagues refer to this disagreement between subject and verb as ‘the linear distance hypothesis’, provided by the presence of a local noun with a similar number than the subject noun. This position views linear distance between words and working memory as determining factors when establishing agreement.

However, several studies comparing agreement in languages with relatively rich and less rich morphologies have shown how semantic interference and notional number may affect the verb marking. In Dutch, for example, Vigliocco and colleagues (1996) have compared notionally singular subjects noun phrases to notionally plural ones (e.g., *the cage with the gorillas* vs. *the signature on the checks*) and observed higher rates of agreement attraction for notionally plural subjects. They argue that languages with rich morphology, such as Spanish (Vigliocco, Butterworth & Garrett, 1996), tend to be more susceptible to the influence of notional number in SVA. An alternative view is provided by Foote & Bock (2012) who argued, in a study comparing subject-verb number agreement in Mexican and Dominican Spanish, that the opposite is in fact the case. That is, languages with richer number morphologies are less affected by number meaning in agreement production. In their study, Dominican Spanish was seen as the Spanish variety with a reduced morphological expression than the Mexican counterpart. The morphological specifications in a rich inflectional system, they state, can minimise the impact of notional marking in
number computation. Although they are diametrically opposed, both approaches factor in effects of notional number on agreement.

Another approach to explaining number agreement and attraction derives from Minimalist syntax, as formalised by Chomsky (1995). This is rooted in the supposed dominance of structural constraints in the production and comprehension of agreement, which is largely governed by the operation ‘AGREE’ (Chomsky, 2000). ‘AGREE’ is purported to move features from one element of syntax to another, although the actual elements involved, the type of movement, and other issues about its formulation remain contested even within the minimalist literature (Smith, Mursell & Hartmann, 2020). One of these issues relate to attraction effects in SVA, where hierarchical distance is seen to exert more influence than linear distance. One interpretation of this is the hierarchical feature-passing hypothesis, by Franck and colleagues (2002), which posits that agreement features are passed up along the syntactic tree, from the subject noun phrase to the verb phrase. They tested this in an experiment and found that plural forms that were structurally closer to the verb had a stronger attraction effect than plurals linearly closer to the verb. Upon closer examination, however, the attractors used in the experiment were different in other ways besides plurality (e.g., pronouns, noun phrases), therefore the attraction effects are not necessarily linked to derivational structures. Furthermore, there is no clear description of where the features come from (i.e., how notional, and lexical sources inform number agreement) or the exact information that dissimilar structures (nouns, pronouns, verbs) contain (i.e., meaning or grammatical features). Others (e.g., Gillespie & Pearlmutter, 2013) have downplayed the importance of hierarchical structures in the production of agreement after observing no differences in the attraction effects of structurally different preambles (i.e., prepositional phrases and relative clauses) on SVA.

One final psycholinguistic account of agreement that is particularly important to discuss in the context of this dissertation is one which relies on memory retrieval (Badecker & Kuminiak, 2007; Lewis & Vasishth, 2005). It favours a backward retrieval process, in which the verb copies or verifies features in the preceding subject, so the heavier the WM burden in the copying/verifying process the more agreement errors that ensue. In Slovak, for example, SVA relies on the retrieval of an
agreement source from ‘content-addressable’ WM, whether retrieval cues are of lexical or phrasal nature (Badecker & Kuminiak, 2007). Therefore, agreement in a sentence such as *The text on the label of the bottles of whisky is blurry* involves retrieving (back movement) the subject when processing the verb, which in this case is challenging because of the distance between verb and subject and the interference of lexical and phrasal cues (e.g., *label, bottles of whisky*) that share relevant morphosyntactic features. One specific way in which WM-retrieval theories depart from previous theories of agreement is by how short and long agreement are conceived. All relations established between subject and verb in SVA are deemed as novel, and therefore, regardless of the distance, memory of the past is necessary to realise this relation. This might be more conspicuous in long agreement than short agreement because of the memory burden, but the principle behind it is fundamentally the same. This is because all computations where new information is processed over time involve interactions with past information (Elman, 1990). This is not unique to language.

In Lewis & Vasishth (2005)’s computational model, the two key factors identified in the role of WM in sentence processing are fluctuating activation and similarity-based interference. For SVA, this translates as chunks intervening in agreement have different ‘values’ of activation based on the user’s familiarity with these chunks. These values tend to change over time (i.e., decay) just as humans forget. More recent studies (e.g., Franck & Wagers, 2020) have evidenced the active role of WM in dealing with agreement interference, as more easily retrievable structures have been found to generate more attraction. One way in which this WM-based model of sentence processing comes to terms with previous theories of agreement is in that the nature of the cues that generate associative interference to SVA can be morphosyntactic or semantic. Accordingly, these potential sources of interference (e.g., semantic plausibility, clitics in Spanish), were controlled for in the design on the sentences in Study 1 (see *Materials* under section 3.3). The key message here, however, is that this view of the processing of SVA is derived from the application of general cognitive principles, not language-specific ones, and it is therefore in better alignment with the theoretical underpinnings of this dissertation (i.e., DCL). Further arguments discussing the role of WM in SVA will be covered in more detail in the next section.
Regarding the broader developmental picture, some aspects of agreement comprehension are not mastered by 5-6 years-of-age (see Johnson et al., 2005 for verbal –s in English and Pérez-Leroux, 2005 for third 3rd person S–V agreement in Dominican Spanish). The precise age of acquisition appears to be very dependent on contextual factors – including which subcomponent of agreement is being tested and exactly how it measured. One important aspect influencing the developmental trajectory of agreement is the way agreement is realised through the unique morpho-phonological conventions of different languages. For example, the /z/ liaison consonant in French is a relatively more salient and reliable cue to plural morphology when compared with the English –s or, to a lesser extent, the Spanish – n. Legendre and colleagues (2014) argue these differences account for the fact French comprehension of this element of agreement is acquired earlier than English or Spanish and more generally, that because a broad range of factors affect acquisition, including saliency and cue reliability, a universal path of acquisition for agreement is unlikely.

In terms of its learning trajectory, some studies have reported successful production before comprehension, reversing the normal age-of-acquisition asymmetry. Johnson, de Villiers, and Seymour (2005) observed that English-speaking children were able to use verbal person/number even without comprehending their usage. Lack of comprehension, in the context of this experiment, means that children aged 3 and 4 showed no sensitivity to -s in 3rd person verbs (sleeps vs. sleep), even when they differentiate in their usage (Brown, 1973) whereas older children, aged 5-6, did show sensitivity. This approach to number agreement was reproduced in Dominican Spanish, a language where agreement is morphologically rich, using comparable sentences and the same results were observed, namely that not until children are at least 5 years old are they able to discriminate accurately between singular and plural number in SVA (Pérez-Leroux, 2005). This perceived delay in verbal agreement (number) in English and Spanish, are theorized to respond to aspects of either morphology, syntax, or semantics. Pérez-Leroux (2005) believes that this gap between production and comprehension of number agreement lies somewhere between the second and latter. What children acquire over time, according to them, is the ability to distinguish between ‘instances’ and ‘kinds’, so a child could learn
these ‘surface’ syntactic structures without understanding their semantic ‘depth’. By this account, the ‘complete’ acquisition of the plural marking cannot occur until all aspects of semantic distribution are understood. Others (Hsin et al, 2021) attribute SVA to a mental representation of ‘singularity’ and ‘plurality’ that is activated whenever a child processes a specific morpheme, which contradicts item-based theories of acquisition (e.g., Tomasello, 2013). This is because usage-based approaches might predict singularity/plurality is acquired on a lexical basis, influenced by the frequency and semantic properties of each word, rather than on a top-down, across-the-board basis.

On the one hand, under closer inspection, when a wider range of factors are considered, including testing non-canonical word orders, the frequency of forms in the input, spontaneous speech versus elicited production, task-specific/stimulus-specific features, and different languages, the gap in production and comprehension of agreement may be much less than previously thought (see Barrière et al, 2016 for French; Hsin et al, 2021 for Mexican Spanish). As for the question of the source of number problems (morphological conflation, syntactic feature-checking, or semantic number neutralisation), a number of studies, many of which were discussed above, have found an agreement error asymmetry, such that plural local nouns (e.g., ‘the key to the cabinets are’) cause more agreement errors than do singular local nouns (e.g., ‘the keys to the cabinet is’) (e.g., Bock & Miller, 1991; Hartsuiker & Barkhuysen, 2006).

2.7 WM and SVA

This section begins to integrate the domain-general account of WM detailed thus far with language, in the spirit of the DCL research questions of this thesis. The aim, as mentioned earlier, is the syntactic processing of grammatical agreement and the extent to which WM is involved.

WM abilities have been identified as a source for individual variation in other forms of syntactic processing. King & Just (1991) famously established that sentences such as *The reporter that the senator attacked admitted the error* are more difficult for
readers with low WM to read and understand. Fiebach et al. (2001) conducted an event-related brain potential (ERP) study and a fMRI study investigating the processing of German wh-questions and concluded that a memory component was crucial for establishing filler-gap dependencies. The German wh-fillers used, unlike those in English, were case-marked and therefore participants were morphologically aided to differentiate between sentences 18 and 19 below which carry different memory demands. Similarly, Bentea & Marinis (2021) have observed that Romanian-speaking children find more difficult to comprehend which-questions with long dependencies than who-questions with long dependencies.

1- Thomas fragt sich, wer am Mittwoch den Doktor verständigt hat.
  =(Thomas asks himself, who (NOM) on Wednesday the (ACC) doctor called has)

2- Thomas fragt sich, wer am Mittwoch nachmittag nach dem Unfall den Doktor verständigt hat.
  =(Thomas asks himself, who (NOM) on Wednesday afternoon after the accident the (ACC) doctor called has)

Superficially, it would seem that, if a WM-syntax transfer were to be potentially established, this type of syntactic structure would make for a good candidate. Yet, based on the morphosyntactic differences observed with English and Spanish, it would not be unreasonable to first define exactly the form of syntax. Throughout this chapter and the previous one, the terminology used to refer to the specific syntactic structure that pertains to this research has fluctuated from sentences with LDD to distal subject-verb relationships, to suspended agreement, etc. In the case of LDD, also called filler-gap dependency, the relationship between who and booked as an object in sentence 20 and who and kicked in sentence 21 reflects a long-distance dependency.

3- The player who the referee booked is on my team (Object relative clause)
4- The player who kicked the referee is on my team (Subject relative clause)
5- The player with the yellow card is on my team
6- The player under the stairs is on my team

While strictly speaking, sentences 22 and 23 do not reflect LDD properties, SVA is suspended in all four cases as the subject needs to be temporarily maintained in WM

---

2 Translation provided by the authors in the original paper, see Fiebach et al. (2001).
before agreement with the verb can be established while the information in between 
is simultaneously processed. Labelling all four sentences as SdV sentences serves to 
underline the memory and processing burden of the distance between subject and 
verb on the WM system. Using a broader classification that subsumes centre-
embedding instances, such as LDD and other less defined syntactic structures, 
increases the possible points of contact between WM and morphosyntax processing, 
which in turn facilitates the understanding of the involvement of WM in suspended 
SVA.

In an error analysis study, Fayol et al. (1994) conducted three experiments to test the 
hypothesis that cognitive overload improves the probabilities of subject-verb 
agreement mistakes in French. To this end, adults were asked to write down the 
sentences they heard. They were also given a concurrent task to contrast isolated 
sentence situations. The sentences were of the types Noun 1 of Noun 2 Verb such as 
[24] and Pronoun 1 Pronoun 2 Verb [25].

7-  *Le chien des voisins arrive*  
*(The neighbours’ dog is arriving)*  
8-  *Il les aime*  
*(He likes them)*

Participants made errors when performing two concurrent tasks yet almost no errors 
when recalling isolated sentences. As expected, errors occurred when N1 (Pr1) and 
N1 (Pr2) mismatched in number. Erroneous agreement with a local noun was more 
likely when the participant was required to remember four unrelated words. This led 
Fayol *et al.* (1994) to conclude that subject–verb agreement is computed 
automatically from the closest preceding noun to the verb. Rapid (automatic) 
agreement occurs when the distance between subject noun and verb is minimal, but 
when there is a preverbal local noun, activation will be affected by this too. In order 
to guarantee correct agreement with the subject noun, a ‘non-automatic checking 
mechanism’ is assumed to be activated, which taxes the WM system. A concurrent 
memory load task would therefore reduce WM resources available for the ‘checking

---

3 Translation provided by the authors in the original paper, see Fayol *et al.* (1994).
mechanism’, and consequently increase the probability of agreement of the verb with a local noun.

This study argues that the cognitive burden observed by King & Just (1991) in sentences with wh-fillers is also present in the processing of agreement in SdV sentences, where the elements intervening between subject and verb tax the learners’ memory by forcing them to keep them active when establishing agreement. Furthermore, even though SVA is a thoroughly investigated grammatical process, developmental variation in agreement has been depicted mostly through a linguistic lens, and the intervention of WM underestimated. This reflects the influence of recursion on several aspects of syntactic processing and grammatical acquisition, given that explanations rooted only in language-specific processes are deemed as all-encompassing nets for linguistic and non-linguistic phenomena.

Explanations concerning the resolution of long-distance dependencies have started to shift within computational linguistics, for example, in more recent years, with the attention moving from language-based to more cognitive-based accounts of locality effects, which are those associated with the processing of the gap. Both dependency locality theory (Gibson, 2000; Gibson et al., 2005), which was specifically conceived to address sentence comprehension, and the activation-based model (Lewis et al., 2006; Lewis & Vasishth, 2005), which is part of a general cognitive model, agree that for SVO syntax increasing the distance between subject and verb hinders agreement (for antilocality effects in SOV syntax see Vasishth & Lewis, 2006). These modern frameworks of sentence comprehension have often focused on the types of long-distance dependencies that will be studied in this dissertation (see a critical review of the predictions made by both models in Nicenboim et al., 2015). They have gone to the extent of making very specific predictions of the WM processes involved in the resolution of long dependencies in cases such as relative clauses (Bartek et al., 2011; Grodner & Gibson, 2005), which are a subgroup of the SdV sentences explored here.

The prediction of these recent theoretical frameworks of WM retrieval and cue-based parsing (e.g., Lewis et al., 2006; Lewis & Vasishth, 2005), is that WM is recruited for successful number agreement between subject and verb precisely when intervening lexical material forces long-distance agreement. For example, in sentence 26, the
plurality of the subject *tareas* needs to be held in mind while simultaneously processing the intervening prepositional phrases before agreeing in number with the downstream *son*.

9- Las *tareas* de la clase de Lengua Española de primer grado *son* para mañana

Encoding  Storage Interval  Retrieval

(The assignments for the first-grade Spanish class are for tomorrow)

The approach adopted here is somewhat inspired by these state-of-the-art models in that it predicts that language processing, and in this case the processing of long-distance agreement, is subject to general processing principles and constraints that govern other domains of memory (Marcus, 2006).

### 2.8 SVA in Spanish and English

As indicated in the previous section, the acquisition of SVA in number has been reported to occur sometime around the 5-year mark, even in languages with different morphosyntax, such as English (Johnson et al., 2005) and Spanish (Pérez-Leroux, 2005). Spanish is notably richer than English in its morphosyntax, and this has been shown to impact how SVA dependencies are processed (Acuña-Fariña, 2018). There are significant aspects of SVA that differ in English and Spanish. Vigliocco et al. (1996), for example, provide a detailed account of studies exploring notional SVA errors, including their own experiments with native-speakers of Spanish and native-speakers of English, seeking to assess the role of semantic information of the head of the subject in SVA errors. Notional concord imposes the notion of number over the actual grammatical marker on the subject, leading to different uses in what we call collective nouns, in sentences such as *The government has/have broken all its/their promises*. In one of a series of four tests, they used English and Spanish sentence preambles or compound subjects such as:

10- Juan junto con María [va/van a la playa]
11- Juan together with Mary goes/go to the beach
Their results showed that English speakers produced 5% plural verbs, while Spanish speakers produced about 60% (Vigliocco et al., 1996), suggesting that number marking in singular is favoured in English and number marking in plural is favoured in Spanish. The morphosyntactic differences between both languages seem to suggest that SVA hinges on more than linguistic resources. These non-linguistic resources, as theorised here, are synonymous of WM in action.

GJTs and acceptability tasks have been used extensively to assess children’s grammatical ‘correctness’. McDonald (2008) has managed to map an order of construction difficulty for children performing GJTs. This apparent order of complexity differs depending on the language, as one would expect, given that crucial aspects of morphosyntax, such as rich/poor inflection, alter the way verbs are conjugated or grammatical agreement is established. For example, intra-phrasal word order errors are easily detected by English-speaking children as opposed to French-speaking children, which is often attributed to the freer word order of the latter compared to the former (McDonald, 2008; Kail, 2004). In the case of verbal agreement, third person singular SVA has been documented in studies using GJTs (Wulfeck, 1993; Kail, 2004; Wulfeck et al., 2004) as being problematic for English-speaking children to learn, even more so than other forms of agreement. This difficulty is almost self-evident in Spanish, a language with a rich inflectional system, where grammatical concord is unique to at least six subject-verb pairs.

2.9 SVA in L1 and L2

A final angle to be explored is whether the processing of agreement in SdV sentences taxes WM differently in a first language than in a second or foreign language. The body of research linking bilingualism and executive functions is extensive (Bialystok et al., 2005; Hernández et al., 2010; Kroll & Bialystok, 2013; Morales et al., 2013). A similar case is that of WM, with numerous studies suggesting links to L2 learning in children’s and adults’ reading (Juffs, 2004; Leeser, 2007), writing (Adams & Guillot, 2008), vocabulary (Papagno & Vallar, 1995), and overall proficiency (van den Noort et al., 2006). Learning and using a second language requires a formidable amount of cognitive energy and resources, with inhibition and attentional control perhaps being some of the most widely studied cases because the non-target language interferes in
the processing of the target language. In the case of WM, several key aspects of its involvement in syntactic processing have been reported in L2 studies. For example, after reviewing a number of studies suggesting links between WM and different forms of linguistic processing, Juffs & Harrington, (2011) found positive effects that some WM measures (reading span) could account for individual variation in L2 syntactic processing. As was the case in L1 studies, the authors also advised future researchers to improve the levels of compatibility between the WM task and the linguistic task for optimal results.

A different way of approaching WM research in L2 has been through correlational studies comparing performance in L1 and L2 tasks. Both Juffs (2004) and Osaka & Osaka (1992) found correlations between L1 and L2 using reading span tests, but once again, the variability in the associations is too statistically significant for a consensus to be reached. Even with similar language pairs and target groups (English and Japanese), Harrington & Sawyer’s (1992) and Osaka & Osaka’s (1992) findings contrasted each other, although other factors such as L2 proficiency could have been responsible for this variation.

Other studies comparing SVA between L1 monolinguals and L2 learners based on participants’ judgments of grammatical/ungrammatical sentences (Sagarra & Herschensohn, 2010, 2013) have established correlations between WM ability and the accurate processing of verbal agreement in Spanish as an L1 and L2. However, it could be argued that there is strategic processing that occurs in GJTs outside of the underlying processing mechanism that is supposed to be measured.

While other online processing tasks such as self-paced reading (SPRT) and word-monitoring task (WMT) have proven to be more reliable measures of processing they violate one of the guiding principles of the transfer study proposed in Study 1, namely the structural compatibility between the trained and the untrained task. The SPRT and WMT share no superficial features with the n-back task and therefore transfer is not predicted to occur. In WMTs, for example, participants first hear a target word (e.g., *cakes*, taken from Chondrogianni & Marinis, 2012; Marinis, 2012) or see a picture representing that word (*cakes*). They then listen to a leading sentence introducing the context, followed by an experimental sentence containing the target
word, as in *Mary really likes to bake. Every day she bakes cakes and sometimes cookies and muffins*. They proceed to pressing a response button as soon as they hear the target word, therefore registering a reaction time. The main advantage of this word monitoring is tapping into online processing, but while it offers a medium for assessing working memory, the task is structurally incompatible with the n-back task if the interest is maximising transfer opportunities (full discussion in Chapter 3), and perhaps more suitable for studies interested in semantic priming, for example (i.e., presence of *bake* in the leading sentence).

Additionally, there is a long tradition of using GJT as measures of grammar comprehension in second language acquisition studies. The validity of the construct has not escaped criticism in the field (see confirmatory factor analysis in Vafaee et al., 2017), especially with regards to timed and untimed modalities of the task and whether they measure implicit or explicit knowledge, a subfield in multilingualism with a tradition of scholarship, receiving much more interest in the past decades due to the advances in terms of accessibility to cognitive processing. That said, there are two main reasons why GJT's still continue to be employed in psycholinguist research, both in L1 and studies of multilingualism. First, in GJT’s there is no need to infer comprehension from production (e.g., elicitation tasks). The task has low performance demands, which might not have been a requirement for testing 7-year-old monolinguals’ grammar, but without which measures of low proficient L2 learners are more difficult to assess. Second, their implementation is relatively simple, especially with large samples of participants, and this makes an experiment testing multiple variables such as the one discussed here logistically feasible.

In summary, the conceptualisation of the SdV sentences discussed here has been motivated by the clear links between WM and some forms of syntactic structures (centre-embedding) that are part of this new classification proposed. These sentences seem to share the same basic cognitive underpinnings in both the L1 and L2 from a developmental perspective. That being said, there are additional variables, ranging from the influence of the L1 on L2 and vice versa, to the level of proficiency of the L2, that need to be controlled for before any comparisons of WM involvement in L1 vs. L2 syntactic processing are discussed.
2.10 Summary

This chapter has covered the key theoretical and methodological issues concerning children’s WM and morphosyntax that will inform the studies discussed hereafter. In this way, it serves as the theoretical referential point for this dissertation, and, to that end, a summary of the critical points reviewed so far is included below.

- The construct of WM, as originally conceived and used in neuroscience and psychology is that of a dynamic system that stores and manipulates the information needed to execute a task, with special emphasis on its structural composition and functional overlap with other forms of domain-general cognition.

- A thorough examination of evidence emerging from multiple lines of research suggests that [1] by the time children are 7 years old, based on key developmental milestones reached in the 6-to-8 period, the WM system is appropriately formed and operating, but [2] further neurophysiological changes that occur in the following years seem to indicate that adult-like performance in complex WM activities is not reached until adolescence.

- The n-back task is a viable way of gauging domain-general cognition via the central executive of WM, as evidenced by results showing correlations with other tasks measuring fluid intelligence, interference resolution and inhibition (Jaeggi et al., 2010, Kane et al., 2007; Oberauer, 2005).

- Theories of cognitive transfers suggest that [1] transfer is more likely to occur if the trained and untrained tasks are unfamiliar, but they recruit similar specific processing mechanisms, and [2] training in the n-back task has led to positive transfer in an untrained task.

- SdV is proposed as a grouping criterion for syntactic structures with suspended SVA where WM is speculated to intervene.
The underlying memory factors that support processing agreement in SdV sentences seem to operate equally in L1 and L2, yet there are L2-specific factors that may account for individual variation.

The four studies that follow all seek to explore the interplay between children’s WM and the morphosyntactic processing of SdV sentences, but they adopt both cross-sectional and longitudinal perspectives, focus on different developmental periods, and use different measuring apparatuses. The literature reviewed above addressed the theoretical underpinnings that are common to these studies. Those theoretical and methodological aspects that are specific to each of these studies will be covered in full depth in their respective chapter.
Chapter 3 – Study 1: experimental study

3.1 Chapter overview

This chapter discusses in depth the first of four studies that explore the nature of the relationship between domain-general cognition and syntax processing. This first study involves a pre-post-test randomised control trial in which 104 native Cuban Spanish-speaking children aged 7 were assigned to one of four groups: [1] L1 training which consisted in assessing the grammaticality of SVA; [2] L2 training that similarly focused on subject-verb agreement; [3] non-linguistic working memory training; [4] normal tuition (control) with no extra training. After 16 training sessions over a 6.5-week period, the children showed evidence of transfer from WM to both L1 and L2 tasks, but no evidence of transfer from language training to WM. The effects were largely significant in respect of the L1, where working memory training was as effective as language training in the performance of the untrained L1 task, and moderately significant in respect of the L2. These results suggest that cognitive transfer was established unilaterally from WM to syntactic processing, and as such, support a connection between cognition and language with implications beyond those offered by correlational evidence. Lastly, the chapter concludes by discussing these findings against the backdrop of developmental cognitive linguistics and its integrated view of developmental language processes and domain-general cognition.

3.2 Introduction

While it is uncontroversial to claim that WM is required in many syntactic operations, this interaction between both systems is not clearly understood. Therefore, this study seeks to move beyond correlational evidence into possible causal links by testing whether WM training can improve children’s syntactic ability and vice versa, as captured by RQ1.

*Can non-linguistic WM training lead 7-year-old children to perform more accurate grammatical concord of L1 and L2 SdV sentences in a GJT transfer task and vice versa?*
The literature reviewed in chapters 1 and 2 inform the following conditions for this study. First, the aim is to explore the possible interconnection between WM and syntax processing. Second, to explore the possible bidirectionality of these effects, to date unexplored, two different trained-untrained task pairs are proposed: WM trained - language untrained and language trained - WM untrained (see Figure 9). A, B, and C support the case for a relationship between domain-general cognition and language, whereas D backs the view of a modular brain. Third, a widely studied WM paradigm (n-back) that has been documented in training studies to show transfer effects to other untrained WM tasks will be employed as trained task. Fourth, this WM task is structurally compatible with a language task (SVA) in the L1 and L2 where WM is significantly taxed.

Figure 9. Four hypotheses regarding the relationship between syntax and WM. Improved working memory performance will transfer into improved syntactic ability, but not the other way around (A); improved syntactic ability will transfer into working memory performance, but not the other way around (B); interactions between working memory and syntax will run in both directions (C); interactions will run in neither direction (D).

The need for a training study responds to the limitations of the type of evidence presented in support for an interaction between WM and syntax. Correlations between working memory tasks and syntactic tasks are abundant but, under the mind modularity canon, the interaction between working memory and syntax is dictated by the language system. Conversely, the DCL approach proposes an interaction where key syntactic tasks are largely conditioned by domain-general cognition. One prediction that follows from the deep integrationist account is that individual differences should be correlated. The reason for this is that if two cognitive capacities share similar cognitive routines, skills and resources, then being good at one ability should make somebody good at the other (Kidd, Donnelly, &
Christiansen, 2018). Many experiments taking this individual variation approach have demonstrated a consistent relationship between WM capacity and various aspects of linguistic ability, including grammatical judgements (Baddeley & Hitch, 1974; Ellis, 1991; McDonald, 2008a, 2008b; Wulseck, 1993) novel word learning (Atkins & Baddeley, 1998), reading comprehension (Baddeley, Logie, Nimmo-Smith, & Brereton, 1985; Daneman & Carpenter, 1980, 1983; Turner & Engle, 1989), understanding complex and ambiguous structures (King & Just, 1991; Miyake et al., 1994) and the ability to draw linguistic inferences (Cochran & Davis, 1987; Masson & Miller, 1983). Furthermore, the correlations between working memory and language have been established for both adults (Daneman, 1987, 1991; Daneman & Green, 1986; Masson & Miller, 1983) and children (Gaulin & Campbell, 1994; Swanson, 1996a).

As interesting as these cases are, they provide a rather weak test for the claim that language is built out of domain-general resources, because the majority of studies of this type have also used measures of working memory which have involved some linguistic component, such as being asked to read aloud a sentence while simultaneously comprehending another (Atkins & Baddeley, 1998; Daneman & Carpenter, 1980, 1983; Masson & Miller, 1983). Thus, they cannot rule out a linguistic-internal explanation for the correlation they establish between memory and language performance. A clearer test of the DCL approach would be to measure how non-verbal working memory, such as that used for shapes, constrains and predicts linguistic ability. Accordingly, the present study employs a non-linguistic, shape test of working memory (see section 3.3).

A more powerful demonstration of the causal interaction between language and cognition than the natural individuation variation approach outlined above would be to show that by training a child on ability $x$ they became better at the untrained ability $y$. That such cognitive transfer exists between domains may not seem particularly remarkable. What is more revealing is when transfer occurs in one direction but not in the other, for example from the general to the specific but not vice versa, then it can be used to infer something of the structural relationship between domains and the extent to which they are encapsulated from each other.
The training-induced transfer proposed here is designed following Gathercole et al. (2019)’s skills-based framework for cognitive transfer, previously discussed in chapter 2. The underlying assumption of the model is that cognitive transfer is established if [1] training leads to developing new practices or ‘routines’ that cue the necessary cognitive processes to successfully complete the untrained task and [2] the transfer task is structurally compatible with the training. If no new skills are learned, nor new routines of existing processes created, transfer and training are more limited in scope, even if they recruit the same cognitive processes. It seems that the complexity of the n-back task is what leads to the development of new cognitive routines for participants to succeed. The task is not only complex because participants have to mentally store and update shapes or letters, but because they also have to suppress the interference of the competing distractors.

This would be aligned with McCall & Mash (1995)’s view of inhibition as a key cognitive mechanism that facilitates language control, which replaces the assumption of mental modules with one of intrinsically interconnected functions. As observed earlier, n-backs seem to provide measurements of more than just WM. Inhibitory control has been linked in the past to language gains for a similar age group to the one involved in this study. For example, in a study testing 5-year-olds, Ibbotson & Kearvell-White (2015) found evidence that inhibitory control was a more accurate predictor of grammatical ability than age or vocabulary. The same is true for L2 learning, with some researchers going even further and claiming that the requirement to continually control their two languages improves bilinguals’ ability to exert executive control in general, as a consequence of which they outperform monolinguals in tasks that depend heavily on inhibition (Bialystok et al., 2004, 2005; De Groot, 2013). Understanding the role of inhibition becomes critical then in the assessment of n-back performance, as its additional demand on the memory task seems to be key in unlocking the new routines needed for training to transfer. When infants are able to inhibit stimuli that have been habituated already, they are able to use these resources to learn new routines effectively, linguistic or otherwise.

An argument could be made against the novelty of the cognitive routine behind grammatical agreement. To reiterate, the main caveat for positive transfer under the skills-based account is that the training/task is mediated by new routines. While
verbal WM is already fully established by the time children are seven years old and they have had vast experience of SVA by that time, they are not used, however, to establishing grammatical agreement over long distances, which is what is being explored here. This is supported by a corpus-based analysis of the frequency of SdV sentences in child speech in naturalistic settings, described in Chapter 5, and by suboptimal performance of participants at baseline (reported in section 3.3).

The study also differentiates the effects of any such cognitive transfer among L1 and L2 learners. The bilingual advantage hypothesis remains a contested idea for executive function as a global concept and even more so for working memory as a subcomponent (Lehtonen et al., 2018; De Groot, 2013; Bialystok, et al., 2004; Paap, 2019). Surveying these groups here serves not as a direct test of the bilingual advantage hypothesis, but rather as a further condition that may differentiate transfer effects between two populations who differ in their familiarity with language and the robustness of how they represent linguistic knowledge. The effects of cognitive transfer (memory to syntax and/or vice versa) might be stronger in the L1 training condition where WM abilities have greater room for development than for L2 learners who recruit WM more often and in greater depth. This would provide additional evidence for the skills-based account of cognitive transfer, namely transfer of WM to other capacities would be predicted to only occur when both the trained and untrained activities impose the same unfamiliar task demands that are not supported by existing WM subsystems.

In the case of the n-back training, because considerable variation has been reported in the strategies used to perform the task (Gathercole et al., 2019b), the prediction is that participants at that age are not equipped with ready-to-go mechanisms to handle the memory demands (see also on chapter 2). The argument presented here is that this will therefore lead them to learn a new complex cognitive skill to complete the task. Nevertheless, two modalities of the task, one factorial and two factorials, were employed to offer enough scope in terms of complexity, given that the second modality demands for more inhibitory control. Because at the 1-back level, targets are dropped immediately after a new item is presented, content is less pervious to interference than in situations with larger cognitive loads (i.e., 2-back).
In summary, we predict that the cognitive demands for processing an n-back task (1 and 2 loads) and SVA in sentences with distal subject-verb relationships (SdV sentences) warrant that new cognitive routines are learned. Success in both tasks is underpinned by participants holding relevant information for a short time, while simultaneously inhibiting competing information needed for a separate task. Therefore, there are strong indications that, provided that the methodological design takes into consideration the superficial compatibility of both tasks, the tasks share enough properties for training-induced transfer to occur under the conditions described in the skills-based theory of transfer (Gathercole et al., 2019b). To our knowledge we provide the most stringent test yet of the DCL account by (a) using a training study to establish cause and effect beyond the well-established natural individual variation approach (b) using a strictly non-verbal test of WM in combination with a strictly verbal (syntactic) one (c) testing the directionality of any cognitive transfer by employing a randomised control trial (d) differentiating these results in the context of first and second language learning. We do this by assigning participants to an L1, L2, WM training or control group (independent variables) and measure their accuracy on tests of syntax and working memory before and after training (outcome variables).

3.3 Methods

Participants

104 Cuban Spanish-speaking children (mean age = 7;2, range = 75-94 months, SD=0.44, F=54, M=50) learning English as a foreign language took part in all stages of the study, of the 116 initially contacted (see figure 7 below for a breakdown of all stages and group composition).

The reason why participants of early school years (1st and 2nd grade) were recruited was that, as discussed in chapter 2, children with typical development have exhibited enough indications that their WM systems, including skills that are critical to other executive functions such as rehearsal, are in place by the time they are 6-7 years old, even if they are not fully formed (Flavell et al. 1966; Gathercole et al., 2004). This means there is sufficient room for training to grow to as children’s WM skills have not yet reached ceiling performance (a critical analysis supporting this claim was
provided in chapter 2). From a linguistic development point of view, the appropriacy of the language task was tested on a pilot study with a smaller sample ($N = 24$, $M = 7;0$, range $= 73-89$ months) which showed a mean baseline target sentence accuracy of 43.06% for the L1 and 52.78% for L2. None of the participants who took part in the pilot took part in the main study.

Additionally, from a second language learning perspective, English proficiency starts to be measured systematically from 1st grade, when English becomes part of the national curriculum of Cuban government-run schools, which means exposure to the L2 through instruction was consistent across the sample. Nevertheless, L2 experience was quantified in months and used in tandem with parents’ reports of the participants exposure to English. Even though the difference between minimum and maximum English exposure among the participants is significant according to parents’ reports, these are just estimations and L2 proficiency was more accurately determined by the placement test the participants took when they enrolled in the private English schools where the study took place.

Participants’ proficiency in English is estimated at the upper end of the pre-A1 level, based on the Common European Framework of Reference for Languages and the latest descriptors of language competences for learners in the 7-10 age range (Szabo, 2018). This means they can recognise and use simple principles of word order in short statements and can understand simple statements and questions delivered slowly face to face.

*Design*

Using a randomised block assignment, four groups were formed by counterbalancing participants’ scores in all three tests. Group [1] received language training in Spanish, and Group [2] training in English, Group [3] non-linguistic WM training, and Group [4] normal tuition (control) with no extra training (see Figure 10). The content of all of these training sessions is detailed below.
Figure 10. Flow diagram of the design and structure of the study. Out of the 116 participants initially contacted, only 104 were included in the analyses of the results, 26 per training group (inc. control). Group allocation followed a randomised block order based on pre-test results in L1, L2 and WM.

**Context**

Participants were recruited from two private English language schools, located in two different municipalities of Havana, Cuba. All children aged 6 and 7 enrolled at the time were considered eligible to be part of the study. They attended two 60-minute English classes per week, in addition to their weekly or twice weekly English classes at school, which take place once or twice a week. While exposure to English has become more widespread in the daily lives of Cubans, mostly through American music, films, and TV shows, and especially after the collapse of the Soviet Union in the early 90s, the level of English proficiency throughout the country is generally low, and English graduates tend to pursue more financially attractive careers in the hospitality industry where multilingualism is always in high demand.

Both private language schools are part of the same academic network, under the same management, so they follow similar pedagogical approaches, shared principles,
and teachers and students often collaborate on classes and projects. Tuition fees are standard for both schools at 250 Cuban pesos per month, which means access is limited to those parents who can afford this.

To provide a clearer economic perspective on the matter, the average monthly state salary for 2017 is approximately 750 pesos, according to the Cuban National Office of Statistics and Information. It should be noted, however, that Cubans often rely on other complementary financial means that range from non-state jobs to international remittances, but the national literature on these is scarce. It also worth mentioning how, as a result of the recent boom of small businesses in Havana, and in Cuba in general, that followed the relaxed governmental policies concerning private property that came into effect in the 2010s, together with the influx of remittances from the Cuban diaspora, purchasing power estimates have shifted significantly in recent years. For instance, Delgado Vazquez (2017) found that the vast majority (89.2%) of the families surveyed considered international remittances to be their most important income, ranging from 2500 to 5000 pesos on a monthly basis, and almost a third (29%) claimed that they drew on these funds for their own or their families’ extra-curricular activities, especially for private language, music and dance classes.

<table>
<thead>
<tr>
<th>Schools</th>
<th>Britannia 1</th>
<th>Britannia 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Municipality</td>
<td>10 de Octubre</td>
<td>Centro Habana</td>
</tr>
<tr>
<td>Median salary of governmental jobs in Cuban pesos</td>
<td>664 pcm</td>
<td>610 pcm</td>
</tr>
<tr>
<td>Size sample</td>
<td>42 (25 females, 17 males)</td>
<td>62 (29 females, 33 males)</td>
</tr>
<tr>
<td>Age (year/months)</td>
<td>7;2</td>
<td>7;2</td>
</tr>
</tbody>
</table>


Given that all groups were counterbalanced, and the absence of dissimilarities between the samples from both schools, no significant influence of socio-economic status (SES) differences on cognitive development was expected. It should be noted
that in January 2021, after this study was conducted, the Cuban government introduced a new national financial reform which intends to raise the average salary approximately five times the figures reported in Table 1. However, the plan is not yet fully operational, and the cost of English private tuition is expected to rise accordingly. Because both schools are located in areas that share similar socioeconomic landscapes where economic access to English language tuition is the same, participants’ daily experiences are not predicted to have confounding effects of statistical significance on the post-training, between-group comparisons of performance.

Materials
During the pre-test stage, participants were tested for WM performance through a series of non-linguistic, shape-based, n-back tests and for syntactic processing through GJT.

The WM tests are n-back tests in which participants received non-linguistic visual input, via individual computer screens, in a sequence of shapes and were asked to indicate when a shape matched its immediate antecedent (1-back), or the one two positions back (2-back). In other words, they had to mentally store images of irregular polygons shown on a screen (visual stimuli) and press a button whenever matching images were presented consecutively (1-back) or alternately. The eight yellow irregular polygons were shown against a black background (see sample in Figure 11), and they randomly changed every three seconds (stimuli and instructions designed by Jaeggi et al., 2010). Participants very often interact with laptops in their regular English classes and at home as many of the coursebook activities are integrated into an interactive DVD. The n-backs were administered via Inquisit Lab, a software package containing a library of psychometric tests used in psychological experiments (Millisecond, 2018).
The GJTs focused on subject-verb agreement (SVA) in number in L1 and L2 sentences. As mentioned in sections 2.7 and 2.8 (Chapter 2), there is a long tradition of using participants’ judgments of sentence (un)grammaticality to infer the underlying processing abilities, especially in L2 studies. The GJTs for the L1 and L2 consisted of 20 and 32 sentences respectively, with half of the sentences being targets with higher WM demands (longer subject-verb distance). The remaining sentences followed a right-branching pattern and were used as fillers, as they do not tax the listeners’ WM in terms of SVA. Half of the target sentences illustrate two instances of mismatches in SVA in number, namely, singular subject – plural verb / plural subject – singular verb.

Local agreement and semantic plausibility were also counterbalanced within the presentation of L1 and L2 sentences. Because the number of the embedded noun can either aid or inhibit long-distance coordination of the head noun with the verb (Franck et al., 2002), this was controlled by allowing the number of the noun proximal to verb to agree with the head noun 50% of the time (e.g., 29) and disagree 50% of the time (e.g., 30). In this way we could be sure participants could not rely on a local coordination strategy to produce correct responses.

12- The book [under the piano] is white
13- The book [under the stairs] are white
The semantic plausibility of this relationship was also controlled such that local relationships (embedded noun-verb) were semantically plausible (e.g., in 31 a piano is the kind of thing capable of being white, as is a book) rather than implausible (e.g., in 32 a book can be written in German but the stairs cannot), which would have caused an unwanted cue to the grammaticality of agreement between the subject and main verb.

14- The book [under the piano] are white
15- The book [under the stairs] is in German

This supports the methodological decision of using GJTs to assess children’s processing of SVA. Accordingly, the L1 sentences used in the experiment contained distal subject-verbal relationships of the type:

16- *Las tareas de la clase de Lengua Española de primer grado es para mañana.
*The assignments of the Spanish language class of first grade is for tomorrow)

In (29), the mismatch between the plural subject tareas and the singular verb es is occluded by the prepositional phrases in between. In the case of the L1 sentences, both simple present and simple past conjugations of the verbs were used, as the latter forms are more phonetically distinctive between the 3rd person singular and the 3rd person plural than most forms in the simple present, where the pronunciation difference between regular 3rd person singular verb forms and plural verb forms is just a final /n/. Compare [30] and [31].

17- El maestro ve... vs. *El maestro ven... (The teacher sees...)
18- El maestro vió... vs. *El maestro vieron... (The teacher saw...)

For L2 sentences, simple present 3rd person verb forms of the verb be were used (is, are) taking into consideration students’ proficiency in English. Sentences also had shorter antecedent-gaps than in L1, like those in (4)

19- The bus in the school is yellow

---

4 The verbatim translation is respecting the sequence of the constituents in Spanish as much as possible. The purpose of the translated version, in this case, was to highlight the distance between subject and verb in the original, not to offer a bona fide equivalence in English.
In English sentences, third person forms of regular verbs are grammatically and phonetically more demanding for participants, as in most cases they are distinct only in a final /s/ or /z/ sound. Therefore, the forms of be (is, are) needed for successful agreement are relatively phonetically distinct from one another. Because more verbs were needed in Spanish to satisfy their greater dependency length, greater care was needed to offer L1 speakers the same key phonemic contrast and thus the same opportunity to succeed at the task. Because person was fixed, and the agreement always relied on number, the phonetic contrast was maximised by only using the past forms of the verb in 3rd person singular and plural (e.g., vio vs. vieron in 37), and thus avoid using the less distinct option where the 3rd person singular and plural of the verb only differ in one sound (final /n/ in 38).

20- El maestro vio ... vs. *El maestro vieron... (The teacher saw...)  
21- El maestro ve... vs. *El maestro ven... (The teacher sees...)

Collocation interferences were also taken into consideration in the design of the target items. For example, in the sentences below (39, 40) both nouns book and stairs collocate with the adjective white but in 34, the adjective long collocates more strongly with the local noun stairs than with the subject book, and this would have an effect on subjects establishing SVA.

22- *The book under the stairs are white  
23- *The book under the stairs are long

Proximity and semantic interferences were controlled as these could assist or hinder participants in their assessments of the target sentences' grammaticality. These errors and the long and short distances between subjects and verbs were tested for both L1 and L2 in two pilot studies conducted in advance. L1, L2 and WM training were matched in terms of time spent on intervention per child resulting in a total of 20 sentences for L1 and 32 sentences for L2. In this way we could ensure that both L1 and L2 were broadly comparable in terms of error rate before the training began: recall the pre-test accuracy of 43.06% for L1 and 52.78% for L2.
Additionally, all test lexical items in both languages were controlled for frequency using the *Corpus del Español: Web/Dialects* for L1 and the Corpus of Contemporary American English for L2. See the summary below.

<table>
<thead>
<tr>
<th>TEST ITEMS</th>
<th>SPANISH (L1) ITEMS</th>
<th>ENGLISH (L2) ITEMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus</td>
<td>Corpus del Español: Web/Dialects (Cuba)</td>
<td>Corpus of Contemporary American English</td>
</tr>
<tr>
<td>Corpus size</td>
<td>67 655 690</td>
<td>560 000 000</td>
</tr>
<tr>
<td>Word tokens</td>
<td>224</td>
<td>96</td>
</tr>
<tr>
<td>Word types</td>
<td>103</td>
<td>34</td>
</tr>
<tr>
<td>TTR (type/token ratio)</td>
<td>0.459</td>
<td>0.354</td>
</tr>
<tr>
<td>Average raw freq. word type</td>
<td>186 633</td>
<td>1 554 068</td>
</tr>
<tr>
<td>Ratio raw freq./corpus</td>
<td><strong>0.002758</strong></td>
<td><strong>0.002775</strong></td>
</tr>
</tbody>
</table>

*Table 2. L1 and L2 frequency analyses in Corpus del Español: Web/Dialects (Cuba) and Corpus of Contemporary American English.*

By balancing the frequency rates of the L1 and L2 lexical items used in the tests, any potential differences to be found between post-test performances in the L1 and L2 are believed not to be driven by how frequently they occur in Spanish and English, but by other factors, including those that concern this study, such as the influence of WM training on L1 and L2 syntactic processing skills.

Figure 2 offers a visual summary of the two main measures (training and untrained task) which highlights the superficial compatibility between the memory task and the grammatical agreement task. In the case of the n-back (left), the red arrow highlights the 3-second window between targets in a 2-back where participants mentally hold an image while dealing with a distractor. On the other hand, the red arrow on the SdV sentence (right) highlights the distance between subject and verb which gave children between 4-5 seconds to process the elements in between while maintaining the subject-verb relationship mentally active. The prediction here is that this structural compatibility increases the chances of the new skills learned during training to be applied into an untrained task.
Figure 12. RQ1 concerns the potential ‘spillover’ that results from training one arm of this hierarchy into the other untrained arm.

Procedure

Baseline measurements were collected over a three-day period for all participants, right before their regular English classes. The GJTs consisted of 20 sentences for L1 and 32 sentences for L2, and each set lasted 20 minutes on average.

Once all the groups were randomised, training began. Groups [1], [2] and [3] received two 25-minute sessions of training in L1, L2 and WM per week respectively for the first two weeks, and three sessions of the same duration per week for the following four weeks, for a total of 16 training sessions (6.67 hours). The training dosage was informed by a multi-level meta-analysis of n-back training sessions by Soveri et al. (2017) using a sample of 33 studies, which established a median level of 6.67 hours and 15 sessions. The language training in [1] and [2] was matched for time accordingly.

All sentences were read by their teacher, and participants had to choose between circling a checkmark (correct) and a cross (incorrect) on a piece of paper when answering the question: Is the sentence correct or incorrect? The checkmark-cross procedure was adapted from listening exercises used in their English coursebook. Because repeated practice could lead participants to work out the purpose of the task (match subject-verb), additional yes/no questions focusing on the central structures were added (see example 41, and the full list of sentences in Appendix 1).
24. The cat on the bed is white

Q – Is the cat under the bed?

The n-back task was framed as a two-level game, where children would progress to a more difficult level 2 (2-back) upon completion of level 1 (1-back). All participants initially had a condensed practice set of both levels, with immediate feedback, which helped minimise procedural errors. Upon completing the practice set, participants engaged with three trials of level 1, each with 20-21 shapes, randomised in order, before moving on to the three trials of level 2, with 22-23 shapes each.

After 6.5 weeks of training, participants were reassessed using a new set of WM and GJT tests.

Ethics

The study was approved by the Human Research Ethics Committee at the Open University UK (HREC 3368). Interested parents and children were briefed on site in Cuba with via oral presentation, with opportunities to ask questions of the lead researcher. Together with a written Participant Information Sheet (in Cuban Spanish), this informed potential participants of the aims of the project, who was conducting the experiment and their affiliations, why they were being invited to take part, what the study involves if they decide to take part, what we would be observing and the main research methods. They were also informed how their data will be used, stored and that no individual will be identifiable from the published results. All communications with parents and children stressed their right to withdraw from the study at any point as well as the lead experimenter monitoring for signs on non-verbal assent or otherwise during the experiment. It was made clear that, should one condition of the study perform significantly and robustly better than another because of an intervention, then that the same intervention would be offered to all participants. The intervention was accommodated to the institutions’ timetables to avoid disrupting regular classes, and the children’s English teachers assisted as co-researchers in all measurements. All briefing documents and consent forms are included in Appendix 1.

3.4 Results
The results are organised according to the dependent variable of interest. First, scores of the Spanish SVA task (L1) are presented, then scores of the English SVA task in (L2), and finally WM. For each set of results, it is reported whether there were any significant differences or not in these dependent measures as a function of the groups children were assigned to [1] L1 training, [2] L2 training, [3] WM training, [4] Control with normal tuition). The analysis was conducted for both pre-test and post-test scores. Age, gender, and second-language exposure (see section 3.3) were included as covariates that could all plausibly account for some variance in the outcome. Significant differences between groups are reported after covariate variance were controlled for.

*L1 pre-test*
To confirm that counterbalancing (see methods) had resulted in groups that were matched for initial working memory and Spanish skills, a 3-way independent ANCOVA was performed with L1, WM and control as main group variables, having age, gender and L2 exposure as covariates and L1 pre-test performance as our dependent variable. There was no main effect of group (F (2, 72) = 2.81, p = .901 ηp 2 = .003).

*L1 post-test*
To find out whether 6.5 weeks of WM and syntax training resulted in cognitive transfer a 3-way independent analysis of covariance (ANCOVA) was implemented with L1, WM and control as main group variables, having age, gender and L2 exposure as covariates and L1 post-test performance as the dependent variable. There was a significant main effect of Group (F (2, 72) = 17.67, p<.001 ηp 2 = .329). Pairwise comparisons (with Sidak correction) revealed a significant difference in L1 scores between L1 training and Control (p<.001, 95% CI [7.87, 20.30]) and between WM training and Control (p < .001, 95% CI [5.35, 17.24]). No significant differences were found between L1 training and WM training on L1 scores.
L2 Pre-test

As was the case with the analysis of L1, a 3-way independent ANCOVA was conducted to weigh for counterbalanced results in English and working memory skills, using L2, WM and control as main group variables, controlling for age, gender and L2 exposure as covariates, and using L2 pre-test performance as the dependent variable. There was no main effect of Group (F (2, 72) = .42, p = .959 ηp² = .001).

L2 Post-test

To determine whether similar transfer effects to those found in L1 took place in L2, we carried out a 3-way independent ANCOVA with L2, WM and Control as main Group variables, Age, Gender and L2 experience as covariates and L2 post-test performance as our dependent variable. There was a significant main effect of Group (F (2, 72) = 40.11, p<.001 ηp² = .527). Pairwise comparisons with Sidak correction, after controlling for covariate variance, revealed a significant difference in L2 scores between L2 training and Control (p<.001, 95% CI [8.96, 15.70]), between WM
training and Control (p < .001, 95% CI [3.27, 9.87]), and between L2 training and WM training (p < .001, 95% CI [2.47, 9.04]).

Figure 14. L2 subject-verb agreement by training group. Dark horizontal bars represent median scores for the group with boxes containing scores > 75% and < 25% quartiles. Small circles represent outliers between 1.5 and 3 times greater than the middle 50% quartile range and asterisks represent those greater than 3.

**WM Pre-test**

All four groups (L1, L2, WM, Control) were included in a 4-way independent ANCOVA to assess the results of counterbalancing for language and working memory skills. L1, L2, WM and control were the main group variables; age, gender and L2 exposure the covariates; and WM pre-test performance the dependent variable. There was no main effect of Group (F (3, 97) = .65, p = .978 ηp² = .002).

**WM Post-test**

Finally, to assess if language and working memory training led to gains in WM, we carried out a 4-way independent ANCOVA with L1, L2, WM and Control as main Group variables; age, gender and L2 experience as covariates; and WM post-test
score as dependent variable. There was a significant main effect of Group (F (3, 97) = 23.07, p < .001, \( \eta^2 = .416 \)). Pairwise comparisons (with Sidak correction) revealed a significant difference in WM scores between WM training and Control (p < .001, 95% CI [8.59, 19.75]), but no significant differences were found among the remaining groups.

Figure 15. Working memory performance by training group. Dark horizontal bars represent median scores for the group with boxes containing scores >75% and <25% quartiles. Small circles represent outliers between 1.5 and 3 times greater than the middle 50% quartile range and asterisks represent those greater than 3.

3.5 Discussion

As noted on the methods (section 3.3), the design of the WM training is consistent with other studies in terms of length and intensity, consisting of 6.67 total hours across 16 training sessions. To recap, a sample of 104 participants were divided into four counterbalanced groups and went on to receive either: linguistic training in SVA sentences in L1 [1] or L2 [2], WM training via non-linguistic 1-back and 2-back training [3], or normal tuition [4], for 6.5 weeks. After the training period, members of [3] (n=26) showed significant improvements in their L1 and L2 post-test
performances when compared to [4], but members of [1] and [2] showed no improvements in their WM post-tests when compared to [4]. Groups [3] and [4] had no significant differences in terms of WM and L1, as reflected by their pre-tests scores, but [3] outperformed [4], and almost matched [1]’s disproportionately high score (due to practice effects). The implications of these results are discussed below.

Strong cross-domain transfer effects were found between the WM training and both untrained language tasks (L1, L2), suggesting that SVA, a core grammatical process, is significantly dependant of working memory. In contrast, no effects were found supporting the case for language training on domain-general cognition, in either L1 or L2. The better performers in the L1 post-test were, unsurprisingly, group [1]. While they did not receive explicit feedback during their training, repeated training was predicted to lead to ‘inflated’ post-test scores due to practice effects. The prediction was to use their high score as benchmark and observe Group [3]’s performance in the spectrum marked by [1] (high) and [4] (low).

One of the key results of the study is that of [3], as it addresses whether domain-general skills are directly involved in syntactic processing. The findings seem to suggest that this is, in fact, the case, as training a group of participants in their WM skills using non-linguistic, visual input led them to improve their WM processing abilities in an aural linguistic task. Therefore, after six weeks of training, they successfully transferred the skills gained when performing a non-linguistic WM task to improve their performance in a compatibly formatted yet unrelated language task. This suggests an overlap or transfer between the underlying cognitive mechanisms recruited to successfully perform the WM task and the language task. In other words, training children’s WM skills improved their syntactic skills, even when language was not part of the training.

This would appear to be the first empirical study in the field where a clear causal association from domain-general cognition to syntactic processing skills has been recognised, suggesting that the involvement of domain-general cognition in prompting syntax is far more decisive than what was previously thought. Furthermore, the study also tested for bidirectional transfer effects (language training to WM) and found no evidence supporting the case for syntactic training
skills transferring to general WM skills. In other words, by revisiting the four hypotheses posited at the beginning of this chapter (see Figure 18 below), the evidence of this training test seems to suggest that the relation between WM and syntax is better illustrated by A, meaning that improved working memory performance led to improved syntactic ability, but not the other way around.

When exploring the bidirectionality of the effect, no transfer was established between language training and WM skills, as no WM gains were observed. Therefore, training language-specific memory skills does not seem to lead to improvements in general WM processing skills, limiting the interaction observed to a unidirectional influence of the general domain on the specific linguistic subdomain. The implications of the observed absence of influence of language on general cognition are as crucial to the unified model of cognition as is the measured influence of general cognition on language. That is to say that the significant effect of WM on syntactic processing observed in this study counters the language-as-module view of restricted involvement of general cognition in syntax inasmuch as the lack of influence of syntactic processing on WM, also evidenced in the results, rejects the singularity and distinctiveness of language from the rest of cognition.

However, if language and cognition are purported to be integrated, transfer should have been observed bilaterally (box C in Figure 16). One possible explanation for this may reside in the differences in the developmental stages of the cognitive resources (e.g., inhibition, attention, analogy, categorisation) and the language functions.
(morphology, phonology, argument-structure constructions) involved. There is good reason to suppose domain-general cognitive abilities, like categorization, event-segmentation, and memory were well established by the time language evolved and have had a longer phylogenetic history than language (Evans & Levinson, 2009). Of course, language-to-cognition effects are also apparent in development once language has emerged; not in as much as language determining the basic categories of thought or overwriting pre-existing conceptual distinctions, rather, as language making some distinctions difficult to avoid – language as a ‘spotlight’ on cognition – as well as augmenting certain types of thinking (Wolff & Holmes, 2011). For the age group tested here ($M = 7;2$), the expectations would be that these channels had been bidirectional for some time so further research is needed with a wider range of language and cognition training measures to fully understand the unidirectionality of the effect.

Furthermore, there are possible limitations to this interpretation of cognitive transfer, especially linked to the use of mnemonic techniques, which have been documented in past WM transfer studies (Holmes et al., 2009; Minear et al., 2016). For example, it could be argued that, although the inclusion of the shapes in the WM training is meant to discourage participants from naming them, and hence facilitating their recollection, there is no evidence that this is in fact so. Participants could have used their own linguistic tags to refer to each shape in the n-back task through associations with similarly shaped familiar objects, so if they are memorising words and not shapes, this would compromise the case for domain-general involvement. However, this approach relies on the assumption that 26 different children created their own shape-word associations, and these were relatively equally successful. If this were the case, it seems more likely that these tags would have come into effect once participants had become more familiar with them, say after a few training sessions, given the narrow time window (3 seconds) between shapes and the fact they would have to determine the total number of different shapes on their own. One way of addressing this in future studies could involve running a separate word-shape matching task at the beginning of the trial to assess possible correlations in participants’ associations.
From this it can be concluded that there is a strong case to be made for a unified view of cognition and language, as exemplified by the unilateral influence of working memory on syntactic processing in both Spanish (L1) and English (L2). This seems to reject the view of a modular brain with a dedicated language module and instead supports one that sees language more connected to the rest of general cognition. The implications of these results and those of the remaining studies reported in the following chapters for the developmental cognitive linguistics approach adopted in this dissertation will be discussed in more detail in chapter 7. This final chapter will examine all the findings that stem from each of the different research studies, with the aim of providing a more global and comprehensive overview of the nature of the relationship between language and cognition.
Chapter 4 – Study 2: reaction times

4.1 Chapter overview

The previous chapter focused on the effects of WM and language training by measuring children’s accuracy in an n-back task and a grammaticality judgment task (GJT) involving distal subject-verb agreement. The main finding suggests that WM training leads to improved performance in the n-back task and the GJT. The purpose of this second study is to explore in more depth the processing of the n-back through a statistical analysis of participants’ reaction times (RTs). This is intended to uncover possible individual differences in the response times to the WM task, especially because there are important antecedents in the literature, many of which report gender effects (Lejbak et al., 2011; Lowe et al., 2003; Pelegrina et al., 2015; Vuontela et al., 2003). The results of the statistical analysis performed on the RTs shows a bimodal distribution of the data that suggests that two main response strategies were used during the n-back, responding early or responding late, and neither had an effect on accuracy. These strategies were tested against gender and boys were found to respond faster and make more mistakes than girls. The implications of the results are of potential relevance to language processing in light of the positive cognitive transfer effects observed between the n-back and the processing of subject-verb agreement in SdV sentences, albeit establishing this connection would require measuring RTs in a comparable computer-based GJT. The full study will be discussed hereafter.

4.2 Individual differences in WM

Individual differences in general intelligence have been observed to respond to either the speed with which humans process information (Jensen, 2006) or the capacity of their WMs (Conway et al., 2002; Kane et al., 2005). While these predictors are often deemed to work independently, positive correlations have been established between both (Schmiedek et al., 2007; Schmitz & Wilhelm, 2016) and with other forms of executive functions (Dang et al., 2015). In the case of processing speed, behavioural indicators such as reaction times are often employed for measurements, and variability in response latencies has been documented extensively for populations of
different ages (Woods et al., 2015). Of particular interest to the previous study and the main motivation for this one are the disassociations that have been reported between accuracy and RTs in the n-back task (Jaeggi, Studer-Luethi, et al., 2010b). If processing speed is intertwined with WM and other forms of executive functioning, variation in the RTs is of special relevance to understanding the underlying cognitive mechanisms that operate during the n-back. This begs the second research question of this dissertation, namely,

*RQ2 – What motivates individual variation, if any, in the reaction times to the n-back task?*

Gender differences in children’s RT to n-back tasks have been identified in the past as a common source for variation in performance. A study by Vuontela et al. (2003) compared three school-age groups (6-8, 9-10, 11-13) looking at boys’ and girls’ reaction times to visual and auditory n-backs (0-, 1- and 2-backs) and found an apparent critical period (6 to 8 years old) when girls outperformed boys’ accuracy whilst adopting a late response strategy. This difference seemed to dissipate as children became older, as was negligible in the oldest group (11 to 13 years old). Visual tasks were performed faster and more accurately than auditory ones in all three age groups. Similarly, the oldest children responded faster and more accurately than the youngest group across all modalities of the n-back. However, barring this developmental difference, responding late seemed to be the strategy that led to fewer mistakes in both auditory and visual tasks for each age group. Late responders were mostly girls who also showed better accuracy rates in fewer responses in total. This finding has been supported by normative data of verbal n-back performance (1-, 2- and 3-back), collected by Pelegrina et al. (2015) from a sample of 3722 Spanish children, aged 7 to 13. Overall, mean response times decreased as children grew older and so did the number of mistakes, calculated as d’, a discrimination parameter which stands for the difference between the z-transforms of the hit rate and the false alarm rate, $z(H) - z(FA)$.

Not all studies have reported similar gender differences, however. Lejbak et al. (2011) compared male and female performances on a verbal, a spatial and an object 2-back task and observed no differences in RTs across all task modalities. Yet, in terms of
accuracy, no sex differences were reported in the verbal n-back, but male participants did outperform females in the spatial and the object n-backs. This echoes the results of other studies using different visuospatial WM tasks with high cognitive loads, which also support a male advantage (Cattaneo et al., 2006; Orsini et al., 2004). More recently, high-load conditions (4-, 5- and 6-backs) have been proposed as a more viable way to explore individual differences in adults' WM performance (Lamichhane et al., 2020). The belief is that inverted-U patterns start to emerge with +3 parameters of an n-back in adults and these functions are shown to signify that the cognitive system is overloaded (Van Snellenberg et al., 2015), and it is precisely under these conditions that the brain-behaviour relationships seems to be stronger. These limits have not been fully tested in children so it remains to be seen whether the findings of Lamichhane et al. (2020) open up an exploratory path for individual variation in younger populations. It should be noted that, unlike those reported by Vuontela et al. (2003) and Pelegrina et al. (2015), these differences, have been reported in adult populations, so it is highly possible that they reflect the ‘before’ and ‘after’ stages of a developmental shift that occurs later in adolescence, previously discussed in chapter 2.

Tracing these gender effects back to their neuropsychological sources is also not straightforward. Several computational approaches have proposed models that are claimed to account for individual differences in n-back performance. Chatham et al. (2011), for example, developed a biologically based computational model of the parieto-fronto-striatal system that performed n-back tasks under diverse conditions. The system is mapped across an ‘executive network’ that is theorized to recruit domain-general functions such as WM and inhibition. One of the peculiarities of the model is that it captured human variation in the n-back performance, which they associated with [a] genetic variation linked to dopaminergic functioning and [b] control strategies and response bias. Another study, by Rac-Lubashevsky & Kessler (2016), examined in more depth the role of control strategies by comparing reaction times during n-back performance (1- and 2-back) to those of the reference-back paradigm, a task generally similar to the n-back but designed to differentiate between matching and updating skills (see Figure 17). The experiment consisted in deconstructing the n-back task into its constituting skills that are theorised to be activated when information is [a] maintained or [b] updated in WM. Their results
support the hypothesis that an input-gating mechanism, responsible for controlling the contents of WM by ‘opening’ and ‘closing’ to internal (perceptual input) and external (long term memory) information, is the main source for individual variation in n-back performance. For other computational models of high-level cognition, see O’Reilly (2006).

Regardless of the source or sources of these individual differences, what the brief overview of the studies above aims to show is that there is sufficient compelling evidence to investigate variation in the response times to the n-back task introduced in chapter 3, especially for this age group. RTs and overall accuracy seem to follow independent developmental trajectories as evidenced by the disparity of both variables in the results of Vuontela et al. (2003) and Pelegrina et al. (2015). Because accuracy and RTs are interpreted as different measurements of performance, reporting both whenever possible, improves the overall understanding of the WM functions in place.

The inclusion of this chapter in the dissertation responds, to a large extent, to the existence of well-documented empirical studies (e.g., Vuontela et al., 2003; Pelegrina et al., 2015) showing the relevance of gender differences to WM through the analysis of RT data. Studying gender differences in children’s cognition can bring insight into when, where, and how these differences might emerge. The why might respond to an unequal distribution of inhibition resources for boys and girls at this age. Differences in WM underpin a wide range of cognitive tasks and developmental outcomes for children. One prediction that follows from the DCL account is that individual differences should be correlated. The reason being that if two cognitive
capacities share similar cognitive routines, skills, and resources, then being good at one ability should make the same person good at the other (Kidd, Donnelly, & Christiansen, 2018). Many experiments taking this individual variation approach have demonstrated a persistent relationship between WM capacity and various aspects of linguistic ability, including grammatical judgements (Baddeley & Hitch, 1974; Ellis, 1991; McDonald, 2008a, 2008b). Given that the previous study (chapter 3) evidenced the strong interaction occurring between WM and language, this study seizes the opportunity to analyse all data related to WM performance in a framework that does not undervalue the potential relevance of WM to language processing. This is not to say that the individual differences observed in the performance of the n-back task, discussed in this chapter, are predicted to have a direct bearing on the processing of grammatical agreement. This section explores individual differences in WM performance motivated by the existing empirical research in the field, in light of the findings of Study 1. Future studies may successfully or unsuccessfully connect the dots between this body of evidence and the findings discussed here.

4.3 Methods

Materials and procedure

Following up on study 1, discussed in chapter 3, this section looks at the RTs of all 104 participants in the n-back tasks during pre-testing and post-testing (208 tests in total). RTs and accuracy measures were obtained for all 121 runs of 1-back and 2-back in the pre-test and the same for the post-test. However, only trials where a button was pressed, referred to in the app as hits (H) and false alarms (FA), were included in this sample.

The version of the n-back administered is that developed by Jaeggi et al. (2010) which is fully incorporated into the app Inquisit 5 (Millisecond, 2018). RTs and accuracy measures were recorded by the app. Participants were seated comfortably in their regular school chairs in front of one of two laptops (13.3-inch screens, resolution: 1366 × 768 pixels). The distance between the eyes and the screen was approximately 50cm. Participants undertook two unrecorded practice trials (one 1-back, and one 2-back) before every session, after which followed three blocks of 1-back trials and three blocks of 2-back trials, in that order, that were part of the
analysis. Unlike the training blocks, the practice trials offered immediate feedback on accuracy, displayed as a percentage of correctness on the screen, which helped to minimise the occurrence of procedural mistakes on the training. All trials presented the stimulus for 500 milliseconds (ms) and offered a 2500 ms window before the next image was shown. Participants had all 3000ms to press the A key on a laptop keyboard whenever they deemed an image to be a target, that is if it matched the same stimulus presented one (1-back) or two (2-back) positions before in the sequence. The images in question were eight irregular, yellow polygons shown in the centre of a black background and the sequence of the stimuli was automatically randomised for each participant. A depiction of the procedure for the 2-back version of the test is presented below in Figure 18.
1. Images were presented for 2700ms and there was a 300ms transition between them. This gave participants a 3000ms-window to press the button A on a laptop keyboard whenever they thought the image presented had been presented one (1-back) or two (2-back) positions before.

4.4 Results

All hits (pressing the A button on a target) and false alarms (pressing the A button on a non-target) were extracted and used to illustrate the distribution of the data. In
total, 7888 responses were recorded between 0 and 2999 ms. The histogram (Figure 19) shows a bimodal distribution of the data, with the first peak occurring between 500 and 1000 ms and the second peak towards the end of the frame, approximately at 2750 ms. The second peak, indicating a slower reaction time, is noticeably larger than the first one, with over one thousand occurrences recorded around 2800 ms.

In terms of accuracy, indicated in red and blue, responding fast or slow does not seem to account for differences in scores. This was calculated by a chi-square accuracy comparison between early and late responses. Responses were labelled as early or late based on a median split of the data, at 1809ms. The chi-square statistic (Yates correction) $X^2 (1, N = 7888) = 2.7105$ and $p=0.99689$ shows no significant statistical differences in accuracy between early and late responses (Table 3).

![Figure 19. All reaction times for hits (red) and false alarms (blue).](image)

<table>
<thead>
<tr>
<th></th>
<th>Correct</th>
<th>Incorrect</th>
<th>Marginal Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>2148 (2111.04)</td>
<td>1797 (1833.96)</td>
<td>3945</td>
</tr>
<tr>
<td>Late</td>
<td>2073 (2109.96)</td>
<td>1870 (1833.04)</td>
<td>3943</td>
</tr>
<tr>
<td><strong>Marginal Column Totals</strong></td>
<td>4221</td>
<td>3667</td>
<td>7888 (Grand Total)</td>
</tr>
</tbody>
</table>

*Table 3. Chi-square table representing accuracy by early/late response. Values stand for frequency, values in parentheses for expected cell totals and values in brackets for the chi-square statistic for the individual cell.*
Because of the skewness of the distribution, with a major mode towards the late responses and the minor mode occurring between 500 and 1000ms, chi-squared comparisons were deemed more apt to explain the dichotomy in the data. An item-based analysis of a random sample was conducted to rule out the influence of specific stimulus. In other words, the aim was to clarify if participants’ early/late responses were a result of differences in processing specific images. Figures 20 and 21 present the RTs for each of the eight images used as stimuli.

![Image 1 RTs](image1.png)

![Image 2 RTs](image2.png)

![Image 3 RTs](image3.png)

![Image 4 RTs](image4.png)

*Figure 20. RTs of hits (red) and false alarms (blue) for stimuli 1 – 4.*
Because figures 20 and 21 show that no image is disproportionally eliciting responses for one of the two peaks, a next step was to explore if the distribution was a reflection of individual differences in response strategies. For this purpose, mean RTs were calculated for all 104 participants and plotted into a scatter graph. The noticeable antimode or gap between 2250ms and 2750ms in figure 22 confirms that participants adopted one of two strategies, responding early (between 750ms and 2250ms) or responding late (around 2800ms).
Figure 22. Mean RTs for all 104 participants. The discontinuance of the scatter approximately between 2250ms and 2800ms confirms the bimodal distribution of the histogram in figure 2.

The next step in the analysis was to address gender representation in these two default response strategies, taking into consideration the existing literature reporting differences in speed and accuracy between girls and boys. Using the median split for all RTs mentioned above, a chi-square test of independence was performed to examine the relation between gender and response strategy (Table 4).

<table>
<thead>
<tr>
<th></th>
<th>Females</th>
<th>Males</th>
<th>Marginal Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>1569 (1881.98) [52.05]</td>
<td>2376 (2063.02) [47.48]</td>
<td>3945</td>
</tr>
<tr>
<td>Late</td>
<td>2194 (1881.02) [52.08]</td>
<td>1749 (2061.98) [47.51]</td>
<td>3943</td>
</tr>
<tr>
<td>Marginal Column Totals</td>
<td>3763</td>
<td>4125</td>
<td>7888 (Grand Total)</td>
</tr>
</tbody>
</table>

Table 4. Chi-square table examining the relationship between gender and early/late responses.

The chi-square calculation with Yates correction showed statistical significance between the variables, \( X^2 (1, N = 7888) = 198.47, p < .00001 \). This indicates that female participants were more likely to respond later than male participants, as
A second chi-square analysis (Yates correction) was conducted to investigate the relation between gender and accuracy (Table 5). The results show statistical significance, $X^2 (1, N = 7888) = 25.10, p < .00001$ and suggest that male participants, who were in general early responders, displayed lower levels of accuracy.

At first glance, there seems to be an apparent contradiction between the first set of results and the second and third sets, namely that responding early or late does not lead to superior accuracy, and yet girls, who in general responded later than boys, were more accurate. Further examination of early (T1) and late (T2) responses in isolation reveals the following composition of gender and accuracy (Figure 24). This was confirmed by two separate chi-square calculations. In the case of T1, $X^2 (1, N = 3945) = 43.6972, p < .00001$, showing girls were significantly more accurate than
boys when responding early. However, in the case of T2, $X^2 (1, N = 3943) = .93, p = .3348$, which means girls and boys were equally accurate when responding late.

The last step in the analysis was disentangling 1-back from 2-back performance to examine, among other issues, if the skewed distribution of the data was a direct reflection of lower or higher memory load conditions. First, a chi-square test, with Yates correction, was performed to examine the relation between gender and early/late responses in the 1-back. This showed statistical significance between the variables, $X^2 (1, N = 4113) = 84.24, p < .00001$. In the case of the 2-back, the relation was also significant between both variables, $X^2 (1, N = 3696) = 99.50, p < .00001$. Figures 25 and 26 show these distributions. To compare the relation between gender and accuracy (i.e., hits and false alarms), a second set of chi-square tests were conducted. There was statistical significance between the variables in both the 1-back, $X^2 (1, N = 4113) = 17.82, p = .000024$, and the 2-back, $X^2 (1, N = 3696) = 9.70, p < .001836$. Figures 27 and 28 show gender by accuracy rates for the 1-back and 2-back, respectively, while Figure 29 shows that for both modalities of the n-back the distribution was bimodal.
Figure 25. Reaction times to the 1-back by gender.

Figure 26. Reaction times to the 2-back by gender.
Figure 27. Reaction times to the 1-back by accuracy.

Figure 28. Reaction times to the 2-back by accuracy.
Figure 29. Reaction times to the 1-back (blue) and 2-back (red).

Figure 30. Median reaction times of females (blue) and males (orange) to the 1-back and 2-back task.

Figure 30 (above) provides a general characterisation of gender performance in the two modalities of the n-back used in the experiment. Median reaction times were used as they reflect skewness slightly better than mean values, albeit less accurately than the raw values used thus far. Generally speaking, both females and males
responded later in the 2-back, while females responded later than males in both n-back modalities.

4.5 Discussion

Overall, the interpretation of the results can be summarised into the following points.

- First, children aged 7 seem to adopt one of two response strategies when undertaking an n-back task (1, 2 factors). They either react early (between 500 and 1500ms) or late (after 2750ms), as evidenced by the lack of recorded responses for the 500ms gap in between these two periods.

- Second, adopting one or the other strategy is positively correlated to participants’ gender. This was supported by a chi-square comparison between speed of response and gender showing that in general female participants opt for a late response strategy while male participants are more likely to respond fast.

- Third, a second chi-square comparison between accuracy and gender shows that female participants are significantly more accurate than male participants. Upon closer examination, this difference in accuracy is significant especially in connection with the early responses, where girls registered the highest accuracy rate of all the RTs. The last two findings confirm results previously observed by Vuontela et al. (2003) and Pelegrina et al. (2015).

- Fourth, all previous findings apply to both the 1-back and the 2-back, with RTs to the latter were slightly higher for both males and females.

The general developmental observation with RTs to the n-back task, and other memory tasks, is that as children grow older, they respond faster, which suggests speed is entwined with accuracy. Even for the same individual, regardless of age, late responses have been attributed to lapses in attention and working memory, and accordingly individuals with improved executive control are believed to experience these/said lapses with less frequency. Unsurprisingly, late responses offer more predictive value of someone’s underlying control mechanisms than early responses. This is what the worst performance rule posits (Larson & Alderton, 1990), namely
that the slowest RTs to a cognitive task tell more about someone’s general intelligence than their fastest responses. During childhood, children’s WM and executive functions are still not fully formed (see developmental discussion in chapter 2, section 2.3) and only in late adolescence and early adulthood do they reach peak performance in several cognitive tasks. However, during adulthood, response times to n-backs and other WM tasks increase as overall cognitive performance decreases with increasing age (Gajewski et al., 2018). It seems to be the case that several processing strategies are employed during an individual’s lifespan, depending on the development and readiness of the mental resources at hand at a given time. The results reported here, and by others, show that for children who are approximately 7 years old, responding either early or late in visuospatial 1-backs and 2-backs tends to be the default strategy. Girls, on average, responded later than boys and were more accurate in their early responses. Boys at this age seem to exhibit more impulsive reactions to visual input, which leads them to being more erratic.

There appear to be two important constraints when comparing these results with the broader literature. First, gender differences are less or more evident depending on the domain of the task. Generally speaking, the literature favours females in some verbal WM tasks (Temple & Cornish, 1993) and males in some visuospatial WM tasks (Lowe et al., 2003). However, this seems to be particularly true during childhood and adolescence as studies reporting between-gender differences are scarcer with adult populations (but see Freides & Avery, 1991), which leads to the second and more important constraint. Second, the findings seem to be relevant only to this age group, as further investigations of gender differences in young adult populations (+18 years old) performing a verbal n-back have found no behavioural or neurological (Schmidt et al., 2009) differences between males and females and some have reported a male advantage in visuospatial tasks (Cattaneo et al., 2006; Orsini et al., 2004). Similarly, studies with younger samples (kindergarteners and pre-schoolers), give boys a significant advantage over girls in visuospatial WM tasks (Robinson et al., 1996). In summary, it seems to be the case that the speed with which information is processed, represented in this study as RTs, and WM, are not only interconnected but constrained by other forms of executive functions (Dang et al., 2015), and as a result, different information processing strategies are employed across the lifespan depending on the attentional, control and memory resources that are available.
In terms of what motivates these differences, individual variation in n-back performance has been linked for both children and adults to multiple neurophysiological and psychosocial factors, ranging from neurodevelopment differences (Vuontela et al., 2003), the adoption of different response strategies (Rac-Lubashevsky & Kessler, 2016; Speck et al., 2000), and differences in educational levels (Patel & Azzam, 2005; Pergher et al., 2020). As discussed above and in chapters 2 and 3, other forms of executive functions apart from working memory are recruited during n-back performance, and because these do not develop at the same rate (see chapter 2 and also Xu et al., 2013), they constitute probable sources of variability especially for children this age. This chapter has given important consideration to individual variability (i.e., gender) in the response times to the n-back because of the interconnection of processing speed and WM, which in turn, and under the principles of the relationship between WM and syntactic processing here, is pertinent to both cognition and language.
Chapter 5 – Study 3: corpus study

5.1 Chapter overview

The purpose of this study as a whole is to closely examine the longitudinal relationship between grammatical and memory development among English- (n=608) and Spanish-speaking (n=43) children aged 2 to 12 years. To this end, three separate measures corresponding to a 10-year developmental period are compared, [1] MaxSV in child speech (CS) utterances, [2] the maximum CS utterance length in syllables, and [3] a normative measure of visuospatial WM. Values for [1] and [2] are expressed as number of syllables given that memory span has been proven to be more sensitive to syllable length than other measures such as word length (Baddeley et al., 1975; Cohen-Mimran & Adwan-Mansour, 2013; Service & Tujulin, 2002). By projecting these three values over the same period, we can understand better how SdV sentences emerge and whether this emergence mirrors the general development of grammar or memory. If, for example, the trajectory of MaxSV resembles that of visuospatial WM more than the typical incremental increase in utterance length, as predicted here, this could mean that the emergence and development of SdV sentences are contingent on extra-linguistic factors.

5.2 Introduction

Study 3 continues to explore the key relation between children’s development of working memory and of their language. In line with the previous experiment on cognitive transfer, discussed in chapters 3 and 4, the prime focus is on the role WM plays in processing SVA. Unlike that experiment, however, this study [a] addresses the involvement of WM in language production, in the form of CS, not language comprehension [b] covers a wider age range, 2-12 years, which enables the pinpointing of critical trajectory loci in language and memory development, [c] uses corpus evidence that reflects more naturalistic language settings, thus addressing validity concerns of language-usage approaches, and [d] establishes possible correlations or disassociations across different languages, namely English and Spanish. Accordingly, the following research question was posited:
(RQ3) Can non-linguistic WM act as a developmental predictor of children’s ability to establish grammatical concord in SdV sentences?

As a reminder, SdV are defined in this dissertation as sentences marked by the distal relationship between subject and verb. The distance between subject and verb leads to grammatical agreement being suspended and WM being taxed in order to establish grammatical agreement (SVA), as in the sentence *The two teams that he likes are the Milwaukee Brewers and the Royals*. The demand in WM is evident because the subject needs to be suspended before agreement can be established, while simultaneously processing the relative clause in between. If the distance grows or becomes more complex in its morphosyntax, so will the demands on working memory (for a cross-linguistic comparison of the effects of syntactic and morphological complexity and length on working memory see Marton et al. 2006).

In chapter 1 centre-embedded sentences, theorised here to be a subset of SdV sentences, were contrasted with sentences that have a decentralised structure and follow the more frequent right-branching pattern (Figure 31). In A, the ‘distance’ between *woman* and *ran* poses an extra processing demand for the memory and attentional resources already concerned with SVA. In B, however, the information can be processed linearly, and agreement, specifically, is never interrupted. This means that if the same speaker were to process longer versions of these two sentences, each bound by their original morphosyntactic distribution, the longest form of sentence A would be considerably shorter than that of sentence B.

![Figure 31. Tree diagrams of [A] centrally embedded and [B] right branching patterns.](image-url)
Generative approaches have used centre-embedding as an example of recursion being constrained by individual performance of what would otherwise be produced and processed ad infinitum (Chomsky, 1995). While recursion is also present in sentences that follow a right-branching pattern, such as B (e.g., The woman ran when the dog bit her after the car scared him...), these iterations are more easily processed than complex ones, such as the one in A (e.g., The woman that the dog that the mouse scared bit ran). In generative models of natural language these constraints are usually recreated by the introduction of extrinsic memory limitations. Because language syntax is conceptualised by Chomsky as context-free, the limitations to the number of embeddings that humans can process must be external to language, hence the memory impediment. However, there is still a mismatch between the processing and the production of centrally embedded structures that generative grammars fail to explain. Even when the memory resources are available, centre-embedding clauses are rarely used. In fact, corpus analyses of several European languages (Karlsson, 2007) have shown that sentences with double centre-embedding are almost non-existent in spoken language.

These examples suggest that the emergence and development of SdV sentences is to some extent contingent of the development of WM, given that an improved WM entails more than storing more items in memory. It involves a superior processing of the simultaneous task, which several studies on individual differences in WM have linked to differences in executive functions (Engle et al., 1999; Miyake, 2001). This is in alignment with the DCL perspective adopted here, where domain-general cognition is intricately connected, and WM is not considered to be operating as an independent unit.

What is in further alignment with the DCL perspective is adopting an approach based on longitudinal data. This complements the cross-sectional analysis discussed in chapters 3 and 4. Moreover, as a corpus-based study, it reflects children’s own language and thus adds a new dimension to the understanding of how SdV sentences are processed by focusing on their emergence. In other words, the answer to the research question broadens the scope of the interpreted results, by offering a developmental outlook of naturalistic language emergence over a 10-year period of
childhood. The aim is that by doing so, we can understand when and how the distance between subject and verb emerges, and when and how it grows.

5.3 SdV sentences [1]

As discussed before, the sentences that were part of the experimental study had been carefully designed to test children’s processing abilities, and for the purposes of Study 1, the distance had been controlled for in terms of length and complexity. The interest, however, has now shifted to understanding the natural emergence of that distance between subject and verb and how its syntactic length and complexity unfolds over a long period of childhood. Furthermore, mapping these trajectories in English and Spanish, with Spanish notably having a richer morphology than English, offers a more detailed account of the developmental relationship between WM and SVA agreement across languages.

5.3.1 English SdV sentences

In the case of English, the sample data used corresponds to children aged 1;6 to 11;11, accessed from three different English corpora (Gillam & Pearson, 2004; Weismer et al., 2013; Wells, 1981) available at the Child Language Data Exchange System (CHILDES). Opting for three different corpora was a necessity as no single English corpus provided data with the appropriate format (morphological line) for such a long developmental period. Transcripts were analysed be means of the Child Language Analysis program (CLAN) (MacWhinney, 2000). A summarised characterisation of each corpus, together with the raw totals of sentences extracted, is included in Table 6.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Age</th>
<th>Sample</th>
<th>Description</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gillam</td>
<td>5;0-11;11</td>
<td>n = 520</td>
<td>Listed as “naturalistic, storytelling”, children listened to stories, answered questions and then created their own stories with/without visual cues, as part of a test of narrative language.</td>
<td>Total CS utterances = 1824</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total SoV sentences = 1443</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SdV sentences (≥ 1 syllable) = 113</td>
</tr>
</tbody>
</table>
Composed of examiner-child (EC, elicitation) and parent-child (PC, naturalistic) interactions. Data collected annually at 2;6, 3;6, 4;6 and 5;6.

<table>
<thead>
<tr>
<th>Ellis Weismer</th>
<th>2;6-5;6</th>
<th>n = 56</th>
<th>Total CS utterances = 1953</th>
<th>Total SoV sentences = 789</th>
<th>SdV sentences (≥ 1 syllable) = 60</th>
</tr>
</thead>
</table>

| Wells | 1;6-5;0 | n = 32 | Naturalistic, spontaneous conversations between children and their caregivers recorded at home. | Total CS utterances = 1001 | Total SoV sentences = 205 | SdV sentences (≥ 1 syllable) = 10 |

Table 6. Characterisation of the three English corpora that were sampled, including age ranges, sample size, a description of the composition of the corpus, and the totals for CS utterances, sentences with a clear subject and verb (SV), and target SdV sentences extracted.

The first step in determining the distance between subject and verb was to regulate the sentences to be included in the sample. By SdV I mean sentences where the distance between subject and verb is measurable in syllables. For example, the subject-verb distance in sentence [42] is zero, as no syllables intervene between subject and verb, whereas [43]’s distance is six. The initial criterion was to extract sentences that followed a similar syntactic pattern to the one used in the experimental study, that is, \textit{the (art.)} + \textit{subject} + \textit{nominal adjunct} + \textit{(copular) verb} + ..., such as [43] to facilitate possible correlations. For these purposes, the first approach consisted in using a string that worked on the grammatical line (%gra) of the target corpus, as the coding at this level reflects syntactic function. Therefore, a string such as the one in [a], for example, was devised. However, the output of such a limited search string, and of other variations to it, returned too few sentences. This could be due to several factors, including the fixed syntactic constraints [44], the nature of CS output [45] and the ways in which CS is annotated [46] which was especially problematic in the case of Spanish corpora, among other reasons. For example, utterances [44] to [46] included below, would be disregarded by a string that only reflects the structure of [43], even when the SV distance is the same. Post-hoc analyses revealed that perhaps the most compelling explanation for the limited output is to do with the low frequency of occurrence of SdV sentences in child speech. This is explained in the discussion section (5.6).

[a] combo +s"\%gra:^DET^SUBJ^NJCT^POBJ^ROOT^*" +t%gra t*CHI *.cha
25- The wet towel is under the bathroom sink
26- The towel under the bathroom sink is wet
27- Towel under the bathroom sink is wet
28- It is wet, mom, the towel under the bathroom sink is wet
29- The towel <&towel> under the bathroom sink is wet

Although it was possible to make additional adjustments to the existing search string to facilitate the integration of utterances similar to [44] to [46], relaxing the constraints proved to be more constructive as it returned more sentences. This was also the case when the morphological line (%mor) was targeted over %gra, as some of the corpora had no annotation on the latter, which means the pool of selection was expanded when the focus shifted to the former. Furthermore, the %mor line tags every word in the main line by part of speech and is perhaps the most accurate CLAN tool for morphosyntactic analysis (MacWhinney, 2000). As expected, by covering more ground, new structurally different SdV sentences were added, and the possibility of fine-tuning post extraction granted more flexibility. This delayed the analysis given that the final filtering steps required manual intervention, but at the same time it increased the total output of SdV sentences, as detailed in the next section. In order to account for the issues illustrated by sentences [44] to [46], the search string was reformulated as follows:

combo +s”\%mor:^[det:*|^*^*n]*^*^* +t%mor t*CHI *.cha

Combo was used as it is particularly suitable for processing specific complex strings of grammatical patterns, words, and characters. The new search parameters target all child speech and child-directed speech (CDS) utterances. CDS was included for future referential analyses of child input vs. output only and was filtered out for the SdV analysis. The new output followed the pattern det + ... + noun + ... which is exceptionally inclusive. The decision to expand the types of determiners beyond articles was a minimal adjustment with respect to the original pattern and meant that sentences like [47], not too dissimilar to [43], were now collected. The distance between the determiner and noun is meant to support sentences like [48] with a prenominal modifier fronting the head subject noun. The code for copular verbs was deliberately excluded from the string to include sentences with all types of verbs like [49].
30- That towel under the bathroom sink is wet  
31- The only towel under the bathroom sink is wet  
32- The towel under the bathroom sink looks wet

In summary, the search string retained the necessary elements to maintain enough resemblance to the sentences used in the cognitive transfer study and gained enough flexibility to maximise the number of returns, as exemplified by the types of sentences in [44] to [49], which were originally excluded from the output even when their SdV quality is the same. The totals for CS output, sentences with zero syllables between subject and verb (SoV henceforth), and SdV sentences are listed in the last column of Table 7.

As a consequence of unhinging the string, the resulting output was categorised based on the nature of the distance between subject and verb, creating a working taxonomy of SdV sentences that, based on the literature reviewed, was inexistent. For example, the literature on the involvement of WM when processing relative clauses is abundant (Diessel, 2004; Diessel & Tomasello, 2005; King & Just, 1991; Larkin & Burns, 1977), and crucial to the understanding of subject-verb dependencies, but the distance between subjects and verbs can be occupied by other non-clausal structures, as evidenced by types 2, 3 and 4 in Table 7. The compilation of evidence of the involvement of WM in processing these structures is not as plentiful as that of relative clauses. It should be noted, however, that the taxonomy was not intended to guide the analysis, but rather helped to disentangle the composition of SdV sentences and to illustrate how distal subject-verb relationships, and their potential link to WM, can be studied in cases beyond those of relative clauses, which have been extensively documented.

<table>
<thead>
<tr>
<th>Type</th>
<th>Syntactic function</th>
<th>Example</th>
</tr>
</thead>
</table>
| 1    | Relative clause    | [50] the two teams that he like [* oes] oare the Milwaukee brewers...  
|      |                    | [51] a sink that floats  
|      |                    | [52] the kid named John is ok |
| 2    | Adverbial modifier | [53] my dad always has breakfast  
<p>|      |                    | [54] my dog really [/] really [/] really did die |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 3 | Subject adjunct | [55] they both did it  
[56] one of the babies is there  
[57] this thing down in her kitchen that has cream...  
[58] that one right there with the white mark was stitches  |
| 4 | Coordinated subjects | [59] the mother and the daughter cook  
[60] one day a boy named Michael and a girl named Eliza  
<were walking in> [/?] were walking in the park |

Table 7. Working taxonomy of SdV sentences including their syntactic function in their contexts. All sample sentences are real examples extracted from the corpora.

**Type 1**

Not all types of relative clauses contain active fillers between subject and verb, so sentences [50] and [51] serve to illustrate an important distinction. The former is a classic example of a relative clause that modifies the main-clause subject, in this case with an object gap (Tavakolian, 1978). However, these types of fully-fledged sentences tend to be less frequent in CS than [51] (Diessel & Tomasello, 2005). Sentences with isolated, relativised or gapped head nouns [51] are in fact more representative of natural CS. On the other hand, [52] is very much like [50] and the only difference lies in how the gap is occupied by a reduced version of a relative clause. Arguably, the agreement of kid and is could be assisted by the presence of John as a subject substitute. However, in other structurally similar sentences, such as the kid dressed in black is OK this would not be valid. Therefore, for general guidance purposes, sentences like [52] were included in the SdV count.

**Type 2**

Another common type of SdV sentence found in the output involves an intervening adverb between subject and verb [53]. Particularly relevant to this type is the exclusion of any instances of retraces and repetitions, as marked by any of the codes [/?], [/~], [/], [//], [///], and [nx]. While this was recurrent in type 2 sentences, the same principle applies to all sentences that were extracted, meaning they were all considered in their simple form and the material in parentheses stripped away. Therefore, in sentences like [54], the measured distance is two syllables (rea-ly) as the iterated forms were ignored. This is consistent with standard mean length of utterance (MLU) output in CLAN and therefore facilitated the comparability of both
maximum length of utterance and SV distance data. The SV distance in sentences like [55], with pre-verbal positioned adverbs that worked as substitutes of the subject in terms of number, was measured as zero, given that SVA could be established between the subject substitute (both) and the main verb (did).

Type 3

Type 3 accounts for subject-modifying prepositional and adverbial phrases such as [56] and [57], respectively. Functionally, they are similar to type 1, as they serve as modifiers of the subject head noun, and as such can work in combined structures with other modifiers. In these cases, unsurprisingly, the subject-verb distance tends to be considerable [58].

Type 4

The final type of SdV sentence targeted by the selection criteria is that of coordinated subjects. Some special considerations were taken in this instance, arising from concerns about the variability in measuring the subject-verb distance. There were two options for measuring the distance in sentences with coordinated subjects such as [59], measuring from the furthest subject noun mother or measuring from the closest subject noun daughter. For the purposes of this study, following the first option was considered to be more relevant because it involved the maximum distance for keeping fillers active in WM. Otherwise, measuring from the closest noun to the verb in sentences like [59] would lead to a zero-syllable subject-verb distance and hence dismiss the hypothesised role played by WM in SVA. In other words, the SV distance in sentences like [60] was measured from the furthest head noun boy to the main verb were, which in this case in particular totalled ten syllables.

Other remarks

In the case of sentences with the code xxx, if the syllable inflection was intelligibly marked in the audio file [61], provided that the corpus had one, the sentences were included in the output count. Otherwise, they were disregarded altogether. The same
conditions were applied to sentences such as [62] where personal names were
excluded from the transcription but were accessible in the audio files.

61- one day a new family from xxx moved in. (xxx was an unintelligible two-syllable word)
62- <the dog who> [///] the [/] the boy who has &-um the dog named Name* is Name (*Ro-sie)

Results

Once all English sentences were extracted and filtered, only sentences with clear SV
relationships were included in the analysis. In total, 2620 sentences were
plotted in a scatter graph to show any possible patterns in the data (see sample in Appendix 2).
The linear trendline in Figure 32 shows that the majority of these sentences had 0
syllables between subject and verb, and while there is a perceived increase over time,
it is minimal. A further scatter plot (Figure 33) confirms the Zipfian distribution of
the data, namely that there is an inverse relationship between the rank (number of
syllables) and the frequency of the data. This means that using mean values of SV
syllable length is not informative to the understanding of WM involvement in SdV
sentences, as these sentences are far less common. Instead, focusing on the trajectory
of the maximum SV syllable length over the 10-year period might prove a more
appropriate approach to explaining what children’s WM can do at different ages. For
example, the maximum syllable values in Figure 32 shows that children become
more adept at establishing long SVA as they grow older. It is also worth noting that
the ascending trend reaches what seems to be children’s maximum SVA performance
at around 120 months (10 years of age) and starts to decrease henceforth. This seems
to suggest that children become fully equipped to establish long SV agreements by
the time they are eight years old, and from there on opt for sentences with shorter SV
distance. This is not surprising. Even when children are able to establish successful
long SVA, this does not imply that they have stopped establishing short ones, as
shown by the trendline in Figure 32. It could be argued that, once children have
reached their ceiling performance, a decline in the trajectory should not be expected,
as they become more efficient in how they use language.
Figure 32. Scatter graph of the 2620 English sentences extracted with clear SV relationships. The trendline (dashed line) illustrates that most of the sentences have 0 syllables between subject and verb, even if this seems to increase over time.

Figure 33. Representation of the skewed distribution of syllable length between subject and verb in English sentences contrasting rank and frequency of syllables.

5.3.2 Spanish SdV sentences [1]
The Spanish data corresponds to two different corpora and was matched for the same period used in English, from 1;6 to 11;11 years of age (Albala & Marrero, 2006; Benedet et al., 2004), accessed via CHILDES and extracted using CLAN (MacWhinney, 2000). The table below offers a key summary of the data and the output obtained.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Age</th>
<th>Sample</th>
<th>Description</th>
<th>Output</th>
</tr>
</thead>
</table>
| BecaCESNo    | 2;10-12;2 | n = 40 | Descriptive study of the development of conversational language skills of Spanish children with TD. Transcripts included are free conversations between one or more children and an adult. | Total CS utterances = 3098  
Total SoV sentences = 1499  
SdV sentences (≥ 1 syllable) = 822 |
| Marrero      | 2;3-4;11  | n = 3  | Longitudinal study of TD Spanish children. Interactions are mostly between the target child and adults (relatives and investigator). Note - one of the children is from the Canary Islands, where people speak a Spanish variant similar in many ways to Cuban Spanish. | Total CS utterances = 153  
Total SoV sentences = 22  
SdV sentences (≥ 1 syllable) = 18 |

Table 8. Characterisation of the two Spanish corpora that were sampled, including age ranges, sample size, a description of the composition of the corpus, the totals for CS utterances, sentences with a clear subject and verb (SV), and target SdV sentences extracted.

In the case of the Marrero corpus, the search string used to extract the sentences was the same used in English, namely combo +s\( \%\text{mor}.^\dagger \%\text{det}.*\*\%\text{n}.*\*\text{t}.*\text{CHI}.^{.}\text{cha}\) which follows the det + noun + ... sentence pattern that was characteristic of the experimental study discussed in the previous section. However, the approach to obtaining the sentences from the BecaCESNo corpus had to be fine-tuned to address the absence of the %mor line and the %gra line from the annotation of the corpus. This adjustment consisted in reformulating the string to work within the constraints of the main tier, which meant targeting the actual words in a sentence, not their grammatical class (%mor line) or syntactic function (%gra line). A comparison of the annotation for the same sentence in all three tiers, main [a], %mor [b], and %gra [c] is included below. Accordingly, the new string was as follows, combo +s\|\text{el}^\dagger +fs +u
+t*CHI *.cha, which returned all sentences containing the article el. Parallel extractions were conducted for sentences with the article la, with los and with las, to match the output from the English corpora and the sentences used in the experimental study. However, the output for el sentences was large enough to enable comparisons (2620 English sentences vs. 2358 Spanish sentences).

[a] the + kid + ran – main line
[b] det:art + noun + verb - %mor line
[c] det + subject + root - %gra line

As was the case with English sentences, the first step when measuring the distance between subject and verb was to provide a rationale for classifying the sentences with one syllable or more between subject and verb (SdV). Spanish, unlike English, is a rich agreement language, which means the system of verb conjugation includes unique inflections for each combined person and number, totalling six or seven specific pairs, in contrast with the two possibilities of plurality for the 3rd person in English. This licenses, among other things, the extensive use of null subjects, forcing the listener to use the verb inflection to decipher the subject of a sentence, while English has only one unique verb agreement inflection (i.e., he/she sings). In other words, while Spanish counts with a rich inflectional system, there is little overt agreement in English, thus, person and number are often not morphologically realised.

Given the high frequency of sentences with null subjects, cases like [63] were labelled as SV sentences as they meet the same requirements outlined for English sentences, most importantly that no interference is observed between subject and verb. A different type of null subject case are impersonal sentences [64] (i.e., Spanish impersonal se sentences) which were not listed as SdV because agreement is not manifested, unlike null subjects where the subject is not overt, but concord is established.

63- *CHI: Corrió en el parque – (She/he/it) ran in the park.

Some Spanish variants use vosotros and its unique concord with the verb, in addition to ustedes, for the second person plural.
The null subject in Spanish is also exemplified by [65], where the SV distance from the last conjugated verb *tiene* is not measured to the subject *aguja* in the first sentence, but rather treated as a sentence with a null subject in isolation. This is consistent with the same type of additional conjunct subject elision that occurs in English, as evidenced by sentence [66], where the subject (*Meg*) is omitted in the coordinated sentence. The corpus annotation of this type of sentence is irregular, and often the elided subject can be found on a separate line, as in [67] [68], which raises questions about the consistency with which sentence length is annotated.

Word order

In Spanish, the most frequent word order for active voice sentences is SVO. However, there is significant flexibility in word order patterns. For example, the passive form is limited to use in formal situations, so a frequent alternative in informal speech is the use of the active voice with an object-dislocation construction, rendering an OVS word order. However, OVS word order is complicated by the obligatory use of object clitics [69].

Inverted order means that the traditional feature-copying approach emphasising the subject as the referential source for concord is unsatisfactory when measuring SV distance. If post-VS sentences can occur, and quite frequently in Spanish there is no reason why the same mental energy needed to hold active information related to grammatical number and person between subject and verb would not be used in VS constructions. Therefore, measuring from V to S is identical to measuring from S to V. In other words, the SV distance in [69] is measured from the verb *volvió* to the
article *la*, this being the shortest distance between two syntactic components that facilitate SVA. This means that both SV and VS measurements are equally possible, but whenever both are possible simultaneously, SV measurement tends to be favoured, as in [70]. This is particularly frequent in copular sentences where both noun phrases at each side of the verb agree with it in person and number. However, in those cases where they differ in number, as in [71], traditionally the verbal agreement is established with the plural noun (RAE, 2005).

69- *CHI: los volvió a repartir la niña .
(*again handed them out the child *)

70- *CHI: el niño más vago de la clase es Eduardo .
(the laziest kid in the class is Eduardo )

71- *CHI: mi preocupación son ustedes .
(*my concern is you all *)

**Clitics**

As pronouns, clitics act as substitutes, and thus most match their source nouns in person and number. However, *se* bridges its source noun/pronoun with the 3rd person conjugation of the verb in both singular and plural, and with the 2nd person verbal conjugation in plural. This means that deprived from its source, *se* offers multiple opportunities for agreement with the verb. For example, in sentences like [72] the distance to the verb is measured from the clitic as the clitic unambiguously references the subject. However, in [73] the proclitic form *se* could refer to the 3rd person singular, 2nd person or 3rd person plural, therefore the SV distance is measured from the noun *cosas*.

72- *CHI: nosotros mañana nos vamos .
(we leave tomorrow )

73- *CHI: estas cosas también se pintan .
(these things are also painted )

As for the proclitic forms that required a verbal host, the distance to the verb was measured from the closest form, noun or clitic, that enabled SVA. For example, in [73] the proclitic *se* does not suffice to establish agreement with the verb *pintan* as it cannot be pluralised. Therefore, the SV distance is measured from the noun *cosas*.
Proclitic forms not referencing the subject or requiring a verbal host, were counted following the typical SV measure, as they offered no assistance in SVA.

Other considerations

In highly inflectional and agreement-rich languages like Spanish, there are syntactic elements apart from the subject and verb that carry information, albeit partial, about the SVA of the sentence. For example, nouns and adjectives in Spanish agree in gender and number, which makes the latter feature, plurality, shared with the verb. Another example is the proclitic se which often acts as a substitute for a 3rd person source, singular or plural, thus conveying half of the information needed for establishing SVA. These two cases, acting simultaneously in the sentence in Figure 34, would propose in this type of sentence a shorter route to verbal agreement R2 than the one between subject and verb R1. However, there is no evidence than children actually use R2 for processing agreement or that a shorter distance entails more efficient processing. That is, although R2 is shorter, it involves mentally holding information from two separate categories, adjective and pronoun, to gain control of the features needed for SVA, whereas R1 involves a far more frequent scenario of shared features between subject and verb. These possibilities have been explored specifically for languages such as Spanish, and the results seem to suggest that gender and number agreement are processed differently but number agreement is computed as an unitary process for SVA and subject-predicative adjective (Anton-Mendez et al., 2002). Nonetheless, cases such as this where multiple cues from other syntactic components could be used by speakers to establish SVA were not common in the sample. It is thus only addressed here because of its relevance for future studies.
Results

The data was first organised into 10-month bins to avoid unnecessary clustering. Then the final step before performing the regression analyses was to remove four extreme outliers that were found in the SdV sentences, indicated by the red asterisks in the box plots below (Figure 35). Mild outliers, indicated by circles, were included in the regression analyses as they provide more relevant bordering measurements concerning SVA at maximum capacity.
Following the same procedure used with the English corpora, the sanitised data was used to plot a scatter graph (Figure 36). Similar to English sentences, the distribution of the data appears to be skewed as most of the occurrences are of sentences with 0 or 1 syllable in between subject and verb. The skewness was confirmed by a separate graph (Figure 37) weighing rank versus frequency of syllables. The maximum values mirror those of English too, suggesting that as children grow, so does the distance between subject and verb, and that once maximum SVA seems to be established by the time children are 10 years old (120 months), the trend hints at a subsequent decline.
Figure 36. Scatter graph of the 2358 Spanish sentences extracted with clear SV relationships. The trendline (dashed line) illustrates that most of the sentences have 0-1 syllables between subject and verb, even if this seems to increase over time.

Figure 37. Representation of the skewed distribution of syllable length between subject and verb in Spanish sentences contrasting rank and frequency of syllables.

The similarity between the English and Spanish data can be observed clearly when they are superimposed onto the same graph (Figure 38) as both trends suggest an increase in syllables as children become older. The only noticeable difference in both
datasets is that the distance between subject and verb is longer in Spanish than in English. This is not surprising considering how rich verbal agreement is in Spanish as compared to English, and specifically how proclitic forms substitute direct and indirect objects in preverbal positions, thus adding to the distance between subject and verb. As mentioned before, this is characteristic of Spanish and has no syntactic equivalence in English.

![Graph showing Maximum SV distance](image)

Figure 38. English and Spanish maximum syllable length between subject and verb for each 10-month bin between 29 and 149 months. There are 13 values per language, but these are represented in the same colour whenever they are the same in an age bin (e.g., 49-month mark).

### 5.4 MAXWD [2]

#### 5.4.1 English MAXWD [2]

As mentioned previously, the longest utterance of each corpus for each 10-month bin was used as an indicator of grammar development. More specifically, the syllable length of the longest utterance was used as a benchmark to which to compare the maximum syllable length of the SdV sentence that matches the same age period. The comparison at hand, thus, is concerned with morphosyntactic length increase and between a type of sentence that grows centrally (SdV) and a type of sentence that follows a rightward trend (right-branching).
Averages were calculated whenever two or three corpora overlapped in time. For example, for the 60-69-month group, three different measures were averaged, one for each corpus, and the resulting value was used to plot the trajectory (blue line in Figure 39). In order to extract these utterances, the following search string was used:

```
maxwd +t*CHI +g1 +c5 +d1 *.cha
```

Maxwd is commonly employed to extract the longest word in a given dataset but when used with the +g switch, it computes utterance lengths. The +c5 switch limits the output to the five longest utterances. By default (+g1), the length is measured in morphemes, with additional switches for words (+g2) and characters (+g3). However, as no syllable count was available, the default switch was used, and the syllables counted manually. It should be noted that two exceptions were made when measuring maximum length of utterance for two age bins, 30-39 and 40-49, of the Wells corpus. In the first case, the top three sentences were irregularly high in length, when compared to other values of the same corpus and the other two corpora. In the second case, the top two sentences showed a similar pattern. Because these were all lyrics to popular nursery rhymes (Twinkle twinkle little star for the first group, and There was a princess for the second one) and therefore were not instances of spontaneous CS, the immediate highest values were used instead. These were then averaged with the matching values from the Ellis Weismer corpus, following the standard procedure.
The key finding thus far, when examining the trajectories plotted in Figure 39, is the difference in the trajectories of both lines. While in both cases the overall pattern is a sustained increase in the number of syllables, it is noticeable how children’s long SVA remains practically unchanged from approximately the 100-month mark, while utterance length keeps its upward trajectory.

It seems that, by the time children are 7-8 they stop coordinating longer dependencies even when their utterances continue to become more complex. This could suggest a possible disassociation between the development of the child’s grammar and their ability to establish agreement between subject and verb, thus hinting at a fundamental reliance on non-linguistic factors. However, this claim can only be supported by a third line, denoting one of the extra-linguistic factors, that parallels the trajectory of maximum SV distance.

5.4.2 Spanish MAXWD [2]
The next step in the analysis was extracting the measurements of the longest utterances (in syllables) to match to each of the age bins. Averages were used whenever corpora overlapped at any given interval. In the case of the Marrero corpus, [a] was used as a search string to isolate the five longest utterances by morphemes using the +g1 switch on the mor line. However, as discussed above, the BecaCESNo corpus had no morphological annotation which rendered all computations on the mor line invalid. However, the +g switch in the maxwd string has two other sets to determine the longest utterances that work on the main line, +g2 for total words and +g3 for total characters. Accordingly, the +g2 switch was used for this corpus and the five utterances counted using an automated syllable counter to determine the longest one by syllable number.

[a] maxwd +t*CHI +g1 +c5 +d1 *.cha
[b] maxwd +t*CHI +g2 +c5 +d1 *.cha

Figure 40. Trajectories of MaxSV (orange) and maximum syllable distance between S and V (blue) in Spanish SdV sentences.
Once the values denoting the longest utterance (in syllables) for each of the age intervals were obtained, these were used to sketch a representation of the developmental trajectory of syntax growth using a logarithmic Y axis (Figure 40). The scale was needed to address skewed values corresponding to the maximum utterance length for some of the age periods and provides a clearer visual comparison of both trajectories.

5.5 WM normative scores [3]

Obtaining comparable, normative WM data that would cover the same age span (~2 to ~12 years old) required using measurements from two different intelligence tests, the Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition (WPPSI-IV) for data covering the 2;6 to 7;7 period and Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) for data from 6;0 to 16;11-year-olds. Both test batteries are part of a family of intelligence tests originally developed by David Wechsler commonly known to psychologists as the ‘Wechsler scales’ and have been widely researched and trialled (see summary in Niileksela & Reynolds, 2019). Although the tests are developed and normed separately, the constructs they set to measure have been shown to be virtually the same (see full factorial invariance study by Niileksela & Reynolds, 2019), which supports the decision to effortlessly integrate the scores for one of the subtests that compose the Working Memory Index (WMI).

In the case of WPPSI-IV, children engage in the Picture Memory (PM) task, where they are asked to hold pictures mentally after seeing them on a page for some time, and indicate on a response page which pictures were shown (Wechsler, 2012). On the other hand, WISC-V’s WM subtest, Picture Span (PS), requires that children do the same but following the same sequential order in which the stimuli were presented, if possible, which is an effective way of adjusting the level of complexity to older children without transforming the task too much. The score scale for both tests is identical, ranging from 1 to 19, with the 10-benchmark representing the average score for each age period. Due to the age overlap between both tests for three intervals between 70 and 99 months, averages of the PM and PS scores had to be calculated. Finally, the regression analysis focused on the age groups from 30 to 149 months as the lowest boundary (20-29 months) was not covered by WPPSI-IV.
Raw score averages for the 10-benchmark were used as the WM measures to compare with SdV and MLU. It should be noted that both SdV syllable length and MLU syllable length are exemplary cases of maximum values which are suggestive of ceiling performance. In contrast, the WM measure is a reflection of average performance and this restricts the comparisons to be made with the other two variables of interest. However, the purpose here is to have two different benchmarks across a 10-year span, one projecting morphosyntactic growth and one projecting visuospatial WM gains, that serve to contextualise the emergence and development of SdV sentences, as visualised by Figures 41 (English) and 42 (Spanish).

Figure 41. The SV distance in English SdV sentences (blue) and the MaxSV of the longest English CS utterance (orange) are plotted on the primary y-axis (syllables). The development of WM (grey) is plotted on the secondary y-axis, indicating the raw average scores for the visuospatial WM subtests of WPPSI-IV and WISC-V.
The last step in the analysis was to conduct a regression using the values for maximum length of utterance [2] and WM [3] as predictors of SdV sentence occurrence [1]. For the English sentences, the analysis of variance showed statistical significance between the independent and the dependent variables (F 11.819, p = .003). However, when weighing the significance of this influence between both predictors, [3] was found to be as equally capable of affecting [1] as [2], as evidenced by their standard coefficients .431 and .432. (Table 9).

### ANOVA

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>45.085</td>
<td>2</td>
<td>22.542</td>
<td>11.819</td>
<td>.003</td>
</tr>
<tr>
<td>Residual</td>
<td>17.165</td>
<td>9</td>
<td>1.907</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>62.250</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### COEFFICIENTS

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardised Coefficients</th>
<th>Std. Error</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
</table>
In the case of Spanish sentences, the regression also revealed statistical significance between the predictors and the dependent variables (F 23.867, p < .001). When comparing the individual influence of the predictors, [3] was found to be a significantly better predictor of [1] (beta = .838) than [2] (beta = .095) (Table 10). The implications of these results will be fully discussed in the next section.

**Table 9. Analysis of variance of syllable distance in English SdV sentences, using WM and maximum length of utterance as predictors. Beta coefficients show no difference between predictors.**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Beta</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>1.076</td>
<td>1.949</td>
<td>.552</td>
<td>.594</td>
<td></td>
</tr>
<tr>
<td>WM</td>
<td>.125</td>
<td>.151</td>
<td>.432</td>
<td>.825</td>
<td>.431</td>
</tr>
<tr>
<td>MLU_EN</td>
<td>.096</td>
<td>.116</td>
<td>.431</td>
<td>.823</td>
<td>.432</td>
</tr>
</tbody>
</table>

In the case of Spanish sentences, the regression also revealed statistical significance between the predictors and the dependent variables (F 23.867, p < .001). When comparing the individual influence of the predictors, [3] was found to be a significantly better predictor of [1] (beta = .838) than [2] (beta = .095) (Table 10). The implications of these results will be fully discussed in the next section.

**Table 10. Analysis of variance of syllable distance in Spanish SdV sentences, using WM and maximum length of utterance as predictors. Beta coefficients show that WM is a far better predictor of variance than MLU.**

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>121.156</td>
<td>2</td>
<td>60.578</td>
<td>23.867</td>
<td>.000</td>
</tr>
<tr>
<td>Residual</td>
<td>22.844</td>
<td>9</td>
<td>2.538</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>144.000</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 10. Analysis of variance of syllable distance in Spanish SdV sentences, using WM and maximum length of utterance as predictors. Beta coefficients show that WM is a far better predictor of variance than MLU.**

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardised Coefficients B</th>
<th>Std. Error</th>
<th>Standardized Coefficients Beta</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>.923</td>
<td>1.256</td>
<td>.735</td>
<td>.481</td>
<td></td>
</tr>
<tr>
<td>WM</td>
<td>.368</td>
<td>.099</td>
<td>.838</td>
<td>3.732</td>
<td>.005</td>
</tr>
<tr>
<td>MLU_SP</td>
<td>.005</td>
<td>.011</td>
<td>.095</td>
<td>.425</td>
<td>.681</td>
</tr>
</tbody>
</table>

**5.6 Discussion**

The above findings have several implications. First, SdV sentences occur less frequently in child speech than what formal generative approaches have posited. Even though centre-embedded clauses, a subgroup of this broader SdV classification, have attracted considerable attention from generative approaches as primary examples of infinite recursive patterns, these were scarce. In fact, even after using several corpora comprising an 11-year span, SdV sentences were significantly rare in
English with only 3.8% of all CS extracted showing at least a syllable between subject and verb. While calculating the percentage occupied by SdV sentences in each corpus proves more difficult, it is sensible to assume that this would be considerably small, given that SdV sentences were only a small subset of the search string output. This supports previous findings by Diessel & Tomasello (2000, 2005), who have observed that children tend to use ‘broken’ versions of relative clauses, which are clauses attached to an isolated head noun or to the predicative nominal of a copular clause (e.g., *This is the sugar that goes in there*), more than fully-fledged relative clauses (e.g., *The sugar that is on the table goes in there*) in spontaneous speech. Furthermore, longer SVA (at least 8 syllables) do not emerge until later in life (circa 5 years) and continue to develop until the child is 12 years old, suggesting that the ability to process recursive patterns is acquired in a piecemeal manner. Christiansen & MacDonald (2009) have concluded the same through their connectionist learning model in their study of centre-embedding and left- and right-recursive structures.

Second, long utterances are significantly longer than SdV utterances, which supports the view that most sentences tend to follow a right-branching pattern. This was illustrated by the comparisons of the longest utterance and the longest SdV utterance in English (fig. 10) and Spanish (fig. 11). This is especially true in the case of the latter, where a logarithmic y axis had to be employed because of the substantial difference in values. Additionally, the lack of SdV occurrences in the 20-29 period suggests that long utterances emerge before SdV utterances in child speech.

Lastly, regression analysis shows that the development curve of SdV sentences in English is equally predicted by WM development and maximum utterance length. In Spanish, WM development was found to be a better predictor of SdV sentences than maximum utterance length. This entails that SVA is heavily dependent on WM, which more broadly suggests that language development is arbitrated by the development of non-linguistic cognitive capacities such as WM.

One important caveat to the understanding of SVA in this study is that not all SdV sentences exhibit morphological agreement in the same way that SdV sentences did in Study 1, where plurality mismatches for subject-verb pairs were controlled for. Many of the sentences extracted from the English corpus, for example, use regular
verbs in past tenses (e.g., *that big bad guy always killed them*) where the choice of tense removes number ambiguities. This, unarguably, would alleviate the processing (here, production) demands of agreement. However, as discussed in section 2.6 of this dissertation, agreement relationships cannot be completely ruled out, as memory retrievals are theorised to pervade the relations between subject and verb in agreement (Lewis & Vasishth, 2005). So, while it can be argued that the type of subject-verb matching in some of these corpus sentences might be well different from the one in SdV sentences in Study 1, both seem to be mediated by WM in some, probably different, way. The focus here was evaluating if the emergence of sentences that contain any information between subject and verb reflected the trajectory of non-linguistic WM for the same period. It was a first attempt to understanding the emergence of children’s SdV sentences in naturalistic settings. The developmental correlations showed only apply to non-linguistic WM and the SdV sentences extracted from the corpora selected, not to all instances of long-distance agreement. Further investigations that [1] examine sentences that do contain subject-verb number mismatches, [2] scrutinise the contents of the *d* in SdV sentences (i.e., syntactic function, semantic complexity), and [3] match the child’s SdV sentences to the CDS SdV sentences they are exposed to, are needed before drawing any conclusions about the role of WM in long-distance agreement for this developmental period.

Until now, we have examined how training children’s WM in a randomised controlled trial using non-linguistic tasks has shown positive transfer effects when they later perform long SVA in English and Spanish. However, one limitation of that study was that the transfer task was artificially created to address compatibility concerns. That is, even when cognitive transfer was established, the sentences including long SVA were not children’s sentences, so there are restrictions to any claims to be made about the influence WM training had on their own language. As a follow-up to the experiment and to address its design constraints, this corpus study sought thus to investigate the emergence and development of a group of SdV sentences over a wide longitudinal period by surveying child speech and established strong correlations with the development of WM. However, even when these sentences with long SVA were not fabricated and are actually child utterances, we know very little about their syntactic nature. If there is something about the structure
of an SdV sentence that supports children’s establishment of grammatical agreement, this would undermine to some extent the case for a deep integration between language and cognition. For instance, if these structures are found to be somewhat cyclic repetitions of rehearsed syntactic chunks, it is not clear to what extent WM is involved in agreement. The next chapter attempts to address this specific question, thereby completing this family of studies exploring the interconnection between syntax and WM.
Chapter 6 – Study 4: SdV formulaicity

6.1 Chapter overview

Earlier chapters in this thesis have argued for a strong role of working memory in both the longitudinal emergence (chapter 5) and in the online processing (chapter 3) of SdV sentences. However, WM is only implicated in these processes to the extent that SdV sentences cannot be readily retrieved from memory as frozen phrases or formulaic chunks. A high formulaicity index for these sentences in comparison with a matched sample would reduce the role of WM and weaken the general argument in favour of deep integration between domain-general processes and language. To rule out this explanation, this chapter takes a deeper look at the composition of the SdV sentences in a corpus of naturalistic speech. First, all SdV sentences were parsed into JavaScript Object Notation (JSON) format to be used as input in a Sankey diagram generator. The resulting visualisations of the syntactic distribution of the English and Spanish SdV sentences showed clear instances of both sentence formulaicity and flexibility. To quantify this relationship, a novel network-measure of sentence formulaicity was devised which revealed that the English and Spanish SdV sentences showed use that was somewhere between complete formulaicity and flexibility. Because the SdV sentences showed some level of flexibility of use – that is, they were not all repetitions of previous use – the results make a further contribution to the claim made in this dissertation that SVA in SdV sentences is contingent on WM. The chapter concludes by recommending a more systematic characterisation of morphosyntactic complexity and a case-by-case analysis of SdV sentences for future studies.

6.2 Introduction

All the studies discussed thus far have produced and reviewed cross-sectional and longitudinal evidence suggesting that SVA in SdV sentences is largely dependent on WM skills. Study 4 was specifically designed to assess whether morphosyntactic repetitiveness could be an intervening factor facilitating WM recall. Analysing the morphosyntactic composition of these sentences serves to confirm or dismiss the
possibility that repetitiveness could have acted as a confounder in studies 1 (chapter 3) and 3 (chapter 5), prompting the following research question:

(RQ4) To what extent are SdV used formulaically and what are the implications for WM?

Two issues ensue from this research question. On the one hand, if the morphosyntactic composition of SdV sentences is flexible, as flexible as it would be in comparable syntax (e.g., utterances by the same speakers with no suspended agreement), then the structures to be held mentally between subject and verb demand enough flexibly not to be readily retrieved from memory as a frozen-phrase or formulaic chunk. WM will be taxed on this occasion as it is required to coordinate the independent elements in the sentence. But if, on the other hand, there is some unusual ‘formulaicity’ to the syntactic structures enclosed by subject and verb, then it could be argued that the WM burden is alleviated by the chunking because they are stored as a whole, unanalysed form. For example:

74- The twins, who John knows are wrong, are gossipers.

If there is no trace that these elements have been combined before in this way, agreement syntax, and arguably WM, need to be involved for successful agreement between subject and verb. If, however, the same elements have frequently co-occurred, there is more chance they will be stored as a relatively unanalysed chunk, as in [75], where the whole utterance is processed and stored more as if it was a word.

75- Thetwinswhojohnknowsarewrongaregossipers.

The model of acquisition that underpins this dissertation poses that children acquire chunks from the input they are exposed to, and the degree to which they acquire these chunks depends on their frequency of occurrence and their future usability. Studies such as the one by Lieven et al. (2009) have traced back children’s utterances to their own previous utterances and observed that their speech contains up to 40% of multiword sequences that they have used before. Moreover, in almost half of the
cases, the utterance was identical to a previous utterance with only one distinct feature, suggesting that in many cases new material could have been inserted into ‘a prefabricated frame to produce a novel utterance’ (Bannard & Lieven, 2012). If children’s language production is so dependent on how successful they are in reusing the language they have been exposed to, it is sensible to explore if these ‘prefabricated frames’ occur more frequently in sentences with distal subject-verb relationships.

Chunking and other forms of strategic processing such as verbal rehearsal are the types of cognitive tools that would facilitate the construction and immediate recall of these frames. As powerful mnemonic techniques, they enable children to optimise information processing and improve subsequent recall. Children resort to chunking from a very early age whenever their WM is overloaded with continuous input (Jones, 2012) and using a strategy for chunking items in WM to link them to multi-chunk items in long-term memory for later retrieval is common to both visual-spatial processing and language (Christiansen & Arnon, 2017; Cowan et al., 2012). Segmenting and compartmentalising new information are complex cognitive skills where EFs, such as attention, categorisation and analogy intervene to make immediate recall possible. Because there is no direct way of testing the extent to which the sampled children employed chunking strategies to process the SdV sentences that they were exposed to, all assumptions about the operationalisation of this cognitive mechanism are based on post-hoc analyses. One of the methods in which this could be explored, the one adopted by this study, is by examining children’s own usage of SdV sentences, more specifically by focusing on morphosyntactic formulaicity as a factor that encourages or discourages chunking. This is supported by studies that have recorded children using chunking strategies in data with specific formatting characteristics.

For example, a recent computational study by McCauley and Christiansen (2019) proposed a model of child language acquisition where, as theorised by DCL, learning occurs in an item-based type of fashion, which contradicts the modular view of learning whereby an inherited mental grammar is activated as a reaction to input. In their model, children showed high sensitivity to the distributional properties of multiword sequences, which is presumed to indicate that chunking strategies come
to the aid of WM given the memory constraints of online language processing. If the effective processing of information, linguistic or other, is limited by WM capacity (Cowan, 2001) and this in turn is easily overloaded by a constant stream of input, there must be complementary mechanisms that parse information into chunks that make post-hoc recall uncomplicated. By means of illustration, a child attempting to remember single-digit numbers in a long sequence by listening to them just once without pauses in between might only recall three or four elements in a correct sequence, depending on their WM capacity (Murdock, 1983). However, implementing chunking strategies would mean first packaging segments in that sequence into multi-component units (e.g., chunk 1 = number 378, chunk 2 = 489) and then storing those units for later recall. In this example, remembering only two multi-component sequences would probably lead to better recall of the whole sequence. In fact, a training study where an adult spent over 200 hours learning how to chunk numbers in a sequence as a mnemonic tool to segment and memorise digits reported an increase in memory span from 7 to 79 digits (Ericsson et al., 1980). Other studies on adults, such as the one by Wiechmann & Kerz, (2016), for example, concluded that adults, like children, are sensitive to distributional cues in syntax and rely on formulaic language as a support strategy that helps compensate high processing syntactic demands. Crucially, chunking depends on the strategic codification of the input into hierarchically organised levels of representation so that old information is not overwritten by the new input and the sequence chain is maintained (Christiansen & Chater, 2015).

Because, as mentioned, children are adept at recognising organisational patterns in the information received, if the nature of the input is characterised by marked distributional cues, then these are read as a blueprint of what can be chunked, and as a result, the processing becomes more efficient. Cowan, for example, has reported that when flexible input is processed by different sensorial systems, such as listening to and seeing numbers on a screen, it is typically more easily chunked than input containing less variability (i.e., only listening to the numbers). The reason for this, according to Cowan, is the role of attentional focus on the salient features of the information to be processed, as they assist the recipient in deciding what to chunk (e.g. aural and visual input) (Cowan, 2001).
Theoretically speaking, in the same way that differences in sensorial input can be major pointers that expedite the decision-making of what to chunk, aspects concerning the content of the input could also act as markers. This has been confirmed by Cowan et al. (2010). In one of several experiments investigating the developmental nature of immediate recall, Cowan and colleagues compared children’s (mean age = 8 years) and adults’ performance in recalling information they had been previously exposed to. During this phase of item familiarisation, words were presented either in isolation or as part of a pair, with the arrangement of the latter staying consistent throughout the experiment. Their results show that children benefited as much as adults from pair familiarisation when recalling this information. The more significant difference between age groups was observed in the recall of single units, suggesting that adults were more adept at processing unstructured information. However, when information was encoded following an organisational pattern, children improved their performance as much as adults improved theirs (within-age-group comparison).

In the case of SdV sentences, which are the focal point of this study, repetitions of the syntactic patterns that conform the distance could act as delimiters or organisational patterns signalling chunks, which would be used as a processing crutch by WM. The results of Cowan et al. (2010) indicate that children will make use of these processing advantages if the input exhibits such patterns. Therefore, understanding the extent to which SdV sentences contain repetitive patterns could serve to establish whether chunking mechanisms were involved in maintaining the elements between subject and verb mentally active. This, in turn, will inform the claims that were made and those to be made about how WM is taxed in sentences where SVA is suspended.

### 6.3 Methods

There are two key components in the methodological design of the study. The first concerns the Sankey representations of the English and Spanish SdV sentences, both used to visualise the syntactic distribution and highlight patterns of repetition. The second relates to the rationale behind using a measurement of node connectivity as an indicator of morphosyntactic repetitiveness.
Sankey representation

Sankey diagrams generally depict flows from a source to a target, highlighting the width of each flow as a representation of the significance or value of the connection (Nuttbohm et al., 2009). The resulting visual is a weighted network that can be used to trace the transitions from $n$ given sources to $n$ ends. While its adoption outside the energy industry is still relatively limited, for language studies, more specifically syntax-based studies, the Sankey visualisation could serve, among other things, to map the distributional dependencies between words within the constraints of the sentences they appear in. For the present study, the Sankey Diagram Generator (Csala, 2014) was used to model a directed acyclic graph that would illustrate the syntactic distribution of the SdV sentences extracted from the English and Spanish corpora.

To optimise the transposition of data (i.e., SdV sentences) to the Sankey Diagram Generator, a parser was custom-built in Python. The Sankey app uses input in JSON format and, while it provides users with an easy-to-use interface, the volume of data is too large for the entry to be undertaken manually. Instead, the parser was programmed to first sanitise each sentence, as its input is the output of the search string computed in each corpus (see section 5.3), and then recode the sentences into JSON. A detailed walkthrough is provided in Appendix 3 and the parser is made available as an open-source file in GitHub for future research projects interested in mapping syntactic distribution using Sankey diagrams.

This flow-like diagram is composed of three prime elements: layers, nodes, and links. Layers contain all words in the same position for any given sentence, whereas each node stands for one specific word, as long as the layer is different. Otherwise, they share the node, and links measure the connectivity of the node with its neighbours. In Figure 43, for example, there are seven layers in total, as marked by the number of words in the longest utterance, and one node for each existing word, except for $a$ in the first layer and that in the third layer, which contain two links each. The visualisation of these five sentences below clearly maps the independent and shared trajectories of the sampled sentences.
Figure 43. Sankey representation of five English CS sentences. Each layer is represented by the vertical alignment of coloured boxes, which stand for nodes. Links are horizontal lines connecting nodes.

**Node connectivity**

By focusing on node transit, one can determine the level of formulaicity of any sample speech. For the purposes of this study, this measurement of connectivity is used to quantify how repetitive SdV sentences are in terms of syntactic distribution. Frequently traversed nodes are larger in size and therefore indicate more syntactic repetition. To measure the connectivity of the whole graph, and thus the degree of repetitiveness of the SdV sentences, several considerations were made.

First, a consistent measurement of repetitiveness, applicable to small and large networks, was needed. This was crucial given the size difference between the English and Spanish data and the interest in testing cross-linguistic correlations. Figure 44 illustrates how this measurement was conceived through two sample sentences *The book is under the table* and *A book is under the table*. If these two sentences are assumed to be the whole sample corpus (Sent2), then the network representation would look like Figure 44.
Figure 44. Example of connectivity measurement using two sentences (Sent2). The values highlighted on the right represent the proportion of dummy nodes to the total number of tokens. Blue nodes indicate no repetition and red nodes +1 repetition.

In this sample, blue nodes are independent nodes and red nodes are slightly larger nodes showing repetition, one instance each. That is, in all five red nodes in total, there were five cases of repetition. In order to measure these instances, dummy nodes (white nodes below) were conceptualised as having zero values and used in the graph above to illustrate that every time a node is larger than one, a dummy node is created to record this. Therefore, five words (dummy nodes) were repeated in the two sentences. The table on the right summarises the frequency of occurring in the corpus and the percentage of repetitiveness (41.7%) given the total of tokens uttered (12 words).

**Frequency Table**

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sent2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid</td>
<td>.00</td>
<td>5</td>
<td>20.8</td>
<td>41.7</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>2</td>
<td>8.3</td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>5</td>
<td>20.8</td>
<td>41.7</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>12</td>
<td>50.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Missing System</td>
<td>12</td>
<td>50.0</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>24</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

**Sent3**

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>.00</td>
<td>10</td>
<td>41.7</td>
<td>55.6</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>3</td>
<td>12.5</td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>5</td>
<td>20.8</td>
<td>27.8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>18</td>
<td>75.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Missing System</td>
<td>6</td>
<td>25.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>24</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

**Sent4**

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>.00</td>
<td>15</td>
<td>62.5</td>
<td>62.5</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>4</td>
<td>16.7</td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>5</td>
<td>20.8</td>
<td>20.8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>24</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Figure 45. Example of connectivity measurement using three and four sentences (Sent3 and Sent4). Blue nodes indicate no repetition, orange nodes +2 repetition, and green nodes +3 repetition.

In Figure 45 there are two additional samples standing for two different corpora. The first one (Sent3) is composed by the sentences in Sent2, plus a new one *This book is under the table*. The second one (Sent4) includes Sent3 in addition to a fourth
sentence *That book is under the table.* In both cases inserting a new sentence has led to the increase, by one, of the same nodes that were shared in Sent 2. In other words, orange nodes show two instances of repetition and green nodes three instances. The dummy nodes above and below illustrate the degree of repetition and the frequency table on the right summarises the distribution of these values in their networks. Thus, in Sent3 there were 10 words repeated (dummies), out of 18 tokens uttered, and in Sent4 15 out of 24.

Second, to frame the actual values of repetitiveness (dummy count) in a scale, theoretical minimums and maximums were formulated to serve as limiters of how repetitive or non-repetitive the syntactic distribution of SdV sentences could have been. In the case of the maximum, the total of cumulative tokens for the whole age period was used. This served as a ceiling or theoretical maximum of repetitiveness because it signifies that the same word is being repeated over time. On the other hand, uttering a word without ever repeating it in the same syntactic position, serves to illustrate a theoretical minimum.

To sum up, the connectivity measurement consists in computing the frequency of dummy nodes in each network. Dummy nodes are created every instance a word is repeated in the same syntactic position, following the layered distribution of the Sankey. This dummy count is plotted next to theoretical maximum and minimum values of repetitiveness used as references.

### 6.4 Results

*Sankey*

To illustrate the Sankey distribution of SdV sentences, the 183 English sentences, where the SV distance was equal or larger than one syllable, extracted from the English corpora in Study 3 were parsed and shown in Figure 46. The high density of nodes and links shows how interconnected and complex syntactic networks are.
Figure 46 is a cumulative representation of SdV sentences over the 30-to-149-month period in question and serves to highlight several points about the syntactic spreading of these sentences. First, unsurprisingly, more popular nodes tend to occur in the first layers, as shown by the size of the coloured blocks. These mostly correspond to determiners such as *the* and *a*. As the flows move right, nodes tend to decrease in size which signifies more independent syntactic paths. The same is true for Spanish sentences, represented in Figure 47. The density is considerably higher in the case of the Spanish data as a result of a larger corpus sample (i.e., 840 sentences) and the distribution more elongated as Spanish SdV sentences are longer, as discussed in Study 3.
Additionally, at a first glance, the syntactic trajectories of SdV seem to show more independence than interdependence. That is, there are more smaller nodes than there are larger nodes, which would suggest that these are not essentially the same SdV structures being repeated continuously. However, in order to assess this, a valid form of connectivity measurement that is applicable to comparable data samples needs to be provided, which is why the connectivity of the graph was calculated using the dummy count explained earlier.

**SdV repetitiveness**

An automated dummy counter was coded in Python as a complement to the sentence parser created to transpose the SdV sentences into the Sankey representations. As discussed in the methods (section 6.3), the counter provided the ratio of dummies compared to the total number of tokens.

The line graphs (English - Figure 48, Spanish – Figure 49) show how repetitive English and Spanish SdV sentences (in blue) become as children grow older, from 30
to 149 months. The maximum (grey) represents a scenario where every token is repeated, that is, a situation where the child utters the same word over the whole period. The minimum (orange), on the contrary, signifies a situation where no word is ever repeated in the same position in their respective sentence.

Figure 48. Linear representation of repetitiveness of 189 English SdV (blue) sentences and 189 English non-SdV (yellow) sentences as measured by the dummy counter. Non-SdV sentences contain fewer words, hence the shorter trajectory.
Because the grey and orange lines offer theoretical maximums and minimums, a new variable representing real CS was introduced in the analysis. To offer a comparable sample of data, the same number of non-SdV sentences were extracted from the same corpora discussed in chapter 5 (sections 5.3.1 and 5.3.2 for English and Spanish, respectively) and analysed for repetitiveness. Both sets were matched for age, meaning that, for both languages, non-SdV sentences were uttered at the same point in time as SdV sentences. Due to the higher frequency of non-SdV sentences (see tables 2 and 4 in chapter 5) the selection pool was larger. Therefore, sentences were picked at random as no other matching criteria was needed.

The graphs in Figures 48 and 49 show how, for both English and Spanish, the SdV sentences and non-SdV sentences are somewhere between being repetitive and non-repetitive, following a progressive increase that represents approximately 40-50% of repetitiveness across all age groups in both English and Spanish sentences. More importantly, the trajectories followed by SdV and non-SdV sentences are virtually the same. It is worth mentioning that the short trajectory of non-SdV sentences (yellow) in both graphs reflects that fewer words were used in these sentences, but not less.
repetition. As discussed in sections 5.3.1 and 5.3.2, the distance in SdV sentences in English and Spanish, respectively, increases the length of an otherwise shorter sentence with no distance (e.g., *The movie is long* / *The movie on the telly is long*). Fewer words lead to fewer nodes.

### 6.5 Discussion

By way of reminder, the purpose of the analysis was twofold, [a] to depict and [b] to measure the syntactic connectivity of SdV sentences, as uttered by differently aged children with typical development. This adds to the existing understanding, fully covered in chapter 5, of how these types of sentences are constructed. By factoring in the qualitative aspect of SdV sentences this study is addressing the potentially confounding influence of recurrent formulaic patterns in the emergence of SdV sentences on account of WM involvement.

The predictions were that if the morphosyntactic composition of SdV sentences is shown to be built off repeated chunks, then it could be argued that the involvement of WM in these sentences is diminished by the effects of rehearsal and strategy. Accordingly, if the paths that follow SdV sentences are found to be significantly interconnected, this could suggest that the sampled SdV sentences are inherently formulaic, which limits the involvement of WM when children establish grammatical agreement over long distances. Conversely, if the syntactic distribution of SdV sentences is based on mostly disconnected, untraversed paths, this would support the case for ‘greater’ WM contribution to SVA, given the limited strategic resources for children to coordinate subject and verb without taxing their WM.

The results show no unusual iterative patterns in the morphosyntactic composition of the SdV sentences that would suggest that recall would be less complicated than in any other situation. This is evidenced by how, over the same 10-year period, morphosyntactic repetitiveness emerges at a similar rate in SdV sentences, whose grammatical agreement rely on WM, and non-SdV sentences, where agreement does not entail memory resources. As such, the explanation that SdV processing is formulaic to the extent that it does not implicate a role for WM can be ruled out, and
consequently, this further strengthens the general argument in favour of deep integration between domain-general processes and language.

It is worth pointing out that the qualitative analysis of morphosyntax discussed in this chapter is by no means comprehensive, as its conception was only complementary to the syntactic distribution analysis of the Sankey diagrams. A more systematic characterisation of the development of SdV sentences is needed, similar to those provided by Diessel & Tomasello (2001, 2005) in their studies of relative clauses, where elements concerning the grammatical role of the gap, or the communicative function of the entire utterance, for example, are considered. Additionally, the results apply specifically to the sentences analysed and are not meant to be generalised to other types of SdV sentences. This is because the subsets that integrate the SdV category are too varied so a case-by-case analysis of repetition of SdV sentences is necessary as there could be specific syntactic structures (e.g., relative clauses) that reuse patterns more than others (e.g., adverbial modifier, see Table 7 in Chapter 5).

What this study hopes to have contributed to future research, however, apart from the findings which are perhaps only relevant to this dissertation, is a full walkthrough of this novel methodological approach to measuring and visualising syntactic repetition, and the tools used to accomplish this (i.e., custom-built parser with a built-in node counter and a Sankey visualiser) which could be employed by future researchers interested in assessing language formulaicity.
Chapter 7 – General discussion and conclusions

7.1 Chapter overview

This dissertation has explored the interaction between WM and the syntactic processing of grammatical agreement through four different studies. Chapter 1 introduced the general theoretical framework underpinning the research questions that motivated each study. Chapter 2 provided a detailed review of the literature pertaining to [1] WM as a psycholinguistic construct - its structure, functions, development, assessment, and training; and [2] its relationship with SVA, as documented by previous research in the field. Chapter 3 reported a causal link running unidirectionally from WM training to SVA processing in Spanish (L1), and English (L2) as a subdominant language. Chapter 4 explored in more depth the online processing of the n-back used in the training study and found that gender played a significant role in individual reaction times to the task. Chapter 5 analysed corpus-based English and Spanish longitudinal data of naturalistic child language and established a positive correlation between the developmental trajectory of the SdV sentences extracted and non-verbal WM. Chapter 6 assessed the degree of formulaicity of these SdV sentences and found that they showed enough syntactic flexibility for WM to be required in their processing. This final chapter will summarise the main findings of the four studies and will integrate the results into a general conversation about the relationship between domain-general cognition and syntactic processing.

7.2 Introduction

This dissertation set out to examine the interaction between children’s WM and long-distance SVA. It did so through four empirical studies, whose research questions and findings are summarised in Figure 50. This visual representation of the entire dissertation was first introduced in Chapter 1 and is now revisited to include an abridged statement of the findings of each study and some of the key discussion points to be covered in the next sections.
7.3 Main findings

The original contributions of the dissertation can be summarised into four points, namely:

1- Non-verbal WM training improved L1 and L2 performance in long-distance agreement, but not vice versa.
2- The reaction times to the n-back task were bimodally distributed, and boys tended to respond earlier but girls were more accurate.
3- Children’s use of some sentences with long-distance agreement is correlated to the development of non-verbal WM.
4- The use of formulaic language in SdV sentences is predicted to have no unexpected influence as a support mechanism on WM processing.

In Study 1 (Chapter 3), the principal interest was testing four different hypotheses of bidirectionality of transfer between WM and the processing of SVA in SdV sentences, both in Spanish (L1) and English (L2). The links between WM and syntactic
processing had been thoroughly examined in Chapter 2, but the conclusion was that much of the evidence reviewed was of correlational nature. If the purpose of the study, and the dissertation, was testing the limits of the relationship between WM and syntax, on the assumption that language processing is deeply integrated with the rest of cognition, then using a training study was one methodological route for attesting causality. The results, in bright green (figure 1), showed that non-linguistic WM training led to gains in the processing of grammatical agreement among children’s first and second language, but the effects ran unidirectionally.

Study 2 served to make the most of data that had become available with the experimental study. Its contribution to the general understanding of the relationship between WM and SVA is more limited than that of Study 1, as it is solely focused on WM processing. However, the exploration of individual differences, by measures of reaction times to the n-back, offers new opportunities of correlational analyses in future studies interested in WM and SVA. For example, one proposal for a replication study could substitute the GJT's used in Study 1, which children took using pen and paper, with computer based GJT's that measure response latencies. This enables cross-comparisons between the gender differences already observed in the RTs to the n-back, and the potential findings in the GJT's testing SVA. The relevance of pursuing this route of inquiry will be explained in depth in section 7.3 of this chapter.

Study 3 offered an entirely new perspective to the types of interaction between WM and syntax that had been explored in Study 1. The arrow in figure 1 from [1] to [3] is used to indicate this shift, which, in fact, is multidimensional. First, the focus was now on longitudinal observations and not cross-sectional data, covering a larger period of childhood (2-12 years) for a larger sample of English and Spanish-speaking participants. Second, the study was designed to test developmental correlations between non-verbal WM and the processing of agreement in SVA, not causal relationships. Third, the results are based on analyses of language production, in the form of CS, not language comprehension, although inferences from the former to the latter are made. The primary finding was that the development of SdV sentences was contingent on the development of WM, as indicated by the positive correlations established.
Study 4 fits into the grand design of the dissertation as a complementary set of analyses conducted on the data obtained in Study 3. There were two focal points in the investigation of the SdV sentences that were extracted from the English and Spanish corpora, both concerned with the degree of morphosyntactic formulaicity of these sentences. The first one related to providing a visualisation of these sentences to highlight any potential anomalies or patterns in the syntactic connectivity of the sentences. The rationale was that higher sentence connectivity would suggest higher formulaicity, which in turn would imply that WM is less burdened because recall from LTM is possible. Sankey diagrams were deemed a novel yet fitting method for representing these patterns of emergence of language use and reuse. The second objective was measuring this connectivity so that a reliable index of formulaicity could be employed to support the findings. Eventually, the developmental trajectory of morphosyntactic repetition in SdV sentences and the one corresponding to non-SdV sentences were compared and no significant differences were found, concluding that SdV sentences exhibited no atypical formulaicity.

One major conclusion, included in figure 1, that can be drawn from interpreting the findings as one, but especially those of Studies 1 and 3 is that there seem to be important processing exchanges between syntax and domain-general cognition. This is evidenced fundamentally by [1] the causal link established between non-linguistic WM and the processing of grammatical agreement, and [3] the developmental correlation between the emergence of SdV sentences and the emergence of non-verbal WM.

7.4 Cross-study examination

Thus far in this dissertation all findings and their implications have been discussed within the context of their respective studies, as the intention has been to highlight their own merits as separate research enterprises. It is time now to consider these results in the broader setting of the dissertation.

Age-related performance
There are other multiple ways in which the four studies complement and contrast each other, one of which concerns age-related performance. In the case of Study 1, studying the possible effects of WM training on SVA for this particular age group (mean = 7 years) was motivated fundamentally by [1] the fact there is ample evidence in the literature suggesting that children aged 5-6 are able to establish SVA in English (Johnson et al., 2005) and Spanish (Pérez-Leroux, 2005); and [2] by the time children are 6-7 years old, the key components of the WM system and subsystems are already in place (Alloway et al., 2004). However, there was little evidence to inform the rationale as to what would be an adequate syllable distance for the distance between subject and verb. After all, setting the right level of processing difficulty (i.e., SV distance) was paramount for transfer of training to occur. If the task was undemanding, no new cognitive routines would be operationalised, and therefore transfer would not be predicted to occur (Gathercole et al., 2019a). If the task was excessively difficult, children would not be expected to be motivated to perform. Therefore, the optimal SV distance would be one right above ceiling performance for transfer of training to take possible effect (more on the following section Syllables, images, and chunks).

To assess ceiling performance, an internal pilot study was conducted prior to the experiment. As mentioned in Chapter 3, due to the novelty of the approach and the absence of referential studies in the literature, the pilot study was conducted specifically to gauge the level of difficulty of the GJTIs of the SdV sentences. The SV distance was settled at 16 syllables once the accuracy rate in L1 and L2 was approximately 50% of hits and the morphosyntactic complexity of the elements between subject and verb (i.e., controlling for vocabulary frequency and syntactic function of the modifiers) had been standardised.

Upon completion of both the experimental and the corpus studies, post-hoc analyses on the maximum SV distance in naturalistic CS sentences facilitated the corroboration of the SV distance used in Study 1. As mentioned earlier, the SV distance of the L1 SdV sentences used was 16 syllables on average (e.g., sentence 76, below), and the maximum SV distance extracted from the Spanish corpora for the 7-year mark was 10 syllables (e.g., sentence 77). This means that the L1 SdV sentences
used in the experiment were above the ceiling performance of TD 7-year-old children, and this provided ample space for training to improve performance.

It is unclear to what extent adjusting the SV distance closer or further from the 10-syllable benchmark would impact the results obtained in the training study, or whether an ‘optimal’ range could be established. Additionally, as L2 processing was a secondary objective, the relationship between SV distances was not verified using bilingual corpora, but the same procedure followed in L1 is encouraged in future studies.

In the case of age-related performance in the n-back task, the primary focus was on the responses to the 2-back in view of the normative data collected by Pelegrina et al. (2015), reviewed in chapter 3, suggesting that this modality explained typical 7-year-old performance best. This is because, similar to the SV distance of the GJTs in Study 1, this level of difficulty is believed to be most appropriate to investigate brain-behaviour relationships. The results of a recent investigation of individual differences in WM function by Lamichhane et al. (2020) can provide an indirect explanation as to why this is the case. In their study, they tested adults’ processing abilities of six different parametric levels of an n-back, divided into low cognitive loads (1- to 3-back) and high loads (4- to 6-back). After comparing participants’ responses in both modalities to independent measures of left lateral prefrontal cortex activity, frequently associated with WM activity, they observed that individuals exhibited greater individual variability in both behavioural performance and brain activity in high-load situations than in low-load conditions. Behavioural data relied on target accuracy and RTs while brain activity was measured via fMRI. It should be noted that the size of the effects of the correlation between RTs and blood-oxygen-level-dependent imaging, used in the fMRI, were considerably larger than those found

---

6 The verbatim translation is respecting the sequence of the constituents in Spanish as much as possible. The purpose of the translated version, in this case, was to highlight the distance between subject and verb in the original, not to offer a bona fide equivalence in English.
between the latter and target accuracy. The clarification highlights the disassociation between target accuracy and RTs that was discussed in Chapter 4, and how this warrants the inclusion of both measures in studies focusing on differentiating individuals’ WMs.

The results reported by Lamichhane et al. (2020), when interpreted in tandem with those by Pelegrina et al. (2015), seem to indicate that n-backs with WM demands near and above ceiling performance (e.g. 2-back for 7 year olds) offer a more suitable pathway to the study of individual differences than n-backs that are less taxing (1-back for the same age group). One caveat to this claim is that the first study listed adult participants and the same finding is yet to be replicated in children. However, there are two well-defined bodies of evidence discussed here, one supported by normative n-back data covering large periods of childhood until late adolescence (Pelegrina et al., 2015; Yaple & Arsalidou, 2018), and the other documenting the developmental changes associated with WM gains during childhood and early adulthood (see section 2.3), that lead to the same presumption, namely that the processing of a 2-back by 7 year olds is generally comparable to the processing of a 4-back by adults. It seems to be that case that brain-behaviour relationships seem to be stronger when the cognitive system is overloaded (Van Snellenberg et al., 2015), and it is exactly this overload that leads humans to create new cognitive routines to complete a task that it is out of the reach of the mechanisms that are readily available (Gathercole et al., 2019). In other words, because the gender effects in RTs reported in Study 2 were more significant in the 2-back than in the 1-back, and the existing literature suggesting that high-load conditions reveal more about age-related performance than less challenging ones, it can be argued that the ‘threshold of difficulty’ for transfer to occur in Study 1 was especially met through the 2-back modality and the 16-syllable SV distance. The latter condition is predicted to have caused a similar cognitive overload to that caused by the 2-back, considering that the ceiling performance for production in that age group is 10 syllables, as reported by Study 3.

Syllables, images, and chunks
As discussed in the previous section, the SdV sentences in the experiment were controlled for syllable length, both for the ‘distance’ and the entire sentence, following the results of the pilot in Study 1. In light of the results and commentary provided in Studies 3 and 4, post-hoc analyses may prompt alternative explanations of the distance between subject and verb in SVA and how it compares to the gap of the 2-back and multi-component chunks. As mentioned before, the results showed that training in 1-back and 2-back led to improved performance in SVA in the SdV sentences. The distance of the gap, in syllables, was 16 on average in Spanish (L1 – sentence 78) and 4 in English (L2 – sentence 79). Therefore, in Study 1, the demands associated with processing the 2-back, as the higher-load condition, posed a similar challenge to the processing of the 16 syllables in Spanish (L1) and the 4 syllables in English (L2). However, while the involvement of chunking strategies was accounted for to some extent, it remains unclear how multicomponent chunks compare to these two other measures (i.e., syllables and images).

While syllables are an apt metric for measuring SV distance given most known languages can divide words into syllables and these have been linked to memory processing, there is no simple way of delineating where one multiword chunk begins and another ends in the sample sentences used in the experiment, especially in the L1 where the 16 syllables can be divided into two, three or four prepositional or adverbial phrases.

78- Los padres del compañero de clase de historia de Julito fueron a la reunión.
79- The parents of the History classmate of Julito went to the meeting

The balls under the chair are yellow.

Figure 51 depicts two different conditions comparing the WM loads of the n-back and the syllable distance between subject and verb. In Study 1, cognitive transfer of WM training is theorised to have occurred because the WM demands of the 2-back and the 16-syllable-long distance in grammatical agreement were near and above the ceiling performance of 7-year-olds, which led them to establish a new cognitive routine to cope with the processing burden. Therefore, matching the level of

---

7 The verbatim translation is respecting the sequence of the constituents in Spanish as much as possible. The purpose of the translated version, in this case, was to highlight the distance between subject and verb in the original, not to offer a bona fide equivalence in English.
difficulty posed by the syllable distance between subject and verb with the number of intervening images of the n-back was crucial for transfer to occur in Study 1, but it remains unclear to what extent chunking mechanisms were used in the experiment, and therefore of interest for future studies.

Figure 51 – Because the conditions in A are less strenuous for WM than those in B, one prediction is that children aged 7 can solve the processing demands of the task with the cognitive and linguistic resources at hand. In B, however, a new cognitive routine is believed to have been created as the existing ones cannot meet the demands of the tasks, hence the disrupting Xs.

7.3 Limitations of the results and further research

Near transfer or far transfer?

The results of the experiment in Study 1 show strong support for the Gathercole account of cognitive transfer, namely transfer should occur when both the trained and untrained activities impose the same unfamiliar task demands that are not supported by existing subsystems (Gathercole et al., 2019). It also shows that, in contrast to meta-analyses failing to find effects of transfer of working memory to measures of intelligence (Melby-Lervåg et al., 2016), far transfer can and does occur if the match between trained and untrained arms of the study demands the same functional resources; the demands of the task are not met by existing resources and there is enough room below ceiling performance for capacities to be trained. Undisputed examples of near-transfer effects are abundant in the WM literature, most notably between tasks sharing the same paradigm (e.g. n-back and arithmetic updating, Linares et al., 2019). Although less documented (Melby-Lervåg & Hulme,
An alternative explanation of the differences between near- and far-transfer effects in the WM training literature has been offered by Swanson & McMurran, (2018), whose own experiment found that near- and far-transfer effects of WM training were moderated by individual differences in fluid intelligence, so controlling for Gf could, according to the authors, explain the distance of the transfer effects. Also of interest for future research is the difference in transfer effects with different age groups, when controlling for difficulty. It should be noted that this dissertation does not address the extent to which these effects can be observed across a life span or at earlier or later stages of childhood. As discussed in Chapter 2 (section 2.3), 7-year-old children were sampled because several developmental milestones that are speculated to increase the chances of transfer to occur have been reached by this stage. Most notably, the levels of plasticity are known to decrease significantly after the critical period during childhood, although notions about plasticity in adults have gone through important changes in recent years, (for recent developments in plasticity in adults see Reh et al., 2020; Voss et al., 2017). Nonetheless, the type of transfer detailed in Study 1 may well depend on the high levels of plasticity exhibited by children during this sensitive period and significant fluctuations on these levels could constraint the formation of the novel cognitive routines needed for transfer to occur.

Another point of interest which moderates the strength of the transfer effects, and which could be further investigated, is that concerning the non-linguistic conditions of the WM training task. As discussed in Chapter 3, the irregular polygons used in the transfer study are meant to discourage participants from using linguistic labels and thus maximise the chances of disassociating visuospatial processing from phonological processing. However, it is impossible to rule out that visual stimuli are not being verbally coded, and as documented by Souza & Skóra (2017) albeit on adult populations, verbal labels help activate long-term categorical information which in turn boosts the retrieval of information. It is unlikely, though, that in this experiment
26 different children created their own working labels for eight irregular polygons to relatively equal success. In fact, a shape-by-shape accuracy analysis of several trials – included in chapter 4 – showed no differences in how participants processed all eight figures, which would suggest that all eight labels had to be employed similarly. A recent study of individual perceptions of visual complexity provides a comprehensive analysis of geometric shapes such as the ones used as stimuli in the n-back task. There, Sun & Firestone (2021) show how adults’ descriptions, in English, of similarly shaped polygons are considerably long and unsystematic. Previously, Lewis & Frank (2016) had shown that participants tend to map longer words to more complex objects in production and comprehension tasks, which means that the verbal labels debated here, if any, would not be predicted to cue visual retrieval efficiently.

In any case, further testing is warranted to determine whether the effects observed in Study 1 are transient or have long term implications, and whether WM training transfers to syntax in different age populations. In the case of the former, it should be noted that extended follow-up measures to many of the near and far transfer conditions reported here have noted that these effects are often not sustained over long periods (see meta-analysis by Schwaighofer et al., 2015).

**Unidirectionality of the transfer effects**

If there is such integration between domain-general cognition and language, the prediction would have been that the training effects would run in both directions, and yet only WM training led to transfer. One possible explanation is that the n-back trained a higher-level system than what was required for the long-distance agreement. As discussed elsewhere in this dissertation, the n-back task has shown inconsistent correlations to other ‘purer’ WM paradigms as the resolution of the task seems to depend on inhibitory mechanisms of control and other attentional resources. Therefore, the unidirectionality observed in Study 1 could reflect transfer of training occurring from the more general cognitive system to the more specific, but not vice versa. While both the n-back and the processing of long-distance agreement rely on the encoding of item and order information and the engagement of a maintenance rehearsal process, the former depends more on general cognitive resources than the latter (Gathercole et al., 2019). This could be one of the factors
that facilitates the development of new cognitive routines, as predicted by Gathercole and colleagues (2019, p.24) “...training on visuo-spatial STM should generate greater transfer to untrained tasks with similar recall demands than the corresponding verbal STM tasks”.

One limitation of using pen and paper for the GJT in the transfer study discussed in Chapter 3 is that it complicates the disentanglement of the interaction between WM and SVA during the processing, and this provides an alternative explanation to why transfer effects were not observed in both directions. Because the GJT is an offline task, the decision of how (un)grammatical the sentence is, is made post completion, so there is no real time measure of the processing of the SdV sentence, and therefore no advantageous window into the cognitive mechanisms involved in the resolution of the grammatical agreement.

Further research is required with a wider range of language (morphology, argument-structure constructions, phonology) and cognition (inhibition, attention, analogy) training measures and a wider age range to fully understand the unidirectionality of the effect for this group.

**Grammaticality effect**

Post-hoc investigations revealed a significant role of grammaticality in the effect of training. Language training improved the accuracy of identifying grammatical sentences (M = 68%, SD = .21) more than it did ungrammatical ones (M = 46%, SD = .23), t(102)=4.91, p<.001. Also, working memory training improved the accuracy of identifying grammatical sentences (M = 75%, SD = .17) more than it did ungrammatical ones (M = .37%, SD = .37), t(102)=9.85, p<.001.

In sentences of the type used in the experiment in Study 1, verbs are thought to trigger retrieval cues such as the grammatical number of the candidate noun phrase it is supposed to agree with as well as its case or syntactic position (Badecker & Kuminiak, 2007). The efficiency of this process is determined by whether there are encodings that match cues, how strongly they match the cues, and how uniquely they do so (Naime, 2002). The greater cognitive cost of processing ungrammatical
sentences caused by the mismatch means it likely that participants had a harder time in general processing ungrammatical sentences, and it may be that starting from a lower base requires more and longer training to achieve the same training effects as the boost to grammatical sentences. Of all sentences in both L1 and L2, all but one local noun appeared in the singular form in this experiment, which is likely to have reduced the over agreement error rate, but unlikely to have caused systematic bias between grammatical and ungrammatical sentences or driven the main effect of transfer.

Whatever the precise mechanism by which transfer occurred, the result was that by improving children’s non-linguistic WM they improved their ability to succeed on a linguistic task that also required WM. In the case of L1, this effect was additionally supported by an individual differences analysis – the children that scored highest on the WM training were also the ones that stood to benefit most in their syntactic performance improvement. As others have argued for in populations with Developmental Language Disorder and Autism Spectrum Disorder, WM limitations are likely to impact the acquisition of complex syntax and improving WM capacities via a dedicated training program could free up cognitive resources to deal more effectively with syntax (Delage et al., 2021a; Delage et al., 2021b; Stanford, et al., 2019).

**L1 versus L2**

As covered in chapter 3, the unidirectionality of the transfer effect reported in the experimental study was most striking in the case of L1, where WM training actually boosted syntactic performance as much as linguistic training did. The same pattern was replicated in L2, but here the boost in performance was less pronounced; language training improved syntax more than WM training, but those who received WM training still outperformed those in the control group. A possible reason for this is that the effects of cognitive transfer are stronger in the L1 training condition where WM abilities have greater room for development than L2 learners who recruit WM resources more often and in greater depth.
And, or it could be the case that despite the best efforts (*L1 and L2 Balancing*, Methods), it was simply not possible to match L1 and L2 on all relevant aspects (or indeed know what all the relevant measures were in advance). For example, while the mean syllable distance was set lower in L2 when compared with L1 (for proficiency reasons), it could have plausibly taxed WM in L1 more despite error rate matching (see Methods) and thus trained a greater potential spill over effect. The morphological component of each language is also different, with Spanish being richer than English, and this has been shown to impact how SVA dependencies are processed (Acuña-Fariña, 2018). Note however, differences between L1 and L2 do not invalidate the main effect of cognitive transfer, nor that it was witnessed in one direction rather than the other. Rather the between language condition differences point to the fact that this effect was demonstrated more clearly in L1 than it was in L2.

**N-back accuracy**

The accuracy measurement for the n-back task used in the experimental study (Study 1 in Chapter 3) offers, arguably, less nuance of performance than other alternatives. For example, signal-detection measures such as $d'$ are based on a proportion of total hits minus false alarms which address differences between omission and commission errors. This has been reported to offer a more reliable measurement of performance than a global measurement of overall accuracy, as it is more sensitive to the types of participants’ processing strategies (i.e., avoiding false alarms) that could misrepresent improvement. Furthermore, distinguishing between omission (missing a target) and commission (false alarm) errors has enabled researchers in the past to seek correlations between accuracy and reaction times to n-back tests. For example, Meule (2016) has reported positive correlations between omission errors and RTs in n-back tests but not between commission errors and RTs. However, prior to this study, Oberauer (2005) had failed to find correlations between omission errors and commission errors or RTs but did report associations between the former and measures of working memory capacity. What these two studies illustrate is that the construct of accuracy in n-back tasks is prone to being misrepresented lest more performance data (accuracy and RTs) from studies that control for different task conditions (single versus dual), modalities (visuospatial, auditory), and cognitive
loads (1-back, 2-back, etc.), for different populations becomes available. Even papers such as those by Jaeggi et al. (2010) and Kane et al. (2007) that have accounted for most of these conditions have only tested adult populations. Therefore, due to the lack of training studies with children populations distinguishing between misses and false alarms and the absence of correct rejections in the experimental conditions of this study (i.e., pressing a different button to indicate non-targets), the decision was to report general accuracy rate, which included all three types of responses, and provide a more nuance analysis of hits vs. false alarms for Study 2 which focused on individual differences.

*Lure performance in the n-back task*

Lure trials are often employed in n-back tasks because they provide more details about conflict resolution than using solely match non-match trials. Lures are strategically placed to match previously presented stimuli in a mismatch position, such as two identical images being shown one after the other, which would be a match in a 1-back, in a 2-back. Study 1 does not provide an analysis of lure performance for a couple of reasons. First, half of the study relies on 1-back trials, where lures are irrelevant. Second, studies in lure performance have referred to cognitive control skills that differ from inhibitory control in the way they assist conflict resolution. Even when SdV sentences are difficult to process because they interrupt syntactic agreement their comprehension does not depend on conflict-control procedures (Farmer, Misyak, & Christiansen, 2012). Training n-back with lures, on the other hand, is theorised to improve the type of cognitive control required in processing garden-path recovery sentences (Hussey et al., 2017) such as [80] but not SdV sentences such as [81]. Therefore, using lures in this study is not compatible for the two structures assessed here (n-back vs. SdV agreement) as the goal is not to test how cognitive control helps to resolve competition.

80- While the thief hid the jewellery that was elegant and expensive sparkled brightly
81- The jewellery on display at the museum in Central London sparkled brightly

*Active vs. passive control*
Adopting an active or passive control group in studies of cognitive training has become a common cause of disagreement over the last years. It has been argued that passive controls (i.e., no-intervention groups) are unfairly disadvantaged in comparison to training groups as placebo effects and expectancy effects (i.e., the expectation that a treatment will improve one’s performance), which are the standard in medical trials, for example, are not taken into consideration. For the field of cognitive training, the adoption of active control is still not generalised and under strenuous scrutiny.

Evidence in favour of the insertion of active controls in WM and n-back training studies stems mostly from studies, or meta-analyses of these studies, conducted with young adults (Dougherty et al., 2016; Foroughi et al., 2016). Boot et al. (2013) summarises the views shared by many advocates for the methodological implementation of active control groups. They believe that only when participants’ expectations of improvement are matched with those of the intervention group, can conclusions be drawn about the efficacy of the treatment, notwithstanding if is meant to promote mental health, cognition, or educational achievement. Lange & Süß (2015), for example, followed this approach in a cognitive training program with older adults. In their study, training older participants’ WM (mean age = 67.97) resulted in post-test gains but no transfer effect was observed. That is, while training effects took place, the performance of the experimental group increased as much as that of the active control group, thus implying that training gains occurred due to unspecific reasons.

This finding on adult populations has been challenged by studies such as Tsai et al. (2018), specifically designed to test placebo and expectancy effects from training young adults in a visual n-back to an untrained auditory n-back. In Tsai et al. (2018), four experimental conditions were tested: positive expectancy, negative expectancy, active control with positive expectancy, and active control with negative expectancy. The positive-negative expectancy was based on participants’ realisation that training could improve their performance in both trained and untrained tasks (i.e., positive beliefs about cross-modal transfer) or solely in the trained task (i.e., negative beliefs about cross-modal transfer versus uni-modal transfer). In the post-test participants showed improvements in the non-trained n-back task, regardless of expectations and
active controls had no significant gains over passive controls. Even in recent, broader non-training studies focusing on cognitive enhancement, placebo and nocebo effects have been linked to perceptions of self-improvement but not to actual cognitive gains (Winkler & Hermann, 2019).

The evidence is scarcer in training studies with children and there are several important distinctions to be made between using active or passive controls in studies with adults or children. First, one of the recommendations when controlling for placebo and expectancy effects is the need to avoid overt and suggestive recruitment. This, however, applies to a much lesser extent in children, and does not apply at all to Study 1. As noted by the briefing documents in Appendix 1, the information about the goals of the project was kept to a minimal and no discussion of expected outcomes was provided.

A final point about active vs. passive control groups is illustrated by a recent meta-analysis published after the experimental intervention of Study 1 was conducted. To test if engaging in an alternate intervention (active control) would yield significant results in studies of cognitive training, Au et al., (2020) conducted two separate meta-analyses. The first one surveyed the existing literature and analysed 34 studies that used one form of control. The second, however, synthesised data from 42 empirical studies that used both forms of control simultaneously. Both meta-analyses showed no meaningful difference between both types of control which could suggest that these effects, if anything, are more subtle than thought and not generalisable. In the meantime, n-back training studies comparing the performance of training and (passive) control groups where transfer effects have been observed continue to be published by top-tier journals (e.g., Li et al. (2021)

Nevertheless, if part of the positive effects of WM training on SVA processing in Study 1 were due to intensive computer-based practice, or other non-specific intervention effects, including placebo and/or expectancy effects, future studies are encouraged to test this hypothesis. As mentioned in chapter 4, employing computer-based GJTs, as opposed to pen and paper, would have enabled additional comparisons to the RTs of the n-back training group. By measuring the latencies of the responses to the GJTs not only do we gain access to the online processing of SVA...
in-situ, but it opens up investigations on individual differences in RTs, which is especially relevant after gender effects were observed in the RTs to the n-back task. While these were avenues worth of exploration, the potential benefits of controlling for placebo and expectancy effects for children this age, under the conditions of this WM training paradigm, were not deemed important enough. This was especially true, not only because of the mixed experiences reported in the literature, but because of methodological constraints. Unfortunately, access to the laptops used in the intervention and the proprietary software needed for the n-back was limited, therefore it was logistically unfeasible to guarantee computer access to any practice outside of the n-back training.

**Limitations of corpus data**

While corpus work enables researchers to understand the emergence of language in naturalistic settings, for example, there are clear limitations to the use of this data. Unlike experimental conditions which can be strictly controlled, the sampling density of corpus data is often sparse, consisting of a few one-hour-long recordings for vast developmental periods. Furthermore, in the case of Study 3, covering such an extended period of childhood (2-12 years of age) precluded the longitudinal observation of the same speakers, which would allow claims regarding individual performance. The fact that two (Spanish) and three (English) different corpora had to be collated for this age span meant that it is not possible to provide an accurate depiction of the developmental trajectories for each participant. These are only snapshots in time of different speakers, therefore the strength of the correlations established between the development of these SdV sentences and non-linguistic WM development is weakened by the nature of the data extracted from corpora. However, the main interest of the study is not generalising these findings in isolation, whether they pertain to WM or SVA, to these participants, but rather providing developmental comparisons between WM and naturalistic SdV sentences for an unusually extensive age period.

**Error analysis**
In the case of SdV sentences, estimates of frequency of occurrence in the CHILDES corpora were calculated and the results show that sentences with long SV distance were rarely used by children. One question that stems from this observation of frequency is the proportion of SVA mistakes found in the CS SdV sentences produced by the same speakers. Conducting an error analysis would make important contributions to the literature reporting multiple sources of semantic and morphological interference in grammatical agreement in different languages (Severens et al., 2008; G. Vigliocco et al., 1995) and is therefore encouraged in future studies.

**Qualitative analysis of the ‘distance’ between subject and verb**

In the case of Study 4, it is worth clarifying that no qualitative analysis of the distance in SdV sentences was conducted. The analysis was in fact of SdV sentences in their entirety. Additionally, the morphological and syntactic complexity of the elements contained between subject and verb is never assessed, as the main emphasis was on ‘syntactic distance’. Several studies have distinguished between distance and complexity when analysing centre-embeddedness, such as the studies on children’s acquisition of relative clauses by Diessel & Tomasello (2001, 2005) and the activation-based model developed by Vasishth & Lewis (2006) and reviewed by Nicenboim et al., (2015). The latter has specifically tested processing speed/accuracy trade-offs in dependencies with the same distance and showed that the materials in the gap and the relation of gap fillers to the head (subject) affect the resolution of matching head-argument. This, too, is encouraged in future studies in order to provide a more comprehensive characterisation of the SdV sentences and the WM costs associated to their SVA.

This is because two sentences with the same syllable SV distance could present different WM challenges based on the complexity of the elements intervening between subject and verb, as shown by the sentences below. One adverb (underlined), with a total 6 syllables, separates subject and verb in sentence [82], while a short clause with a local interfering noun stands between subject and verb in sentence [83].
7.4 Final remarks

This dissertation documents the main research activities carried out over the past three years. It started with a general enquiry about language processing and cognition which was later fine-tuned into what is now RQ1. The study, at that time, was designed so that the experimental intervention discussed in Chapter 3 was followed by a second intervention 6 months after. The main purpose of this second intervention was to assess, among other issues, the longevity of the transfer effects, if any, of the first experiment. Due to the impact of the pandemic, these plans had to be revised and the decision made was to adopt a strategy that would rely on secondary data. The aim was still to complement the cross-sectional results of Study 1 with longitudinal data, and this became a possibility with the corpus study (Study 3) which was based on the CHILDES database. Studies 2 and 4 complemented the initial experiment and the subsequent corpus study by providing detailed, analytical insights into [a] individual differences of WM and [b] syntactic formulaicity as a WM processing crutch, respectively.

Together, the four studies have explored and tested the limits of the relationship between children’s WM and their ability to process and establish SVA. On the one hand, the focus on non-linguistic visuospatial WM tasks, with special emphasis on n-back tasks, as examples of domain-general cognition was motivated by both the vast literature documenting WM involvement in language processing – a comprehensive review is provided in chapter 2 – and the lack of published studies interested in pushing the understanding of this involvement on syntax. The development of the WM system seems to regulate, to some extent, children’s ability to establish SVA and this is theorised to remain the case until late adolescence, when WM performance in complex tasks reaches adult-like levels.

On the other hand, the interest in SVA stemmed not only from it being widely considered a ‘core’ grammatical process and present in many languages, but because it is also a milestone in grammatical development. This ability is acquired at different
times for different languages and at different times within a language for different aspects of subject-verb agreement (e.g., person, number, gender) (Gxilishe et al., 2008; Johnson et al., 2005; Pérez-Leroux, 2005). Although the precise age of acquisition is often contested, many of the explanations given for the shape of development are linguistic in nature, for example the poverty of the English present tense agreement system (Johnson et al., 2005); difficulty of interpreting verbal agreement markers due to their formal nature (de Villiers & Gxilishe, 2008); the challenge of mastering the mapping from syntax to semantics (Pérez-Leroux, 2005); and differences in processing facilitation of null compared to overt marking (Pérez-Leroux, 2005). From the results discussed here, the non-linguistic influence of WM in shaping the course of subject-verb agreement acquisition can be added to this list of linguistic factors.

The image of a deeply integrated cognitive system where linguistic processing is more heavily dependent on the WM resources available than previously thought is becoming more vivid in recent years. This is evidenced in mathematical word-problem-solving accuracy in L1 and L2 (Swanson et al., 2021), speech fluency and phonological ability (Howell et al., 2020), language and drawing interactions in toddlers (Panesi & Morra, 2021), and in sentence processing in L1 and L2 (Andreou et al., 2021).

As stated in the introduction, two different views of what language is and how it is acquired both agree that language interacts with domain general cognitive processes at some level (Tomasello, 2003; Bybee, 2010; Goldberg, 2005; Dabrowska & Divjak, 2015, 2019; Ibbotson, 2020; Chomsky, 1980; Fodor, 1983; Hauser, Chomsky, & Fitch, 2002; Pinker & Jackendoff, 2005). The debate is over the extent of this interaction, the causal direction, the developmental periods in which it occurs, and the cognitive domains in which it happens. The results contribute to this debate by showing deep interaction between WM and syntax in children and that the causal direction is from WM to syntax.

The findings add to a growing body of research that shows widespread integration between language function and general cognitive processes, for example, inhibitory control and the acquisition of morphological inflection (Ibbotson & Kearvell-White,
domain-general categorisation processes and the acquisition of grammatical constructions (Goldberg, 2005; Ibbotson, Theakston, Lieven, & Tomasello, 2012); social cues and grammatical subject acquisition (Ibbotson, Lieven, & Tomasello, 2013); executive function and the acquisition of parsing strategies (Woodward et al., 2016).

If future research continues in this trend, then the language-unique parts of language are in retreat to such an extent that it becomes more parsimonious to use the framework of general cognition to understand linguistic processes, rather than call upon processes unique to language (see Ibbotson, 2020 for review). The implications of this impacts both developmental theory and applied practice. Theoretical accounts that emphasise deep integration between language and cognition see language acquisition as a process of repurposing general cognition for linguistic ends. Experience with and sensitisation to language input starts prenatally and continues at birth (e.g., Abboub, Nazzi T, Gervain, 2016). In addition, over the first year of life, before children are fully productive with language or can fully comprehend the communicative intentions of others, there are a range of categorization, memory, inhibition, analogy, attention, and social cognition skills also emerging in development, and these resources are available to the child at the time or before they are beginning to use language more productively. Under theoretical accounts that propose language-unique structure and processes, there is no serious role for domain-general cognition in constructing language itself, rather, developing cognition applies a performance filter through which the encapsulated, and largely non-dynamic, language competence emerges. Whether it is cognition ‘all the way down’ or whether there are processes unique to language is a question future research can continue to address as the answer informs our understanding of the nature of language and how it is acquired.

For applied practice, deep integration between domains holds the promise that interventions targeting one domain will spill over into the untrained domain, and ultimately improve developmental outcomes for children. To the recent evidence that has demonstrated the efficacy of this approach for children with Developmental Language Disorder (Delage et al., 2021b; Stanford, et al., 2019) and children with Autistic Spectrum Disorder (Delage et al., 2021a), this research adds evidence from a
typically developing population and a strong causal role of WM in syntax (see Delage et al., 2021a for the lack of control group).

Hitherto, this chapter has reviewed the main findings of the four studies, their respective limitations, and their co-existence in this dissertation. Taken together these data represent some of the most comprehensive and strongest evidence to date that that domain-general cognition is deeply integrated into developmental language processes, as the DCL approach would predict (Tomasello, 2003; Bybee, 2010; Goldberg, 2006; Ibbotson, 2020). The outset of the driving question of this research project was initially framed in a contrast between integrationist accounts and encapsulated or modular views of language. While these views used to reflect radicalised positions in the field, the dichotomy is less pronounced nowadays because of advances in the understanding of brain and language stemming from experimental psychology, neuroimaging, computational linguistics, evolutionary biology, etc. This makes the relationship between working memory and language during development interesting in and of itself. However, modular ‘strongholds’ still exist, mostly in the northeast of the United States of America, and the influence of the views of Fodor and Chomsky on future generations of psychologists and linguists is by no means insignificant. Whether conventionally subscribed or not to theory of modularity, there are many who still reject that working memory plays such a critical role in syntax processing, alluding to many reasons among which feature more prominently the apparent absence of far-transfer effects documented in the literature (Soveri et al., 2017), the loose experimental conditions and lack of scientific rigor in many WM training studies (Melby-Lervåg & Hulme, 2013), and the conflation of WM with other aspects of domain-general cognition (Shain et al., 2021). Therefore, asserting that the dichotomy between an integrationist view of language and domain-general cognition and one that sees both as being more disconnected is outdated is perhaps premature.

The implications of the findings are far reaching as they address language comprehension and production in both formal instruction and naturalistic settings. While variability control is more challenging in the context of a school, the results are more ecologically grounded as they account for more external factors that affect internal processing, and as such become more accessible to teachers, researchers and
educational policy makers in the development and adaption of materials for English for Speakers of Other Languages (ESOL) courses. This was evidenced by the fact that English teachers acted as co-researchers in the experiment without receiving special or intensive training. The experiment offers a full tour of the implementation of an effective age-specific training game that leads to significant in-situ gains in WM and L1 performance and promising results in L2 performance. These are not strict prescriptions directed at teachers, but rather a collaborative demonstration of a well-defined, theory-informed experiment of language learning conducted within the environmental constraints and richness of the context. Considering that most teachers are not second language acquisition researchers, it may seem counterproductive to resort to language acquisition research executed in a laboratory as source for teaching methodologies. While the debates surrounding modular versus integrationist accounts are less fervent these days, generative second language acquisition research is still presented to many ESOL teachers, for example, as a body of evidence to draw on for classroom practice (Whong et al., 2013) which is problematic. To deter language pedagogies from basing their practice on outdated theories of language learning, more empirical studies such as the ones discussed here are encouraged for other age groups. There is enough reason to believe that the design of future syntax interventions could incorporate more insights from the memory-based explanations provided here, especially in view of the correlations established between WM and syntax for such an extensive period of childhood. In any case, future explorations of many of these developmental issues will undoubtedly progress the understanding of the aspects of WM and domain-general cognition that are most relevant to language learning and development.
References


Casey, B. J., Cohen, J. D., Jezzard, P., Turner, R., Noll, D. C., Trainor, R. J., Giedd, J.,


Cowan, N., & Alloway, T. P. (2009). The development of working memory in childhood. In M. L. Courage & N. Cowan (Eds.), The development of memory in...
infancy and childhood. Psychology Press.


for higher reaction time variability for children with ADHD on a range of
cognitive tasks including reward and event rate manipulations.


*Science, 208*(4448), 1181–1182. https://doi.org/10.1126/science.7375930

diversity and its importance for cognitive science.*
https://doi.org/10.1017/S0140525X0999094X

diversity and its importance for cognitive science.*
https://doi.org/10.1017/S0140525X0999094X


Familiarization and Type of Discrimination Task. *Child Development, 45*(2),
351. https://doi.org/10.2307/1127955


Errors: When Cognitive Overload Enhances Subject–Verb Agreement Errors. A
Study in French Written Language. *The Quarterly Journal of Experimental
Psychology Section A, 47*(2), 437–464.
https://doi.org/10.1080/14640749408401119

Memory and the Establishment of Filler-Gap Dependencies: Insights from ERPs
https://doi.org/10.1023/A:1010447102554

https://doi.org/10.1126/science.1089401

https://doi.org/10.1016/j.cognition.2005.02.005

https://doi.org/10.1093/brain/124.5.849


https://doi.org/10.1515/cogl.2005.16.2.313


transfer effects of working memory training: A comparison of two programs focused on working memory updating. *PLOS ONE, 14*(2), e0211321. https://doi.org/10.1371/journal.pone.0211321


Minear, M., Brasher, F., Guerrero, C. B., Brasher, M., Moore, A., & Sukeena, J. (2016). A simultaneous examination of two forms of working memory training:


https://doi.org/10.1080/09658219408258959


https://doi.org/10.1016/B978-0-12-009735-7.50007-4


Pelegrina, S., Lechuga, M. T., García-Madruga, J. A., Elosúa, M. R., Macizo, P.,


Piaget, J. (1950). *The psychology of intelligence.* Harcourt, Brace


https://doi.org/10.3758/bf03197273

https://doi.org/10.3758/bf03197273

https://doi.org/10.1080/09297049.2017.1280142

http://rm.coe.int/09000016808b1688

https://doi.org/10.1515/revneuro-2012-0035


Appendix 1

English subject-verb agreement sentences (pre-test)
*This is an English translation of the original instructions in Spanish

I’m going to read out a group of sentences. I will read them one by one and will ask you a question about each sentence.

If the answer to that question is YES, you’ll circle the checkmark next to the number of the sentence in your paper. If the answer to the question is NO, you will circle the X.

For example, if I read *Dog yellow is and I ask Is the sentence correct or incorrect*, what would you circle? (Teacher checks answers). Another example, if I say *The bag is blue, and I ask Is the bag yellow, what would you circle?* (Teacher checks answers and clarifies with further examples if necessary).

I will read each sentence only once, without repeating it, so you need to pay close attention.

Ok, we will begin now.

1. The cat is under the table.
   Q1. Is the cat on the table?

2. The photos on the desk is new.
   Q2. Is the sentence correct or incorrect?

3. The bicycles are green.
   Q3. Is the sentence correct or incorrect?

4. The games in the shop are new.
   Q4. Is the sentence correct or incorrect?

5. The book under the table are big.
   Q5. Is the sentence correct or incorrect?

6. The boat is beautiful.
   Q6. Is the sentence correct or incorrect?

7. The bottle in the bag are big.
   Q7. Is the sentence correct or incorrect?

8. The books are on the desk.
   Q8. Are the books on the desk?

9. The ball in the bag is red.
   Q9. Is the sentence correct or incorrect?

10. The planes are in the shop.
    Q10. Are the planes in the house?
11. The dog is on the book.
Q11. Is the sentence correct or incorrect?

12. The pictures are yellow.
Q12. Are the pictures blue?

13. The computer is small.
Q13. Is the sentence correct or incorrect?

14. The games in the computer are old.
Q14. Are the games new?

15. The trains on the table is short.
Q15. Is the sentence correct or incorrect?

16. The balls under the chair are yellow.
Q16. Are the balls under the chair?

17. The monkey is on the table.
Q17. Is the monkey under the table?

18. The pencils in the bag is blue.
Q18. Is the sentence correct or incorrect?

19. The dogs are small.
Q19. Is the sentence correct or incorrect?

20. The ball in the bedroom is small.
Q20. Is the ball in the bedroom?

21. The notebooks under the bed are blue.
Q21. Are the notebooks under the bed?

22. The game in the shop are new.
Q22. Is the sentence correct or incorrect?

23. The bedroom is big.
Q23. Is the sentence correct or incorrect?

24. The games are old.
Q24. Are the games new?

25. The book in the classroom is big.
Q25. Is the sentence correct or incorrect?

26. The frogs under the bed is big.
Q26. Is the sentence correct or incorrect?

27. The computer is yellow.
Q27. Is the computer blue?

28. The girl is fast.
Q28. Is the girl fast?
29. The photo in the book are old
Q29. Is the sentence correct or incorrect?

30. The trains are fast
Q30. Is the sentence correct or incorrect?

31. The pencils on the table are red
Q31. Is the sentence correct or incorrect?

32. The boys are strong
Q32. Is the sentence correct or incorrect?
Spanish subject-verb agreement sentences (pre-test)
I’m going to read out loud a group of sentences. I will read them one by one and will ask you a question about each sentence.

If the answer to that question is YES, you’ll circle the checkmark. If the answer to the question is NO, you will circle the X.

For example, if I read *Dog yellow is* and I ask *Is the sentence correct or incorrect*, what would you circle? (Teacher checks answers). Another example, if I say *The bag is blue*, and I ask *Is the bag yellow*, what would you circle? (Teacher checks answers and clarifies with further examples if necessary).

I will read each sentence only once, without repeating it, so you need to pay close attention.

Ok, we will begin now.

1. El gato blanco debajo de la mesa de estudio de Cecilia duerme tranquilo.
   Q1 – ¿Duerme el gato debajo de la cama de Cecilia?

2. La foto de los animales en las montañas de Cuba son del libro de Alina.
   Q2 - ¿La oración es correcta o incorrecta?

3. El jardín de la casa amarilla detrás de la escuela de Carlos necesita flores.
   Q3 – ¿Es la casa, detrás de la escuela, azul?

4. Los miembros de la familia hablaron sobre la música de la película de ayer.
   Q4 - ¿La oración es correcta o incorrecta?

5. El camión recoge a los trabajadores todos los días a las nueve de la mañana.
   Q5 - ¿La oración es correcta o incorrecta?

6. Las palabras en la foto de la portada de la revista parecen de otro idioma.
   Q6 - ¿La oración es correcta o incorrecta?

7. El grupo de jóvenes visitará el parque de juegos mañana después de clases.
   Q7 – ¿Es la visita al parque antes de clases?

8. Las plantas con flores amarillas en los jardines frente al museo crece rápido.
   Q8 - ¿La oración es correcta o incorrecta?

9. La escuela de Verónica está muy lejos del parque más grande de La Habana.
   Q9 – ¿Estudia Verónica muy lejos del parque más grande de la Habana?

10. Las nubes negras cubrieron el sol por varios días durante el mes de septiembre.
    Q10 – ¿Cubrieron las nubes el sol por varios días en septiembre?

11. Los padres del compañero de clase de historia de Julito fueron a la reunión.
    Q11 - ¿La oración es correcta o incorrecta?

12. Los compañeros de la clase de español del hermano de Andrés visitó el cine.
    Q12 - ¿La oración es correcta o incorrecta?

13. Los animales del campesino comieron la comida a las nueve de la mañana.
Q13 – ¿Comieron los animales a las 10 de la mañana?

14. El libro cerca de la ventana al final del aula tienen imágenes de la ciudad.
Q14 - ¿La oración es correcta o incorrecta?

15. Las plantas del bosque movieron sus hojas y ramas con el paso del fuerte viento.
Q15 – ¿La oración es correcta o incorrecta?

16. La abuela de Alejandro tenía un vestido de flores amarillas en el cuarto.
Q16 - ¿La oración es correcta o incorrecta?

17. Los alumnos de la clase de español de la maestra de Claudia estudia a diario.
Q17 – La oración es correcta o incorrecta?

18. El mercado de alimentos está cerrado desde hace semanas por reparaciones.
Q18 – ¿Está cerrado el mercado desde hace meses?

19. La tarea para la clase de Historia de Cuba de la semana próxima es bastante fácil.
Q19 – ¿Es la tarea para la semana próxima?

20. El video de la canción tiene muchos bailarines y músicos en una fiesta.
Q20 - La oración es correcta o incorrecta?
English subject-verb agreement sentences (post-test)

*This is an English translation of the original instructions in Spanish*

I’m going to read out a group of sentences. I will read them one by one and will ask you a question about each sentence.

If the answer to that question is YES, you'll circle the checkmark next to the number of the sentence in your paper. If the answer to the question is NO, you will circle the X.

For example, if I read Dog yellow is and I ask Is the sentence correct or incorrect, what would you circle? (Teacher checks answers). Another example, if I say The bag is blue, and I ask Is the bag yellow, what would you circle? (Teacher checks answers and clarifies with further examples if necessary).

I will read each sentence only once, without repeating it, so you need to pay close attention.

Ok, we will begin now.

1. The buses in the school is yellow.
   Q1 - Is the sentence correct or incorrect?
2. The monkey on the tree is small.
   Q2 – Is the monkey on the tree?
3. The books under the sofa are open.
   Q3 - Is the sentence correct or incorrect?
4. The glasses are on the kitchen table.
   Q4 – Are the glasses under the table?
5. The balls under the chair are blue.
   Q5 - Is the sentence correct or incorrect?
6. The book in the bag are small.
   Q6 - Is the sentence correct or incorrect?
7. The cat is under the big stairs.
   Q7 – Is the cat under the desk?
8. The students are in the big park.
   Q8 – Are the students in the park?
9. The pictures are in the black book.
   Q9 - Is the sentence correct or incorrect?
10. The bag is in the small closet.
    Q10 – Is the bag in the closet?
11. The toys in the bedroom are new.
   Q11 – Are the toys new?
12. The radio is on the blue table.
   Q12 - Is the sentence correct or incorrect?
13. The red t-shirt is on the table.
   Q13 – Is the t-shirt black?
14. The boxes under the closet is old.
   Q14 - Is the sentence correct or incorrect?
15. The room under the stairs are big.
   Q15 - Is the sentence correct or incorrect?
16. The apple on the table is red.
   Q16 - Is the sentence correct or incorrect?
17. The football classes are in the morning.
   Q17 – Are the classes in the morning?
18. The dresses on the bed is white.
   Q18 - Is the sentence correct or incorrect?
19. The tall man is in the park.
   Q19 – Is the man tall?
20. The chair on the table is black.
   Q20 – Is the chair under the table?
21. The blue chair is in the kitchen.
   Q21 - Is the sentence correct or incorrect?
22. The car in the park is fast.
   Q22 – Is the sentence correct or incorrect?
23. The new games are in the computer.
   Q23 - Is the sentence correct or incorrect?
24. The oranges on the desk are big.
   Q24 – Are the oranges on the bed?
25. The houses in the city is new.
   Q25 - Is the sentence correct or incorrect?
26. The shoes are under the old sofa.
   Q26 - Is the sentence correct or incorrect?
27. The dish in the kitchen are dirty.
   Q27 - Is the sentence correct or incorrect?
28. The big computers are on the desk.
   Q28 – Are the computers under the desk?
29. The old shoes are under the closet.
   Q29 - Is the sentence correct or incorrect?
30. The face in the photo are old.
   Q30 - Is the sentence correct or incorrect?
31. The table is in the big bedroom.
   Q31 - Is the sentence correct or incorrect?
32. The small box is under the bed.
   Q32 - Is the sentence correct or incorrect?
Spanish subject-verb agreement sentences (post-test)
*This is an English translation of the original instructions in Spanish

I’m going to read out a group of sentences. I will read them one by one and will ask you a question about each sentence.

If the answer to that question is YES, you'll circle the checkmark next to the number of the sentence in your paper. If the answer to the question is NO, you will circle the X.

For example, if I read Dog yellow is and I ask Is the sentence correct or incorrect, what would you circle? (Teacher checks answers). Another example, if I say The bag is blue, and I ask Is the bag yellow, what would you circle? (Teacher checks answers and clarifies with further examples if necessary).

I will read each sentence only once, without repeating it, so you need to pay close attention.

Ok, we will begin now.

1. La computadora de Verónica tiene muchos juegos y programas en inglés.
   Q1 - ¿Tiene la computadora juegos en inglés?

2. Las luces de la ciudad brillan con más fuerza en las largas noches de invierno.
   Q2 - ¿Es la oración correcta o incorrecta?

3. Las tareas de la clase de Lengua Española de primer grado es para mañana.
   Q3 - ¿Es la oración correcta o incorrecta?

4. La maestra de Educación Física de los estudiantes de primer grado es muy querida.
   Q4 - ¿Es la maestra de Matemática?

5. Las canciones de la película de aventuras del sábado en la noche eran muy largas.
   Q5 - ¿Es la película de aventuras?

6. El televisor necesita reparaciones en la pantalla después de su fuerte caída.
   Q6 - ¿Necesita reparaciones el televisor?

7. La señora alta con espejuelos en los anuncios de la televisión trabaja todos los días.
   Q7 - ¿Es la oración correcta o incorrecta?

8. Las escuelas están cerradas desde la cinco de la tarde hasta las siete de la mañana.
   Q8 - ¿Están cerradas desde las seis de la tarde?

9. Las camisas de Javier están siempre muy sucias después de los juegos de pelota.
   Q9 - ¿Están limpias las camisas después de la pelota?
10. El compañero de clases de Matemática de los hermanos de Claudia hicieron la tarea.
Q10 - ¿Es la oración correcta o incorrecta?

11. El parque recibe siempre muchos visitantes de diferentes países en estos meses.
Q11 - ¿Es la oración correcta o incorrecta?

12. El patio grande detrás de la casa de los abuelos de Carlos necesita pintura.
Q12 - ¿Es la oración correcta o incorrecta?

13. La maestra revisa siempre la tarea antes del comienzo de la clase de Español.
Q13 - ¿Es la oración correcta o incorrecta?

14. Los estudiantes de segundo prepararon una actividad para el fin de curso escolar.
Q14 - ¿Es la oración correcta o incorrecta?

15. Los jardines de flores cerca de la escuela de inglés de Anabel crecen cada día.
Q15 - ¿Es la oración correcta o incorrecta?

16. Los invitados a la actividad del teatro por el fin del curso escolar llegó temprano.
Q16 - ¿Es la oración correcta o incorrecta?

17. La escuela primaria cerca del estadio de los jugadores de fútbol están grandes.
Q17 – ¿Es la oración correcta o incorrecta?

18. Los jugadores del equipo de pelota de la primaria de Alejandro tiene pocos años.
Q18 - ¿Es la oración correcta o incorrecta?

19. Los aviones vuelan desde el aeropuerto de la ciudad temprano en la mañana.
Q19 - ¿Vuelan los aviones en la noche?

20. El sofá verde de la sala lleva muchos años en el mismo posición junto al piano.
Q20 - ¿Está el sofa junto al piano?
Information leaflet for parents

Does working memory training improve syntactic ability? A test of cognitive transfer with childhood bilingualism

What are the aims of the project?
The aim is to explore the influence of working memory on language learning and vice-versa. Working memory is what we use when we temporarily store information to resolve a situation. For example, in the sentence: *The student is always early to class because she never misses the bus*, we need to keep the image of the student in mind, to understand who *she* is referring to. This study uses memory and language exercises in English and Spanish to understand how memory and language influence each other in children aged 6-7.

Who conducts the experiment and what is the affiliation?
Ernesto Roque Gutierrez conducts this project as part of his PhD at the Open University (UK). Ernesto has received and continues to receive training in scientific research. Ernesto will work closely with your current English teacher.

Why am I being invited to take part?
Because you are the parent/guardian of a child that is part of the group of English students who are 6-7 years old, and the study is interested in that age group.

If I decide to join, what does that involve?
First, the child will be taking some exercises. This will be memory exercises (20 minutes), English exercises (20 minutes) and Spanish exercises (20 minutes). The memory exercises involve pressing a key on a computer when they see similar images. The language exercises involve answering a question about sentences that they will listen to. After this first stage, they will either a) continue practicing the memory exercises for the following 6 weeks; b) continue practicing the English sentences for 6 weeks; c) continue practicing the Spanish sentences for 6 weeks; d) carry on with their regular classes. This organisation is random, and all groups will continue with their regular classes, so this will not compete with their classes. Even if they are in group D, their participation is vital to the study. Only the researcher and the teacher will be leading the study, so children’s safety and wellbeing will not be compromised. All students who decide to participate in the study will receive one month of free English tuition as a token of appreciation for their time, along with some English posters, stickers, and other similar materials. Their participation is entirely voluntary, and they are free to stop being part at any time during the duration of the study.

What are the research methods?
The memory test involves pressing a key whenever an image they see on the screen matches a previous image. The exercise uses different shapes and lasts
approximately 20-25 minutes. This will take place twice a week, on the same days of regular English classes. The language sentences in English and Spanish are similar to the following: *The book on the table is black / The book on the table in the room is black.* They will be asked questions such as: *What colour is the book?* This will also take 20 minutes on the same day of English classes.

**What will you be observing?**
The development of their memory and/or language skills, assessed initially and then at the end of the research using a similar test to the first one. We will not be able to provide any clinical assessment of your child’s memory or language ability from these tests.

**Is it confidential?**
Their participation will be treated in **strict confidence** in accordance with the UK Data Protection Act. No personal information will be passed to anyone outside the research team. We will write a report of the findings from this study, but no individual will be identifiable in published results of the research.

**What if I have other questions?**
*If you have any other questions about the study we would be very happy to answer them.* Please contact **Ernesto Roque Gutierrez 78782030** or by email to **ernesto.roque-gutierrez@open.ac.uk.** Alternatively, you can contact **Paul Ibbotson paul.ibbotson@open.ac.uk.**
Information leaflet for children

Memory and Language Training

What are the aims of the project?
This study will help us understand how your memory works when you use Spanish and English. We're going to have different activities where you can answer language questions and do exercises using a laptop where you will see different images.

Who is doing the study?
Ernesto Roque Gutierrez is doing this project. He works for the Open University, in the UK, and together with your English teacher, he will help you with the activities.

Why am I being invited to take part?
Because you are an English student and you are 6 or 7 years old. This helps us understand how other 6-7-year-old children learn languages.

If I decide to join, what does that involve?
Having extra activities from your normal English classes. These activities will be either responding to some images that you see on a screen or answering some questions about sentences that your teacher will read to you. This will happen for 6-7 weeks.

As a thank you for your participation, your parents will receive a discount for your English classes, and you will be given some English posters, stickers, and other materials. You do not have to participate if you do not want to and you can stop taking part in the study at any time.

What are the activities?
One activity involves pressing a key whenever a shape you see on the screen matches a previous image. This lasts approximately 20-25 minutes, twice a week, on the same days of regular English classes. Other students will have different language activities, where they listen to a sentence such as: The book on the table is black and are asked: What colour is the book? This will also take 20 minutes on the same day of English classes.

What will you be observing?
How you perform along the way.

Will other people know my results?
No. Only Ernesto and his supervisors Paul and Lina. No one else will be able to know which results are yours, as your name will never appear.

What if I have other questions?
If you have any other questions about the study we would be very happy to answer them. Ask your parent/carer to contact Ernesto Roque Gutierrez 78782030 or by email to ernesto.roque-gutierrez@open.ac.uk or Paul Ibbotson paul.ibbotson@open.ac.uk
CONSENT/ASSENT FORM FOR CHILDREN
(to be completed by the child with the help of a parent/carer)

Please indicate YES or NO for each of the questions below
(Parent/guardian to read to the child and complete if the child is unable)

Have you read (or had read to you) what the project is about?        YES     NO
Has someone explained this project to you?                        YES     NO
Do you understand what this project is about?                     YES     NO
Have you asked all the questions you want?                        YES     NO
Have you had your questions answered in a way you understand?     YES     NO
Do you understand it’s OK to stop taking part at any time?         YES     NO
Are you happy to take part?                                       YES     NO

If any answers are ‘no’ you can ask more questions. But if you don’t want to take part, don’t write your name.

If you want to take part, please write your name and today’s date

Your name       ___________________________
Date                 ___________________________

Your parent or carer must write their name here too if they are happy for you to do the project

Print name       ___________________________
Sign               ___________________________
Date              ___________________________

The researcher who explained this project to you needs to sign too:

Print Name       ___________________________
Sign               ___________________________
Date              ___________________________

Thank you for your help.
Appendix 2

A representative sample of the sentence used in Study 3 (chapter 5) is included below. In the case of the English sentences, they were extracted from three different corpora (Gillam & Pearson, 2004; Weismer et al., 2013; Wells, 1981) at CHILDES, and they show child speech from age 1;6 to 11;11. As mentioned in chapter 5, transcripts were analysed by means of the Child Language Analysis program (CLAN) (MacWhinney, 2000). The Spanish data corresponds to two different corpora and was matched for the same period used in English, from 1;6 to 11;11 years of age (Albala & Marrero, 2006; Benedet et al., 2004).

Sample sentences (English corpora)

1- *CHI: that sort of thing won’t go on
2- *CHI: a sink that floats
3- *CHI: <makes car noise> [/?] this case that come>
4- *CHI: a dame who lost her . (39 – 1)
5- *CHI: the mac(kintosh) what’s in my drawer . (39 - 1)
6- *CHI: a girl called gallant prince came riding riding . (42 - 4)
7- *CHI: that little boy who was on Peter Pan . (42 - 1)
8- *CHI: that doctor who looks after you . (42 - 1)
9- *CHI: your leg always fall off (4.) . (42 - 2)
10- *CHI: the one that &+i is out of the swing . (1)
11- *CHI: my dad always has breakfast . (2)
12- *CHI: <the &+fir> [/] three because it only has one mom and (2)
13- *CHI: a little table for a baby goes right here . (4) 3;8
14- *CHI: <the girl and the &b> [/] one of the babies is gonna take a bath . (4)
15- *CHI: my finger kind of hurts . (2)
16- *CHI: a purse that goes with [*] . (1)
17- *CHI: one of them’s one of the spots’ not on the leg . (58 – 3)
18- *CHI: the one in school was (3.) . (57 – 2).
19- *CHI: one side of him is sad again . (56 – 2)
20- *CHI: this thing down in her kitchen that [/] that’s [/] that has cream in it <and there's a little> [/] and it can only put out very tiny fires [*] . (6)
21- *CHI: the ambulance just goes like that . (1)
22- *CHI: the one with the bag hasta +... (3)
23- *CHI: my mom just isn’t very happy with me sometimes when I do rude things xxx . (1)
24- *CHI: the one that has shoes stick to it because it has matching shoes [*] . (1)
25- *CHI: one time we even took a family picture . (2)
26- *CHI: my grandma and grandpa kind of live near the cabin . (2)
27- *CHI: one of my friends is Childsname . (3)
28- *CHI: my family just goes . (1)
29- *CHI: (...) <no well> [//] the only thing about him is he stares at the tv . (3)
30- *CHI: my dad he [/] he kind of gets grumpy . (2)
31- *CHI: my dog really [/] really [/] really did die . (2)
32- *CHI: my grandma and grandpa and Devin and Derek and some of my neighbors come to our birthday's . (15)
33- *CHI: my other bear only has a tie on . (2)
34- *CHI: my [/] my &-um [//] the first grandma I went to house is the grandma Name [*] . (4)
35- *CHI: my fish just swims around and we feed it once a day too . (1)
36- *CHI: my other shoe always likes to do that . (2)
37- *CHI: <oh my cats> [//] my dog just eats my cat food for my cats . (1)
38- *CHI: <the only thing that his he can well> [//] one xxx is one thing and the other who is two things [*] . (1)
39- *CHI: <I think he> [//] two teams that he like [* oes] oare the Milwaukee brewers and the [*] (3)
40- *CHI: my mom always wants me to help plant flowers (2)
41- *CHI and the other thing that I do for her is to help her well (5)
42- *CHI: the thing I gave her for mother's day is flower dress xxx [*] . (7)
43- *CHI: my dad wants to wash the rug and get [/] get a car but we already have a black car [*] . (2)
44- *CHI: <go some place we> [//] the only thing [: place] [*] we go is go to the store [*] .
45- *CHI: <the always kill> [//] that big bad guy always killed them . (2)
46- *CHI: a Johndeere tractor that has a lot of stickers on it . (1)
47- *CHI: the [/] the catcher guy that catches the animals . (1)
48- *CHI: the little lady that wasn't scared of anything . (1)
49- *CHI: the ones in the middle > the little pieces < are the special > [//] are the special blocks . (4)
50- *CHI: <we have &-like some> [//] a lot of them died . (2)

Sample sentences (Spanish corpora)

1. *CHI: <e(l) catito> [/] e(l) catito [: gatito] [*] no muerde . (1)
2. *CHI: (l)a cinta se cayó ayer [*] . (1)
3. *CHI: le [: el] [*] pote [: bigote] [*] se va ! (1)
5. *CHI: la [% 'a' velarizada] ropa [*] que tiene aquí yo .
6. *CHI: (l)a fie(s)ta no e(s) aquí , e(s)tá allí .
7. *CHI: e(l) bidé no e(s)tá (.) e(l) bidé no e(s)tá .
8. *CHI: el colchó(n) no (es)tá .
9. *CHI: la raja [: raya] que se me (.) o(l)vida .
10. *CHI: e(l) cuento [*] no (es)tá .
11. *CHI: un tren , y Daniel se iba a subir [*] a(l) tren a.
   a. [% vibrante dentalizada] .
12. *CHI: e(l) coche (.) de abuela [/] de Soraya e(s)tá en e(l) garaje [*] .
13. *CHI: el vaso ya está , chie@o .
14. *CHI: los que [/] los que (.) colorean las fotos +...
15. *CHI: los que (.) hacen [*] el color de la foto , verdad ?
17. *CHI: mi color favorito es el verde .
18. *CHI: el coche de Tanina tiene gafas .
19. *CHI: cuando va el &le &le leon que se cab(r)ea con el tigre .
20. *CHI:  era un niño pequeño que se llama [*] Besitos y el [/] estaba en un huevo.
21. *CHI:  el belén ya estaba pintado. (1)
22. *CHI:  el patito pobrecito que tiene frío.
23. *CHI:  el perro no le va a matar.
24. *CHI:  el pájaro no es bueno (.) bueno. (1)
25. *CHI:  Yasmin (.) el tigre que se llama Rajá. (2)
26. *CHI:  mira (.) el murcielago otra vez esta ahí eh ? (3)
27. *CHI:  el sol cada día se va agotando (.)
28. *CHI:  a Eduardito (.) que el día de San Eduardo cumpl(i)ó
29. *CHI:  a un chico de esos de lata que esta lleno de lata y luego cuando van paseando los dos con el gatito
30. *CHI:  un niño que se llama Sergio (.) todo el rato me estaba echando escupitajos
31. *CHI:  lo que pasa es que te metes cada culetazo ahí en el castillo elástico +//.
32. *CHI:  la hija de un rey se quería casar con un príncipe pero el príncipe se tuvo que ir.
33. *CHI:  el profesor (.) tiene una mala leche que yo que sé .
34. *CHI:  el mejor de todos es David Latorre (.) mi primo .
35. *CHI:  el que la liga tiene una vida (.) y si tiene una vida y si a ese le dá como tiene una vida puede seguir jugando .
36. *CHI:  el quiste mira como está .
37. *CHI:  el hermano de mi padre tiene dos hijos , no , que se llaman Juan Angel y Vanesa (.)
38. *CHI:  la hija le convenció porque era a quien amaba ella (.)
39. *CHI:  el chulo es un hombre que vive de las mujeres .
40. *CHI:  el objeto del juego es colocar cinco de tus piramides en línea tanto vertical (.) horizontal como diagonal . (3)
41. *CHI:  el niño se imaginaba que no estaba [>].
42. *CHI:  el de Silvia es en mayo (.) el día veintiuno (.) pero lo celebra (.) el veinte que cae en viernes (.) y ya tengo una cosa para ella (.) es un diario +/. 
43. *CHI:  a Eva todavía no lo se (.) como xxx falta tiempo y el de Mari Carmen es el veintiuno de junio .
44. *CHI:  el que es mucho (.) pero de verdad (.) mucho mejor que yo (.) saco ocho y es que se distrae mucho (.) y no +//.
45. *CHI:  el sotano del miedo (.) y mama no me dejo verla .
46. *CHI:  el arado (.) también fue. (2)
47. *CHI:  el año que viene cuando vaya a sexto (.) ya entro en el patio de los mayores .
48. *CHI:  el looping star también me gusta y la montaña rusa de agua también me gusta .
49. *CHI:  el juego <se termina a la una> [/] se terminó a la una y media .
50. *CHI:  el mio se llama Fray Luis de Leon y el de mi hermana El Sagrado Corazon de Jesus .
Appendix 3
Parser walkthrough

There are two different processes used in Study 4 (Chapter 6) that involve the custom-built parser. This appendix offers a walkthrough, a sort of README explanation to its usage, which was created in collaboration with Javier Moreno, Leandro Valdes, and Leandro Casuso. The app has been made available as an open-source repository at GitHub (link included below).

The main purpose of the first one is to sanitise the data extracted from the CHILDES corpora using the CLAN string before reformatting it into JavaScript Object Notation (JSON), which is the input used by Sankey Diagram Generator app (Csala, 2014).

The screenshot above shows part of the code used (left) and the input it uses (right) which is exactly in the same output format of the MLU string used in CLAN. While the image in right shows a .txt document, any Excel-supported file extensions could have been used instead. The sanitisation of the lines is particularly important as the sentences extracted usually come with their accompanying annotations (e.g., parentheses, brackets, asterisks, etc.) which would compromise the reliability of the
depiction of the nodes as the Sankey app would assume that every unit is a key/node.

Other researchers interested in running Sankey visualisations of comparable corpus-extracted syntax (i.e., format shown in right image), can access the app.py file shared in GitHub and follow the next steps.

1- Download python from this link. For reference, the parser was coded in Python3.
2- Download the app.py file using this GitHub link.
3- Create a folder and copy the app and the file containing the data to be analysed (e.g., a file named data.txt)
4- In iOS, press command + space and type terminal and press return.
   In Windows, press windows + R to open the run box. Type cmd and then click OK to open a regular command prompt.
5- Access the folder containing the data using CMD commands. For example, if the target folder is in Documents, type cd documents and press return.
6- Once in the folder, type the following: python3 app.py data.txt | pbcopy. Depending on the version of python you’ve installed, the number after python may vary. The pbcopy ... allows you to copy the output of the command into the clipboard.
7- Access the Sankey Diagram Generator using this link and click on Load.
8- You will see an empty box where you can paste the output of step 6. It should show the JSON needed to visualise the Sankey.
9- Click on Done and voilà, the visualisation of your data will be there.