This thesis is presented for the degree of Doctor of Philosophy

The Influence of Linguistic Structure on Memory Span: Repetition Tasks as a Measure of Language Ability

Kamila Polišenská

City University London
Department of Language and Communication Science

July 2011

Table of Contents

CHAPTER 1
The Relationship between Repetition Tasks and Language **16**

1.1 Introduction 16

1.2 Outline of the thesis 19

1.3 Repetition tasks and their relation to language 20

1.3.1 Span tasks 21

1.3.2 Nonword repetition tasks 22
1.3.2.1 Nonword repetition and vocabulary development 23
1.3.2.2 Nonword repetition and syntactic skills 24
1.3.2.3 NWR as a clinical marker 26
1.3.2.4 Nonword repetition summary 28
1.3.2.5 Nonword repetition vs. span tasks: Same or different? 28

1.3.3 Sentence repetition tasks 29
1.3.3.1 Sentence repetition as a measure of language development 30
1.3.3.2 Language assessment through Sentence repetition 32
1.3.3.3 Sentence repetition as a clinical marker 32
1.3.3.4 Sentence repetition summary 34

1.3.4 The trade-off between language and memory in repetition tasks 35

CHAPTER 2
The Influence of Long-Term Memory Knowledge on Immediate Verbal Repetition **37**

2.1 Introduction 37

2.2 STM capacity and the role of linguistic knowledge 37

2.2.1 Length and the notion of capacity 37

2.2.2 Influences of sublexical knowledge on immediate repetition of nonwords 39
2.2.2.1 Wordlikeness 39
2.2.2.2 Phonotactic probability 41
2.2.2.3 Neighbourhood density 42
2.2.2.4 Prosodic structure 42
2.2.2.5 Sublexical properties reflected in nonword repetition tests 44

2.2.3 The influence of frequency, imageability and lexical status on span tasks 46

2.2.4 Evidence of content/function-word distinction in span tasks 46

2.2.5 Is the capacity of immediate recall linked to the nature of items? 49

2.2.6 Influence of LTM on immediate recall: beyond single-item level 51
2.2.6.1 Grouping within a list of digits/words 52
2.2.6.2 Beyond the phonetic/phonological level of grouping 55
2.2.6.3 Summary 61

2.3 How STM models reflect the role of linguistic knowledge 61

2.3.1 Issue of clarification: Short-term memory versus working memory 62

	2.3.2	Multicomponent model: STM and language viewed as two separate constructs	63
		2.3.2.1 Introducing the working memory model	63
		2.3.2.2 The phonological loop	63
		2.3.2.3 The visuospatial sketchpad	65
		2.3.2.4 Central Executive: Baddeley (2000; 2003)	65
		2.3.2.5 Episodic buffer	65
	2.3.3	Jackendoff's remarks on working memory	67
	2.3.4	STM as an activation of long-term memory linguistic representations	67
		2.3.4.1 STM as 'focus of attention'	67
		2.3.4.2 MacDonald and Christiansen (2002)	68
		2.3.4.3 Acheson and MacDonald (2009)	69
	2.3.5	Summary	69
2.4	Implications for language acquisition, assessment and impairment		70

CHAPTER 3
The Contribution of Linguistic Knowledge to Immediate Repetition: Research Questions **71**

3.1	Introduction		71
3.2	Interconnections		71
	3.2.1	Lexical items and morphosyntactic devices	71
		3.2.1.1 Function words	72
		3.2.1.2 Content words	73
		3.2.1.3 Word order and inflections	74
3.3	Research questions		74
	3.3.1	RQ 1: Conceptual structure	74
	3.3.2	RQ 2: Morphosyntax	75
	3.3.3	RQ 3: Lexical items	76
	3.3.4	RQ 4: Prosody	76
	3.3.5	RQs 5 – 7: Function words versus content words	78
	3.3.6	RQ 8: Relations between repetition performance and a vocabulary task	78
	3.3.7	RQ 9: Cross-linguistic comparisons	78
3.4	Implications of the interrelations between linguistic domains		79

CHAPTER 4
The Design of the Experimental Task: A Memory Span Measure beyond the Word Level **80**

4.1	Introduction		80
	4.1.1	Task	82
	4.1.2	Non-experimental factors	85
	4.1.3	Relation between repetition and receptive vocabulary	86
	4.1.4	Overview of chapter	86

4.2	Experimental conditions: Creation of materials		86
	4.2.1	Foundation for all conditions: Condition A	86
	4.2.1.1	Lexical content	87
	4.2.1.2	Function words	88
	4.2.1.3	Balance of content words and function words	88
	4.2.1.4	Structure	89
	4.2.1.5	Prosodic properties	89
	4.2.2	Semantically implausible targets: Condition C	91
	4.2.2.1	Lexical content	91
	4.2.2.2	Function words	93
	4.2.2.3	Structure	94
	4.2.2.4	Prosodic properties	94
	4.2.3	Ungrammatical targets: Condition D	95
	4.2.3.1	Structure	95
	4.2.3.2	Lexical content and function words	96
	4.2.3.3	Prosodic properties	96
	4.2.4	Hybrid targets	98
	4.2.4.1	Condition F	98
	4.2.4.2	Condition I	99
	4.2.5	Nonlexical targets	101
	4.2.5.1	Condition G	101
	4.2.6	Conditions with a list prosody	103
	4.2.6.1	Condition E	104
	4.2.6.2	Condition B	105
	4.2.6.3	Condition H	108
4.3	Experimental conditions: Creation of Czech materials		110
	4.3.1	Foundation for all conditions: Condition A	111
	4.3.1.1	Lexical phonology	111
	4.3.1.2	Morphosyntax	111
	4.3.1.3	Lexical items	114
	4.3.2	Semantically implausible targets: Condition C	114
	4.3.3	Ungrammatical conditions	117
	4.3.4	Non-lexical and hybrid conditions	122
	4.3.5	Conditions with a list prosody versus sentence-like prosody	123
4.4	Recording the stimuli		124
	4.4.1	Editing the recorded stimuli	125
4.5	Order of presentation		125
4.6	Procedure		126
	4.6.1	'Game' format of the presentation	126
	4.6.2	Familiarising the participants with the task	127
	4.6.3	Starting points for conditions in English and in Czech	128
	4.6.4	Rationale behind the 'game' presentation	130

4.7 Scoring of the experimental task 130

4.7.1 Scoring criteria 130
4.7.1.1 English 130
4.7.1.2 Czech 133
4.7.1.3 Acceptability of sentence-like versus list prosody conditions 135

4.7.2 Maximum span 135

4.8 Czech version of the British Picture Vocabulary Scale 136

4.9 Changes following the pilot study 137

4.9.1 Theoretically motivated changes in conditions and materials 137

4.9.2 Problem items 137

4.9.3 Presentation 138

CHAPTER 5
The Study **140**

5.1 Introduction 140

5.2 Participants 140

5.3 Procedure 142

5.3.1 Procedure for children 142

5.3.2 Procedure for adults 143

5.4 Reliability 143

5.4.1 Intra-rater reliability 144

5.4.2 Inter-rater reliability 145

5.5 Results 146

5.5.1 Introduction 146

5.5.2 Between-subject factors 146
5.5.2.1 General language skills – Picture vocabulary task 146
5.5.2.2 Order of presentation 147
5.5.2.3 Age 148
5.5.2.4 Gender differences 150
5.5.2.5 Summary 151

5.5.3 Descriptive statistics: Means and standard deviations for experimental conditions 151

5.5.4 Planned comparisons of experimental conditions 153
5.5.4.1 Introduction 153
5.5.4.2 Research question 1 154
5.5.4.3 Research question 2 155
5.5.4.4 Research question 3 156
5.5.4.5 Research question 4 158
5.5.4.6 Research question 5 159
5.5.4.7 Research question 6 160
5.5.4.8 Research question 7 161
5.5.4.9 Summary of the planned comparisons of experimental conditions 162

5.5.5 Relation between repetition and receptive vocabulary: Research question 8 164

5.5.6 Cross-linguistic comparisons: Research question 9 165

CHAPTER 6
Discussion **167**

6.1 Introduction 167

6.2 Main findings 167

6.2.1 Repetition performance as a reflection of language domains 167

6.2.2 Repetition performance: familiarity of CWs vs. FWs 169

6.3 Factors to be considered in the interpretation of the results 172

6.3.1 Possible confounding factors: Duration 172

6.3.2 Possible confounding factors: Non-experimental linguistic factors 175
6.3.2.1 Different sets of CWs in semantically plausible and implausible sentences 175
6.3.2.2 Balance of CWs/FWs 176

6.3.3 Span as a testing method: results confounded with attention failures? 177

6.4 Differences between Czech and English 178

6.4.1 Differences in creating the stimuli in English and Czech 179

6.4.2 Span in words vs. length in morphemes 180

6.4.3 Word order: Scoring 182

6.4.4 Cross-linguistic comparison of similarities/differences 182
6.4.4.1 Cross-linguistic similarities/differences in children's data 182
6.4.4.2 Cross-linguistic differences in adults' data 183
6.4.4.3 Sentence repetition tasks: sensitivity to language-specific properties 184

6.5 Differences across age groups 185

6.5.1 Evidence from the current study 185

6.5.2 Developmental changes documented in the literature 187

6.6 Implications: Theories of verbal memory and the language-memory debate 189

6.7 Limitations of the study and future research 191

6.7.1 Results in relation to other memory and language tasks 191

6.7.2 Error analyses 193

6.7.3 Immediate vs. delayed recall 193

6.7.4 Limitations in the stimuli and conditions 196

6.8 Repetition tasks re-examined 196

6.8.1 Potential for assessment 196

6.8.2 Conclusions 197

APPENDICES **199**

REFERENCES **251**

List of Tables

Table 2.1 Mean percentages of correctly repeated phonemes in nonwords
Table 2.2 An overview of experimental categories with examples
Table 2.3 Overview of experimental conditions and examples
Table 4.1 Nine experimental conditions in English, span 5
Table 4.2 Overview of the experimental conditions
Table 4.3 Syntactic structures across conditions
Table 4.4 Stimuli in condition A
Table 4.5 Stimuli in condition C
Table 4.6 Stimuli in condition D
Table 4.7 Stimuli in condition F
Table 4.8 Stimuli in condition I
Table 4.9 Stimuli in condition G
Table 4.10 Stimuli in condition E
Table 4.11 Stimuli in condition B
Table 4.12 Stimuli in condition H
Table 4.13 Nine experimental conditions in Czech, span 4
Table 4.14 Czech stimuli in condition A
Table 4.15 The full set of Czech stimuli for condition C
Table 4.16 Ungrammatical conditions for short span in Czech
Table 4.17 Overview of the starting points for presenting stimuli
Table 4.18 Examples of correct responses according to scoring criteria in English
Table 4.19 Examples of incorrect responses according to scoring criteria in English
Table 4.20 Examples of correct responses according to scoring criteria in Czech
Table 4.21 Examples of incorrect responses according to scoring criteria in Czech
Table 5.1 Overview of participants – children
Table 5.2 Overview of participants – adults
Table 5.3 Educational background of parents of the participants compared to population samples
Table 5.4 Ethnic background of parents of the participants compared to population samples
Table 5.5 Overview of intraclass correlation coefficient between online and offline scoring
Table 5.6 Mean differences between online and offline scoring across all conditions in English and Czech
Table 5.7 Overview of the intraclass correlation coeficient between rater 1 and rater 2
Table 5.8 Czech-speaking 5-year-olds vs. adults
Table 5.9 English-speaking 5-year-olds vs. adults
Table 5.10 Means and standard deviations for the experimental conditions
Table 5.11 Paired-samples t-tests comparing maximum spans for conditions A and C
Table 5.12 Paired-samples t-tests comparing maximum spans for conditions A and D

Table 5.13 Paired-samples t-tests comparing maximum spans for conditions B and E
Table 5.14 Paired-samples t-tests comparing maximum spans for conditions A and G
Table 5.15 Paired-samples t-tests comparing maximum spans for conditions B and H
Table 5.16 Paired-samples t-tests comparing maximum spans for conditions E and H
Table 5.17 Paired-samples t-tests comparing maximum spans for conditions A and B
Table 5.18 Paired-samples t-tests comparing maximum spans for conditions D and E
Table 5.19 Paired-samples t-tests comparing maximum spans for conditions G and H
Table 5.20 Paired-samples t-tests comparing maximum spans for conditions F and I
Table 5.21 Paired-samples t-tests comparing maximum span for conditions D and I
Table 5.22 Paired-samples t-tests comparing maximum span for conditions D and F
Table 5.23 Mean differences in span for the key comparisons
Table 5.24 Correlations for English- and Czech-speaking children between receptive vocabulary scores and span scores on sentences, lists of words and lists of nonwords
Table 6.1 Durational differences for stimuli entered into planned comparisons
Table 6.2 Nonlexical stimuli with a list prosody: overview of stressed and unstressed syllables in spans 2-4 for English vs. Czech
Table 6.3 Comparisons of sentence vs. list stimuli (conditions D x E and G x H) across three age groups in English and Czech
Table 6.4 Simple and partial correlations between vocabulary measures and memory tasks

List of Figures

Figure 2.1 An example of syntagmatic and paradigmatic relations in language
Figure 2.2 Pitch contours for the monotone and natural intonation conditions (Frankish, 1995)
Figure 2.3 Results for the 6 categories in terms of mean trials to the perfect criterion (Epstein, 1961)
Figure 2.4 Median per cent of complete strings correct for each of the 4 types over 5 trials (Marks & Miller, 1964)
Figure 2.5 The model of working memory proposed by Baddeley and Hitch (1974)
Figure 2.6 The model of working memory (Baddeley, 2000)
Figure 3.1 The interrelatedness of lexical and morphosyntactic information available in input sentences
Figure 3.2 Plausibility of a sentence determined by selection of lexical items
Figure 3.3 Nonlexical status of all items in a sentence
Figure 3.4 Independence of prosodic information in a sentence
Figure 4.1 Linguistic domains present on a sentence level and manipulated in the present study
Figure 4.2 Speech signal for sentence with a sentence prosody (condition A)
Figure 4.3 Speech signal for sentence with a list prosody (condition B)
Figure 4.4 The two orders of presentation of the nine conditions (A-I)
Figure 4.5 Playing field for condition B consisting of 32 ovals representing 32 stimuli
Figure 4.6 Images attached to the practice trials
Figure 5.1 Distribution of the score for BPVS for 50 English speaking children
Figure 5.2 Distribution of the score for the Czech adaptation of BPVS
Figure 5.3 Mean span for experimental conditions A – H in the three English age groups
Figure 5.4 Mean span for experimental conditions A – H in the three Czech age groups
Figure 5.5 Overview of the means for each experimental condition for the four groups
Figure 5.6 Boxplots for English-speaking children from nine experimental conditions
Figure 5.7 Boxplots for English-speaking adults from nine experimental conditions
Figure 5.8 Boxplots for Czech-speaking children from nine experimental conditions
Figure 5.9 Boxplots for Czech-speaking adults from nine experimental conditions
Figure 5.10 The mean differences in span between the key comparisons
Figure 5.11 Correlations between BPVS and sentence span in English-speaking children
Figure 5.12 Correlations between adapted BPVS and sentence span in Czech-speaking children
Figure 5.13 Correlations between BPVS and nonword span in English-speaking children
Figure 5.14 Correlations between adapted BPVS and nonword span in Czech-speaking children
Figure 6.1 Gains in span due to familiarity and/or position of FWs/CWs
Figure 6.2 Prosodic structure of a sentence (Gerken, 1996)
Figure 6.3 Developmental differences and individual differences in digit span (Dempster, 1981)
Figure 6.4 Developmental differences and individual differences in word span (Dempster, 1981)

List of Appendices

A. The full set of stimuli in English

B. The full set of stimuli in Czech

C. List of errors in stimuli

D. List of function words included in the English set of stimuli

E. List of content words included in the English set of stimuli

F. Duration of stimuli (in sec)

G. Stimuli - Number of syllables

H. Visual background for nine experimental conditions

I. Tests of normality

J. Non-parametric analyses addressing research questions 1-7

K. Overview of the results from online and offline scoring

L. Ethical approval of the study

M. Information leaflet for parents, consent form and questionnaire

N. Czech adaptation of BPVS

Acknowledgements

I would like to thank the Department of Language and Communication Science for the PhD Studentship I was awarded which allowed me to conduct this research and complete this thesis. I am especially grateful to my supervisors Shula Chiat and Penny Roy for so generously giving their time, experience, wisdom and support (and biscuits) over the past four years. Many thanks to the children and adults who participated in the experiment and the schools and nurseries who kindly allowed me to carry out the study.

Several people directly helped with particular aspects of this project. Filip Smolík made it possible to obtain ethical approval in the Czech Republic. Anne Zimmer-Stahl helped extensively with planning the recording of the stimuli and was a supportive colleague and friend over the last four years. Alice Lockhart and Dáša Polišenská patiently recorded and rerecorded the experimental stimuli. Thank you to Dáša and my brother Zbyněk for always welcoming me in Prague during my data collection trips. Bernie Coulthrust, Maruška Geráková, Jana Dankovičová and my Father did a great deal to encourage and cajole the schools, nurseries and adult participants. Míša Hejná, Veronika Foltýnová and my mother patiently checked all my materials in Czech and Jo Piper, James Fenton and Linda Doleží did the reliability scoring. Shula, Penny and James also kindly helped to edit and improve my English text and filled in and crossed out many determiners. All remaining errors are mine. Thank you to all of you who encouraged me to come to London to do a PhD and for encouraging me to finish it!

Thank you to all my friends and family for their support.

Abstract

This thesis evaluated the extent to which long-term memory linguistic representations (syntactic, semantic, prosodic and lexical) affect immediate verbal repetition performance. The effects of these linguistic factors on short-term memory span were explored through an experiment with 140 English- and Czech-speaking participants.

The experiment employed nine experimental conditions which varied the presence/well-formedness of linguistic information in four domains: the lexicon, morphosyntax, semantics, and suprasegmental phonology. This resulted in a spectrum of stimuli with semantically, syntactically and prosodically well-formed sentences with real lexical items at one end, through to a list of nonwords with lexical, semantic, prosodic and syntactic information removed. One hundred typically developing children (50 Czech-speaking; 50 English-speaking) aged 4-5 years and 40 adults (20 Czech-speaking; 20 English-speaking) participated in the study. In each condition, participants were asked to repeat blocks of successively longer stimuli to establish their maximum spans.

The results were similar between age groups and across languages. Each linguistic factor had a significant effect on short-term memory span. The presence of nonwords and syntactic violations dramatically reduced memory span, while semantic implausibility and the removal of sentence prosody played a smaller yet significant role. Despite the typological differences between Czech and English, the same robust differences between conditions were found in both languages. The results provide further evidence that immediate verbal repetition is highly sensitive to the linguistic structures present in the stimuli. It is argued that theories which aim to account for data from immediate repetition should not be limited to lexical phonology but also need to address how syntactic, prosodic, semantic and lexical representations contribute to repetition performance. The findings of this thesis support the theoretical framework of verbal short-term memory emerging from a temporary activation of long-term memory representations and reinforce the view that language and memory are inextricable.

Abbreviations

1.SG.	1st person Singular
3.PL.	3rd person Plural
ACC	Accusative
ADHD	Attention Deficit Hyperactivity Disorder
ADJ	Adjective
AOA	Age of Acquisition
AUX	Auxiliary
BPVS	British Picture Vocabulary Scale
CELF	Clinical Evaluation of Language Fundamentals
CNREP	Children's Test of Nonword Repetition
CV	Consonant Vowel Syllable
CW	Content Word
CZ	Czech
ENG	English
F	Female
FEM	Feminine (grammatical gender)
FW	Function Word
ICC	Intraclass Correlation Coefficient
IP	Intonational Phrase
IPA	International Phonetic Alphabet
LOC	Locative
LTM	Long-Term Memory
M	Male
MASC	Masculine (grammatical gender)
MLU	Mean Length of Utterance
MLUm	Mean Length of Utterance in Morphemes
N	Noun
NEUT	Neuter (grammatical gender)
NOM	Nominative
NP	Noun Phrase
NRT	Nonword Repetition Test
NS	Nonsignificant
NWR	Nonword Repetition
PCC	Pearson Correlation Coefficient
PL	Plural
PP	Prepositional Phrase
PRO	Pronoun
PT	Past Tense
R	Repetition
RQ	Research Question
REFL	Reflexive
SD	Standard Deviation
SIG	Significant
SLI	Specific Language Impairment
SR	Sentence Repetition
STM	Short-Term Memory
T	Target
TD	Typically Developing
TROG	Test For Reception of Grammar
V	Verb
VP	Verb Phrase
WM	Working Memory
WMTB-C	Working Memory Testing Battery - Children

CHAPTER 1

The Relationship between Repetition Tasks and Language

1.1 Introduction

Verbal short-term memory is typically measured by exposing participants to linguistic material and then measuring how successfully it can be recalled. This type of measure dates back to the end of the 19th century when Ebbinghaus (1885/1913) conducted the first recorded experiments on memory span (the longest length at which individuals can accurately recall a sequence). Since then, extensive research has explored the relationship between verbal short-term memory (STM) and language in typically developing children, children with language impairment, and adults. A relationship has been found between children's ability to copy nonsense words and their language abilities (Ellis Weismer et al., 2000) and it has been found that children with language disorders have greater difficulty with copying nonsense words than their typically developing peers (Graf Estes et al., 2007). Moreover, strong correlations have been found between the ability to repeat nonwords and later vocabulary development (Gathercole & Baddeley, 1989). Children with low phonological STM, as measured by performance on nonword repetition and digit span, have been reported to show different syntactic profiles with shorter, less grammatically complex utterances (Adams & Gathercole, 1995) and poorer sentence comprehension (Montgomery, 1995). These findings suggest that the abilities reflected in repetition tasks also form some part of the language acquisition process, particularly in the earlier stages.

But while research has highlighted the existence of relations between STM measures and language abilities, their precise nature remains unclear. 50 years of extensive research into memory following Ebbinghaus's first experiments led to limited progress in understanding what abilities are measured by memory spans (Blankenship, 1938). STM measures, and their relation to other cognitive abilities including language, remain a puzzle today. In particular, it is unclear whether performance on verbal STM measures is driven by language competence or whether functioning of verbal STM largely determines language performance. Also unclear is whether constructs such as verbal STM and language can be separated at all. In the case of language-impaired children, there is extensive debate about whether language difficulties are a consequence of poor verbal STM functioning, or if poor performance on standard STM tasks is due to their language difficulties (see Gathercole, 2006 and commentaries on this paper). It was originally suggested that poorer performance on nonword repetition tasks (NWR), and particularly longer nonwords, by children with specific language impairment (SLI) is a reflection of reduced phonological STM capacity. Since Gathercole and Baddeley (1990) proposed nonwords as a pure measure of STM (eliminating, in their view, the lexical and syntactic information present in other repetition tasks such as digit recall or word recall), the deficit on longer nonwords was interpreted as support for the reduced capacity hypothesis. However, recent research has shown that nonword repetition is bound to language knowledge stored in long-term memory (LTM), being affected by factors such as phonotactic frequency, neighbourhood density and prosodic structure (see section 2.2.2). This influence of the nature of to-be-remembered stimuli on verbal recall performance highlights the relationship between immediate recall and linguistic knowledge and suggests that the capacity of STM (if such a construct exists) cannot be defined in purely quantitative terms and that language deficits are unlikely to arise purely from a reduced capacity of STM.

The current study is concerned with the relationship between language and memory in repetition performance in typically developing (TD) individuals. The aim was to determine the effect of key linguistic properties (lexical, prosodic, semantic, and syntactic) on repetition span. To this end, a novel set of tasks was created and given to participants who were asked to repeat verbal stimuli. In total, there were nine experimental conditions, varying the presence/well-formedness of linguistic information in four domains: lexicon, morphosyntax, semantics-conceptual structure, and suprasegmental phonology. This resulted in semantically, syntactically and prosodically well-formed sentences with real lexical items at one end of the spectrum and a list of nonwords where lexical, semantic, prosodic and syntactic information was removed at the other end of the spectrum. Examples are shown below:

Well-formed sentence:	He sent us a letter.
Well-formed sentence with list prosody:	he, sent, us, a, letter
Semantically implausible sentence:	He sang us a kettle.
Ungrammatical sentence with sentence prosody:	A sent he letter us.
Ungrammatical sentence with list prosody:	a, sent, he, letter, us
Pseudosentence with content words replaced by nonwords:	He /fɪnt/ us a /lɔpə/.
Pseudosentence with function words replaced by nonwords:	/vi/ sent /əʃ ʊ/ letter.
Pseudosentence with words replaced by nonwords:	/vi fɪnt əʃ ʊ lɔpə/.
List of nonwords:	/vi//fɪnt//ɒʃ//u//lɔpə/

The contribution of the lexicon, prosody, syntax and semantics was determined by comparing span for sentences in contrasting conditions, e.g. span for semantically implausible vs. semantically plausible sentences, and syntactically ill-formed vs. syntactically well-formed sentences. Most studies have investigated the effects of word-sized items on repetition, with variables specified through the properties of individual items (e.g. high frequency vs. low frequency words, or short vs. long nonwords). A novel aspect of the current study lies in manipulating sequences of items and the relations between the items, rather than primarily the items themselves, and then using span as an outcome measure. In other words, the properties of sequences were varied while holding constant the relevant characteristics of the constituent items (e.g. familiarity, phonological complexity, imageability) to investigate effects on memory capacity and investigate repetition beyond the word level.

Mainela-Arnold and Evans (2005, p.906) suggested that: '*if the processing capacity of children with SLI is directly affected by their long-term linguistic knowledge, this has direct implications on language intervention*'. If the deficit is dependent on LTM, then understanding and addressing the needs of individuals with language impairment may well benefit from better insight into the interplay between language and memory in typically developing populations where long-term linguistic knowledge is less impaired. The current study aims to investigate the relations between long-term linguistic representations and processing through the repetition performance of typically developing children and adults. The main focus was 4- and 5-year-old children, as this age-range is more likely than any other to be receiving treatment from a speech and language therapist (Bercow review, 2008). Adults were included to see if the findings with children could be replicated, and more specifically, whether the children and adults differed quantitatively and/or qualitatively. It has been suggested that phonological STM plays a role in vocabulary/syntax development and therefore the focus is on first language acquisition (see section 1.3.2.1 and 1.3.2.2).

Most studies in the area of verbal recall have been conducted in English. Recently, there has been growing interest in cross-linguistic studies of child language in the hope of discovering more universal mechanisms and processes that drive the course of language development (Leonard, 1998). Findings from studies which focus on one language may primarily reflect specific properties of that language. In order to overcome this limitation, the present study tested analogous stimuli in two languages, Czech and English, which differ in their expression of grammatical relations. While English relies on word order to determine grammatical roles, this is achieved in Czech through the assignment of case in the form of bound morphemes attached to lexical roots.

An additional motivation for conducting this research in two languages was to explore the contribution of repetition tasks as assessment tools in a language other than in English. Repetition tasks are easy to administer, have been found to reliably differentiate typically developing children from children with language impairment in a range of languages and could potentially serve as assessment tools universally. This would be especially useful in languages where no standardised assessments are available. Understanding the contribution of language in these tasks across languages is important for informing the use of sentence repetition (SR) tasks in clinical assessments cross-linguistically.

1.2 Outline of the thesis

The following sections of this chapter introduce established verbal repetition tasks which are relevant to this study (digit span, word span, nonword repetition and sentence repetition) and provide an overview of previous studies which demonstrate an association between repetition performance and language abilities. The potential of nonword repetition and sentence repetition as measures of language development, as well as clinical markers to identify impairment, is discussed and the trade-off between language and memory is addressed.

Chapter 2 highlights the influence of long-term memory knowledge on repetition and reflects on how this might determine STM capacity. Specific consideration is given to investigations of effects of LTM on immediate repetition of single items, sequences of items and sentences. In particular, the effects of syntactic, semantic, and prosodic well-formedness and availability of lexical information (lexical status of words vs. nonwords) on memory span are reviewed. Data from studies on immediate verbal recall are presented and theoretical accounts such as Baddeley's working memory model, Cowan's focus of attention and MacDonald and Christiansen's concept of STM are discussed.

In *Chapter 3,* the different types of linguistic information available in a sentence are discussed and the research questions addressed in the current study are introduced.

Chapter 4 presents the design of the current study's experimental task. Details on materials, experimental conditions, and the motivations behind them are discussed. The procedures for the presentation of the materials and scoring are also outlined.

Chapter 5 describes the participants in the current study in detail and presents the results. Details on differences between experimental conditions are provided along with correlations between a test of receptive vocabulary and i) performance on nonword span, ii) word span and iii) sentence span.

In *Chapter 6,* the main findings are summarised and cross-linguistic similarities and differences and their implications for recall are considered. It is concluded that STM is linguistically structured, and that this language-memory mechanism is in place from a very young age. Limitations of the study are discussed and implications for future research are outlined.

1.3 Repetition tasks and their relation to language

A number of procedures have been devised to measure verbal STM. The modality of presentation, the timing of recall and the properties of the to-be-recalled stimuli vary, but involve the presentation of verbal stimuli that the subjects are asked to verbally repeat. Serial recall requires that items are recalled in order, while in free recall, items can be recalled in any order. Serial recall tasks are the primary focus of this study due to the well documented relationship between serial recall tasks and language ability (see section 1.3.1).

Tehan and Lalor (2000, p.1012) comment on the ubiquity of serial recall tasks and make the point that: '*At face value, recalling a short string of items in order appears to be a very simple task. However, recent research suggests that the task is far from simple.*' The stimuli first have to be heard, perceived, and the phonological forms segmented. The phonological representations then need to be encoded and retained before executing the output. Despite the possible challenges of the perception, storage and production of verbal material involved, most participants are easily able to follow the necessary instructions and the majority of children and adults can effortlessly repeat the stimuli. However, some individuals perform the task with great difficulties and the task taps core skills which seem to be affected in children/adults with language impairment (see section 1.3.2.3 and 1.3.3.3). Interestingly, repetition tasks appear to tap these skills clearly and quickly without the involvement of expensive equipment, reminding of the potential of repetition tasks in identifying language impairment.

The purpose of this section is to provide an overview of verbal STM measures and examine the relations between these measures and language.The sections is divided into 4 further subsections. The first three introduce relations between language abilities and i) serial recall tasks such as digit span and word span, ii) nonword repetition, iii) sentence repetition. In the final section, the trade-off between language and memory revealed by these tasks is discussed.

1.3.1 Span tasks

The digit span task requires that presented items are verbally recalled in order. The participants hear a series of digits (e.g., 8, 3, 4) and must repeat them. If successful, the list is increased by one item (e.g., 9, 2, 4, 1). The length of the longest list a person can successfully remember is that person's digit span. Letters or syllables have sometimes been used as stimuli instead of digits, and in some versions of the task, participants have been asked to repeat sequences of unrelated lexical items. However, the association of words with semantic content introduces an extra variable which is largely avoided with digit or letter recall. Digits have the benefit of having less semantic context than words, are familiar and high-frequency. Moreover, English digits are mostly monosyllabic, helping to avoid potential durational confounds. Because of these properties, digit span is more popular than word or letter recall. Digit span also seems to be a stable measure; it has performed well on test-retest reliability, showing significant correlations at each time of administration (Gray, 2003).

Correlations have been found between span task performance and a wide variety of language skills for TD children as well as children with SLI. Daneman and Case (1981) investigated the learning of an artificial grammar by 2- to 6-year-old typically developing children and they demonstrated that word span was a better predictor of performance than age. Similarly, Adams and Gathercole (1995) found that typically developing children's scores on digit span tasks significantly correlated with morphosyntactic skills. Montgomery et al. (2009) found that digit span significantly correlated with performance on a narrative comprehension task. For children with SLI, Ellis Weismer (1996) reported a link between performance on a word span task and the ability to produce novel morphemes. Gathercole and Baddeley (1990) reported that children with language impairment had an immediate memory deficit for lists of words.

These findings show that performance on span tasks is related to receptive as well as to productive language abilities. Digit span is a stable measure, as documented by strong correlations between scores at different times, and it also correlates highly with other STM measures such as nonword repetition (e.g. Gathercole, Willis, Baddeley & Emslie, 1994). Interestingly though, the correlation between NWR performance and vocabulary level has been reported to be even stronger than between digit span and vocabulary level (e.g. Baddeley et al., 1998), suggesting that the links between language and NWR performance might be even stronger than those between span tasks and language. In addition, poor performance on nonword repetition was a better predictor of language impairment than performance on digit span, and only nonword repetition, but not digit span, was an independent predictor of language impairment (Conti-Ramsden & Hesketh, 2003), demonstrating the limited potential of a span task to discriminate TD groups from groups with language impairment. Furthermore, van der Lely (1990 in Howard & van der Lely, 1995) found no difference on digit span between 14 children with SLI and 18 language-matched controls. Although there are clearly relations between serial recall tasks and language performance, the differences on these tasks do not seem large enough to discriminate between groups of children with language impairment and TD children. Performance on NWR tests, its relation to language and the potential to discriminate the TD and SLI groups are addressed in the following section.

1.3.2 Nonword repetition tasks

The nonword repetition task gained popularity following Gathercole and Baddeley's research in the late 80s and early 90s[1]. Gathercole and Baddeley (1989, p.201) explained that nonword repetition was used as it was simple to administer and '*has the crucial advantage that it uses nonlexical material; the results are not confounded by varying degrees of familiarity with the to-be-remembered material*'. In other words, the nonword repetition task was especially designed to control for the effects of lexical familiarity. By excluding the lexical, semantic and syntactic knowledge associated with lexical items, nonword stimuli eliminated information stored in LTM, and provided what was seen as a pure measure of phonological STM (Gathercole & Baddeley, 1990).

While recall of letters, digits and real words is largely measured in span tasks, nonword repetition tasks more commonly involve repeating single items, although the Working Memory Test Battery-Children (WMTB-C, Pickering & Gathercole, 2001) employs a nonword span task as do some experimental studies (e.g. Hulme et al., 1999; Gathercole & Pickering, 1999). As shown above, performance on span tasks is associated with a variety of language skills, but links between NWR and language skills appear to be even stronger. Evidence for the relations between NWR scores and i) vocabulary development, and ii) syntactic knowledge is examined in the following sections. In addition, the potential of NWR as a clinical marker is discussed.

---

[1] Nonword repetition was used before this time, for instance the Sound Mimicry Test (part of Goldman-Fristoe-Woodcock Auditory Skills Test Battery, 1974) consists of repeating nonwords. However, this test was assumed to assess auditory rather than linguistic/memory skills. Similarly, Snowling and Stackhouse (1983) used NWR for examining motor planning deficits. For a summary of the 'evolution of non-word repetition tasks', see Coady and Evans (2008).

1.3.2.1 Nonword repetition and vocabulary development

A number of studies have investigated the relationship between NWR and vocabulary development. Gathercole and Baddeley (1989) suggested that STM played an important role in acquiring vocabulary. Research on TD children found that even when general intelligence and chronological age were partialled out, scores on nonword repetition at age 4 were significantly correlated with vocabulary knowledge at the age of 5 (r=0.572). In other words, performance on NWR predicted later vocabulary attainment and this was the case even when initial vocabulary knowledge was taken into account (Gathercole & Baddeley, 1989; Gathercole et al., 1992). Similarly, Gupta et al. (2003) reported a significant correlation (r=0.615) between NWR performance and a word-learning task for a group of 70 TD children, aged 5-10 years. But while the close link between NWR and vocabulary knowledge at the age of 4 and 5 years was established, the relationship was little understood.

In a follow-up study, Gathercole et al. (1992) extended their research to older children and reported strong correlations between phonological memory and vocabulary at the ages of 4, 5, and 6 years. The relationship was still present, at age 8, but was less strong. The same pattern of results emerged when age and nonverbal intelligence were controlled for. As in Gathercole and Baddeley (1989), nonword repetition performance at the age of four was significantly related to vocabulary at 5 years. However, age 5 repetition scores were not significantly associated with age 6 vocabulary scores; age 5 vocabulary scores, on the other hand, were significantly linked with age 6 nonword repetition scores. The researchers concluded that although the relationship between vocabulary and nonword repetition remains significant across the age groups, the directionality changes:

> *'The data thus illustrate a shift toward vocabulary becoming the more important pacemaker in the developmental relationship between vocabulary and phonological memory, in contrast to the earlier period of between 4 and 5 years of age when memory skills appear to play the primary causal role' (p.895).*

In other words, the nature of the relation changes once children reach 5 years of age and it appears that beyond this age it is children's linguistic knowledge which has the greatest influence on phonological STM performance.

Most studies have focused on children aged 4 and older and research is limited on younger children, when the greatest amount of vocabulary is being acquired. A recent study by Hoff and colleagues (2008) investigated relations between NWR and vocabulary development in a group of children, aged 1;8–2;0. They tested the hypothesis that '*children who are more advanced in phonological development have more robust phonological representations and are therefore better able to remember new word forms as they encounter them, which in turn supports word learning*' (p. 905). Two experiments from Hoff et al.'s study revealed that NWR accuracy and vocabulary knowledge were significantly correlated. In addition, the study showed that NWR is an easy task to administer and grasp, even with children as young as 20 months. However, it should be noted that the compliance rate differed across Hoff et al.'s two experiments. In the first experiment only 1 out of 16 children refused to participate, but the second experiment had a much higher non-compliance rate of 6 out of 27 children. It remains unclear whether the refusal reflects 'general reticence' or '*a child's awareness that they will be unable to repeat accurately*' (Hoff et al., 2008, p.912). Chiat and Roy (2007), who also administered a nonword repetition task to young TD children (aged 2;0-4;0), had a refusal rate of 6% (18 out of 333 children). In general, very young children are able to perform repetition tasks and most of them cooperate, although there is a proportion of children who refuse to participate. The refusal is difficult to interpret and may mean something other than just a child being shy or unwilling to take part in the study.

To sum up, previous studies show that there is a link between the skills underlying NWR and the skills involved in vocabulary acquisition, and this link is already present in children under 2 years. This link appears to be present also in the later stages of acquisition, but around the age of 5 the nature of this relation appears to change. While in the early stages of language development the ability to repeat nonwords predicts vocabulary attainment, the opposite is true after 5 years of age, i.e. vocabulary knowledge appears to support NWR performance.

1.3.2.2 Nonword repetition and syntactic skills

The issue of whether the link found between NWR and vocabulary extends to other linguistic domains is discussed in this section. The relation between NWR performance and measures of syntactic skills is explored in both production and comprehension. The study by Adams and Gathercole (1995) investigated if NWR is related to measures of productive syntax, while Montgomery (1995) looked at the link between NWR and sentence comprehension.

Adams and Gathercole (1995) carried out a study with 3-year-old TD children who had relatively high or low phonological memory skills as assessed by NWR tasks and digit span. There were 19 children in each group, with a mean age of 3;7 for both groups. For all children, spontaneous language samples were analysed for the diversity of vocabulary items, the length of utterances, and syntactic complexity of the utterances. A quantitative analysis of the transcripts looked at the number of different word roots, the type-token ratio, and the mean length of utterance (MLU) in morphemes. A qualitative analysis of the grammatical complexity of the utterances was obtained using the Index of Productive Syntax (Scarborough, 1990). The results showed that children who scored higher on the STM measures produced a larger set of different words, and their utterances (as measured by MLU in morphemes) were longer than for the low memory group. In addition, the children with poor scores on STM measures produced spontaneous speech that was significantly less grammatically complex. Thus, the findings from this study extended the established relation between repetition performance and vocabulary knowledge and showed that there is also a link between productive syntax and repetition performance. In particular, the study demonstrated that children who scored better on STM measures were also able to produce a wider range of grammatical structures in their spontaneous speech. Adams and Gathercole proposed that the skills reflected by NWR may be vital for learning new syntactic constructions.

Another link between phonological STM measures and language processing at a syntactic level was explored by Montgomery (1995). His study investigated the relationship between measures of phonological STM and sentence comprehension. Participants were 14 children with SLI (mean age = 98.4 months) and 13 TD children (mean age = 81.3 months). The groups were matched on the Test for Reception of Grammar (TROG, Bishop, 1989) to ensure that their receptive grammatical skills, as measured by this standardised test, were equal. All children participated in a NWR task and a sentence comprehension task. In the experimental comprehension task, children were presented with a sentence and four pictures and then asked to match the sentence to the correct picture. A positive correlation was found between NWR performance and performance on the sentence comprehension task. Children scoring lower on the NWR task also scored lower on the sentence comprehension measure. Thus, a relationship between NWR and language was found in yet another area - sentence comprehension.

Together with the findings from Adams and Gathercole (1995) which provided evidence on the links between NWR and production of syntactic structures, the results from Montgomery's study demonstrate that the connections between NWR performance and language also exist beyond the word level. Based on these results, phonological STM was assigned an additional role in language development beyond just vocabulary acquisition and was now seen as a prerequisite for learning new syntactic constructions.

1.3.2.3 NWR as a clinical marker

Given the close links between NWR and language skills reviewed above, individuals with poor language skills may be predicted to perform poorly on NWR. Conversely, if skills involved in NWR are also relevant or even critical in language learning/processing, it can be hypothesised that individuals with poor NWR will score poorly on language assessments, suggesting that NWR has the potential to detect individuals with language impairment. So far, it has been established that NWR performance is linked to vocabulary and morphosyntactic skills and it has been predicted that this link is likely to be reflected in scores on NWR tasks and language assessments. In the following section, the potential of NWR as a clinical marker is explored.

A significant relationship between deficits in NWR and deficits in language has been found in numerous studies and across many languages: English (see Gathercole, 2006, Graf Estes et al., 2007 and Coady & Evans, 2008 for reviews), Dutch (de Bree, 2007; Rispens & Parigger, 2010), Italian (Bortolini et al., 2006), Swedish (Sahlen et al., 1999), Spanish (Girbau & Schwartz, 2007), Icelandic (Thordardottir, 2008), and French (Thordardottir et al., 2010).[2] Due to the potential to discriminate between groups of TD children and children with SLI, NWR tasks have been proposed as a clinical marker. A *good* marker should show the behaviour in question to be present in individuals who have the disorder and be absent in those who do not. NWR has been found to be a relatively reliable clinical marker, showing a high level of sensitivity and specificity[3] in differentiating children with language impairment from typically developing peers (Bishop et al., 1996; Conti-Ramsden et al., 2001). Gray (2003), who investigated nonword repetition in 4- and 5-year-old English-speaking children at three points in time and reported sensitivity above 90% and specificity above 86% across all three testing sessions. Conti-Ramsden et al. (2001) reported sensitivity of 78% and specificity of 87% for English-speaking 11-year-olds. Similar findings have emerged for languages other than English, e.g. Bortolini et al. (2006) reported both sensitivity and specificity of 81% for Italian-speaking children (age range 3;7-5;6). Sensitivity and specificity of 80% and above are regarded as acceptable and values of at least 90% as good (e.g. Bortolini et al., 2006). What should not be overlooked though is the fact that although there are significant differences between TD groups and groups with

---

[2] However, Cantonese-speaking children with language impairment did not score significantly worse than typically developing age-matched children on a nonword repetition task (Stokes et al., 2006).

[3] Sensitivity measures the proportion of actual positives which are correctly identified as such (i.e. the percentage of language impaired children who are correctly identified as having the condition). A sensitivity of 100% means that all language impaired children are identified as language impaired. Specificity measures the proportion of negatives which are correctly identified (i.e. the percentage of unimpaired children who are correctly identified as not having the condition). A specificity of 100% means that all unaffected children are identified as typically developing.

language impairment, there are some children with language impairment who score in the normal range on NWR (Bishop et al., 1996). Bishop et al. (1996, p.401) stressed: '*demonstrating that a test is a good discriminator between impaired and control groups is not the same as demonstrating that it can be used as a criterion for LI in an unselected sample where the base rate of disorder is low'.* Taken together, the results from studies with different age groups and languages showed that both sensitivity and specificity are close to 80% or above, however, it was also shown that not all individuals with language impairment score poorly on NWR.

In contrast, Dollaghan and Campbell (1998) reported that there was no overlap at 99% confidence intervals between groups with language impairment and the TD group on their NWR test. However, Ellis Weismer et al. (2000), who used the same test items as Dollaghan and Campbell (1998) but with a larger sample of 581 children, found an overlap between TD children and children with SLI. This could be due to the nature of the groups with language impairments: the children with SLI in Dollaghan and Campbell's study were children who had been previously diagnosed as SLI and had been referred by speech and language therapists while the children with language impairment in the Ellis Weismer et al. study were identified from a population sample on the basis of a lower score on language assessments administered as part of the study. Ellis Weismer et al. (2000, p.872) concluded that '*poor nonword repetition performance can help to identify children who also perform poorly on standardized language measures but is not sufficient to make this classification on its own'.* Interestingly, Ellis Weismer also found that NWR was more accurate in distinguishing children with language impairment who had received intervention from those without treatment. This finding might be related to the severity of language impairment. While all children with impairment scored at least -1.25 SD on standardised language tests, children receiving therapy tended to be the more severe cases.

NWR seems to be good at confirming the diagnosis of SLI (as in Dollaghan and Campbell's study of children who had previously been diagnosed with SLI), but its sensitivity decreases once it is applied to a random population which includes children with language impairment (according to the criteria for SLI in the study) but have not previously been diagnosed. This finding might indicate that NWR is better at identifying children with severe language impairment. However, it remains the case that NWR tasks have been found to reliably distinguish language impaired groups from unimpaired groups across many languages and that NWR sensitivity and specificity remain high across different times of testing and age groups.

1.3.2.4 Nonword repetition summary

Nonword repetition has been found to correlate strongly with language skills, such as receptive vocabulary, productive syntax and sentence comprehension. It also demonstrates the potential to serve as a clinical marker, showing high levels of accuracy in distinguishing groups of children with language impairment from typically developing children. However, it is not enough on its own to be used as a tool for ruling in/out language impairment in an individual. The potential to discriminate between groups with/without language impairment appears to increase when the language impairment is severe, suggesting that in cases with mild to moderate language impairment there is still enough language knowledge and/or phonological STM capacity to perform well on NWR, while in cases with severe impairment the language knowledge and/or memory capacity is not enough to achieve 'normal' levels on NWR. Then again the heterogeneity of SLI has to be borne in mind: it is possible that some children who have been diagnosed with specific language impairment who score within the normal range on NWR have a different form of language impairment (Bishop et al., 1996).

Although NWR is a very promising marker and remains a strong indicator of SLI, the underlying processes are unclear. NWR may be related to language abilities, which could be because language abilities rely on memory, or vice versa, or both options might be the case. Elimination of the language information as was originally intended (see section 1.1) is in fact not possible in a verbal task. Instead, it might be more informative to turn the argument around and observe what happens when the language information is fully provided. Sentence repetition, which offers this opportunity, is discussed in the section 1.3.3.

1.3.2.5 Nonword repetition vs. span tasks: Same or different?

As previously mentioned, scores on digit span and nonword repetition highly correlate in TD children and adults (Gathercole et al., 1994), but it is not clear if nonword repetition tasks and serial recall tasks tap the same skills. Archibald and Gathercole (2007a) compared serial recall of lists of monosyllabic nonwords and repetition of the same syllables produced as a multisyllabic nonword. Performance of the TD children was compared on matched stimuli in two conditions: sequences of CV syllables were presented in isolation (string condition: e.g. *fow…moy…chee*) or as a single nonword (nonword condition: e.g. *fowmoychee*). Recall was found to be more accurate in nonword repetition than in serial recall and the finding was attributed to additional mechanisms which facilitate recall in NWR. Similar results were found in Archibald et al. (2009), with better performance by TD children on naturally produced nonwords than concatenated syllables with the same phonetic material.

Archibald and Gathercole (2006a) investigated if the same findings held for children with language impairment by comparing performance on measures of STM in 20 children with SLI, aged 7 – 11 years. Recall performance was compared on lists of digits, words and nonwords from the Working Memory Test Battery for Children (Pickering & Gathercole, 2001) and nonword repetition from The Children's Test of Nonword Repetition (Gathercole & Baddeley, 1996). Impairments were found in all of the above mentioned measures, but '*non-word repetition was more sensitive to the deficit in SLI than serial recall*' (Archibald & Gathercole, 2006a, p.687).

Finally, the performance of TD children and children with SLI on these tasks was compared within one study. Archibald and Gathercole (2007b) compared performance on nonword repetition and serial recall in a group of children with SLI and age-matched TD children, aged 7-13 years. There was a significant interaction between task and group: children with SLI scored lower than TD children on both tasks, but the group mean differences were larger for NWR. While presentation of syllables within a nonword facilitated performance for TD children, children with SLI showed a greater deficit in this type of task compared to serial recall. This suggests that serial recall is less sensitive to SLI deficit and that nonword repetition is a more discriminating task.

1.3.3 Sentence repetition tasks

While digit span, word recall and nonword repetition are considered classic measures of STM (Pickering & Gathercole, 2001; Richardson, 2007), the case for sentence repetition as a measure of STM is less clear-cut. Although the task procedure is the same (i.e. auditory recall of verbal material in the presented serial order), sentence repetition is not usually considered or discussed with other STM measures. All repetition techniques discussed above operate at a word level and aim to minimize the contribution of LTM knowledge. As sentences contain more LTM information, it is not unreasonable to hypothesise that sentence recall might draw on LTM knowledge more than STM measures. For these reasons, sentence repetition is considered a language measure rather than a STM measure, as indicated by inclusion of many sentence recall subtests in standardised language assessments (see section 1.3.3.2). However, when evaluating SR performance, concerns are often expressed about the involvement of STM. For example, McDade et al. (1982, p.19) stressed that sentence repetition has been endorsed as an effective method of evaluating grammatical performance, but STM has been demonstrated to be confounding variable by numerous investigators. Thus, it is unclear to what extent sentence repetition reflects language skills and to what extent it is a memory measure. Opinions on this matter vary. For example, Lust et al. (1996, p.56) see sentence repetition (also referred to as sentence imitation, sentence recall or elicited imitation) as a method of assessing children's syntax as SR involves an active reconstruction of the presented model, rather than just being a passive copy. More recently, Ratner (2000) and Vinther (2002) have provided reviews of elicited imitation, discussing the history of the paradigm and considering certain aspects of administration. However, they do not addresses the question of precisely what the technique is testing or the possible confounds from the involvement of memory.

As the task involves verbatim recall of verbal material, it is not dissimilar to other STM measures such as digit or word recall. In addition, performance on sentence repetition tasks has been found to correlate significantly with nonword repetition scores for both TD children and children with SLI (e.g. Conti-Ramsden et al., 2001). Although SR is not typically considered to be a STM measure, STM is often feared to be a confound in the interpretation of the results. For all these reasons SR will be discussed here alongside other STM measures. First, the potential of a sentence repetition task as a language measure is assessed and then its relation to performance on traditional STM measures and language measures is reviewed. Second, the role of sentence repetition in language assessments and its potential as a clinical marker is explored.

1.3.3.1 Sentence repetition as a measure of language development

Although there is no agreement on the merits of verbal repetition as a measure of language ability, let alone the underlying skills the repetition might be testing, elicited imitation is used in language acquisition research (e.g. Valian & Aubry, 2005; Bannard & Matthews, 2008; Kidd et al., 2007; Doleží, 2008; Chiat & Roy, 2008; Riches et al., 2010) and assessments (see below). Sentence repetition tasks have also been used as a language measure in research with a number of clinical populations including: specific language impairment (Montgomery, 2000; Conti-Ramsden et al., 2001; Rispens, 2004), Williams syndrome (Grant, Valian, & Karmiloff-Smith, 2002), phonological disorder (Seeff-Gabriel, Chiat, & Dodd, 2005), Down syndrome (Seung & Chapman, 2004), autism (Riches et al, 2010), ADHD (Redmond, 2005), reading difficulties (Alloway & Gathercole, 2005, Rispens, 2004), Alzheimer's Disease (Small et al., 2000) and aging (Kemper, 1986). Bley-Vroman and Chaudron (1994, p.259) pointed out that elicited imitation appears to be a valuable experimental method in language acquisition and called for further research into the '*the process itself*'.

Sentence repetition as a measure of children's language ability has been systematically evaluated in experimental studies comparing measures of spontaneous speech and scores on SR, for example in early studies with English-speaking children conducted by Carrow (1974) and McDade and colleagues (1982). Carrow's study included a large sample of 475 children, aged 3;0-7;11, while McDade et al.'s study included only nine children, aged 4;4-5;1. Despite differences in the samples, both studies found significant correlations between spontaneous data and SR scores. More recently, Devescovi and Caselli (2007) investigated sentence repetition in young, TD Italian-speaking children. Children aged 2;0-4;0 were tested on a range of sentences of varying length, syntactic complexity and morphology to evaluate whether SR performance was sensitive to changes in language development. As predicted, SR scores increased with age: from 2.13 words at the age of 2;0, to 4.42 words at the age of 4;0. In addition, repetition showed similar patterns to those reported for spontaneous speech: 2-year-old children's repetition was prone to the omission of function words. This is in line with the findings that function words such as articles and object clitics are often omitted in the speech of young Italian-speaking children reported by Bortolini et al. (2006) and Leonard et al. (1992). Moreover, the study showed that SR performance significantly correlated with spontaneous speech measures such as MLU, omission of articles, and number of produced verbs.

Evidence from early studies on English-speaking participants and Devescovi and Caselli's recent study on Italian-speaking participants suggest that sentence repetition performance is in line with scores on measures of spontaneous speech, but there is little research on the relation between measures of STM and sentence repetition performance. One exception is a study by Alloway and Gathercole (2005) with TD English-speaking children. Children were divided into two groups according to their scores on STM measures (digit recall, word recall and nonword repetition) and the two groups were matched on non-verbal IQ scores. Quantitative and qualitative differences were found between the two groups on their SR performance. The low STM group produced a different language profile from the high STM group, making a significantly greater proportion of non-substitution errors (i.e., lexical omissions, additions, no-responses, and order errors), in contrast to the high memory children who made a greater proportion of lexical substitution errors. Based on these findings, it was suggested that *'phonological short-term memory assists the preservation of the structure of the sentence, such as the word order and inflectional markers'* (Alloway & Gathercole, 2005, p.217). Thus, the role of STM was once again pushed far beyond the word level.

Converging evidence suggests that spontaneous speech measures are strongly correlated with sentence repetition scores and that these are correlated with age: as children get older and their language develops, their SR scores increase. However, this finding cannot be unambiguously interpreted in favour of sentence repetition as a language rather than STM measure. If we assume that verbal STM capacity exists, it appears to increase with age, as evidenced by increasing scores on digit, word and nonword span tasks (e.g. Pickering & Gathercole, 2001). Thus, increasing scores on sentence repetition might reflect the developing language, but might also reflect the increasing capacity of STM. Although the correlations between SR and spontaneous speech measures are strong, they cannot resolve the dilemma of SR being a language vs. memory measure. Returning to Alloway and Gathercole (2005), the qualitative differences they found on SR tasks may be more informative in this regard, suggesting that the SR task is sensitive to linguistic structure.

1.3.3.2 Language assessment through Sentence repetition

Sentence repetition tasks have long been part of standardised language assessments, such as *Clinical Evaluation of Language Fundamentals Revised (CELF-R;* Semel et al., 1994) and CELF *UK*-Preschool 2 (Wiig et al., 2004) and the *Test of Language Development-Primary* (TOLD-P; Newcomer & Hammill, 1997). The *Grammar and Phonology Screening* (GAPS) also contains a sentence repetition task (Gardner et al., 2006). The *Sentence Imitation Test* (SIT), part of *Early Repetition Battery,* has been published more recently and comprises sentences controlled for length and syntactic complexity (Seeff-Gabriel et al., 2008). However, little is understood about how typically developing participants deal with sentence repetition tasks and what the relationship is between language and memory in performing these tasks. The current study will address this issue.

1.3.3.3 Sentence repetition as a clinical marker

Although SR has received less attention than NWR, its potential as a clinical marker has been explored, and shown some early signs of promise, in English (Conti-Ramsden et al., 2001; Poll et al., 2010; Seeff-Gabriel et al., 2010), Dutch (Rispens, 2004, van Daal et al., 2008) and Cantonese (Stokes et al., 2006).

Conti-Ramsden et al. (2001) compared four potential clinical markers in a sample of 160 11-year-old English-speaking children with SLI: i) a third person singular task ii) a past tense task iii) a NWR task and iv) a SR task. SR proved to be the most reliable marker, with sensitivity 90% and specificity 85%. The three other markers varied in their accuracy in identifying children with and without SLI: the least useful task was third person singular, followed by past tense and NWR. Furthermore, SR was the only task which was highly correlated with all other markers. The researchers pointed out that it may be clinically useful to use two markers in combination to establish a diagnosis of SLI. However, when the sensitivity and specificity of pairs of markers were examined, the overall accuracy increased by only 1% compared to using a SR task alone. Equally high accuracy was achieved by combining a past tense task and SR (81%), or NWR and SR (81%), while the accuracy for SR itself was 80%.

A similar study was conducted with adult native speakers of English with and without language impairment (Poll et al., 2010). The study evaluated three different clinical markers, NWR, SR and grammaticality judgments, all of which had previously been reported as challenging for populations with language impairment. As with children, scores on SR produced the highest overall accuracy (87%), with both sensitivity and specificity exceeding 80%, compared with overall accuracy for NWR (71%) and grammaticality judgement tasks (ranging from 61% on missing progressive to 77% on omitted finiteness). Again, the markers were combined in order to determine if the accuracy was increased by taking account of performance on more markers. A logistic regression model including SR, the omitted finiteness grammaticality judgment, and the three-syllable NWR measure correctly classified 90% of cases, with a sensitivity of 92% and a specificity of 89%. Thus, the overall accuracy of the model increased by 3% compared to SR accuracy alone (87%) It was concluded that a three-predictor model explained more variability than models based on a single predictor.

Taken together, studies have found SR to be the most reliable marker for English-speaking children and adults with language impairment. In addition, SR has been found to provide almost as much information on its own as when combined with past tense or nonword repetition. The small differences between the accuracy rates in different studies might be due to differences in population (children vs. adults) or due to the markers themselves. Conti-Ramsden et al.'s study tested NWR and productive tense marking in addition to SR, while Poll et al. entered NWR on 3-syllable items only and grammatical judgments into the regression model, and it was these which seemed to yield the additional information. However, both studies indicated that SR has the greatest potential as an individual clinical marker.

Stokes and colleagues (2006) investigated whether these findings could be replicated in Cantonese. They compared the NWR and SR performance of three groups: Cantonese-speaking children with language impairment, TD age-matched and TD language-matched children. The children with SLI (mean age 4;11) did not significantly differ from TD children on NWR, but their scores on SR were significantly lower. Thus, SR was suggested as a clinical marker for Cantonese. Rispens (2004) found that Dutch-speaking 8-year-old children with SLI performed significantly worse on a SR task than their TD peers. Van Daal and colleagues (2008) found that this held for younger children, with significant differences between the SR performance of 5-year-old children with SLI and

TD children. Thus, sentence repetition appears to be a reliable marker of SLI across different age groups and speakers of different languages.

Given that SR is part of many language assessments, it is surprising that most of the sentence repetition tasks do not employ fine-grained scoring which take into account linguistic categories or structures. Typically, scoring is purely quantitative, e.g. CELF scoring: 3 points if no errors, 2 points if one error etc. However, more fine-grained and linguistically motivated scoring may be more informative as it could better reflect the strengths and weaknesses of the participants' language. In a study by Seeff-Gabriel and colleagues (2010), a SR task with fine-grained scoring was used to investigate expressive morphosyntactic skills. Performance of a group of 4- and 5-year-old children diagnosed with SLI and a group of TD age-matched children on the Sentence Imitation Test (SIT-61) was compared and their repetition of content words, function words and inflections was analysed. The results of the recall showed that not all parts of the sentence were equally affected across the groups of children with SLI. While the TD group's means were above 99% for all three syntactic categories, with extremely small standard deviations, the group of SLI children scored significantly lower on all categories: content words 90.1% (SD=10.4), inflections 80.8% (SD=17) and function words 76.8% (SD=17.3). Within the SLI group, content words were recalled significantly better than function words and inflections, and function words and inflections did not differ significantly from each other.

The study showed that SR is sensitive to morphosyntactic categories, with participants with SLI often preserving the content words the best and omitting more function words and inflections. This finding seems to mirror the fact that children with SLI often experience difficulties with functional categories (e.g. Eyer & Leonard, 1995). If one assumes that SR is a language measure, then it is expected that function words and inflections would be challenging for children with language impairment and most affected in sentence recall, and that was indeed found in Seeff-Gabriel et al.'s study. Their findings show that it is not only the number of words, but also the nature of these words, that will determine whether they are preserved or omitted in recall.

1.3.3.4 Sentence repetition summary

The nature of the task is similar in both sentence repetition and traditional measures of STM with both presenting verbal material for subsequent verbatim recall. However, differences emerge with respect to the LTM (lexical, semantic and syntactic) contribution which is apparently absent in STM measures, but present in SR. In classic STM tasks such as digit span or NWR the contribution of LTM is by design diminished, but in SR the contribution of LTM is obviously present. SR is often considered to be somewhere between a STM measure and language assessment and often viewed as a confounded measure with little understanding of what it is actually testing. However, it has been found to correlate with spontaneous speech measures, be a stable measure across time with high levels of test-retest reliability, and its potential as a method of measuring language development has been recognised in language acquisition research. It has also emerged as the most reliable clinical marker of language impairment, showing high levels of accuracy across different populations and speakers of different languages, and providing more information than any other proposed marker. Interestingly, fine-grained scoring which distinguishes between functional and lexical categories in SR in English, revealed a well-established deficit of functional categories present in children with SLI. This suggests that SR is a language-sensitive measure which goes beyond only measuring a quantitative capacity. Yet it is rarely used as a language assessment in its own right and its use remains controversial.

1.3.4 The trade-off between language and memory in repetition tasks

There is broad evidence that measures of verbal STM and language skills are closely related. This link has been documented for both receptive and productive language in children and although originally established only on a word level (vocabulary), evidence now extends to other linguistic domains (morphology, syntax). However, it remains unclear what verbal STM measures are actually testing. Is it memory or language, and are memory and language just two labels/perspectives on one construct? The term 'verbal/phonological STM' suggests one construct embracing both language and STM. When verbal STM was measured with NWR, it was argued that NWR eliminates language information and what is left is STM, so therefore NWR is measuring STM rather than language. In addition, the finding that the repetition of longer nonwords was particularly difficult suggested a quantitative nature of STM capacity. This led to the view that STM is separate from language and that repetition performance is limited by phonological storage capacity. The nonword repetition deficit was viewed as an impairment of phonological storage (Gathercole & Baddeley, 1990). While Gathercole and Baddeley (1990, p. 358) suggested that '*disordered language development may be a relatively direct consequence of phonological memory impairments*', Gathercole (2006, p.514) revised her view and argued that '*an impairment of phonological storage typically accompanies but may not be the sole causal factor in a key disorder of language learning*'.

An alternative account would be that language and memory are one construct and a deficit in NWR would therefore suggest fragile language representations (i.e. in long-term memory knowledge). Gathercole and Pickering (1999) stressed the importance of developing methods that would permit evaluation of whether the low scores on STM measures *'reflect diminished capacities in the phonological short-term memory system or reduced contribution of long-term knowledge in immediate memory' (p.379).* Although the data presented above provide extensive evidence for the close link between language and memory, they cannot answer the question of whether the low scores reflect reduced STM memory capacity, reduced contribution of LTM or whether these are in fact single or separate constructs. The more fine-grained, morphosyntactically motivated scoring used in Seeff-Gabriel et al.'s study indicated that repetition tasks are sensitive to the linguistic nature of items and that repetition is not purely a quantitative measure. But there is great deal of evidence, in particular findings demonstrating the influence of long-term linguistic knowledge on immediate repetition that could reveal more about the memory-language relation. Findings on repetition of single items and studies looking at performance beyond the single item level will be presented together with various theoretical accounts of verbal STM in the following chapter.

CHAPTER 2

The Influence of Long-Term Memory Knowledge on Immediate Verbal Repetition

2.1 Introduction

The relationship between measures of STM and language that was summarised in chapter 1 has been documented by numerous studies. This chapter will more closely examine STM measures and the effects of various types of linguistic knowledge on immediate repetition. First, the notion of capacity of STM is discussed, with consideration of how span can be expanded depending on the nature of items to be recalled. Second, the sublexical and lexical properties of single items and their impact on immediate recall are discussed, and the role of LTM prosodic, semantic and syntactic structures in immediate recall is then considered. Finally, the issue of how models of STM account for these findings is examined, in light of: i) the most prominent model of STM in developmental literature, Baddeley's Working Memory model (STM separate from LTM) and ii) alternatives to Baddeley's model, including Cowan's and MacDonald and Christiansen's account.

2.2 STM capacity and the role of linguistic knowledge

2.2.1 Length and the notion of capacity

As pointed out in chapter 1, nonwords were designed to serve as a pure measure of STM, and therefore were thought to be a suitable measure of STM capacity (Gathercole & Baddeley, 1989; 1993). It was noted that the problems of English-speaking children with SLI in nonword repetition were more marked on longer items with more syllables (e.g. Archibald & Gathercole, 2006b), and similar findings have been reported for other languages (for Dutch, see de Bree, 2007; Spanish, Girbau & Schwartz, 2007; Italian, Bortolini et al., 2006). Some studies have found that children with SLI score significantly lower on nonwords with more syllables. For example, Dollaghan and Campbell (1998) reported that there were significant differences on longer items (3- and 4-syllable), but no differences on shorter nonwords (1- and 2-syllable) between children with language impairment and TD children. Only a significant interaction between group and length was found: in the group of children with language impairment, performance on 3-syllable nonwords was significantly lower than at the shorter lengths. Overall, it seems that the more syllables there are in the nonwords, the more pronounced the differences are between children with SLI and TD children.

Graf Estes et al. (2007) carried out a meta-analysis of studies investigating NWR and compared children's performance on different lengths across the studies. The meta-analysis showed that children with SLI performed significantly below TD children at all nonword lengths, though deficits were greater on 3- to 4-syllable items than on 1- to 2-syllable items. To account for the differences between the studies, Graf Estes and colleagues pointed out that the non-significant differences in some studies might have been the result of a lack of statistical power in the small samples investigated. Their point was supported by Ellis Weismer et al. (2000) who tested 581 children (mean age 95 months), including TD and children with language impairment, and found a significant NWR deficit on all item lengths: children with language impairment performed below the TD group at each syllable length, with larger group differences on 3- and 4-syllable items than 1- and 2-syllable nonwords. Taken together, the longer nonwords seem to be more challenging than the shorter nonwords, with some studies documenting the significantly poorer performance of SLI children only on 3-syllable and longer nonwords. The findings on longer nonwords were interpreted as a reflection of '*a capacity limitation of the phonological component of working memory*' (Gathercole & Baddeley, 1990, p.344) and this view was reinforced by strong correlations between span measures and NWR (Gathercole et al., 1999; Gupta et al., 2003). Gathercole and Baddeley (1990, p. 357) suggested that the reduced capacity of the phonological store could lead either to fewer items being reproduced or the same number of items being held but stored '*less richly*', producing a less adequate memory trace.

Traditionally more associated with span measures, the notion of capacity had been around long before these findings on nonwords. The storage limit was originally defined either as the number of items which can be held in STM (Miller, 1956) or as a time window (Baddeley et al., 1975). The popularity of the item-based approach arose from an article *The magical number seven, plus or minus two: Some limits on our capacity for processing information* (Miller, 1956). The title reveals a view of memory as a system limited in the number of items an individual can recall, in this case seven, with +/- two-item leeway reflecting individual differences. Miller calls the items a 'chunk' but acknowledges that it is not entirely clear what constitutes a chunk of information. The recall of digits, letters or unrelated units appears unproblematic, but difficulties can arise even with digits or letters, since these are not necessarily viewed as independent items. To illustrate the point, consider, for example, a string of letters: /E P H D C/. These can be counted as five unrelated items, but it is possible that the string of letters may be interpreted as /E PHD C/if the abbreviation PhD had been stored in the LTM. Miller talked about 'recoding', or organising the input sequence into chunks. Memory span can be defined as a fixed number of chunks, but the size of the chunk can vary. The advantage provided by chunking is even more obvious with longer sequences. Miller claimed that the capacity of STM is constant when measured by the number of chunks, with a chunk being a subjectively meaningful unit. In contrast to Miller, Baddeley et al. (1975) proposed that STM capacity is limited by temporal duration and they suggested a two-second span (for more details see section 2.2.5).

Thus, findings on nonword repetition and span tasks, and the correlations between them have led to the assumption that there is a verbal STM capacity, but the notion of capacity appears to be more complex than simply the number of items which can be recalled or a fixed time-window. The following section will show that capacity is closely tied to the nature of items and this holds even for nonwords which were supposed to be free of LTM knowledge. In addition, the evidence on the influence of sublexical and lexical knowledge on immediate recall is presented and the close links between LTM and immediate recall are examined.

2.2.2 Influences of sublexical knowledge on immediate repetition of nonwords

Factors related to nonword repetition such as wordlikeness, phonotactic probability, neighbourhood density and prosodic structure will be first discussed in this section, and the effects of these factors on recall performance in TD and on children with SLI will be demonstrated. Second, tests of NWR which differ in some of these factors will be discussed and the implications for assessment will be considered. Third, factors related to lexical items such as frequency, imageability and lexical status will be considered and the question of whether TD children and children with SLI are affected to the same extent by these factors will be explored.

2.2.2.1 Wordlikeness

Wordlikeness was one of the first factors to be investigated. Wordlikeness is based on a judgement/rating which is a product of multiple factors: presence of derivational and inflectional affixes, presence of syllables that are themselves real words, phonotactic frequency, neighbourhood density, and prosodic structure (see below). For instance, the high-wordlike nonword *defermication* contains the derivational prefix <u>de-</u> and the derivational suffix <u>–ation</u>, and its prosodic structure is highly predictable. The item *glistering* contains an inflectional affix <u>–ing</u> and has a close neighbour (i.e. highly resembles the real word *blistering*). Archibald and Gathercole (2006b) compared i) a group of SLI children, ii) an age-matched TD group and iii) a language-matched TD group on low- and high-wordlike nonwords. All items were taken from the *Children's Test of Nonword Repetition* (for more details see section 2.2.2.5) and were rated on a 5-point scale from 1 (very unlike a real word) to 5 (very like a real word). All three groups tended to repeat the high-wordlike items better than the low-wordlike items, although the difference was minimal for the TD age-matched group (see Table 2.1). Statistically, however, the interaction between group and wordlikeness failed to reach significance.

	SLI group	TD language-matched	TD age-matched
High wordlikeness	88.85	91.11	94.43
Low wordlikeness	83.38	86.26	92.74

Table 2.1 Mean percentages of correctly repeated phonemes in nonwords, based on data from Archibald and Gathercole (2006b)

It should be recalled that wordlikeness is based on ratings and that all the items used in this experiment were in fact more wordlike than, for example, items from another widely used NWR test, the *Nonword Repetition Test* by Dollaghan and Campbell (1998). If the groups were compared on items with a greater difference between the low-wordlike nonwords and high-wordlike nonwords, it is possible that there would be a differential effect for the SLI, TD age-matched and TD language-matched groups. The results would be expected to show differences in the same direction as seen with Archibald & Gathercole (2006b) data, i.e. the greatest difference occurring in the SLI group, with the TD language-matched group showing similar effects but with a smaller difference in the TD age-matched children. Archibald and Gathercole (2006b) acknowledged that the TD age-matched group approached ceiling levels, and that this could have masked a greater benefit from high wordlikeness. A recent study by Jones et al. (2010) was carried out with a TD group (mean age 6;0) and an SLI group (mean age 6;1), using items from *Children's Test of Nonword Repetition* for high-wordlike stimuli and newly designed low-wordlike items. A significant effect of wordlikeness was found, indicating that high-wordlike nonwords were repeated more accurately than low-wordlike nonwords. In contrast to Archibald and Gathercole, Jones et al. (2010) found a significant interaction between wordlikeness and group: repetition accuracy for TD children was almost identical: 65% for high-wordlike nonwords and 64% for low-wordlike nonwords while for the children with SLI it was 60% for high-wordlike nonwords but only 33% for the low-wordlike nonwords.

Together, these results show that SLI children were more affected by wordlikeness, and low-wordlike nonwords were found to be more difficult (provided that high-wordlike and low-wordlike items were clearly different as in Jones et al.'s study). It may be of interest to see how children of a similar language level (e.g. matched for vocabulary) perform on these items. Such a comparison could shed more light on the issue of whether performance on wordlikeness can be accounted for by long-term memory language knowledge, e.g. vocabulary size.

2.2.2.2 Phonotactic probability

Closely related to wordlikeness is phonotactic probability. Phonotactics refers to the combinations of phonemes found in a given language. For instance, there are no words in English beginning with /dn/, but this sequence is possible in other languages, e.g. Czech /dnes/, meaning 'today'. Similarly, German allows /kn/ word-initially, but English does not. Vitevitch and Luce (2005, p.193) defined phonotactic probability as '*the frequency with which phonological segments and sequences of phonological segments occur in words in a given language*'. For instance, sequence /mt/ rarely occurs, while sequence /st/ is much more common. Munson (2001) tested TD children and adults on repetition of nonwords containing more and less frequent phoneme combinations. Both groups performed better on nonwords which contained high-frequency combinations in English.

The effect of phonotactic probability was also examined in three groups: children with SLI, TD age-matched children, and TD receptive vocabulary-matched children (Munson et al., 2005). A main effect of probability was found: the nonwords containing high-frequency phoneme sequences were repeated more accurately than those containing low-frequency sequences. A significant main effect of group was also found: the TD age-matched children performed better than the other two groups and the SLI and vocabulary-matched groups did not differ in overall repetition accuracy. Moreover, when the TD age-matched and SLI groups were compared, a significant interaction between group and probability was found: the children with SLI showed a larger effect of probability on NWR accuracy than their TD peers. For the TD groups the interaction was not significant, but there was a tendency for a larger difference in repetition accuracy between high- and low-probability nonwords in the younger TD children. No significant interaction between group and probability was found between the SLI and vocabulary-matched groups (even when age was accounted for).

To sum up, phonotactic probability influenced NWR accuracy, but the groups were affected differentially. The SLI group was influenced more by phonotactic probability than age-matched children, but showed a similar frequency effect to the vocabulary-matched group. Moreover, TD younger children demonstrated a larger effect of phonotactic probability than older TD children. Together, the findings suggest that the differences between the groups could be accounted for by long-term memory language knowledge, in particular vocabulary size, and that the smaller the vocabulary the larger the probability effects on NWR. It appears that as children acquire more vocabulary, they become less affected by phonotactic probability in nonword repetition.

2.2.2.3 Neighbourhood density

Neighbourhood density has been defined as the number of similar-sounding words in the mental lexicon. Coady and colleagues (2010, p. 497) pointed out that '*Neighbourhood density is an indirect measure of phonotactic frequency in that words with many phonological or lexical neighbours contain frequently occurring phonotactic patterns, and vice versa'*. Metsala and Chisholm (2010) examined the effect of neighbourhood density on children's NWR. Their group of English-speaking TD children aged 3;0-7;0 was asked to repeat 24 nonwords, varying in length from 2 to 4 syllables, and neighbourhood density (sparse/dense). It was found that children repeated significantly more nonwords with syllables from dense neighbourhoods than from sparse neighbourhoods. The interaction between density and length was significant: syllables from a dense neighbourhood were repeated more accurately for 3- and 4-syllable items, but density did not affect repetition of 2-syllable nonwords.

2.2.2.4 Prosodic structure

The few studies which have considered prosodic cues in NWR have revealed that prosody also affects repetition. The evidence comes from studies on English and Dutch, and the main focus of these NWR studies is variation in stress and related error patterns.

De Bree (2007) carried out a study on NWR performance with Dutch-speaking children, controlling the length and stress patterns of the nonwords. She investigated patterns of errors in the performance of TD children (mean age 4;5), children with SLI (mean age 4;7) and children at-risk for dyslexia (mean age 4;4). The study used nonwords designed to minimise wordlikeness and speech production difficulties, and varied in the number of syllables (from 2 to 5) and prosodic structure (pre-stressed, stressed, post-stressed syllables). The scores for correct syllables, syllable substitution and syllable omission were calculated for weak and strong syllables. The analyses revealed an effect of group and length: syllable omission and substitution increased as target length increased. For all groups, the majority of syllable omissions were for weak syllables that occurred in pre-stressed positions. In three-syllable nonwords with a weak-strong-weak pattern, syllable omission always fell on the pre-stressed syllable but never on the post-stressed syllable. Likewise, in the case of substitution, it was mostly weak syllables that were substituted: for TD children, 78% substitutions were for weak syllables and 22% for strong syllables. The results for children with SLI were virtually identical: 79% of substitutions were for weak syllables and 21% were for strong syllables. However, differences were found when pre-stressed and post-stressed syllable substitutions were compared: TD children committed more pre-stressed substitutions (75%), but this bias was not present for children with SLI (pre-stressed substitution 51%).

These results are matched by findings by Roy and Chiat (2004) and Chiat and Roy (2007) for English-speaking 2;00-3;11 old children. Accuracy on word- and nonword-items with varied prosodic structure was compared in a group of TD children and a group of children referred to clinical services with concerns about their language. A significant difference was found between the mean rate of syllable loss in the clinic sample compared to the TD group (11.9% vs. 5.5%). Effects of prosodic structure on syllable loss were found in both groups. However, there were differences with respect to the prosodic position of the lost syllables between the typical and clinic sample. TD children rarely omitted a syllable in a stressed position or post-stressed position in 2-syllable items, while this was the case in the clinic sample. Together, these findings from English and Dutch suggest that whether a syllable is prone to omission/substitution is not exclusively determined by length; it is clear that stress also plays a crucial role (stressed vs. unstressed syllables), as does the position of the syllable within an item (pre-stressed vs. post-stressed), TD children and children with language impairment were affected differentially by the prosodic structure. Syllables in strong positions (stressed or post-stressed in 2-syllable items) were rarely affected in TD children, while syllables in these positions were affected for the clinic sample.

The role of prosodic features in NWR was also investigated by Archibald and Gathercole (2007a), discussed also in section 1.3.2.5. This study examined the effect of prosodic contour, coarticulation and acoustic-phonetic salience on repetition of nonwords in TD English-speaking children. Stimuli were produced in two conditions: sequences of CV syllables presented i) in isolation (e.g. *fow…moy…chee*) and ii) as a single nonword (e.g. *fowmoychee*). Multisyllabic nonwords had a natural prosodic contour and coarticulation, which was absent in the serial sequence of syllables. On the other hand, syllables pronounced in isolation had a higher level of acoustic-phonetic salience. The syllabic content in both tasks was kept identical, so the STM load was equivalent. It was found that NWR was more accurate than syllable recall. The storage capacity alone was not sufficient to account for the differences as it was clear that the prosodic contour and coarticulation contributed to the superior performance on nonwords.

It has been demonstrated that the prosodic structure of nonwords affects repetition performance, with weak syllables being prone to omission or substitution. The advantage provided by a prosodic contour was observed when the stimuli were equal in number of phonemes and syllables, thus the difference could not be attributed to storage capacity. Together the findings suggest that having more of the prosodic and phonetic information typically present in naturally spoken language improves performance on nonwords.

2.2.2.5 Sublexical properties reflected in nonword repetition tests

Numerous sublexical properties can be manipulated in nonwords and several sets of nonwords are available for research and/or clinical purposes. The two most commonly used NWR tests, the *Nonword Repetition Test* (NRT) by Dollaghan and Campbell (1998) and the *Children's Test of Nonword Repetition* (CNRep) by Gathercole and Baddeley (1996), will be reviewed here and it will be discussed how the properties of their items might influence performance.

The two tests differ in the number of test items, length of items, scoring method and, most importantly, in the linguistic characteristics of the nonwords. The CNRep consists of highly wordlike items (e.g. *contramponist*) and low-wordlike (e.g. *ballop*), while the NRT only includes items with low-wordlikeness. As pointed out above, wordlikeness is based on subjective and relative ratings and there are no absolute measures for it. While Archibald and Gathercole (2006b) divided the items from CNRep according to a 5-point scale ranging from 1 (*very unlike a real word*) to 5 (*very like a real word*), Jones et al. (2010) considered all items from the CNRep as highly wordlike. Generally, the CNRep is considered to be more wordlike than the NRT, and items from the CNRep, therefore, offer a possibility of drawing on morphological, sublexical and prosodic knowledge stored in LTM to a greater extent than is possible with the NRT. In addition, while items in the CNRep include clusters, weak syllables and reduced vowels, the NRT only has CV or CVC syllable structures, with late-acquired phonemes excluded. Only tense vowels are included, which are longer in duration than lax vowels and '*inherently less susceptible to being reduced to schwa*' (Dollaghan & Campbell, 1998, p. 1138). Items in the CNRep vary in prosodic structure and all resemble the prosodic structure of real English words. Taken together, the CNRep items allow for more support from LTM language knowledge.

Graf Estes et al. (2007) addressed the question of whether the NRT and the CNRep are 'interchangeable'; in addition, sets of nonwords from experimental tasks were included in the comparison. The meta-analysis revealed that effect sizes on the tests were significant and large, indicating that TD children performed better than children with SLI on all versions of the task. However, the CNRep yielded the largest effect size, and the NRT the smallest. In other words, the magnitude of the NWR deficit was associated with the type of task. This result is in line with a study by Archibald and Gathercole (2006b) which compared performance of groups of TD (age-matched, mean age=9;9; language-matched, mean age=6;1) and children with SLI (mean age = 9;8) on the CNRep and the NRT. Both the group of TD children and the children with SLI performed better on the CNRep than on the NRT, and the CNRep also discriminated better between TD children and children with SLI.

The hypothesis that '*nonword repetition is limited by phonological storage capacity*' (Gathercole, 2006, p.520) will be reviewed again here in light of the data presented above. Do the findings presented above reflect reduced capacity of phonological STM or has the performance pattern occurred due to presence of sublexical factors which differentially affect TD children and children with SLI? The studies above demonstrated that sublexical factors significantly influence repetition performance, and in general this held for TD children as well as children with SLI. However, it was also demonstrated that children with SLI were more affected by prosodic structure, wordlikeness and phonotactic probability. In other words, the more opportunity to rely on LTM nonwords offer, the more discriminating they are. This is in line with findings of Graf Estes et al. (2007), see above. They reported that CNRep, which offers a greater possibility of relying on LTM knowledge due to prosodic structure and wordlikeness of its items, yielded the largest effect sizes for performance of children with SLI and TD children.

Performance in nonword repetition tasks appears to rely on LTM knowledge and this has been supported across different age groups and both typically/atypically developing children. A variety of experimental variables have been examined including stress, phonotactic frequency, wordlikeness and neighbourhood density in the nonword repetition tasks and the key role of LTM knowledge has been confirmed. The differences between TD children and children with SLI can be explained by the interaction of sublexical properties of nonwords and knowledge stored in LTM. When the sublexical properties of nonwords are well represented in a language (e.g. the high frequency sequences of phonemes), both TD and SLI children seem to be benefiting from them in NWR. However, TD children seem to be able to better cope with nonwords based on sublexical properties that are less well represented in the input (e.g. low-frequency sequences of phonemes, less word-like nonwords). This leads to a wider gap in the recall performance of children with SLI and TD children. The differences are more notable on nonwords with lower frequency sublexical properties (e.g. nonwords with low-frequency phonemes, nonwords without prosodic structure etc.). This may be because TD children have a more robust language LTM. Less difference between TD/SLI children is found with nonword items when there is a reduced potential for support from linguistic knowledge (e.g. NRT), hindering the contribution of LTM.

2.2.3 The influence of frequency, imageability and lexical status on span tasks

Both Watkins (1977) and Hulme et al. (1991; 1997) investigated the frequency effect on memory span in adults. Watkins compared spans for lists of high and low frequency words and mixed lists of both, with the aim of evaluating the affect of each of the lists on memory span. The mean memory span for high-frequency words was 5.82, while it was significantly lower for low-frequency words, mean=4.24 and similar findings emerged from Hulme et al. (1991; 1997).

Recently, Bannard and Matthews's (2008) study suggested that frequency effects can be found in repetition performance from a young age. They reported that children aged 2 to 4 years repeated frequently occurring lexical-item sequences significantly more successfully than lexically infrequent sequences. Interestingly, Mainela-Arnold and Evans (2005) found that the recall of high-frequency words in children with SLI was similar to TD children, but differed on the low-frequency words. This suggests that the recall of low-frequency words was disproportionately difficult for the SLI population. The researchers investigated whether this finding could be explained by vocabulary knowledge. Once receptive vocabulary scores were entered as a covariate, the significant interaction between group and frequency disappeared, indicating that the difference between the groups was more likely related to differences in receptive vocabulary knowledge.

There is also evidence that concrete words with greater imageability are better recalled than abstract or nonimaginable items (Romani et al., 2008; Acheson et al., 2010). The lexical status of items to-be-repeated also influences repetition performance. For instance, children accurately repeated significantly more words than nonwords in span tasks on the Working Memory Test Battery – Children (Pickering & Gathercole, 2001). A lexicality advantage is also found on the single item level. The repetition of single words is more successful than the repetition of nonwords. For instance, Chiat and Roy (2007) compared the repetition performance of TD children and clinically referred children on 18 real words and 18 phonologically matched nonwords. A significant advantage of words over nonwords was found for both groups.

2.2.4 Evidence of content/function-word distinction in span tasks

Words can be categorised at different levels: i) syntactically: e.g. nouns, verbs, conjunctions; ii) semantically: things, events, states; iii) phonologically: stressed, unstressed (Black & Chiat, 2003). A further categorisation crosses all of these levels: 'open' vs. 'closed' word classes. Open word classes can have new words added to them; in contrast, no new members are usually added to closed word classes. For example, with the progress of technology, many new open class words have emerged: we know what it means to *google* and to receive an *email*. However, it is highly unlikely that new determiners or pronouns will emerge.

When the closed/open word classes are defined through their syntactic and semantic properties, they are often termed *content words* (CWs) and *function words* (FWs). Content words typically carry meaning, while function words involve little or no meaning but rather express meaning in relation to other words. In most languages, open classes include nouns, verbs, adjectives, and adverbs, whereas function words are the 'little' words such as determiners, auxiliary verbs, or conjunctions. Selkirk (1996) pointed out that closed class words also display phonological properties significantly different to those of open class words. Function words bear little or no stress, whereas content words typically carry stress.

Tehan and Humphreys (1988) investigated whether a word-class effect would occur in a recall task. Their experiments were carried out with adults. In their study, the words were matched for number of syllables and spoken length and then divided into six different categories: high frequency adjectives, nouns and function words; low frequency adjectives, nouns and function words. Stimuli were presented visually under a rehearsal condition, and then the same material was presented under a suppression condition. The order of presentation was counterbalanced regarding rehearsal/suppression conditions, as was the presentation of the six categories. Reliable effects were found for all three factors: suppression impaired performance; span for high-frequency words was higher than for low-frequency words; and function words were recalled less well than adjectives and nouns, which did not differ from each other. These word-class differences appeared under both rehearsal and suppression conditions. The researchers commented on the results: '*finding span differences without the corresponding differences in pronunciation rate presents difficulties for an articulatory loop explanation of span or, for that matter, any theory that views span as a simple race between pronunciation rate and trace decay. The data suggest that processes or mechanisms in addition to the articulatory loop are having an effect on span'* (p. 296).

Similarly, Bourassa and Besner (1994) examined serial recall performance in 16 adult participants in an experiment with word class and articulatory suppression as the two variables. The stimuli consisted of 12 content words and 12 function words matched for word frequency, number of syllables and number of letters[1]. Participants were asked to read the words from a computer screen. Half of the subjects were required to count out loud from 1 to 10 while they viewed the words, while the other half was asked to be quiet.

[1] The authors do not explain why they chose to match the two groups of words on number of letters rather than number of phonemes. It may be that letters were chosen because the stimuli were presented visually rather than auditorily.

After reading each list, the participants were asked to write the words down in the presented order. Analyses revealed a main effect of condition (the quiet condition showing better recall than the suppression condition) as well as of the word type (content words were recalled better than function words). The interaction between condition and word type was not significant. Once again, the differences between word classes were found both under suppression and in the quiet condition, which suggests that the content/function word distinction goes beyond phonological properties of the words and reflect more than just the properties of a phonological loop. This was reinforced by the researchers who suggested that '*semantic information plays a role in span performance that is independent of any contribution from the articulatory loop*' (Bourassa & Besner, 1994, p. 124).

Searching for an explanation, Bourassa and Besner (1994) conducted a second experiment where they also matched the stimuli for imageability. The same procedure as in experiment one was used. Suppression still drastically impaired serial order recall but the word-class effect was eliminated when the words were matched for imageability. The researchers inferred that the distinction content/function words did not seem to be reflected in STM as measured in their experiments, but rather that it is imageability that plays a role in serial order recall: '*Experiment 2 showed that this putative word-class effect was eliminated when the words were matched for imageability. Although we are open to the view that some cognitive tasks tap the semantic/syntactic dimension presumed to be elemental to the content-functor distinction, there is no evidence that this distinction is reflected in memory performance as measured here*.' (Bourassa & Besner, 1994, p. 124).

The stimuli used in their study raise questions. First of all, the function words tested in these experiments were not among the most frequent types of function words with no determiners, pronouns or auxiliary verbs. The items chosen would mostly be classified as prepositions (*above, around, among*, *without*), adverbs (*always, really, perhaps, maybe)* or *conjunctions (because*) which have more semantic content than typical function words such as determiners or auxiliary verbs. This was necessarily the case to match content/function words on imageability. Secondly, some of the words are clearly homonyms. For instance, *second* can be an ordinal number as well as a noun. Other homonyms included were *state*, *place*, *while*, or *like*. Most crucially, in some cases the homonyms cut across content/function word classes. Take for example *like*, which can act as a function word (large cities *like* London) or a content word too (I *like* music). *Like* can also be part of a verbal construction (It *looks like* you are going home). Depending on the context, the frequency, imageability and the word class of the stimuli can change, but the words were tested in serial recall without any context. Therefore, it is questionable how valid the finding that function words are recalled equally well once matched for imageability is beyond the stimuli used in this study. The results only seemed to reflect the properties of the tested set of stimuli rather than the properties of function words.

The findings from both studies suggest that differences in the recall of function and content words may reflect the distinctions present on levels other than just the prosodic. While the authors pointed out the possibility that it is not only phonology that accounts for differences, they give little consideration to the other factors which might contribute. They argue for semantic properties being the only defining factors for the function/content distinction, but it is possible that phonological, syntactic and semantic properties together play a role in defining the function/content word distinction.

2.2.5 Is the capacity of immediate recall linked to the nature of items?

Baddeley et al.'s (1975) study is often cited as evidence for the time-based account of STM. The aim of their study was to test whether capacity is limited by a constant number of items or if it is time-limited. If the 'chunk' claim holds, then the length of the words held in STM should not matter. In other words, an individual should be able to hold the same number of long words as short words. To test the hypothesis, participants were given lists of monosyllabic words and lists of five-syllable words to recall, with the words matched for frequency. A very clear advantage was found for all short-word sets. In order to rule out the possibility that the item/chunk is a syllable, two sets of disyllabic words which differed in duration were compared. Once again, the results were inconsistent with the hypothesis that STM holds a constant number of items. The last comparison was carried out on sets of words which were matched for number of phonemes, but differed in duration. In this case the results were also in line with the 'time' theory, i.e. temporal duration accounted better for the results than did the number of items being tested. Overall, Baddeley and colleagues (1975, p. 584) concluded that '*short-term memory capacity, as measured by memory span, is constant when measured in units of time, not in units of structure*'.

It is striking how influential this paper has become in the face of a number of potential shortcomings. First, there was a relatively small number of participants: for half of the experiments it was eight subjects, for the other half the number was slightly higher. Second, there are some concerns about the characteristics of the words chosen for the experiment. The stimuli in different lists were not matched in several key respects: words containing derivational morphemes such as *-tion* (*association*), *-able* (*considerable*) or the adverbial suffix *-ly* (*immediately*) were compared to non-derived words. There is good reason to assume that these derivational morphemes are stored as chunks in LTM (c.f. findings on wordlikeness) which makes the comparison even more difficult. In a similar vein, some words from the longer list were compounds consisting of two (lexical) morphemes

(*Czechoslovakia, Somaliland*), but were counted as one word. In addition, the words were not matched for parts of speech, e.g. the number of nouns was not the same across the lists. Lovatt et al. (2000) attempted to overcome the shortcomings in the selection of stimuli in Baddeley's study. Lovatt et al.'s study compared the immediate serial recall of disyllabic words that differed in spoken duration. Two sets of long- and short-duration words were matched for frequency, familiarity, phonological similarity, number of phonemes, and controlled for semantic associations. In addition, they administered the stimuli from Baddeley et al. (1975) to see if they could replicate the earlier findings. Lovatt and colleagues failed to find consistent word-duration effects with the new stimuli. However, the effects for words taken from Baddeley et al. (1975) were strikingly similar to the effects found in the original study. Lovatt et al. suggested that '*there is no general effect of word duration on disyllabic-word recall, and that the differences originally observed arose as an accident of item selection*' (2000, p. 15).

Although it is appealing to explain capacity in terms of duration, and many researchers have settled for this approach, the supporting evidence is not conclusive. As Lovatt et al. (2000) suggested, the differences initially attributed to duration effects seem to result from the linguistic properties of the chosen stimuli. Or, stated slightly differently, capacity seems to be tightly bound to long-term knowledge. The point that capacity is defined by the nature of items is reinforced by the evidence that properties of the items to be recalled, such as frequency or imageability (see section 2.2.3), influence memory span. Watkins (1977, p. 533) pointed out that scores on span tasks could '*be interpreted in terms of a simple limited capacity system with a rate of capacity consumption that varies with the nature of the item*'. Since language-memory processes are fairly complex, we can anticipate that their interactions will not be easily disentangled: it is necessary to do empirical research to investigate these. One task which could play an important empirical role is repetition beyond the word level. If the stimuli are carefully designed and administered, the task could potentially shed more light on the interactions between language and memory.

The measures discussed so far are all concerned with a single item level. There has been extensive research on words, digits and nonwords and this research points to the conclusion that long-term memory influences and supports immediate verbal repetition performance. Single items offer an opportunity to manipulate their lexical, sublexical and phonological properties. However, the phrase and sentence levels potentially offer much more. As items on these levels are also bound together by means of syntactic, semantic and prosodic relations, the phrases and sentences can provide a better platform for exploring the influence on immediate recall of long-term linguistic knowledge beyond just the lexical and sublexical properties of single items.

2.2.6 Influence of LTM on immediate recall: beyond single-item level

Much has already been written about the recall of single items. However, findings from serial recall studies have provided only limited evidence of the engagement of specific linguistic domains in recall. Even though some studies using serial recall have explored the relations between items in span tasks, the focus is usually on paradigmatic relations, as opposed to syntagmatic relations. A syntagmatic relation refers to the combination of items in a sequence (i.e. 'this-*and*-this-*and*-this'), while a paradigmatic relation can be described as a selection of items which exist in parallel (i.e. 'this-*or*-this-*or*-this'). Figure 2.1 below illustrates both types of relations.

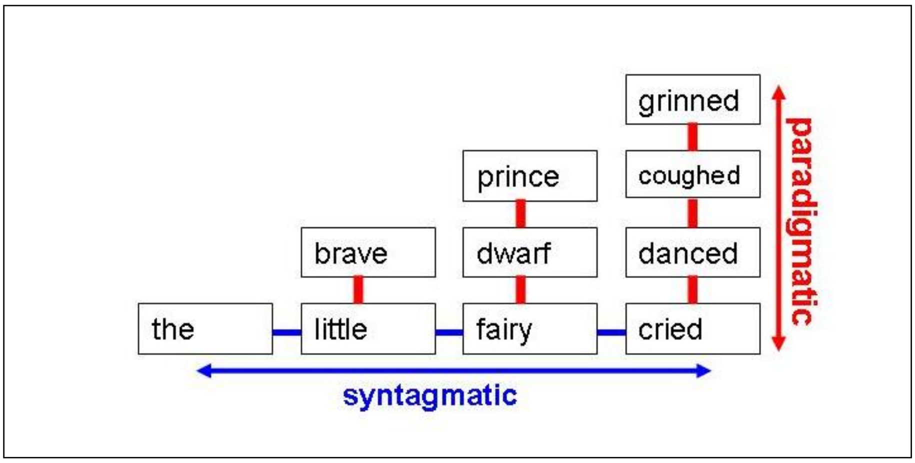


Figure 2.1 An example of syntagmatic and paradigmatic relations in language

Span tasks usually investigate performance on strings of items that are paradigmatically related and compare strings that vary in phonological, syntactic, semantic and lexical characteristics, e.g. phonologically similar versus different, nouns versus verbs, imageable versus nonimageable or real words versus nonwords. However, effects of variation in syntagmatic relations have received little attention in recall tasks to date.

One exception to the neglect of syntagmatic relations is studies on grouping (often referred to as chunking) in serial recall. While studies exploring grouping at a sentence level (i.e. syntactic and semantic) are rare, there are studies which explore the effect of prosodic organization such as rhythm or intonation on serial recall. Although a series of digits or syllables are far from being a sentence, imposing an intonational contour or a rhythmical pattern introduces a syntagmatic relationship between items as some items become more closely related to each other thereby forming a unit.

2.2.6.1 Grouping within a list of digits/words

Frankish (1989, 1995) investigated how prosody might be used to organise items into groups. In particular, Frankish explored whether recall was influenced either by inserting pauses or alternating pitch in the list of items to be recalled. In an experiment focusing on the role of pauses, participants were presented with nine-digit lists in two conditions, grouped and monotonous. In the monotonous condition, all pauses between digits were of the same duration, creating a nine-digit monotonous list. A template would be: 1 2 3 4 5 6 7 8 9. In the grouped condition, a list of 9 digits was divided into 3 groups of 3 digits by inserting extended pauses between the groups, e.g.: 1 2 3 _ 4 5 6 _ 7 8 9. The lists were constructed in a way that ensured no list contained two consecutive digits in ascending or descending order. Participants were asked to write down the digits in the correct order on the response sheets provided. The results demonstrated that pauses (the grouped condition) enhanced recall compared to the monotonous condition.

Frankish (1995) altered pitch within a list of digits in order to impose grouping. In this experiment, digits were presented in four conditions: (1) monotone ('ungrouped'), (2) natural intonation ('intonation grouped'), (3) pauses ('pause grouped'), and (4) intonation plus pauses ('double grouped'). The speakers who recorded the natural intonation condition were instructed to use a strongly-accented intonation to group the sequences of digits into threes without pausing between groups. The speakers were naïve to the aim of the experiment. The pitch contour used for the 'intonation grouped' and 'ungrouped' conditions is illustrated in Figure 2.2 below.

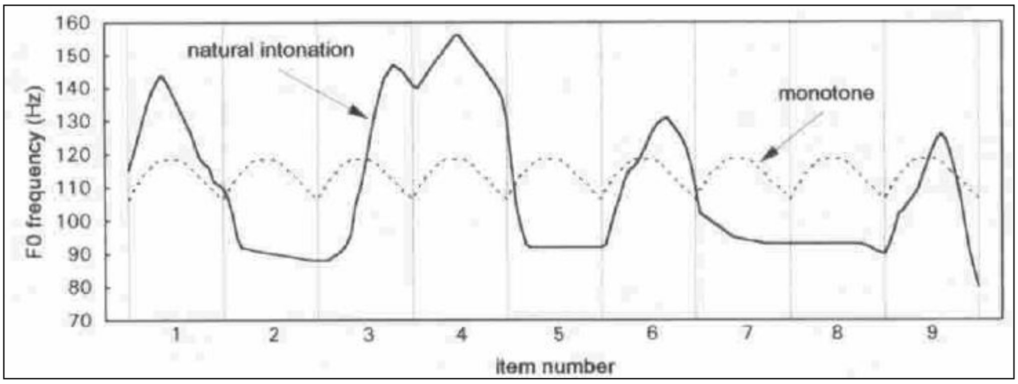


Figure 2.2 Pitch contours for the monotone and natural intonation conditions, reproduced from Frankish (1995, p.9)

Participants listened to the 'grouped' lists (conditions 2-4) and monotone lists (condition 1) and were required to recall the digits in writing. Items in the correct positions were scored as correct. The mean percentage errors at each serial position were counted. It was shown that participants scored better on the grouped conditions than on the monotonous condition, and that the grouped conditions did not significantly differ from each other. Hence, when the grouping methods were combined (intonation + pauses), no further improvement was achieved. The findings from Frankish's studies indicate that prosody positively contributes to STM recall and the advantage appears to arise from the organisation of items (grouping). Combining different methods of grouping had few additional benefits. What seems crucial for superior recall is the organisation of items into groups, no matter what the nature of grouping might be.

A more recent study by Reeves and colleagues (2000) focused on grouping through the addition of stress. The experiments reported in this study investigated whether changing rhythmical patterns affected performance in an immediate serial recall task. If performance improved due to the addition of stress, then at least two explanations may be possible: stress increases the salience of particular items and/or it helps to organise the speech stream into perceptual groups. Performance on stress-patterned and monotone lists was compared. The stimuli were either digits or monosyllabic nouns organised into 9-item lists, either with or without an added stress pattern. A visual representation of conditions is provided in examples (1) to (3) below: a full circle represents stressed items; an empty circle unstressed items. Stressed items were produced at a higher pitch, with greater intensity, and with increased duration relative to unstressed items.

(1) anapaest: ○○●○○●○○●

(2) dactyl: ●○○●○○●○○

(3) monotonous: ○○○○○○○○○

The anapaest pattern consisted of two unstressed items followed by a stressed item, the dactyl pattern consisted of a stressed item followed by two unstressed items, and the items in the monotonous condition were of the same pitch, the same loudness and the same relative duration. Apart from stress patterns, the duration of pauses was manipulated to ensure that rehearsal opportunities were the same for monotone and stress-patterned lists. This study differs from previous ones by not explicitly introducing the notion of grouping to the participants. In Frankish (1989), the grouping boundaries were indicated on the response sheets so the participants were consciously aware of them. Thus, the second aim of Reeves et al.'s study was to investigate whether subjects used stress to place stimuli into groups in the absence of instructions to do so.

Participants were auditorily presented with the lists of items in the conditions described above, and required to recall them in writing as completely and accurately as possible. Data confirmed the effect of stress patterning, revealing an advantage for the dactyl and anapaest over the monotone condition. An assessment of performance at specific serial positions indicated that stress patterns mainly improved recall by organising stimuli into groups, rather than just making certain items more prominent (Reeves et al., 2000). Moreover, the results showed that participants segmented patterned lists into groups during recall without explicit instructions and that rehearsal was not necessary for perceptual grouping effects to occur. Further evidence that grouping was not based on the insertion of pauses was provided by Frankish (1985). This study showed that grouping effects were also produced by having alternate groups of stimuli uttered by male and female speakers and when presented to left or right ears. The insertion of pauses is clearly not the only way to induce a grouping effect.

Reeves et al. (2000) provided a starting point for an investigation into the role of stress patterns in recall. However, note that the auditory modality of the presentation in this study differed from the written modality of the recall. This switch between modalities may have required additional processing resources. Secondly, participants were presented with lists which were nine-item long, exceeding the average capacity of STM (e.g. Miller, 7±2; Cowan 4±1; Baddeley et al., 2 seconds. For more details see section 2.3). This raises the issue of how STM recall can be measured. Reeves et al. deliberately exceeded capacity, as they sought to analyse the participants' performance on items in certain positions and aimed to identify more/less error-prone positions in relation to the stress patterns used.

Quantitative measures can provide an alternative method of assessing repetition performance. Zimmer-Stahl and Polišenská (2009) investigated how prosodic structuring affects immediate STM recall through the presentation of stimuli of increasing lengths. Adult native speakers of British English were asked to repeat strings of nonsense CV syllables in three rhythmical conditions: 'monotonous', 'trochaic' and 'dactylic'. In the monotonous condition, all of the syllables were of identical duration (300 ms), pitch level and intensity. In the trochaic condition a stressed syllable was followed by an unstressed syllable and in the dactylic condition a stressed syllable was followed by two unstressed syllables. The unstressed syllables were created by shortening the vowel durations of the stressed syllables and their intensity was scaled down compared to the stressed syllables. It was found that prosodic structure had a positive effect on recall, with recall performance for dactylic patterning superior to trochaic patterning and both the dactylic and trochaic conditions superior to the monotonous condition. This confirms again that grouping through the production of prosodic structure provided an advantage in recall and immediate repetition was positively influenced by grouping.

2.2.6.2 Beyond the phonetic/phonological level of grouping

Studies on groups of digits/syllables demonstrated that participants show superior performance when items for recall are presented in groups rather than as a series of unconnected items. Sentences are sequences of items that are characterised by multiple levels of organization. In Lasnik's words (2005, p. 1068) '*a sentence is not merely a string of sounds. Minimally, the sounds must be grouped into words. But a sentence is not just a string of words either. The words are grouped together into units, called constituents*'. Grouping is reflected in syntactic, semantic and phonological relations between words. In the following sections, empirical evidence from studies with sentence recall will be presented.

2.2.6.2.1 Syntactic and semantic structuring

In this section, the results of experiments investigating the role of syntax and semantics in recall in adults will be reviewed, as measured by string, sentence and pseudosentence repetition. An early set of experiments carried out by Epstein (1961; 1962) investigated the influence of syntactic structure on learning. He assumed that grammatical structure could be recognised easily, even in an arrangement of nonsense syllables. It should therefore be possible to study the influence of syntactic structure independently of meaningfulness, familiarity, or sequential probability. Epstein (1961, p. 80) pointed out that '*a sentence usually entails a high degree of transitional probability among its components. Nonetheless, a sentence cannot be defined simply as a highly probable sequence of words*.' The hypothesis was that learning nonsense sequences that retain the structure of English sentences would be easier than learning random nonsense sequences. In order to test the hypothesis, stimuli were created to cover six categories of structure, defined by presence/absence of function words/grammatical morphemes, and correct/random order as shown in Table 2.2 below:

Category	Description + example	Syntax
I) Nonlexical	Nonsense syllables in combination with 2 function words; grammatical morphemes present: *A vapy koops desaked the citar molently um glox nerfs*	✓
II) Nonlexical	The same order as in 1; grammatical morphemes absent: *a vap koop desak the citar molent um glox nerf*	×
III) Nonlexical	Material from category 1 presented in random order: *koobs vapy the um glox citar nerfs a molently*	×
IV) Nonlexical	Material from category 1 with grammatical morphemes shifted: *A vapy koobed desaks the citar molents um glox nerfly*	×
V) Lexical	Lexical items in line with syntax; semantically anomalous: *Cruel tables sang falling circles to empty bitter pencils:*	✓
VI) Lexical	Same material as 5 but in random unstructured order: *sang tables bitter empty cruel to circles pencils falling*	×

Table 2.2 An overview of experimental categories with examples, based on examples from Epstein (1961, p. 82)

In total, 192 adults took part in the experiment. They were randomly assigned to six equal groups with each group learning the material of one of the six categories. The materials were presented visually, with each sequence of nonsense syllables/words shown for seven seconds, followed by a 30-second break when the participants were asked to write down what they remembered. If they made a mistake, the process was repeated until the sequence was reproduced perfectly. To be scored as perfect, a reproduction had to include all of the items in the correct order. Errors in spelling were overlooked if they did not affect the pronunciation of the item. Both presentation and recall were visual, and the dependent variable was the number of trials needed for the perfect reproduction of strings. An ANOVA revealed that the main effect of structure was highly significant, and post-hoc tests showed that the following planned comparisons were significant: I vs. II, I vs. III, I vs. V, and V vs. VI. Figure 2.3 illustrates the results, with arrows indicating a significant difference (level of significance .05) between categories.

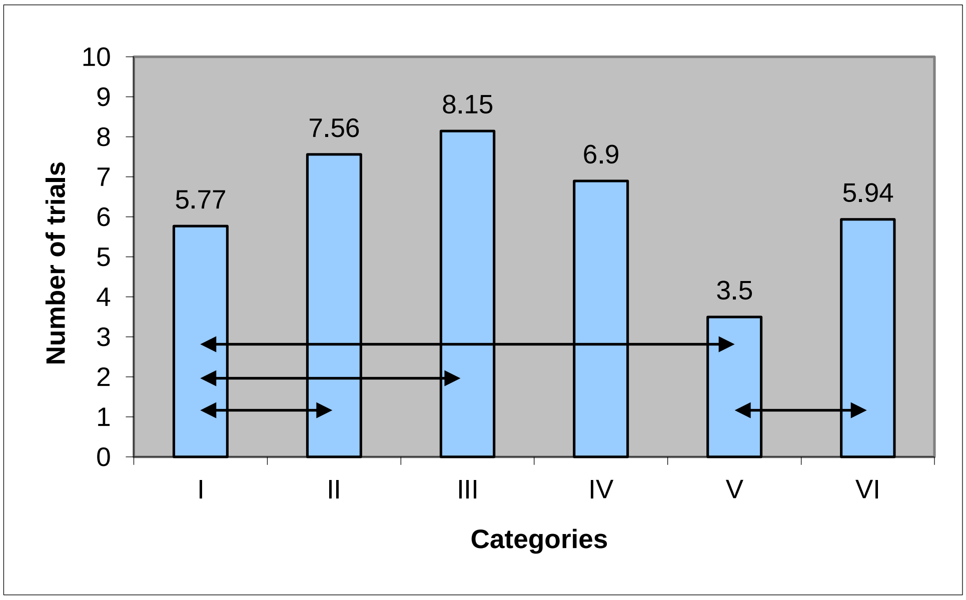


Figure 2.3 Results for the 6 categories in terms of mean trials to the perfect criterion, based on results from Epstein (1961, p. 82)

The results clearly indicated that syntactic structure facilitated learning. Of interest is an absence of a significant difference between categories I and VI, suggesting that the facilitating effects of syntax compensated for the advantages provided by meaning and lexical familiarity. In summary, the experiment showed that syntactic structure facilitated verbal learning apart from the contributions of meaning and familiarity. Epstein (1961, p. 84) pointed out that one of the roles syntax might have is to organise the items into chunks: '*Material which is not syntactically structured may be harder to learn than structured material* *because the latter is already organized whereas the former can be organized into more efficient chunks only through the intentional efforts of the learner*'. This finding was replicated by Epstein with new material (1962). In this experiment the number of categories was reduced to four, but the main characteristics remained the same. The least errors were produced by syntactically structured sentences with real words, then learning of real words in an unstructured order, followed by nonsense words with function words and grammatical morphemes, and finally nonsense words in random order produced most errors. Once again it was shown that syntax played a major role in the remembering of verbal material.

The role of semantic and syntactic constraints in the memorisation of English sentences by adults was also taken up by Marks and Miller (1964). These investigators created normal sentences of five words each, all with identical syntactic structure (adjective - plural noun - verb - adjective - plural noun). These sentences served as a basis for the creation of semantically anomalous sentences which copied the syntactic structure from the normal sentence. These anomalous sentences were created with words from the five normal sentences: the first word was taken from the first sentence, the second word from the second sentence, etc. Since the sentences all had identical syntactic structure, nouns were always substituted with another noun, verbs with verb and so on. Two other types of sequences were derived: anagram strings containing the same words as normal sentences, but with words presented in scrambled order, and word-lists formed by scrambling the word order of the anomalous sentences. Notice that there were no function words in these sequences. Examples (4) – (7) illustrate the four conditions:

(4) Normal sentences:	Noisy parties wake sleeping neighbours.
(5) Anomalous sentences:	Noisy flashes emit careful floods.
(6) Anagram strings:	Neighbours sleeping noisy wake parties.
(7) Word lists:	Floods careful noisy emit flashes.

The strings were recorded and played to 96 adult participants in a quiet room using delayed recall. On each trial they listened to five strings of one type (e.g. 5 normal sentences) and at the end of the block were required to write them down as accurately as possible. In this case, the presentation was auditory, but recall was again in writing and involved a change of modality. Participants were divided into twenty-four groups of four, each receiving strings in a different order to control for order effects.

Unsurprisingly, when the recall of complete strings was scored, normal sentences were found to be far superior to the other three conditions. Anomalous sentences were recalled better than anagram strings, which were in turn recalled better than word lists[2] (see Figure 2.4 for details). Thus, the results demonstrated a beneficial effect of syntax and semantics on recall.

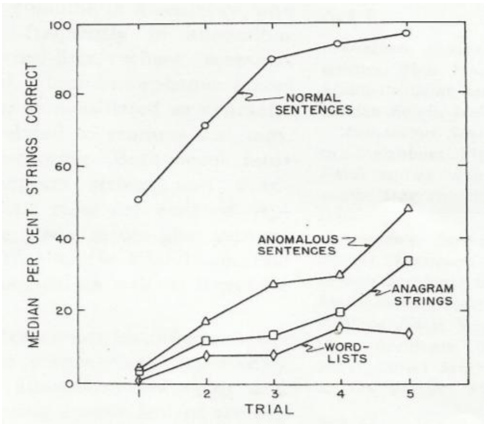


Figure 2.4 Median per cent of complete strings correct for each of the 4 types over 5 trials (Marks & Miller, 1964, p. 3)

2.2.6.2.2 Structuring by means of prosody and morphosyntax

The effect of linguistic constraints on the learning of verbal material was explored by O'Connell et al. (1968). Based on Miller's notion of chunking (see section 2.2.1 for more details), the authors assumed that any stimulus characteristics which provide potential grounds for organising strings into groups should benefit learning. It was thought that both syntax and prosody would offer this potential as both involve hierarchical structures and it was hypothesised that hierarchical structures would facilitate immediate recall: the greater the degree of structure, the better performance on recall would be. The experiments explored this possibility by testing performance on three levels of grammaticality: high structure - morphology and syntax; low structure - morphology alone with nonstructured strings; and no structure; and two levels of prosody: monotone and English sentence intonation[3]. The stimuli were derived from an English grammatical sentence and consisted of 15-syllable nonsense strings (except for the function words in the high and low structured conditions). Examples are presented in Table 2.3.

---

[2] No statistical analyses were carried out and results were given only in percentages.
[3] Although the authors refer to prosody, only the role of intonation was investigated in their study.

English sentence:
A LITTLE BOY LIFTED WITH EFFORT THE BIGGEST WOODEN BLOCKS
High structure (meaningful syllables substituted for nonsense syllables): A HAKY KEX RECILED OU DISON THE KOOTEST PAVA DEEBS
Low structure (groups of syllables which corresponded to meaningful words randomized): KOOTEST HAKY OU RECILED PAVA THE KEX DISON A DEEBS
No structure (randomizing the phonemes in the string while maintaining 15 nonsense syllables): ZUTH VA PA DEEB SEK TOOT DI SON SUH DEE RYL OU KEX HEE KAY

Table 2.3 Overview of experimental conditions and examples, adapted from O'Connell et al. (1968)

Two experiments were carried out, differing by mode of recall (spoken vs. written). In the first, the strings were presented auditorily to 60 adult participants who were asked to repeat them as well as they could. As soon as the recording stopped, the participant was asked to repeat the string. Each string was presented 25 times, reflecting a test of learning as well as recall. In the second experiment, subjects were required to write down the strings to the best of their ability each time a string was presented, and each string was again presented 25 times. Both written and spoken responses were scored for the number of syllables recalled correctly.

Analyses revealed that both the effect of prosody and the effect of structure were significant as was the interaction between them. Both low and high levels of structure significantly facilitated recall in comparison with the nonstructured condition. Intonation was found to facilitate recall in both low- and high-structure conditions. In the second experiment which required written responses, only the effect of prosody was significant and post-hoc analyses showed that intonation facilitated recall at the high level of structure only. The researchers pointed out that the grammatical morphemes were only congruent with the intonation contour of the sentence at the high level of structure and this could explain why intonation was only found to benefit recall at this level. For instance, a pseudosentence with a high level of structure would be easier to relate to sentence intonation than a string of syllables in random order with no structure.

2.2.6.2.3 The content/function word distinction in children's recall

Imitation studies looking at the content/function word distinction in children's recall performance are rare. Scholes (1969, 1970) and Gerken and colleagues (1990) investigated verbal imitation in young children. Scholes (1969) investigated the effect of semantic and syntactic well-formedness in a group of adults and a group of children (age range 3;4-5;10). All participants were presented with word strings in five different conditions:

a) grammatical and meaningful string:	*food is eaten by the cat*,
b) grammatical but semantically anomalous string:	*cat is eaten by the food*,
c) order of the phrases permuted:	*by the cat is eaten food*,
d) order within phrases changed but the order of the phrases kept:	*food eaten is cat the by*,
e) both order within the phrases and order of phrases changed:	*cat eaten food by is the*

The error patterns were analysed and it was found that omission errors highly correlated with word class, i.e. content or function word. Function words were omitted more often (31%) than content words (10%). Especially for the younger children, the omissions of FWs applied across all the conditions. However, there seemed to be a clear developmental pattern: the youngest group (mean age 45.5 months) showed the same percentages of errors in the grammatical strings as the ungrammatical strings. A slightly older group (mean age 47.8 months) made 1.25 times as many errors for ungrammatical than grammatical sentences. The third group (mean age 56.3 months) showed about 1.66 as many errors for ungrammatical as for grammatical sentences and the oldest group (mean age 59.1) showed over twice as many errors for ungrammatical as for grammatical sentences. Adults were error free on the grammatical strings, regardless of their semantic well-formedness. Taken together, function words were more prone to omission than content words and younger children omitted function words regardless of the grammaticality of the string they were in. As ages increased, function word omission decreased in the grammatical sentences and disappeared in the adult performance.

Gerken et al. (1990) investigated whether the omission of function words in children's imitation is mainly motivated by the phonological properties of function words, i.e. being a weak syllable, or by their functioning as morphemes. They tested 16 children aged 23-30 months of age on four string types (p.205):

a) English CW, English FW:	*Pete pushes the dog*
b) English CW, nonword FW:	*Pete pusho na dog*
c) nonword CW, English FW:	*Pete bazes the dep*
d) nonword CW, nonword FW:	*Pete bazo na dep*

The children were divided into two groups according to their MLU: low MLU vs. high MLU. The responses were coded for FW omission, CW omission and CW accuracy. Low

MLU children omitted significantly more FWs (32%) than high MLU children (11%). In addition, low MLU children omitted more English FWs than nonword FWs; no significant difference was found for high MLU children. Gerken and colleagues pointed out that this might be '*because they treated English functors as separate morphemes that added to the morpho-syntactic complexity of the sentence, whereas they treated nonsense functors as simply extra syllables that did not increase structural complexity (p.209)'*. These results were replicated with synthesised speech stimuli, where the phonetic properties of nonwords could be carefully controlled.

Taken together, children tended to retain strongly stressed elements and tended to omit weakly stressed elements. Both low and high MLU children imitated more English CWs than nonword CWs, and low MLU children omitted more English FWs than nonword FW, with no significant difference for high MLU children. Gerken and colleagues explained that as children grow older, their phonology and morphosyntax develop and they become better equipped to produce FWs which are unstressed and carry morphosyntactic information.

2.2.6.3 Summary

The experiments examined in this section have all demonstrated the effects of structure on recall and, more specifically, the effects of sentential syntactic structure/meaning and lexical familiarity on sentence recall. Studies using different methods have demonstrated that each of the core linguistic domains, i.e. prosody, lexicon, semantics and syntax (order plus grammatical morphology), all contribute uniquely to processing input for later recall. It is worth recalling that i) the studies used delayed recall (except for the child studies), ii) stimuli substantially exceeded memory limits, iii) the modalities varied: in most cases either presentation or recall or both were visual, and iv) the participants were adults (with the exception of Gerken and colleagues, and Scholes). Although the current study investigates similar factors to those discussed in the above studies, i.e. the effect of lexical, syntactic, semantic and prosodic factors on recall, there are crucial differences: i) the use of immediate recall, ii) the targeting of capacity by using a span task rather than substantially exceeding it, iii) the consistent use of auditory modality in presentation and recall and iv) the main focus is on children. Additionally, in contrast to the studies above, my stimuli are controlled for other non-experimental variables which are known to affect recall, e.g. frequency, familiarity, imageability (for more details see section 2.2.3).

2.3 How STM models reflect the role of linguistic knowledge

There have been many attempts to produce a theoretical account which would account for the data from recall tasks (for an overview of the various models and approaches attempting to explain the findings from repetition performance, see *Interactions between Short-Term and Long-Term Memory in the Verbal Domain* edited by Annabel Thorn and Mike Page, 2009). Perhaps the most influential[4] theory dealing with verbal STM is the one originally proposed by Baddeley and Hitch (1974) and later developed by Baddeley (1986, 2000, 2003).

This section does not aim to provide a comprehensive overview of STM models. Readers interested in an overview could consult Thorn and Page's (2009) recent publication mentioned above or a review by Jonides et al. (2008). The aim here is to discuss the data presented in sections 2.2.2 - 2.2.6 in light of the two following theoretical accounts of STM: 1) Baddeley and colleagues' model which assumes a separate, modality-specific short-term storage and 2) models assuming that STM consists of activated long-term representations, as represented by Cowan (1988, 1995). Cowan's theory does not specifically apply to language; for application of the theory to the verbal domain, work by MacDonald and Christiansen (2002) and Acheson and MacDonald (2009) is discussed. Both of these theories are in line with Cowan's work, but specifically developed to account for language data. Before discussing the theories themselves, the terminology involved is considered briefly, in particular the distinction between STM and working memory.

2.3.1 Issue of clarification: Short-term memory versus working memory

There is a great deal of disagreement in the literature about how to characterise memory, and what the precise nature of STM and working memory (WM) is. STM strictly refers to the ability to remember a small number of items over a relatively short period. Several 'small numbers' have been proposed, the most popular being the magical number seven (Miller, 1956; see section 2.2.1), but recently the number four has received a great deal of attention (Cowan, 2001). There are also divergent views over the nature of the distinction between STM and working memory. STM is often understood as a passive holding device, while WM is the combination of that holding device along with attentional processes (Engle et al., 1999). In other words, the items are merely stored in STM, while in WM the information is stored but also processed and manipulated (Baddeley, 2000). For example, forward digit span is seen as a measure of STM, while backwards digit span (where participants are asked to remember the digits and recall them backwards), is seen as a working memory measure (Pickering & Gathercole, 2001). Archibald and Gathercole (2006a, p. 677) talk about STM and WM as *'two aspects of immediate memory'*. Cowan (2008)

[4] Between 1980-2006, 7339 of the 16154 papers that cited 'working memory' in their titles or abstracts included citations to Alan Baddeley (Jonides et al., 2008, p. 195).

pointed out that there are differences between recall tasks and other tasks requiring memory and processing. But he also added that *'Whether to use the term working memory for the latter set of tasks, or whether to reserve that term for the entire system of short-term memory preservation and manipulation, is a matter of taste'* (Cowan, 2008, p. 335). Most important here is ascertaining why the repetition tasks discussed in chapter 1 correlate with language skills and what this reveals about language, so for the purpose of this thesis, the terminology as it is used by the authors while describing their theories will be employed.

2.3.2 Multicomponent model: STM and language viewed as two separate constructs

2.3.2.1 Introducing the working memory model

Baddeley and Hitch's (1974) model divides the memory system into three separable components which are assumed to work together (See Figure 2.5). These include a subsystem which deals with verbal and acoustic information, the *phonological loop*, and a parallel visual subsystem for storage and manipulation, the *visuospatial sketchpad*. Both subsystems depend on a control system called the *central executive* which is assumed to be responsible for the attentional control of working memory. This model has been developed and modified as new data have emerged (Baddeley, 1986; 2000).

Figure 2.5 The model of working memory proposed by Baddeley and Hitch (1974), reproduced from Baddeley (2003, p.191)

2.3.2.2 The phonological loop

The most relevant part of the working memory model with respect to language is the phonological loop. The loop comprises two components: a phonological store and an articulatory rehearsal system. '*Traces within the store were assumed to decay over a period of about two seconds unless refreshed by rehearsal, a process akin to subvocalization and one that is dependent on the second component, the articulatory system*' (Baddeley & Hitch, 1974 in Baddeley, 2002, p.86). Evidence for the phonological store includes the '*phonological similarity effect'*, whereby immediate serial recall of items that are similar in sound, e.g. *man, cat, map, cab, can*, is poorer than that of dissimilar items, e.g. *pit, day, cow, sup, pen* (Baddeley, 1966). It was also shown that semantic similarity did not produce the same results. Another effect discussed with regard to the phonological loop is the '*word length effect*' which is connected to the articulatory rehearsal component. Baddeley, Thomson and Buchanan (1975) showed that sequences of shorter items are more likely to be recalled correctly than sequences of longer items, and this was attributed to the fact that longer items take longer to rehearse and will decay more. When subvocal rehearsal is suppressed, e.g. by repeating the word 'the', the word length effect disappears (for more details on this study see section 2.2.5). Baddeley interpreted the word length effect as evidence for decay of information unless refreshed by rehearsal.[5]

As we are dealing with language development, it is worth considering the role of the phonological loop across age span. Baddeley, Gathercole and Papagno (1998) associate the phonological loop with the acquisition of language, proposing that the loop developed in order to facilitate the acquisition of language by maintaining the representation of new words in order to optimise learning (Baddeley, 2003). Baddeley and colleagues (1998) suggested that the function of the phonological loop serves more for learning new words than remembering familiar words. Gathercole and Baddeley (1993 in Baddeley et al., 1998, p. 159) proposed that '*the function of the phonological loop is to provide temporary storage of unfamiliar phonological forms while more permanent memory representations are being constructed*'. This hypothesis was investigated in a study on word learning in 65 5-year-old children conducted by Gathercole et al. (1997). The children were assessed on their ability to learn either pairs of familiar words such as *table-rabbit* or word-nonword pairs such as *fairy-bleximus*. It was found that phonological loop ability (indexed by scores on NWR and digit span tasks) was highly associated with the rate of learning the word-nonword pairs but not with word-pair learning. Thus, it was concluded that the ability to associate pairs of familiar words was quite independent of a phonological loop function. In contrast, the ability to learn new word forms was constrained by phonological loop capacity. Similarly, Jarrold et al. (2009) investigated a sample of typically developing 5- to 8-year-old children and found a relationship between verbal STM measures and ability to learn new phonological forms, but scores on STM measures were not related to learning referents of new words. In summary, the phonological loop did not account for word learning in general, which involves linking the form to a referent in addition to just learning the form. Although the phonological loop seemed to support the first aspect of word learning, it was not associated with the latter.

[5] Jacquemot et al. (2011, p.485) pointed out that '*the standard length effect derives from the number of phonological features to be stored and because of capacity limits, trace decay or interference, longer words are at a disadvantage in recall tasks because more units have to be retained. But, in addition, there is an influence of lexico-semantic factors (Brown and Hulme, 1995; Hulme et al., 1997; Romani et al., 2005).*'

2.3.2.3 The visuospatial sketchpad

The other subsystem proposed by Baddeley and Hitch (1974) was the *visuospatial sketchpad* which is supposed to temporarily maintain and manipulate visuospatial information. The visual and spatial distinction is based on neuropsychological studies. The spatial span is usually measured by the Corsi block task which requires participants to copy a sequence of blocks (Pickering & Gathercole, 2001). The visual aspect is assessed by a pattern task in which subjects are shown matrices of cells, of which a random 50% are filled, and the participants are then asked to recall the patterns.

2.3.2.4 Central Executive: Baddeley (2000; 2003)

The *central executive* is assumed to be responsible for the attentional control of WM. It was originally defined as a limited capacity pool of general processing resources and it decided whether the visuospatial sketchpad or the phonological loop was used and how they interacted with each other. A fourth component, the *episodic buffer*, has recently been added to the model (Baddeley, 2000) and its responsibilities will be addressed in the next section.

2.3.2.5 Episodic buffer

The motivation for the new component was a range of data which could not be accounted for by the previous version of the model. First of all, there is a contrast between immediate memory for sentences and for unrelated words. Word span for unrelated words is typically about five items, while sentence span can reach up to 16 items (Baddeley et al., 1987). Moreover, performance on immediate recall reflects characteristics of items which are supposed to be stored in LTM (see sections 2.2.2 - 2.2.6). The three-component model could not account for the interface of STM and LTM or how information from the two subsidiary systems (the *visuospatial sketchpad* and *phonological loop)* could be bound together. The episodic buffer was proposed as a solution (see Figure 2.6): '*a limited-capacity temporary storage system that is capable of integrating information from a variety of sources. It is assumed to be controlled by the central executive, which is capable of retrieving information from the store in the form of conscious awareness, of reflecting on that information and, where necessary, manipulating and modifying it'* (Baddeley, 2000, p. 421). Adding this new component has drawn attention to the interface between verbal STM and LTM and it has been acknowledged that the phonological loop alone is not enough to account for verbal repetition beyond the word level. All of the parts together create the model of the working memory system. '*The theoretical concept of working memory assumes that a limited capacity system, which temporarily maintains and stores information, supports human thought processes by providing an interface between perception, long-term memory and action*' (Baddeley, 2003, p. 829).

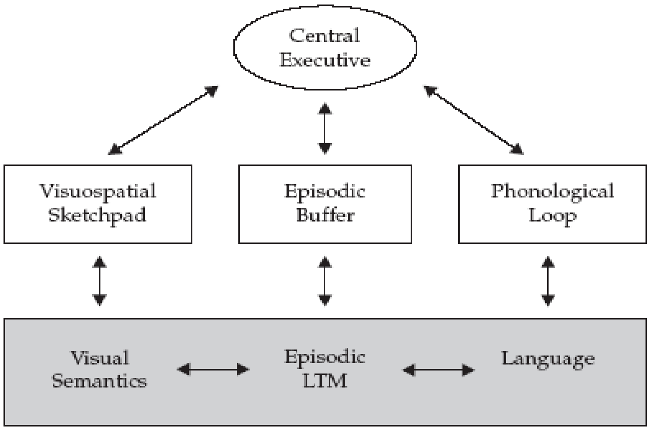


Figure 2.6 The model of working memory (Baddeley, 2000, p.421)

In summary, it was acknowledged that recall advantages due to information stored in LTM were difficult to account for in the older version of the model. The most recent component, the episodic buffer, was added in order to solve the problem and to provide an interface between STM and LTM. However, it is not clear exactly how the information from LTM is made available for immediate recall. The episodic buffer seems to be another store along with the phonological loop, but it is not clear how the information from LTM is stored in this buffer. It also seems that the episodic buffer and language stored in LTM are not directly connected. Yet the model suggests that language and the phonological loop are directly linked. According to the latest version of the model, the interface between LTM and STM should occur in the episodic buffer, but it is not clear why LTM language is connected to the phonological loop and not to the episodic buffer and also when any interaction actually takes place. When discussing data from sentence recall, Allen and Baddeley (2009, p. 78) argued that '*the episodic buffer is involved in retention, as phonological STM would not possess the capacity to store long sequences, nor the capability for semantic or syntactic coding*'.

Although the latest version of Baddeley's model attempted to account for the findings presented in section 2.2.6, acknowledging that LTM has to be taken into account (Allen & Baddeley, 2009), it is not clear how LTM language knowledge contributes to recall and what contributions, if any, are made by specific language domains such as syntax, semantics, prosody and lexicon. Furthermore, it is not clear in what form or code the information in the episodic buffer would be stored (c.f. the phonological loop retains a phonological code consisting of strings of phonemes). The model has been defined as 'a limited capacity system', but it is not clear how the capacity limits are determined and what accounts for individual differences.

2.3.3 Jackendoff's remarks on working memory

Both Jackendoff's concept of WM (2002; 2007) and Baddeley's model reject the notion that WM is part of long-term memory and instead argue that WM is separate from LTM. However, Jackendoff offers an alternative by proposing that WM is '*a set of indices or pointers or transient bindings to long-term memory*' (Jackendoff, 2002, p.205) rather than just a means to copy material from LTM into WM as suggested by Baddeley. Jackendoff (2007) seems to accept that Baddeley's phonological loop may be adequate to describe the processes behind nonword repetition, but the phonological loop in itself does not provide an adequate foundation for what is underlying meaningful speech. Instead, Jackendoff's model suggests that the phonological loop is just one aspect of WM, possibly working alongside additional syntactic and conceptual structures (2002, p.199).

2.3.4 STM as an activation of long-term memory linguistic representations

2.3.4.1 STM as 'focus of attention'

Although Baddeley's model has been popular, the assumption of an STM separate from an LTM has been questioned by many. Cowan (1988, 2001), offered an alternative account where STM is seen as an activated portion of long-term memory rather than a separate construct. Attention plays an important role in Cowan's theory and what can be recalled in memory tasks is determined by the number of chunks held in the '*focus of attention'*. The capacity of the focus of attention is limited for Cowan to four items/chunks (compared to Miller's magical number seven). However, while the number is limited to four, the size of each chunk is not restricted. As pointed out with the definition of a chunk in Miller's theory, a chunk is a subjectively meaningful unit (see section 2.2.1) and it is therefore theoretically possible to increase the capacity simply by packing more information into one chunk. There is more linguistic knowledge stored in LTM relating to the sentence level (compared to the single-item level) and thus a sentence offers more opportunities for forming chunks. This would explain why span for sentences is much longer than for unrelated words in performance on recall tasks. In addition, effects of sublexical and lexical properties such as neighbourhood density, phonotactic frequency and lexical frequency found in immediate recall are to be expected as they reflect properties of LTM. STM is seen as an activated part of LTM and therefore shows identical properties to LTM.

However, difficulties with this theory arise with the recall of items not stored in LTM, a problem noted by Cowan (1995) and also by Baddeley (2000). According to Cowan's account, only information stored in LTM can be activated. For instance, nonwords are not part of our lexicon, but extensive research has demonstrated that they can be repeated. Cowan (2000) offered a solution in the form of rapid learning: '*Immediate memory performance can be accounted for by a combination of information in the focus of attention and activated long-term memory, if the latter includes the results of rapid long-term learning of new associations'* (Cowan & Chen 2009, p. 91). Once rapid learning was added, the theory was also able to account for data which were not part of the LTM. With rapid learning, the new information could be acquired, instantly becoming part of the LTM and could therefore be recalled by activation of long-term memory representations.

Cowan's account is not specific to language, and he also discussed the visual/spatial domain. As the focus of this thesis is on the verbal domain, two theories which were developed specifically for language will be discussed below, but they both pursue Cowan's view of STM as an activation of LTM representations.

2.3.4.2 MacDonald and Christiansen (2002)

As in Cowan's account, MacDonald and Christiansen claim that there is no real distinction between WM and language and that working memory and language form a single construct, making separate impairment impossible. Individual variations in verbal memory tasks do not come from reduced WM capacity, but are instead attributed to: i) variation in exposure to language and ii) biological differences which affect '*processing accuracy, such as differences in the precision of phonological representations*' (MacDonald & Christiansen, 2002, p.36). MacDonald and Christiansen's account is closely linked to connectionist models of language processing and the idea that speed and accuracy of performance is strongly related to language exposure. For instance, frequency effects in word recall are explained through the greater exposure to high-frequency words which in turn leads to more rapid and accurate recognition. Capacity in their view is *'not independent of knowledge'* (p.42). Differences that cannot be explained by experience alone are accounted for by differences in phonological representations, some of which may be inaccurate or incomplete.

How could MacDonald and Christiansen's theory account for the data presented in sections 2.2.2 - 2.2.6 and how does it explain the differences on recall tasks between TD children and children with SLI? In their view, having more experience with language will produce higher spans on STM measures if biological factors are equal. Thus, it would be expected that older TD children who have had more exposure to language will achieve higher scores on STM measures. But if the biological factors which might affect the precision of phonological representations are not equal, individuals will differ in their language knowledge and therefore in repetition performance, in spite of having the same language exposure. Thus, children with SLI would be expected to perform below typically developing peers, even when language exposure was the same, due to different biological factors affecting their ability to build up precise phonological representations. MacDonald and Christiansen (2002) only refer to phonological representations. When it comes to immediate verbal recall, questions remain about the contribution of other linguistic information, e.g. syntactic structure, and if and how linguistic domains other than phonology may contribute.

2.3.4.3 Acheson and MacDonald (2009)

Acheson and MacDonald (2009) insist that language and memory are inseparable and propose that the mechanisms used in recall tasks are the same as mechanisms used during language production. They pointed out that errors in language production often parallel errors in recall tasks, for instance nonimageable words are more likely to be affected than imageable words (Martin et al., 1996). This account predicts that variation in knowledge will affect recall performance, as shown for example in a computational model of nonword repetition by Gupta and Tisdale (2009). In this model, vocabulary size was varied, but the 'memory' (e.g. decay rate) did not vary, and variation in LTM alone could explain variation in NWR performance. Similarly, a study using connectionist networks by Joanisse and Seidenberg (2003) treated working memory and phonology as two '*inseparable and indistinct components of cognitive processing*' (p.54). This study showed that a syntactic deficit in children with SLI could arise from a perceptual deficit which would affect the quality of phonological representations and that deficit would, in turn, affect syntactic processing. Both Gupta and Tisdale (2009) and Joanisse and Seidenberg (2003) treat language and memory as one inseparable construct.

2.3.5 Summary

The interaction of language and memory in language processing appears complex and remains little understood. There is evidence that immediate repetition is influenced by linguistic properties on different levels, even in the case of single words. In spite of this, the contribution of each linguistic domain to repetition performance has been little explored or elaborated on in current models of verbal STM. The interaction between LTM and STM is difficult to capture and this is reflected in the inadequate models of verbal memory (Baddeley, 2000; 2003), which do not explain how lexical, syntactic, semantic or prosodic information contributes to immediate recall. An attempt to resolve this problem was made by treating language and memory as one construct (Cowan, 2001; MacDonald & Christiansen, 2002; Acheson & MacDonald, 2009). However, this explanation is still confined to phonology and it remains unclear how other types of LTM knowledge interact and contribute to immediate recall.

2.4 Implications for language acquisition, assessment and impairment

The conceptualisation of STM has implications for language acquisition, impairment and assessment. According to the models which posit STM and language as two separate constructs, it is possible to develop STM and language independently, with impairment occurring in only one of the constructs so that language impairment can exist without STM impairment and vice versa. This scenario is not possible if STM and language are treated as one construct, suggesting that STM impairment cannot be present without language impairment. Baddeley's account views deficits on STM measures as a reflection of impaired STM, which might result in language impairment (Gathercole, 2006). On the other hand, theories such as Cowan's or MacDonald and Christiansen's would interpret poor performance on verbal recall tasks as a reflection of language knowledge, which could be impaired, missing or at least imprecise. Depending on the view taken, scores on STM measures such as NWR or digit span are likely to be interpreted accordingly.

CHAPTER 3

The Contribution of Linguistic Knowledge to Immediate Repetition: Research Questions

3.1 Introduction

A number of studies have indicated that linguistic representations stored in LTM have an influence on immediate repetition performance (see sections 2.2.2, 2.2.3 and 2.2.6). If single item repetition performance draws on available LTM knowledge, it is expected that sentence repetition should have an even greater reliance on LTM knowledge as the availability of linguistic knowledge at a sentence level is increased. This is because the stimuli in a sentence repetition task will also include some or all of the following in addition to the lexical familiarity of single items: i) word order and/or grammatical morphemes, ii) prosodic hierarchy and iii) conceptual structure encoded by lexical items together with word order and grammatical morphemes. This chapter introduces the research questions which address the contributions of specific language domains to immediate recall and discusses whether these domains can be investigated separately.

3.2 Interconnections

3.2.1 Lexical items and morphosyntactic devices

All languages have lexical items and a variety of morphosyntactic devices which enable the items to be combined to express the conceptual structures[1] found at a sentence level. A lexical item entails a relationship between lexical phonology (form) and lexical meaning (function). In a sentence, lexical items are combined with available morphosyntactic devices to express relations between words. Both the lexical items and the devices (inflections, function words and word order) expressing relations between them need to be acquired for each specific language. After the relation between the form of a lexical item

[1] Jackendoff (2002, p. 123) defined conceptual structure as 'the thoughts expressed by language' which are structured in levels of cognitive organisation.

and its meaning has been stored in LTM, it is expected that this greater familiarity will boost sentence repetition performance.

Lexical items carry more information than merely the form-meaning relation. The properties of lexical items determine these items' roles in morphological and syntactic structures, e.g. the English verb *give* requires three arguments (SOMEONE gives SOMETHING to SOMEONE); the Czech verb *číst* 'to read' requires a direct object in accusative form (in *číst knihu*, 'to read a book', the noun *kniha* 'book' receives a morpheme -u which carries the feature of feminine, accusative, and singular). The lexical and morphosyntactic information available in input sentences and the interrelatedness between these types of information is illustrated by Figure 3.1. Lexical items and morphosyntactic devices are mutually dependent and reveal information about each other. As a result, the presence of information from each item can support performance on other items.

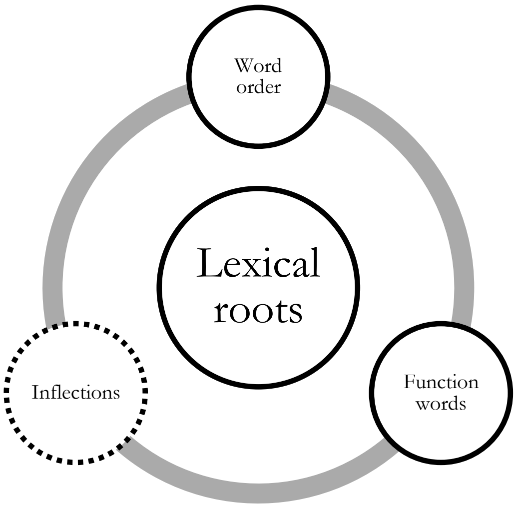


Figure 3.1 The interrelatedness of lexical and morphosyntactic information available in input sentences. The dotted lines signal less independence for inflections compared to function words which can operate independently. The outer circle depends on the inner circle and shows the interrelatedness between inflections, function words and word order in the expression of grammatical relations.

3.2.1.1 Function words

As discussed in section 2.2.4, lexical items can be classified as being function or content words. Function words (FWs) are closed classes of words (not usually added to) and their form and position are often dependent on content words. Christophe et al. (1997) pointed out that the recognition of function words may assist initial segmentation of the speech signal because the forms are familiar and part of a closed class of highly frequent items. This recognition may also help with the division of sentences into syntactic constituents because function words usually require a complement, e.g. the presence of a determiner signals the presence of a noun phrase; an auxiliary entails a VP headed by a lexical verb. In the case of unfamiliar input, function words may help in discovering the forms of content words. While function words have the benefit of being a closed class with a limited number of members, the form of the function word may differ depending on the context and ambient lexical items. Examples include: i) phonologically motivated changes: e.g. the form of the indefinite determiner in English: an apple vs. a pear; or prepositions in Czech: v nemocnici 'in hospital' vs. ve škole 'in school'; ii) syntactically motivated changes: e.g. auxiliary *be* in English am 1.SG. vs. are 2.SG. or the auxiliary in conditional in Czech which also changes according to number and person: bych (1.SG.) vs. bys (2.SG.) vs. by (3.SG./PL.) 'would'.

3.2.1.2 Content words

Content words (CWs) are open word classes and offer many more combinatorial and positional possibilities than function words. Knowledge of CWs may further include subcategorisation information, e.g. where a verb selects a preposition heading an argument: e.g. look *for,* glance *at.* CWs also carry information about the types of inflections they can have, e.g. both English and Czech verbs can be modified by tense inflections. Compared to FWs, CWs in English are less likely to change their form with nouns and verbs that take regular inflections having a consistent form. However, this is less clear-cut in Czech where some CWs also change their surface form according to context. Certain inflectional affixes will trigger consonant alteration in lexical items[2]: e.g.

*uče**s**at 'comb' infinitive*	*vs.*	*uče**š**u 'comb' 1.SG.pres.*
*u**ch**o 'ear' NOM.SG.*	*vs.*	*u**š**i 'ear' NOM.PL.*
*ru**k**a 'hand' NOM.SG.*	*vs.*	*ru**c**e 'hand' NOM.PL.*
*Pra**h**a 'Prague' NOM.SG.*	*vs.*	*Pra**z**e 'Prague' DAT./LOC.SG.*

Similarly, vowels can be changed:

*n**ů**ž 'knife' NOM.SG.*	*vs.*	*n**o**že 'knife' GEN.SG.*
*d**ů**m 'house' NOM.SG.*	*vs.*	*d**o**mu 'house' GEN./DAT./LOC.SG.*
*p**e**s 'dog' NOM.SG.*	*vs.*	*psem 'dog' INST.SG.*

Furthermore, all voiced obstruents lose their voicing word-finally, e.g.

/vjɛʒ-ɛ/ 'tower' GEN.SG.	vs.	/vjɛʃ#/ 'tower' NOM./ACC.SG.
/plod-u/ 'fruit' GEN.SG.	vs.	/plot#/ 'fruit' NOM./ACC.SG.

[2] A similar process can be found in a case of morphologically triggered allomorphy in English, e.g. *critic - criticise, analog - analogy* (Spencer & Zwicky, 2001).

3.2.1.3 Word order and inflections

In English, the subject precedes the verb and the object normally follows it (SVO). The establishment and recognition of order helps to determine grammatical and thematic roles (who did what to whom) and is crucial in interpreting the verb-argument relations. In contrast, Czech allows freer word order and relies on case markings to disambiguate subjects from objects. For instance, 'Peter kicked Paul' could be either *Petr kopl Pavla* (SVO) or *Pavla kopl Petr* (OVS) and both are grammatical expressions of the same thematic structure. Case marking, rather than word order, assigns the grammatical roles (nominative for subject, marked by zero morpheme *Petr-*Ø, and accusative for direct object, marked by -a *Pavl-a*). The relation between inflections and lexical items on a sentence level (e.g. how some inflections trigger morphophonological changes within content words) was mentioned in the previous section.

It has been shown how lexical items and morphosyntactic devices are interdependent and how together they determine the conceptual structure. The following section introduces the research questions targeting the impact of linguistic familiarity on immediate recall and discusses how it is possible to investigate the separate contribution of these different domains to sentence recall despite their interdependence.

3.3 Research questions

Four research questions (RQs) were formulated to investigate the contribution to immediate verbal recall of i) plausibility of conceptual structure, ii) morphosyntax, iii) lexicon and iv) prosody. Three further research questions extended this investigation by exploring the content/function word distinction in sentence recall.

3.3.1 RQ 1: Conceptual structure

(I) *Is span for semantically plausible sentences greater than span for semantically implausible sentences?*

While in sentence production one starts with an idea and encodes it through language, the idea is already present in a sentence repetition task's input sentence. Conceptual structure is a product of, and inseparable from, the lexical roots and morphosyntactic structure of the sentence. The way the lexical items are combined within a sentence creates sentential meaning or 'conceptual structure'. However, this structure can vary in the degree of plausibility and the extent to which it is '*incorporating pragmatic considerations and "world knowledge*"' (Jackendoff, 2002, p.123). So, what can be separated from the sentence syntax and semantics is the semantic plausibility or probability of the sentence meaning, which relies largely on the combination of lexical roots. The combination of specific lexical items will determine how plausible or semantically anomalous a sentence is. Therefore, lexical items impact on the plausibility/conceptual structure, as illustrated by Figure 3.2. The figure also illustrates the relative independence of the conceptual structure, compared to the interdependence of lexical items and morphosyntactic devices (and semantics determined by these).

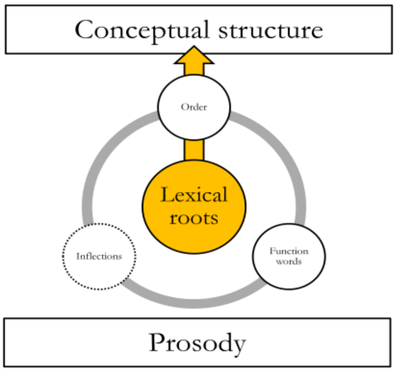


Figure 3.2 Plausibility of a sentence determined by selection of lexical items

The degree of plausibility can be varied independently of variations in syntax or prosody. The lexical items, syntactic structure and prosodic organisation might be familiar, while the meaning of the whole sentence might be implausible and unfamiliar, hence the novelty of the conceptual structure. For instance, the sentence 'He sang us a kettle' invokes an implausible and unfamiliar idea, although the lexical items 'sing' and 'kettle', plus function words 'he', 'a', 'us' plus the argument structure '*someone sings someone something*', and phrasal structure, e.g. '*determiner + noun*' are all familiar and produced with familiar prosodic organisation. The first research question explores whether reduced familiarity and/or plausibility of conceptual structure expressed by the input sentence impacts on repetition performance.

3.3.2 RQ 2: Morphosyntax

(II) *Is span for grammatically well-formed pseudo/sentences*[3] *greater than span for grammatically ill-formed pseudosentences?*

Separating morphosyntax while keeping other linguistic domains intact is problematic[4]: manipulating morphosyntax (e.g. by disrupting the order within and between phrases) affects the syntactic relations and also the semantics and prosody. The meaning relations of

[3] The term pseudo/sentences is used as an abbreviation for 'sentences + pseudosentences', more on this distinction in section 4.1.1.
[4] It might be possible to some extent by violating agreement, but this was not applied in the current study for reasons outlined in section 4.2.3.1.

the sentence inevitably change and are disrupted, though lexical and pragmatic information in English and information expressed by grammatical cases in Czech may allow some relational meaning to be computed. Therefore, it is not possible to completely separate morphosyntax from sentence semantics. Due to disrupting the order within and between phrases, prosodic structuring is altered since stressed and unstressed items change their positions. Given that familiarity of lexical items benefited verbal recall (see section 2.2.3), it is expected that familiarity of syntactic structures will affect sentence repetition performance. This issue is addressed by the second RQ. However, it should be recognised that any effects of morphosyntactic disruption necessarily entail semantic disruption.

3.3.3 RQ 3: Lexical items

(III) *Is span for pseudo/sentences comprised of real lexical items greater than for pseudosentences made up of nonwords?*

A sentence with nonlexical items necessarily affects the lexical semantics and the meaning of a sentence (Figure 3.3). It must be recognised that if all of the lexical information is unavailable, the syntax and semantics of the sentence cannot be preserved either, as lexical roots and function words would be unfamiliar.

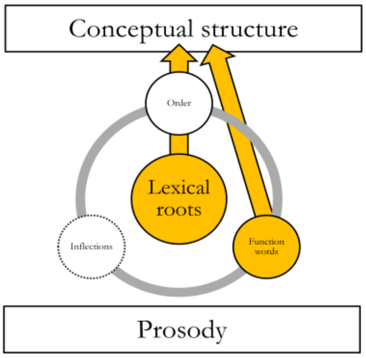


Figure 3.3 Nonlexical status of all items in a sentence

3.3.4 RQ 4: Prosody

(IV) *Is span for pseudo/sentences presented with sentence prosody greater than for pseudosentences presented as a list?*

Spoken sentences are organised into hierarchical prosodic structures rather than simply being a linear sequence of morphosyntactically organised lexical roots. For instance, an English utterance consists of sequences of stressed and unstressed syllables: *the* ***sky*** *is* ***clear*** *and the* ***clouds*** *are* ***gone.*** Similarly, stressed and unstressed syllables alternate in a Czech utterance: ***Vče**ra jsem **vi**děl **no**vý **film.*** 'I saw a new movie yesterday'. Prosodic structure relates to the syntactic organisation of lexical items: function words are usually unstressed, while content words carry lexical stress. Furthermore, the stress pattern of a sentence depends on the position of function/content words within the phrasal and sentence structure. The example below (Jackendoff, 2007; bold font is mine) illustrates how the stress and intonation of prosodic organisation directly relates to the lexical items and syntactic organisation of the sentence[5] and when this is not respected, ill-formed output results:

Phonology:*[Sesa**me**][**Street is** a][**pro**duc**tion of** the Children's][Te**le**vi**sion** Work**shop**]

The noun 'production' would usually carry stress on the second syllable, the function word 'of' would typically be unstressed, and Sesame Street would syntactically be one constituent: [Sesame street].

While order, inflections and lexical roots interact and provide information about each other and determine normal prosody, prosody does not impact on morphosyntax and it is possible to separate prosodic structuring from the morphosyntactic organisation of lexical roots. For instance, instead of an utterance with a natural prosody such as [The **li**ttle **girl**] [**lost** [her **doll**] [at **school**]], the sequence can be presented with each item stressed, as in [**the**] [**little**] [**girl**] [**lost**] [**her**] [**doll**] [**at**] [**school**], with morphosyntactic and semantic relations preserved. Thus, in contrast to lexical items and morphosyntax, prosody can be manipulated without 'side effects', as illustrated in Figure 3.4.

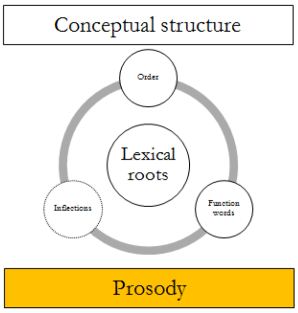


Figure 3.4 Independence of prosodic information in a sentence

It is nevertheless possible that repetition performance may benefit if the input sentence is prosodically organised in a familiar way, i.e. i) in line with the properties of lexical items

[5] However, phonological and syntactic structuring do not necessarily overlap exactly (Jackendoff, 2007):
Syntax: [This is [the cat [that chased [the rat [that ate the cheese]]]]]
Phonology: [This is the cat] [that chased the rat] [that ate the cheese]

(function words unstressed and content words stressed) and ii) intonational and prosodic phrasing respecting syntactic structure. This is investigated by the fourth research question.

3.3.5 RQs 5 – 7: Function words versus content words

(V) *Is span differentially affected by lexical familiarity of content words vs. function words?*

(VI) *Is span differentially affected by the familiarity of form vs. the familiarity of position of function words?*

(VII)*Is span differentially affected by the familiarity of form vs. the familiarity of position of content words?*

If familiar function words are preserved in a model sentence, syntactic structure can be maintained to some extent, permitting computation of meaning at an abstract level, e.g. *The /skaɪs/ is /maɪdɪŋ/ in the /krɑf/*: function words indicate NP $_{VP}$[V $_{PP}$[P [NP]]. If only content words are familiar, conceptual structure is affected as the familiar content words are deprived of the devices indicating their interrelations. Knowledge of lexical items and potential pragmatic relations between these and preserved order may also contribute to the computation of sentence meaning, as seen in the example */tʃə/ snake /ʊz/ hiding /ɪp tʃə/ grass.* Familiarity of lexical items, which also encode syntactic and semantic information, will inevitably influence sentence repetition performance. RQs 5-7 investigate the contribution of FWs and CWs to recall, in particular the contribution of familiar form and position.

3.3.6 RQ 8: Relations between repetition performance and a vocabulary task

(VIII) *Is performance on a receptive vocabulary task related to sentence, word and nonword spans?*

Relations between receptive vocabulary and a variety of repetition tasks have been found in numerous studies and the evidence was reviewed in section 1.3.2.1. These studies usually focus on NWR tasks and how scores on these tasks correlate with receptive vocabulary knowledge. The relation between span tasks (word span, nonword span, and sentence span) and receptive vocabulary will be explored here.

3.3.7 RQ 9: Cross-linguistic comparisons

(IX) Do the planned comparisons addressing research questions 1-7 produce different results in English and in Czech?

The rationale for cross-linguistic research was discussed in section 1.1 and differences between the English and Czech languages are discussed further in chapter 4. The ninth research question investigates whether the patterns of results differ between the two languages, which would suggest effects of typological differences between English and Czech.

3.4 Implications of the interrelations between linguistic domains

As shown above, the plausibility of a model sentence and its prosody can be investigated separately, but grammatical relations are expressed by an interdependent combination of familiar lexical items and morphosyntactic devices. It follows that it should be possible to determine the contribution of plausibility and prosody to memory span, as these two factors are relatively easy to separate on a sentence level. However, it is not possible to unravel the contribution of each of the language domains, in particular the syntax-semantics interface. A sentence is a combination of information from all of the above mentioned sources: prosody, conceptual structure, syntax and the lexicon. These were not treated as separate variables. Instead the contribution of linguistic domains to repetition was determined by comparing spans for sentences in contrasting conditions, e.g. spans for semantically implausible vs. semantically plausible sentences, or syntactically ill-formed vs. syntactically well-formed sentences (see chapter 4 for details). Therefore, the results will also be presented in the form of differences between the experimental conditions.

CHAPTER 4

The Design of the Experimental Task: A Memory Span Measure beyond the Word Level

4.1 Introduction

The current study addressed the interaction between memory capacity and language knowledge/processing in a novel way by systematically manipulating linguistic factors in an immediate repetition task. The aim of the study was to determine how the availability of lexical, semantic, syntactic and prosodic structures affects memory span and thereby evaluate the importance of long-term memory representations for short-term storage, as measured by immediate repetition of sentences, sequences of words, and sequences of nonwords. The stimuli were created by manipulating the following: i) the lexical status of constituent items occupying content and/or function word slots in a sentence (words vs. nonwords), ii) the well-formedness of sentence syntax (grammatical vs. ungrammatical), iii) sentential meaning (semantically plausible vs. semantically implausible relations), and iv) sentence prosody (sentence-like vs. list prosody), as shown in Figure 4.1

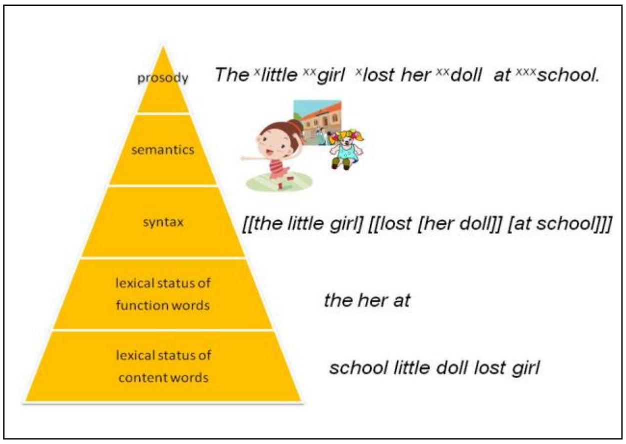


Figure 4.1 Linguistic domains present on a sentence level and manipulated in the present study

Manipulation of these linguistic properties resulted in 9 different conditions, ranging from collections of nonwords to well-formed sentences. The same conditions were applied in

English and Czech, two typologically different languages. The nine conditions (A-I) in English are exemplified by the examples for span 5 (stimuli consisting of five items), which were presented in the introduction and reproduced for convenience in Table 4.1 below.

Label	Description	Example
A	Well-formed sentence:	He sent us a letter.
B	Well-formed sentence with list prosody:	he, sent, us, a, letter
C	Semantically implausible sentence:	He sang us a kettle.
D	Ungrammatical sentence:	A sent he letter us.
E	Ungrammatical sentence with list prosody:	a, sent, he, letter, us
F	Pseudosentence with content words replaced by nonwords:	He /fɪnt/ us a /lɔpə/.
I	Pseudosentence with function words replaced by nonwords:	/vi/ sent /əʃ ʊ/ letter.
G	Pseudosentence with words replaced by nonwords:	/vi fɪnt əʃ ʊ lɔpə/
H	List of nonwords:	/vi/ /fɪnt/ /ɒʃ/ /u/ /lɔpə/

Table 4.1 Nine experimental conditions in English, span 5

Table 4.2 shows how linguistic information was progressively removed across conditions. Conditions A-E are formed with lexical items, thus designated as 'lexical'; conditions F and I are labelled as 'hybrid' as they involve a combination of lexical items and nonwords; and conditions G and H are 'non-lexical' conditions consisting of nonwords. ✓ indicates well-formedness at the relevant level, ✗ indicates violation at the relevant level, – indicates that the relevant level is not applicable (for items in hybrid and non-lexical conditions that have no meaning or function assigned to them).

CONDITIONS	**Lexical**					**Hybrid**		**Non-lexical**	
	A	B	C	D	E	F	I	G	H
Prosody	✓	✗	✓	✓	✗	✓	✓	✓	✗
Syntax	✓	✓	✓	✗	✗	✓	-	-	-
Semantic plausibility	✓	✓	✗	✗	✗	-	-	-	-

Table 4.2 Overview of the experimental conditions

The typological differences of English and Czech were reflected in the stimuli. English has a more fixed word order and relies more on function words, while Czech has a richer morphology and freer word order. In Czech, the relations between phrases in a sentence are mostly expressed by bound morphemes rather than word order. As a result, the number of bound morphemes in Czech sentences is high. English was the starting-point for this study, but the aim was to ensure that the Czech stimuli matched the English versions in certain key respects. Hence, the Czech and English stimuli were matched for the number of words, and the number of content words and function words was also equal. Bound morphemes (inflections) were not manipulated, however, in either English or Czech which led to the Czech stimuli having a greater number of bound morphemes reflecting this inherent tendency of the language. More detail about the relevant differences between the English and Czech stimuli is presented in the sections on the creation of English and Czech stimuli (see sections 4.2 and 4.3).

4.1.1 Task

A 'span' task was used to determine the effects of the target linguistic factors on short-term storage. Span tasks are the standard method for measuring short-term memory capacity for lists of items, with digits being the most popular item used. The task establishes an individual's span, which is the number of the longest list of items an individual can repeat in the correct order immediately after presentation. The current study extends the use of the span task beyond a single-item level: rather than measuring span for sequences of unrelated items, it uses the span technique for sequences of items that are related to varying degrees, as illustrated by the examples in the conditions above (see Table 4.1). Some of these meet the criteria for 'being called' a sentence, while others do not. Where the term 'sentence' does not apply, but the stimuli are still related to each other in prosody, semantics and/or syntax, the term 'pseudosentence' is adopted. The term pseudo/sentences is used as an abbreviation for 'sentences + pseudosentences'. An extensive literature search uncovered no study which had used a span technique with pseudo/sentences.

Although the use of a span task with related items is novel, the span procedure itself is well established. The present experiment followed the procedure used in the Working Memory Test Battery for Children (WMTB-C; Pickering & Gathercole, 2001) which assesses working memory in 5 to 15 year olds. It is a standardised battery designed to test Baddeley's multicomponent working memory model (see section 2.3.2) which measures the central executive, visuospatial sketchpad and phonological loop, but not the episodic buffer. The subtests assessing the phonological loop are digit recall, word list recall, nonword list recall and word list matching. In the digit recall task, participants are asked to recall successively longer series of digits. After correctly repeating four out of six trials per span, participants proceed to a greater span length, up to a maximum span of nine items. This procedure is identical for word list recall, with monosyllabic words used for the stimuli instead of digits and a maximum span of seven. Finally, the stimuli for the nonword list recall task are monosyllabic nonwords and the maximum span is six.

The WMTB-C employs three rules: 'a move on rule', 'a discontinue rule' and 'a reverse rule'. The move on rule states the tester is to skip trials of the same length as soon as a child has repeated four of the trials correctly. The discontinue rule means the tester stops when a child makes errors in three or more stimuli in a particular block. The reverse rule requires that the tester administers shorter stimuli when a child has failed a block (if testing began at any point other than the first trial of the test).

While the current study used a span task, the nature and number of stimuli differed from those in standard memory span tasks which lead to certain adjustments. As noted above, the stimuli in the WMTB-C consist of digits, words and nonwords and they are presented as a list with no variation in structure, i.e. one digit is always followed by another digit or one word is followed by another word. Consequently, if four digit lists have been repeated correctly, the administration of a fifth digit list is unlikely to yield new information. Provided the stimuli are of the same structure, as is the case in list tasks, four correct examples are considered enough to 'pass' a span and the move on rule can be applied to speed up the process of determining span. However, in the current study, pseudo/sentences of the same length varied in linguistic structure so the move on rule was not applied. Consider, for example, (1) – (4) below. These are all stimuli for one condition at one span (well-formed sentences at span 4), but which differ in linguistic structure. They also differ in the number of function and content words, with one function word in (1), but two function words in (2) and (3) and three function words in (4). As a result, each stimulus has distinct characteristics.

(1) *Apples grow on trees.*	(N V PP)
(2) *The fairy was crying.*	(NP aux V)
(3) *I hurt my knee.*	(Pro V NP)
(4) *Give it to me.*	(V pro PP)

The linguistic structure of the stimuli varied for the following reasons. First, unlike the WMTB-C, the current task was not exclusively a quantitative measure. The experiment used a variety of syntactic structures which were intended to allow for generalisations which could be extended to the syntactic domain. If only one structure was used, e.g. NP-V-NP, and an advantage was found for grammatical rather than ungrammatical sentences, this might be due to a property of the structure rather than the syntax. On the other hand, the same result across different structures would suggest that the advantage is due to preservation of well-formedness in a specific linguistic domain and not just a specific structure. Second, using a single structure of the same type would trigger a priming effect. Again, this would not allow for generalisations about each linguistic domain (for more details about the priming effect see section 4.5).

Third, the balance of function and content words was critical for a comparison of the hybrid conditions which investigated the effect of changing the lexical status of content and function words. In condition F, all function words were real words, while all of the content words were nonwords; condition I reversed this by using real content words, but with all of the function words replaced with nonwords. Spans with an odd number of words would not allow for the use of identical structures. In the case of 5-word stimuli, for example, Condition F would have 3 real function words (He /fɪnt/ us a /lɔpə/), but condition I would have 2 real content words (/vi/ sent /əʃ ʊ/ letter.). The problem arises as an equal number of content/function words would no longer be possible. By using different structures, even the spans 3, 5, 7 and 9 could be included and the balance of content and function words across the condition as a whole could be maintained. Different structures were then chosen to ensure that the number of CWs/FWs remained equal within the spans. In addition, it was necessary to maintain the balance of content and function words in order to avoid a confound between the number of nonwords and the content/function status of nonwords. If, for instance, the span for condition F was five and the number of words vs. nonwords was unequal, it would not be possible to decide whether this span was achieved due to the number of familiar words or the number of items. On the other hand, if the number of function and content words were equal, it could be assumed that it was not the number of familiar words that distinguished performance on condition F from I, but rather the familiarity (i.e. lexical status) of the function or content words.

Fourth, the aim was to create a test for children that would maintain their attention and cooperation. Repeatedly presenting the same structures may have led to the children losing interest and concentration. It should be noted that while structures varied across stimuli, the way that different conditions were created ensured that the structures were consistent across conditions (unless violation of structure was the crucial factor differentiating a condition). Examples are shown in Table 4.3 below:

Span 5	A	C	D	F
5.1	The teacher read a story.	The driver bit a tower.	Story a read the teacher.	The /mitʃə/ /kɛd/ a /spɔlɪ/.
5.2	I have seen an angel.	I have read an uncle.	Seen I an have angel.	I have /gin/ an /ɛmbəl/.
5.3	The red bus was late.	The red grass was brave.	Bus the was late red.	The /lup/ /tʌs/ was /dut/.
5.4	He sent us a letter.	He sang us a kettle.	A sent he letter us.	He /fɪnt/ us a /lɔpə/.

Table 4.3 Syntactic structures across conditions: The four sentences in span 5 (5.1, 5.2, 5.3, 5.4) differ in structure, but the structure of 5.1 in condition A was identical to the structure of 5.1 in conditions C and F. Structure of 5.1 in condition D was not identical in structure to 5.1 in A, C and F because the structure was the manipulated variable in condition D.

The move on rule was not employed because structures varied and generalisations about domains, rather than just the specific structures, were sought.

The WMTB-C format was followed in the current study, but differed in a number of conditions and adjustments were made to the testing rules as a result. In the recall tasks of the WMTB-C, only three kinds of stimuli are presented: words, nonwords and digits, which together yield three conditions. In contrast, the current experiment included stimuli in nine different conditions. Due to the number of conditions, it was decided to reduce the number of trials for each span to avoid overloading participants so each span in only contained four trials, compared to six trials in the WMTB-C and the rule for passing a block and the discontinue and reverse rules were adjusted accordingly. To pass the span, three out of four trials had to be repeated correctly, and testing stopped or moved backwards if the participant failed to repeat three out of four stimuli within a block (as opposed to four out of six in the WMTB-C).

4.1.2 Non-experimental factors

When constructing the stimuli, efforts were made to control the factors which are known to influence recall but which were not the target of the experiment (see section 2.2.3). The selected lexical items were high frequency, imageable, familiar and age-appropriate for the target age group of four- and five-year-old children (see section 4.2.1.1). For English, information about familiarity, imageability and age of acquisition of lexical items was obtained from the MRC Psycholinguistic Database (Coltheart, 1981). In order to minimise potential difficulties in speech output, the intention was to keep the stimuli phonologically simple. To this end, the words/nonwords in English contained no more than two adjacent consonants within a single syllable, were no more than disyllabic and carried trochaic stress. All nonwords respected the phonotactics of English. For Czech, no database of lexical items with information on familiarity, imageability and age of acquisition was available so the choice of items was discussed with Czech native speakers who were linguists and nursery teachers. Words/nonwords contained no more than three adjacent consonants within a single syllable, were trisyllabic or less and carried trochaic or dactylic stress. The differences between the English and Czech stimuli reflected the different lexical characteristics of the two languages (see section 4.3.1).

4.1.3 Relation between repetition and receptive vocabulary

Research questions were concerned with the effects of well-formedness and the relationship between the availability of linguistic information and memory performance. A pseudo/sentence span task was designed to address these questions, with conditions carefully targeting performance on well-formed vs. ill-formed pseudo/sentences and manipulating stimuli in linguistic domains such as syntax, semantics, lexicon and prosody. A receptive vocabulary test (BPVS-II, Dunn et al., 1997) was included as a 'warm up' task and to test the relationship between repetition skills, particularly list of words, list of nonwords and sentence repetition, and receptive vocabulary knowledge. This issue was investigated by the eighth research question (see chapter 3). Due to the lack of a standardised vocabulary test in Czech, the BPVS-II was used as a foundation and a Czech adaptation was designed (see section 4.8).

4.1.4 Overview of chapter

The following four sections of this chapter provide greater detail of the experimental method. The first section details the experimental conditions, their linguistic characteristics and the procedures for the recording of stimuli and the editing processes. The second section provides more information about the span procedure, including the presentation format and scoring. The third section describes the assessment of receptive vocabulary in English and Czech. The final section describes modifications to the original design of the task which were made following pilot studies.

4.2 Experimental conditions: Creation of materials

In this section, the creation of the English language conditions is discussed in more detail. Condition A was created as a foundation for the other conditions and therefore it will be presented first. Conditions B-I were derived from condition A.

4.2.1 Foundation for all conditions: Condition A

The motivation behind the choice of the lexical content, function word content, structures and prosody of the sentences is detailed below. Manipulation of each of these components gave rise to the different experimental conditions. More specifically, the lexical status of content and/or function words was manipulated; prosodic properties of function words and phrases were manipulated; and word order within and between phrases was manipulated. The following sections describe the selection process for each of these stimuli.

4.2.1.1 Lexical content

Previous research has identified a number of lexical variables that affect recall performance. Stadthagen-Gonzalez and Davis (2006) pointed out that some variables, such as word length, are intrinsic to each word and can be determined directly from the surface structure of the word without reference to anything else. Other variables, such as word frequency, are determined from the relationship between the target item and a larger corpus of words. Aspects such as age of acquisition, familiarity and imageability are also known to affect performance in tasks such as word recall, word recognition, or word naming. Age of acquisition (AoA) refers to the age at which a word was learnt; imageability measures how easy it is for a word to arouse mental images; and familiarity ratings have often been interpreted as a measure of frequency of exposure to a word (Coltheart, 1981). These properties are usually established by asking people to rate words on a scale. Gernsbacher (1984) suggested that familiarity is a better predictor of word performance than word frequency. A recent study by Stadthagen-Gonzalez and Davis (2006) examined the relations between written/spoken frequency and familiarity. They found relatively strong correlations between familiarity and both written/spoken frequency, and this was interpreted by the authors as support for the idea that subjective familiarity ratings reflect frequency of exposure. The correlation between familiarity and the British National Corpus (Burnard, 1995) spoken frequency measure was 0.72 ($p<0.001$), while it was lower (0.57, $p<0.001$) for the British National Corpus written frequency measure[1]. As a result, words were controlled for familiarity rather than for word frequency as it can be assumed that children are exposed more to spoken language and therefore familiarity ratings were found to better reflect spoken frequency. Information about age of acquisition, familiarity and imageability was obtained from the MRC Psycholinguistic database.

The eighty-eight content words included in the present study were all monosyllabic or disyllabic, had a trochaic stress pattern, and had no more than two adjacent consonants

[1] Stadthagen-Gonzalez and Davis also showed that the possibility that familiarity is tapping into spoken rather than written frequency is supported by their multiple regression analyses in which familiarity is the criterion variable. They reported that the British National Corpus written frequency explains 33% of the variance in the familiarity norms and that adding British National Corpus spoken frequency to the equation accounts for an additional 20% of variance ($R^2 = .53$). When the order of entry of these variables was reversed, the British National Corpus spoken frequency accounts for 51% of variance in familiarity by itself, and the addition of the British National Corpus written frequency only explains an extra 2% of variance.

in a syllable. Only words which met these criteria were considered. The next step was to obtain information about AoA, imageability and familiarity from the database. These properties were rated on a scale of 1 to 7 for AoA, with 1 representing age 0 to 2 years of acquisition and 7 equalling acquisition at 13 years or more. A score of 1 corresponded to low imageability and 7 indicated high imageability. 1 was assigned to words of the least familiarity that the raters had never heard and 7 to words that they had been seen the most often ('nearly every day'). In all cases, ratings on the 1-7 scale were subsequently multiplied by 100 and rounded to the nearest integer in order to present all the ratings as integers on a scale from 100 to 700. Appendix E presents full information about familiarity and imageability and, where available, the AoA of content words used in the conditions.

4.2.1.2 Function words

A variety of function words were included in the stimuli: personal pronouns, possessive pronouns, prepositions, determiners and auxiliaries. Unlike the content words, function words could occur more than once in a set of sentences. For the full list of function words used in the stimuli, see Appendix D.

4.2.1.3 Balance of content words and function words

The equal number of function and content words was critical for the comparisons of certain conditions. Psycholinguistic research suggests that function words and content words are processed in different ways and the better performance with CWs than FWs found with repetition tasks taken by children with language impairment (e.g. Seeff-Gabriel et al., 2010) was an additional focus of the current research. The number of content and function words were equalised for each span. As illustrated by examples in the introduction to this chapter and repeated again below, the number of content/function words for each sentence may differ, but the total number for the whole block (i.e. over four sentences of identical length) is always equal:

SPAN 4		SPAN 7	
Apples grow on trees.	1 FW/3 CW	*The snake is hiding in the grass.*	4 FW/3 CW
The fairy was crying.	2 FW/2 CW	*The clown did tricks with a monkey.*	3 FW/4 CW
I hurt my knee.	2 FW/2 CW	*My friend will buy the new book.*	3 FW/4 CW
Give it to me.	3 FW/1 CW	*It has been snowing for one week.*	4 FW/3 CW
Total	8 FW/8 CW	Total	14 FW/14 CW

4.2.1.4 Structure

The main aim of the study was to determine the contribution of different linguistic domains (see chapter 3) rather than different syntactic structures, so simple structures were used (see below) which were consistent across the conditions (apart from the cases in condition D where violation of structure was the variable manipulated). Even though clausal and phrasal structures varied within spans and across spans (e.g. verbs took different numbers of arguments, the number of prepositional phrases ranged from zero to two, noun phrases could be pronominal or lexical), the foundation sentences were always a simple, declarative main clause. Negatives, passives, quantifiers, comparatives, questions and subordinate clauses were avoided since these structures could increase complexity and could be more difficult to compare across the experimental conditions. For example, a passive construction would introduce a further linguistic factor (movement) and this would lead to two undesirable consequences in the design of the current study.

The first undesirable outcome relates to the comparison between conditions, and the second to performance within a condition. If a passive sentence in condition A is compared to an ungrammatical sentence in condition D, the factor of movement introduced into condition A would not be in D, since the passive construction cannot even be recognised as being a passive construction in an ungrammatical sentence, where, once the word order is disrupted, the remaining list of words would be no different from a list derived from an active sentence. Thus, comparing a span for grammatical sentences with a passive construction and a span for ungrammatical sentences which had originally contained a passive construction would introduce an extra variable in condition A. This was not desirable as the aim was to manipulate just one variable at a time and in the case of the comparison between conditions A and D, the phrasal structure was the focus.

The second reason to avoid movement is the greater difficulty of processing and producing structures with movement (Gibson, 1998). Therefore, performance on structures with/without movement should not be compared (unless this is a factor to be investigated). The only exception to limiting sentences to a simple, declarative main clause was the use of imperative constructions in the short pseudo/sentences. Imperatives were included because of their lack of movement and to help provide an alternative possible sentence structure.

4.2.1.5 Prosodic properties

In English, stress refers to the relative prominence of a lexical item within its domain (Ewen & van der Hulst, 2001) and is primarily realised through changes in three phonetic parameters: the fundamental frequency, duration, and amplitude (Lieberman, 1960). The domains at which stress is assigned are i) lexical item, ii) phrase, and iii) sentence (Ewen & van der Hulst, 2001). The sentences in condition A were produced as prosodically neutral (i.e. no emphatic stress), with stress assigned according to lexical and phrasal structure, as illustrated by the following example:

A: The	little	girl	lost	her	doll	at	school.	
	x	x	x		x		x	lexical stress
		x			x		x	phrasal stress

Just as lexical words contain one word stress, on the next level up a phrase contains one phrasal stress, i.e. within a phrase, one word will be more prominent than others. But note that '*one particular syllable can be accented with reference to several inclusive domains*' (Ewen & van der Hulst, 2001, p.200). This is illustrated by the example above where 'girl' receives lexical and phrasal stress and these accumulate, while 'little' only receives lexical but not phrasal stress. Stimuli in condition A are presented in Table 4.4.

Block	Label	Stimulus
2	2A1	She cooks.
	2A2	They played.
	2A3	Help him.
	2A4	Boys do.
3	3A1	Ice is cold.
	3A2	Dance with us.
	3A3	We like him.
	3A4	Watch the race.
4	4A1	Apples grow on trees.
	4A2	The fairy was crying.
	4A3	I hurt my knee.
	4A4	Give it to me.
5	5A1	The teacher read a story.
	5A2	I have seen an angel.
	5A3	The red bus was late.
	5A4	She sent us a letter.
6	6A1	John eats rice with his meat.
	6A2	The picture was full of colours.
	6A3	He threw a stone at me.
	6A4	Come out of the dirty water.
7	7A1	The snake is hiding in the grass.
	7A2	The clown did tricks with a monkey.
	7A3	My friend will buy the new book.
	7A4	It has been snowing for one week.
8	8A1	The little girl lost her doll at school.
	8A2	They have waited there for a long time.
	8A3	The train only stopped in the thick fog.
	8A4	I met my old aunt at her farm.
9	9A1	The white cat was chasing him in the park.
	9A2	We can see the stars on a clear night.
	9A3	The young dancer was looking at the shiny mirror.
	9A4	I have been blowing pretty bubbles the whole day.

Table 4.4 Stimuli in condition A

4.2.2 Semantically implausible targets: Condition C

4.2.2.1 Lexical content

Condition C was based on condition A. Real lexical items replaced the content words from condition A in order to create an implausible sentence where the implausibility arises not from the meaning of individual words but from their combination. The criteria for the lexical items which were used were the same as in condition A: the words selected were rated high in familiarity and imageability and were also age-appropriate (information obtained from MRC database). In addition, the words were mono- or disyllabic and did not contain more than two adjacent consonants within a syllable. Content words in semantically implausible sentences (condition C) were matched for number of syllables in the corresponding lexical items found in the corresponding semantically plausible sentence (condition A), e.g.:

A:	The	little	girl	lost	her	doll	at	school.
C:	The	middle	sock	brushed	its	eye	at	home.

A:	The	picture	was	full	of	colours.
C:	The	flower	was	sick	of	ladders.

Efforts were made to match the syllabic structure of the lexical items in the semantically implausible sentences to the syllabic structures of lexical items in semantically plausible counterparts, e.g.:

little	middle	colours	ladders
CV.CC	CV.CC	CV.CVC	CV.CVC

However, this matching could not be applied in all cases since other criteria such as familiarity or imageability took priority. Vocabulary at the age of four is still limited and priority was given to ensure that the words should be familiar, real words for children. It was therefore not possible to satisfy all the criteria and the syllable structure was considered only after the other criteria were met. This is illustrated by the example below.

doll	eye
CVC	V

In order to understand how the content words were chosen in creating implausible sentence meanings, the notion of selectional restrictions will be discussed. In his section on 'Degrees of grammaticalness', Chomsky (1965) discussed sentences such as '*colorless green ideas sleep furiously*', which break selectional rules. The notion of selectional restrictions is used to characterise what arises from the combination of words which are not compatible with each other. For instance, words can carry features such as [+animate], [+human], [+abstract] and if words with incompatible features are put together, the violation of selectional rules results. Chierchia and McConnell-Ginet (2000) illustrate the point with the verb *drink*: the verb carries the information that its object designates something drinkable, something concrete rather than abstract and its subject designates an animate being. Thus, 'drink' should only select arguments satisfying such restrictions. Violations of selectional restrictions would arise from mismatches between the features, giving rise to an implausible sentence. This is not to say that the sentence is uninterpretable. Chomsky used examples such as '*the book who you read was a best seller*' and '*a very walking person appeared*' to illustrate that sentences such as these '*are uniquely, uniformly, and immediately interpretable, no doubt, although they are paradigm examples of departure from well-formedness'* (1965, p.150-151).

With the current stimuli, violations most often involved animacy and concreteness restrictions on verb arguments, e.g. *sock brushed its eye, flower was sick, plane swam, she snows, wash the voice, he sang us a kettle.* Another type of violation involved the attribute expressed by an adjective being incompatible with the properties of the noun it modified, such as *sweet salt* or *sunny hammer*. Many anomalous combinations could be plausible in specific contexts, *e.g. red grass, spotty needles,* while some were infrequent in the real world but compatible e.g. *pink light, wet pencil.* Such sentences involve some kind of semantic incompatibility: logically incompatible, i.e. incompatible with the principles of the real world, or at least describing situations that are unlikely. All are potentially interpretable in the sense that the syntax is intact and the interpretations range from unusual to fantastic. For instance, we can easily understand *The snail is laughing in the sky,* but it does not describe an experience from the real world. While the nature of implausibility varied, the stimuli in condition C consistently differ from the semantically plausible sentences. The sentences were novel and one can presume that the children had never encountered the particular word combination before. Although the frequency of co-occurrence of constituent items cannot be calculated, it can be assumed to be low relative to that of plausible sentences.

The content words in the semantically implausible sentences differed from the content items used in the semantically plausible sentences, raising the question of whether the content words in both conditions were matched for lexical characteristics. One way of avoiding possible confounds arising from the introduction of new lexical items would be to swap the same lexical items between conditions A and C. However, if words were swapped, maintaining an equal number of syllables may not have been possible. Secondly, the aim was to violate selectional restrictions in order to create semantic implausibility without altering the sentence structure. The option of recycling words in semantically plausible and semantically implausible sentences has been used by others (Marks & Miller, 1964), but only where all of the stimuli had the same structure. This is not the case in the current experiment and therefore familiar and imageable lexical items which differed from those in condition A were chosen. In total, eighty-eight content words were replaced to create thirty-two semantically implausible sentences.

4.2.2.2 Function words

The function words in condition C were not manipulated. The same pool of function words was used in condition C as in condition A.

4.2.2.3 Structure

As only the content words were replaced, the original syntactic structure of condition C was retained. The function words remained in their original positions within the sentences.

4.2.2.4 Prosodic properties

The prosodic structures were copied from the original semantically plausible sentences, maintaining the items' lexical status and phrasal structures, e.g.:

The	middle	sock	brushed	its	eye	at	home	
	x	x	x		x		x	lexical stress
		x			x		x	phrasal stress

When the sentences were recorded, the original sentences were written on the same page as the semantically implausible sentences for the speaker. This provided a template, making it easier to match the stimuli on sentence prosody when spoken for the recording.

To sum up, condition C was created by substituting semantically incompatible lexical items for the original items in semantically plausible sentences. As a result, the syntactic and prosodic properties of the sentences were kept constant, while the sentence meaning was manipulated (from plausible to implausible). Hence, the individual words and the syntactic constructions are semantically viable, but the semantic relations they encode are less so. The stimuli in condition C are presented in Table 4.5.

Block	Label	Stimulus
2	2C1	She snows.
	2C2	They rained.
	2C3	Mix him.
	2C4	Beans do.
3	3C1	Ice is sad.
	3C2	Rain with us.
	3C3	We bake him.
	3C4	Wash the voice.
4	4C1	Tables grow on cars.
	4C2	The cherry was cooking.
	4C3	I dug my tea.
	4C4	Break it to me.
5	5C1	The driver bit a tower.
	5C2	I have read an uncle.
	5C3	The red grass was brave.
	5C4	He sang us a kettle.
6	6C1	John feels noise with his seat.
	6C2	The flower was sick of ladders.
	6C3	He flew a ring at me.
	6C4	Walk out of the fluffy candle.
7	7C1	The snail is laughing in the sky.
	7C2	The bike got bread with a donkey.
	7C3	My frog will toast the short book.
	7C4	It has been sneezing for all time.
8	8C1	The middle sock brushed its eye at home.
	8C2	They have hated there for a sweet salt.
	8C3	The plane only swam in the thick road.
	8C4	I wrote my blue child at her room.
9	9C1	The green pig was washing him in the fork.
	9C2	We can hear the skies on a pink light.
	9C3	The wet pencil was resting at the sunny hammer.
	9C4	I have been fighting spotty needles the whole boy.

Table 4.5 Stimuli in condition C

4.2.3 Ungrammatical targets: Condition D

English largely relies on word order to mark grammatical relations; therefore, violations to word order result in ungrammatical sentences. In this condition, ungrammatical sentences were created by altering the order of words in condition A.

4.2.3.1 Structure

Grammatical relations were disrupted between verbs/prepositions and their noun phrase complements, and between determiner/adjective specifiers and the nouns they specify. The advantage of this was that no phonological material was altered or lost and both the grammatical and ungrammatical sentences contained the same lexical items and affixes.

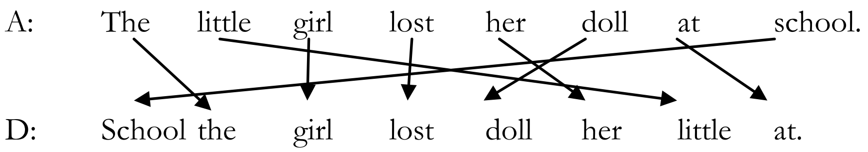
Another way to create an ungrammatical sentence in English would be to violate agreement between syntactic constituents. Given that English does not have a rich morphology, agreement is largely limited to subject-verb. This means that there would be only one possible violation in most sentences. Furthermore, a primary reason for controlling the non-experimental characteristics of the stimuli was to maintain a constant number of syllables and phonemes across the conditions. Adding or deleting plural *-s* or 3.SG. *-s*, as required for the violation of subject-verb agreement, would change the number of phonemes and in some cases even syllables.

The final option was to violate both word order and agreement, which came with the same issues that accompanied the first and second options. Furthermore, it is possible that once grammatical relations in the sentence are lost to the point where the words become little more than a list, there will be no further effect from any additional manipulation of grammaticality.

Based on these considerations, only word order violations were investigated in this experiment. It may be worthwhile for future research to investigate the other forms of ungrammaticality which were not covered.

4.2.3.2 Lexical content and function words

Since only word order was manipulated in this condition, both lexical items and function words could be matched perfectly with the lexical items and function words from semantically plausible sentences (condition A).



4.2.3.3 Prosodic properties

Like condition A, the ungrammatical sentences in condition D were produced with lexical, phrasal and sentence stress. However, due to changes in word order, stress could not be assigned in accordance with the phrasal structure. Templates corresponding to the experimental sentences were created in order to elicit an appropriate sentence prosody with stressed content words, unstressed function words, and copied phrasal stress. The template was a grammatical sentence with the same sequence of content and function words that was found in the ungrammatical sentences used for this condition. For example, the stimulus 'In snake the grass is the hiding' was matched with a template sentence 'I got the room with the mirror'. The number of syllables in content/function words was matched and the function and content words in the template sentence occupied the same positions as in the stimulus sentence. This was all done so that when the stimuli were recorded, the speaker could apply the same prosodic patterns to the experimental sentences in condition D[2]: To sum up, the lexical items and their phonological properties matched condition A, but the order of the items was different which in turn altered the prosody. Stimuli in condition D are presented in Table 4.6.

Block	Label	Stimulus
2	2D1	Cooks she.
	2D2	Played they.
	2D3	Him help.
	2D4	Do boys.
3	3D1	Is cold ice.
	3D2	Us with dance.
	3D3	Like we him.
	3D4	Race the watch.
4	4D1	Trees on grow apples.
	4D2	The crying fairy was.
	4D3	Hurt my I knee.
	4D4	Me to it give.
5	5D1	Story a read the teacher.
	5D2	Seen I an have angel.
	5D3	Bus the was late red.
	5D4	A sent he letter us.
6	6D1	Rice with eats his John meat.
	6D2	The full was colours of picture.
	6D3	A threw me at stone he.
	6D4	Of dirty out water the come.
7	7D1	In snake the grass is the hiding.
	7D2	A tricks the with clown did monkey.
	7D3	Friend will new the buy book my.
	7D4	Been has one it week for snowing.
8	8D1	School the girl lost doll her little at.
	8D2	For time have they long waited a there.
	8D3	Fog the stopped the in thick train only.
	8D4	I old her met farm my at aunt.
9	9D1	Park in chasing the him cat the was white.
	9D2	Night see can the a clear on stars we.
	9D3	Dancer the was shiny young the at mirror looking.
	9D4	Bubbles the pretty have day whole been I blowing.

Table 4.6 Stimuli in condition D

[2] Other options were available for recording the sentence-like prosodic contour on a grammatically ill-formed sentence as required for condition D. For example, the pseudosentence could have been resynthesised by pasting the prosodic contour from a grammatical sentence onto it. An easier, 'natural' option can be supported by giving the speaker the opportunity to practice and acquire the appropriate contour from a normal sentence and this could then be exported to the grammatically ill-formed sentence.

4.2.4 Hybrid targets

The hybrid conditions were created by replacing items in either the content word or function word slots with nonwords. The syntactic structure and prosodic properties of the original items were retained.

4.2.4.1 Condition F

4.2.4.1.1 Lexical content

The stimuli in condition F were created by replacing the lexical items in the content word slots with nonwords. The nonwords were phonotactically possible sequences which matched the number of syllables in the original content words from condition A, e.g.:

A:	The	little	girl	lost	her	doll	at	school
F:	The	/rɪkəl/	/bɜl/	/kɛst/	her	/mɒl/	at	/brul/

The nonwords' syllabic structure reflected the syllabic structure of the original lexical items:

angel	dance	snake
/ɛmbəl/	/bɔns/	/skaɪs/
VC.CC	CVCC	CCVC

4.2.4.1.2 Function words

The function words from condition A were preserved in this condition and were not manipulated.

4.2.4.1.3 Structure

Since the lexical status of content words was the only factor which was manipulated, the structure did not change and was matched to the structure of the sentences in condition A.

4.2.4.1.4 Prosodic properties

The stimuli were matched for the prosody of the original sentences, with function words unstressed and pseudocontent words stressed at word and phrasal levels, as were their real word counterparts. When recording the sentences, the original sentences from condition A were written on the same page as the hybrid sentences. This offered a template and made it possible to match the sentence prosody stimuli. The stimuli in condition F are presented in Table 4.7.

Block	Label	Stimulus
2	2F1	She /tups/.
	2F2	They /blɔɪd/.
	2F3	/nɪlt/ him.
	2F4	/fɔɪz/ do.
3	3F1	/ɔɪs/ is /pɛld/.
	3F2	/bɔns/ with us.
	3F3	We /rɔɪk/ him.
	3F4	/jʌtʃ/ the /rəʊk/.
4	4F1	/ɛməlz/ /drɑʊ/ on /prɔz/.
	4F2	The /sɛlɪ/ was /blaɪɪŋ/.
	4F3	I /vɜt/ my /ri/.
	4F4	/dʊf/ it to me.
5	5F1	The /mitʃə/ /kɛd/ a /spɔlɪ/.
	5F2	I have /gin/ an /ɛmbəl/.
	5F3	The /lup/ /tʌs/ was /dut/.
	5F4	She /fɪnt/ us a /lɔpə/.
6	6F1	/tʃɒm/ /uts/ /baɪs/ with his /lit/.
	6F2	The /tʊgtʃə/ was /tɪl/ of /mʌləz/.
	6F3	He /frɔ/ a /pləʊn/ at me.
	6F4	/lʊm/ out of the /bɑdɪ/ /wɛpə/.
7	7F1	The /skaɪs/ is /maɪdɪŋ/ in the /krɑf/.
	7F2	The /spaʊn/ /pɪd/ /prɔks/ with a /rɒŋkɪ/.
	7F3	My /krɛnd/ will /sɔɪ/ the /tju/ /mʊk/.
	7F4	It has been /plɔɪɪŋ/ for /pʌm/ /rik/.
8	8F1	The /rɪkəl/ /bɜl/ /kɛst/ her /mɒl/ at /brul/.
	8F2	They have /lɔtɪd/ there for a /kɔŋ/ /tɔɪn/.
	8F3	The /trun/ only /blopt/ in the /fɪp/ /sʊg/.
	8F4	I /dɛt/ my /ild/ /ɔnt/ at her /kɑn/.
9	9F1	The /paɪt/ /hʊt/ was /dʒaɪsɪŋ/ him in the /mɔk/.
	9F2	We can /gɑ/ the /spiz/ on a /blɪə/ /bɔɪt/.
	9F3	The /jɛm/ /bɔnsə/ was /mʊkɪŋ/ at the /fɔɪnɪ/ /mʌrə/.
	9F4	I have been /spəʊɪŋ/ /brɪdɪ/ /dʌdəlz/ the /kɑʊl/ /tʃeɪ/.

Table 4.7 Stimuli in condition F

4.2.4.2 Condition I

4.2.4.2.1 Lexical content

The content words from condition A were preserved and were not manipulated.

4.2.4.2.2 Function words

The stimuli in condition F were created by replacing the items in function word slots with nonwords. The nonwords were phonotactically possible sequences and matched for number of syllables with the original function words from condition A:

A:	The	little	girl	lost	her	doll	at	school
I:	/tʃə/	little	girl	lost	/ʃə/	doll	/əb/	school.

The syllabic structure of the nonwords copied the syllabic structure of the original lexical items, as in examples below:

was	the	I
/rɪs/	/tʃə/	/ɔɪ/
CVC	CV	V

4.2.4.2.3 Structure

Since the lexical status of the function words was the only factor which was manipulated, the syntactic structure did not change and was matched to the structure of sentences from condition A. It should be noted that structure is not recognisable without recognisable function words.

4.2.4.2.4 Prosodic properties

The stimuli were matched with the prosody of the original sentences, with pseudofunction words unstressed and content words stressed. Again, templates in the form of sentences from condition A were available when recording the stimuli. The stimuli in condition I are presented in Table 4.8.

Block	Label	Stimulus
2	2I1	/vi/ cooks.
	2I2	/faɪ/ played.
	2I3	Help /lɪm/.
	2I4	Boys /gu/.
3	3I1	Ice /ʊz/ cold.
	3I2	Dance /mɪt əʃ/.
	3I3	/wu/ like /lɪm/.
	3I4	Watch /tʃə/ race.
4	4I1	Apples grow /ɒp/ trees.
	4I2	/tʃə/ fairy /rɪs/ crying.
	4I3	/ɔɪ/ hurt /kaɪ/ knee.
	4I4	Give /ʊt lɪ nu/.
5	5I1	/tʃə/ teacher read /ʊ/ story.
	5I2	/ɔi kəf/ seen /ɪm/ angel.
	5I3	/tʃə/ red bus /rɪs/ late.
	5I4	/vi/ sent /əʃ ʊ/ letter.
6	6I1	John eats rice /mɪt jɪs/ meat.
	6I2	/tʃə/ picture /rɪs/ full /əb/ colors.
	6I3	/fi/ threw /ʊ/ stone /əb nu/.
	6I4	Come /ɔɪt əb tʃə/ dirty water.
7	7I1	/tʃə/ snake /ʊz/ hiding /ɪp tʃə/ grass.
	7I2	/tʃə/ clown did tricks /mɪt ʊ/ monkey.
	7I3	/kɑ/ friend /bəl/ buy /tʃə/ new book.
	7I4	/ʊt kəs gin/ snowing /zə/ one week.
8	8I1	/tʃə/ little girl lost /ʃə/ doll /əb/ school.
	8I2	/faɪ kəf/ waited /sə zə ʊ/ long time.
	8I3	/tʃə/ train /inlɪ/ stopped /ɪp tʃə/ thick fog.
	8I4	/ɔɪ/ met /kaɪ/ old aunt /əb ʃə/ farm.
9	9I1	/tʃə/ white cat /rɪs/ chasing /lɪm ɪp tʃə/ park.
	9I2	/wu lən/ see /tʃə/ stars /ɒp ʊ/ clear night.
	9I3	/tʃə/ young dancer /rɪs/ looking /əb tʃə/ shiny mirror.
	9I4	/ɔɪ kəf gin/ blowing pretty bubbles /tʃə/ whole day.

Table 4.8 Stimuli in condition I

4.2.5 Nonlexical targets

4.2.5.1 Condition G

This condition was nonlexical and consisted only of nonwords.

4.2.5.1.1 Content and function words

Both the content and function words were replaced by nonwords. Nonwords from condition F were used to replace the content words, and nonwords from condition I replaced the function words. This provided the match described in conditions F and I, i.e. nonwords and real words matched for the number of syllables and syllabic structure, e.g:

A:	The	little	girl	lost	her	doll	at	school
I:	/tʃə/	little	girl	lost	/ʃə/	doll	/əb/	school
F:	The	/rɪkəl/	/bɜl/	/kɛst/	her	/mɒl/	at	/brul/
G:	/tʃə	rɪkəl	bɜl	kɛst	ʃɜ	mɒl	əb	brul/

4.2.5.1.2 Structure

The nonwords which replaced both the function and content words were matched to the original sentence in their prosodic structure/syllable structure and number of syllables. As a result, the nonwords then occupied the same positions as the words from the original grammatical sentence, i.e. the nonwords took up a serial order identical to the order in the stimuli with real words. However, as all the information was nonlexical, it is not possible to classify words into syntactic categories and the syntactic structure of the stimuli is therefore irrelevant.

4.2.5.1.3 Prosodic properties

The stimuli were matched with the prosody of the original sentences. Pseudofunction words were unstressed and pseudocontent words were stressed and both carried the same phrasal stress as their real word counterparts. Sentences from condition A were again used as templates. The stimuli from condition G are presented in Table 4.9.

Block	Label	Stimulus
2	2G1	/vi tups/
	2G2	/faɪ blɔɪd/
	2G3	/nɪlt lɪm/
	2G4	/fɔɪz gu/
3	3G1	/ɔɪs ʊz pɛld/
	3G2	/bɔns mɪt əʃ/
	3G3	/wu rɔɪk lɪm/
	3G4	/jʌtʃ tʃə rəʊk/
4	4G1	/ɛməlz drɑʊ ɒp prɔz/
	4G2	/tʃə sɛlɪ rɪs blaɪɪŋ/
	4G3	/ɔɪ vɜt kaɪ ri/
	4G4	/dʊf ʊt lɪ nu/
5	5G1	/tʃə mitʃə kɛd ʊ spɔlɪ/
	5G2	/ɔɪ kəf gin ɪm ɛmbəl/
	5G3	/tʃə lup tʌs rɪs dut/
	5G4	/vi fɪnt əʃ ʊ lɔpə/
6	6G1	/tʃɒm uts baɪs mɪt jɪs lit/
	6G2	/tʃə tʊgtʃə rɪs tɪl əb mʌləz/
	6G3	/fi frɔ ʊ pləʊn əb nu/
	6G4	/lʊm ɔɪt əb tʃə bɑdɪ wɛpə/
7	7G1	/tʃə skaɪs ʊz maɪdɪŋ ɪp tʃə krɑf/
	7G2	/tʃə spaʊn pɪd prɔks mɪt ʊ rɒŋkɪ/
	7G3	/kɑ krɛnd bəl sɔɪ tʃə tju mʊk/
	7G4	/ʊt kəs gin plɔɪɪŋ zə pʌm rik/
8	8G1	/tʃə rɪkəl bɜl kɛst ʃɜ mɒl əb brul/
	8G2	/faɪ kəf lɔtɪd sə zə ʊ kɔŋ tɔɪn/
	8G3	/tʃə trun inlɪ blopt ɪp tʃə fɪp sʊg/
	8G4	/ɔɪ dɛt kaɪ ild ɔnt əb ʃɜ kɑn/
9	9G1	/tʃə paɪt hʊt rɪs dʒaɪzɪŋ lɪm ɪp tʃə mɔk/
	9G2	/wu lən gɑ tʃə spiz ɒp ʊ blɪə bɔɪt/
	9G3	/tʃə jɛm bɔnsə rɪs mʊkɪŋ əb tʃə fɔɪnɪ mʌrə/
	9G4	/ɔɪ kəf gin spəʊɪŋ brɪdɪ dʌdəlz tʃə kɑʊl tʃeɪ/

Table 4.9 Stimuli in condition G

4.2.6 Conditions with a list prosody

The aim of comparing the stimuli with a sentence-like prosody versus stimuli with a list prosody was to evaluate the extent prosody affects memory span. The list prosody could be described as sequences of words with a trochaic stress pattern, produced with monotonous presentation and avoiding any changes to the final item. Of the nine conditions created, six had a sentence-like prosody (described above) and three had a list prosody. Condition B contained grammatical, semantically plausible pseudosentences which differed from condition A only in their prosody. Condition E contained ungrammatical sentences with a sequence of lexical items which were taken from its prosodic counterpart, condition D.

Condition H was comprised of lists of nonwords which were the counterpart to the nonlexical pseudosentences in condition G.

A:	The little girl lost her doll at school.							
B:	The	little	girl	lost	her	doll	at	school
D:	School the girl lost doll her little at.							
E:	School	the	girl	lost	doll	her	little	at.
G:	/tʃə rɪkəl bɜl kɛst ʃə mɒl əb brul/							
H:	/tʃə/	/rɪkəl/	/bɜl/	/kɛst/	/ʃɜ/	/mɒl/	/æb/	/brul/

In summary, the choice of lexical content, function words and structures for the list conditions was identical to their prosodic counterparts – the otherwise identical stimuli with a sentence-like prosody. The following sections detail the creation of the conditions with a list prosody. Condition E is discussed first, since it was used for the creation of the stimuli in condition B, which is discussed second, and this set of experimental conditions is completed by the nonlexical condition H, which consisted of lists of nonwords.

4.2.6.1 Condition E

Condition E copied the lexical items and structure of condition D, with each word in a sentence receiving stress (see the example below) to form a list of words without a predictable order. Prior to recording the lists, the speaker counted aloud to ten in an attempt to gain the idea of the lack of phrasal or sentence stress needed for the list. The speaker was then asked to produce each item in the list with stress. In addition, in order to prevent changes in the final item such as final lengthening or a change in pitch, an extra item was added at the end of each list. This item was later cut using a software program designed for sound editing (for details see section 4.4.1) to achieve a more monotonous presentation and to eliminate any changes which might have occurred in the last item[3].

3 Another possibility would be to record all of the words individually in carrier sentences, then cut the required words out and then paste them together. This option may provide a more accurate acoustic representation but as the experiment was not primarily concerned with the role of acoustic cues in recall, it was decided that the first option would suffice.

D:	School	the	girl	lost	doll	her	little	at	*sentence prosody*
E:	school,	the,	girl,	lost,	doll,	her,	little,	at	*list prosody*

Unlike the other items, the definite determiner 'the' was produced with a schwa rather than a full vowel as the speaker found the unreduced vowel 'unnatural', although the other elements of stress were maintained. The stimuli in condition E are presented in Table 4.10.

Block	Label	Stimulus
2	2E1	cooks, she
	2E2	played, they
	2E3	him, help
	2E4	do, boys
3	3E1	is, cold, ice
	3E2	us, with, dance
	3E3	like, we, him
	3E4	race, the, watch
4	4E1	trees, on, grow, apples
	4E2	the, crying, fairy, was
	4E3	hurt, my, I, knee
	4E4	me, to, it, give
5	5E1	story, a, read, the, teacher
	5E2	seen, I, an, have, angel
	5E3	bus, the, was, late, red
	5E4	a, sent, he, letter, us
6	6E1	rice, with, eats, his, John, meat
	6E2	the, full, was, colours, of, picture
	6E3	a, threw, me, at, stone, he
	6E4	of, dirty, out, water, the, come
7	7E1	in, snake, the, grass, is, the, hiding
	7E2	a, tricks, the, with, clown, did, monkey
	7E3	friend, will, new, the, buy, book, my
	7E4	been, has, one, it, week, for, snowing
8	8E1	school, the, girl, lost, doll, the, little, at
	8E2	for, time, have, they, long, waited, a, there
	8E3	fog, the, stopped, the, in, thick, train, only
	8E4	I, old, her, met, farm, my, at, aunt
9	9E1	park, in, chasing, the, him, cat, the, was, white
	9E2	night, see, can, the, a, clear, on, stars, we
	9E3	dancer, the, was, shiny, young, the, at, mirror, looking
	9E4	bubbles, the, pretty, have, day, whole, been, I, blowing

Table 4.10 Stimuli in condition E

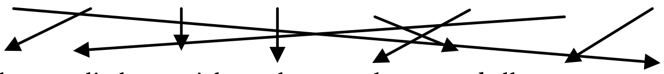
4.2.6.2 Condition B

The stimuli in condition B were lists of the same lexical items and syntactic structures found in condition A and were in a predictable order. They were morphosyntactically well-formed but lacked the corresponding prosody found in the stimuli from condition A.

A:	The	little	girl	lost	her	doll	at	school	*sentence prosody*
B:	the,	little,	girl,	lost,	her,	doll,	at,	school	*list prosody*

The stimuli in condition B consisted of the same lexical items as the stimuli in condition E and could therefore be created by splicing together the items from condition E. Each lexical item was cut and then placed in an appropriate position in a stimulus string. The pauses were minimised, and the last items which were not part of the stimuli, but had been added to prevent the final lengthening, were cut out (in this case, the item 'only').

E:	school,	the,	girl,	lost,	doll,	her,	little,	at,	~~only~~



B:	the,	little,	girl,	lost,	her,	doll,	at,	school

Stimuli in condition B are presented in Table 4.11. Figures 4.2 and 4.3 represent the speech signal for examples from conditions A (natural prosody) and B (list prosody).

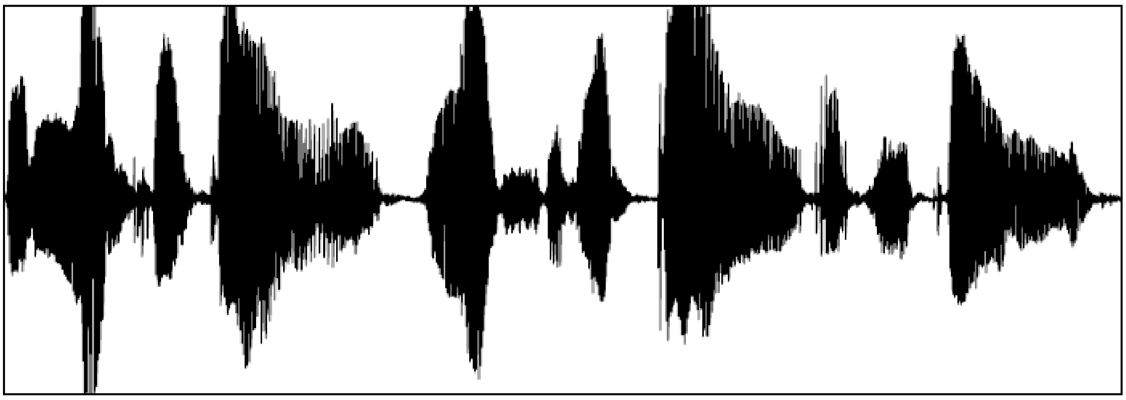
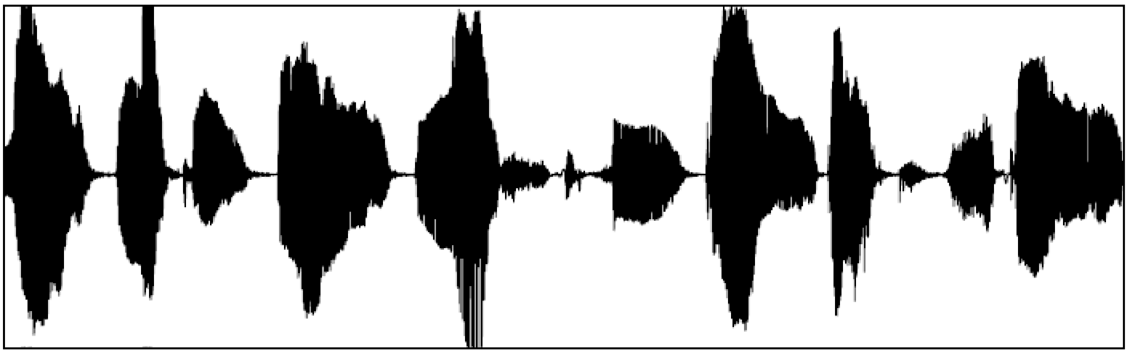


Figure 4.2 Speech signal for sentence with a sentence prosody (condition A)



The little girl lost her doll at school

Figure 4.3 Speech signal for sentence with a list prosody (condition B)

Block	Label	Stimulus
2	2B1	she, cooks
	2B2	they, played
	2B3	help, him
	2B4	boys, do
3	3B1	ice, is, cold
	3B2	dance, with, us
	3B3	we, like, him
	3B4	watch, the, race
4	4B1	apples, grow, on, trees
	4B2	the, fairy, was, crying
	4B3	I, hurt, my, knee
	4B5	give, it, to, me
5	5B1	the, teacher, read, a, story
	5B2	I, have, seen, an, angel
	5B3	the, red, bus, was, late
	5B4	he, sent, us, a, letter
6	6B1	John, eats, rice, with, his, meat
	6B2	the, picture, was, full, of, colors
	6B3	he, threw, a, stone, at, me
	6B4	come, out, of, the, dirty, water
7	7B1	the, snake, is, hiding, in, the, grass
	7B2	the, clown, did, tricks, with, a, monkey
	7B3	my, friend, will, buy, the, new, book
	7B4	it, has, been, snowing, for, one, week
8	8B1	the, little, girl, lost, her, doll, at, school
	8B2	they, have, waited, there, for, a, long, time
	8B3	the, train, only, stopped, in, the, thick, fog
	8B4	I, met, my, old, aunt, at, her, farm
9	9B1	the, white, cat, was, chasing, him, in, the, park
	9B2	we, can, see, the, stars, on, a, clear, night
	9B3	the, young, dancer, was, looking, at, the, shiny, mirror
	9B4	I, have, been, blowing, pretty, bubbles, the, whole, day

Table 4.11 Stimuli in condition B

4.2.6.3 Condition H

The stimuli in condition H were lists of nonwords. This condition was a prosodic counterpart to condition G, nonlexical pseudosentences with sentence-like prosody. As discussed in chapter 2, most function words take a reduced form in a sentential context. While the nonwords in function word slots in condition G were phonologically reduced, the nonwords in function word slots in condition H (nonlexical pseudosentences with a list prosody) were all in a full form.

G:	/tʃə	rɪkəl	bɜl	kɛst	ʃə	mɒl	əb	brul/	
H:	/tʃə/	/rɪkəl/	/bɜl/	/kɛst/	/ʃɜ/	/mɒl/	/æb/	/brul/	

In order to match the lexical and nonlexical conditions, the nonword form for the definite determiner appeared with a schwa rather than a full vowel in condition H, being parallel to conditions E and B.

Condition H was produced as a list of nonwords, with extra nonwords again added to the list for recording and then removed in the final version of the lists.

H: /tʃə/ rɪkəl/ /bɜl/ /kɛst/ /ʃɜ/ /mɒl/ /æb/ /brul/ ~~/kaɪ/~~

The stimuli in condition H are presented in Table 4.12.

Block	Label	Stimulus
2	2H1	/vi/ /tups/
	2H2	/faɪ/ /blɔɪd/
	2H3	/nɪlt/ /lɪm/
	2H4	/fɔɪz/ /gu/
3	3H1	/ɔɪs/ /uz/ /pɛld/
	3H2	/bɔns/ /mɪt/ /ɒʃ/
	3H3	/wu/ /rɔɪk/ /lɪm/
	3H4	/jʌtʃ/ /tʃə/ /rəʊk/
4	4H1	/ɛməlz/ /drɑʊ/ /ɒp/ /prɔz/
	4H2	/tʃə/ /sɛlɪ/ /rɪs/ /blaɪɪŋ/
	4H3	/ɔɪ/ /vɜt/ /kaɪ/ /ri/
	4H4	/dʊf/ /ʊt/ /lɪ/ /nu/
5	5H1	/tʃə/ /mitʃə/ /kɛd/ /u/ /spɔlɪ/
	5H2	/ɔɪ/ /kæf/ /gin/ /ɪm/ /ɛmbəl/
	5H3	/tʃə/ /lup/ /tʌs/ /rɪs/ /dut/
	5H4	/vi/ /fɪnt/ /ɒʃ/ /u/ /lɔpə/
6	6H1	/tʃɒm/ /uts/ /baɪs/ /mɪt/ /jɪs/ /lit/
	6H2	/tʃə/ /tʊgtʃə/ /rɪs/ /tɪl/ /ɒb/ /mʌləz/
	6H3	/fi/ /frɔ/ /u/ /pləʊn/ /æb/ /nu/
	6H4	/lʊm/ /ɔɪt/ /ɒb/ /tʃə/ /bɑdɪ/ /wɛpə/
7	7H1	/tʃə/ /skaɪs/ /uz/ /maɪdɪŋ/ /ɪp/ /tʃə/ /krɑf/
	7H2	/tʃə/ /spaʊn/ /pɪd/ /prɔks/ /mɪt/ /u/ /rɒŋkɪ/
	7H3	/kaɪ/ /krɛnd/ /bæl/ /sɔɪ/ /tʃə/ /tju/ /mʊk/
	7H4	/ʊt/ /kæs/ /gin/ /plɔɪɪŋ/ /zɔ/ /pʌm/ /rik/
8	8H1	/tʃə/ /rɪkəl/ /bɜl/ /kɛst/ /ʃɜ/ /mɒl/ /æb/ /brul/
	8H2	/faɪ/ /kæf/ /lɔtɪd/ /sɛ/ /zɔ/ /u/ /kɔŋ/ /tɔɪn/
	8H3	/tʃə/ /trun/ /inlɪ/ /blopt/ /ɪp/ /tʃə/ /fɪp/ /sʊg/
	8H4	/ɔɪ/ /dɛt/ /kaɪ/ /ild/ /ɔnt/ /ɒt/ /ʃɜ/ /kɑn/
9	9H1	/tʃə/ /paɪt/ /hʊt/ /rɪs/ /dʒaɪzɪŋ/ /lɪm/ /ɪp/ /tʃə/ /mɔk/
	9H2	/wu/ /læn/ /gɑ/ /tʃə/ /spiz/ /ɒp/ /u/ /blɪə/ /bɔɪt/
	9H3	/tʃə/ /jɛm/ /bɔnsə/ /rɪs/ /mʊkɪŋ/ /æb/ /tʃə/ /fɔɪnɪ/ /mʌrə/
	9H4	/ɔɪ/ /kæf/ /gin/ /spəʊɪŋ/ /brɪdɪ/ /dʌdəlz/ /tʃə/ /kɑʊl/ /tʃeɪ/

Table 4.12 Stimuli in condition H

In summary, there were six conditions with a sentence-like prosody and three conditions with a list prosody. There was no prosodic counterpart for the hybrid conditions F and I, nor for the semantically implausible sentences in condition C. The lack of list counterparts for F, I and C was motivated theoretically and by the practical aspects of the experiment. First, investigating the effects of prosody in A vs. B, D vs. E, and G vs. H was considered to be enough to allow for generalisations about the contribution of the prosodic domain to performance on a pseudo/sentence span task. It was not clear that inclusion of prosodic counterparts to conditions F, I and C would add to the findings from the other sentence-like vs. list-prosody comparisons. Second, having included nine conditions, the experiment already involved extensive repetition and it was feared that adding even more conditions would have made the experiment too demanding for the participants. The other alternative would have been to split the experiment into two sessions. However, this would have meant meeting each participant at least twice and it was decided that this was too time-consuming and disruptive of children's classroom activities.

4.3 Experimental conditions: Creation of Czech materials

The aim was to investigate the same nine conditions in Czech as had been examined in English, as illustrated in Table 4.13.

Label	Description	Example			
A	Well-formed sentence:	Zranil	jsem	si	koleno
		Hurt	aux	reflexive	knee
	English translation:	*'I hurt my knee'*			
B	Well-formed sentence with list prosody:	zranil,	jsem,	si,	koleno
		Hurt	aux	reflexive	knee
C	Semantically implausible sentence:	Pískal	jsem	si	rameno
		whistle	aux	reflexive	shoulder
	English translation:	*'I whistled a shoulder'*			
D	Ungrammatical sentence:	Si	jsem	koleno	zranil
		reflexive	aux	knee	hurt
E	Ungrammatical sentence with list prosody:	si,	jsem,	koleno,	zranil
		reflexive	aux	knee	hurt
F	Pseudosentence with content words replaced by nonwords:	Dvonil	jsem	si	vylono
		nonword	aux	reflexive	nonword
I	Pseudosentence with function words replaced by nonwords:	Zranil	tem	so	koleno
		hurt	nonword	nonword	knee
G	Pseudosentence with words replaced by nonwords:	Dvonil	tem	so	vylono
		nonword	nonword	nonword	nonword
H	List of nonwords:	dvonil,	tem,	so,	vylono
		nonword	nonword	nonword	nonword

Table 4.13 Nine experimental conditions in Czech, span 4

Measures were taken to address the typological differences between Czech and English in the creation of the conditions and how this specifically applied to Czech is discussed in this section. As with the English stimuli, Czech condition A provided the basis for the other conditions and conditions B-I were then derived from condition A.

4.3.1 Foundation for all conditions: Condition A

4.3.1.1 Lexical phonology

Generally, Czech words are longer and have more consonant clusters and of a greater range of complexity than in English. This is true even in words which are acquired early (Pačesová, 1968) and was reflected in the stimuli. In addition, while the length of words did not exceed two syllables in the English stimuli, the Czech stimuli contained words which were trisyllabic e.g.:

zpí.va.la	'*she sang*' – verb, past tense, imperfective, singular, feminine
stu.de.né	'*cold*' – adjective, genitive, singular, feminine
mi.min.ko	'*baby*' – noun, nominative, singular, neuter

Similarly, there were a maximum of two adjacent consonants within one syllable as in the English stimuli, but in one instance in Czech, three adjacent consonants arose from a combination of a preposition and a noun and this was allowed:

v trávě	'*in the grass*'
CCCV.CCV	v – preposition '*in*'; trávě - noun, locative, sg., feminine '*grass*'

The rules for the distribution of non-syllabic versus syllabic prepositions in Czech are complex. Chládková (2009, p.6) suggested that '*the nature of the onset cluster that results from linking the non-syllabic preposition to the word is a primary determiner of whether the preposition will take the vocalized form or not, while the prosodic structure of the whole sequence formed by the preposition and the immediately following word may as well play a role in assigning the form of the preposition*'. When the preposition is non-syllabic, it is aligned to the following onset (Palková, 1994). This explains the three adjacent consonants in '*v trávě*'.

4.3.1.2 Morphosyntax

As in English, the sentences for the stimuli were all declarative main clauses with no subordinate clauses, and negation, passives, quantifiers, and questions were not included. Czech has a relatively free word order and a rich morphology: nouns, adjectives and pronouns are marked for gender, animacy, case and number; verbal morphemes express person, number, tense, aspect, and gender. Thus, the number of morphemes in Czech sentences generally exceeds the number of morphemes in English sentences, while the number of words in English sentences is typically higher. This is illustrated by the example below:

The	teacher	read	a	story.	(5 words)
	Učitel	četl		příběh.	(3 words)
	nom.sg.masc.	PT. impf. 3sg. masc.		accus.sg.masc.	
	'teacher'	'read'		'story'	

The first priority was to create stimuli that satisfied the requirements of the experimental design. Since span in words was the outcome measure, stimulus length had to be equal within a block, and increase between blocks. Since content and function words were being compared, a balance between the words from the two classes was necessary. Both these requirements precluded a direct translation of the stimuli given Czech's rich morphology and fewer function words. Notably, Czech is a pro drop language and has no definite/indefinite articles. In order to control length and the balance of function/content words, a modified set of stimuli were created. Rather than translating the English stimuli into Czech, the aim was to create natural Czech sentences with semantics broadly similar to the English stimuli. However, sentence content had to meet other criteria such as familiarity and phonological constraints on the clusters and syllabic structures of words. While it was sometimes possible to create stimuli that were semantically similar to English stimuli, satisfying familiarity and phonological constraints meant that not all the Czech sentences were matched closely to the semantic content of the English sentences. For the full set of stimuli in condition A see Table 4.14.

Block	Label	Stimulus
2	2A1	Ona vaří. *'She cooks'*
	2A2	Hráli si. *'They played.'*
	2A3	Pomoz mu. *'Help him.'*
	2A4	Pojď sem. *'Come here.'*
3	3A1	Led je studený. *'Ice is cold.'*
	3A2	Tancuj s námi. *'Dance with us.'*
	3A3	Máme ho rádi. *'We like him'*
	3A4	Podej nám to. *'Pass it to us.'*
4	4A1	Schovám se za dveře. *'I will hide behind the door.'*
	4A2	Hodil po mně kámen. *'He threw a stone at me.'*
	4A3	Zranil jsem si koleno. *'I hurt my knee.'*
	4A5	Polož to na stůl. *'Put it on the table'*
5	5A1	Já si hraju s autem. *'I am playing with a car'*
	5A2	Had se schoval v trávě. *'The snake has hidden in the grass.'*
	5A3	Zpívala jsem si veselou písničku. *'I was singing a happy song'*
	5A4	My jsme ho včera viděli. *'We saw him yesterday.'*
6	6A1	Budu si malovat na velký papír. *'I will be painting on a large sheet of paper'*
	6A2	V naší školce se mi líbí. *'I like it in our nursery'*
	6A3	Ptal jsem se tě dnes ráno. *'I asked you this morning'*
	6A4	Pojď ven z té studené vody. *'Come out of the cold water'*
7	7A1	Tamto miminko se směje na svou mámu. *'That baby is smiling at her/his mum.'*
	7A2	Tu písničku jsem slyšel asi před týdnem. *'I heard the song about a week ago'*
	7A3	Koláč se peče v naší nové troubě. *'The cake is being baked in our new oven'*
	7A4	Můj tatínek si koupil tu drahou knihu. *'My dad has bought the expensive book.'*

8	8A1	Od září se budu učit anglicky ve škole. *'I am going to be learning English at school from September on.'*
	8A2	Po obědě jsem se napil čaje s medem. *'After lunch I drank a tea with honey.'*
	8A3	V létě jsme chodili na procházku se psem. *'In summer we used to walk our dog'*
	8A4	Zítra si postavím stan u nás na zahradě. *'I will build a tent in our garden tomorrow'*
9	9A1	Potkal jsem se se svou tetou na naší louce. *'I have met my aunt at our meadow.'*
	9A2	Ten mladý herec se díval do toho velkého zrcadla. *'The young actor was looking at the big mirror.'*
	9A3	Pavel se s chutí zakousl do dortu od babičky. *'Pavel bit the cake from his grandma.'*
	9A4	Přes den jsem si stavěl ty krásné barevné kostky. *During the day I was building with the nice colourful blocks.*

Table 4.14 Czech stimuli in condition A

4.3.1.3 Lexical items

The aim was to include items that were high frequency, imageable, familiar and age-appropriate. However, no tool like the MRC Psycholinguistic database was available for Czech. Instead, lexical items which were believed to meet the criteria were chosen, and the stimuli were then checked by four Czech native speakers, including two nursery school teachers, in order to judge whether the items were age-appropriate and familiar to very young children. In addition, a pilot study was conducted with 13 children and items which were consistently challenging were replaced. For more details see section 4.9 on the pilot study in Czech.

4.3.2 Semantically implausible targets: Condition C

The semantically implausible targets in condition C were created following the same principles as in English: stimuli were created by violating selectional restrictions while matching the original stimuli for the number of words and syllables. The full set of stimuli for condition C with the translations into English word-by-word are provided in Table 4.15.

Block	Label	Stimulus
2	2C1	Ona prší. she, rain *'She rains'*
	2C2	Sněžili si (they) snow, refl *'They snowed'*
	2C3	Létej mu. fly, him *'Fly for him'*
	2C4	Spi sem. sleep-imperative, here *'Sleep to here'*
3	3C1	Led je vařený. ice, be, boiled *'Ice is boiled'*
	3C2	Odskoč s námi. jump away, with, us *'Jump away with us'*
	3C3	Známe ho sami. know, him, alone *'We know him alone'*
	3C4	Odlet´ nám to. Fly off, us, it *'Fly it off to us'*
4	4C1	Uspím se za nůžky. (I) fall asleep, refl, for, scissors *'I will fall asleep for scissors'*
	4C2	Snědl po mně kufr. (he) eat, after, me, luggage *'He's eaten luggage after me'*
	4C3	Pískal jsem si rameno. (I) whistle, aux, refl, shoulder *'I whistled a shoulder'*
	4C4	Napiš to na krok. write-imperative, it, on, step *'Write it on the step'*
5	5C1	Já si píšu s oknem. I, refl, write, with, window *'I've been writing with a window'*
	5C2	Med se umyl v noze. honey, refl, wash, in, leg *'Honey has washed in a leg'*
	5C3	Volala jsem si vysokou mastičku. (I) call, aux, refl, tall, cream *'I have called a tall cream for myself'*
	5C4	My jsme ho zítra vařili. We, aux, him, tomorrow, boil *'We boiled him tomorrow'*

6	6C1	Budu si stěžovat na bílý komín. (I) will, refl, complain, on, white, chimney *'I will complain about the white chimney'*
	6C2	V naší misce se mi daří. in, our, bowl, refl, me, go well *'It is going well in our bowl for me'*
	6C3	Bál jsem se tě hned zítra. (I) fear, aux, refl, you, only, tomorrow *'I was scared of you only tomorrow.'*
	6C4	Běž pryč z té mražené mouky. Go-imperative, away, out, that, frozen, flour *'Go away from the frozen flour'*
7	7C1	Tamto koťátko se mračí na svou vílu. that, kitten, refl, frown, at, his/her, fairy *'That kitten was frowning at his fairy'*
	7C2	Tu travičku jsem barvil asi před nosem. That, grass, aux, (I) dye, about, in front of, nose *'I have dyed that grass in front of the nose'*
	7C3	Kámen se pere v naší dlouhé knize. stone, refl, wash, in, our, long, book *'The stone has been washed in our long book'*
	7C4	Můj domeček si našel tu slanou krávu. My, house, refl, find, that, salty, cow *'My little house has found the salty cow'*
8	8C1	Od mokra se budu divit anglicky ve městě. Since, wet-adverb, refl, (I) will, wonder, English, in, town *'Since wet conditions I will be surprised in English in the town'*
	8C2	Po výletě jsem se bál lesa s chlebem. After, trip, aux, refl, (I) afraid, forest, with, bred *'After the trip I was scared of the forest with bread'*
	8C3	V okně jsme tančili na omáčku se lvem. in, window, aux, (we)dance, on, sauce, with, lion *'In the window we danced to the sauce with a lion'*
	8C4	Potom si nakreslím park u nás na silnici. Afterwards, refl,(I) draw, park, at, us, on, road *'Afterwards I will draw a park on the road at ours'*
9	9C1	Loučil jsem se se svou duhou na naší mouce. (I) say goodbye, aux, refl, with, my-refl, rainbow, at, our, flour *'I said good-bye to my rainbow at our flour'*
	9C2	Ten slepý lovec se díval do toho slaného sedadla. This, blind, hunter, refl, look, at, that, salty, seat *'The blind hunter was looking at the salty seat'*
	9C3	Pavel se s mastí zatoulal do hradu od žehličky. Pavel, refl, with, cream, wander off, to, castle, from, iron *'Pavel with the cream has wandered off to the castle from the iron'*
	9C4	Přes dům jsem si koupil ty dlouhé voňavé myšky. Through, house, aux, refl, (I) buy, these, long, fragrant, mouse *'Through the house I have bought the long fragrant little mice'*

Table 4.15 The full set of Czech stimuli for condition C

4.3.3 Ungrammatical conditions

As in English, there were a range of options for creating ungrammatical sentences in Czech. For English, this was done through violations of word order. However, word order in Czech is relatively free and subject/object positions relative to the verb can vary. A basic SVO sentence can undergo all six possible permutations: SVO, SOV, OVS, OSV, VSO, VOS, although SVO, OVS and SOV are the most common. Examples of all the possible orders are provided below (Janda & Towsend, 2000), with the meaning of the sentence always 'Anna has already written the dissertation', and comprised of the elements *Anna* ('Anna', a nominative subject), *už* ('already', an adverb), *napsala* ('wrote', a verb), and *tu disertaci* ('the dissertation', an accusative object):

SVO:	Anna	už	napsala	tu	disertaci.
OVS:	Tu	disertaci	už	napsala	Anna.
SOV:	Anna	tu	disertaci	už	napsala.
OSV:	Tu	disertaci	Anna	už	napsala.
VSO:	Napsala	už	Anna	tu	disertaci.
VOS:	Napsala	už	tu	disertaci	Anna.

The possibility of altering word order in Czech is traditionally associated with its rich case system which includes Nominative, Genitive, Dative, Accusative, Vocative, Locative and Instrumental cases. Many word order changes in English disrupting the grammatical relations between constituents would possibly be equivalent to changing the cases of subject and complement noun phrases in Czech. However, this option was not selected for a number of reasons:

1) It is often phonologically minor changes which mark the differences between cases, so case violations could easily be missed by participants. For instance, noun 'man' in accusative, singular, masculine, animate and in dative, singular, masculine, animate differ by one vowel sound: *muže* vs. *muži*. Such a change is likely to be less salient than a change of word order.
2) Importantly, the lexical content and length of the stimuli were matched across the conditions. Changing the case could affect length, e.g. noun 'sir' in nominative singular masculine animate and in dative, singular, masculine differ by two syllables: *pán* (1 syllable) vs. *pánovi* (3 syllables).
3) Changing case markers alone may have limited effect, as some stimuli (especially short ones) contain few nouns/pronouns. In addition, the subject is phonetically unrealised in many stimuli as Czech is a pro drop language. Since there is no overt

case marking on subjects in many instances, subject-verb agreement could not be violated.

The number of instances of ungrammaticality could be increased by violating agreement in addition to case. However, agreement violation was rejected because its effects would still be limited: although Czech distinguishes seven cases, the endings of some are homonymous, with certain paradigms more homonymous than others, e.g.: the neuter case where nominative = genitive = accusative = vocative. Agreement violation was also rejected for creating the ungrammatical sentences for the reasons outlined in point (2) above, i.e. changing the nominal inflections of the stimuli was likely to result in a different number of syllables.

4) Violation of case agreement while retaining phrasal structure in the Czech stimuli did not take place as this would contrast with the English versions where word order changes were used and all chunks disrupted. If words were kept in the same order in either language, the phrasal structure would not be disrupted, so there would still be chunks within the sentences. The sentence is still structured and even if the chunks are not grammatical, one can still perceive them as chunks. Consequently, the semantic relations in this kind of ungrammatical sentence might be easier to understand than the meaning relations in the English sentence when words are out of their positions and placed quite randomly.

Word order is not entirely free in Czech, and through exploration of different possible violations it is clear that changes in word order disrupt chunking more than violations of case marking and agreement. Since chunking is crucial for grammatical relations, it was decided to make the Czech sentences ungrammatical through word order violations (i.e. phrase structure violations). Moreover, changing word order in Czech makes it possible to retain the same phonological material across grammatical and ungrammatical conditions, as required for the aim of comparing spans across conditions without contamination of other factors.

Full phrases can be freely ordered in Czech as outlined above, but certain constraints apply to the order within phrases and the distribution of clitics[4]. Prepositions always precede their complement noun phrase. Typically, adjectives precede the nouns they modify and quantifiers precede the nouns they modify. Janda and Townsend (2000)

[4] Clitics are lexical items that cannot bear stress by themselves and become attached to stress-bearing lexical items.

pointed out that the placement of clitics in Czech is governed by both phonology and syntax. A clitic (or where there are more clitics in one clause- a clitic cluster) occupies the so-called Wackernagel position. Dotlačil (2004) explains that this is a position after the first syntactic constituent in the clause. If there are more clitics within a sentence, their order is fixed: the first position within the group of clitics is taken by an inflected form of the auxiliary verb be, then reflexive se/si, which would precede pronouns. If there is more than one object expressed by a pronoun, a dative precedes an accusative. The order of the clitics is illustrated by the examples below:

Viděli	jsme	se	v	září.
see-verb,	auxiliary	reflexive	in-preposition	September-
PT, 1.sg., impf.				loc., sg., neuter

'We saw each other in September'

Líbilo	se	mi	to.
please-verb,	reflexive	me-pronoun, dative	it-pronoun, nom.
PT, 3sg, neuter			

'I liked it'

Dal	nám	ji	zadarmo
give-verb, PT, 3sg,	us-pronoun, dative	it-pronoun, fem. acc.	free-adverb
masc.			

'He gave it to us for free'

Based on these observations, it was decided to create ungrammaticality in Czech stimuli by violating both types of constraint on word order: i) disrupting the phrases, e.g. a noun was separated from its specifier or preposition from its complement and ii) misplacing clitics. This is illustrated by the following example:

Grammatical sentence (condition A)

Budu	si	malovat	na	velký	papír.
will-verb	reflexive	paint- verb	on-	big-adjective	sheet-noun,
1.sg.		infinitive	preposition	acc.sg.masc.	acc.sg.masc.

'I will be painting on a big sheet.'

Ungrammatical sentence (condition D)

Velký	si	malovat	na	budu	papír.
big- adjective,	reflexive	paint- verb,	on-	will-verb	sheet-noun,
nom/acc.sg.masc.		infinitive	preposition	1.sg.	acc.sg.masc.

In the grammatical sentence, the noun *papír* is preceded by *velký* (adj) and the adjective + noun *velký papír* are preceded by the preposition *na*. The order of these items is fixed in the grammatical sentence. In the ungrammatical sentence, these items were split: preposition *na* is separated from its complement by the verb *budu* and adjective *velký* is separated from the noun it modifies. In addition, once the noun phrase is separated from the preposition it complements, the marking of grammatical case becomes ambiguous: *papír* as well as *velký* could be nominative or accusative, so in this case there is no agreement marker indicating the relationship disrupted by the word order change.

However, prepositions were not always separated from their complements. As pointed out in section 4.3.1.1, Czech has both syllabic and non-syllabic prepositions, with non-syllabic prepositions aligned to the onset of the following word (Palková, 1994). For example, the prepositional phrase v lese 'in the forest' is pronounced as /vlɛsɛ/. Syllabic prepositions, on the other hand, form a whole syllable and are not dependent on the following syllable ve vlaku 'on the train' /vɛ vlakʊ/. Since the phonological form of the preposition is determined by the item which follows it within a sentence, it was decided to keep the non-syllabic prepositions and the item which followed together. This was necessary to maintain the number of syllables in grammatical and ungrammatical sentences. However, syllabic prepositions were separated from their complements which did not affect either their phonological form or the number of syllables. This is illustrated by the following example of syllabic and non-syllabic prepositions in a grammatical sentence and its ungrammatical counterpart:

Grammatical sentence: **V** létě jsme chodili **na** procházku **se** psem.
in summer aux go-PT for stroll with dog
'*In summer we used to go for a stroll with a dog.*'

V létě jsme chodili **na** procházku **se** psem.

Ungrammatical sentence: Procházku **se** chodili **na** psem **v** létě jsme.

There are three prepositions in the sentence above. V 'in' is non-syllabic; na 'for' and se 'with' are syllabic. In total, there are twelve syllables in the grammatical sentence. As the aim was to keep the number of syllables constant across the conditions, the prepositions retained their phonological form. In order to allow the non-syllabic preposition v to retain the same form and be recognised as a preposition, it was not separated from its complement. If this non-syllabic preposition was separated from the noun in the ungrammatical sentence, it would have to be vocalised and the number of syllables would increase, i.e. there would be twelve syllables in a grammatical sentence but thirteen in an ungrammatical sentence. On the other hand, the syllabic prepositions na and se can stand on their own and therefore it was not necessary for them to stay attached to their complements.

While word order violations were selected as the best option for the Czech stimuli due to the language's relative free word, it was not always possible to create ungrammatical sentences in this way. In particular, the possible permutations were limited within the shorter stimuli, and an ungrammatical sentence could not be built this way in some cases. The grammatical sentence from condition A provides an example below:

Podej	nám	to.
Pass-verb, imperative, 2.sg.	us-pronoun, dative	it-pronoun, acc.sg. neuter

'Pass it to us'

The sentence allows all permutations possible:

Nám to podej.
To nám podej.
To podej nám.
Podej to nám.
Nám podej to.

As none of these permutations resulted in an ungrammatical sentence, it was decided to create the ungrammatical stimuli in 2- to 3-word stimuli by recombining words from sentences of the same length. For instance, for stimuli at span two, the function word *ona* 'she' from example 2.1 was combined with content word *hráli* 'played - 3.pl.' from example 2.2. This maintained the same lexical and phonological material within one span, consistent with the original aim to retain the same material across conditions A and D as illustrated in the Table 4.16:

	Grammatical sentences at span 2	Ungrammatical sentences at span 2
2.1	Ona vaří	Vaří sem
2.2	Hráli si	Ona hráli
2.3	Pomoz mu	Si pomoz
2.4	Pojď sem	Pojď mu

Table 4.16 Ungrammatical conditions for short span in Czech

4.3.4 Non-lexical and hybrid conditions

The same principles which applied to English nonwords were followed in creating the Czech nonwords: nonwords matched the lexical items from the grammatical sentences in their number of syllables, phonemes and in their syllabic structure. However, as outlined above, Czech has more morphemes and this fact was inevitably reflected in the Czech nonwords. The lexical roots were replaced, but the grammatical morphemes remained part of the nonwords. Similarly, if morphemes were present in the English real words, they were also preserved in the nonwords. However, as the number of morphemes in Czech is inherently higher than in English, the Czech nonwords showed more wordlikeness (see section 2.2.2.1), as illustrated in the examples below:

Czech

nonword *oupem* — lexical counterpart: *autem* 'car'

-em is an inflectional morpheme marking instrumental, singular, masculine/neuter

nonword *besličku* — lexical counterpart: *písničku* 'song'

-ička is a derivational morpheme, creating a diminutive of feminine gender

-u is an inflectional morpheme marking accusative, sg., feminine; nonword has the form *besličku*, i.e. -u instead of -a ending, because the context of the sentence requires accusative

Nonword *bavovat* — lexical counterpart: *malovat* 'paint'

-ovat is an inflectional morpheme, marking the non-finite form of a verb.

English

nonword /ɛməlz/ — lexical counterpart: apples

-s is a plural morpheme

nonword /mitʃə/ lexical counterpart: teacher
-er is a derivational morpheme

The creation of hybrid conditions followed the same principles as in English: content words were replaced by the nonwords in condition F and function words were replaced by nonwords in condition I.

4.3.5 Conditions with a list prosody versus sentence-like prosody

Primary stress in Czech is always fixed to the first syllable of a stressed unit. The stress-bearing words can be nouns, adjectives, verbs (except auxiliaries), most adverbs, prepositions of two or more syllables, numerals, non-clitic pronouns, and some conjunctions. All of these words can and normally do bear stress but there are certain exceptions which cannot be stressed. These include all clitics, discussed above in the section on word order, and prepositions, although these sometimes carry stress. Monosyllabic prepositions form a unit with their complement if the following word is not longer than three syllables. Within this unit, the stress is placed on the preposition rather than the following word. For instance, in the sentence *Děti šly* ***do*** *školy* 'Children have gone to school', stress is carried by the preposition *do* 'to' and not by the noun *škola* 'school'. However, in contexts without a preposition, the stress would be placed on the first syllable of the noun: *Dnes není* ***ško****la* 'There is no school today'.

Unlike in English, vowels in Czech are not reduced in unstressed syllables and both long and short vowels can occur in either stressed or unstressed syllables. Palková and Volín (2003, p.1783) pointed out that '*the stress contrast is not reflected on the segmental level: the components of unstressed syllables are reduced neither in their quality nor in their quantity*'. Thus, stressed vowels are not lengthened phonetically as stress is not marked by duration (Palková, 1994).

As in English, an extra item was added at the end of the list to avoid final lengthening. This extra item was removed when the stimuli were edited. Dankovičová (2001) described how phrase-final lengthening has been found in many languages (for an overview see Vaissière, 1983). Dankovičová's findings also reveal that the phrase-final lengthening in Czech more commonly occurs in the last word of the final intonation phrase than in the last words of non-final intonation phrases. It is not yet clear if this is a universal phenomenon, but the evidence so far appears strong and therefore final lengthening was taken into account in the current study.

4.4 Recording the stimuli

The stimuli were recorded in a quiet room where only the experimenter and the speaker were present. For English, the speaker was a female speech and language therapist who was a native speaker of Standard Southern British English. For Czech, the stimuli were recorded by a female native speaker of Czech. The recording procedure for Czech resembled the recording of English stimuli. What was different was the use of Czech orthography instead of the *International Phonetic Alphabet* (IPA), even for the nonlexical stimuli. This is primarily because Czech orthography is transparent with no ambiguity in the pronunciation of nonwords and the Czech speaker was unfamiliar with IPA.

For the English stimuli, a Marantz PMD660 recorder and an external microphone were used; for the Czech version a Marantz Professional PMD620 digital recorder was used[5]. The recordings for each language took place over four sessions. The recording of the stimuli took place as follows: lexical conditions A, C and E were recorded first; the more demanding condition D which required additional practice was recorded in the second session; in the third session, the hybrid and non-lexical conditions were recorded. The fourth session was dedicated to re-recording stimuli which were not clear enough or where an error occurred (either the instructions were not followed properly or a phoneme/word was deleted or substituted).

Stimuli were printed out, with each condition on a separate list. The lexical stimuli were written in English orthography and the non-lexical stimuli were phonemically transcribed. The speaker was given the list of stimuli one condition at a time, starting with the normal English sentences. She was asked to read each stimulus to herself to gain familiarity and then repeat it out loud twice. When a mistake was made, she was asked to correct the stimulus immediately. After the whole list of 32 stimuli was read out (each stimulus twice), the whole list was repeated again. This way, each stimulus was recorded at least three times. When the experimenter noticed a mistake during the recording session, more recordings of the stimulus in question were made. Apart from the experimental stimuli, four practice stimuli were recorded. The recordings were transferred to a computer after each session and any further editing was carried out using the software programmes Audacity (Audacity Team, 2008) and Praat (Boersma, 2001).

[5] A different recording device was used for Czech because the recording had to be made in the Czech Republic and it was not possible to use the same equipment as for English. For English, recordings took place at the Department of Language and Communication Science, City University, where all the equipment was available.

4.4.1 Editing the recorded stimuli

Each stimulus was identified in the recorded string, cut out with Audacity and then saved with an appropriate label. The label consisted of the name of the condition, span and number of the example, e.g. 2A3 stands for a stimulus of span 2, in condition A, the third example. The volume was set to 80db for all of the stimuli using the Praat function 'Modify – Scale Intensity'. The practice stimuli were edited and saved in the same way as the experimental stimuli were. The final set of stimuli had 288 experimental stimuli and 4 practice stimuli in each language for a total of 584 stimuli which were all saved in a .wav format.

4.5 Order of presentation

Because the same lexical items were used in different conditions, a priming effect[6] was likely (Bock, 1986). In addition to the priming of lexical items, syntactic structure could have been primed (Pickering & Branigan, 1999; Ferreira & Bock, 2006) since the stimuli in syntactically well-formed conditions were deliberately matched on the syntactic level (see e.g., the stimuli for span 5 presented in Table 4.1). The content words were identical in conditions A, B, D, E and I. These similarities could lead to effects of order of presentation, hence confounding the order with target variables. Rather than creating different random orders for each participant, it was thought preferable to establish whether order had an effect on performance. Thus, two list orders were generated in order to control for learning effects, fatigue or boredom and also to check for possible order effects. Participants were randomly assigned to one of the two list orders.

The lists always started with grammatical and plausible sentences (i.e. the lexical items, semantics and syntax were in line with English/Czech grammar) with either sentence prosody (List A) or with list prosody (List B). In addition, lexical, nonlexical and hybrid conditions were alternated. There were five lexical, two hybrid and two nonlexical conditions, with the lexical conditions alternating with the hybrid and nonlexical conditions. Both lists started with a lexical condition, followed by a hybrid, then lexical again and so on. The specific comparisons of experimental conditions involved in the research questions were taken into account when constructing the order. For example, research question 1 requires a comparison of conditions A and C, therefore condition A preceded C in List 1, whereas condition C preceded A in List 2.

[6] Priming refers to an increased sensitivity to certain stimuli due to prior experience, i.e. exposure to a stimulus at time 1 influences the response to a related stimulus at time 2. It is believed that exposure at time 1 activates associations in memory, reducing the resources required to execute the action or task.

This principle also applied to the other comparisons required by the research questions. All pairs to be compared were included except for the pair entailed in research question 2 (grammatical vs. ungrammatical sentences), where only one of the possible orders (A followed by D) was satisfied. The other sequence (D followed by A) could not be realised unless another list was added or the sequences motivated by other research questions were violated. Although A always preceded D, the positions of these two conditions in the two lists differed: in List 1, A was presented first and D followed only after four other conditions; in List 2, condition A appeared in the fifth position and condition D followed after two other conditions. As a result, the number and the nature of the conditions presented between A and D differed. The resulting two lists of the nine conditions are presented in Figure 4.4 below.

Figure 4.4 The two orders of presentation of the nine conditions (A-I).

4.6 Procedure

4.6.1 'Game' format of the presentation

The experiment was administered via Windows *Microsoft PowerPoint*, 2003. In order to create a child-friendly task, pictures of animals and colourful creatures were provided (see Appendix H) as a background to sound files and short animations were included as fillers. The combinations of pictures and sounds shown in the PowerPoint slides will be referred to as a 'playing field'. Each condition was represented by one playing field, consisting of 32 coloured ovals which corresponded to the 32 stimuli for that condition. An example of a 'playing field' is provided below for condition B (see Figure 4.5). Section 4.6.2 describes the instructions to children, practice trials and how the experiment proceeded. The procedure was the same for English and Czech, except for the starting points which are detailed for each language in section 4.6.3.

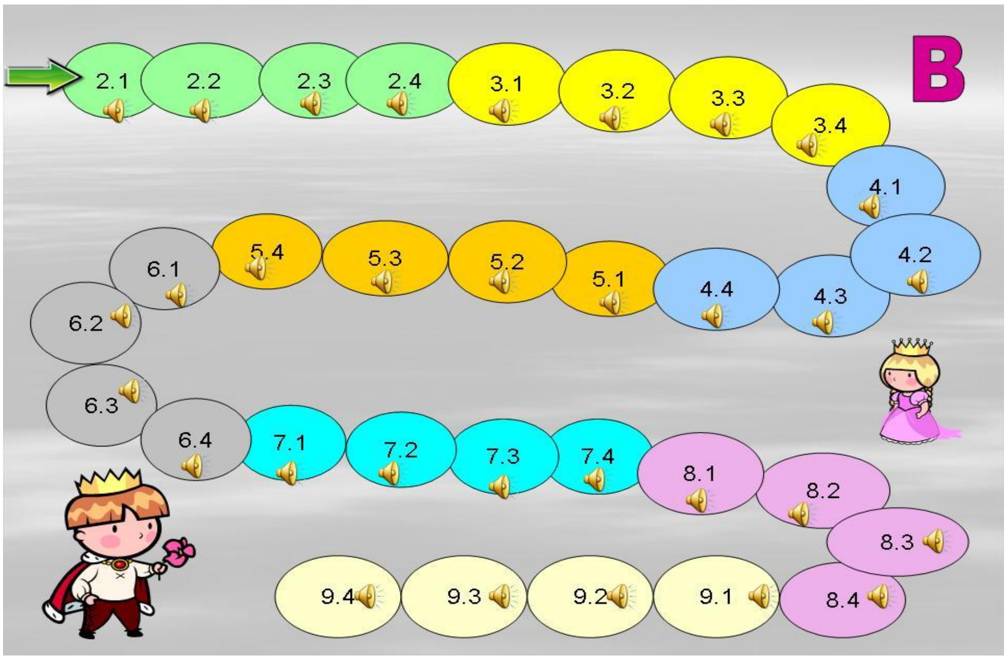


Figure 4.5 Playing field for condition B consisting of 32 ovals representing 32 stimuli (length 2- length 9, with four examples at each length).The small speakers indicate the sound file associated with each oval.

4.6.2 Familiarising the participants with the task

The participant was told that the game consisted of listening to many words and repeating them as accurately as possible. The presentation started with four practice trials. Four pairs of animals appeared on the laptop screen, with the animals forming a pair which were mirror images of each other, differing only in size (see Figure 4.6).

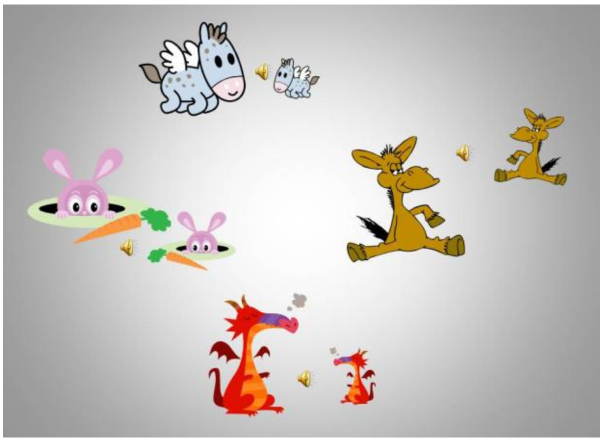


Figure 4.6 Images attached to the practice trials

The child was told that the larger of the two characters 'is a mummy' and the smaller one 'is a baby', and that 'the baby has to repeat what the mummy says'. The experimenter then played the 'mum's voice' and the child was asked to help the baby to repeat it. If the child did not repeat a stimulus, the experimenter asked the child again to repeat the stimulus and played the same stimulus once more. An example of the script that the experimenter used in the practice section and the responses of one of the child participants is detailed below:

Experimenter (E): *Let's look at these animals. There is a mummy and a baby. Can you see them?*
Child (C): *Yes.*
E: *Good. The mummy is going to say something and the baby has to repeat it. Can you help the baby to repeat it?*
C: *Yeh.*
E: *Ok. That's very nice of you. Now listen carefully and try to say what mummy is going to say. Are you ready? Which one would you like to listen to first?*
C: Carrot one [the pink rabbit next to the larger carrot].

There were four pairs of animals and the child could choose which animal they wanted to listen to. It was observed that the opportunity to influence events seemed to encourage the participants' involvement. After the experimenter was sure that each child understood the task, further instructions were given and the experiment was presented, e.g.:
E: *You're doing really well so I have some more listening and repeating for you. We're going to listen to more words and some of them are really easy, some of them are a bit difficult, some of them are funny and some of them will be new for you, you have never heard them before. But we will just try, ok? I will play the words for you and you will repeat them. Are you ready?*

There were two list orders and participants were randomly assigned to one of them, with half of the participants receiving list A and half list B. The practice trials and instructions were identical for the two lists. The order of conditions and motivation for this order were detailed in section 4.5.

The stimuli for each condition were accompanied by different pictures (see Appendix H for the backgrounds for all conditions). The experimenter controlled when and which stimulus was played. Immediately after the participant finished the repetition of one stimulus, the next stimulus was played; breaks were placed only between conditions. In addition, each condition was followed by an animated sequence. In total, participants watched 9 different animations.

4.6.3 Starting points for conditions in English and in Czech

Following the practice trials for each condition, testing started with a pseudo/sentence length intended to be well within the participant's capacity. English and Czech differed on the starting points, with testing in English starting with slightly longer stimuli than in Czech. This was motivated by the results from pilot studies carried out in both languages (see section 4.9).

In English, conditions A, B and C started at span length 7 for four-year-olds and at span 8 for five-year-olds. Conditions D and E started at span 4 for four-year-olds and at span 5 for five-year-olds. The starting-point for the presentation of non-lexical stimuli was set to span length 2 for all participants, and for hybrid conditions at span 2 for four-year-olds and span 3 for five-year-olds.

In Czech, conditions A, B and C started at span length 6 for four-year-olds and at span 7 for five-year-olds. Conditions D and E began at span 3 for four-year-olds and at span 4 for five-year-olds. The starting-point for the presentation of non-lexical stimuli was set to span length 2 for all participants, and for the hybrid conditions span 2 for four-year-olds and span 3 for five-year-olds. See Table 4.17 for an overview. Starting points for adults were 1 span longer than for five-year-olds. Depending on the participant's performance on the starting span, testing proceeded either to the next higher or lower span.

Condition	Four-year-olds		Five-year-olds	
	English	Czech	English	Czech
A	7	6	8	7
B	7	6	8	7
C	7	6	8	7
D	4	3	5	4
E	4	3	5	4
F	2	2	3	2
I	2	2	3	2
G	2	2	2	2
H	2	2	2	2

Table 4.17 Overview of the starting points for presenting stimuli

If the test started at any block other than the first and the subject failed to complete three out of four stimuli within the presented block, the preceding block of stimuli was presented and the test was continued from there without the failed block being administered again. If the participant succeeded on the block, testing progressed through successive lengths until he or she was unable to repeat the majority (3 out of 4) of the stimuli in a block of the same length and condition, at which point the discontinue rule (see section 4.1.1) applied. The aim was to establish the maximum span (the greatest length at which at least three out of four stimuli within one block were repeated correctly). This procedure provided a measure of each participant's span without an excessive number of trials by targeting the subject's threshold and avoiding presentation of items that were either too difficult or too easy.

When a participant failed to correctly repeat at least two lists at length 2, span was scored as length 1[7] (for a similar approach see Adams & Gathercole, 1996).

4.6.4 Rationale behind the 'game' presentation

Efforts were made to introduce variety into the experiment to compensate for the more repetitive nature of the experimental activities of listening to and recalling the stimuli. The presentation of stimuli as a game in PowerPoint was chosen for two reasons. First, to make it more interesting for the participants, but not so distracting that the participant's attention could not be primarily focused on the stimuli. Second, as a span task was employed, it was necessary to choose a method of presentation which made it easy for the experimenter to control the administration of the stimuli. The game presentation allowed progress to the next block if the participant succeeded in recall, but also backtracking if the stimuli were too challenging.

After each playing field was completed, a short animation was presented on a new slide. The appearance of the animation was controlled by the experimenter. The aim of the animation was twofold. First, it was intended to reward the participants for completing each playing field and also provide motivation to attempt the next playing field (i.e. next experimental condition). Second, the animation was meant to serve as a kind of filler. It allowed for participants to have the opportunity to have a short conversation about events on the screen and so break the monotony of repetition. The animations presented after each condition differed and this also provided novelty.

4.7 Scoring of the experimental task

4.7.1 Scoring criteria

Each stimulus was scored as correct or incorrect. The scoring criteria for each English and Czech condition are provided below. Since span was the outcome measure, the whole stimulus was scored as correct/incorrect.

4.7.1.1 English

Lexical conditions A B C D E were considered to be correct if:

- ✓ all words were present in the correct order
- ✓ all syllables of the words were present in the correct order
- ✓ all inflections were present (e.g. 3.sg. present tense -s, plural -s, past tense -ed)

[7] Out of the 140 participants, this only happened in 1 case in condition D, 1 case in condition E, 2 cases in condition G, and 1 case in condition I. It was more frequent in condition H (the list of nonwords) when this occurred on five occasions (2 English-speaking children and 3 Czech-speaking children).

Hybrid and non-lexical conditions F I G H were considered to be correct if:

- ✓ all items were present in the correct order
- ✓ all syllables were present in the correct order

Since precise repetition of nonwords was not a key issue, some allowances were made for the syllables within nonwords which were considered to be correct if:

- ✓ at least one consonant from a syllable onset was present (e.g. for syllable structure CCV, the syllable was scored correct if only one member of the consonant cluster was present: e.g. target proz – 'roz' ✓, 'poz' ✓)
- ✓ the onset and the nucleus were preserved, i.e. the coda could be omitted: e.g. lup: 'lu' ✓ bais: 'bai' ✓
- ✓ the nucleus vowel was identical, or was within a similar vowel space. Vowels were acceptable if they differed in no more than one feature (front vs. back, high vs. low, tense vs. lax etc).

The experimental tasks in the present study were scored less stringently than nonwords in repetition tests such as CNRep, NRT and WMTB-C (see sections 1.3.2 and 2.2.2.5), largely because the focus of the current study is on the relations *between* items rather than items themselves. The CNRep and NRT are concerned with the recall of word-level phonology, so phonological precision of the response is critical (with allowances only for systematic substitutions that are due to developmental phonological processes). For the purposes of the present study, the key issue was not the phonological accuracy of each item, but the preservation of all items and their sequence. It was therefore sufficient to determine whether each item was attempted in the response, and the more lenient scoring criteria achieved this. In the WMTB-C span task, items are largely restricted to CVC monosyllables, so their phonology is less challenging than the phonology of items in the present span task and is critical for determining whether the item has been produced and therefore for scoring the string. If more stringent criteria were employed in the present study, floor effects could occur and the planned comparisons may not be meaningful. Allowances for limited phonological deviations in the present study made it possible to identify differences between nonword conditions, and made for more conservative evaluation of differences between word and nonword conditions.

For examples of correct and incorrect responses, see Table 4.18 and Table 4.19.

CORRECT			
All items present in correct order	T		apples, grow, on, trees
	R	✓	apples, grow, on, trees
All syllables present	T		I have gin an embel.
	R	✓	I have gin an embel.
At least one consonant from the onset present	T		They bloid.
	R	✓	They loid.
Nucleus vowel copied/matched to target vowel in length + close in vowel space	T		/jʌtʃ tʃə rəʊk/
	R	✓	[jɒtʃ tʃə rəʊk]

Table 4.18 Examples of correct responses according to scoring criteria in English

INCORRECT			
All items present, wrong order	T		apples, grow, on, trees
	R	×	apples, **trees**, on, **grow**
Items in different morphological form	T		I have seen an angel.
	R	×	I have seen angels**.**
Word/syllable omitted	T		My friend will buy the new book.
	R	×	My friend will buy the ___ book.
Word/syllable added	T		The young dancer was looking at the shiny mirror.
	R	×	The young dancer was looking at the **nice** shiny mirror.
Word/syllable substituted	T		I have gin an embel.
	R	×	I have **loon** an embel.

Table 4.19 Examples of incorrect responses according to scoring criteria in English

In summary, in order to be scored as correct: i) all syllables in the string had to be present ii) at least one consonant from the syllable onset had to be present and iii) the nucleus vowel had to be preserved. In cases when a vowel could not be clearly classified, it had to match the target vowel in length and had to be within a similar vowel space in order to be scored as correct. Strings were marked as correct if all items were produced in a correct order. If all items were recalled but did not follow the presented order, the response was counted as incorrect. In addition, a string was scored incorrect when a whole word or syllable was added, a word was missing or extra phonetic material was inserted between target items. The string was also scored as incorrect when a word was substituted or put in a different morphological form, e.g. bus instead of buses.

4.7.1.2 Czech

Due to the free word order in Czech, the scoring criteria slightly differed from the criteria described above for English. As discussed in section 4.3.1.2, the order of the clausal elements in Czech is free, while the order within phrases is relatively fixed. In addition, the ungrammatical sentences were created by violation of word order within phrases. These characteristics of Czech led to certain allowances for violation in the order of words, but only in the grammatical conditions. As some Czech-speaking children changed the word order spontaneously (which no English-speaking child did) the changes appeared to reflect the properties of the language (see section 6.3.3). Therefore, it was decided to allow word order changes in the Czech grammatical sentences. The word order scoring in Czech is discussed further in section 6.4.3. Scoring for the Czech stimuli is detailed below:

Lexical conditions A B C were considered to be correct if:

- ✓ all words were present
- ✓ all syllables were present in the correct order
- ✓ all inflections were present (e.g. drah-ou: 'expensive' –adj.,accus.,sg.,fem. zpíval-a: 'sing' – verb, PT, imperf., sg. fem.)
- ✓ words were in the presented order or in a different order but with grammatical relations preserved so that the sentence was still grammatical with the meaning relations preserved

For examples of correct and incorrect responses in Czech, see Table 4.20 and Table 4.21.

CORRECT			
All items present in correct order	T		*Můj tatínek si koupil tu drahou knihu.* my dad refl bought that expensive book 'My dad has bought that expensive book'
	R	✓	*Můj tatínek si koupil tu drahou knihu.*
Word order changed, but grammatical relations are preserved	T		*Můj tatínek si koupil tu drahou knihu.*
	R	✓	*Můj tatínek si tu drahou knihu koupil.*
	R	✓	*Tu drahou knihu si koupil můj tatínek.*

Table 4.20 Examples of correct responses according to scoring criteria in Czech

INCORRECT			
Word order changed; sentence has become ungrammatical	T		*Můj tatínek si koupil tu drahou knihu.* my dad refl bought that expensive book 'My dad has bought that expensive book'
	R	×	**Můj **si tatínek** koupil tu drahou knihu.*
	R	×	**Můj tatínek **koupil si** tu drahou knihu.*
Word omitted	T		*Můj tatínek si koupil tu drahou knihu.*
	R	×	*Můj tatínek si koupil ___ drahou knihu.*
Inflection substituted	T		*Můj tatínek si koupil tu drahou knihu.*
	R	×	*Můj tatínek si koupil tu drah<u>ý</u> knihu.*
Word added	T		*Můj tatínek si koupil tu drahou knihu.*
	R	×	*Můj **drahý** tatínek si koupil tu drahou knihu.* 'My dear dad has bought that expensive book'

Table 4.21 Examples of incorrect responses according to scoring criteria in Czech

Lexical conditions D E were considered to be correct if:

✓ all criteria from lexical conditions A, B, C applied, other than the allowance for word order changes. In the ungrammatical conditions, words had to be produced in the presented order.

Target	Tatínek drahou tu si knihu koupil můj.	
Response	Tatínek drahou tu si knihu koupil můj.	✓ all words copied in correct order
Response	Tatínek drahou <u>si tu</u> knihu koupil můj.	× word order changed

Hybrid and non-lexical conditions F I G H were considered to be correct if:

✓ all items were present in the correct order

✓ all syllables were present in the correct order

As in English, span was the outcome measure and the whole stimulus was scored as correct/incorrect. Therefore, some allowances were made for the syllables within nonwords which were considered to be correct if:

✓ at least one consonant from a syllable onset was present (e.g. syllable structure CCV.CVC 'dvonil' produced either as donil CV.CVC ✓ or vonil CV.CVC ✓)

✓ the onset and the nucleus were preserved, but the coda could be omitted: e.g. CVC.CVC 'duncuj' produced as CV.CVC = ducuj ✓

✓ the nucleus vowel was copied, or matched the target vowel in length and was within a similar vowel space: /ɪ/ - /ɛ/; /ɛ/- /a/; /a/ - /o/; /o/ - /ʊ/

4.7.1.3 Acceptability of sentence-like versus list prosody conditions

The aim of these conditions was to investigate effects of prosodic structure on preservation of lexical items and morphosyntax in the model, but not the preservation of prosody itself. Therefore the imitation of sentence versus list prosody was not scored. Whether the participants reproduced the list stimuli with a sentence-like prosody or vice versa, a correct response was awarded if other scoring criteria were met.

Likewise, where the target morphemes/words have prosodic variants in Czech and English, the child was credited for correct production for any of the variant forms. For instance, in the Czech list conditions B and E, some children produced the non-syllabic prepositions as consonant + schwa (e.g. /və/ 'in') or consonant + full vowel (e.g. /vɛ/ 'in') and both instances were scored as correct (for more details about the non-syllabic prepositions in Czech see section 4.3.1.1. In English, children sometimes reduced auxiliaries, e.g. 'I've been blowing pretty bubbles' instead of 'I have been blowing pretty bubbles' and this was allowed.

4.7.2 Maximum span

The highest span at which the participant could successfully repeat at least three out of four strings of the same length was considered to be the maximum span (for more details, see section 4.1.1). When the participant correctly produced *two* stimuli at the next highest/first failed span, a half point score was awarded. For instance, if she/he repeated three stimuli correctly at span 5, two at span 6 and only one at span 7, the maximum span would be 5.5 (i.e. half point given for span 6). This modification was implemented to allow for more fine-grained scoring.

4.8 Czech version of the British Picture Vocabulary Scale

As there is no standardised vocabulary test in Czech, the BPVS was adapted and translated into Czech for the purpose of this study (the full version used is presented in Appendix N). The main purpose of the vocabulary test used in this study was to see if there is any relation between performance on repetition tasks and vocabulary knowledge. In other words, the vocabulary test was not meant to establish whether vocabulary knowledge is age-appropriate, but rather to deliver enough variation between children to allow analysis between vocabulary knowledge and scores on repetition tasks. Input on the translations was received from several native speakers of Czech, along with the comments from the two nursery teachers from the pilot study and the final version of the test was modified accordingly.

In most cases, items were easy to translate from English to Czech and matched the pictures available in the British booklet. However, issues arose when the Czech equivalents were established:

- Where English has two words, Czech often only has one: 'ruka' is equivalent to English '*hand*' and '*arm*', 'krk' stands for '*neck*' and '*throat*', 'želva' covers both '*turtle*' and '*tortoise*', 'prst' corresponds to '*toe*', '*finger*' as well as '*digit*', 'skákání' covers both '*diving*' and '*jumping*'.
- Czech adjectives are marked for gender, which only applies to personal pronouns in English. The Czech adjectives were presented in a masculine form which is considered to be unmarked
- Some English words are expressed by two items in Czech: '*orbit*' – 'oběžná dráha'; '*nostril*' - 'nosní dírka', '*easel*' – 'malířský stojan'
- Some words (e.g. '*feline*') are unfamiliar, low frequency and morphologically unrelated to their semantic real world equivalents ('*cat*'). This is not the case in Czech: '*feline*' – 'kočičí' is related to 'kočka' – '*cat*'.

There were several BPVS items which were not suitable for the Czech adaptation. This was because in some cases target and distractor pictures in the BPVS corresponded to the same word in the Czech language, e.g. '*skákání*' is used for both '*diving*' and '*jumping*'. In such cases, one of the other pictures on the page which corresponded to a distinct lexical item in Czech was used as the target. This was done to ensure that only one out of the four possibilities would be correct. This also allowed for the same visual material to be used for both languages.

The Czech version of the BPVS was an adaptation intended to serve the needs of this study and the main aim was not to compare children's performance across the two languages. From the examples above, it is clear that the sets of words in the two languages differed on many levels.

4.9 Changes following the pilot study

A pilot study was run to find out if the methodology outlined above would address the research questions outlined in chapter 2. The pilot study had two aims: to test the use of the span procedure at a pseudo/sentence level and to identify problematic items and/or factors. In total, three pilot studies were carried out with: (1) English-speaking adults, (2) English-speaking children, and (3) Czech-speaking children. A pilot study with adults was conducted first as the technique itself was novel. The results from the pilot studies will not be presented here since testing the methodology was the primary aim. Instead, several issues related to the materials and presentation which were revealed by the pilot study are discussed below.

4.9.1 Theoretically motivated changes in conditions and materials

The pilot study only had 8 conditions. Following the pilot, experimental condition I was added (nonwords inserted only into function word slots, with content words and sentence-like prosody preserved). This addition arose from the literature review and research questions presented in chapter 3. Condition I is a counterpart to condition F, where nonwords are inserted only into content word slots, with function words and sentence-like prosody preserved. The comparison between conditions F and I allows an evaluation of the extent to which familiarity of function vs. content words affects recall. In the original set of stimuli, content words were in the majority and thus modification of the stimuli was required. The stimuli were modified to achieve an equal number of function and content words within each span (for more details see section 4.2.1.3). In the pilot study, all participants received the stimuli in an identical order. For reasons outlined in section 4.5, two orders of presentation were created and used in the main study.

4.9.2 Problem items

When produced verbally, it was found that the placement of several items in the syntactically ill-formed sentences made them open to misperception which resulted in a syntactically 'more correct' sentence. In particular, when 'the' was placed in the final position of a sentence, it could be perceived as the adverb '*there*' which is liable to reduction in this position. Take for example the pseudosentence '*Story a read teacher the*': many participants repeated the final word in this ungrammatical sequence as '*there*'. Similarly, 'an' following an auxiliary verb was often repeated as a reduced negative particle *'n't'*, as in the pseudosentence '*I angel have an seen'*. Therefore, these stimuli were changed for the main experiment.

4.9.3 Presentation

There were no specific practice stimuli in the pilot study, and participants were given the short stimuli from condition A as practice trials instead. In order to give participants more opportunity to grasp the task and make sure that none of the stimuli would be excluded from having been used as test items, practice trials with newly designed stimuli were added to the main study (for more details on practice trials see section 4.6.2).

The pilot study established appropriate starting-points for the experimental conditions. As a general rule, four-year-old children were given more of the shorter stimuli, while five-year-old children scored slightly higher which meant they were administered more of the longer stimuli. The starting-points for each condition were adjusted according to the children's age (see section 4.6.3).

In order to establish whether the administration of the BPVS should precede or follow the experimental task, both options were tested on the children. It was found that children engaged better with the experimental task if the BPVS was presented first. This may be because children are not required to speak during the BPVS and it allowed them time to become comfortable with the experimenter before they were required to produce output.

Obtaining a high-quality recording with children can be more demanding than with adults, as the children may have less understanding of the process and reduced attention spans. Adults may better understand that their actions can influence the quality of the recording and it is therefore desirable that they speak loudly enough, move little, and do not produce additional sounds from tapping their feet or touching the mouth while speaking. In addition, some schools and nurseries could potentially be noisier environments than those used for adults. Therefore, several different types of recorders, microphones and headphones were tested.

The best results were achieved with a Marantz Professional handheld recorder PMD620. This device has two built-in microphones and an automatic gain control system, which can make a quiet signal louder and loud signal quieter. This was useful when recording children as it was able to accommodate better the unexpected changes in the distance between the participants and the microphone arising from the children's movement.

In order to provide high-quality sound in the input and prevent background noise, headphones designed for children were used. These headphones were easy to adjust and decorated with cartoon pictures and the child participants willingly wore them.

In between the conditions, the children's performance was praised and they were verbally encouraged to complete the tasks. If more encouragement was needed, children were given stickers. The stickers were only available between the experimental conditions to minimise disruption the performance on the 'playing field' and were intended to encourage the children to proceed to another condition and complete further tasks. Overall, it was found that the four- and five-year-old participants readily cooperated with the tasks.

CHAPTER 5

The Study

5.1 Introduction

This chapter presents the participants, procedure and results for the experimental task outlined in chapter 4. The reliability of the data is assessed first and then the results are presented in three sections: section 5.5.2 focuses on between subject factors including age, gender and order of presentation. Section 5.5.3 presents the means and standard deviations for the experimental conditions across different age and language groups, followed by the results of the planned comparisons in section 5.5.4 which addressed the research questions introduced in chapter 3. The final section 5.5.5 presents the results of the correlational analyses which assessed the relation between receptive vocabulary and repetition tasks. The formal ethical approval for the study was granted by *City University Research Ethics Committee* and by Research ethics board of the *Institute of Psychology, Academy of Science of the Czech Republic* (see Appendix L for the documents).

5.2 Participants

There were a total of 140 participants in the study. They all had English (N=70) or Czech (N=70) as a first language and were either children (4;0-5;11) or adults (18-40 years) as shown in Table 5.1 and Table 5.2.

Children	Number of participants	Gender M/F	Mean age	Age range	SD
English	50	17/33	4yrs; 11mths	4;0-5;11	6 mths
Czech	50	23/27	4yrs; 11mths	4;1-5;11	6 mths

Table 5.1 Overview of participants – children

Adults	Number of participants	Gender M/F	Age range
English	20	8/12	22-40 years
Czech	20	8/12	18-40 years

Table 5.2 Overview of participants – adults

The whole sample consisted of 140 child and adult participants, of which 54 were males and 84 were females. The highest proportion of females was found in the sample of English-speaking children (33 out of 50). The classrooms the child participants came from had a more even distribution of boys and girls, but more of the girls' parents consented to their child's participation.

The children were recruited via nurseries and primary schools in the London area and nurseries in the Czech Republic. The nurseries and schools approached were from a range of areas both in London and the Czech Republic in an attempt to provide participants from different socioeconomic backgrounds, ethnicities and neighbourhoods. Parents received an invitation for their children to participate in the research with further information about the experiment, a consent form and the parental questionnaire. The information leaflets, consent forms and parental questionnaires were distributed by the schools and nurseries. Only children whose parents gave signed consent were included. In the Czech sample, approximately 90% of the parents who had been approached about the project agreed for their children to participate. In the English sample, the return rate varied between the schools and ranged from 10% to 60%. Seven children from the English sample were tested at their home. Differences between Czech and English schooling patterns are discussed further in section 6.4.

Children were only included if no concerns about their language had been reported by their teachers or parents and they had not been referred for speech and language therapy; they had no known hearing loss or neurological illness; and they spoke either English or Czech as a first language. Information about their language use, health and family background was obtained from a parental questionnaire (see appendix M).

In both samples, the proportion of parents with a higher degree was higher than seen in the London or Czech populations as a whole, although the Czech sample more closely resembled the proportion of individuals with a higher degree in the Czech population (see Table 5.3). Ethnically, both samples reflected their larger populations they were coming from (see Table 5.4). The Czech population is more homogeneous and this was reflected in the sample for the current study. 94% of the children were monolingual Czech and 6% had an additional language. 86% of the children were monolingual English and 14% had an additional language. General population statistics was taken from *Census 2001, Office for National Statistics* for the UK population and from *Sčítání lidu 2001* [Census 2001], *Český statistický úřad* [Czech Statistical Office].

	London			Czech		
	mothers	fathers	General population	mothers	fathers	General population
No qualification	2	6		2	4	
Higher degree	57	53	31	26	20	16

Table 5.3 Educational background of parents of the participants compared to population samples presented in percentages

	London		Czech	
	white	other	white	other
Current sample	72	28	96	4
Population	69	31	94.2	4.1*

Table 5.4 Ethnicity background of parents of the participants compared to population samples presented in percentages * 1.7% did not respond.

In total, 53 English-speaking and 53 Czech-speaking children participated in the study. Three children were excluded from the English sample: two children did not finish the task and one child had speech problems and unintelligible speech. Three children were excluded from the Czech sample: one child did not finish the task and interaction with two children revealed speech problems. All participants from the adult sample were included in the study.

5.3 Procedure

The experimental task was described in chapter 4, see section 4.2 for the English version of the task and section 4.3 for the Czech version. The main focus of this study was 4- and 5-year-olds, and the experiment was designed accordingly. The procedure of the main experiment was the same for children and adults, but minor adjustments were made with the adult participants. Section 5.3.1 describes the procedure for children in detail and section 5.3.2 points out where the procedure was different for adults.

5.3.1 Procedure for children

The child participants were tested in a quiet room at their school/nursery or at their homes. All participants received the British Picture Vocabulary Scale or the Czech adaptation, followed by the experimental repetition task. At the beginning of each session, participants were given a blank sheet of paper which they wrote their names on, with help from the experimenter if necessary. They were allowed to choose one sticker which was put on their sheet before the assessment started. The administration of the vocabulary test took between 10 to 15 minutes. Upon completion of the task, children were praised for their participation and given another sticker. The participant then proceeded to the computer-administered task which was auditorily presented on a Sony VGN-NS20Z laptop computer equipped with two sets of headphones (for full details on the administration of the experimental task see section 4.6). All instructions for the task were given orally by the experimenter (for more details see section 4.6.2).

The stimuli were presented over headphones. Before the testing started, background noise such as fans, air-conditioning or noise coming from open doors and windows was checked and minimised when possible. The volume on the headphones was set to 55db. During the practice trials, the children were offered an opportunity to raise or lower the volume and it was changed if requested, but 55db was suitable for the majority of the children. The experimenter was equipped with a second set of headphones and could monitor what the children were listening to.

Children were randomly assigned to one of the two list orders (25 Czech children received order A and 25 received order B; 25 English-speaking children received order A and 25 received order B), and this was recorded on the scoring sheet. As the experiment proceeded, participants' responses after each stimulus were recorded by the experimenter on a scoring sheet. The administration was manually controlled by the researcher and each stimulus was only played once unless a child did not hear the stimulus either because of external interference such as someone entering the room, or because the child spoke as the stimulus was being played. The children's responses were recorded on a Marantz Professional PMD620 digital recorder which was set to record WAV files at a 44.1 kHz sample rate. The files were saved on an SD card and then transferred to a computer for later offline scoring.

Throughout the task, participants were verbally praised regardless of their performance and encouraged to continue the task until stimuli from all of the experimental conditions had been administered. There were small breaks between the administration of the conditions (for more details see section 4.6) where participants could watch a short animation and choose a sticker to place on their name sheet. At the end of the session, participants were thanked for their participation.

5.3.2 Procedure for adults

The adult participants only participated in the experimental task. They were not given the vocabulary test or the stickers, but could watch the short animations if they wanted. The procedure was the same in all other respects.

5.4 Reliability

The reliability of the data needs to be established before analysing the data and addressing the key experimental questions. An assessment is reliable if it produces the same results across a number of possible situations, e.g. different raters or different times. In this section, the reliability of scoring across two situations (online and offline scoring) and across two raters is assessed. For each experimental condition, the dependent variable was a span. As span was measured on a continuous scale, intraclass correlation coefficient (ICC), a measure suitable for continuous data, was used to assess reliability. ICC was preferred over the Pearson correlation coefficient (PCC) because PCC primarily detects linear associations between two measures (Streiner & Norman, 1995). A reliability measure should estimate the extent to which raters agree, not a linear function between the scores of raters (e.g. rater A always marks 2 points higher compared to rater B) and therefore the ICC was used as a reliability measure.

There were two types of scoring. Online scoring was carried out immediately after each response, and offline scoring was done later from recordings. Two types of reliability were measured: intra-rater reliability (see section 5.4.1) and inter-rater reliability (see section 5.4.2).

5.4.1 Intra-rater reliability

The performance of all children was scored online by the experimenter and later scored independently by the same experimenter offline. The intraclass correlation coefficient between the online and offline scores was calculated for both languages across all conditions. The results from both the English and in Czech stimuli are reported for each condition. The ICC with 95% confidence interval, and type-absolute agreement, is presented in Table 5.5. Values from 0.61 to 0.80 indicate substantial agreement, and from 0.81 to 1.00 indicate almost perfect agreement (Landis & Koch, 1977).

Intraclass correlation coefficient	**A**	**B**	**C**	**D**	**E**	**F**	**I**	**G**	**H**
English	.96	.96	.93	.83	.74	.93	.76	.78	.87
Czech	.97	.90	.95	.84	.85	.91	.88	.69	.83

Table 5.5 Overview of intraclass correlation coefficient between online and offline scoring

The agreement for the majority of experimental conditions was perfect by this measure and a small number of conditions showed substantial agreement. The mean differences between online and offline scoring for all conditions in both languages are presented in Table 5.6 below.

Mean differences between online and offline scoring	Mean for English	Mean for Czech
A	-0.03	-0.07
B	0.05	0.09
C	0.04	0.11
D	0.13	-0.05
E	-0.06	-0.02
F	-0.01	0.11
G	0.01	-0.04
H	0.08	-0.01
I	-0.04	-0.14

Table 5.6 Mean differences between online and offline scoring across all conditions in English and Czech

Table 5.6 shows that the biggest mean difference between the online and offline scoring in English was 0.13 in condition D (SD=.6) and in Czech, 0.14 in condition I (SD=0.51). The figures for the mean differences between online and offline scoring presented in Table 5.6 demonstrate no bias towards either type of scoring. The negative values indicate that offline scoring led to slightly higher spans than online scoring, while the positive values indicate the reverse.

The results from online and offline scoring were almost identical. An ANOVA did not reveal any significant differences between online and offline scores at a value $p<0.01$ for any conditions either in English or Czech. However, at a value $p<0.05$ a significant difference between online and offline scoring was detected in condition C in Czech. As the results from both types of scoring were almost identical, it was decided to primarily use the data from the online scoring method for the main analyses. Online scoring was chosen as it reflects the way that this kind of testing would usually take place in a clinical setting. The summary of results based on means from both types of scoring (online and offline) can be found at Appendix K.

5.4.2 Inter-rater reliability

Ten percent of the samples were given to a second rater for independent scoring. This selection consisted of the results of 5 English-speaking children or 5 Czech-speaking children, randomly selected from the original sample. The rater for English had an MA in linguistics and was a native speaker of English; the rater for the Czech sample had a PhD in linguistics and was a native speaker of Czech. Inter-rater reliability was assessed for offline scoring. The results are summarised in Table 5.7.

Intraclass correlation coefficient	A	B	C	D	E	F	I	G	H
English	---	.97	.90	.80	---	.72	.79	.93	.89
Czech	.96	.96	.93	.73	.73	.98	.95	.83	.96

Table 5.7 Overview of the intraclass correlation coeficient between rater 1 and rater 2. ICC was not computed in English conditions A and E where the scale had zero variance

As shown in Table 5.5 and Table 5.7, the values fell between 0.72-1.0 for both raters. It can thus be concluded that the data were reliable.

5.5 Results

5.5.1 Introduction

The results section is divided into three parts. Section 5.5.2 explores factors which were not experimental variables but could influence the results. In order to investigate between subject factors, the effects of receptive vocabulary (section 5.5.2.1), order of presentation (section 5.5.2.2), age (section 5.5.2.3) and gender (section 5.5.2.4) were assessed. The second part of the results section focuses on the experimental conditions. Means and standard deviations for each condition are presented (section 5.5.3), followed by the results of the planned comparisons outlined in section 5.5.4. Finally, section 5.5.5 investigates the relation between receptive vocabulary and i) repetition of sentences, ii) lists of words and iii) lists of nonwords.

5.5.2 Between-subject factors

5.5.2.1 General language skills – Picture vocabulary task

The rationale for the administration of the BPVS and a Czech adaptation was provided in section 4.1.3. The scores for the Czech-speaking and English-speaking children are presented in this section. Standardised scores could be obtained for the English sample as the BPVS is standardised for British English. The mean of the English sample was 109.36 points (SD=11.58), compared to the population mean which is set to 100 with a range from 88 to 131. No norms were available for the adaptation of this test created with Czech lexical items for the Czech sample (see section 4.8). English norms were used to obtain standard scores for children's performance on the Czech adaptation to gain information about the distribution of scores in the Czech sample and compare these scores with the English sample. The mean for the Czech sample was 113.66 (SD=18.52), with a range from 75 to 138 points, with five children scoring below 85 and twelve children scoring above 130. The distribution for the English sample is depicted below in Figure 5.1 and in Figure 5.2 for Czech.

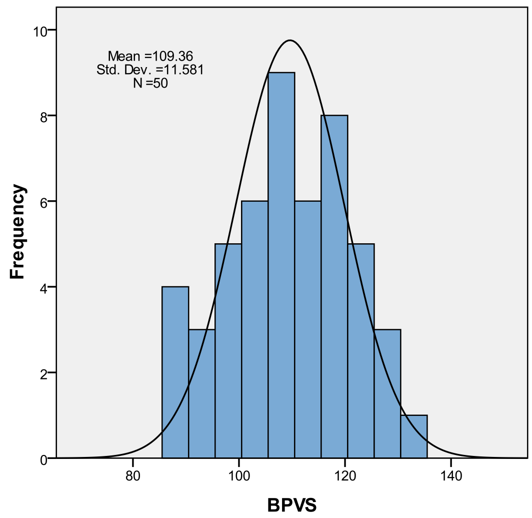


Figure 5.1 Distribution of the score for BPVS for 50 English-speaking children

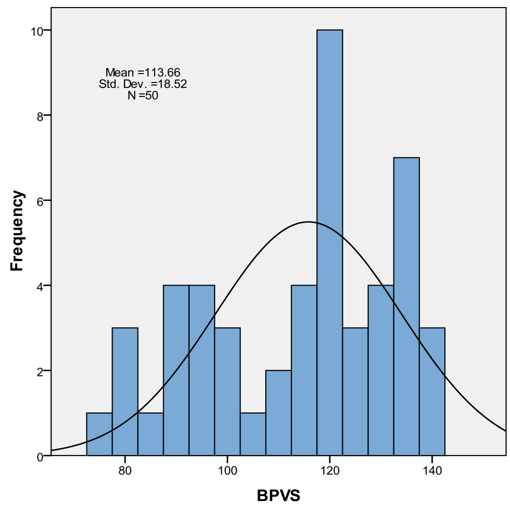


Figure 5.2 Distribution of the score for the Czech adaptation of BPVS for 50 Czech-speaking children

Both samples performed at an above average level for the age range. There was a range of scores in both languages, with the English data clustering closer to the mean and the Czech data more widely spread. Univariate Analysis of Variance with age controlled revealed no significant differences between Czech-speaking and English-speaking children on their BPVS raw scores (F(1)=2.61, ns.), and this result was replicated with the standard scores (F(1)=1.94, ns.), suggesting the samples were broadly comparable in their receptive vocabulary knowledge. The data from BPVS were also used for analysis of order effects and correlational analyses (see section 5.5.5).

5.5.2.2 Order of presentation

A repeated measures design has the advantage of controlling for multiple inter-individual confounding variables (Dancey & Reidy, 2004). However, using the same participants in several conditions might introduce order effects, but these effects can be diminished through the introduction of counterbalancing. By doing this, any practice, learning, fatigue and boredom could spread across conditions and would be less likely to become a confounding variable. In the current study, two different orders of presentation for the experimental conditions were generated and participants were randomly assigned to one of them (list A or list B), as discussed in section 4.5.

In order to establish whether any advantage found for performance on experimental conditions was a function of their order of presentation, the performance of participants from the two order groups was compared. Order effects were assessed separately for each age group and each language, yieldeding 4 groups: English-speaking children (N=50), English-speaking adults (N=20), Czech-speaking children (N=50) and Czech-speaking adults (N=20). The analysis involved multiple comparisons so the significance level was set to $p<0.006$ due to the Bonferroni correction. The independent t-test comparing scores of participants from lists A and list B revealed no significant differences between the conditions for Czech-speaking children and adults and for English-speaking adults. Among the English-speaking children, participants from group B achieved a significantly higher span than participants from group A on experimental condition A ($t(48)=-3.22$, $p<0.01$). The results were replicated with non-parametric tests: English-speaking children from group B scored higher on condition A (Mann-Whitney, $z=-2.97$, $p<0.01$), but differences in all other groups and conditions failed to reach significance.

The advantage from group B could be due to order effects after having received condition A after condition B, but could also be due to a higher overall ability of the participants from group B. Although no other comparison reached significance, there was a trend for English-speaking children from group B to outperform children from group A on all experimental conditions, regardless of the order of administration. The two groups' scores on BPVS were compared using an independent t-test. In the English sample, children in group A performed significantly lower on the receptive vocabulary test than children in group B ($t(48)=-3.92$, $p<0.001$). There was no significant difference in the Czech sample ($t(48)=-.43$, ns.). When scores on the BPVS were entered as a covariate, the significant differences between groups A and B on experimental condition A in the English-speaking children was reduced to non-significance.

It was concluded that the order of conditions could not account for the results presented in the following sections (see section 5.5.4). No differences were found in groups of Czech-speaking children and Czech- and English-speaking adults. A significant difference was only found in one condition in the group of English-speaking children, however, once the BPVS score was taken into account, the difference failed to reach significance, suggesting the effect was due to the overall higher performance of children in group B rather than order effects. As a result, the data from participants receiving list A and list B were collapsed into one file and were not analysed separately in subsequent analyses.

5.5.2.3 Age

There were three age groups for each language: 4-year-olds, 5-year-olds and adults. The age range for the English-speaking 4-year-olds was 48-59 months, mean age 54.04 (SD=2.84). This was closely matched with the Czech-speaking 4-year-olds: age range 49-59 months, mean age 54.69 (SD=2.78). The English-speaking sample of 5-year-olds ranged between 60-71 months, mean age 66.22 (SD=3.9) and for Czech-speaking 5-year-olds it was 60-71, mean age 65.33 (SD=3.92). The age range for adults was 18 to 40 years (the exact date of birth was not collected). Of primary interest was the pattern of results across all age groups in both languages shown in Figure 5.3 for English and Figure 5.4 for Czech.

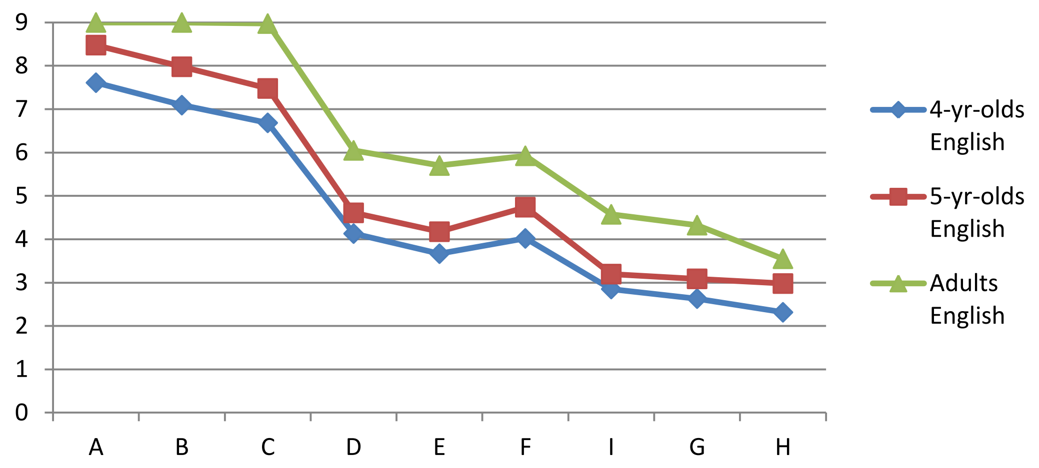


Figure 5.3 Mean span for experimental conditions A – H in the three English age groups

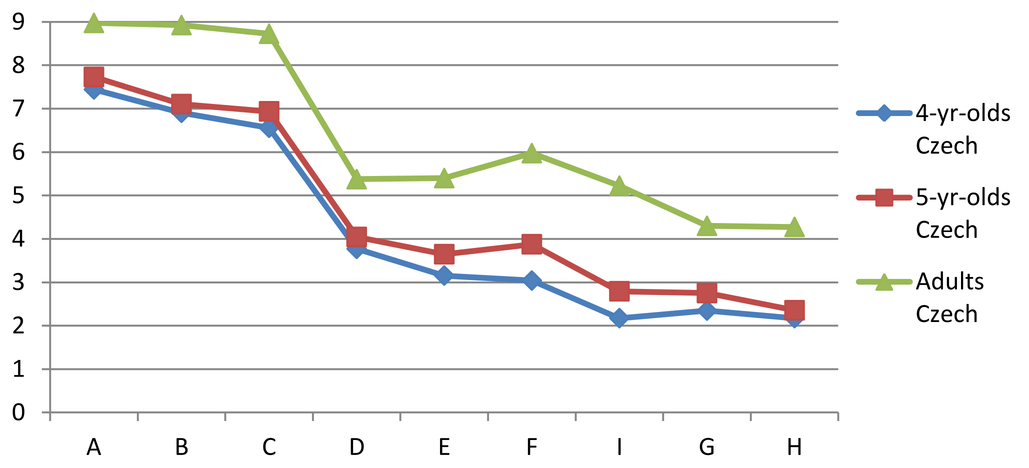


Figure 5.4 Mean span for experimental conditions A – H in the three Czech age groups

While adults gained higher scores on all conditions, Figures 5.3 and 5.4 show that the pattern of performance is strikingly similar across all age groups and in both languages. The adult results are similar to the children's in both languages, although minor differences can be seen in conditions A, B and C, possibly due to ceiling effects in the adult data (see table 5.6 for more details). In the Czech sample, the pattern is slightly different between conditions E and F, with adults and 5-year-olds performing better on F than E, while the reverse was true for 4-year-olds. The comparison between conditions E and F was not of interest in this study, however. There were also slight differences for conditions D and F in Czech: 4-year-olds and 5-year-olds performed better on D than F, but the reverse was true for adults. This difference is further analysed in section 5.5.4.7.

Data from the 5-year-old group and adults were compared using the Mann-Whitney test (non-parametric tests were preferred as the number of participants was not equal across the groups and Levene's test was significant for some conditions). The analysis involved multiple comparisons so the significance level was set to $p<0.006$ due to the Bonferroni correction. The analyses revealed that the differences between Czech adults and 5-year-olds were significantly different across all experimental conditions (see Table 5.8).

	A	B	C	D	E	F	G	H	I
Mann-Whitney U	44.00	37.00	38.50	40.50	27.00	58.00	47.00	16.50	31.00
Wilcoxon W	344.00	337.00	338.50	340.50	327.00	358.00	347.00	316.50	331.00
Z	-5.01	-5.08	-4.88	-4.77	-5.10	-4.32	-4.62	-5.34	-5.02
Asymp. Sig.	.000	.000	.000	.000	.000	.000	.000	.000	.000

Table 5.8 Czech-speaking 5-year-olds vs. adults

The differences between English-speaking adults and 5-year-olds were significant in all conditions except conditions A and H (see Table 5.9).

	A	B	C	D	E	F	G	H	I
Mann-Whitney U	170.00	90.00	64.50	64.00	60.50	100.50	59.50	147.50	75.00
Wilcoxon W	446.00	366.00	340.50	340.00	336.50	376.50	335.50	423.50	351.00
Z	-2.43	-4.10	-4.50	-4.09	-4.19	-3.19	-4.21	-2.04	-3.82
Asymp. Sig.	.015	.000	.000	.000	.000	.001	.000	.041	.000

Table 5.9 English-speaking 5-year-olds vs. adults

In contrast, a comparison between the 4-year-old and 5-year-old group failed to reach significance in any condition in English. In the Czech-speaking children, 4-year-olds did not significantly differ from 5-year-olds in any condition apart from condition F ($z=-2.85$, $p=.004$).

Since the 4- and 5-year-olds resembled each other more closely than either group resembled the adult data, these groups were combined into one file to increase the statistical power of further tests (N=50 English-speaking children and 50 Czech-speaking children).

5.5.2.4 Gender differences

There were more females than males in the sample for the reasons outlined in section 5.2. Gender differences were assessed separately for children and adults in each language which yielded 4 groups: English-speaking children (N=50), English-speaking adults (N=20), Czech-speaking children (N=50) and Czech-speaking adults (N=20). The analysis involved multiple comparisons so the significance level was set to $p<0.006$ due to the Bonferroni correction. As there were more female participants in each group, non-parametric Mann-Whitney tests were run to compare the spans for each experimental condition for males and females in each of the four groups. The tests revealed no significant differences for any of the nine experimental conditions, in any of the four groups. As a result, the data from male and female participants were not analysed separately in subsequent analyses.

5.5.2.5 Summary

An analysis of the between-subjects variables failed to reveal any order or gender effects and a remarkably similar pattern of performance for both 4-year-old and 5-year-old children emerged. As a result, the data from the following groups were collapsed and not analysed separately in further analyses: sex (M/F), age (4/5 year olds) and list order (list A/B). Four participant groups remained for further analyses: i) English-speaking children, ii) English-speaking adults, iii) Czech-speaking children and iv) Czech-speaking adults.

5.5.3 Descriptive statistics: Means and standard deviations for experimental conditions

The means and standard deviations for each experimental condition are shown in Table 5.10 and illustrated in the graphs from Figure 5.5. The maximum score possible was 9 and the minimum score was 1.

	ENG CHILDREN		CZ CHILDREN		ENG ADULTS		CZ ADULTS	
	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
A	8.01	1.31	7.58	1.08	9.00	0.00	8.98	0.11
B	7.50	1.37	7.00	1.27	9.00	0.00	8.93	0.24
C	7.05	1.45	6.74	1.14	8.98	0.11	8.73	0.55
D	4.35	1.00	3.90	0.68	6.05	0.97	5.38	0.70
E	3.90	0.81	3.39	0.83	5.70	0.89	5.40	0.85
F	4.35	1.05	3.44	0.95	5.93	1.10	5.98	1.63
I	3.01	0.87	2.47	1.01	4.58	0.96	5.23	0.80
G	2.84	0.79	2.54	0.74	4.33	0.73	4.30	0.91
H	2.62	0.89	2.26	0.56	3.55	0.86	4.28	0.83

Table 5.10 Means and standard deviations for the experimental conditions

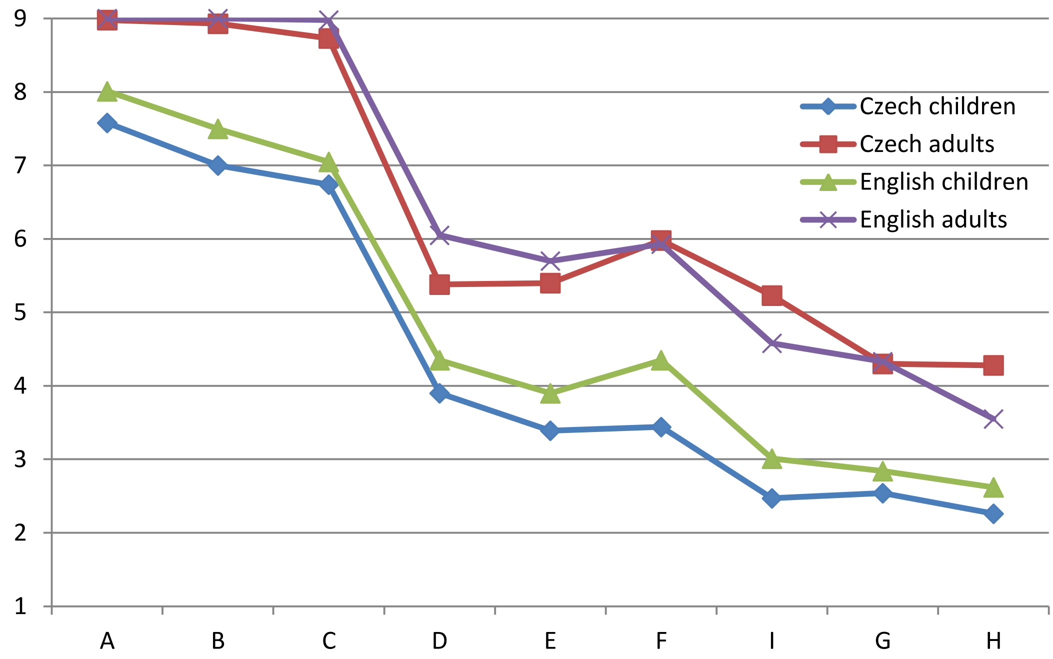


Figure 5.5 Overview of the means for each experimental condition for the four groups

As observed above, the pattern of results was broadly similar across both English and Czech groups and across children and adults, although the adults' performance was consistently better than the children's performance in both languages. Looking more closely, the English children's performance was better than the Czech children without exception, but the picture was more mixed between the two adult language groups. English- and Czech-speaking adults performed at ceiling in conditions A and B and reached similar spans in conditions F and G. English-speaking adults scored higher in conditions C, D, and E but the pattern was reversed in conditions I and H with English-speaking adults scoring lower than their Czech-speaking counterparts. The boxplots below (Figure 5.6 - Figure 5.9) show the distribution of the scores for English-speaking and Czech-speaking children and adults across the nine experimental conditions.

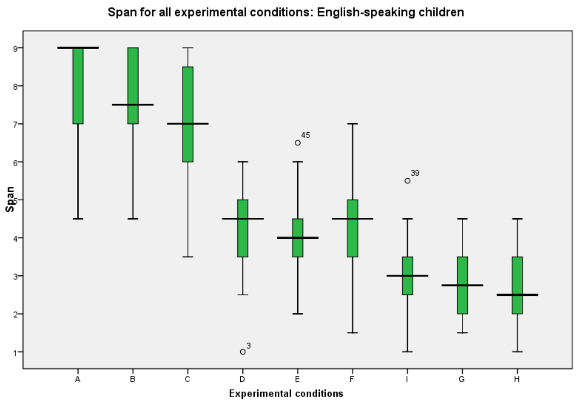


Figure 5.6 Boxplots for English-speaking children from nine experimental conditions

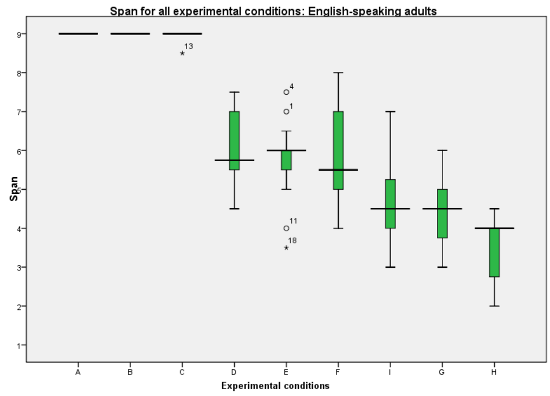


Figure 5.7 Boxplots for English-speaking adults from nine experimental conditions

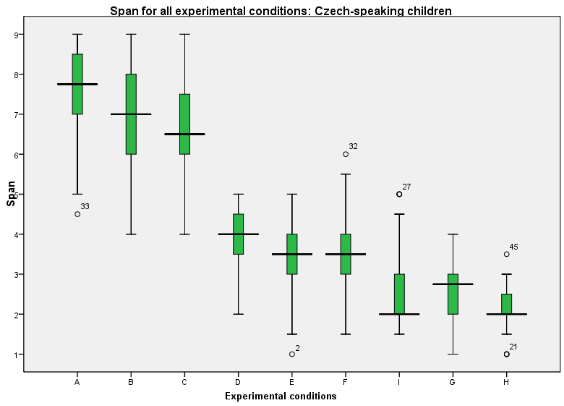


Figure 5.8 Boxplots for Czech-speaking children from nine experimental conditions

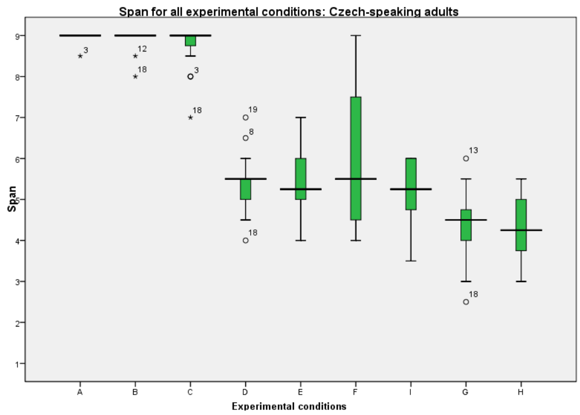


Figure 5.9 Boxplots for Czech-speaking adults from nine experimental conditions

A similar pattern of distribution was found across the four groups. There was little variability in conditions A, B and C for the adult data, possibly due to ceiling effects. The boxplots for the majority of conditions do not appear to be symmetrical, indicating an absence of a normal distribution (see Tests of Normality in Appendix I). One Czech-speaking adult (participant 18) repeatedly scored lower than other participants in conditions B, C, D and G. The outliers found in English-speaking children and adults and Czech speaking-children, however, were different participants in each of the conditions, suggesting that no single condition was responsible for extreme performance.

5.5.4 Planned comparisons of experimental conditions

5.5.4.1 Introduction

The current study used a repeated measures design, testing the same individuals on all nine experimental conditions in order to determine a maximum span for each condition. However, only planned comparisons between certain conditions were of theoretical interest (see the research questions in chapter 3). Therefore, paired-samples t-tests were carried out to make the specific comparisons raised by the research questions for each group of participants. The differences are reported to be significant at values $p<0.001$, $p<0.01$ and $p<0.05$ (see further details below). However, statistical significance does not provide full information about the size of the effect so the effect size, which quantifies the size of the difference between two groups, is also reported. Effect sizes in the current study were calculated with software G-power 3.0.10 (Faul et al., 2007) and were interpreted according to Cohen's guidelines (1969) as small, medium, or large: d of .2 = small, .5 = medium, .8 = large.

The assumption of normality was violated in some of the experimental conditions (see the distributions in Figure 5.6 - Figure 5.9). This was confirmed by the Kolmogorov-Smirnov test of normality (see Appendix I for results). Non-normality can affect the probability of making a wrong decision, either rejecting the null hypothesis when it is true (Type I error) or accepting the null hypothesis when it is false (Type II error), meaning the results of parametric t-tests can be misleading if the normality assumption is violated. However, Hubbard (1978) argued that assuming there are no marked departures from the normal distribution, parametric tests can be used. Similarly, Howell (1997) noted that '*moderate departures from normality are not usually fatal*' (in Pring, 2005, p.119). The potential problem of non-normal distributions was addressed by repeating all analyses with a non-parametric alternative to the paired-samples t-test: the non-parametric Wilcoxon signed ranks test. This test allows for a comparison of repeated measures on a single sample to assess whether the means differ and it is used when a population cannot be assumed to be normally distributed. The assumption was that if the results for the parametric tests were replicated by the non-parametric tests, this would help to demonstrate that non-normality had not affected the results. In all cases, the results from the parametric tests were replicated with non-parametric tests. It was therefore decided to report the data from the parametric analyses, with the information about the results from nonparametric tests detailed in Appendix J. It is noted whenever parametric/nonparametric analyses and online/offline scoring produced different results (for an overview see appendix K).

5.5.4.2 Research question 1

Is span for semantically plausible sentences greater than for semantically implausible sentences?

Conditions A x C

A paired-samples t-test was conducted to compare the span for semantically plausible (condition A) and semantically implausible (condition C) sentences. The results are summarised in Table 5.11.

L1	Participants	Mean A	SD	Mean C	SD	A-C mean	Df	t	Sig.	Effect size (d)
EN	children	8.01	1.31	7.05	1.45	0.96	49	5.68	< 0.001	0.8
	adults	9	0	8.98	0.11	0.02	19	1	ns.	0.18
CZ	children	7.58	1.09	6.74	1.14	0.84	49	7.16	<0.001	1.01
	adults	8.98	0.11	8.73	0.55	0.25	19	2.13	<0.05	0.47

Table 5.11 Paired-samples t-tests comparing maximum spans for conditions A and C

The Czech- and English-speaking children, along with Czech-speaking adults, achieved a longer mean span for semantically plausible sentences than for semantically implausible sentences. In contrast, the same comparison did not reveal a significant difference in the English-speaking adults. However, as can be seen in Table 5.11, participants in the English adult group performed at ceiling on both conditions. The effect size was large for the children (d=0.8 for English-speaking children and d=1.01 for Czech-speaking children). There was a small effect for the Czech-speaking adults (d=0.47).

5.5.4.3 Research question 2

Is span for grammatically well-formed pseudo/sentences greater than for grammatically ill-formed pseudosentences?

Two comparisons were conducted in order to investigate the role of syntactic structure in recall. The first comparison involved grammatical sentences vs. ungrammatical pseudosentences with a sentence prosody and the second set of tests compared the performance on grammatical vs. ungrammatical pseudosentences which had a list prosody.

Conditions A x D

The first set of paired-samples t-tests was conducted to compare the span for grammatically well-formed sentences which had a sentence prosody (condition A) and grammatically ill-formed pseudosentences with a sentence prosody (condition D). The results are summarised in Table 5.12 below.

L1	Participants	Mean A	SD	Mean D	SD	A-D mean	df	T	Sig.	Effect size (d)
EN	children	8.01	1.31	4.35	1	3.66	49	23.54	<0.001	3.33
	adults	9	0	6.05	0.97	2.95	19	13.57	<0.001	3.03
CZ	children	7.58	1.08	3.9	0.68	3.68	49	27.04	<0.001	3.83
	adults	8.98	0.11	5.38	0.7	3.6	19	23.66	<0.001	5.29

Table 5.12 Paired-samples t-tests comparing maximum spans for conditions A and D

The comparisons revealed that all groups showed a superior performance on grammatical sentences vs. ungrammatical ones. The difference was highly significant for all four groups (see Table 5.12). The effect sizes for all groups were large: d=3.03 and above. The difference in spans between grammatically well-formed and ill-formed sentences was 3-4 words, indicating that grammatical well-formedness affected span more than semantic plausibility.

Conditions B x E

A second set of paired-samples t-tests was conducted to compare the span for grammatically well-formed pseudosentences with a list prosody (condition B) and grammatically ill-formed pseudosentences with a list prosody (condition E). The results are summarised in Table 5.13.

L1	Participants	Mean B	SD	Mean E	SD	B-E mean	df	t	Sig.	Effect size (d)
EN	children	7.5	1.37	3.9	0.81	3.6	49	21.37	<0.001	3.02
	adults	9	0	5.7	0.89	3.3	19	16.5	<0.001	3.7
CZ	children	7	1.27	3.39	0.83	3.61	49	20.97	<0.001	2.96
	adults	8.93	0.24	5.4	0.85	3.52	19	20.5	<0.001	4.58

Table 5.13 Paired-samples t-tests comparing maximum spans for conditions B and E

Paired samples t-tests revealed a highly significant difference between the grammatical and ungrammatical pseudosentences in all groups. As with comparison A x D, the drop in span was about 3-4 words. Effect sizes were also large for all groups, ranging between d=2.96 for the Czech-speaking children to d=4.58 for the Czech-speaking adults. The drop in span between grammatically well-formed and ill-formed pseudo/sentences with a sentence prosody (A x D) was almost identical to the drop between grammatically well-formed and ill-formed pseudosentences with a list prosody (B x E): 3.66 vs. 3.6 for the English-speaking children, 2.95 vs. 3.3 for the English-speaking adults, 3.68 vs. 3.61 for the Czech-speaking children and 3.6 vs. 3.52 for the Czech-speaking adults.

5.5.4.4 Research question 3

Is the span for pseudo/sentences comprised of real lexical items greater than for pseudosentences made up of nonwords?

Conditions A x G, and Conditions B x H

The span for pseudo/sentences consisting of lexical items (conditions A and B) vs. the span for pseudo/sentences made up of nonwords (conditions G and H) was compared in order to investigate the role of lexical status in recall. Again, two different comparisons were made. First, grammatical sentences comprised of lexical items and with a natural sentence prosody (condition A) versus pseudosentences comprised of nonwords with a sentence prosody that matched the original sentence (condition G). Second, grammatical pseudosentences comprised of a list of lexical items with a list prosody (condition B) versus pseudosentences comprised of a list of non-lexical items (condition H). The results are summarised in Table 5.14 and Table 5.15.

L1	Participants	Mean A	SD	Mean G	SD	A-G mean	Df	t	Sig.	Effect size (d)
EN	children	8.01	1.31	2.84	0.79	5.17	49	26.49	<0.001	3.74
	adults	9	0	4.33	0.73	4.67	19	28.62	<0.001	6.4
CZ	children	7.58	1.08	2.54	0.74	5.04	49	33.36	<0.001	4.71
	adults	8.98	0.11	4.3	0.91	4.68	19	23.82	<0.001	5.32

Table 5.14 Paired-samples t-tests comparing maximum spans for conditions A and G

L1	Participants	Mean B	SD	Mean H	SD	B-H mean	Df	t	Sig.	Effect size (d)
EN	children	7.5	1.37	2.62	0.89	4.88	49	26.63	<0.001	3.77
	adults	9	0	3.55	0.86	5.45	19	28.45	<0.001	6.23
CZ	children	7	1.27	2.26	0.56	4.74	49	28.75	<0.001	4.07
	adults	8.92	0.24	4.28	0.83	4.65	19	26.67	<0.001	5.96

Table 5.15 Paired-samples t-tests comparing maximum spans for conditions B and H

A clear advantage for the lexical stimuli over the non-lexical stimuli was shown for all participants. The difference was highly significant for all four groups both in conditions with sentence prosody (A>G) and with list prosody (B>H). Effect sizes were large for all of the groups, ranging between d=3.74 for English-speaking children to d=6.4 for English-speaking adults. The drop in span length of 4-5 was almost identical for conditions with and without prosody. The effect sizes across the groups were also very similar. The size of the difference for A x G and B x H was d=3.74 vs. d=3.77 for the English-speaking children; d=6.4 vs. d=6.23 for the English-speaking adults; d=4.71 vs. d=4.07 for the Czech-speaking children; and d=5.32 vs. d=5.96 for the Czech-speaking adults.

Conditions E x H

The span for lists of words (condition E) and lists of nonwords (condition H) was compared to investigate the role of lexical status in recall of a random selection of words vs. nonwords. The results are summarised in Table 5.16.

L1	Participants	Mean E	SD	Mean H	SD	E-H mean	df	t	Sig.	Effect size (d)
EN	children	3.9	0.81	2.62	0.89	1.28	49	10.62	<0.001	1.50
	adults	5.7	0.89	3.55	0.86	2.15	19	10.46	<0.001	2.34
CZ	children	3.39	0.84	2.26	0.56	1.13	49	8.49	<0.001	1.20
	adults	5.4	0.85	4.28	0.83	1.13	19	6.79	<0.001	1.52

Table 5.16 Paired-samples t-tests comparing maximum spans for conditions E and H

A clear advantage for the lexical stimuli over the non-lexical stimuli was shown for all groups. Effect sizes were large for all groups, ranging between d=1.20 for the Czech-speaking children to d= 2.34 for the English-speaking adults. However, the drop in span length of 1-2 was much smaller than in comparisons A x G and B x H, where the drop in span was 4-5.

5.5.4.5 Research question 4

Is span for pseudo/sentences presented with sentence prosody greater than for pseudosentences presented as a list?

Conditions A x B, Conditions D x E, and Conditions G x H

Three comparisons were carried out to evaluate the effect of prosody on memory span: i) grammatical pseudo/sentences with vs. without sentence prosody (conditions A vs. B); ii) ungrammatical pseudosentences with vs. without sentence prosody (conditions D vs. E); and iii) nonlexical pseudosentences with vs. without sentence prosody (conditions G vs. H). Results for the paired-samples t-tests for A vs. B, D vs. E, and G vs. H are summarised in Table 5.17, Table 5.18 and Table 5.19[1].

L1	Participants	Mean A	SD	Mean B	SD	A-B mean	Df	t	Sig.	Effect size (d)
EN	children	8.01	1.31	7.5	1.37	0.51	49	3.35	0.05	0.47
	adults	9	0	9	0	0	19	-	ns.	-
CZ	children	7.58	1.08	7	1.27	0.58	49	4.09	0.001	0.58
	adults	8.98	0.11	8.93	0.24	0.05	19	0.81	ns.	0.18

Table 5.17 Paired-samples t-tests comparing maximum spans for conditions A and B

L1	Participants	Mean D	SD	Mean E	SD	D-E mean	Df	T	Sig.	Effect size (d)
EN	children	4.35	1	3.9	0.81	0.45	49	3.96	0.001	0.56
	adults	6.05	0.97	5.7	0.89	0.35	19	1.7	ns.	0.38
CZ	children	3.9	0.68	3.39	0.83	0.51	49	4.88	0.001	0.69
	adults	5.38	0.7	5.4	0.85	0.03	19	0.213	ns.	0.05

Table 5.18 Paired-samples t-tests comparing maximum spans for conditions D and E

[1] Although the pattern of results was identical across scoring methods (online vs. offline), the level of significance differed within the group of English-speaking children. In the comparison A x B, data from online scoring produced (t(49)=3.35, p=0.002) while data from offline produced lower level of significance (t(49)=3.67,p=0.001). A reverse pattern was found in comparisons D x E. D x E: data from online scoring (t(49)=3.96,p<0.001) and from offline scoring (t(49)=2.03, p=0.048). G x H: data from online scoring (t(49)=2.22,p=0.03) and from offline scoring (t(49)=2.87, p=0.006).

L1	Participants	Mean G	SD	Mean H	SD	G-H mean	df	T	Sig.	Effect size (d)
EN	children	2.84	0.79	2.62	0.89	0.22	49	2.22	<0.05	0.31
	adults	4.33	0.73	3.55	0.86	0.77	19	5.62	<0.001	1.26
CZ	children	2.54	0.74	2.26	0.56	0.28	49	2.85	<0.05	0.4
	adults	4.3	0.91	4.28	0.83	0.02	19	0.188	ns.	0.04

Table 5.19 Paired-samples t-tests comparing maximum spans for conditions G and H

The differences between list and sentence prosody conditions were significant in all three comparisons (A x B, D x E, and G x H), in both the English- and Czech-speaking children. However, none of the comparisons reached significance with the Czech-speaking adults. With the English-speaking adults, the span for nonlexical pseudosentences with a sentence prosody was significantly greater than the span for nonlexical pseudosentences with a list prosody (G x H). Effect sizes for these prosodic comparisons were small or medium, and were comparable across the children's groups, ranging between d=0.31 to d=0.69. The drop in span for lexical items was about half a point (A>B and D>E) and about 0.2 for nonlexical items (G>H) in the children's data, demonstrating a minimal advantage for the presence of prosody as seen in Table 5.17 - Table 5.19. The pattern for adult data was more mixed. Due to the ceiling effects in conditions A and B, no advantage for the presence of natural prosody could be detected. However, it is possible that the advantage would not be present even in the absence of ceiling effects. A comparison of conditions D and E revealed a lack of significant differences in either English-speaking or Czech-speaking adults and this was in the absence of ceiling effects, suggesting that adults were less sensitive to prosody in the span task and it provided less advantage for them than it did for the children. However, a significant difference of the effect of prosody in the English-speaking adults was found in the nonlexical conditions (G>H). The span difference in this condition was 0.77, exceeding the span differences in other comparisons involving prosody. This suggests that the English-speaking adults could take advantage of prosody with non-familiar stimuli (i.e. nonwords), but not with familiar lexical items.

5.5.4.6 Research question 5

Is recall differentially affected by lexical familiarity of content words versus function words?

The last three research questions investigated the effects of word classes and the extent that the familiarity and position of function words/content words contribute to immediate recall was evaluated. Three comparisons were carried out: conditions F x I, D x I and D x F.

Conditions F x I

The first set of paired-samples t-tests looked at the effect of the lexical status of function words and content words on memory span, when both function and content words were correctly positioned. The span for pseudosentences with real function words but nonwords instead of content words (FW + non-CW, condition F) was compared to the span for pseudosentences with real content words but nonwords instead of function words (non-FW + CW, condition I). The results are summarised in Table 5.20.

L1	Participants	Mean F	SD	Mean I	SD	F-I mean	df	t	Sig.	Effect size (d)
EN	children	4.35	1.05	3.01	0.87	1.34	49	9.18	<0.001	1.3
	adults	5.93	1.1	4.58	0.96	1.35	19	7.74	<0.001	1.73
CZ	children	3.44	0.95	2.47	1.01	0.97	49	6.99	<0.001	0.99
	adults	5.98	1.63	5.23	0.80	0.75	19	2.33	<0.05	0.52

Table 5.20 Paired-samples t-tests comparing maximum spans for conditions F and I

The span for pseudosentences with real function words and non-content words was greater than the span for pseudosentences with the reverse pattern. Paired-samples t-tests revealed significant differences in performance for both children and adults and in both languages, indicating an advantage for familiar function words. The effect size was medium (d=0.52) for the Czech-speaking adults and large for the other three groups, ranging between d=0.99 to d=1.73. The drop in mean span for English-speaking children and adults was identical (1.3 item) and was higher than the drop found with Czech-speaking participants, where it was below 1 item for both children and adults.

5.5.4.7 Research question 6

Is span differentially affected by the familiarity of form vs. the familiarity of position of function words?

Conditions D x I

The second set of tests focused on conditions where the position and familiarity of *function words* were manipulated, evaluating whether it was the position or familiarity of function words in a sentence that had a greater effect on performance. In order to answer this question, the span for ungrammatical pseudosentences with a sentence prosody, where function words are familiar but in a wrong position (condition D), was compared to the span for pseudosentences where function words are in a correct position but are unfamiliar (i.e. function words are nonwords – condition I). The results are detailed in Table 5.21.

L1	Participants	Mean D	SD	Mean I	SD	D-I mean	df	t	Sig.	Effect size (d)
EN	Children	4.35	1	3.01	0.87	1.34	49	9.6	<0.001	1.36
	Adults	6.05	0.97	4.58	0.96	1.47	19	6.39	<0.001	1.43
CZ	Children	3.9	0.68	2.47	1.01	1.43	49	11.56	<0.001	1.63
	Adults	5.38	0.7	5.22	0.8	0.15	19	0.88	ns.	0.2

Table 5.21 Paired-samples t-tests comparing maximum span for conditions D and I

Czech-speaking and English-speaking 4- and 5-year olds achieved a significantly greater span for pseudosentences with familiar function words out of their correct position (condition D) than for pseudosentences with unfamiliar function words in their correct position (condition I). A significant difference was also found in the data of English-speaking adults but not in Czech-speaking adults. The effect sizes in children and English-speaking adults were large, ranging between d=1.36 to d=1.63. The significant drop in span was around 1.4 items on average.

5.5.4.8 Research question 7

Is span differentially affected by the familiarity of form vs. the familiarity of position of content words?

Conditions D x F

The third set of t-tests focused on conditions where the familiarity and position of *content words* were manipulated to evaluate whether it was the familiarity of content words or their position in a sentence that had a greater impact on performance. In order to answer this question, the spans for ungrammatical pseudosentences with a sentence prosody where familiar content words are in a wrong position (condition D) were compared to the spans for pseudosentences where unfamiliar content words are in a correct position (condition F)[2]. The results are summarised in Table 5.22 below.

L1	Participants	Mean D	SD	Mean F	SD	D-F mean	Df	t	Sig.	Effect size (d)
EN	children	4.35	1	4.35	1.05	0	49	0	ns.	0
	adults	6.05	0.97	5.93	1.1	0.13	19	0.75	ns.	0.17
CZ	children	3.9	0.68	3.44	0.95	0.46	49	3.3	<0.01	0.47
	adults	5.38	0.7	5.98	1.63	0.6	19	1.94	ns.	0.43

Table 5.22 Paired-samples t-tests comparing maximum span for conditions D and F

Neither of the comparisons reached significance for English-speaking participants. Repeating pseudosentences with familiar content words in an incorrect position did not produce better results than repeating pseudosentences with unfamiliar content words in

[2] For Czech-speaking children, the comparisons from online and offline scoring achieved a different level of significance. The results for the comparison with online data were (t(49)=3.31, p=0.002) and with offline data (t(49)=4.35 , p<0.001).

correct positions. The same pattern was found for the Czech-speaking adults. However, a significant difference was found with the Czech-speaking children, with an advantage for pseudosentences with familiar content words in incorrect positions. The effect size for this difference was small and the drop in span was less than 0.5 item.

5.5.4.9 Summary of the planned comparisons of experimental conditions

Large effects were found across a number of comparisons. The comparisons of grammatical versus ungrammatical pseudosentences with natural prosody (A x D) or with list prosody (B x E) revealed large effects for all groups. The comparison of lexical vs. nonlexical stimuli, both with natural prosody (A x G) and with a list prosody (B x H) also showed large effects for all groups. The comparison of semantically plausible and semantically implausible sentences (A x C) yielded large effects for both groups of children. Large effects of lexicality of content versus function words (F x I) were found with both the English-speaking children and adults, and also with the Czech-speaking children. The comparison of the role of the function words' position vs. familiarity (D x I) showed a large effect for both children's groups and English-speaking adults. A large effect was also found with the English-speaking adults in the comparison of nonlexical stimuli with a natural prosody versus a list prosody (G x H).

Medium effects were found in the comparison of prosodic conditions of ungrammatical strings (D x E) for both children's groups and for grammatical strings (A x B) in the English-speaking children. There was a medium effect for Czech-speaking adults in the comparison of conditions F x I where there was greater recall of sequences with real function words and nonwords in content word slots than the reverse (nonwords in function word slots and real content words).

Small effects were found in a nonlexical prosodic comparison (G x H) in both children's groups and in a prosodic comparison in lexical grammatical conditions (A x B) with the English-speaking children. Small effects were also detected in the comparison of semantically plausible and semantically implausible sentences (A x C) with the Czech-speaking adults.

Most of the comparisons yielded the same pattern of results across the age and language groups, showing quantitative differences rather than a qualitatively different pattern. However, there were a few qualitatively different results between age and language groups. No difference was found for the English-speaking adults in the comparison of semantically plausible vs. semantically implausible sentences, while large effects were found for both children's groups and a small effect was found for the Czech-speaking adults. This lack of a difference in the English-speaking adults could be due to ceiling effects in conditions A and C. The ceiling effects in adults also prevented evaluation of the effects of prosody in grammatical conditions (A x B) in both English and Czech.

Some comparisons yielded different results across the age and language groups which could not have been due to ceiling effects. Small or medium effects were found with the children in ungrammatical strings with natural prosody vs. list prosody (D x E). Neither the English nor Czech-speaking adults benefited from prosodic organisation where no significant differences were found. In addition, the Czech-speaking adults did not benefit from prosody in the nonlexical comparison (G x H), compared to a small effect found with the Czech-speaking children and a large effect in the English-speaking adults. The comparison of the role of function words' position vs. familiarity (D x I) showed large effects for the Czech-speaking children, and for the English-speaking children and adults, but no effect was found for the Czech-speaking adults. The comparison of the role of content words' position vs. familiarity (D x F) was non-significant for both English-speaking children/adults and also Czech-speaking adults, but produced a small effect in the Czech-speaking children.

The mean differences in span for the twelve key comparisons discussed above are summarised in Table 5.23, including standard deviations and information about the significance of comparisons. These mean differences in span are illustrated in Figure 5.10.

	English children		English adults		Czech children		Czech adults	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
A-C	**0.96**	1.19	0.03	0.11	**0.84**	0.83	**0.25**	0.53
A-D	**3.66**	1.10	**2.95**	0.97	**3.68**	0.96	**3.60**	0.68
B-E	**3.60**	1.19	**3.30**	0.89	**3.61**	1.22	**3.53**	0.77
A-G	**5.17**	1.38	**4.68**	0.73	**5.04**	1.07	**4.68**	0.88
B-H	**4.88**	1.30	**5.45**	0.86	**4.74**	1.17	**4.65**	0.78
E-H	**1.28**	0.85	**2.15**	0.92	**1.13**	0.94	**1.13**	0.74
A-B	**0.51**	1.08	0.00	0.00	**0.58**	1.00	0.05	0.28
D-E	**0.45**	0.80	0.35	0.92	**0.51**	0.74	-0.03	0.53
G-H	**0.22**	0.70	**0.78**	0.62	**0.28**	0.69	0.03	0.60
F-I	**1.34**	1.03	**1.35**	0.78	**0.97**	0.98	**0.75**	1.44
F-D	0.00	1.08	-0.13	0.74	**-0.46**	0.98	0.60	1.38
D-I	**1.34**	0.99	**1.48**	1.03	**1.43**	0.87	0.15	0.76

Table 5.23 Mean differences in span for the key comparisons (differences in bold were significant)

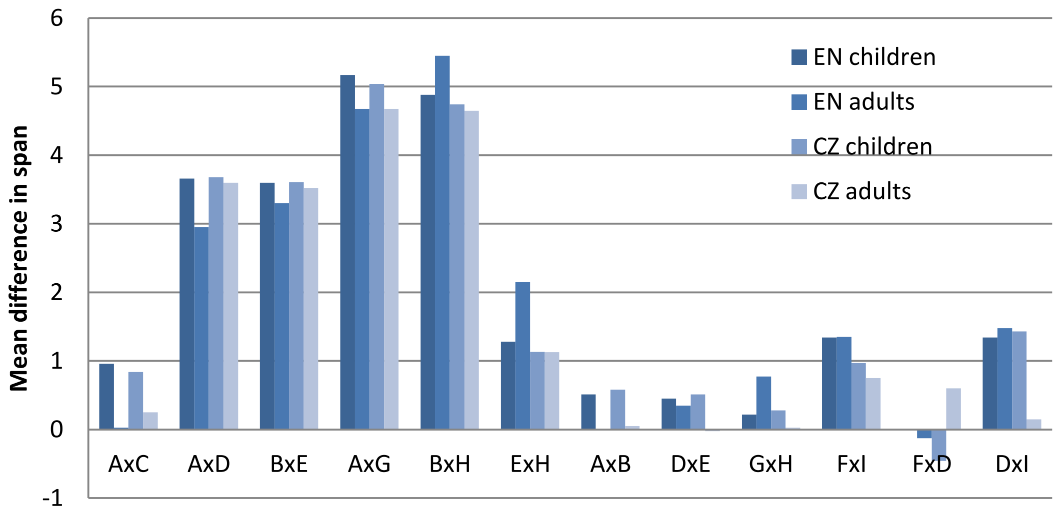


Figure 5.10 The mean differences in span between the key comparisons

5.5.5 Relation between repetition and receptive vocabulary: Research question 8

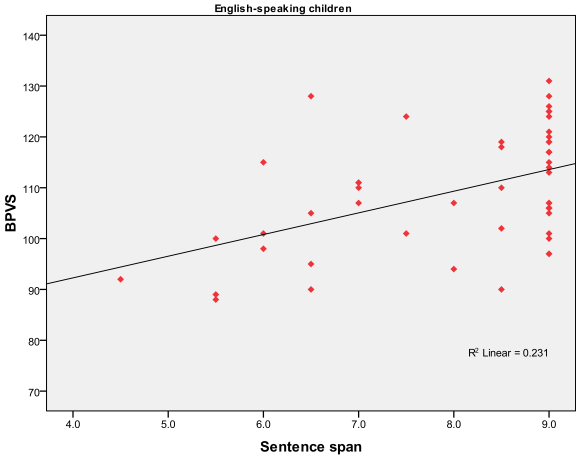
Is performance on a receptive vocabulary task related to sentence, word and nonword spans?

A set of correlational analyses were run on the data from the English- and Czech-speaking children in order to investigate the relations between receptive vocabulary and a selected set of tasks: i) sentence span (condition A), ii) word span (condition E) and iii) nonword span (condition H).

A significant relation between receptive vocabulary and nonword span and between receptive vocabulary and sentence span was found for both English and Czech (see Table 5.24 for the Pearson correlation) and the results were replicated with Spearman correlational analyses. However, vocabulary scores did not significantly correlate with word span in either language, and the pattern was the same across both languages. The significant correlations are depicted in Figure 5.11 - Figure 5.14.

L1		Sentences	List of words	List of nonwords
English	BPVS	**.481****	.172	**.431****
Czech	Adaptation of BPVS	**.386****	.207	**.349***

Table 5.24 Correlations for English- and Czech-speaking children between receptive vocabulary score and span scores on sentences, lists of words and lists of nonwords (** Significant at the 0.01 level, * significant at the 0.05 level)



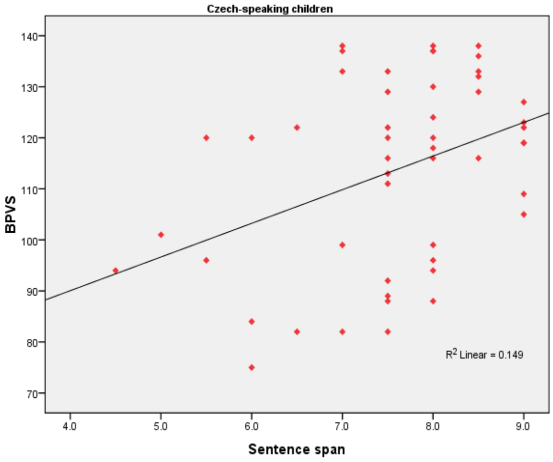
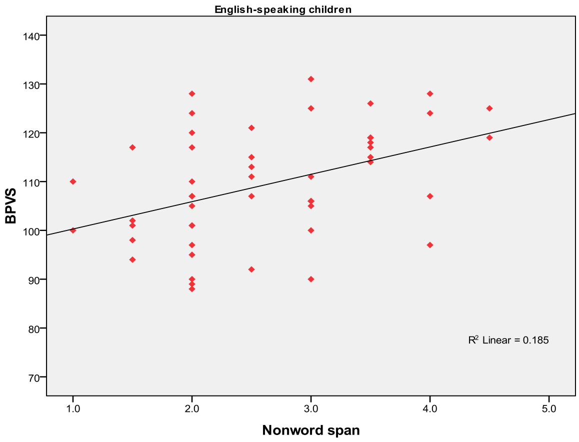


Figure 5.11 Correlations between BPVS and sentence span in English-speaking children

Figure 5.12 Correlations between adapted BPVS and sentence span in Czech-speaking children



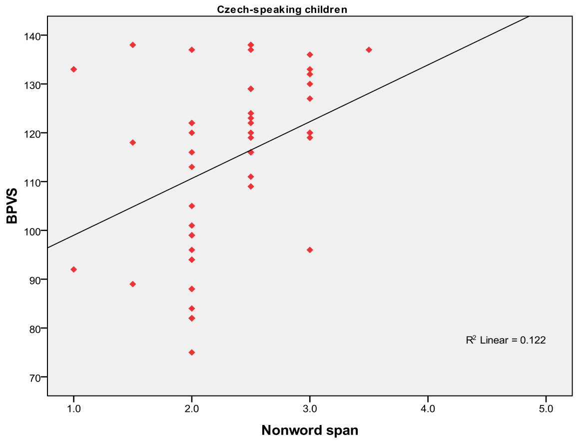


Figure 5.13 Correlations between BPVS and nonword span in English-speaking children

Figure 5.14 Correlations between adapted BPVS and nonword span in Czech-speaking children

5.5.6 Cross-linguistic comparisons: Research question 9

The results across both English and Czech have already been presented in sections 5.5.4.2 – 5.5.4.8. For children, the pattern of results was identical across the languages, apart from one comparison (condition F x D) which was significant in Czech, but not in English, as shown in Table 5.22.

Differences between the English and Czech adult groups emerged in several comparisons. Spans for semantically plausible and semantically implausible sentences significantly differed with Czech-speaking adults, while no difference was found with English-speaking adults (conditions A x C). The results are reported in Table 5.11. A different pattern was also found between spans for sentence vs. list prosody in nonlexical stimuli (conditions G x H). While there was a significant difference in English, the difference failed to reach significance in Czech. The results are reported in Table 5.19. Another cross-linguistic difference in the groups of adults was found between spans for pseudosentences with all familiar items but in incorrect positions vs. sentences with items in correct positions but with unfamiliar function words (conditions D vs. I). As sumarised in Table 5.21, the English-speaking participants benefited more from familiar forms of the content words (condition D) than from familiar positions (condition I), but the difference was not significant in Czech.

CHAPTER 6

Discussion

6.1 Introduction

This chapter reviews the experimental results and the factors that may have contributed to them. The implications of the findings for sentence repetition and the language/memory debate are discussed. The limitations of the study are explored and the merits of repetition tasks are re-examined in light of the current findings.

6.2 Main findings

6.2.1 Repetition performance as a reflection of language domains

The current study investigated the impact of the well-formedness of linguistic information on memory span and demonstrated that increasing the amount of linguistic information that was removed resulted in a corresponding drop in span. The biggest drop occurred between well-formed sentences from condition A and the nonlexical stimuli from condition G (5.17 for English-speaking children, 5.04 for Czech-speaking children, 4.68 for both English- and Czech-speaking adults). As discussed in chapter 3, the information contained in the language domains of lexicon, syntax and semantics is interconnected. A sentence from condition A, for example, carries lexical items which are combined to grammatically express the conceptual structure. Pseudosentences in condition G, on the other hand, were deprived of the lexical information which, in turn, leads to a lack of syntactic and semantic information and no opportunity to express conceptual structure. Thus, the drop between conditions A and G reflects the presence of syntactic, semantic and lexical well-formedness in A, and a lack of this well-formedness in G. Comparison B vs. H showed a similar drop in span, with the only difference being that both B and H were presented with list prosody.

The second biggest drop in span was found in comparison A vs. D between grammatically well-formed and grammatically ill-formed pseudosentences. The drop was smaller than in the previous comparison: 3.66 for English-speaking children, 3.68 for Czech-speaking children, 2.95 for English-speaking adults and 3.60 for Czech-speaking adults. Sentences in condition D were made ungrammatical by disrupting the correct word order. As pointed out in chapter 3, disrupting the syntax makes it impossible to identify grammatical roles which can also affect the semantics. Thus, the drop in span for comparison A vs. D reflects the ill-formedness of syntax and lack of interpretable semantic relations. Comparison B vs. E showed a similar drop in span, with the only difference between A vs. D, and B vs. E, being that both B and E were presented with list prosody.

The next largest drop in span involved comparison E vs. H, where lexical information was present vs. absent with no other information provided. This drop, reflecting only lexical information, was again smaller than in the previous comparisons (1.28 for English-speaking children, 1.13 for Czech-speaking children, 2.15 for English-speaking adults and 1.13 for Czech-speaking adults).

An even smaller drop was found between spans for semantically plausible and semantically implausible sentences (0.96 for English-speaking children, 0.84 for Czech-speaking children, 0.03 for English-speaking adults and 0.25 for Czech-speaking adults). This comparison involved changes in the sentences' semantics, while syntactic structure remained constant. However, different sets of lexical items were used for conditions A and C and the possible implications of this are discussed further in section 6.3.2.1.

The last domain which influenced repetition performance was prosody. Although the drops in span in the prosodic comparisons were the smallest of all of the linguistic domains, the difference between prosodic and list conditions was still significant in all of the children's comparisons, but failed to reach significance in either adult group except for the comparison G x H in English (see section 6.4.4.2 for further discussion). In all comparisons, the drop in span was smaller than 0.8 items and for the majority of comparisons it was smaller than 0.5 (see table 5.23 for an overview).

In summary, the presence of nonwords and violation of syntax led to the greatest reduction in memory span (both large effects), followed by removal of semantic plausibility (large effect) and then removal of sentence prosody (medium to small effects). These findings are in line with the results reported by Marks and Miller (1964) and Epstein (1961, 1962), discussed in detail in section 2.2.6.2.1. Familiar lexical items and well-formed syntactic structure were essential for successful immediate recall in the current study. Notably, familiarity of the lexical items and syntactic well-formedness was also found to be important to learning (i.e. the number of trials needed to correctly recall stimuli) in the above-mentioned studies. Although the effect of plausibility in the current study was statistically large, it was less important than syntax or the lexicon. Similarly, in Marks and Miller's study, syntactic well-formedness in the implausible sentences was advantageous for recall, but the advantage was much smaller compared to the benefit from well-formed syntactic structure. O'Connell et al. (1968) also found that presenting stimuli with prosody was beneficial, but that the benefit was much smaller than the gains from the presence of morphosyntactic well-formedness. The above mentioned studies used learning rather than immediate recall as a method of evaluating the contribution of specific language domains to recall performance. But the role of the specific language domains in language processing appears to be so robust that the findings could be replicated with different methods (learning vs. immediate recall), participant groups (children vs. adults) and in typologically different languages (English vs. Czech).

6.2.2 Repetition performance: familiarity of CWs vs. FWs

Spans for pseudosentences with familiar FWs and unfamiliar CWs (condition F) were compared to spans with unfamiliar FWs and familiar CWs (condition I). Participants produced longer spans for pseudosentences with familiar FWs in both languages and age groups. The advantage seems to be greater for English-speaking participants which could be due to the greater importance of FWs in English.

It seems that only knowing content words, without having indicated relations between them, is not helpful for recall, and span for familiar CWs in a potentially meaningful order but mixed together with nonwords (condition I) showed strikingly little benefit over a completely meaningless combination of nonwords (condition G):

	Benefit: EN children	Benefit: CZ children
G: /ɔɪ vɜt kaɪ ri/ I: /ɔɪ/ hurt /kaɪ/ knee.	0.17	0.1

In comparison, familiar FWs seem to aid recall with span for real FWs in correct positions mixed together with nonwords (condition F) showing greater benefit:

	Benefit: EN children	Benefit: CZ children
G: /ɔɪ vɜt kaɪ ri/ F: I /vɜt/ my /ri/.	1.51	0.9

A possible interpretation is that FWs help to identify and recall novel CWs. Note that this assistance comes in the condition where FWs are in correct positions relative to each other and potential CWs. This raises the question of whether FWs provide similar benefit when they are in the wrong position. There was no condition with the same items as F but in random order, e.g. /vɜt/ my I /ri/. However, the random position of real function words was compared with nonwords which revealed a relatively small advantage of a similar size to the advantage gained from having real function words in correct positions. When the benefits from spans with lexical items in incorrect positions (condition D) over non-lexical items (condition G) were compared, the gain was rather small:

	Benefit: EN children	Benefit: CZ children
G: /ɔɪ vɜt kaɪ ri/ D: Hurt my I knee.	1.51	1.36

For English, the gain from condition D compared to condition G was equal to the one found with having real FWs in correct positions; for Czech providing familiar lexical items was slightly more beneficial.

However, a large gain only occurred when both CWs and FWs were real lexical items and occurred in appropriate order (condition A):

	Benefit: EN children	Benefit: CZ children
G: /ɔɪ vɜt kaɪ ri/ A: I hurt my knee.	5.17	5.04

And the gain was large even when the CWs and FWs are in semantically implausible sentences (condition C):

	Benefit: EN children	Benefit: CZ children
G: /ɔɪ vɜt kaɪ ri/ C: I dug my tea.	4.21	4.2

The biggest leap was found where function words are in their expected place relative to known content words (all morphosyntactic properties satisfied), regardless of meaning relations. These relative gains are illustrated in Figure 6.1.

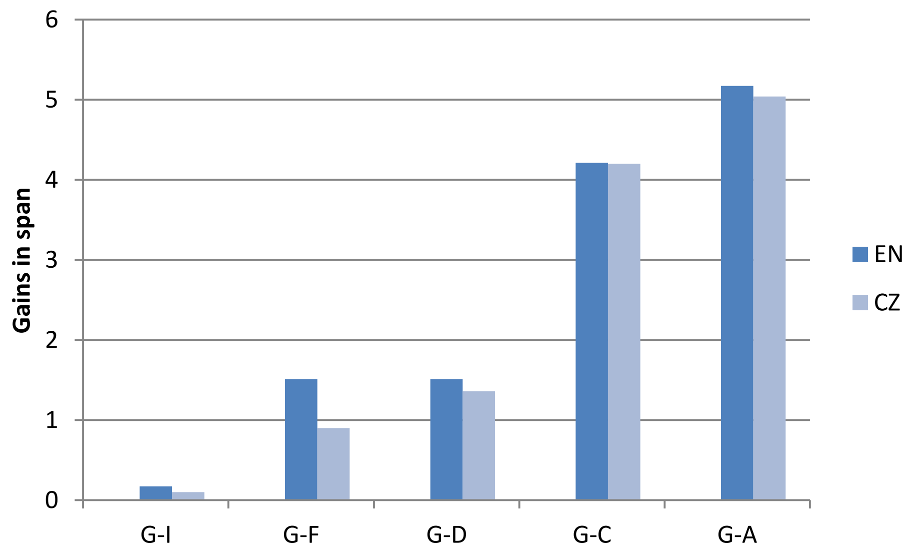


Figure 6.1 Gains in span due to familiarity and/or position of FWs/CWs

While FWs seem to be important in the early stages of language development (Christophe et al., 1997), and have also been found to support immediate repetition of pseudosentences more than CWs, FWs appear to be challenging in early production (Brown, 1973) and their omission is often seen as an indication of language impairment (see section 1.3.3.1). In a comparison of children with and without language impairment, Seeff-Gabriel et al. (2010) suggested that FWs, along with inflections, were the discriminating factors on a sentence repetition task, showing a clear disadvantage for the clinical sample. The current study showed that providing real function words was more beneficial than providing real content words. Knowing all of the lexical items, i.e. both FWs and CWs, improved performance as did having FWs in the correct position. The greatest gains were due to familiar FWs and CWs in correct positions, regardless of meaning.

Shi, Werker and Cutler (2003) reported that English-speaking infants begin to recognise function words in continuous phrases between 8 months and 13 months of age. Höhle and Weissenborn (2003) suggested that the ability to detect closed-class words might be language specific, and their study revealed that German-speaking infants are already able to recognise closed-class words at 7 to 9 months. Shady, Gerken and Jusczyk (1995) found that infants are not only sensitive to the presence of function words but also to their position within a sentence. In their study, 10-month-old infants listened longer to sequences that contained pauses at the end of a phrase rather than within the phrase (between a determiner and a noun). Infants appear to be sensitive not only to the familiarity of the form of function words, but also to their position within a sentence. This ability to recognise the form and position of function words is important for language acquisition. First, function words might help with the initial segmentation of the signal into syntactic constituents and with their identification. Second, the recognition of function words is likely to be an important first step in segmenting content words and establishing their form. The importance of the form and position of functions words in acquisition may clarify the differential findings for content and function words in this study.

Recall performance was more affected in the current study when FWs were replaced with nonwords than when CWs were replaced with nonwords. This cannot be explained by the reduced hypothesis account (see sections 1.1.1 and 1.3.4). FWs are mainly monosyllabic and shorter than CWs, so when FWs were replaced with nonwords, there was *less* novel material to recall, yet recall performance decreased more.

Nonwords replacing FWs were assigned the same phonological properties as real FWs, i.e. they were unstressed and therefore perceptually less salient. Could this account for the greater difficulty when function words were replaced with nonwords? Two findings go against this explanation. When content words were replaced with nonwords, the familiarity of function words appeared to aid recall even though they were unstressed; in contrast, when function words were misplaced, performance was severely affected even though they were familiar. It might be inferred that once the position and form of function words are established, they are extremely robust *despite* being phonologically weak: they can be recognised even when they are surrounded by unfamiliar phonology provided they are in appropriate positions in the prosodic structure. It might further be inferred that familiar function words enable the recognition of syntactic frames which in turn facilitates the processing and recall of content words. These inferences echo inferences about their role in language acquisition discussed above. Scholes (1969) found that omission rates of FWs in recall varied not only with children's age but also with the grammaticality of the strings to be recalled. Younger children omitted FWs regardless of the grammaticality of the string, but as children got older, FWs omission decreased in the grammatical sentences and disappeared altogether in the adult performance. Despite the lack of phonological salience, adults can effortlessly recall familiar FWs when they are in their expected positions. In other words, if FWs are familiar and in their expected positions their processing appears automatic and requires little attention. This interpretation is supported by a study by Haber and Schindler (1981) who reported that participants instructed to circle all misspellings in a text detected fewer misspellings in FWs than in CWs of equal length. Once acquired, the perceptual non-saliency of FWs does not seem to increase the difficulty of processing these in sentences. This is in line with inferences from findings in the present study.

6.3 Factors to be considered in the interpretation of the results

6.3.1 Possible confounding factors: Duration

Before discussing the implications of the results, potential confounding factors will be examined. Stimuli were carefully matched across the conditions for the number of words and syllables (see sections 4.2 and 4.3), but differences in duration appeared. This may have been because nonwords were less familiar to those recording the stimuli which led to slower output: e.g. Condition A 'Ice is cold' (duration 1.43 sec) vs. Condition G: /ɔɪs ʊz pɛld/ (duration 2.0 sec).

Three comparisons were used to examine the durational differences between the lexical vs. nonlexical stimuli: conditions: A x G, B x H and E x H. The durations of the stimuli in these conditions were then compared using paired samples t-tests. Applying a Bonferroni correction, the significance level was set to $p<0.02$. The comparison yielded a significant difference between A x G for English: ($t(31)=-10.46$, $p<0.001$), and Czech ($t(19)=-9.33$, $p<0.001$), with stimuli from condition A having a shorter duration than G. A smaller but significant difference was found between B x H in English: ($t(31)=-3.248$, $p=0.003$), again with the lexical stimuli being of shorter duration. However, the lexical and nonlexical stimuli in Czech did not differ significantly in duration: ($t(19)=1.65$, $p=0.115$). The final comparison between E x H failed to reach significance in English ($t(31)=-1.712$, $p=0.1$) and Czech ($t(19)=2.3$, $p=0.03$). The hybrid Conditions F and I, where half of the items were real words and the other half nonwords, did not significantly differ in duration, either in English ($t(31)=-0.08$, $p=0.93$), or in Czech ($t(19)=0.71$, $p=0.49$).

Although the lexical stimuli tended to be shorter than the nonlexical, it is unlikely that duration alone could account for the large differences between the lexical and nonlexical conditions. There was no significant difference in duration between lexical condition E and nonlexical H in either language, nor between B and H in Czech, but the recall advantage for the lexical stimuli was still present in these two comparisons. This suggests that the effect was more likely to have been primarily due to the lexical semantics which were available in the lexical conditions (but lacking in the nonlexical conditions) than to the differences in duration.

Duration could also have impacted on the prosodic comparisons, such as A x B, D x E, G x H (see section 5.5.4.5). The duration of the list conditions was necessarily longer than the natural prosody conditions as the rate of presentation was not manipulated. Stress is phonetically realized by the combination of three factors: duration, amplitude and frequency (Lieberman, 1960). In English, vowels tend to be longer in stressed syllables but reduced in unstressed syllables (Ewen & van der Hulst, 2001). Placing stress on all of the items in the list conditions (including function words) led to more stressed syllables and fewer reduced syllables than in the conditions with a sentence prosody (A, D and G), and the strings of stimuli therefore took longer to produce. The second factor which contributed to the longer duration of the list stimuli (conditions B, E and H) were pauses. In natural speech, pauses do not always appear between words, but this was the case with the list conditions where a pause followed each item. As a result, the list prosody had longer stimuli and there were significant differences in all three comparisons in English and Czech. However, as duration is an inherent part of the lexical stress which was manipulated, it can be argued that the differences found between conditions with natural vs. list prosody may be due to the presence of linguistic structures rather than due to duration alone[1].

[1] See Oberauer and Lewandowsky (2008) and Lewandowsky et al. (2009) for a discussion of the impact of duration effects on STM tasks.

The overview of the durational differences for all of the planned comparisons is presented in Table 6.1.

		English: Duration in seconds			Czech: Duration in seconds		
		Mean difference	SD	Sig.	Mean difference	SD	Sig.
RQ1	AxC	0.28	0.29	0.001	0.10	0.21	0.009
RQ2	AxD	0.33	0.35	0.001	0.16	0.18	0.000
	BxE	0.10	0.26	0.05	0.14	0.36	0.03
RQ3	AxG	0.98	0.53	0.001	0.40	0.19	0.001
	BxH	0.20	0.35	0.003	0.13	0.35	0.12
	ExH	0.12	0.36	0.1	0.16	0.30	0.03
RQ4	AxB	1.62	0.91	0.001	1.90	0.73	0.001
	DxE	1.38	0.88	0.001	1.88	0.95	0.001
	GxH	0.84	0.58	0.001	0.98	0.47	0.001
RQ5	FxI	0.01	0.40	0.93	0.03	0.16	0.49
RQ6	DxI	0.44	0.28	0.001	0.00	0.18	0.95
RQ7	DxF	0.44	0.37	0.001	0.07	0.21	0.09

Table 6.1 Durational differences for stimuli entered into planned comparisons

As is evident from the table, the marked advantage of the syntactically well-formed list of words in comparison B x E (RQ2) cannot be explained by duration. Similarly, the lexical advantage (RQ3) found in the comparisons E x H in English, and B x H in Czech, and the differences found in comparison F x I (RQ5) cannot be attributed to duration as no durational differences were present. The same applies to comparisons D x I (RQ6) and D x F (RQ7) in Czech. Where duration could have confounded the results is in comparisons A x C (RQ1) and prosodic comparisons (RQ4). The table shows that the biggest durational differences were present in the prosodic comparisons, but returning to the results from A x B, D x E, G x H (see section 5.5.4.5), the drops in span for prosody were small compared to other linguistic factors and the effect sizes were small or medium in magnitude. If duration was the critical source of the differences, then the differences should be greater for prosody than any other comparisons. In fact, the opposite was true.

Another potential confounding variable was intonational phrasing. The speaker was explicitly instructed which words/nonwords were intended to be stressed/unstressed, but the distribution of intonational phrases (IP) was not controlled. Utterances are usually structured into intonational phrases and one utterance can be assigned a variable number of these phrases. e.g.

a. $[_I$ My friend's baby hamster always looks for food in the corners of its cage$]_I$

b. $[_I$ My friend's baby hamster$]_I$ $[_I$ always looks for food in the corners of its cage$]_I$

c. $[_I$ My friend's baby hamster$]_I$ $[_I$ always looks for food$]_I$ $[_I$ in the corners of its cage$]_I$

(examples from Nespor & Vogel, 2007, p. 194)

The number of IPs is determined by a number of other non-linguistic factors: '*restructuring often occurs to yield somewhat shorter constituents, perhaps for physiological reasons having to do with breath capacity and for reasons related to the optimal chunks for linguistic processing*' (Nespor & Vogel 2007, p.194). Although the number of IPs may have varied across conditions, e.g. condition A carried three IPs: '[The little girl] [lost her doll] [at school]', but a matched sentence in condition C only carried two IPs: '[The middle sock] [brushed its eye at home]'. The distribution of IPs was not biased for any particular condition (e.g. Condition A producing more IPs than Condition C) so it is thought unlikely that this would have significantly influenced the results. Given the small effects of prosody/duration, the number of IPs is unlikely to be the source of the differences observed between the conditions. However, it would be useful to control the number of IPs in future studies to better avoid potential confounding effects.

6.3.2 Possible confounding factors: Non-experimental linguistic factors

6.3.2.1 Different sets of CWs in semantically plausible and implausible sentences

Care was taken to avoid any confounds when designing the task, but factors which could have impacted on the results are discussed here. As described in section 4.2.2.1, the content words in condition A and condition C differed. The conditions were matched for syntactic structure, number of words, syllables and phonemes and prosodic properties, but involved different lexical items. The properties bound to these items, such as frequency, familiarity, age of acquisition or imageability were not necessarily identical to the properties in condition A. Thus, a possibility remains that the significant difference in span between conditions A and C is not only due to the implausibility factor of the sentences, but could also have emerged because of the differences in the properties of the items.

The fact that the results were replicated with both the English and Czech stimuli makes it less likely that the effect emerged from the properties of individual items, an interpretation supported by Marks and Miller's (1964) study (see section 2.2.6.2.1.). In this study, the same lexical items were used in semantically plausible and anomalous sentences, and a marked difference between these two sentence types was still found (see Figure 2.4). This suggests that the differences that emerged were due to the combinations of items, rather than the properties of the items themselves. Marks and Miller's findings therefore support the claim that sentence plausibility was responsible for the observed effect on repetition performance, rather than the properties of the individual items.

6.3.2.2 Balance of CWs/FWs

Section 4.2.1.3 highlighted the importance of having a balance between content/function words. The number of CWs and FWs was matched across spans, but not for each stimulus. As a result, individual stimuli within a block of the same length did not have the same balance of FWs/CWs. Performance was only analysed as a maximum span, i.e. a span for the whole block rather than as responses to each particular stimulus.

In English, spans 2, 3, 5, 7 and 9 were balanced for FWs and CWs, so it is unlikely that participants were unable to achieve a target span just because a stimulus contained a higher number of FWs. In the remaining spans 4, 6 and 8, there was always one stimulus with a higher number of FWs. In span 4, for example, one stimulus had 3 FWs and 1 CW, two stimuli were balanced with 2FWs and 2CWs and the last one contained 1 FW and 3CWs (see section 4.2.1.3). Even if the stimulus with a higher proportion of FWs led to failure in the repetition task, this fact alone could not have solely been the source of the failure, as the discontinue criterion required at least three errors within a span. Moreover, balanced and unbalanced spans alternated, e.g. span 3 was balanced, 4 was unbalanced, 5 balanced etc. This also ensured that any imbalance between CWs and FWs could not lead to failure on its own.

The Czech stimuli were all balanced between FWs and CWs, with the exceptions of spans 6 and 9. Span 6 contained one stimulus with 4FWs and 2CWs, one with 2FWs and 4CWs and two balanced with 3FWs and 3CWs. Span 9 contained 3 stimuli with 4FWs and 5CWs and one stimulus with 6FWs and 3CWs. This last stimulus differed more from the other stimuli within span 9. In future, it may be beneficial to perform single-stimulus analysis to identify if there were any particular stimuli which were possibly responsible for lower performance.

Item based analysis was not conducted, and it is possible that there may have been errors with particular stimuli, e.g. where more function words were present. However, even if that were the case, it was unlikely to have affected the results as the number of FWs and CWs was matched across conditions on a single-stimulus basis, e.g.

condition A, length 4, example 1: '*Apples grow on trees*' 1 FW/3CW,

condition C, length 4, example 1: '*Tables grow on cars*' 1 FW/3CW,

condition D, length 4, example 1: '*Trees on grow apples*' 1 FW/3CW, Any imbalances within the blocks occurred in all of the relevant conditions, since these were matched for every stimulus. Therefore, it is thought unlikely that the FWs/CWs balance was an issue.

6.3.3 Span as a testing method: results confounded with attention failures?

Span has been widely used across different populations and materials, including: unimpaired adults (Bourassa & Besner, 1994; Hulme et al., 1991), children (Roodenrys et al., 1993; Dempster, 1978) and children and adults with brain damage (Baddeley & Wilson, 1993). Gathercole and Baddeley (1995, p.464) have suggested that children, in particular, can fail a span task for many reasons, including attentional failure. In the current study, span was used with 4- and 5-year-olds and adults. While fluctuations in attention are possible, there is no reason to assume that attention losses would only occur with certain conditions. Two list orders were employed (see section 4.5) in an attempt to reduce the likelihood of attention lapses occurring at the same places and with the same material. No deterioration in the results was noted at any particular point or with any condition in either of the different list orders, and it is thus unlikely that attention failure influenced the results to any significant degree. Although Gathercole and Baddeley (1995) expressed concerns about the reliability of span testing, both researchers have used this technique in their own studies and Pickering and Gathercole (2001) published a comprehensive testing battery of working memory tasks for children largely based on span tasks. High levels of validity and reliability were reported from these tasks, suggesting that attention failure was unlikely to have affected the stability of the results.

However, in the future, testing could be extended to exceed capacity. In the current study, testing stopped when there were three errors within a block. Following Gathercole and Baddeley's (1995) argument, testing could go one length further after the current discontinue rule applied to determine if testing stopped due to attention problems. If a child had failed because of attention failure, they could still regain attention and score successfully on a longer block confirming that the original failure was due to attention. In addition, slightly exceeding the participants' capacity could reveal different types of processing of the specific conditions, for example performance on grammatically correct but semantically implausible sentences might be affected much more by exceeding the capacity than performance on sentences that are grammatically correct and semantically plausible. Different types of errors might also emerge, e.g. more paraphrases and substitutions for normal or semantically implausible sentences and more omissions/order errors for lists of words, suggesting different strategies and/or processing.

Several findings from the current study have suggested that attention is unlikely to be a determining factor in the current study: i) as pointed out above, no order differences were found, and ii) a consistency of results was found across children and adults and across two different languages. An alternative possibility may be that attention failure was likely in certain conditions (for instance with condition D, resulting in a shorter span in this condition) and the same findings were shown in the four participant groups. This would reflect the view that attention is inextricable from linguistic factors, as advocated by MacDonald and Christiansen (2002, p.50): '*the attention, working memory, and computational capacity of a system may be unified with the system's computational processes, rather than viewed as separate independent entities*.'

6.4 Differences between Czech and English

Pattern of performance within languages was the main theoretical question, but the comparison of the patterns between the languages was of interest too. The Czech and English data were not compared quantitatively for a number of reasons. Comparing raw scores from spans from experimental conditions would not be very informative and interpreting the potential group differences would be problematic as there were many confounds between the Czech/English stimuli, such as durational differences, number of morphemes, number of syllables, lexical properties and the samples-related variables such as the return rate of the consent forms, maternal education etc. However, the confounds can be dismissed in within-language comparisons as these were based on a repeated measures design. Therefore, the patterns of results obtained from each language rather than the spans from English and Czech were compared. A few points which might have contributed to the differences found between the languages are discussed below.

In the English sample, the return rate of the consent form was much lower. The reasons why many parents did not agree are unknown. However, the return rate in the Czech sample was 90%, which possibly provided a more representative sample.

A cultural difference between the Czech and English samples is the age when children start attending school, including when they usually begin to learn to read and might have experience of language assessments. All the children in the Czech sample were attending nursery where they are not normally being formally instructed or assessed on reading or writing. The majority of the children in the English sample were attending primary school (either reception class or Year 1) where they are more likely to be assessed and formally exposed to written language.

The children's samples were closely matched for age (see section 5.5.2.3). However, as Figure 5.5 shows, the English sample scored consistently higher than the Czech sample.

This difference could be due to the sampling. Across the two languages, a test of receptive vocabulary was the only other measure which was taken. Comparisons on this measure yielded no significant differences, suggesting that the Czech- and English-speaking children (see section 5.5.2.1) had broadly comparable receptive vocabulary knowledge.

The Czech and English samples are comparable in age and vocabulary knowledge, but the participants in each group showed a difference in experience of being assessed and there was a potential bias in the samples due to the differing return rates of the consent forms. However, as the research questions involved within subject comparisons, the above mentioned factors could not influence the results of the planned comparisons (see section 5.5.4). Therefore, the cross-linguistic differences discussed in section 6.4.4 could not emerge due to sampling but rather due to language differences.

6.4.1 Differences in creating the stimuli in English and Czech

The creation of ungrammatical sentences was discussed in section 4.2.3 (English) and in 4.3.3 (Czech). The Czech ungrammatical sentences differed from the English stimuli in two aspects:

i) stimuli in spans 2 and 3 were created by altering words within a block (Czech) rather than changing the word order within a sentence (English)

ii) Czech nonsyllabic prepositions were not separated from their complements

The first difference is unlikely to have influenced the results as most Czech child participants (96%) proceeded to span 4, which was constructed in the same way as in English. Sixty percent of the Czech children achieved span 4 or higher for condition D, suggesting the different construction of the stimuli was not relevant to the results.

The second difference might have affected the results more, as these stimuli were more frequently administered than the stimuli relevant to point i). In English, all prepositions were separated from their complements which resulted in the disruption of more chunks, so it is possible that the Czech stimuli with non-syllabic prepositions were easier than the English equivalents. Therefore a closer analysis of the stimuli with unseparated non-syllabic prepositions was carried out. Non-syllabic prepositions s 'with', v 'in', z 'from' occurred zero times within span 2 and 4, once within span 3, and twice within spans 5 and 6 (spans 7-9 are not relevant here as no child participant achieved a span higher than 5). In other words, in spans 3, 5 and 6 in the Czech stimuli, chunking was less disrupted than in English. However, significant differences were found between spans in conditions D and A even in Czech, where the recall of condition D could have been easier due to smaller chunk disruption and the effect sizes were comparable across languages (d=3.33 for English-speaking children, and d=3.83 for Czech-speaking children). Together, these observations suggest that neither of the differences found in the Czech stimuli impacted on the effect of syntactic ill-formedness of immediate recall, as similar and large differences between conditions A and D were found in both languages.

The stimuli in English and Czech were matched on number of words but not on the number of syllables because the aim was to provide stimuli that reflected the properties of the tested language. There were more syllables in the Czech stimuli due to the language's greater number of inflectional affixes, (see 4.3.1.1 and Appendix G). Although the number of syllables differed across languages, the number of words and number of syllables were highly correlated within each language (English: $r=0.955$, $p<0.01$; Czech: $r=0.952$, $p<0.01$). This shows that although the word span measure was matched for number of words, it was also closely matched for syllable length: as the number of words increased, the number of syllables did so too.

6.4.2 Span in words vs. length in morphemes

Span is widely used in experimental studies and in standardised tests. It is usually defined as the number of items (e.g. digits, words, nonwords) that can be correctly recalled in serial order. The current study employed span on a sentence level and the span was defined as the number of words correctly recalled. Mukherjee (2005, p.1184) stated that '*words are regarded as basic units in the psychological reality of language acquisition, production and processing*'. However, in linguistic analyses, the notion of 'word' may capture different concepts:

- <u>Prosodic words</u>: A prosodic word contains only one main stress in English. For example, the sentence '*My brother likes bananas*' consists of 3 prosodic words. The prosodic structure of this sentence is illustrated in Figure 6.2 below:

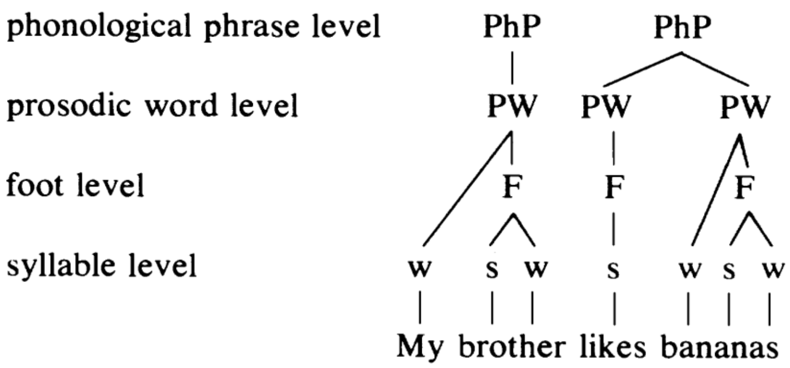


Figure 6.2 Prosodic structure of a sentence (Gerken, 1996, p.684). PhP=phonological phrase, PW=phonological word, F=foot, w=weak syllable, s=strong syllable

- <u>Syntactic words</u>: Positional freedom within the phrase or the clause and indivisibility are necessary and crucial criteria for something to be a word. Julien (2002, p.34) stated that *'with phrases it is normally the case that they can be broken up by additional words and phrases, whereas words that consist of words cannot be interrupted in this way'*. This is demonstrated by the contrast between the phrase in (1) and the compound in (2) and between the phrase in (3) and the phrasal word in (4) in the following examples taken from Julien (2002):

(1)	*a black bird*	*a black and beautiful bird*
(2)	*a blackbird*	**a black-and-beautiful-bird*
(3)	*a jack in the box*	*a jack in the little box*
(4)	*a jack-in-the-box*	**a jack-in-the-little-box*

- Lexical words: A lexical item is an abstract unit which is comprised of a connection between phonological form, semantic representation, and syntactic function. For instance, *brother* and *brothers* are the two possible forms of the lexical item BROTHER. The lexical item LIKE can be represented by any of the grammatical forms *like, likes, liked* and *liking*.

In the examples of the above three criteria, items appear to be words in some instances, but not others. Returning to the sentence '*My brother likes bananas*', there are three prosodic words, but four lexical words. In accordance with the syntactic criterion, the four syntactic words can be separated by other elements, e.g. '*My older brother really likes ripe bananas*'.

It is now clear that linguistic analyses of the items used in a span test, particularly when using items that may be syntagmatically related, require a more precise idea of what constitutes a word. One possibility is to determine span length through the number of lexical and/or grammatical morphemes, rather than through the number of words. Measurement of mean length of utterance in morphemes (MLUm) in English has been discussed by Brown (1973). An adaptation of MLUm to Slovak, a highly inflected language closely related to Czech, has been suggested by Kapalková (2003) and Kapalková et al. (2010), but MLUm has been discussed in relation to language samples, not sentence repetition. It is possible, that MLUm may provide a better reflection of ability on sentence repetition tasks and thus morphosyntactic development in languages with richer morphologies. However, counting span in MLUm would not be appropriate for the experimental task in the current design as it would be impossible to determine the number of morphemes in the nonlexical or ungrammatical sequences and therefore to match them to the normal sentence. If cross-linguistic comparisons are required, stimuli could be matched on the number and nature of lexical morphemes without having to count FWs or inflections, which will vary across languages. Language-specific properties could then be maintained for each language, while allowing for effective cross language comparisons. Future research will hopefully explore the advantages/disadvantages of this approach.

6.4.3 Word order: Scoring

As discussed in section 4.7, sequences were scored as correct or incorrect. It was necessary to preserve the order that the items had been presented in to obtain a correct score in all conditions except for the grammatically correct Czech sentences, which allow relatively freer word order (see section 4.7.1.2). Later, data were rescored using more stringent criteria which did not allow for word order changes. With the more stringent scoring, 82% (41 out of 50) of the participants achieved identical span in both scorings for Czech grammatical sentences with sentence prosody. Fourteen percent (7 out of 50), had a span half a point less and four percent (2 out of 50) had a span which was 1.5 points less. A paired samples t-test comparing spans from the lenient and stringent scoring methods revealed no significant differences ($t(49)=1.23$, $p=0.22$).

6.4.4 Cross-linguistic comparison of similarities/differences

The contribution to memory span of specific linguistic domains (syntax, lexicon, semantics and prosody) appears to be almost identical across the two languages. Nevertheless, certain comparisons yielded subtle differences between the languages. These will first be discussed within the children's data and then the adults' data. Finally, the implications of these findings and a more general discussion on how language-specific knowledge might be reflected in repetition tasks are presented.

6.4.4.1 Cross-linguistic similarities/differences in children's data

Within the children's group, the pattern of results was identical across the languages, apart from one comparison (condition F x D) which was significant in Czech, but not in English. This comparison aimed to establish if the familiarity of form vs. the familiarity of position of content words would differentially boost performance. No difference was found in the English data, but Czech-speaking children produced a longer span when all items were familiar and occupied incorrect positions, than when items were in correct positions but with nonwords in content word slots. This finding might be related to the freer word order of Czech, but the same pattern was not found with Czech-speaking adults. In order to further investigate the nature of this finding and because of the difference between 4- and 5-year olds on condition F (section 5.5.2.3), the children's group was split into 4-year-olds and 5-year-olds and the analyses addressing comparison F x D were run again. The difference remained significant in the group of 4-year-olds ($t(25)=3.88$, $p=0.001$), while it failed to reach significance in the 5-year-olds (t(23)=0.87, p=0.39). The group of 5-year-olds resembled the adult group where no significant difference was found and the results were replicated with nonparametric tests. This suggests a developmental change rather than cross-linguistic differences (developmental trends are discussed further in section 6.5). However, caution is needed in the interpretation of this finding due to the number of participants and numbers of comparisons. More research is needed to confirm these age-related differences and to address the precise impact of the cross-linguistic differences involved in these two conditions.

6.4.4.2 Cross-linguistic differences in adults' data

Differences emerged between the English and Czech adult groups in several comparisons. Spans for semantically plausible and semantically implausible sentences significantly differed with Czech-speaking adults, while no difference was found with English-speaking adults. This finding for English was likely due to ceiling effects in conditions A and C (see section 5.5.3) rather than cross-linguistic differences.

A difference was also found between spans for sentence vs. list prosody in nonlexical stimuli (conditions G x H). While there was no difference in Czech-speaking adults, English-speaking adults produced significantly shorter spans in nonlexical stimuli with list prosody (group mean was 3.55, SD=0.86) compared to sentence prosody (group mean was 4.33, SD=0.73). The majority of the stimuli in English spans 2–4 were monosyllabic items. When each item was stressed (condition H), there was no alteration between stressed and unstressed syllables, unlike the more usual iambic/trochaic patterns which would normally be found, and it may be that the stimuli's lack of familiarity made repetition more difficult (see chapter 2). In support of this, adult speakers of English appeared to be more sensitive to prosody with nonlexical stimuli (conditions G x H) than with lexical stimuli (conditions D x E). This may have been because they did not have to rely exclusively on the prosody in the lexical stimuli where there was also lexical information to take advantage of. In Czech, on the other hand, items in the stimuli were disyllabic or trisyllabic as well as monosyllabic. Even when each item was stressed, the di- and trisyllabic stimuli could still contain stressed and unstressed syllables, preserving more of the normal alteration of stressed and unstressed syllables. This may account for the lack of statistical difference in the Czech prosodic comparisons, previously illustrated in Table 6.1.

Span	English stimuli	Stress pattern	Czech stimuli	Stress pattern
2	/vi tups/	●●	Ila toří.	●○●○
	/faɪ blɔɪd/	●●	Mláli so.	●○●
	/nɪlt lɪm/	●●	Kanoz hu.	●○●
	/fɔɪz gu/	●●	Beď zim.	●●
3	/ɔɪs ʊz pɛld/	●●●	Kot me ludený.	●●●○○
	/bɔns mɪt əʃ/	●●●	Duncuj š kámi.	●○●○
	/wu rɔɪk lɪm/	●●●	Káme no bůdi.	●○●●○
	/jʌtʃ tʃə rəʊk/	●●●	Pacej kám lo.	●○●●
4	/ɛməlz drɑʊ ɒp prɔz/	●○●●●	Slotám ču ba dreje.	●○●●●○
	/tʃə sɛlɪ rɪs blaɪɪŋ/	●●○●●○	Madil pa bě sůmen.	●○●●●○
	/ɔɪ vɜt kaɪ ri/	●●●●	Dvonil tem so vylono.	●○●●●○○
	/dʊf ʊt lɪ nu/	●●●●	Pokož lo mo brůl.	●○●●●

Table 6.2 Nonlexical stimuli with a list prosody: overview of stressed and unstressed syllables in span 2-4 for English vs. Czech. ● stressed syllable, ○ unstressed syllable

Another cross-linguistic difference in the groups of adults was found between spans for pseudosentences with all familiar items but in incorrect positions vs. sentences with items in correct positions, but with unfamiliar function words. While the difference was not significant in Czech, the English-speaking participants benefited more from familiar forms of the content words (condition D) than from familiar positions (condition I). This is possibly because the unfamiliar CW in condition I in Czech still carried inflections which may have boosted performance, while in English there were fewer inflections attached to the nonlexical CW and hence fewer cues to relations between items. This comparison was not only different cross-linguistically but also developmentally (for more details see section 6.5.1).

6.4.4.3 Sentence repetition tasks: sensitivity to language-specific properties

All of the morphosyntactic devices discussed in chapter 3 can be found in languages other than Czech and English. Languages differ in how they use these devices, with some languages tending towards a richer morphology and freer word order; others a stricter word order and more limited morphology. When children are acquiring language, they have to recognise the devices their language uses. It was assumed that children would first focus on word order, even in languages which mark case (e.g. Pinker, 1982). However, Hakuta (1982) showed that Japanese children acquire word order and inflectional cues simultaneously in Japanese, a language which relies on case. Similar patterns have been found for Polish (Weist, 1983), Hungarian and Turkish (Slobin & Bever, 1982) which all use case.

This reminds of Bates et al. (1984, p.341) cautioning that: '*the bulk of our knowledge of acquisition derives from studies of children acquiring English. As a result of these limitations in the data base, we run the risk of elevating idiosyncratic facts about English to the status of language universals'*. Many studies have examined typical and atypical acquisition in languages other than English since Bates' warning. However, the English language still dominates research involving verbal imitation tasks, and little research, particularly with tasks beyond the word level, has been done on recall tasks by speakers of other languages.

Evans and Levinson (2009, p.447) critically reviewed claims about language universals and stated that '*the child's mind can learn and the adult's mind can use, with approximately equal ease, any one of this vast range of alternative systems'*. The repetition performance of speakers of different languages suggests that the diversity and availability of morphosyntactic devices are reflected in immediate verbal repetition responses. It is likely that morphosyntactic devices are retained in recall in proportion to their importance in a given language. For instance, if a language relies on word order rather than inflections to encode grammatical relations, as in English, the same recall order needs to be retained for a succesful reproduction of the model. If, on the other hand, a language primarily determines grammatical relations through inflections, order does not have to be preserved in repetition for the original grammatical roles to be retained. This was confirmed by the Czech data in the current study (see section 6.4.3 on word order changes).

Research which has taken place more recently has suggested that repetition performance reflects the properties of the participant's language (Hale et al. 1995). In their study on speakers of Warlpiri, a language with free word order and greater use of inflections than English, participants often changed the order of items that they were asked to repeat. English speakers, on the other hand, have preserved the order they were provided with, at least for predicate-argument relations and intra-phrasal relations. Changing word order in sentence repetition tasks in languages with free word order has been noted in a number of languages, e.g. Kannada (Nag, unpublished data) and Slovak (Mikulajová, unpublished data). These findings further may well support the argument that immediate verbal repetition directly reflects the language-specific properties of the stimuli.

6.5 Differences across age groups

6.5.1 Evidence from the current study

The differences between children and adults in the current study were quantitative rather than qualitative (see Figure 5.3 for English and Figure 5.4 for Czech). The pattern of results was similar across the age groups and most of the comparisons which were significant in children were also significant in adults. After eliminating the differences which emerged due to ceiling effects (see section 6.4.4.2), just one comparison in the English data produced a different result between age groups: English-speaking children benefited from sentence prosody in ungrammatical strings as opposed to lists of words (comparison D x E), but English-speaking adults showed no sensitivity to this difference. The same pattern was found in Czech with the children who showed a significant difference in comparison D x E, while the Czech adults did not. In addition, the Czech children showed a significant difference for the nonlexical prosodic comparison (conditions G x H), but the adults did not. In order to investigate the developmental trends in the prosody comparisons, the children's groups were further divided into 4-year-old and 5-year-old groups and an overview of these results is presented in Table 6.3.

	D x E		G x H	
	English	Czech	English	Czech
4-year-olds	Z=-2.44, **p=0.015**	Z=-3.18, **p=0.001**	Z=-2.17, **p=0.03**	Z=-1.06, p=0.29
5-year-olds	Z=-2.26, **p=0.024**	Z=-2.95, **p=0.003**	Z=-0.76, p=0.45	Z=-2.71, **p=0.007**
adults	Z=-1.55, p=0.122	Z=-.25, p=0.805	Z=-3.5, **p<0.001**	Z=-0.18, p=0.858

Table 6.3 Comparisons of sentence vs. list stimuli (conditions D x E and G x H) across three age groups in English and Czech

While the comparison D x E was significant in both groups of children, comparisons of nonlexical conditions with and without prosody yielded mixed results. The significant result in the English-speaking groups of children seems to be driven by the performance of the 4-year-olds, while the opposite pattern was found in Czech. Due to the mixed pattern of results, relatively small samples and number of comparisons which were not planned, conclusions are problematic.

The comparison D x F (already mentioned in section 6.4.4.1) was only significant in the group of Czech-speaking 4-year-olds, who attained a longer span in condition D than condition F. This suggests that the younger children showed a better response to familiar forms of CWs and FWs (even in incorrect positions) than to familiar FWs and unfamiliar CWs in correct positions. In this case, familiarity of form appeared to take precedence over familiarity of position. One possibility is that the syntactic structure provided by function words in condition F in English is as important for performance as the familiar form of lexical items (condition D) in the earlier stages in development, while in Czech it is more the familiarity of lexical items (longer span for condition D) which is important at these stages of development. Again, order is less fixed in Czech and the lexical items carry inflections which can play the role played in English by canonical word order and familiar FWs.

The comparison D x I was significant in both children's groups, but failed to reach significance with the Czech-speaking adults. Czech inflections were included in both conditions, but condition D also provided familiar forms of CWs, as opposed to the familiar positions in condition I. Developmentally, the familiar form might be more important than correct position, therefore children could benefit more from the information provided in condition D. As children get older, familiarity of position becomes as important as familiarity of form, producing the lack of significant difference shown in comparison D x I in the Czech-speaking adult group. This difference is only present in the Czech data and has not been found in English, possibly due to the more fixed order and lack of inflections which could have boosted the performance of the Czech adults (see section 6.4.4.2).

6.5.2 Developmental changes documented in the literature

It is well-documented that scores on span tasks (digit, word or nonword) increase with age (Dempster, 1981; Gathercole, 1998; Pickering & Gathercole, 2001; Gathercole et al., 2004; Ottem et al., 2007). This is illustrated by Figure 6.3 and Figure 6.4 below:

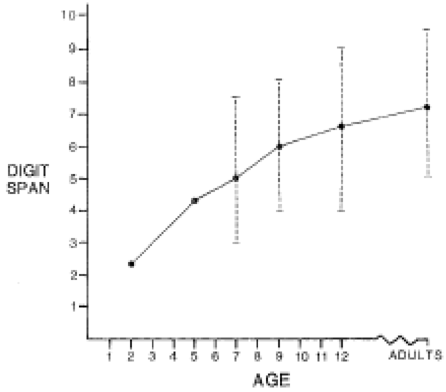


Figure 6.3 Developmental differences (solid line), and individual differences expressed as ranges (dashed lines) in digit span (from Dempster, 1981, p.66)

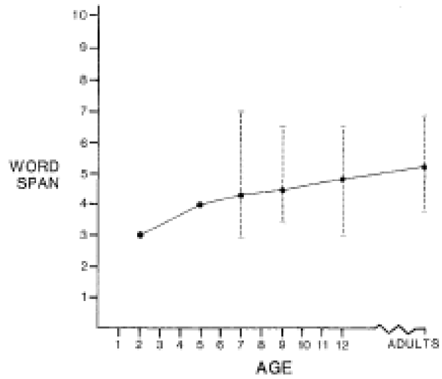


Figure 6.4 Developmental differences (solid line), and individual differences expressed as ranges (dashed lines) in word span (from Dempster, 1981, p.67)

Baddeley et al. (1998) compared data across groups of children between 3 to 13 years of age and showed that there are strong correlations between vocabulary measures/NWR and vocabulary measures/digit span. The majority of these correlations remained significant even after partialling out nonverbal IQ measured by Raven's matrices (see Table 6.4).

Age	N	Nonword repetition		Digit span	
		Simple	Partial	Simple	Partial
3.00	54	**0.34**	**0.31**	0.15	0.16
4.01	70	**0.49**	**0.47**	**0.28**	**0.21**
4.07	80	**0.56**	**0.46**	—	—
4.09	57	**0.41**	**0.41**	**0.28**	**0.29**
5.03	70	**0.34**	**0.36**	0.2	0.18
5.06	48	**0.48**	**-**	—	—
5.07	80	**0.52**	**0.5**	—	—
5.09	51	**0.41**	**0.31**	**0.38**	**0.28**
5.09	65	**0.61**	**0.53**	**0.44**	**0.38**
6.07	80	**0.56**	**0.48**	**0.44**	**0.33**
8.07	80	**0.28**	0.22	**0.36**	0.23
13.10	60	—	—	**0.49**	**0.46**

Table 6.4 Simple and partial correlations between vocabulary measures and memory tasks (NWR and digit span) across different studies, adapted from Baddeley et al. (1998, p. 159). Coefficients printed in bold are significant at $p<0.05$.

Although researchers agree that span increases with age, their explanations for this fact vary. In her article 'The Development of Memory', Gathercole (1998) and Gathercole et al. (1994), following Baddeley's model (see section 2.3.2), argued that children below 7 years do not subvocally rehearse and the increase in span below this age is a reflection of the phonological store. When individuals use rehearsal, developmental increases in memory span can be partially explained by increased rates of articulation (Gathercole, 1998; Gathercole et al., 2004). However, this is only a partial explanation. As Hulme and Roodenrys (1995) pointed out, if Gathercole's theory was true, articulating words more quickly should lead to an increase in span, but this was not the case when children were trained to do this (Hulme & Muir, 1985). Cowan et al. (2005) suggested that mnemonic or chunking strategies might contribute to the increase in span.

Ottem et al. (2007) suggested that storage capacity actually remains constant throughout childhood and increases in span are derived from improved language abilities. This was supported by their study of 100 TD Norwegian children split into two age groups: 65 children from the third grade (mean age 8.5 years, SD=0.42) and 35 children from grade seven (mean age 12.5, SD=0.88). Language (CELF screening), nonverbal IQ and memory span were assessed. When age and language scores were simultaneously entered into a regressional analysis, there was no significant effect of age on memory span, but a significant effect of language score.

6.6 Implications: Theories of verbal memory and the language-memory debate

LTM linguistic information was systematically removed from the stimuli and the effects on memory span were measured. The removal of prosodic, semantic, lexical and syntactic information directly affected repetition performance and clearly demonstrated that the ability to immediately recall verbal material is highly dependent on the linguistic characteristics of the items to-be-recalled. This suggests that it is language knowledge, rather than word/syllable length or duration, which has the greatest impact on immediate repetition performance. The same striking and robust differences between conditions were found in both languages despite the inherent differences between Czech and English. The study demonstrated that span measures on a sentence level are sensitive to all levels of linguistic structure, in particular morphosyntactic and phonological information.

Verbal STM is measured by the immediate repetition of verbal material (see section 1.3). Gathercole and Baddeley (1990) argued that reduced STM capacity can lead to language impairment, and Gathercole (2006) stressed that it typically accompanies language impairment. However, the question remains if it is possible to identify what a 'basic STM' capacity might be. As argued in section 2.2.2, the repetition of nonwords will draw on phonological and sublexical LTM knowledge, even when the linguistic information is removed. Repeating verbal material always involves language, and it is arguably contradictory to attempt to identify 'pure' memory capacity while using language stimuli. Memory always must be a memory for *something*, whether it is a string of phonemes, words or phrases. Without providing verbal stimuli, no capacity can be measured. But once the participants are asked to recall verbal stimuli, their LTM is engaged, making the quantification of 'pure' memory theoretically implausible.

The close relationship between LTM language knowledge and STM needs to be included in any theoretical model which attempts to account for the range of linguistic structures that have been shown to play a role in immediate recall. The most popular treatment of STM, Baddeley's model (see section 2.3.3), might struggle to accommodate these findings with its view of STM as being primarily phonological in nature and separate from LTM. Although the more recently added episodic buffer (Baddeley, 2000) was designed to accommodate the LTM's contribution, it is not clear i) how the information from LTM might be transferred into this buffer, ii) what kind of code is used for the storage of LTM material and iii) whether there is any limit on how much can be stored. As Postle (2006, p.23) states: '*the standard model has been a victim of its own success, and can no longer accommodate many of the empirical findings of studies that it has motivated.*'

Jackendoff (2002; 2007) proposes that WM should consist of syntactic and conceptual structures, rather than just a phonological loop. This concept of WM seems to better reflect the differences found between the experimental conditions in this study (semantic plausibility and syntactic well-formedness significantly affected recall). However, Jackendoff's theory does not make any predictions about the contribution of the different WM structures he refers to, e.g. does an increased amount of syntactic information and/or semantic plausibility improve performance and to what degree? Due to the lack of specificity on capacity limits, it is not clear how Jackendoff's theory would account for the current findings.

The theoretical account of MacDonald and Christiansen (2002) seems to better accommodate the current study's finding that LTM is tightly bound to STM performance. These researchers propose that language and memory are a single construct, with capacity not independent of knowledge. It follows that the factors underlying capacity cannot be manipulated without also affecting knowledge.

MacDonald and Christiansen (2002) treat recall tasks as 'special' language-processing tasks which require activation of phonological representations to maintain a set of verbal items, as would be the case in speech production. They suggest that the activation of phonological representations is also important for comprehension, in particular with complex structures. How much phonological information is important in processing syntax is linked to experience and less experienced comprehenders are believed to depend on the phonological information even more. In their view, sentence comprehension is '*intertwined with'* and dependent on phonological representations. Furthermore, not all sentences require involvement of phonological information to the same extent. Sentences with rare or complex syntax, sentences presented out of context, or those with conflicting cues will draw on phonological information during comprehension more than simpler sentences will. This highlights the importance of phonological information in sentence processing, not only at the lower levels of recognising lexical items, but also for the higher levels of processing, e.g. with syntactic structures.

MacDonald and Christiansen (2002, p.38) argue that success of the activation of LTM linguistic representations depends on 'biological factors' as well as experience. They claim that '*innate differences in phonological representations affect processes well beyond the representation of speech sounds, and that they have important consequences in development for the accrual of linguistic experiences.*' Newman et al. (2006, p.644) point out that despite many researchers documenting infant processing abilities and *'theorizing about how such skills would be necessary for language acquisition',* studies which experimentally examine the relation between early processing skills and later language outcomes are rare. Newman et al. (2006) showed that early processing skills have been found to relate to later language skills and similar findings were reported by Tsao et al. (2004) and Chiat and Roy (2008).

In MacDonald and Christiansen's view, verbal STM is an activation of LTM representations and it has been demonstrated that representations which are already stored in LTM can be repeated. But if short-term memory only activates what is already there, one difficulty is how to arrive at new representations in long-term memory and how the representations may be learned. Statistical learning has been offered as a general mechanism for acquiring language. Pelucchi et al. (2009) reminded us that infants can exploit numerous segmentation cues for this kind of learning, including phonotactic regularities, prosodic patterns, allophonic variation and transitional probability between syllable sequences (see also Friederici & Wessels, 1993; Mattys & Jusczyk, 2001; Jusczyk et al. 1999; Morgan, 1996; Christophe et al., 1994; Saffran et al., 1996). Hence, infants are sensitive to the acoustic-phonetic properties of speech. They are also sensitive to its context: *'The child's understanding of that context and of the connection between utterance and context are essential if she is to have some chance of discovering the meaning of the utterance'* (Chiat, 2001, p. 118-119).

How deficits in early processing skills can disrupt the mapping process which guides the acquisition of semantic and syntactic structures is discussed in detail in Chiat (2001). The mapping theory outlines what the consequences of impaired phonological processing might be for language acquisition and shows how impaired processing can lead to language impairment. It predicts that acquisition of lexical, morphological and syntactic structures will be differentially affected according to the phonological complexity of the forms. The current study's finding on CWs and FWs are in line with the mapping theory's predictions that phonologically strong forms (CWs) should be more successfully retained than weak forms (FWs). Returning to MacDonald and Christiansen (2002), the best theoretical account for the current data so far, the innate differences in phonological representations they suggest would also be likely to affect language learning, a proposition in line with the predictions of Chiat's mapping theory (2001).

6.7 Limitations of the study and future research

6.7.1 Results in relation to other memory and language tasks

A test of receptive vocabulary was the only other measure employed in the current study apart from the experimental task (see section 4.1.3 for the rationale behind the inclusion of the vocabulary task). Scores on this test highly correlated with performance on sentence repetition (condition A) and lists of nonwords (condition H), but test scores were not found to be significantly correlated to recall of lists of words (condition E). The findings were identical for English and Czech and were in line with the studies presented in section

1.3.2.1, showing a significant relation between nonword repetition and vocabulary development. The current results from English and Czech showed that both sentence repetition and nonword repetition better reflect vocabulary differences than word span.

No other measures reflecting language, STM/WM or cognitive ability were included. The participants were typically developing children and adults. Participants with language, cognitive or developmental impairments were not included (background information obtained from teachers and parental questionnaires, see Appendix M). Therefore, testing for nonverbal IQ was not considered necessary. For the same reason, comprehensive language measures were not used to investigate language profiles in detail. The aim was to compare conditions to shed light on typical language processing/memory, rather than investigate how the conditions related to language abilities.

Another option for future research would be to include traditional measures of STM, such as digit span, word span and NWR. Two of these measures were included as experimental conditions: condition E was a list of words and condition H was a list of nonwords. Although on the surface these tasks are similar to classic verbal lexical and nonlexical span tasks, two differences should be pointed out:

i) Word span tasks are usually based on one category, e.g. lists of nouns, adjectives or content words. Condition E from the current experiment, on the other hand, included words within each list which were mixed and not limited to one category so any list included at least nouns, verbs and function words: '*seen, I, an, have, angel*'

ii) Items were not equal in their number of syllables, e.g. there were monosyllabic and disyllabic (trisyllabic in Czech) words within one list, e.g. '*the, crying, fairy, was*'

Thus, it is not possible to say how performance on the experimental conditions relates to performance on more traditional word and nonword span tasks, such as those from the WMTB-C.

Most importantly, the research questions of the current study did not set out to address any possible relations between STM measures and the experimental task so there was no obvious reason to include other measures. But it may be interesting to see how performance in different conditions relates to language abilities, and in extreme cases of language acquisition disruption such as SLI, whether relations between conditions of the experimental task and language measures for such populations may differ from patterns found in typical populations. It is hoped that future research may address how children with SLI would perform on the experimental task. Future studies could also address the question of whether performance of children with SLI would differ only quantitatively or also qualitatively. In other words, would they score lower on all conditions (i.e. showing the same pattern, just quantitatively different) or would they not be sensitive to the linguistic characteristic manipulated in the experimental conditions and therefore show a qualitatively different pattern (e.g. score at span 4 on all lexical conditions, regardless of their well-formedness)?

6.7.2 Error analyses

The current study was designed as a span task, which does not lend itself to error analysis. The aim is to determine maximum span in each condition starting from likely span and increasing or decreasing the length of targets according to performance on the starting span. Consequently, different participants might receive different numbers and different lengths of stimuli, and all participants received different lengths in different conditions, so that opportunities for numbers and types of error differed between conditions and participants. It would be interesting to know what types of errors are made, e.g. if FWs or inflections are more vulnerable, and what type of error is most common, e.g. omission, substitution, order errors. It would also be interesting to know if the types of errors differed across conditions, e.g. i) if lexical substitutions are more likely to occur in semantically implausible sentences (condition C) than in semantically plausible grammatical (condition A) and ungrammatical sentences (condition D), or ii) if omissions are more likely in ungrammatical sentences (condition D) than in grammatical sentences (condition A). In the case of the nonlexical responses, there may be differences in how often participants lexicalised the nonwords and whether there was a greater tendency to lexicalise FWs or CWs. The different theoretically motivated conditions in this study open up many potentially informative questions for error analysis. These questions could be addressed systematically using a different methodology which targeted and scored specific elements, rather than span.

6.7.3 Immediate vs. delayed recall

This study only investigated immediate recall. As pointed out in section 2.2.6.2, many studies of verbal recall involve delay. Sentence repetition can be via immediate recall (as in the current study) or delayed recall. In addition, the delayed recall could contain a filler/distractor or just be filled with silence. Whether the tasks rely on the same skills in all three possibilities (Immediate, Delayed-Silent, Delayed-Filler) is unclear.

Potter and Lombardi (1990; 1998) and Lombardi and Potter (1992) explored delayed sentence recall and argued that performance is based on a propositional-conceptual structure which is generated during comprehension of the sentence. The sentence is thought to be regenerated via mechanisms involved in normal sentence production, instead of retaining the sentence as a surface string of words. Two priming mechanisms are posited which add to the regeneration process. First, the syntactic structure is primed while processing the sentence. Second, the entries in the mental lexicon that have been part of the presented sentence are activated and therefore have a higher probability of being selected during the repetition. Thus, a propositional representation and syntactic/lexical priming provide a basis for verbatim sentence recall without assuming surface short-term storage of phonological representations. Key here is that phonological information is unlikely to be present when a distractor is included and disrupts the phonological trace. This suggests that phonological information is not always necessary for successful recall.

The findings of Potter and Lombardi (1990) have been challenged by Rummer and Engelkamp (2003) who stressed that the amount of surface phonological information involved in sentence recall depends on the delay between sentence presentation and recall. What influences the availability of phonological STM representation is the interval between presentation and sentence recall. Rummer and Engelkamp conducted an experiment comparing adult performance on immediate vs. delayed recall both with and without a distractor. Sentence recall was more accurate for the immediate than the delayed conditions. The results suggested that the phonological information is encoded in both the delayed and the immediate condition, but is only available if there is no distractor.

Together, these findings suggest that phonological information is retained in immediate recall and also in delayed recall when no extra task is present. Since the current study's findings come from immediate recall, it is not possible to comment on how the different conditions would affect span/errors in delayed recall. Once anything is inserted between the model sentence and recall, the surface phonological information seems to disappear and recall is based more on the reconstruction of meaning and priming of specific syntactic structures and lexical items. Future research could address these issues by presenting the same conditions and same material for both immediate and delayed recall, with and without a distractor. This would reveal whether immediate and delayed recall produce the same results within/between conditions, and error analyses would show if the errors across the immediate and delayed recall are similar, providing more evidence of processes involved in immediate/delayed recall. Theoretically, the distinction between immediate/delayed recall is overlooked in standard memory models and more empirical findings could inform the theories and improve the models.

Another issue is whether immediate or delayed recall better captures language abilities. A possible clue is provided by a study conducted by McDade and colleagues (1982) who examined a group of six 4-year-old children. In a first experiment, the timing of recall and the filler between sentence presentation and recall was varied. Children's comprehension of the sentences was tested via a picture-pointing task. When children did not understand the sentences, they performed significantly better in immediate recall condition than on delayed recall conditions with a distractor. When only correctly understood sentences were examined, the differences between immediate and delayed recall were not significant. These findings suggest that when a sentence is understood, repetition performance can draw on syntax, semantics as well as phonological surface forms and the presence of a distractor does not interfere with performance. However, without understanding the sentence, recall relies on phonological representations more than when understanding is present. The distractor might interfere with the phonological trace and so significantly impact on the performance. This is in line with above discussion highlighting the role of phonology in immediate recall, and suggests that phonological representations might play a crucial role in recall, especially in recalling sentences which are beyond the comprehension of the participants (e.g. complex relative clauses) or when language knowledge is impaired. Similar suggestions have been made by Wells et al. (2009), MacDonald and Christiansen (2002) and Acheson and MacDonald (2011). This idea also points to the importance of phonological processing and its role in language acquisition as proposed by Chiat (2001).

A second experiment compared data from spontaneous production using *Developmental Sentence Scoring (DSS)* guidelines (Lee, 1974), and the *Carrow Elicited Language Inventory (CELI)* in nine children aged 4;4-5;1. The CELI was administered in three different conditions: immediate recall, recall with a 3-second delay and recall with a 5-second delay. The retention interval was unfilled. The correlations between DSS and the CELI were significant in all conditions: immediate (r=.77), 3-second delay (r=.90) and 5-second delay (r=.78). Children produced significantly fewer errors in immediate repetition than with either the 3-second or 5-second delay, with no difference between the delayed conditions. It was stressed that elicited imitation reflects language knowledge best when it is presented with a 3-second delay (indicated by the highest correlation). As immediate recall produced a significantly greater number of accurate responses than delayed recall, McDade and colleagues suggested that immediate recall might slightly overestimate language knowledge. As pointed out in chapter 3, spontaneous production fundamentally differs from repetition tasks in the fact that the speaker has to move from meaning (conceptual structure) to structure in spontaneous production, while in immediate recall the meaning is already provided in the model sentence. As the findings of the current study showed, immediate recall relies mostly on morphosyntax and phonology so it may not be a reflection of language processing as a whole. It is worth noting that the correlation between scores on immediate recall and DSS was also very high (r=.77), suggesting a good reflection of language scores obtained through DSS. A larger scale study which employs a variety of language measures and verbal repetition in immediate and delayed conditions may reveal more.

6.7.4 Limitations in the stimuli and conditions

As pointed out in section 6.3.1, duration could have confounded the results. Table 6.1 showed that the durational differences between the conditions were very small, with the exception of the prosodic comparisons. Duration could have been controlled more tightly by ensuring the speakers who recorded the stimuli were more familiar with the ungrammatical, semantically implausible and nonlexical conditions, or by acoustically manipulating the stimuli. Even the stimuli in prosodic comparisons could have been matched for duration, but then the list conditions may have been produced at a faster rate than the sentence-like prosodic conditions. As pointed out in 6.3.1, duration is an inherent part of prosody and therefore impossible to manipulate independently.

Despite carefully considering and controlling the stimuli, a few errors were noted after testing, but these were infrequent and unlikely to have affected the results as they were within spans that participants were unlikely to have received (see Appendix C for a full list).

As discussed in 6.2.2, it would have been useful to compare condition F, where FWs were familiar and in correct positions, to a condition where FWs are familiar but in wrong positions. Future research could compare condition F vs. a new condition: list of real FWs + nonlexical CWs. The same questions and comparisons would apply to CWs, condition I vs. a new condition: list of real CWs + nonlexical FWs. These further conditions could clarify whether it is position or familiarity which is of greater importance in recall.

6.8 Repetition tasks re-examined

6.8.1 Potential for assessment

Adlof et al. (2010, p.341) pointed out that SR: '*may in some ways be considered an all-purpose measure for predicting both language and reading performance, offering fairly accurate prediction with a single measure'*. This optimistic statement was followed by a reminder that test developers do not consider it useful enough to include in assessments, with the reason offered that sentence repetition does not reveal specific weaknesses nor offer guidance on selecting the right intervention. However, it is proposed here that repetition tasks can provide information on specific strengths and weaknesses when the stimuli are carefully constructed. If the stimuli include a full range of function words and inflections representative of a participant's language and an appropriate range of syntactic structures deploying these, the scores on a repetition task are likely to be an adequate indication of morphosyntactic development. SR offers the possibility of including structures which might be difficult to elicit or too lengthy and tedious to observe in spontaneous speech. In addition, knowing what a child is targeting can be particularly helpful with children whose speech is difficult to understand (Seeff-Gabriel et al. 2010).

A standardised test developed by Seeff-Gabriel et al. (2008) highlighted that SR mainly revealed weaknesses with function words and inflections within clinic groups. Together with the findings from the current study on FWs and CWs, it is clear that familiar FWs in correct positions are crucial for successful repetition performance and that FWs help to bootstrap CWs. It may be inferred that familiar FWs will help with segmenting and identifying unfamiliar CWs and also facilitate the learning of new CWs. Conversely, lack of familiarity of FWs will have an impact on the repetition of CWs. If SR reveals difficulties with repetition of FWs and inflections, it suggests that a child will also have greater difficulty with CWs.

However, SR does not provide a full profile of children's language. For example, the current results suggest that SR does not effectively assess or provide indicators for intervention in cases of pragmatic deficit. It is also not entirely clear to what extent children's comprehension is assessed in immediate recall tasks. It has been reported that children can repeat sentence without understanding them in immediate recall but not when delayed recall was used (McDade, 1982). This suggests that delayed recall has more potential than immediate recall for assessing comprehension. Future investigation is needed to explore the processes involved in delayed recall and its potential for assessment.

6.8.2 Conclusions

It is proposed that immediate verbal recall tasks rely on LTM. Based on the participants' knowledge, the linguistic representations of repetition tasks' stimuli are activated from representations for phonology and also sublexical, lexical, semantic, syntactic and prosodic structures. Differences between speakers are due to differences in their LTM knowledge which, in turn, varies according to their experience with language and the precision of their representations. Better specified representations (e.g. early acquired words/structures and high-frequency items) lead to more effective repetition. Thus, immediate verbal repetition is viewed as part of a process where linguistic representations are activated on the basis of LTM knowledge. An underspecified, unavailable or impaired linguistic representation will be reflected in repetition performance. The impoverished representation could be due to experience, biological factors (or even an interaction between the two), and can be seen in examples including limited exposure in L2 acquisition, language not having developed yet in infants or speech and language impairments.

It appears likely that stimuli which offer more potential to draw on LTM knowledge (morphological, lexical, syntactic and prosodic) will yield better discrimination between normal and impaired/underdeveloped language systems. The more dependent a recall test is on LTM knowledge, the greater the disadvantage may be for speakers with an impaired/underdeveloped language system, clearly revealing their disadvantage. Conversely, participants who can draw on LTM knowledge will perform better on items that are more dependent on LTM knowledge. It is hoped that future studies will be assisted by the suggestions for further research and novel findings of this thesis.

APPENDICES

A. Appendix: The full set of stimuli in English

Block	Label	Stimulus
2	2A1	She cooks.
	2A2	They played.
	2A3	Help him.
	2A4	Boys do.
3	3A1	Ice is cold.
	3A2	Dance with us.
	3A3	We like him.
	3A4	Watch the race.
4	4A1	Apples grow on trees.
	4A2	The fairy was crying.
	4A3	I hurt my knee.
	4A4	Give it to me.
5	5A1	The teacher read a story.
	5A2	I have seen an angel.
	5A3	The red bus was late.
	5A4	She sent us a letter.
6	6A1	John eats rice with his meat.
	6A2	The picture was full of colours.
	6A3	He threw a stone at me.
	6A4	Come out of the dirty water.
7	7A1	The snake is hiding in the grass.
	7A2	The clown did tricks with a monkey.
	7A3	My friend will buy the new book.
	7A4	It has been snowing for one week.
8	8A1	The little girl lost her doll at school.
	8A2	They have waited there for a long time.
	8A3	The train only stopped in the thick fog.
	8A4	I met my old aunt at her farm.
9	9A1	The white cat was chasing him in the park.
	9A2	We can see the stars on a clear night.
	9A3	The young dancer was looking at the shiny mirror.
	9A4	I have been blowing pretty bubbles the whole day.

Block	Label	Stimulus
2	2B1	she, cooks
	2B2	they, played
	2B3	help, him
	2B4	boys, do
3	3B1	ice, is, cold
	3B2	dance, with, us
	3B3	we, like, him
	3B4	watch, the, race
4	4B1	apples, grow, on, trees
	4B2	the, fairy, was, crying
	4B3	I, hurt, my, knee
	4B5	give, it, to, me
5	5B1	the, teacher, read, a, story
	5B2	I, have, seen, an, angel
	5B3	the, red, bus, was, late
	5B4	he, sent, us, a, letter
6	6B1	John, eats, rice, with, his, meat
	6B2	the, picture, was, full, of, colours
	6B3	he, threw, a, stone, at, me
	6B4	come, out, of, the, dirty, water
7	7B1	the, snake, is, hiding, in, the, grass
	7B2	the, clown, did, tricks, with, a, monkey
	7B3	my, friend, will, buy, the, new, book
	7B4	it, has, been, snowing, for, one, week
8	8B1	the, little, girl, lost, her, doll, at, school
	8B2	they, have, waited, there, for, a, long, time
	8B3	the, train, only, stopped, in, the, thick, fog
	8B4	I, met, my, old, aunt, at, her, farm
9	9B1	the, white, cat, was, chasing, him, in, the, park
	9B2	we, can, see, the, stars, on, a, clear, night
	9B3	the, young, dancer, was, looking, at, the, shiny, mirror
	9B4	I, have, been, blowing, pretty, bubbles, the, whole, day

Block	Label	Stimulus
2	2C1	She snows.
	2C2	They rained.
	2C3	Mix him.
	2C4	Beans do.
3	3C1	Ice is sad.
	3C2	Rain with us.
	3C3	We bake him.
	3C4	Wash the voice.
4	4C1	Tables grow on cars.
	4C2	The cherry was cooking.
	4C3	I dug my tea.
	4C4	Break it to me.
5	5C1	The driver bit a tower.
	5C2	I have read an uncle.
	5C3	The red grass was brave.
	5C4	He sang us a kettle.
6	6C1	John feels noise with his seat.
	6C2	The flower was sick of ladders.
	6C3	He flew a ring at me.
	6C4	Walk out of the fluffy candle.
7	7C1	The snail is laughing in the sky.
	7C2	The bike got bread with a donkey.
	7C3	My frog will toast the short book.
	7C4	It has been sneezing for all time.
8	8C1	The middle sock brushed its eye at home.
	8C2	They have hated there for a sweet salt.
	8C3	The plane only swam in the thick road.
	8C4	I wrote my blue child at her room.
9	9C1	The green pig was washing him in the fork.
	9C2	We can hear the skies on a pink light.
	9C3	The wet pencil was resting at the sunny hammer.
	9C4	I have been fighting spotty needles the whole boy.

Block	Label	Stimulus
2	2D1	Cooks she.
	2D2	Played they.
	2D3	Him help.
	2D4	Do boys.
3	3D1	Is cold ice.
	3D2	Us with dance.
	3D3	Like we him.
	3D4	Race the watch.
4	4D1	Trees on grow apples.
	4D2	The crying fairy was.
	4D3	Hurt my I knee.
	4D4	Me to it give.
5	5D1	Story a read the teacher.
	5D2	Seen I an have angel.
	5D3	Bus the was late red.
	5D4	A sent he letter us.
6	6D1	Rice with eats his John meat.
	6D2	The full was colours of picture.
	6D3	A threw me at stone he.
	6D4	Of dirty out water the come.
7	7D1	In snake the grass is the hiding.
	7D2	A tricks the with clown did monkey.
	7D3	Friend will new the buy book my.
	7D4	Been has one it week for snowing.
8	8D1	School the girl lost doll her little at.
	8D2	For time have they long waited a there.
	8D3	Fog the stopped the in thick train only.
	8D4	I old her met farm my at aunt.
9	9D1	Park in chasing the him cat the was white.
	9D2	Night see can the a clear on stars we.
	9D3	Dancer the was shiny young the at mirror looking.
	9D4	Bubbles the pretty have day whole been I blowing.

Block	Label	Stimulus
2	2E1	cooks, she
	2E2	played, they
	2E3	him, help
	2E4	do, boys
3	3E1	is, cold, ice
	3E2	us, with, dance
	3E3	like, we, him
	3E4	race, the, watch
4	4E1	trees, on, grow, apples
	4E2	the, crying, fairy, was
	4E3	hurt, my, I, knee
	4E4	me, to, it, give
5	5E1	story, a, read, the, teacher
	5E2	seen, I, an, have, angel
	5E3	bus, the, was, late, red
	5E4	a, sent, he, letter, us
6	6E1	rice, with, eats, his, John, meat
	6E2	the, full, was, colours, of, picture
	6E3	a, threw, me, at, stone, he
	6E4	of, dirty, out, water, the, come
7	7E1	in, snake, the, grass, is, the, hiding
	7E2	a, tricks, the, with, clown, did, monkey
	7E3	friend, will, new, the, buy, book, my
	7E4	been, has, one, it, week, for, snowing
8	8E1	school, the, girl, lost, doll, the, little, at
	8E2	for, time, have, they, long, waited, a, there
	8E3	fog, the, stopped, the, in, thick, train, only
	8E4	I, old, her, met, farm, my, at, aunt
9	9E1	park, in, chasing, the, him, cat, the, was, white
	9E2	night, see, can, the, a, clear, on, stars, we
	9E3	dancer, the, was, shiny, young, the, at, mirror, looking
	9E4	bubbles, the, pretty, have, day, whole, been, I, blowing

Block	Label	Stimulus
2	2F1	She /tups/.
	2F2	They /blɔɪd/.
	2F3	/nɪlt/ him.
	2F4	/fɔɪz/ do.
3	3F1	/ɔɪs/ is /pɛld/.
	3F2	/bɔns/ with us.
	3F3	We /rɔɪk/ him.
	3F4	/jʌtʃ/ the /rəʊk/.
4	4F1	/ɛməlz/ /drɑʊ/ on /prɔz/.
	4F2	The /sɛlɪ/ was /blaɪɪŋ/.
	4F3	I /vɜt/ my /ri/.
	4F4	/dʊf/ it to me.
5	5F1	The /mitʃə/ /kɛd/ a /spəlɪ/.
	5F2	I have /gin/ an /ɛmbəl/.
	5F3	The /lup/ /tʌs/ was /dut/.
	5F4	She /fɪnt/ us a /lɔpə/.
6	6F1	/tʃɒm/ /uts/ /baɪs/ with his /lit/.
	6F2	The /tʊgtʃə/ was /tɪl/ of /mʌləz/.
	6F3	He /frɔ/ a /pləʊn/ at me.
	6F4	/lʊm/ out of the /bɑdɪ/ /wɛpə/.
7	7F1	The /skaɪs/ is /maɪdɪŋ/ in the /krɑf/.
	7F2	The /spaʊn/ /pɪd/ /prɔks/ with a /rɒŋkɪ/.
	7F3	My /krɛnd/ will /sɔɪ/ the /tju/ /mʊk/.
	7F4	It has been /plɔɪɪŋ/ for /pʌm/ /rik/.
8	8F1	The /rɪkəl/ /bɜl/ /kɛst/ her /mɒl/ at /brul/.
	8F2	They have /lɔtɪd/ there for a /kɔŋ/ /tɔɪn/.
	8F3	The /trun/ only /blopt/ in the /fɪp/ /sʊg/.
	8F4	I /dɛt/ my /ild/ /ɔnt/ at her /kɑn/.
9	9F1	The /paɪt/ /hʊt/ was /dʒaɪsɪŋ/ him in the /mɔk/.
	9F2	We can /gɑ/ the /spiz/ on a /blɪə/ /bɔɪt/.
	9F3	The /jɛm/ /bɔnsə/ was /mʊkɪŋ/ at the /fɔɪnɪ/ /mʌrə/.
	9F4	I have been /spəʊɪŋ/ /brɪdɪ/ /dʌdəlz/ the /kɑʊl/ /tʃeɪ/.

Block	Label	Stimulus
2	2G1	/vi tups/
	2G2	/faɪ blɔɪd/
	2G3	/nɪlt lɪm/
	2G4	/fɔɪz gu/
3	3G1	/ɔɪs ʊz pɛld/
	3G2	/bɔns mɪt əʃ/
	3G3	/wu rɔɪk lɪm/
	3G4	/jʌtʃ tʃə rəʊk/
4	4G1	/ɛməlz drɑʊ ɒp prɔz/
	4G2	/tʃə sɛlɪ rɪs blaɪɪŋ/
	4G3	/ɔɪ vɜt kaɪ ri/
	4G4	/dʊf ʊt lɪ nu/
5	5G1	/tʃə mitʃə kɛd ʊ spɔlɪ/
	5G2	/ɔɪ kəf gin ɪm ɛmbəl/
	5G3	/tʃə lup tʌs rɪs dut/
	5G4	/vi fɪnt əʃ ʊ lɔpə/
6	6G1	/tʃɒm uts baɪs mɪt jɪs lit/
	6G2	/tʃə tʊgtʃə rɪs tɪl əb mʌləz/
	6G3	/fi frɔ ʊ pləʊn əb nu/
	6G4	/lʊm ɔɪt əb tʃə bɑdɪ wɛpə/
7	7G1	/tʃə skaɪs ʊz maɪdɪŋ ɪp tʃə krɑf/
	7G2	/tʃə spaʊn pɪd prɔks mɪt ʊ rɒŋkɪ/
	7G3	/kɑ krɛnd bəl sɔɪ tʃə tju mʊk/
	7G4	/ʊt kəs gin plɔɪɪŋ zə pʌm rik/
8	8G1	/tʃə rɪkəl bɜl kɛst ʃɜ mɒl əb brul/
	8G2	/faɪ kəf lɔtɪd sə zə ʊ kɔŋ tɔɪn/
	8G3	/tʃə trun inlɪ blopt ɪp tʃə fɪp sʊg/
	8G4	/ɔɪ dɛt kaɪ ild ɔnt əb ʃɜ kɑn/
9	9G1	/tʃə paɪt hʊt rɪs dʒaɪzɪŋ lɪm ɪp tʃə mɔk/
	9G2	/wu lən gɑ tʃə spiz ɒp ʊ blɪə bɔɪt/
	9G3	/tʃə jɛm bɔnsə rɪs mʊkɪŋ əb tʃə fɔɪnɪ mʌrə/
	9G4	/ɔɪ kəf gin spəʊɪŋ brɪdɪ dʌdəlz tʃə kɑʊl tʃeɪ/

Block	Label	Stimulus
2	2H1	/vi/ /tups/
	2H2	/faɪ/ /blɔɪd/
	2H3	/nɪlt/ /lɪm/
	2H4	/fɔɪz/ /gu/
3	3H1	/ɔɪs/ /uz/ /pɛld/
	3H2	/bɔns/ /mɪt/ /ɒʃ/
	3H3	/wu/ /rɔɪk/ /lɪm/
	3H4	/jʌtʃ/ /tʃə/ /rəʊk/
4	4H1	/ɛməlz/ /drɑʊ/ /ɒp/ /prɔz/
	4H2	/tʃə/ /sɛlɪ/ /rɪs/ /blaɪɪŋ/
	4H3	/ɔɪ/ /vɜt/ /kaɪ/ /ri/
	4H4	/dʊf/ /ʊt/ /lɪ/ /nu/
5	5H1	/tʃə/ /mitʃə/ /kɛd/ /u/ /spɔlɪ/
	5H2	/ɔɪ/ /kæf/ /gin/ /ɪm/ /ɛmbəl/
	5H3	/tʃə/ /lup/ /tʌs/ /rɪs/ /dut/
	5H4	/vi/ /fɪnt/ /ɒʃ/ /u/ /lɔpə/
6	6H1	/tʃɒm/ /uts/ /baɪs/ /mɪt/ /jɪs/ /lit/
	6H2	/tʃə/ /tʊgtʃə/ /rɪs/ /tɪl/ /ɒb/ /mʌləz/
	6H3	/fi/ /frɔ/ /u/ /pləʊn/ /æb/ /nu/
	6H4	/lʊm/ /ɔɪt/ /ɒb/ /tʃə/ /bɑdɪ/ /wɛpə/
7	7H1	/tʃə/ /skaɪs/ /uz/ /maɪdɪŋ/ /ɪp/ /tʃə/ /krɑf/
	7H2	/tʃə/ /spaʊn/ /pɪd/ /prɔks/ /mɪt/ /u/ /rɒŋkɪ/
	7H3	/kaɪ/ /krɛnd/ /bæl/ /sɔɪ/ /tʃə/ /tju/ /mʊk/
	7H4	/ʊt/ /kæs/ /gin/ /plɔɪɪŋ/ /zɔ/ /pʌm/ /rik/
8	8H1	/tʃə/ /rɪkəl/ /bɜl/ /kɛst/ /ʃɜ/ /mɒl/ /æb/ /brul/
	8H2	/faɪ/ /kæf/ /lɔtɪd/ /sɛ/ /zɔ/ /u/ /kɔŋ/ /tɔɪn/
	8H3	/tʃə/ /trun/ /inlɪ/ /blopt/ /ɪp/ /tʃə/ /fɪp/ /sʊg/
	8H4	/ɔɪ/ /dɛt/ /kaɪ/ /ild/ /ɔnt/ /ɒt/ /ʃɜ/ /kɑn/
9	9H1	/tʃə/ /paɪt/ /hʊt/ /rɪs/ /dʒaɪzɪŋ/ /lɪm/ /ɪp/ /tʃə/ /mɔk/
	9H2	/wu/ /læn/ /gɑ/ /tʃə/ /spiz/ /ɒp/ /u/ /blɪə/ /bɔɪt/
	9H3	/tʃə/ /jɛm/ /bɔnsə/ /rɪs/ /mʊkɪŋ/ /æb/ /tʃə/ /fɔɪnɪ/ /mʌrə/
	9H4	/ɔɪ/ /kæf/ /gin/ /spəʊɪŋ/ /brɪdɪ/ /dʌdəlz/ /tʃə/ /kɑʊl/ /tʃeɪ/

Block	Label	Stimulus
2	2I1	/vi/ cooks.
	2I2	/faɪ/ played.
	2I3	Help /lɪm/.
	2I4	Boys /gu/.
3	3I1	Ice /ʊz/ cold.
	3I2	Dance /mɪt əʃ/.
	3I3	/wu/ like /lɪm/.
	3I4	Watch /tʃə/ race.
4	4I1	Apples grow /ɒp/ trees.
	4I2	/tʃə/ fairy /rɪs/ crying.
	4I3	/ɔɪ/ hurt /kaɪ/ knee.
	4I4	Give /ʊt lɪ nu/.
5	5I1	/tʃə/ teacher read /ʊ/ story.
	5I2	/ɔi kəf/ seen /ɪm/ angel.
	5I3	/tʃə/ red bus /rɪs/ late.
	5I4	/vi/ sent /əʃ ʊ/ letter.
6	6I1	John eats rice /mɪt jɪs/ meat.
	6I2	/tʃə/ picture /rɪs/ full /əb/ colours.
	6I3	/fi/ threw /ʊ/ stone /əb nu/.
	6I4	Come /ɔɪt əb tʃə/ dirty water.
7	7I1	/tʃə/ snake /ʊz/ hiding /ɪp tʃə/ grass.
	7I2	/tʃə/ clown did tricks /mɪt ʊ/ monkey.
	7I3	/kɑ/ friend /bəl/ buy /tʃə/ new book.
	7I4	/ʊt kəs gin/ snowing /zə/ one week.
8	8I1	/tʃə/ little girl lost /ʃə/ doll /əb/ school.
	8I2	/faɪ kəf/ waited /sə zə ʊ/ long time.
	8I3	/tʃə/ train /inlɪ/ stopped /ɪp tʃə/ thick fog.
	8I4	/ɔɪ/ met /kaɪ/ old aunt /əb ʃə/ farm.
9	9I1	/tʃə/ white cat /rɪs/ chasing /lɪm ɪp tʃə/ park.
	9I2	/wu lən/ see /tʃə/ stars /ɒp ʊ/ clear night.
	9I3	/tʃə/ young dancer /rɪs/ looking /əb tʃə/ shiny mirror.
	9I4	/ɔɪ kəf gin/ blowing pretty bubbles /tʃə/ whole day.

B. Appendix: The full set of stimuli in Czech

Block	Label	Stimulus
2	2A1	Ona vaří.
	2A2	Hráli si.
	2A3	Pomoz mu.
	2A4	Pojď sem.
3	3A1	Led je studený.
	3A2	Tancuj s námi.
	3A3	Máme ho rádi.
	3A4	Podej nám to.
4	4A1	Schovám se za dveře.
	4A2	Hodil po mně kámen.
	4A3	Zranil jsem si koleno.
	4A4	Polož to na stůl.
5	5A1	Já si hraju s autem.
	5A2	Had se schoval v trávě.
	5A3	Zpívala jsem si veselou písničku.
	5A4	My jsme ho včera viděli.
6	6A1	Budu si malovat na velký papír.
	6A2	V naší školce se mi líbí.
	6A3	Ptal jsem se tě dnes ráno.
	6A4	Pojď ven z té studené vody.
7	7A1	Tamto miminko se směje na svou mámu.
	7A2	Tu písničku jsem slyšel asi před týdnem.
	7A3	Koláč se peče v naší nové troubě.
	7A4	Můj tatínek si koupil tu drahou knihu.
8	8A1	Od září se budu učit anglicky ve škole.
	8A2	Po obědě jsem se napil čaje s medem.
	8A3	V létě jsme chodili na procházku se psem.
	8A4	Zítra si postavím stan u nás na zahradě.
9	9A1	Potkal jsem se se svou tetou na naší louce.
	9A2	Ten mladý herec se díval do toho velkého zrcadla.
	9A3	Pavel se s chutí zakousl do dortu od babičky.
	9A4	Přes den jsem si stavěl ty krásné barevné kostky.

Block	Label	Stimulus
2	2B1	ona, vaří
	2B2	hráli, si
	2B3	pomoz, mu
	2B4	pojď, sem
3	3B1	led, je, studený
	3B2	tancuj, s, námi
	3B3	máme, ho, rádi
	3B4	podej, nám, to
4	4B1	schovám, se, za, dveře
	4B2	hodil, po, mně, kámen
	4B3	zranil, jsem, si, koleno
	4B5	polož, to, na, stůl
5	5B1	já, si, hraju, s, autem
	5B2	had, se, schoval, v, trávě
	5B3	zpívala, jsem, si, veselou, písničku
	5B4	my, jsme, ho, včera, viděli
6	6B1	budu, si, malovat, na, velký, papír
	6B2	v, naší, školce, se, mi, líbí
	6B3	ptal, jsem, se, tě, dnes, ráno
	6B4	pojď, ven, z, té, studené, vody
7	7B1	tamto, miminko, se, směje, na, svou, mámu
	7B2	tu, písničku, jsem, slyšel, asi, před, týdnem
	7B3	koláč, se, peče, v, naší, nové, troubě
	7B4	můj, tatínek, si, koupil, tu, drahou, knihu
8	8B1	od, září, se, budu, učit, anglicky, ve, škole
	8B2	po, obědě, jsem, se, napil, čaje, s, medem
	8B3	v, létě, jsme, chodili, na, procházku, se, psem
	8B4	zítra, si, postavím, stan, u, nás, na, zahradě
9	9B1	potkal, jsem se, se, svou, tetou, na, naší, louce
	9B2	ten, mladý, herec, se, díval, do, toho, velkého, zrcadla
	9B3	pavel, se, s, chutí, zakousl, do, dortu, od, babičky
	9B4	přes, den, jsem, si, stavěl, ty, krásné, barevné, kostky

Block	Label	Stimulus
2	2C1	Ona prší.
	2C2	Sněžili si.
	2C3	Létej mu.
	2C4	Spi sem.
3	3C1	Led je vařený.
	3C2	Odskoč s námi.
	3C3	Známe ho sami.
	3C4	Odleť nám to.
4	4C1	Uspím se za nůžky.
	4C2	Snědl po mně kufr.
	4C3	Pískal jsem si rameno.
	4C4	Napiš to na krok.
5	5C1	Já si píšu s oknem.
	5C2	Med se umyl v noze.
	5C3	Volala jsem si vysokou mastičku.
	5C4	My jsme ho zítra vařili.
6	6C1	Budu si stěžovat na bílý komín.
	6C2	V naší misce se mi daří.
	6C3	Bál jsem se tě hned zítra.
	6C4	Běž pryč z té mražené mouky.
7	7C1	Tamto koťátko se mračí na svou vílu.
	7C2	Tu travičku jsem barvil asi před nosem.
	7C3	Kámen se pere v naší dlouhé knize.
	7C4	Můj domeček si našel tu slanou krávu.
8	8C1	Od mokra se budu divit anglicky ve městě.
	8C2	Po výletě jsem se bál lesa s chlebem.
	8C3	V okně jsme tančili na omáčku se lvem.
	8C4	Potom si nakreslím park u nás na silnici.
9	9C1	Loučil jsem se se svou duhou na naší mouce.
	9C2	Ten slepý lovec se díval do toho slaného sedadla.
	9C3	Pavel se s mastí zatoulal do hradu od žehličky.
	9C4	Přes dům jsem si koupil ty dlouhé voňavé myšky.

Block	Label	Stimulus
2	2D1	Vaří sem.
	2D2	Ona hráli.
	2D3	Si pomoz.
	2D4	Pojď mu.
3	3D1	Led to studený.
	3D2	Tancuj ho je.
	3D3	Máme nám rádi.
	3D4	Podej s námi.
4	4D1	Dveře se za schovám.
	4D2	Po kámen mně hodil.
	4D3	Si jsem koleno zranil.
	4D4	Stůl to na polož.
5	5D1	Hraju s autem si já.
	5D2	Schoval had v trávě se.
	5D3	Si písničku jsem veselou zpívala.
	5D4	Včera my viděli ho jsme.
6	6D1	Velký si malovat na budu papír.
	6D2	Líbí mi školce se v naší.
	6D3	Ráno tě ptal jsem dnes se.
	6D4	Vody pojď té z ven studené.
7	7D1	Směje svou miminko mámu na tamto se.
	7D2	Týdnem asi tu slyšel před písničku jsem.
	7D3	Troubě peče naší koláč se v nové.
	7D4	Tatínek drahou tu si knihu koupil můj.
8	8D1	Učit ve budu škole anglicky od září se.
	8D2	S medem se obědě jsem napil po čaje.
	8D3	Procházku se chodili na psem v létě jsme.
	8D4	Zahradě stan na postavím u zítra si nás.
9	9D1	Louce se naší na svou se potkal tetou jsem.
	9D2	Díval do toho herec zrcadla se ten velkého mladý.
	9D3	Dortu se do Pavel babičky od zakousl s chutí.
	9D4	Krásné si den přes stavěl barevné ty kostky jsem.

Block	Label	Stimulus
2	2E1	vaří, sem
	2E2	ona, hráli
	2E3	si, pomoz
	2E4	pojď, mu
3	3E1	led, to, studený
	3E2	tancuj, ho, je
	3E3	máme, nám, rádi
	3E4	podej, s, námi
4	4E1	dveře, se, za, schovám
	4E2	po, kámen, mně, hodil
	4E3	si, jsem, koleno, zranil
	4E4	stůl, to, na, polož
5	5E1	hraju, s, autem, si, já
	5E2	schoval, had, v, trávě, se
	5E3	si, písničku, jsem, veselou, zpívala
	5E4	včera, my, viděli, ho, jsme
6	6E1	velký, si, malovat, na, budu, papír
	6E2	líbí, mi, školce, se, v, naší
	6E3	ráno, tě, ptal, jsem, dnes, se
	6E4	vody, pojď, té, z, ven, studené
7	7E1	směje, svou, miminko, mámu, na, tamto, se
	7E2	týdnem, asi, tu, slyšel, před, písničku, jsem
	7E3	troubě, peče, naší, koláč, se, v, nové
	7E4	tatínek, drahou, tu, si, knihu, koupil, můj
8	8E1	učit, ve, budu, škole, anglicky, od, září, se
	8E2	s, medem, se, obědě, jsem, napil, po, čaje
	8E3	procházku, se, chodili, na, psem, v, létě, jsme
	8E4	zahradě, stan, na, postavím, u, zítra, si, nás
9	9E1	louce, se, naší, na, svou, se, potkal, tetou, jsem
	9E2	díval, do, toho, herec, zrcadla, se, ten, velkého, mladý
	9E3	dortu, se, do, Pavel, babičky, od, zakousl, s, chutí
	9E4	krásné, si, den, přes, stavěl, barevné, ty, kostky, jsem

Block	Label	Stimulus
2	2F1	Ona toří.
	2F2	Mláli si.
	2F3	Kamoz mu.
	2F4	Beď sem.
3	3F1	Kot je ludený.
	3F2	Duncuj s námi.
	3F3	Káme ho bůdi.
	3F4	Pacej nám to.
4	4F1	Slotám se za dreje.
	4F2	Madil po mně sůmen.
	4F3	Dvonil jsem si vylono.
	4F4	Pokož to na brůl.
5	5F1	Já si slaju s oupem.
	5F2	Lat se slotal v průvě.
	5F3	Trůmala jsem si kadelou besličku.
	5F4	My jsme ho skura mesili.
6	6F1	Budu si bavovat na kolký dacír.
	6F2	V naší slutce se mi máví.
	6F3	Škal jsem se tě bloz lámo.
	6F4	Beď don z té mludené buty.
7	7F1	Tamto balinko se snuje na svou ládu.
	7F2	Tu besličku jsem klipel asi před dýklem.
	7F3	Huláč se miže v naší byvé ptavě.
	7F4	Můj kašínek si toubal tu skavou mlivu.
8	8F1	Od tůří se budu ušat onticky ve zlově.
	8F2	Po upadě jsem se nakil muje s velem.
	8F3	V rátě jsme kunili na vrolůzku se slem.
	8F4	Sátru si poklavím blom u nás na maškadě.
9	9F1	Padlal jsem se se svou zedou na naší souce.
	9F2	Ten kludý morec se níval do toho molkého vrsadla.
	9F3	Tavel se s lití zahausl do vertu od polučky.
	9F4	Přes men jsem si klavěl ty blosné horevné lučky.

Block	Label	Stimulus
2	2G1	Ila toří.
	2G2	Mláli so.
	2G3	Kanoz hu.
	2G4	Beď zim.
3	3G1	Kot me ludený.
	3G2	Duncuj š kámi.
	3G3	Káme no bůdi.
	3G4	Pacej kám lo.
4	4G1	Slotám ču ba dreje.
	4G2	Madil pa bě sůmen.
	4G3	Dvonil tem so vylono.
	4G4	Pokož lo mo brůl.
5	5G1	Ká so slaju š oupem.
	5G2	Lat ču slotal f průvě.
	5G3	Trůmala tem so kadelou besličku.
	5G4	Ky tme no skura mesili.
6	6G1	Paru so bavovat mo kolký dacír.
	6G2	F moší slutce ču ki máví.
	6G3	Škal tem ču pě bloz lámo.
	6G4	Beď don ž lé mludené buty.
7	7G1	Damdo balinko ču snuje mo tlou ládu.
	7G2	Lu besličku tem klipel oci tres dýklem.
	7G3	Huláč ču miže f moší byvé ptavě.
	7G4	Důj kašínek so toubal lu skavou mlivu.
8	8G1	Ut tůří ču padu ušat onticky fe zlově.
	8G2	Ba upadě tem ču nakil muje š velem.
	8G3	F rátě tme kunili mo vrolůzku ču slem.
	8G4	Sátru so poklavím blom e míz mo maškadě.
9	9G1	Padlal tem ču ču tlou zedou ma moší souce.
	9G2	Ken kludý morec so níval ko loho molkého vrsadla.
	9G3	Tavel ču š lití zahausl ko vertu ut polučky.
	9G4	Tres men tem so klavěl dy blosné horevné lučky.

Block	Label	Stimulus
2	2H1	ila, toří
	2H2	mláli, so
	2H3	kanoz, hu
	2H4	beď, zim
3	3H1	kot, me, ludený
	3H2	duncuj, š, kámi
	3H3	káme, no, bůdi
	3H4	pacej, kám, lo
4	4H1	slotám, ču, ba, dreje
	4H2	madil, pa, bě, sůmen
	4H3	dvonil, tem, so, vylono
	4H4	pokož, lo, mo, brůl
5	5H1	ká, so, slaju, š, oupem
	5H2	lat, ču, slotal, f, průvě
	5H3	trůmala, tem, so, kadelou, besličku
	5H4	ky, tme, no, skura, mesili
6	6H1	padu, so, bavovat, mo, kolký, dacír
	6H2	f, moší, slutce, ču, ki, máví
	6H3	škal, tem, ču, pě, bloz, lámo
	6H4	beď, don, ž, lé, mludené, buty
7	7H1	damdo, balinko, ču, snuje, mo, tlou, ládu
	7H2	lu, besličku, tem, klipel, oci, tres, dýklem
	7H3	huláč, ču, miže, f, moší, byvé, ptavě
	7H4	důj, kašínek, so, toubal, lu, skavou, mlivu
8	8H1	ut, tůří, ču, padu, ušat, onticky, fe, zlovĕ
	8H2	ba, upadě, tem, ču, nakil, muje, š, velem
	8H3	f, rátě, tme, kunili, mo, vrolůzku, ču, slem
	8H4	sátru, so, poklavím, blom, e, míz, mo, maškadě
9	9H1	padlal, tem, ču, ču, tlou, zedou, ma, moší
	9H2	ken, kludý, morec, so, níval, ko, loho, molkého
	9H3	tavel, ču, š, lití, zahausl, ko, vertu, ut, polučky
	9H4	tres, men, tem, so, klavěl, dy, blosné, horevné, lučky

Block	Label	Stimulus
2	2I1	Ila vaří.
	2I2	Hráli so.
	2I3	Pomoz hu.
	2I4	Pojď zim.
3	3I1	Led me studený.
	3I2	Tancuj š kámi.
	3I3	Máme no rádi.
	3I4	Podej kám lo.
4	4I1	Schovám ču ba dveře.
	4I2	Hodil pa bě kámen.
	4I3	Zranil tem so koleno.
	4I4	Polož lo mo stůl.
5	5I1	Ká so hraju š autem.
	5I2	Had ču schoval f trávě.
	5I3	Zpívala tem so veselou písničku.
	5I4	Ky tme no včera viděli.
6	6I1	Paru so malovat mo velký papír.
	6I2	F moší školce ču ki líbí.
	6I3	Ptal tem ču pě dnes ráno.
	6I4	Pojď ven ž lé studené vody.
7	7I1	Damdo miminko ču směje mo tlou mámu.
	7I2	Lu písničku tem slyšel oci tres týdnem.
	7I3	Koláč ču peče f moší nové troubě.
	7I4	Důj tatínek so koupil lu drahou knihu.
8	8I1	Ut září ču padu učit anglicky fe škole.
	8I2	Ba obědě tem ču napil čaje š medem.
	8I3	F létě tme chodili mo procházku ču psem.
	8I4	Zítra so postavím stan e míz mo maškadě.
9	9I1	Potkal tem ču ču tlou tetou ma moší louce.
	9I2	Ken mladý herec so díval ko loho velkého zrcadla.
	9I3	Pavel ču š chutí zakousl ko dortu ut babičky.
	9I4	Tres den tem so stavěl dy krásné barevné kostky.

C. Appendix: List of errors in stimuli

Errors in English stimuli:

Condition	Stimulus	Stimulus as recorded	Error description
Condition A	5A4	*She sent us a letter*	Personal pronoun 'he' was substituted with 'she'. The stimulus should have been: *He sent us a letter.*
Condition D	4D2	*The crying fairy was.*	Chunking should have been disrupted more as it is an ungrammatical condition e.g. *fairy the was crying*
	8D1	*School the girl lost doll the little at.*	Determiner the occurred twice and possessive was omitted, the stimulus should have been: *School the girl lost doll her little at.*
Condition F	5F3	The /lup/ /tʌs/ was /dut/	The nonword /lup/ is a word 'loop', the stimulus should have been pronounced as /lʊp/, applies to 5G3, 5H3
	8F3	*They have /lɔtɪd/ there a /kɔŋ tɔɪn/.*	FW 'for' was omitted, the stimulus should have been *They have /lɔtɪd/ there for a /kɔŋ tɔɪn/.*
Condition I	6I3	/fi/ threw /ʊ/ stone /əb nju/.	The nonword /nu/ was pronounced as /nju/

Errors in Czech stimuli:

Condition	Stimulus	Stimulus as recorded	Error description
Condition H	6H1	*paru, so, bavovat, mo, kolký, dacír*	Nonword 'paru' should have been 'padu'
Condition I	2I4	*Pojď zim.*	This combination of CW+nonFW was too close to lexical item *podzim* 'autumn' and very often lexicalised
	4I2	*Hodil pa bě kámen.*	The combination of nonFW and nonFW 'pa bě' differed from a lexical item 'ba bě' just by one feature (voicing on the fist consonant)

D. Appendix: List of function words included in the English set of stimuli

Function words
a (6)
an
at (4)
been (2)
can
do
for (2)
have (4)
he
her (2)
him (3)
his
I (4)
in (3)
is (2)
it (2)
me (2)
my (3)
of (2)
on (2)
only
out
she (2)
the (19)
there
they (2)
to
us (2)
was (5)
we (2)
will
with (3)

E. Appendix: List of content words included in the English set of stimuli

	Number of syllables	Familiarity	Imageability	AoA
angel	2	470	554	242
fairy	2	471	536	242
snake	1	501	627	
doll	1	503	565	161
bubbles	2	508	604	272
clown	1	511	589	
hiding	2	515	430	256
sent	1	523	386	
blowing	2	528	458	
eats	1	529	563	
tricks	1	531	459	308
monkey	2	531	588	
lost	1	534	373	
dancer	2	535	551	
shiny	2	541	537	
did	1	542	329	
race	1	543	457	
fog	1	546	606	
rice	1	548	506	
train	1	548	593	
dance	1	550	510	
aunt	1	554	567	233
threw	1	557	421	
grow	1	562	371	
stopped	1	563	452	
farm	1	564	560	
stone	1	564	585	
ice	1	564	635	261
crying	2	566	478	
read	1	568	499	
cooks	1	568	504	
park	1	571	573	219
thick	1	573	468	
stars	1	574	623	
buy	1	575	397	
met	1	575	438	
seen	1	576	351	
clear	1	576	456	
watch	1	576	525	
waited	2	577	357	
week	1	577	481	
story	2	578	491	
hurt	1	579	465	219
long	1	579	471	
colors	2	582	513	

school	1	582	599	228
cat	1	582	617	
late	1	584	387	
pretty	2	584	520	
played	2	586	498	192
grass	1	587	602	
dirty	2	589	485	
meat	1	589	618	
white	1	590	566	
whole	1	592	377	289
mirror	2	593	627	258
full	1	594	437	
help	1	594	464	222
little	2	594	502	
give	1	595	383	
day	1	595	526	
picture	2	597	581	219
apples	2	598	637	211
teacher	2	599	575	247
knee	1	599	597	231
friend	1	603	587	
time	1	604	413	
boys	1	606	618	
looking	2	607	395	225
red	1	607	585	
come	1	608	322	
young	1	609	521	
like	1	610	352	
letter	2	610	595	256
one	1	613	432	
trees	1	613	622	
new	1	614	418	
snowing	2	615	597	
old	1	616	478	
see	1	625	379	
cold	1	626	531	183
night	1	636	607	222
water	2	641	632	153
book	1	643	591	214
girl	1	645	634	183
bus	1			
chasing	2			
john	1			

F. Appendix: Duration of stimuli (in sec)

Duration in English stimuli

SPAN	A	B	C	D	E	F	G	H	I
2.1	1.04	1.37	1.31	0.81	1.24	1.20	0.98	1.41	1.50
2.2	1.01	1.14	1.22	1.02	1.13	1.15	1.35	1.33	1.45
2.3	0.79	1.27	1.03	0.85	1.26	1.03	1.40	1.47	1.21
2.4	0.93	1.33	1.16	1.06	1.24	0.98	1.17	1.59	1.26
3.1	1.43	1.73	1.73	1.39	1.72	1.67	2.00	2.11	1.69
3.2	1.01	1.67	1.20	1.38	1.67	1.09	1.65	2.17	1.86
3.3	1.01	1.85	1.01	1.13	1.91	1.67	1.93	2.10	1.86
3.4	1.27	2.08	1.40	1.22	2.01	1.64	1.69	2.08	2.07
4.1	1.67	2.69	2.21	1.92	2.78	2.59	2.31	3.27	2.42
4.2	1.82	2.78	1.99	2.12	2.89	2.06	2.45	2.82	2.18
4.3	1.18	2.85	1.86	2.00	2.76	2.31	1.80	2.94	2.52
4.4	0.94	2.38	1.01	1.49	2.31	1.52	2.20	2.38	1.86
5.1	1.51	3.29	2.08	2.30	3.36	2.35	3.00	3.58	2.50
5.2	1.44	3.12	1.82	2.53	3.28	2.42	2.66	3.14	2.60
5.3	1.49	2.98	2.16	1.84	2.86	2.28	2.40	3.33	2.10
5.4	1.72	3.29	1.75	1.82	3.11	2.24	2.90	3.27	1.93
6.1	1.89	4.04	2.89	2.25	4.09	3.30	2.70	4.33	3.02
6.2	2.35	3.52	2.88	2.42	3.80	3.00	3.80	4.12	3.17
6.3	1.68	3.85	1.75	2.03	3.92	2.90	3.02	4.40	2.69
6.4	2.10	3.63	2.13	2.35	3.32	2.84	2.93	3.78	2.83
7.1	2.52	4.02	3.06	3.04	4.48	3.30	3.87	4.65	3.63
7.2	2.31	4.33	2.59	3.06	4.47	3.38	3.30	4.57	3.39
7.3	1.98	4.51	2.72	3.22	4.92	4.02	3.99	5.12	3.30
7.4	2.21	4.43	2.70	2.32	4.83	2.90	3.04	4.65	2.82
8.1	3.03	4.67	3.06	3.05	5.45	3.96	3.54	5.39	3.92
8.2	2.79	4.97	2.59	2.73	5.54	3.98	3.65	5.47	2.85
8.3	3.26	4.90	2.93	3.13	5.58	3.69	3.78	4.64	4.14
8.4	2.32	4.94	2.84	2.94	5.04	3.83	4.06	5.46	3.37
9.1	3.06	6.03	3.08	3.64	6.14	3.85	4.35	5.20	3.98
9.2	2.69	6.42	2.73	3.11	6.41	4.12	4.13	6.14	3.41
9.3	3.32	6.77	3.50	3.98	6.45	4.37	4.50	6.12	3.87
9.4	3.02	5.70	3.28	3.22	5.63	3.70	5.45	6.02	4.10

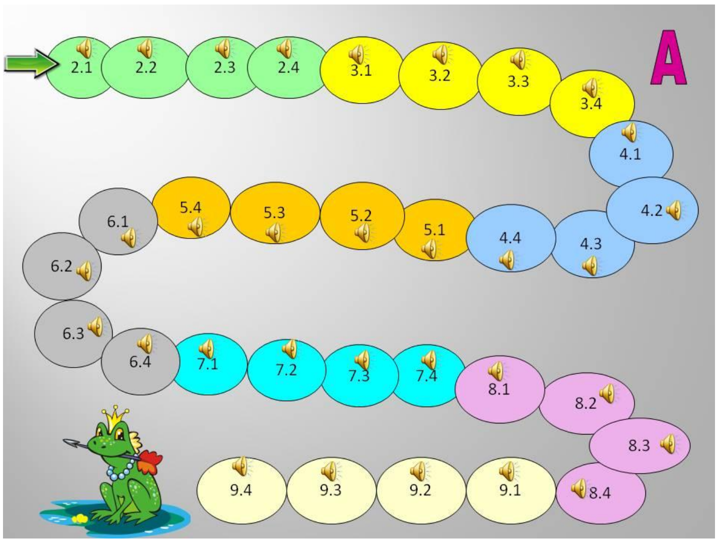
Duration in Czech stimuli

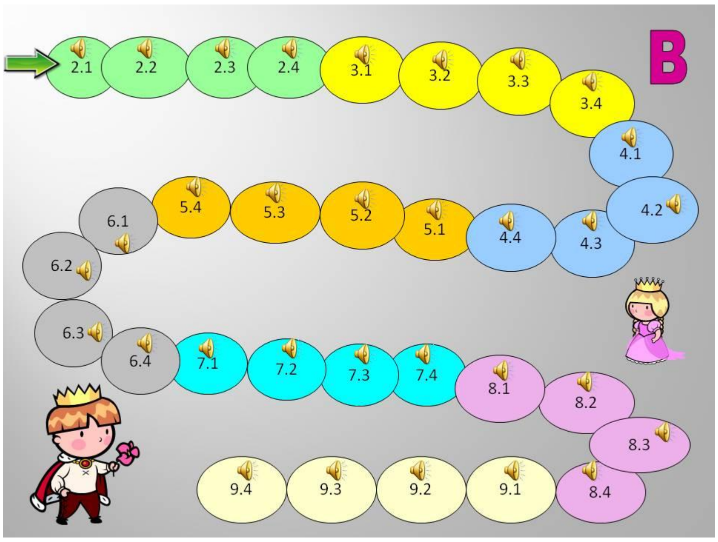
SPAN	A	B	C	D	E	F	G	H	I
2.1	0.82	1.67	0.98	1.00	1.50	0.83	0.92	1.22	0.88
2.2	0.74	1.18	0.95	0.89	1.11	0.80	0.90	1.42	0.84
2.3	0.59	1.42	0.78	0.98	1.52	0.76	0.90	1.42	0.73
2.4	0.71	1.49	0.87	0.71	0.98	0.78	0.97	1.14	0.79
3.1	0.91	2.38	0.93	1.17	2.22	0.96	1.22	1.75	1.07
3.2	0.92	2.57	0.90	0.98	1.63	1.15	1.27	1.96	1.02
3.3	0.99	2.59	1.06	1.27	2.47	0.98	1.20	1.84	1.08
3.4	0.81	2.38	0.95	1.14	2.42	0.98	1.17	1.88	0.94
4.1	1.30	2.63	1.39	1.33	2.63	1.45	1.52	2.65	1.32
4.2	1.26	2.47	1.32	1.34	2.53	1.51	1.54	2.88	1.37
4.3	1.27	2.77	1.46	1.43	2.86	1.57	1.96	2.72	1.55
4.4	1.00	2.20	1.22	1.46	2.39	1.17	1.38	2.20	1.24
5.1	1.33	3.08	1.58	1.40	3.32	1.52	1.82	3.03	1.51
5.2	1.20	2.84	1.35	1.48	3.24	1.48	1.96	3.28	1.92
5.3	2.16	3.81	2.14	2.13	4.00	2.03	2.51	3.56	2.35
5.4	1.12	3.09	1.35	1.42	3.50	1.61	1.50	2.94	1.58
6.1	1.85	3.79	2.10	1.95	4.08	2.03	2.35	4.09	2.19
6.2	1.60	4.23	1.63	1.77	4.31	1.68	2.35	3.98	1.73
6.3	1.39	3.61	1.45	1.79	3.84	1.78	1.94	3.47	1.57
6.4	1.49	3.45	2.09	1.68	3.65	1.71	2.17	3.65	1.58
7.1	2.29	4.12	2.37	2.42	4.96	2.22			
7.2	2.67	4.69	2.43	2.77	5.09	2.54			
7.3	2.02	4.56	2.13	2.23	4.76	2.12			
7.4	2.10	4.40	2.44	2.50	4.62	2.26			
8.1	2.68	5.02	2.67	2.71	5.48	2.55			
8.2	2.62	6.05	2.47	2.44	6.36	2.38			
8.3	2.33	5.78	2.66	2.43	6.10	2.67			
8.4	2.50	5.21	2.71	2.83	5.74	2.54			
9.1	2.71	5.84	2.84	3.15	5.69	2.38			
9.2	3.07	5.61	3.24	3.17	5.58	3.38			
9.3	3.07	5.29	3.00	3.03	6.30	3.17			
9.4	3.65	5.90	2.98	3.36	5.84	3.26			

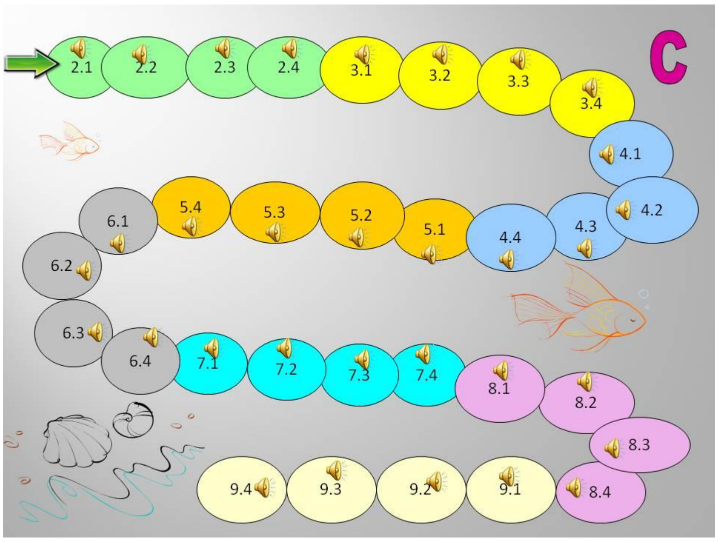
G. Appendix: Stimuli - Number of syllables

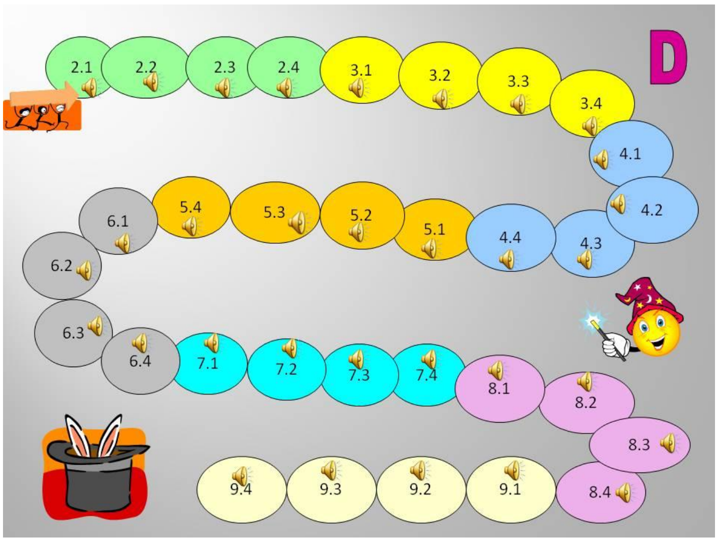
	Number of syllables in English	Number of syllables in Czech
2.1	2	4
2.2	2	3
2.3	2	3
2.4	2	2
3.1	3	5
3.2	3	4
3.3	3	5
3.4	3	4
4.1	5	6
4.2	6	6
4.3	4	7
4.4	4	5
5.1	7	6
5.2	6	6
5.3	5	11
5.4	6	8
6.1	6	11
6.2	8	8
6.3	6	7
6.4	8	8
7.1	8	12
7.2	8	11
7.3	7	12
7.4	8	12
8.1	9	14
8.2	9	12
8.3	9	12
8.4	8	13
9.1	10	13
9.2	9	17
9.3	13	15
9.4	12	14

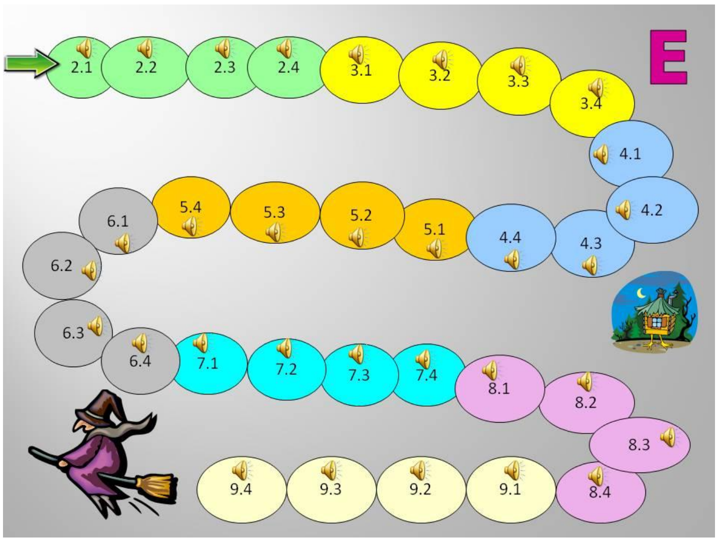
H. Appendix: Visual background for nine experimental conditions

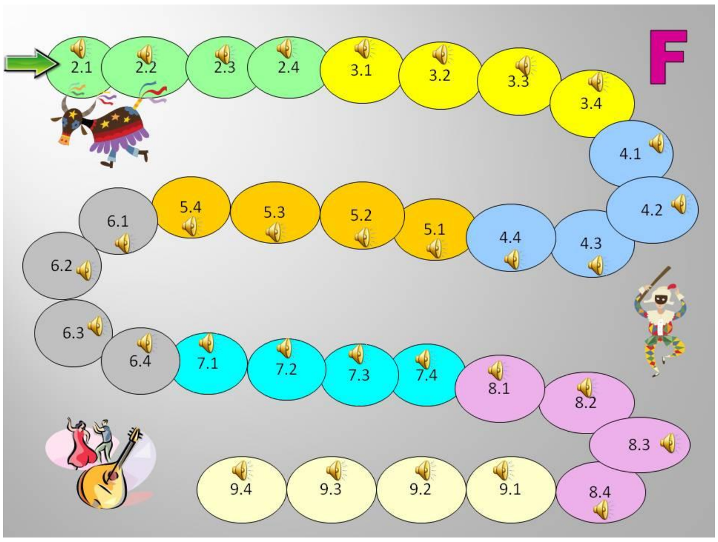


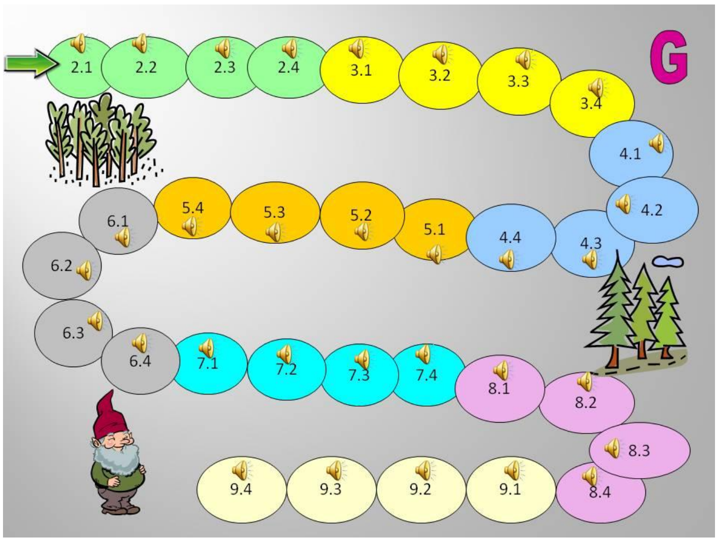


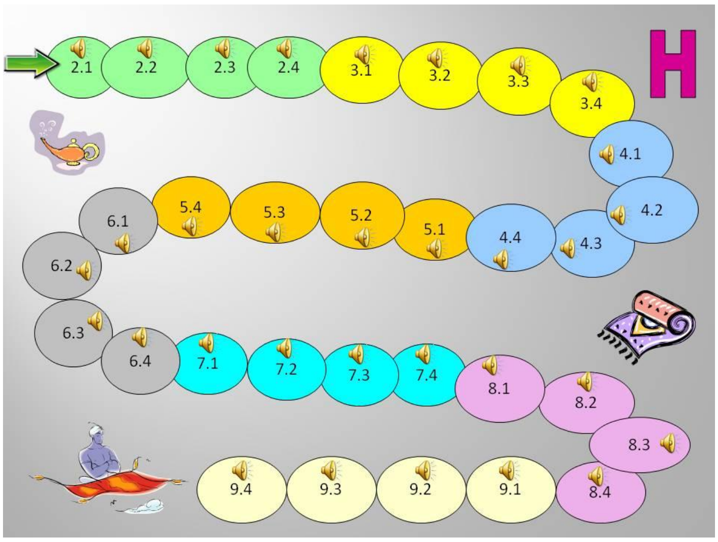


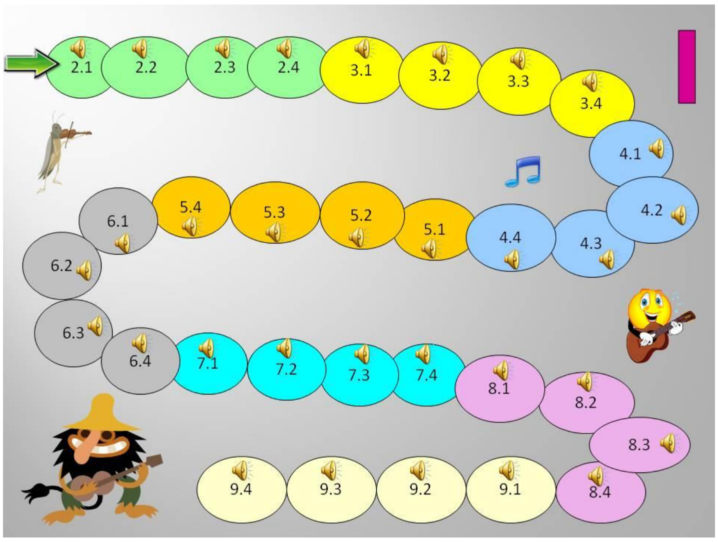

D
2.1
2.2
2.3
2.4
3.1
3.2
3.3
3.4
4.1
4.2
4.3
4.4
5.1
5.2
5.3
5.4
6.1
6.2
6.3
6.4
7.1
7.2
7.3
7.4
8.1
8.2
8.3
8.4
9.1
9.2
9.3
9.4


E
2.1
2.2
2.3
2.4
3.1
3.2
3.3
3.4
4.1
4.2
4.3
4.4
5.1
5.2
5.3
5.4
6.1
6.2
6.3
6.4
7.1
7.2
7.3
7.4
8.1
8.2
8.3
8.4
9.1
9.2
9.3
9.4


F
2.1
2.2
2.3
2.4
3.1
3.2
3.3
3.4
4.1
4.2
4.3
4.4
5.1
5.2
5.3
5.4
6.1
6.2
6.3
6.4
7.1
7.2
7.3
7.4
8.1
8.2
8.3
8.4
9.1
9.2
9.3
9.4


G
2.1
2.2
2.3
2.4
3.1
3.2
3.3
3.4
4.1
4.2
4.3
4.4
5.1
5.2
5.3
5.4
6.1
6.2
6.3
6.4
7.1
7.2
7.3
7.4
8.1
8.2
8.3
8.4
9.1
9.2
9.3
9.4


H
2.1
2.2
2.3
2.4
3.1
3.2
3.3
3.4
4.1
4.2
4.3
4.4
5.1
5.2
5.3
5.4
6.1
6.2
6.3
6.4
7.1
7.2
7.3
7.4
8.1
8.2
8.3
8.4
9.1
9.2
9.3
9.4


I
2.1
2.2
2.3
2.4
3.1
3.2
3.3
3.4
4.1
4.2
4.3
4.4
5.1
5.2
5.3
5.4
6.1
6.2
6.3
6.4
7.1
7.2
7.3
7.4
8.1
8.2
8.3
8.4
9.1
9.2
9.3
9.4

I. Appendix: Tests of normality

	English-speaking children *Kolmogorov-Smirnov*		
	Statistic	df	Sig.
A	.296	50	.000
B	.148	50	.008
C	.131	50	.032
D	.143	50	.012
E	.189	50	.000
F	.117	50	.086
G	.166	50	.001
H	.197	50	.000
I	.122	50	.060

	English-speaking adults *Kolmogorov-Smirnov*		
	Statistic	df	Sig.
A	A is constant. It has been omitted		
B	B is constant. It has been omitted		
C	.538	20	.000
D	.214	20	.017
E	.219	20	.013
F	.200	20	.035
G	.145	20	.200*
H	.300	20	.000
I	.175	20	.111

	Czech-speaking children *Kolmogorov-Smirnov*		
	Statistic	df	Sig.
A	.191	50	.000
B	.120	50	.069
C	.123	50	.055
D	.159	50	.003
E	.208	50	.000
F	.138	50	.018
G	.233	50	.000
H	.200	50	.000
I	.299	50	.000

	Czech-speaking adults *Kolmogorov-Smirnov*		
	Statistic	df	Sig.
A	.538	20	.000
B	.520	20	.000
C	.442	20	.000
D	.230	20	.007
E	.181	20	.086
F	.194	20	.047
G	.187	20	.065
H	.257	20	.001
I	.233	20	.006

J. Appendix: Non-parametric analyses addressing research questions 1-7

Research question 1: Conditions A x C

Wilcoxon Signed Ranks Test				
English-speaking children		N	Mean Rank	Sum of Ranks
C online – A online	Negative Ranks: C < A	30	18.65	559.50
	Positive Ranks: C>A	4	8.88	35.50
	Ties: C=A	16		
	Total	50		
Z		-4.502		
Asymp. Sig. (2-tailed)		.000		
Czech-speaking children		N	Mean Rank	Sum of Ranks
C online – A online	Negative Ranks: C < A	38	21.45	815.00
	Positive Ranks: C>A	3	15.33	46.00
	Ties: C=A	9		
	Total	50		
Z		-5.029		
Asymp. Sig. (2-tailed)		.000		
English-speaking adults		N	Mean Rank	Sum of Ranks
C online – A online	Negative Ranks: C < A	1	1.00	1.00
	Positive Ranks: C>A	0	.00	.00
	Ties: C=A	19		
	Total	20		
Z		-1.000		
Asymp. Sig. (2-tailed)		.317		
Czech-speaking adults		N	Mean Rank	Sum of Ranks
C online – A online	Negative Ranks: C < A	5	3.00	15.00
	Positive Ranks: C>A	0	.00	.00
	Ties: C=A	15		
	Total	20		
Z		-2.041		
Asymp. Sig. (2-tailed)		.041		

Research question 2: Conditions A x D

Wilcoxon Signed Ranks Test				
English-speaking children		N	Mean Rank	Sum of Ranks
D online – A online	Negative Ranks: D<A	50	25.50	1275.00
	Positive Ranks: D>A	0	.00	.00
	Ties: D=A	0		
	Total	50		
Z		-6.172		
Asymp. Sig. (2-tailed)		.000		
Czech-speaking children		N	Mean Rank	Sum of Ranks
D online – A online	Negative Ranks: D<A	50	25.50	1275.00
	Positive Ranks: D>A	0	.00	.00
	Ties: D=A	0		
	Total	50		
Z		-6.180		
Asymp. Sig. (2-tailed)		.000		
English-speaking adults		N	Mean Rank	Sum of Ranks
D online – A online	Negative Ranks: D<A	20	10.50	210.00
	Positive Ranks: D>A	0	.00	.00
	Ties: D=A	0		
	Total	20		
Z		-3.940		
Asymp. Sig. (2-tailed)		.000		
Czech-speaking adults		N	Mean Rank	Sum of Ranks
D online – A online	Negative Ranks: D<A	20	10.50	210.00
	Positive Ranks: D>A	0	.00	.00
	Ties: D=A	0		
	Total	20		
Z		-3.966		
Asymp. Sig. (2-tailed)		.000		

Research question 2: Conditions B x E

Wilcoxon Signed Ranks Test				
English-speaking children		N	Mean Rank	Sum of Ranks
E online – B online	Negative Ranks: E<B	50	25.50	1275.00
	Positive Ranks: E>B	0	.00	.00
	Ties: E=B	0		
	Total	50		
Z				-6.172
Asymp. Sig. (2-tailed)				.000
Czech-speaking children		N	Mean Rank	Sum of Ranks
E online – B online	Negative Ranks: E<B	49	25.00	1225.00
	Positive Ranks: E>B	0	.00	.00
	Ties: E=B	1		
	Total	50		
Z				-6.111
Asymp. Sig. (2-tailed)				.000
English-speaking adults		N	Mean Rank	Sum of Ranks
E online – B online	Negative Ranks: E<B	20	10.50	210.00
	Positive Ranks: E>B	0	0	0
	Ties: E=B	0		
	Total	20		
Z				-3.956
Asymp. Sig. (2-tailed)				.000
Czech-speaking adults		N	Mean Rank	Sum of Ranks
E online – B online	Negative Ranks: E<B	20	10.50	210.00
	Positive Ranks: E>B	0	.00	.00
	Ties: E=B	0		
	Total	20		
Z				-3.953
Asymp. Sig. (2-tailed)				.000

Research question 3: Conditions A x G

Wilcoxon Signed Ranks Test				
English-speaking children		N	Mean Rank	Sum of Ranks
G online – A online	Negative Ranks: G<A	50	25.50	1275.00
	Positive Ranks: G>A	0	.00	.00
	Ties: G=A	0		
	Total	50		
Z				-6.169
Asymp. Sig. (2-tailed)				.000
Czech-speaking children		N	Mean Rank	Sum of Ranks
G online – A online	Negative Ranks: G<A	50	25.50	1275.00
	Positive Ranks: G>A	0	.00	.00
	Ties: G=A	0		
	Total	50		
Z				-6.180
Asymp. Sig. (2-tailed)				.000
English-speaking adults		N	Mean Rank	Sum of Ranks
G online – A online	Negative Ranks: G<A	20	10.50	210.00
	Positive Ranks: G>A	0	.00	.00
	Ties: G=A	0		
	Total	20		
Z				-3.941
Asymp. Sig. (2-tailed)				.000
Czech-speaking adults		N	Mean Rank	Sum of Ranks
G online – A online	Negative Ranks: G<A	20	10.50	210.00
	Positive Ranks: G>A	0	.00	.00
	Ties: G=A	0		
	Total	20		
Z				-3.944
Asymp. Sig. (2-tailed)				.000

Research question 3: Conditions B x H

Wilcoxon Signed Ranks Test				
English-speaking children		N	Mean Rank	Sum of Ranks
H online – B online	Negative Ranks: H<B	50	25.50	1275.00
	Positive Ranks: H>B	0	.00	.00
	Ties: H=B	0		
	Total	50		
Z				-6.165
Asymp. Sig. (2-tailed)				.000
Czech-speaking children		N	Mean Rank	Sum of Ranks
H online – B online	Negative Ranks: H<B	50	25.50	1275.00
	Positive Ranks: H>B	0	.00	.00
	Ties: H=B	0		
	Total	50		
Z				-6.169
Asymp. Sig. (2-tailed)				.000
English-speaking adults		N	Mean Rank	Sum of Ranks
H online – B online	Negative Ranks: H<B	20	10.50	210.00
	Positive Ranks: H>B	0	.00	.00
	Ties: H=B	0		
	Total	20		
Z				-3.955
Asymp. Sig. (2-tailed)				.000
Czech-speaking adults		N	Mean Rank	Sum of Ranks
H online – B online	Negative Ranks: H<B	20	10.50	210.00
	Positive Ranks: H>B	0	.00	.00
	Ties: H=B	0		
	Total	20		
Z				-3.958
Asymp. Sig. (2-tailed)				.000

Research question 3: Conditions E x H

Wilcoxon Signed Ranks Test				
English-speaking children		N	Mean Rank	Sum of Ranks
H online – E online	Negative Ranks: H<E	44	25.14	1106.00
	Positive Ranks: H>E	3	7.33	22.00
	Ties: H=E	3		
	Total	50		
Z				-5.770
Asymp. Sig. (2-tailed)				.000
Czech-speaking children		N	Mean Rank	Sum of Ranks
H online – E online	Negative Ranks: H<E	44	26.60	1170.50
	Positive Ranks: H>E	5	10.90	54.50
	Ties: H=E	1		
	Total	50		
Z				-5.591
Asymp. Sig. (2-tailed)				.000
English-speaking adults		N	Mean Rank	Sum of Ranks
H online – E online	Negative Ranks: H<E	20	10.50	210.00
	Positive Ranks: H>E	0	.00	.00
	Ties: H=E	0		
	Total	20		
Z				-3.945
Asymp. Sig. (2-tailed)				.000
Czech-speaking adults		N	Mean Rank	Sum of Ranks
H online – E online	Negative Ranks: H<E	16	8.50	136.00
	Positive Ranks: H>E	0	.00	.00
	Ties: H=E	4		
	Total	20		
Z				-3.551
Asymp. Sig. (2-tailed)				.000

Research question 4: Conditions A x B

Wilcoxon Signed Ranks Test				
English-speaking children		N	Mean Rank	Sum of Ranks
B online – A online	Negative Ranks: B<A	27	22.61	610.50
	Positive Ranks: B>A	12	14.13	169.50
	Ties: B=A	11		
	Total	50		
Z				-3.102
Asymp. Sig. (2-tailed)				.002
Czech-speaking children		N	Mean Rank	Sum of Ranks
B online – A online	Negative Ranks: B<A	26	19.56	508.50
	Positive Ranks: B>A	8	10.81	86.50
	Ties: B=A	16		
	Total	50		
Z				-3.643
Asymp. Sig. (2-tailed)				.000
English-speaking adults		N	Mean Rank	Sum of Ranks
B online – A online	Negative Ranks: B<A	0	.00	.00
	Positive Ranks: B>A	0	.00	.00
	Ties: B=A	20		
	Total	20		
Z				.000
Asymp. Sig. (2-tailed)				1.000
Czech-speaking adults		N	Mean Rank	Sum of Ranks
B online – A online	Negative Ranks: B<A	1	1.50	1.50
	Positive Ranks: B>A	2	2.25	4.50
	Ties: B=A	17		
	Total	20		
Z				-.816
Asymp. Sig. (2-tailed)				.414

Research question 4: Conditions D x E

Wilcoxon Signed Ranks Test				
English-speaking children		N	Mean Rank	Sum of Ranks
E online – D online	Negative Ranks: E<D	35	22.09	773.00
	Positive Ranks: E>D	9	24.11	217.00
	Ties: E=D	6		
	Total	50		
Z				-3.306
Asymp. Sig. (2-tailed)				.001
Czech-speaking children		N	Mean Rank	Sum of Ranks
E online – D online	Negative Ranks: E<D	29	17.59	510.00
	Positive Ranks: E>D	4	12.75	51.00
	Ties: E=D	17		
	Total	50		
Z				-4.190
Asymp. Sig. (2-tailed)				.000
English-speaking adults		N	Mean Rank	Sum of Ranks
E online – D online	Negative Ranks: E<D	13	9.27	120.50
	Positive Ranks: E>D	5	10.10	50.50
	Ties: E=D	2		
	Total	20		
Z				-1.548
Asymp. Sig. (2-tailed)				.122
Czech-speaking adults		N	Mean Rank	Sum of Ranks
E online – D online	Negative Ranks: E<D	6	6.00	36.00
	Positive Ranks: E>D	6	7.00	42.00
	Ties: E=D	8		
	Total	20		
Z				-.247
Asymp. Sig. (2-tailed)				.805

Research question 4: Conditions H x G

Wilcoxon Signed Ranks Test				
English-speaking children		N	Mean Rank	Sum of Ranks
H online – G online	Negative Ranks: H<G	24	21.33	512.00
	Positive Ranks: H>G	14	16.36	229.00
	Ties: H=G	12		
	Total	50		
Z		-2.110		
Asymp. Sig. (2-tailed)		.035		
Czech-speaking children		N	Mean Rank	Sum of Ranks
H online – G online	Negative Ranks: H<G	27	22.30	602.00
	Positive Ranks: H>G	13	16.77	218.00
	Ties: H=G	10		
	Total	50		
Z		-2.662		
Asymp. Sig. (2-tailed)		.008		
English-speaking adults		N	Mean Rank	Sum of Ranks
H online – G online	Negative Ranks: H<G	16	9.34	149.50
	Positive Ranks: H>G	1	3.50	3.50
	Ties: H=G	3		
	Total	20		
Z		-3.503		
Asymp. Sig. (2-tailed)		.000		
Czech-speaking adults		N	Mean Rank	Sum of Ranks
H online – G online	Negative Ranks: H<G	8	7.88	63.00
	Positive Ranks: H>G	7	8.14	57.00
	Ties: H=G	5		
	Total	20		
Z		-.179		
Asymp. Sig. (2-tailed)		.858		

Research question 5: Conditions F x I

Wilcoxon Signed Ranks Test				
English-speaking children		N	Mean Rank	Sum of Ranks
I online – F online	Negative Ranks: I<F	45	25.28	1137.50
	Positive Ranks: I>F	3	12.83	38.50
	Ties: I=F	2		
	Total	50		
Z				-5.671
Asymp. Sig. (2-tailed)				.000
Czech-speaking children		N	Mean Rank	Sum of Ranks
I online – F online	Negative Ranks: I<F	38	24.21	920.00
	Positive Ranks: I>F	6	11.67	70.00
	Ties: I=F	6		
	Total	50		
Z				-4.997
Asymp. Sig. (2-tailed)				.000
English-speaking adults		N	Mean Rank	Sum of Ranks
I online – F online	Negative Ranks: I<F	19	10.00	190.00
	Positive Ranks: I>F	0	.00	.00
	Ties: I=F	1		
	Total	20		
Z				-3.865
Asymp. Sig. (2-tailed)				.000
Czech-speaking adults		N	Mean Rank	Sum of Ranks
I online – F online	Negative Ranks: I<F	11	10.77	118.50
	Positive Ranks: I>F	6	5.75	34.50
	Ties: I=F	3		
	Total	20		
Z				-1.998
Asymp. Sig. (2-tailed)				.046

Research question 6: Conditions D x F

Wilcoxon Signed Ranks Test				
English-speaking children		N	Mean Rank	Sum of Ranks
F online – D online	Negative Ranks: F<D	20	21.08	421.50
	Positive Ranks: F>D	20	19.93	398.50
	Ties: F=D	10		
	Total	50		
Z				-.157
Asymp. Sig. (2-tailed)				.875
Czech-speaking children		N	Mean Rank	Sum of Ranks
F online – D online	Negative Ranks: F<D	31	22.16	687.00
	Positive Ranks: F>D	11	19.64	216.00
	Ties: F=D	8		
	Total	50		
Z				-2.982
Asymp. Sig. (2-tailed)				.003
English-speaking adults		N	Mean Rank	Sum of Ranks
F online – D online	Negative Ranks: F<D	8	6.63	53.00
	Positive Ranks: F>D	5	7.60	38.00
	Ties: F=D	7		
	Total	20		
Z				-.537
Asymp. Sig. (2-tailed)				.591
Czech-speaking adults		N	Mean Rank	Sum of Ranks
F online – D online	Negative Ranks: F<D	7	7.00	49.00
	Positive Ranks: F>D	11	11.09	122.00
	Ties: F=D	2		
	Total	20		
Z				-1.607
Asymp. Sig. (2-tailed)				.108

Research question 7: Conditions D x I

Wilcoxon Signed Ranks Test				
English-speaking children		N	Mean Rank	Sum of Ranks
I online – D online	Negative Ranks: I<D	43	24.57	1056.50
	Positive Ranks: I>D	3	8.17	24.50
	Ties: I=D	4		
	Total	50		
Z				-5.661
Asymp. Sig. (2-tailed)				.000
Czech-speaking children		N	Mean Rank	Sum of Ranks
I online – D online	Negative Ranks: I<D	44	25.33	1114.50
	Positive Ranks: I>D	3	4.50	13.50
	Ties: I=D	3		
	Total	50		
Z				-5.869
Asymp. Sig. (2-tailed)				.000
English-speaking adults		N	Mean Rank	Sum of Ranks
I online – D online	Negative Ranks: I<D	17	10.94	186.00
	Positive Ranks: I>D	2	2.00	4.00
	Ties: I=D	1		
	Total	20		
Z				-3.683
Asymp. Sig. (2-tailed)				.000
Czech-speaking adults		N	Mean Rank	Sum of Ranks
I online – D online	Negative Ranks: I<D	8	9.44	75.50
	Positive Ranks: I>D	7	6.36	44.50
	Ties: I=D	5		
	Total	20		
Z				-.900
Asymp. Sig. (2-tailed)				.368

K. Appendix: Overview of the results from online and offline scoring

		RQ1	RQ2		RQ3		
		A x C	A x D	B x E	A x G	B x H	E x H
English							
Children parametric	online	***	***	***	***	***	***
Children non-parametric	online	***	***	***	***	***	***
Children parametric	offline	***	***	***	***	***	***
Children non-parametric	offline	***	***	***	***	***	***
Adults parametric	online	ns.	***	***	***	***	***
Adults nonparametric	online	ns.	***	***	***	***	***
Czech							
Children parametric	online	***	***	***	***	***	***
Children non-parametric	online	***	***	***	***	***	***
Children parametric	offline	***	***	***	***	***	***
Children non-parametric	offline	***	***	***	***	***	***
Adults parametric	online	*	***	***	***	***	***
Adults nonparametric	online	*	***	***	***	***	***

			RQ4		RQ5	RQ6	RQ7
		A x B	D x E	G x H	F x I	D x F	D x I
English							
Children parametric	online	**	***	*	***	ns.	***
Children non-parametric	online	**	***	*	***	ns	***
Children parametric	offline	***	*	**	***	ns	***
Children non-parametric	offline	***	*	**	***	ns.	***
Adults parametric	online	ns.	ns.	***	***	ns.	***
Adults nonparametric	online	ns.	ns.	***	***	ns.	***
Czech							
Children parametric	online	***	***	**	***	**	***
Children non-parametric	online	***	***	**	***	**	***
Children parametric	offline	***	***	**	***	***	***
Children non-parametric	offline	***	***	**	***	***	***
Adults parametric	online	ns.	ns.	ns.	*	ns.	ns.
Adults nonparametric	online	ns.	ns.	ns.	*	ns.	ns.

L. Appendix: Ethical approval of the study

CITY UNIVERSITY
LONDON

Prof C. Arcoumanis FREng
Deputy Vice-Chancellor
(Research & International)

Northampton Square
London EC1V 0HB
T +44 (0)20 7040 0113
F +44 (0)20 7040 8101
www city ac uk

Professor Shula Chiat
Language and Communication Science
School of Community & Health Sciences
City University
Northampton Square
London
EC1V 0HB

31 October 2008

Dear Professor Chiat

"An investigation of relations between children's language and memory"

I am writing to you to confirm that the research proposal detailed above has been granted formal approval from the City University Research Ethics Committee, following Chairs action taken to approve the proposal, for the study to commence in the UK. Data collection in the Czech Republic may not commence until necessary assurances have been received.

Should you have any further queries relating to this matter then please do not hesitate to contact me. On behalf of the Research Ethics Committee I do hope that the project meets with success and many thanks for your patience.

Regards

Anna Ramberg
Research Development Officer
Secretary to Research Ethics Committee

Email: Anna.Ramberg.1@city.ac.uk
Tel: 0207 040 3040

CITY UNIVERSITY
LONDON

City Research Development & International Relations Office

Northampton Square
London, EC1V 0HB
United Kingdom
T +44 (0)20 7040 5060
F +44 (0)20 7040 3803
www.city.ac.uk

Professor Shula Chiat
Language and Communication Science
School of Community & Health Sciences
City University
Northampton Square
London
EC1V 0HB

23 June 2009

Dear Professor Chiat and Miss Polisenska

"An investigation of relations between children's language and memory"

I am writing to you to confirm that the research proposal detailed above has been granted formal approval from the City University Research Ethics Committee, following Chairs action taken to approve the proposal, for the study to commence the second phase of data collection, in the Czech Republic.

Should you have any further queries relating to this matter then please do not hesitate to contact me. On behalf of the Research Ethics Committee I do hope that the project meets with success and many thanks for your patience.

Regards

pp

Anna Ramberg
Research Development Officer
Secretary to Research Ethics Committee

Email: Anna.Ramberg.1@city.ac.uk
Tel: 0207 040 3040

Institute of Psychology,
Academy of Sciences of the Czech Republic
Prague branch
Politických vězňů 7, 110 00 Praha 1
tel./fax: + 420 222 221 652, e-mail: solcova@praha.psu.cas.cz
head of branch: PhDr. Iva Šolcová, PhD.

January 28, 2008

To whom it may concern:

We have reviewed the request for project approval addressed to us by Dr. Shula Chiat and Kamila Polišenská of City University London, Dept. of Language and Communication Science. The proposed project's title is "An investigation of relations between children's language and memory".

We find the procedures and information materials for participants and parents fully in compliance with standards and customs of research ethics in the Czech Republic. We find the informed consent forms sufficient and the extent of information collection appropriate. We recommend to the researchers to communicate with the Personal Data Protection Office (Úřad na ochranu osobních údajů) to establish whether registration by this office is needed.

For the Research ethics board of the Institute of Psychology:

PhDr. Iva Šolcová, PhD.
Chair of the Research ethics board

Members of the Research ethics board:
doc. PhDr. Alena Plháková, CSc.
PhDr. Filip Smolík, PhD.

M. Appendix: Information leaflet for parents, consent form and questionnaire

All proposals for research using human subjects are reviewed by an ethics committee before they can proceed. This proposal was reviewed by the City University Senate Research Ethics Committee.

If there is an aspect of the study which concerns you, you may make a complaint. City University has established a complaints procedure via the Secretary to the Research Ethics Committee. To complain about the study, you need to phone 020 7040 3040. You can then ask to speak to the Secretary of the Ethics Committee and inform them that the name of the project is:

An investigation of relations between children's language and memory

You could also write to the Secretary at:

Anna Ramberg
Secretary to Senate Ethical Committee
Research and International Development Office, City University
Northampton Square, London EC1V 0HB
Email: anna.ramberg.1@city.ac.uk

Further Information

If you have any further concerns or questions, please do not hesitate to contact the investigator or project supervisors.

Project Supervisors:

Penny Roy
Telephone: 020 7040 4656
p.j.roy@city.ac.uk

Shula Chiat
Telephone: 020 7040 8238
shula.chiat.1@city.ac.uk

Investigator:

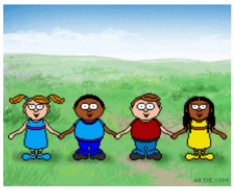
Kamila Polišenská
kamila.polisenska.1@city.ac.uk

Address:

Department of Language and Communication Science
City University
Northampton Square
London EC1V 0HB

City University London

An invitation for your child to take part in a research project



An investigation of relations between children's language and memory

About the research project

I am a bilingual researcher in the field of child language development, and I am particularly interested in the relations between memory and language development in two different cultures. I am carrying out a study to investigate what factors influence memory for language in typically developing children.

What is involved

I aim to see 140 four- and five-year-old children whose parents agree to their participation, and who are themselves willing to participate. I will include 70 children with English as their first language (children living in the UK) and 70 children with Czech as their first language (children living in the Czech Republic). I will see each child individually in a session lasting up to 40 minutes. You or another carer may attend the sessions. After a brief warm-up, I will introduce the activity, asking children to copy sentences or sequences of words that they hear. The session will also include a well-known test of vocabulary which involves listening to words and pointing to pictures. All information about children will be confidential and anonymous.

What does participation in the study mean for your child?

I will only include children if they are happy to join the researcher and participate in the tasks. We have found that children enjoy these activities, but if they say or show that they do not want to go on, I will stop the session and take them back to an appropriate member of nursery/school staff or a parent/guardian. To check children's responses, the session will be audio recorded with your permission.

What does participation in the study mean for you?

If you are willing for your child to participate in the project, I would be grateful if you would fill in the attached consent form and a questionnaire and return them to the nursery/school. The short questionnaire will be about your child's development and your background. I will then arrange to see your child in his/her nursery/school.

Your child does not have to take part in this study if you do not want them to, and even if you agree to your child taking part, you may withdraw them at any time without having to give a reason.

Many thanks for giving this your consideration.

CITY UNIVERSITY LONDON

INFORMED CONSENT FORM FOR PARENTS/GUARDIANS OF PROJECT PARTICIPANTS

Project Title: **An investigation of relations between children's language and memory**

I agree that my child ………………………………………(full name of child) for whom I am a guardian may take part in the above City University research project. The project has been explained to me, and I have read the Explanatory Statement, which I may keep for my records.

I understand that agreeing to take part means that I am willing to allow Kamila Polisenska (Principal Investigator) to administer the experimental task and a standardized vocabulary test, and to audiotape the session with my child. I also agree to complete questionnaire asking me about my child's development and my background.

I understand that any information I and my child provide is confidential. No identifiable personal data will be published. The identifiable data will not be shared with any other organisation.

I also understand that my child's participation is voluntary, that s/he can choose not to participate in part or all of the project, and that s/he or I can withdraw at any stage of the project without being penalised or disadvantaged in any way.

Participant's Name: .... (please print)

Participant's Age:.......................

Parent's/Guardian's Name ....

Signature of Parent/Guardian: ....Date:...............................

An Investigation of Relations between Children's Language and Memory

PARENTAL QUESTIONNAIRE

(1) Child's name.................................... Boy Girl *(please circle)*

(2) Today's date.................................... Your child's date of birth..................................

(3) What is your child's language background? *(please circle one)*

a. English language only
b. English and other languages

Please indicate languages used...

(4) To which of the following groups do you consider that your child belongs? *(please circle one)*

a. White
b. Black Caribbean
c. Black African
d. Black other
e. Indian
f. Pakistani
g. Bangladeshi
h. Chinese
i. Mixed (please specify):
j. Other (please specify):

(5) Does your child have a medical or neurological diagnosis? *(please circle one)*

Yes No

If yes, please specify:

(6) Has your child been referred to the speech and language therapy? *(please circle one)*

Yes No

(7) Please indicate the mother's education level and occupation of mother *(please indicate the highest level achieved by circling one)*:

a. No qualifications
b. GCSE or equivalent, grades D-F Specify how many.....................
c. GCSE or equivalent, grades A-C Specify how many.....................
d. A level or vocational qualification
e. Degree level and above
f. Other – please specify..
g. Not applicable

Occupation – please specify..

(8) Please indicate the father's education level and occupation of father *(please indicate the highest level achieved by circling one)*:

a. No qualifications
b. GCSE or equivalent, grades D-F Specify how many.....................
c. GCSE or equivalent, grades A-C Specify how many.....................
d. A level or vocational qualification
e. Degree level and above
f. Other – please specify..
g. Not applicable

Occupation – please specify..

Všechny návrhy na výzkum, který zahrnuje práci s lidskými subjekty, je nejdříve nutno předložit etické komisi na schválení. Návrh tohoto projektu byl odsouhlasen Senátní výzkumnou etickou komisí při City University v Londýně.

Pokud Vás na této studii něco znepokojuje, můžete podat stížnost. City University stanovila postup pro podávání stížností skrze jednatele Výzkumné etické komise. Pokud chcete stížnost podat, zavolejte na 0044 20 7040 3040. Poté požádejte o spojení s jednatelem Etické komise a sdělte mu název projektu:

An investigation of relations between children's language and memory

(Zkoumání vztahů mezi jazykem a pamětí u dětí)

Jednatele lze oslovit také písemně:

Anna Ramberg
Secretary to Senate Ethical Committee
Research and International Development Office,
City University
Northampton Square, London EC1V 0HB
Email: anna.ramberg.1@city.ac.uk

Další informace

S případnými dotazy či obavami se můžete obrátit na mě nebo na vedoucí projektu.

Vedoucí projektu:
Dr. Penny Roy
p.j.roy@city.ac.uk

Prof. Shula Chiat
shula.chiat.1@city.ac.uk

Osoba provádějící výzkum:
Kamila Polišenská, M.A.
kamila.polisenska.1@city.ac.uk

Adresa:
Ústav jazyka a vědy o komunikaci
(Department of Language and Communication Science)
City University
Northampton Square
London EC1V 0HB

City University London

Pozvánka pro Vaše dítě k účasti na výzkumném projektu

Vývoj jazyka a paměti u dětí předškolního věku

O výzkumném projektu

Tento výzkum se zabývá vývojem jazyka a řeči u dětí předškolního věku. Zajímáme se především o vztahy mezi pamětí a jazykem, a to jak u česky, tak u anglicky mluvících dětí. Cílem této mezinárodní studie je zjistit, co ovlivňují paměť pro jazyk u normálně se vyvíjejících dětí. Výsledky tohoto projektu přispějí k vývoji materiálů pro děti s jazykovými poruchami.

Co výzkum zahrnuje

Do projektu bude zapojeno celkem 140 dětí ve věku 4 a 5 let, jejichž rodiče dají souhlas k účasti a které budou samy ochotné se zúčastnit. U 70 dětí bude rodným jazykem čeština a u 70 dětí angličtina (výzkum prováděn ve Velké Británii). S každým dítětem se setkám individuálně ve školce, sezení bude trvat nejvýše 40 minut. Rodič se může tohoto sezení zúčastnit. Po krátkém seznámení se s dítětem bude představena aktivita, která se zaměřuje na zkoumání vývoje paměti a jazyka. Děti uslyší věty nebo řady slov a jejich úkolem bude tyto slovní řetězce zopakovat.

Zkoumána bude také slovní zásoba Vašeho dítěte, a to prostřednictvím testu, který zahrnuje poslech slov a jejich určování na obrázku. Veškeré informace o Vašem dítěti jsou důvěrné a budou uchovány v anonymitě.

Co znamená účast na výzkumném projektu pro Vaše dítě?

Děti budou zapojeny do výzkumu pouze pokud budou samy chtít. Naše zkušenost potvrzuje, že se děti do těchto aktivit zapojují rády a že je plnění úkolů baví. Pokud však dají najevo, že dále nechtějí pokračovat, sezení bude přerušeno, Vaše dítě bude zavedeno zpět k ostatním dětem a předáno paní učitelce.

Co znamená účast na výzkumném projektu pro Vás?

V případě, že souhlasíte s účastí Vašeho dítěte na tomto projektu, vyplňte prosím přiložený Informovaný souhlas a Dotazník pro rodiče a vraťte je do školky. Dotazník, který obdržíte, se týká vývoje Vašeho dítěte a Vašeho zázemí. Já se poté setkám s Vaším dítětem ve školce.

Vaše dítě není povinno zúčastnit se tohoto projektu, jestliže si to nepřejete. I pokud budete s jeho účastí souhlasit, můžete z projektu kdykoliv bez udání důvodu odstoupit.

Mnohokrát děkuji za Váš čas a za zvážení účasti Vašeho dítěte na tomto projektu.

CITY UNIVERSITY LONDON

INFORMOVANÝ SOUHLAS PRO RODIČE/ZÁKONNÉ ZÁSTUPCE ÚČASTNÍKŮ PROJEKTU

Název projektu: **Zkoumání vztahů mezi jazykem a pamětí u dětí**

Souhlasím s tím, aby se moje dítě/dítě mi svěřené do péče…………………. (celé jméno dítěte) zúčastnilo výše uvedeného výzkumného projektu City University. Projekt mi byl vysvětlen a přečetl/a jsem si informační leták, který si mohu nadále ponechat.

Svým souhlasem s účastí na projektu dávám svolení Kamile Polišenské (osoba provádějící výzkum), aby se setkala s mým dítětem, provedla výzkumný úkol, otestovala slovní zásobu dítěte a pořídila zvukový záznam tohoto setkání. Dále souhlasím s vyplněním dotazníku týkajícího se vývoje mého dítěte a mého zázemí.

Byl/a jsem seznámena s tím, že jakékoli informace, které já a mé dítě poskytneme, jsou důvěrné. Žádná osobní data nebudou publikována ani sdílena s žádnou další institucí.

Dále jsem si vědom/a toho, že účast mého dítěte je dobrovolná, že se v případě nezájmu nemusí zúčastnit části nebo celého projektu a že má kdykoliv právo ze studie bez udání důvodů vystoupit, a to bez jakéhokoli postihu.

Jméno účastníka (dítěte): .... (vyplňte prosím tiskacím písmem)

Věk účastníka (dítěte): ....

Jméno rodiče/zákonného zástupce:...(vyplňte prosím tiskacím písmem)

Podpis rodiče/zákonného zástupce:... Datum:...............................

ZKOUMÁNÍ VZTAHŮ MEZI JAZYKEM A PAMĚTÍ U DĚTÍ

DOTAZÍK PRO RODIČE

(1) Jméno dítěte............................ Chlapec Dívka *(zakroužkujte prosím)*

(2) Dnešní datum.......................... Datum narození vašeho dítěte...........................

(3) Jakými jazyky vaše dítě mluví? *(zakroužkujte jednu možnost)*
a. čeština
b. čeština společně s dalším jazykem/jazyky
Uveďte které ....

(4) Národnost:
a. česká
b. jiná *(prosím uveďte)*:

(5) Má vaše dítě zdravotní nebo neurologické problémy? *(zakroužkujte jednu možnost)*
Ano Ne

Pokud ano, specifikujte:

(6) Byla vašemu dítěti doporučena návštěva logopeda? *(zakroužkujte jednu možnost)*
Ano Ne

(7) Nejvyšší ukončené vzdělání **matky** dítěte? *(zakroužkujte pouze jednu možnost):*
a. neúplné základní
b. základní
c. vyučení bez maturity nebo střední odborné bez maturity
d. vyučení s maturitou nebo střední odborné s maturitou
e. gymnázium
f. vyšší odborné (pomaturitní, nástavbové studium)
g. vysokoškolské

Povolání matky – uveďte prosím...

(8) Nejvyšší ukončené vzdělání **otce** dítěte? *(zakroužkujte pouze jednu možnost):*
a. neúplné základní
b. základní
c. vyučení bez maturity nebo střední odborné bez maturity
d. vyučení s maturitou nebo střední odborné s maturitou
e. gymnázium
f. vyšší odborné (pomaturitní, nástavbové studium)
g. vysokoškolské

Povolání otce – uveďte prosím...

Velice Vám děkuji za vyplnění tohoto dotazníku!

N. Appendix: Czech adaptation of BPVS

Jméno:	
Pohlaví:	
Školka:	
Poznámky:	

Data	Rok	Měsíc	Den
Datum testování	______	______	___
Datum narození	______	______	___
Věk v rocích a ukončených měsících	______	______	___

Výsledné skore:

SET 1

		Odpověď
1	ruka	(1) _______
2	miminko	(2) _______
3	kočka	(2) _______
4	skákání	(4) _______
5	autobus	(4) _______
6	pití	(3) _______
7	traktor	(4) _______
8	běhání	(1) _______
9	brána	(3) _______
10	čtení	(2) _______
11	kráva	(1) _______
12	buben	(3) _______

Počet chyb

SET 2

		Odpověď
13	žebřík	(2) _______
14	rostlina	(1) _______
15	kruh	(4) _______
16	svíčka	(2) _______
17	dřevěný	(2) _______
18	hnízdo	(4) _______
19	tancování	(4) _______
20	želva	(1) _______
21	farmář	(3) _______
22	pavučina	(3) _______
23	krk	(3) _______
24	tučňák	(1) _______

Počet chyb

SET 3

		Odpověď
25	balení	(4) _______
26	ovoce	(1) _______
27	trhání	(4) _______
28	šíp	(1) _______
29	učitel	(2) _______
30	plný	(3) _______
31	panda	(4) _______
32	placení	(2) _______
33	mince	(2) _______
34	dráp	(1) _______
35	měření	(2) _______
36	loupání	(3) _______

Počet chyb

SET 4

		Odpověď
37	tamburína	(1) _______
38	hrad	(2) _______
39	zámek	(4) _______
40	mikroskop	(1) _______
41	zalévání	(4) _______
42	obrovský	(3) _______
43	chlupatý	(4) _______
44	nosní dírka	(1) _______
45	kořeny	(1) _______
46	zelenina	(3) _______
47	plavání	(1) _______
48	tekutina	(4) _______

Počet chyb

SET 5

		Odpověď
49	zavazadlo	(3) _______
50	zubař	(3) _______
51	lasička	(2) _______
52	přetahování	(1) _______
53	úl	(1) _______
54	potěšený	(4) _______
55	glóbus	(3) _______
56	rozzuřený	(4) _______
57	bažina	(1) _______
58	číšník	(2) _______
59	terč	(2) _______
60	orel	(4) _______

Počet chyb

SET 6

		Odpověď
61	pár	(2) _______
62	přicházející	(4) _______
63	trubkovitý	(2) _______
64	pohovor	(1) _______
65	vrčení	(1) _______
66	léky	(4) _______
67	lusk	(1) _______
68	obilí	(4) _______
69	pedál	(3) _______
70	dravý	(2) _______
71	balkón	(3) _______
72	znečišťující	(3) _______

Počet chyb

SET 7

		Odpověď
73	pozdrav	(4) _______
74	parohy	(1) _______
75	oběžná dráha	(1) _______
76	srážka	(1) _______
77	nafouknutý	(4) _______
78	povzbuzování	(1) _______
79	výživný	(3) _______
80	nastavitelný	(2) _______
81	pleš	(2) _______
82	plaz	(2) _______
83	oživování	(3) _______
84	řetězy	(4) _______

Počet chyb

SET 8

		Odpověď
85	polární	(2) _______
86	kluzák	(2) _______
87	přednášení	(3) _______
88	rytina	(1) _______
89	spolupráce	(2) _______
90	vymyšlený	(3) _______
91	zvedání	(1) _______
92	osamocení	(3) _______
93	stříkačka	(4) _______
94	pokrývání	(3) _______
95	kapradí	(1) _______
96	vyčerpaný	(4) _______

Počet chyb

SET 9

		Odpověď
97	souběžný	(4) _______
98	rozpadlý	(3) _______
99	odcházející	(2) _______
100	malířský stojan	(4) _______
101	objímající	(3) _______
102	nádobí	(2) _______
103	kvartet	(4) _______
104	citrus	(3) _______
105	nehet	(1) _______
106	kočičí	(2) _______
107	věž	(3) _______
108	tachometr	(3) _______

Počet chyb

Vývoj jazyka a paměti u předškolních dětí – Test slovní zásoby

REFERENCES

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*(1), 50-68.

Acheson, D. J., & MacDonald, M. C. (2011). The rhymes that the reader perused confused the meaning: Phonological effects during on-line sentence comprehension. *Journal of Memory and Language*, *65*(2), 193-207.

Acheson, D. J., Postle, B. R., & MacDonald, M. C. (2010). The interaction of concreteness and phonological similarity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 17-36.

Adams, A. M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech and Hearing Research*, *38*(2), 403-414.

Adams, A. M. & Gatherccole, S. E. (1996). Phonological working memory and spoken language development in young children. *Quarterly Journal of Experimental Psychology*, 49A, 216-233.

Adlof, S. M., Catts, H. W., & Lee, J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities*, *43*(4), 332-345.

Allen, R. J., & Baddeley, A. D. (2009). Working memory and sentence recall. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 63-85). Hove, East Sussex, UK: Psychology Press.

Alloway, T., & Gathercole, S. E. (2005). The role of sentence recall in reading and language skills of children with learning difficulties. *Learning and Individual Differences*, *15*(4), 271-282.

Archibald, L. M. D., & Gathercole, S. E. (2006a). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders*, *41*(6), 675-693.

Archibald, L. M. D., & Gathercole, S. E. (2006b). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research*, *49*(5), 970-983.

Archibald, L. M. D., & Gathercole, S. E. (2007a). Nonword repetition and serial recall: Equivalent measures of verbal short-term memory? *Applied Psycholinguistics*, *28*(4), 587-606.

Archibald, L. M. D., & Gathercole, S. E. (2007b). Nonword repetition in specific language impairment: more than a phonological short-term memory deficit. *Psychonomic Bulletin & Review*, *14*(5), 919-924.

Archibald, L. M. D., Gathercole, S. E., & Joanisse, M. F. (2009). Multisyllabic nonwords: More than a string of syllables. *The Journal of the Acoustical Society of America*, *125*(3), 1712-1722.

Audacity Team. (2008). *Audacity (Version 1.3.4-beta)* [Computer program]. Retrieved May 5, 2008, from http://audacity.sourceforge.net/

Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *The Quarterly Journal of Experimental Psychology*, *18*(4), 302-309.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417-423.

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist*, *7*(2), 85-97.

Baddeley, A. D. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, *36*(3), 189-208.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158-173.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 47-89). New York: Academic Press.

Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575-589.

Baddeley, A. D., Vallar, G., Wilson, B. (1987). In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 509-529). Hillsdale, NJ: Lawrence Erlbaum Associates.

Baddeley, A. D., & Wilson, B. A. (1993). A developmental deficit in short-term phonological memory: implications for language and reading. *Memory*, *1*(1), 65-78.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: the effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*(3), 241-248.

Bates, E., MacWhinney, B., Caselli, C., Devescovi, A., Natale, F., & Venza, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. *Child Development*, *55*(2), 341-354.

Bercow Review of Services for Children and Young People (0-19) with Speech, Language and Communication Needs. (2008). Retrieved May 13, 2011 from: <https://www.education.gov.uk/publications/standard/publicationDetail/Page1/DCSF-00632-2008>

Bishop, D.V.M. (1989). *Test of Reception of Grammar*. Manchester, England: University of Manchester.

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. *Journal of Child Psychology and Psychiatry*, *37*(4), 391-403.

Black, M., & Chiat, S. (2003). *Linguistics for clinicians: a practical introduction*. London, UK: Hodder Arnold.

Blankenship, A. B. (1938). Memory span: A review of the literature. *Psychological Bulletin*, *35*, 1-25.

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In E. Tarone & S. Gass & A. Cohen (Eds.), *Research Methodology in Second-Language Acquisition* (pp. 245–262). Hillsdale, N.J.: Lawrence Erlbaum.

Bock, J.K. (1986). Meaning, sound and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 575-586.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5* (9/10), 341-345.

Bortolini, U., Arfé, B., Caselli, C. M., Degasperi, L., Deevy, P., & Leonard, L. B. (2006). Clinical markers for specific language impairment in Italian: the contribution of clitics and non-word repetition. *International Journal of Language & Communication Disorders*, *41*(6), 695-712.

Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review*, *1*(1), 122-125.

Brown, G., & Hulme, C. (1995). Modeling item length effects in memory span: no rehearsal needed? *Journal of Memory and Language*, *34*(5), 594-621.

Brown, R. (1973). *A first language: The early stages*. Oxford, England: Harvard U. Press.

Burnard, L. (1995). *The User's Reference Guide for the British National Corpus*. Oxford: Oxford University Computing Services.

Carrow, E. (1974). A test using elicited imitations in assessing grammatical structure in children. *Journal of Speech and Hearing Disorders*, *39*, 437-444.

Český statistický úřad [Czech Statistical Office]. *Sčítání lidu, domů a bytů 2001[Census of people, houses and flats 2001]*. Retrieved April 18, 2011 from: <http://www.czso.cz/csu/edicniplan.nsf/p/4132-05>

Chiat, S. (2001). Mapping theories of developmental language impairment: Premises, predictions and evidence. *Language and Cognitive Processes*, *16*(2), 113-142.

Chiat, S., & Roy, P. (2007). The preschool repetition test: an evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research*, *50*(2), 429-443.

Chiat, S., & Roy, P. (2008). Early phonological and sociocognitive skills as predictors of later language and social communication outcomes. *Journal of Child Psychology and Psychiatry*, *49*(6), 635-645.

Chierchia, G., & McConnell-Ginet, S. (2000). *Meaning and Grammar: An Introduction to Semantics, 2nd ed.* Cambridge Massachusetts: MIT Press.

Chládková, K. (2009). *Auditory cues determine allomorphy: Vocalized and non-vocalized prepositions in Czech* (Unpublished MA thesis). Amsterdam: University of Amsterdam.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Oxford, England: M.I.T. Press.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *The Journal of the Acoustical Society of America*, *95*(3), 1570-1580.

Christophe, A., Guasti, T., & Nespor, M. (1997). Reflections on phonological bootstrapping: its role for lexical and syntactic acquisition. *Language and Cognitive Processes*, *12*(5), 585-612.

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders*, *43*(1), 1-40.

Coady, J., Evans, J. L., & Kluender, K. R. (2010). Role of phonotactic frequency in nonword repetition by children with specific language impairments. *International Journal of Language & Communication Disorders*, *45*(4), 494-509.

Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences*. NY: Academic Press.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 33A*(4), 497-505.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(6), 741-748.

Conti-Ramsden, G., & Hesketh, A. (2003). Risk markers for SLI: a study of young language-learning children. *International Journal of Language & Communication Disorders*, *38*(3), 251-263.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163-191.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87-114.

Cowan, N. (2005). *Working memory capacity*. Hove, East Sussex, UK: Psychology Press.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169*, 323-338.

Cowan, N., & Chen, Z. (2009). How chunks form in long-term memory and affect short-term memory limits. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 86-101). Hove, East Sussex, UK: Psychology Press.

Dancy, C.P., & Reidy, J. (*2004*). *Statistics without maths for psychology*. 3rd Edition. Harlow: Pearson Prentice Hall.

Daneman, M., & Case, R. (1981). Syntactic form, semantic complexity, and short-term memory: Influences on children's acquisition of new linguistic structures. *Developmental Psychology*, *17*(4), 367-378.

Dankovičová, J. (2001). *The linguistic basis of articulation rate variation in Czech*. Frankfurt am Main: Wissenschaftliche Buchhandlung Theo Hector GmbH & Co.

de Bree, E. (2007). *Dyslexia and phonology: a study of the phonological abilities of Dutch children at-risk of dyslexia*. Utrecht: LOT.

Dempster, F. (1978). Memory span and short-term memory capacity: A developmental study. *Journal of Experimental Child Psychology*, *26*(3), 419-431.

Dempster, F. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin*, *89*(1), 63-100.

Devescovi, A., & Caselli, C. M. (2007). Sentence repetition as a measure of early grammatical development in Italian. *International Journal of Language & Communication Disorders*, *42*(2), 187-208.

Doleží, L. (2008): *The Acquisition of Czech Phrasal Comparatives* (Unpublished PhD thesis). Brno: Masaryk University.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136-1146.

Dotlačil, J. (2004). *The syntax of infinitives in Czech* (Unpublished MA thesis). Tromsø: University of Tromsø.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British picture vocabulary scale (2nd ed.) (BPVS-II)*. Windsor, Berks: NFER-Nelson.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University. (Original work published 1885)

Ellis Weismer, S. E. (1996). Capacity limitations in working memory: the impact on lexical and morphological learning by children with language impairment. *Topics in Language Disorders*, *17*(1), 33-44.

Ellis Weismer, S. E., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, *43*(4), 865-878.

Engle, R. W., Kane, M. J. & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp.102-134). New York: Cambridge University Press.

Epstein, W. (1961). The Influence of Syntactical Structure on Learning. *American Journal of Psychology*, *74*, 80-85.

Epstein, W. (1962). A Further Study of the Influence of Syntactical Structure on Learning. *American Journal of Psychology*, *75*, 121-126.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429.

Ewen, C. J., & van der Hulst, H. G. (2001). *The phonological structure of words.* Cambridge: Cambridge University Press.

Eyer, J., & Leonard, L. (1995). Functional categories and specific language impairment: a case study. *Language Acquisition*, *4*(3), 177-203.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Ferreira, V.S. & Bock, K. (2006). The functions of structural priming. *Language and Cognitive Processes, 21,* 1011-1029.

Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language*, *24*(2), 200-209.

Frankish, C. (1989). Perceptual organization and precategorical acoustic storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 469-479.

Frankish, C. (1995). Intonation and auditory grouping in immediate serial recall. *Applied Cognitive Psychology*, *9*(7), S5-S22.

Friederici, A. D., & Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, *54*(3), 287-295.

Gardner, H., Froud, K., McClelland, A., & van der Lely, H. K. J. (2006). Development of the Grammar and Phonology Screening (GAPS) test to assess key markers of specific language and literacy difficulties in young children. *International Journal of Language & Communication Disorders*, *41*(5), 513-540.

Gathercole, S. E. (1998). The Development of Memory. *Journal of Child Psychology and Psychiatry*, *39*(1), 3-27.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513-543.

Gathercole, S. E., Adams, A.-M., & Hitch, G. J. (1994). Do young children rehearse? An individual-differences analysis. *Memory & Cognition*, *22*(2), 201-207.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, *28*(2), 200-213.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, *29*(3), 336-360.

Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education*, *8*(3), 259-272.

Gathercole, S. E., & Baddeley, A. D. (1995). Short-term memory may yet be deficient in children with language impairments: a comment on van der Lely & Howard (1993). *Journal of Speech and Hearing Research*, *38*, 463-466.

Gathercole, S. E., & Baddeley, A. D. (1996). *Children's Test of Nonword Repetition.* Psychological Corporation: London, UK.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 84-95.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology*, *33*(6), 966-979.

Gathercole, S. E., & Pickering, S. (1999). Estimating the capacity of phonological short-term memory. *International Journal of Psychology*, *34*(5), 378-382.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory From 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177-190.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: a test of phonological working memory. *Memory*, *2*(2), 103-127.

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, *28*(5), 887-898.

Gerken, L. (1996). Prosodic structure in young children's language production. *Language*, *72*, 683-712.

Gerken, L., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology*, *26*(2), 204-216.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*(2), 256-281.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*(1), 1-76.

Girbau, D., & Schwartz, R. G. (2007). Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). *International Journal of Language & Communication Disorders*, *42*, 59-75.

Goldman, R., Fristoe, M., & Woodcock, R.W. (1974). *Goldman-Fristoe-Woodcock Auditory Skills Test Battery*. Circle Pines, Minn: American Guidance Service.

Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: a meta-analysis. *Journal of Speech, Language, and Hearing Research*, *50*(1), 177-195.

Grant, J., Valian, V., & Karmiloff-Smith, A. (2002). A study of relative clauses in Williams syndrome. *Journal of Child Language*, *29* (2), 403-416.

Gray, S. (2003). Diagnostic accuracy and test–retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*, *36* (2), 129-151.

Gupta, P., MacWhinney, B., Feldman, H. M., & Sacco, K. (2003). Phonological memory and vocabulary learning in children with focal lesions. *Brain and Language*, *87*(2), 241-252.

Gupta, P., & Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language*, *61*(4), 481-502.

Haber, R. N., & Schindler, R. M. (1981). Error in proofreading: Evidence of syntactic control of letter processing? *Journal of Experimental Psychology: Human Perception and Performance*, *7*(3), 573-579.

Hakuta, K. (1982). Interaction between particles and word order in the comprehension and production of simple sentences in Japanese children. *Developmental Psychology*, *18*(1), 62-76.

Hale, K., Laughren, M. & Simpson, J. (1995). Warlpiri. In J. Jacobs, A. von Stechow, W. Sternefeld & T. Vennemann (Eds.). *Syntax. Ein internationals Handbuch zeitgenossischer Forschung. [An international handbook of contemporary research] (* pp. 1430–51). Berlin: Walter de Gruyter.

Höhle, B., & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed class elements in continuous speech. *Developmental Science, 6,* 122–127.

Hoff, E., Core, C., & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20- to 24-month-olds. *Journal of Child Language*, *35*(4), 903.

Howard, D., & van der Lely, H. K. J. (1995). Specific language impairment in children is not due to a short-term memory deficit: response to Gathercole & Baddeley. *Journal of Speech and Hearing Research*, *38*, 466-472.

Howell, D. C. (1997). *Statistical methods for psychology (4th ed.)*. Belmont, CA: Wadsworth.

Hubbard, R. (1978). The probable consequences of violating the normality assumption in parametric statistical analysis. *Area*, *10*(5), 393–8.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*(6), 685-701.

Hulme, C., & Muir, C. (1985). Developmental changes in speech rate and memory span: A causal relationship? *British Journal of Developmental Psychology*, *3*(2), 175-181.

Hulme, C., Newton, P., Cowan, N., Stuart, G., & Brown, G. (1999). Think before you speak: pauses, memory search, and trace redintegration processes in verbal memory span. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(2), 447-463.

Hulme, C., & Roodenrys, S. (1995). Practitioner review: verbal working memory development and its disorders. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *36*(3), 373-398.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1217-1232.

Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution*. New York: Oxford University Press.

Jackendoff, R. (2007). A Parallel Architecture perspective on language processing. *Brain Research*, *1146*, 2-22.

Jacquemot, C., Dupoux, E., & Bachoud-Lévi, A.-C. (2011). Is the word-length effect linked to subvocal rehearsal? *Cortex*, *47*(4), 484-493.

Janda, L., & Townsend, C.E. (2000). *Czech. Number 125 In Languages Of The World/Materials*. Munich, Newcastle: Lincom Europa.

Jarrold, C., Thorn, A., & Stephens, E. (2009). The relationships among verbal short-term memory, phonological awareness, and new word learning: Evidence from typical development and Down syndrome. *Journal of Experimental Child Psychology*, *102*(2), 196-218.

Joanisse, M. F., & Seidenberg, M. S. (2003). Phonology and syntax in specific language impairment: evidence from a connectionist model. *Brain and Language*, *86*(1), 40-56.

Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., & Pine, J. M. (2010). Lexicality and frequency in specific language impairment: accuracy and error data from two nonword repetition tests. *Journal of Speech, Language, and Hearing Research*, *53*(6), 1642-1655.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*(1), 193-224.

Julien, M. (2002). *Syntactic heads and word formation*. Oxford, New York: Oxford University Press.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*(8), 1465-1476.

Kapalková, S. (2003). Pokus o aplikáciu MLU na podmienky slovenského jazyka [Adaptation of MLU to the properties of the Slovak language]. In V. Lechta (Ed.), *Logopaedica VI. Zborník Slovenskej asociácie logopédov (*pp. 35-39). Bratislava: Vydavateľstvo Liečreh Gúth.

Kapalková, S., Slančová,D., Bónová,I., Kesselová, J., & Mikulajová, M. (2010). *Hodnotenie komunikačných schopností detí v ranom veku [The evaluation of children's communication skills in early childhood]*. Bratislava: Slovenská asociácia logopédov.

Kemper, S. (1986). Imitation of complex syntactic constructions by elderly adults. *Applied Psycholinguistics*, *7*(3), 277-287.

Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes*, *22*(6), 860-897.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174.

Lasnik, H. (2005). Syntax. In P. Strazny (Ed.), *Encyclopedia of Linguistics, Volume 2* (pp.1068-1071). NY&Oxon: Fitzroy Dearborn.

Lee, L. (1974). *Developmental sentence analysis*. Evanston, Illinois: Northwestern University Press.

Leonard, L. (1998). *Children with specific language impairment*. Cambridge Mass.: MIT Press.

Leonard, L., Bortolini, U., Caselli, M.C., McGregor, K., & Sabbadini, L. (1992). Morphological deficits in children with specific language impairment: The status of features in the underlying grammar. *Language Acquisition*, *2*, 151-179.

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences*, *13*(3), 120-126.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, *32*(4), 451.

Lombardi, L., & Potter, M. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language*, *31*(6), 713-733.

Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *The Quarterly Journal of Experimental Psychology*, *53A*(1), 1-22.

Lust, B., Flynn, S. & Foley, C. (1996). What children know about what they say: elicited imitation as a research method for assessing children's syntax. In D. McDaniel, C. McKee & H. Cairns (Eds.), *Methods for assessing children's syntax* (p.55-76). Cambridge, MA: MIT Press.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*(1), 35-54.

Mainela-Arnold, E., & Evans, J. L. (2005). Beyond Capacity limitations: determinants of word recall performance on verbal working memory span tasks in children with SLI. *Journal of Speech, Language, and Hearing Research*, *48*(4), 897-909.

Marks, L., & Miller, G. (1964). The role of semantic and syntactic constraints in the memorization of English sentences. *Journal of Verbal Learning and Verbal Behavior*, *3*(1), 1-5.

Martin, N., Saffran, E. M., & Dell, G. S. (1996). Recovery in deep dysphasia: evidence for a relation between auditory - verbal STM capacity and lexical errors in repetition. *Brain and Language*, *52*(1), 83-113.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91-121.

McDade, H. L., Simpson, M. A., & Lamb, D. E. (1982). The use of elicited imitation as a measure of expressive grammar: a question of validity. *The Journal of Speech and Hearing Disorders*, *47*(1), 19-24.

Metsala, J. L., & Chisholm, G. M. (2010). The influence of lexical status and neighborhood density on children's nonword repetition. *Applied Psycholinguistics*, *31*(03), 489-506.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81-97.

Montgomery, J. W. (1995). Sentence comprehension in children with specific language impairment: the role of phonological working memory. *Journal of Speech and Hearing Research*, *38*(1), 187-199.

Montgomery, J. W. (2000). Verbal working memory and sentence comprehension in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *43*(2), 293-308.

Montgomery, J. W., Polunenko, A., & Marinellie, S. A. (2009). Role of working memory in children's understanding spoken narrative: A preliminary investigation. *Applied Psycholinguistics*, *30*(3), 485.

Morgan, J. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, *35*(5), 666-688.

Mukherjee, J. (2005). Word. In Philipp Strazny (Ed.), *Encyclopedia of Linguistics, Volume 2* (pp. 1184-1186). NY&Oxon: Fitzroy Dearborn.

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research*, *44*(4), 778-792.

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, *48*(5), 1033-1047.

Nespor, M., & Vogel, I. (2007). *Prosodic phonology. With a new foreword*. Berlin: Mouton de Gruyter.

Newcomer, P., & Hammill, D. (1997). *Test of Language Development-Primary 3rd Edition*. Austin, TX: PRO-ED.

Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Developmental Psychology*, *42*(4), 643-655.

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review*, *115*(3), 544-576.

O'Connell, D., Turner, E., & Onuska, L. (1968). Intonation, grammatical structure, and contextual association in immediate recall. *Journal of Verbal Learning and Verbal Behavior*, *7*(1), 110-116.

Office for National Statistics. *Census (2001)*. Retrieved April 18, 2011 from <http://www.statistics.gov.uk/census2001/census2001.asp>

Ottem, E., Lian, A., & Karlsen, P. (2007). Reasons for the growth of traditional memory span across age. *European Journal of Cognitive Psychology*, *19*(2), 233-270.

Pačesová, J. (1968). *The development of vocabulary in the child*. Brno: Universita J. E. Purkyne.

Palková, Z. (1994). *Fonetika a fonologie češtiny [Phonetics and phonology of Czech]*. Prague: Karolinum.

Palková, Z. & Volín, J. (2003). The role of F0 contours in determining foot boundaries in Czech. In *Proceedings of the 15th ICPhS, Vol. II* (pp. 1783-1786). Barcelona: Organizing Committee.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674-685.

Pickering, M., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences*, *3*(4), 136-141.

Pickering, S.J., & Gathercole, S.E. (2001). *Working Memory Test Battery for Children*. London: Psychological Corporation Europe.

Pinker, S. (1982). A theory of the acquisition of lexical interpretive grammars. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.

Poll, G. H., Betz, S. K., & Miller, C. A. (2010). Identification of Clinical Markers of Specific Language Impairment in Adults. *Journal of Speech, Language, and Hearing Research*, *53*(2), 414-429.

Postle, B. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23-38.

Potter, M., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*(6), 633-654.

Potter, M., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, *38*(3), 265-282.

Pring, T. (2005). *Research Methods in Communication Disorders*. London; Philadelphia: Whurr.

Ratner, N. B. (2000). Elicited imitation and other methods for the analysis of trade-offs between speech and language skills in children. In L. Menn & N. Ratner (Eds.), *Methods for studying language production (*pp. 289-309). Mahwah, N.J.: Lawrence Erlbaum Associates.

Redmond, S. M. (2005). Differentiating SLI from ADHD using children's sentence recall and production of past tense morphology. *Clinical Linguistics & Phonetics*, *19*(2), 109-127.

Reeves, C., Schmauder, A. R., & Morris, R. K. (2000). Stress grouping improves performance on an immediate serial list recall task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1638-1654.

Richardson, J. T. E. (2007). Measures of short-term memory: a historical review. *Cortex, 43*(5), 635-650.

Riches, N. G., Loucas, T., Baird, G., Charman, T., & Simonoff, E. (2010). Sentence repetition in adolescents with specific language impairments and autism: an investigation of complex syntax. *International Journal of Language & Communication Disorders*, *45*(1), 47-60.

Rispens, J. (2004). *Syntactic and phonological processing in developmental dyslexia*. Groningen: Rijksuniversiteit Groningen.

Rispens, J., & Parigger, E. (2010). Non-word repetition in Dutch-speaking children with specific language impairment with and without reading problems. *British Journal of Developmental Psychology*, *28*(1), 177-188.

Romani, C., Mcalpine, S., & Martin, R. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology*, *61*(2), 292-323.

Romani, C., Mcalpine, S., Olson, A., Tsouknida, E., & Martin, R. (2005). Length, lexicality, and articulatory suppression in immediate recall: Evidence against the articulatory loop. *Journal of Memory and Language*, *52*(3), 398-415.

Roodenrys, S., Hulme, C., & Brown, G. D. A. (1993). The development of short-term memory span: separable effects of speech rate and long-term memory. *Journal of Experimental Child Psychology*, *56*(3), 431-442.

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: evidence from typically developing children. *Journal of Speech, Language, and Hearing Research*, *47*(1), 223-234.

Rummer, R., & Engelkamp, J. (2003). Phonological information in immediate and delayed sentence recall. *The Quarterly Journal of Experimental Psychology: Section A*, *56*(1), 83-95.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926-1928.

Sahlén, B., Reuterskiöld-Wagner, C., Nettelbladt, U., & Radeborg, K. (1999). Non-word repetition in children with language impairment--pitfalls and possibilities. *International Journal of Language & Communication Disorders*, *34*(3), 337-352.

Scarborough, H. S. (1990). Very Early Language Deficits in Dyslexic Children. *Child Development*, *61*(6), 1728-1743.

Scholes, R. (1969). The role of grammaticality in the imitation of word strings by children and adults. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 225-228.

Scholes, R. (1970). On functors and contentives in children's imitations of word strings. *Journal of Verbal Learning and Verbal Behavior*, *9*(2), 167-170.

Seeff-Gabriel, B., Chiat, S., & Dodd, B. (2005). The relationship between speech disorders and language. In B. Dodd (Ed.), *Differential diagnosis and treatment of children with speech disorder* (pp. 100 - 116). Whurr Publishers: London.

Seeff-Gabriel, B., Chiat, S., & Dodd, B. (2010). Sentence imitation as a tool in identifying expressive morphosyntactic difficulties in children with severe speech difficulties. *International Journal of Language & Communication Disorders*, *45*(6), 691-702.

Seeff-Gabriel, G., Chiat, S. & Roy, P. (2008). *Early Repetition Battery (ERB)*. London, UK: Pearson Ltd.

Selkirk, E. (1996). The prosodic structure of function words. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 187–214). Mahwah, NJ: Lawrence Erlbaum Associates.

Semel, E., Wiig, E. & Secord, W. (1994). *Clinical Evaluation of Language Fundamentals - Revised*. San-Antonio, TX: Psychological Corporation.

Seung, H.-K., & Chapman, R. (2004). Sentence memory of individuals with Down's syndrome and typically developing children. *Journal of Intellectual Disability Research*, *48*(2), 160-171.

Shady, M. E., Gerken, L. A., & Jusczyk, P. W. (1995). Some evidence of sensitivity to prosody and word order in ten-month-olds. In D. MacLaughlin & S. McEwan (eds), *Proceedings of the 19th Boston University Conference on Language Development (Vol. 2)*. Somerville, MA: Cascadilla Press.

Shi, R.,Werker, J. F., & Cutler, A. (2003). Function words in early speech perception. In *Proceedings of the 15th International Conference of Phonetic Sciences* (pp. 3009–3012). Adelaide, Australia: Causal Productions.

Slobin, D., & Bever, T. G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, *58*, 265-289.

Small, J. A., Kemper, S., & Lyons, K. (2000). Sentence repetition and processing resources in Alzheimer's disease. *Brain and Language*, *75*(2), 232-258.

Snowling, M., & Stackhouse, J. (1983). Spelling performance of children with developmental verbal dyspraxia. *Developmental Medicine and Child Neurology*, *25*(4), 430-437.

Spencer, A., & Zwicky, A. M. (Eds.). (2001). *The Handbook of Morphology*. Oxford, UK: Blackwell Publishing Ltd.

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*(4), 598-605.

Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: the case of Cantonese. *Journal of Speech, Language, and Hearing Research*, *49*(2), 219-236.

Streiner, D. L., & Norman, G. R. (1995). *Health Measurement Scales: A Practical Guide to their Development and Use*. Oxford: Oxford University Press.

Tehan, G., & Humphreys, M. S. (1988). Articulatory loop explanations of memory span and pronunciation rate correspondences: a cautionary note. *Bulletin of the Psychonomic Society*, *26*, 293-296.

Tehan, G., & Lalor, D. M. (2000). Individual differences in memory span: the contribution of rehearsal, access to lexical memory, and output speed. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *53*(4), 1012-1038.

Thordardottir, E. (2008). Language-specific effects of task demands on the manifestation of specific language impairment: a comparison of English and Icelandic. *Journal of Speech, Language, and Hearing Research*, *51*(4), 922-937.

Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., et al. (2010). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research*, *54*(2), 580-597.

Thorn, A., & Page, M. (Eds). (2009). *Interactions between short-term and long-term memory in the verbal domain*. Hove, East Sussex; New York: Psychology Press.

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Development*, *75*(4), 1067-1084.

Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler & R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 53-66). Hamburg: Springer-Verlag.

Valian, V., & Aubry, S. (2005). When opportunity knocks twice: two-year-olds' repetition of sentence subjects. *Journal of Child Language*, *32*(3), 617-641.

van Daal, J., Verhoeven, L., van Leeuwe, J., & van Balkom, H. (2008). Working memory limitations in children with severe language impairment. *Journal of Communication Disorders*, *41*(2), 85-107.

van der Lely, H. K. J. (1990). *Sentence comprehension processes in specifically language impaired children* (Unpublished PhD thesis). London: University of London.

Vinther, T. (2002). Elicited imitation: a brief overview. *International Journal of Applied Linguistics*, *12*(1), 54-73.

Vitevitch, M., & Luce, P. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language*, *52*(2), 193-204.

Watkins, M. J. (1977). The intricacy of memory span. *Memory & Cognition*, *5*(5), 529-534.

Weist, R. M. (1983). The word order myth. *Journal of Child Language*, *10*(1), 97-106.

Wells, J., Christiansen, M., Race, D., Acheson, D., & Macdonald, M. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*(2), 250-271.

Wiig, E. H., Secord, W., & Semel, E. M. (2004). *Clinical evaluation of language fundamentals – preschool 2*. San Antonio, TX: Psychological Corporation.

Zimmer-Stahl, A. & Polišenská, K. (2009). *Short-term memory in speech perception and prosodic structuring*. Presented at Rhythm beyond the word Osnabrück, Germany.