BIROn - Birkbeck Institutional Research Online

Enabling Open Access to Birkbeck's Research Degree output

Mechanisms of intervention and the interface with clinical practice: the example of word-finding difficulties

https://eprints.bbk.ac.uk/id/eprint/50425/

Version: Full Version

Citation: Alireza, Hala Hussein (2022) Mechanisms of intervention and the interface with clinical practice: the example of word-finding difficulties. [Thesis] (Unpublished)

© 2020 The Author(s)

All material available through BIROn is protected by intellectual property law, including copyright law.
Any use made of the contents should comply with the relevant law.

Deposit Guide
Contact: email

Mechanisms of Intervention and the Interface with Clinical Practice:

The Example of Word-Finding Difficulties

Hala Hussein Alireza

Department of Psychological Sciences

Birkbeck, University of London

A thesis submitted for the degree of Doctor of Philosophy

April 2022

DECLARATION OF ORIGINALITY

I, Hala H. Alireza, declare that the work presented in this thesis is my own and that it has not been submitted before to any institution for assessment purposes.

Acknowledgement

I wish to express my gratitude to Professor Michael S. C. Thomas for supplying the computational model and for his assistance, unlimited patience and valuable advice throughout the course of this work.

I would like to thank Professor Wendy Best and the other members of the WORD project for their support and for allowing me to do secondary analysis on their data.

Abstract

This thesis investigated mechanistic approaches for understanding the effectiveness of interventions for developmental disorders of language and cognition, in order to contribute to narrowing the gap between theories of underlying causes of language deficits and intervention practices. Taking word-finding difficulties as an example language disorder, the project includes three methodological strands: (i) *Qualitative*: A questionnaire in which clinicians (speech and language therapists) were asked to identify ways in which research about causes of deficits did or did not inform their practice. (ii) *Quantitative*: secondary data analyses of two datasets collected as part of the WORD project (Best et al., 2015, 2017, 2021). Developmental trajectories were constructed for phonological, semantic, naming and comprehension skills of 20 children with WFD and 100 typically developing children (ages 4 years to 8;7). Regression analyses were used to assess whether phonological or semantic interventions were more effective in the group of children with WFD, and whether the patterns could be predicted from the children's language profiles. (iii) *Computational*: The Best et al. (2015) connectionist model of WFD was modified, and new training sets constructed, to run a sequence of simulations investigating optimal conditions for intervention. These tested whether better outcomes were obtained when intervention was targeted at remediating an area of weakness (the naming pathway) or buttressing a child's strengths (the structure of phonological and semantic representations) (Alireza, Fedor & Thomas, 2017; Thomas, Fedor, Davis, Yang, Alireza, Charman, Masterson & Best, 2019). Remediating weakness immediately accelerated development of the system but along the same atypical trajectory; improving strengths produced long-term increases in final vocabulary size; a combination yielded the best outcome. Overall, the thesis argues that both convergent methodologies, and a dialogue between basic researchers and clinicians, are necessary to narrow the gap between theories of deficit and clinical practice.

Table of Contents

Chapter 1

Introduction

This research project investigated the possibility of narrowing the gap between theories and treatment of developmental disorders. It fell within a larger project whose aim was to identify ‘what works’, in terms of intervention for developmental disorders of language and cognition, and to narrow the gap between theoretical understanding of the neurocognitive mechanisms underlying deficits, and clinical practice (Economic and Social Research Council funded WOrd Retrieval and Development [WORD] project; https://sites.google.com/site/wordfinding/). The broad strategy was to first investigate the mechanisms behind one type of disorder, Word-Finding Difficulties (WFD); and second, to explore the most suitable methods of intervention for WFD, with a view to develop general principles for effective methods of intervention for developmental disorders. This thesis focused on building links between computational models of vocabulary development and intervention methods for WFD, as well as exploring clinicians’ views on the relation of children’s language profiles to interventions.

By word-finding difficulties, we refer to children who have difficulty in the lexical retrieval of words they already know and understand. It is not regarded as a difficulty in comprehension since the words are usually in the child’s vocabulary (as assessed by standardised vocabulary tests). Dockrell, Messer, George and Wilson (1998) proposed that the difficulty may arise from mechanisms involved in retrieving the word at the time that it is needed. This difficulty can have a big impact on education outcomes and self esteem in particular (Best et al., 2015; Bishop, 2004).

Research on this topic looks at *semantic* and *phonological representations* and *processing speed* as the cognitive processes behind lexical retrieval difficulties (Messer & Dockrell, 2006). Semantics refers to the meaning of a word and phonology refers to the fundamental speech sounds of a language. For example, in this context, the key question is whether the problem is primarily due to difficulty in accessing semantic or phonological representations of the word when a child is presented with a corresponding picture. Data for behavioural intervention from both practitioners' accounts and a computational model are taken into consideration in this research.

Although there are emerging theories of underlying causes of language deficits, researchers such as Law, Campbell, Roulstone, Adams and Boyle (2008) and Michie and Prestwich (2010) have pointed out that intervention practices do not seem to be guided by those same theories. For example, in a survey exploring intervention practices for WFD, Best (2003) reported that 46% of therapists associated difficulties in phonological awareness with WFD and only 13% of therapists thought difficulties in semantics co-occurred with WFD. However 79% of therapists said they routinely used semantic approaches whereas only 13% said they routinely used phonological approaches in terms of intervention for WFD. At the very least, this suggests a tendency to bolster strengths rather than address weaknesses. In the same survey, when therapists were then asked, whether they would choose to work on a weakness or strength of the child, answers were quite varied, yet a distinct 21% of therapists indicated that they would use interventions that targeted the area of difficulty for the child - semantic work for children with apparently semantic difficulties and phonological work for children with phonological difficulties.

The apparent lack of guided procedures for adopting intervention strategies according to theories of causes of deficits may be due to the fact that the process of testing out behavioural interventions for children is usually very time consuming and is not practical if one wishes to test one component of a theory. A range of other reasons for the disparity between theories of intervention and theories of causes of deficits are discussed in a review paper by Thomas, Fedor, Davis, Yang, Alireza, Charman, Masterson and Best (2019). These include the characteristics of each child, which tend to guide therapists in selecting their intervention strategy; other comorbid difficulties the child may have and the availability of resources including time and money. A study by Alireza, Fedor and Thomas (2017) pointed out that other reasons for the gap could be that the primary focus of intervention is behavioural outcomes, for which it is not always necessary to understand the underlying cause of difficulty. In addition, in reality therapeutic situations involve treatment of the whole child, therapy is a complex social interaction with the therapist; and, even when there is a theory of deficit, the route to the best treatment is not necessarily obvious. The study was hence run to investigate the approach of ignoring the deficit and its causes and instead focus on using the strengths of the child for compensatory strategies.

A major focus of this thesis is the utility of computational modelling methods. The Alireza, Fedor and Thomas (2017) study is one example of employing computational models of development based on artificial neural networks (ANN) to investigate the mechanisms underlying language disorders, and explore the effects of specific types of intervention. This study is described in more detail in a later chapter. Computational modelling may be one way to narrow the gap between theories of deficits and clinical practice because it can establish and test causes of developmental deficits and then allow evaluation of possible interventions. Its main drawback as a method in psychology is its dependence on the simplifications required to implement working models. The Thomas et al. (2019) paper offers an extended review of computational modelling of interventions for developmental language disorders focussing at the mechanistic level as well as the behavioural level. The authors explore modelling interventions in four main sections: (i) long-term outcomes of intervention, (ii) single network models to remediate developmental weaknesses, (iii) interventions that encourage alternative compensatory pathways, and (iv) individual differences in response to intervention. This review is described in more detail in Chapter 3.

An early example of such computational models of development, which use ANN to investigate the effects of specific types of intervention, is the Harm, McCandliss, and Seidenberg (2003) model of reading development which was extended to assess an intervention for developmental dyslexia. More recently, Best and colleagues (Best et al., 2015) used a model based on the developmental trajectories approach where the disorder is set in the context of how the specific skills develop for typically developing children (Thomas & Karmiloff-Smith, 2002; Thomas et al., 2009). It was a simple connectionist model that was first set to simulate the mechanisms of vocabulary acquisition for normal or typical development. It was then modified to display atypical development and then a series of simulations were run to explore the effectiveness of intervention on different components of the theoretical cognitive processes. Results of the model were assessed against empirical data provided in the same study by running a randomised control trial with two participants who met the criteria for WFD. Both these models are described in more detail in a later section.

While the WORD project provides the context for the current thesis, the thesis has four important unique goals. First, it builds on data collected during the WORD project to characterise the developmental trajectories of typical and atypical vocabulary development through secondary analysis. Second, similarly through secondary data analysis, it evaluates the extent to which language profiles can predict response to interventions in children with WFD. Third, it leverages computational modelling methods to explore the implications of targeting weaknesses vs. strengths in language interventions. And fourth, it constructs a questionnaire study to draw on the expertise of speech and language therapists to help narrow the gap between theories of deficit and theories of treatment. Taking the developmental disorder of WFD, the thesis triangulates by revealing the underlying principles by which therapists select intervention strategies, characterising individual language profiles that predict the effectiveness of intervention, and exploring computational principles that shape the effectiveness of interventions. The ultimate aim of this research is, then, to seek to narrow the putative gap between intervention practices and the various theories of underlying causes of language deficits (Law et al., 2008).

This research project therefore had three methodological strands; (i) Qualitative: A questionnaire in which clinicians were asked to identify ways in which research about causes of deficits does or does not inform their practice was constructed. This was distributed to Speech and Language Therapists (SLT), special needs teachers and other relevant professionals. (ii) Empirical: Secondary data analyses of developmental trajectories of phonology, semantics, naming and comprehension skills of children with WFD and typically developing children, were made for the purpose of predicting which intervention type (semantic or phonological) would be most effective for treating each child with WFD. (iii) Computational: The Best et al. (2015) model of vocabulary development was modified and appropriate training sets were developed. A sequence of simulations were run on the modified Best et al. model to evaluate the relative merits of bolstering developmental strengths versus addressing developmental weaknesses to remediate problems of WFD.

The structure of this thesis will proceed as follows. Chapter 2 is a literature review on theoretical models of vocabulary development, developmental language disorders, word-finding difficulties and related intervention methods. Chapter 3 is a literature review of computational models of language development, computational models of developmental language disorders and computational models of interventions for developmental language disorders. Our focus will be on models of vocabulary development and the key mental representations underpinning it, namely semantics and phonology.

Chapter 4 presents the questionnaire study that aims to uncover any implicit principles applied by SLT when choosing intervention practices for children with WFD. The questionnaire was developed after having reviewed both a small survey of intervention methods for children with WFD (Best, 2003); and a series of email responses to a pilot questionnaire asking: “what techniques do you typically use to support children with word-finding difficulties”. Once inferences were made, based on the responses of the questionnaire, they were integrated into the computational modelling simulations of intervention part of the investigations of this study.

In Chapters 5 and 6, two sets of secondary data analyses of core language skills and relevant interventions are presented. These establish the target empirical effects to be captured by computational simulations employing the modified Best et al. (2015) model (Best et al., 2021). The first set of data in Chapter 5 compares cross-sectional developmental trajectories of typically developing and atypically developing language skills for children in each of four core skills namely: word production, word comprehension, semantics and phonology. In Chapter 6, the second set of analyses comprise regression analyses of intervention results from each of phonological and semantic interventions given to 20 children with WFD that were taken from the Best et al. (2021) study. Using a within participant design to compare interventions, all 20 children were given both a semantic and a phonological intervention for the same amount of time with a washout period in between the two types of intervention; half the children received phonological intervention followed by semantic intervention; and the other half received semantic intervention followed by phonological intervention. The regression analyses were carried out in pursuit of answering an extended list of questions given in more detail later in Chapter 6, with the broad goal of establishing the extent to which the most effective intervention for each child can be predicted from their profile of core skills.

Chapter 7 presents a description of the further development work completed on the Best et al. (2015) model, initially the building of appropriate training sets and comparing alternative architectures. Two sources were used to construct an appropriate training set of 397 words for the modified Best et al. (2015) model: a set of 1029 speaker generated feature norms for 456 words collected by Vinson and Vigliocco (2008) from 280 adults; and the Children's Printed Word Database (CPWD), which is an online database of the vocabulary in reading materials used by 5-9 year old children in the United Kingdom (Masterson, Stuart, Dixon & Lovejoy, 2010). A smaller set of 57 items taken from Small et al. (1996) was initially used to compare two different architectures of the improved model in order to choose the best option for the purpose of our research.

Chapter 8 provides an account of the Alireza, Fedor and Thomas (2017) study, in which intervention simulations were run on the modified Best et al. (2015) model to explore the notion of intervening to target strengths versus weaknesses to remediate problems of WFD. The goal was to simulate mechanisms of vocabulary acquisition for typical development, then modify the model by changing one or more start state components assigned to represent certain mechanistic elements (for example the number of internal processing units) thereby generating atypical development. A series of simulations are then presented to explore the effectiveness of different behavioural interventions implemented mainly on the training sets. A key question explored using this method was: Should therapy be focussed on a weak skill to try to remediate a weakness or focus on the child's strengths to try to find a compensatory skill to work with instead? For example would working on repetition of phonological elements produce better results than working on the semantic aspect, the latter generally believed to be a relative strength for children with WFD?

Finally, Chapter 9 considers what has been learned in the three strands, and the prospect of narrowing the gap between basic research and clinical practice in the area of developmental language disorders.

Chapter 2

Empirical Literature

Vocabulary Development

This chapter starts with a brief overview of typical language development. Since the focus of this study is intervention for WFD, the emphasis will be on vocabulary development, and its underpinning mental representations, and will not include grammatical aspects of language development like morphology and syntax. Hence a literature review on theoretical models of vocabulary development and a section on key factors that influence word acquisition will follow. Next will be a section on developmental language disorders and an explanation of who should be included under this diagnosis, along with proposed subgroups. The review will then narrow the focus to describe word-finding difficulties and to discuss their potential causes and effective intervention methods.

Overview of Typical Language Development. At 7 months babies start to babble, by one year they can usually say some meaningful words, by 18 months they have an average of 50 words in their vocabulary and by the age of 2 years they can produce multiword utterances (Levelt, 1999). In their third year, they use syntactic properties (begin expressing syntactic relations by means of prepositions, auxiliaries, inflections and word order) and improve sentences and word order. The architecture of this cognitive system is usually in place by the age of 5 or 6 years (Levelt, 2001). Here our focus is on vocabulary acquisition and its key properties.

Models of Vocabulary Acquisition. Lexical retrieval is generally described as ‘the process of getting from a concept to a spoken word’ (Friedmann, Biran & Dotan, 2013, p. 350). We acquire a mental lexicon of around 50,000 words (Miller, 1991), we say on average 2-3 words per second and we make 1 or 2 errors on average every 1000 words (Levelt, 1999). Traditionally, studies of word production focussed on speech error analysis and chronometric (accurate time measurement) studies of naming latencies (Levelt, 1999). The tradition of speech error analysis goes back to 1895 when Meringer and Mayer collected and published a corpus of German speech errors. Their theoretical analyses made an important distinction between substitutions that are based on meaning and those that are based on form (Levelt, 1999). They noted that meaning-based errors regularly included phonological based connections. The tradition of measuring latencies to naming objects has also been around since1885 when Cattell found that naming a list of 100 line drawings of objects took twice as long as naming a list of the corresponding printed names of the object (Levelt, 1999).

Analysis of error types and reaction times led to the development of a number of models of adult lexical retrieval (Dell, 1986; Fromkin, 1973; Levelt, 1999). Most of these models postulate that lexical retrieval is a multi-stage process with distinct components and the general agreement is that the process comprises three main areas: 'conceptualization', 'formulation' and 'encoding for articulation' (Friedmann, Biran & Dotan, 2013). Conceptualization involves deciding what to say, and is sometimes referred to as the message level of representation. Formulation involves two parts, selecting the word that we want to say and then selecting the appropriate syntax as required for the semantic context to put into a sentence. Finally, the process of encoding involves turning the word into a phonological sound using the appropriate syntactic form.

An early example of an adult model of word production is Garrett's (1975) serial processing model. On the basis of error analyses, Garrett proposed that each stage of word production takes place only after the previous stage has finished. Thus, syntactic planning at the *lexeme* (or lexical) level only starts after the word has been determined at the *lemma* (or comprehension) level. Most models of lexical retrieval refer to the lemma level as the stage of word production for which the word is semantically and syntactically specified but not yet phonologically assigned; the next stage is referred to as the lexeme level and this is where the phonological form of the word is theoretically specified (Harley, 2001; Levelt, 1992). While Levelt's lexical access theory was inspired by speech error evidence, it has subsequently been developed largely based on analyses of reaction time data, for example using picture naming. The architecture of the model has two distinct components: one for what he calls 'lexical selection' and one for 'form encoding' (Levelt, 2001, p.13464). Dell's (1986) Interactive activation model is based on the mechanism of spreading activation from one level to another. Activation spreads down from a syntax level to a morphology level and finally to a phonological level. Several units may be activated at the same time and there is feedback between levels; thereby the model gives a good account of speech errors, such as phonological errors (for example *cancake* instead of *pancake*), semantic errors (for example saying *spoons* instead of *knives*) and mixed errors (for example *stop* instead of *start*) (Dell, 1986).

Key Properties of Vocabulary Acquisition. Factors found to influence word naming (or lexical retrieval) in typically developing children include word frequency, age of acquisition, lexical neighbourhood and stress pattern (Coady, 2013; Newman & German, 2002). For example, a study by Newman and German (2002) included 320 children aged 7 to 12 years, of which 273 children had typically developing language abilities and 58 children were diagnosed with word-finding difficulties. Newman and German assessed accuracy of naming on the picture-naming task and on the sentence completion task (the child was asked to complete each sentence with one word). They found that words that occur more frequently in the language are more easily accessible than less frequent words and that this phenomenon does not appear to change with age. Age of acquisition also influenced word retrieval but the effect decreased over time.

Words that were learned earlier in life or were known for a longer period were found to be more easily accessible than words learnt more recently but as the children became older this has less impact for typical development. The implication is that the longer the word has been in the child's vocabulary the more complex the lexical access and the more lexical pathways to the word. However, this pattern was less obvious for children with WFD. Children with WFD did not find it easier to access words that they have known for longer periods, suggesting some other limiting factor on lexical access.

The study by Newman and German (2002) also showed that lexical neighbourhood had an effect on retrieval but in three different ways. (i) Words which have more neighbours are apt to be harder to access than words with fewer neighbours; (ii) At the same time average neighbourhood frequency had a facilitative effect on lexical retrieval; And (iii) relative to the target word, neighbouring words that were moderately frequent had a facilitative effect whereas neighbouring words that were more frequent competed with target word for selection and were proposed to have an inhibitory effect. The study considered stress pattern effects on lexical retrieval and suggested from preliminary results that words that had a typical stress pattern were easier to retrieve than those that did not. For example, a typical stress pattern is that for a two-syllable word the stress falls on the first syllable, e.g. '*o*bject' = physical entity; a less frequent pattern is for the stress to fall on the second syllable, e.g. 'ob*ject'* = argue with. The presence of word frequency and neighbourhood frequency effects for children with WFD that were similar to typically developing children was taken to imply that difficulties could be in the retrieval process rather than in the storage mechanism (German & Newman, 2004).

Developmental Language Disorder and Word-Finding Difficulties

Developmental language Disorder (DLD) has been studied for 200 years and was previously known by various names such as developmental aphasia, developmental language impairment and most recently specific language impairment (SLI). The term SLI has until recently been used, mostly by researchers, to refer to children who have significant limitations in language ability that cannot be attributed to problems in hearing, neurological status, nonverbal intelligence, or other known factors (Leonard, 2014). In reality children with language difficulties tend to have a range of comorbid developmental disorders, which lead to difficulties in diagnosis (Bishop, 2004). The term SLI has thus been scrutinized for the past few years because of the implication that there are no other problems except language (Bishop, 2004, 2014; Ebbels, 2014a). In a review paper, Bishop (2004) pointed to the enormous difficulty, in practice, of deciding who should be diagnosed with SLI and to the notion that different diagnostic labels could be given to the same child depending on the type of specialist implementing the assessment. As Bishop (2004) noted, 'The same child might receive a label of SLI from a speech-language pathologist, dyslexia from a school psychologist, ADHD from a pediatrician, PDDNOS from a child psychiatrist, right-hemisphere learning disability from a neuropsychologist, and developmental coordination disorder from a physical therapist' (p.316). Although a label is important in order for children to receive the most appropriate intervention, she also noted that one should be aware not to exclude children, who have co-morbid disorders with developmental language difficulties, from interventions that would be available for children diagnosed only with SLI (Ebbels, 2014a). This issue was addressed by an international panel of 57 professionals (the CATALISE panel) recruited from ten disciplines which were relevant to the field such as speech and language therapists, specialist teachers, psychologists, psychiatrists, pediatricians, and educational psychologists (Bishop, Snowling, Thompson, Greenhalgh & The CATALISE Consortium, 2016). A consensus was reached to use the term *Developmental Language Disorder* instead of *Specific Language impairment* for children

who have language difficulties that cannot be explained by anything else. The panel also agreed on a number of guidelines to facilitate the diagnosis of DLD for professionals and to allow access to intervention by more children as needed. The two most important guidelines included were: (1) DLD can be used as a diagnosis for children who's language difficulties are associated with other neurodevelopmental disorders such as attention (e.g., Attention Deficit Hyperactivity disorder [ADHD]) or motor (e.g., dyspraxia); (2) DLD can be diagnosed even if there is no discrepancy between verbal and nonverbal ability (Bishop et al., 2016; Ebbels, 2016). A detailed account of the procedure for reaching the consensus on DLD and other diagnostic guidelines are provided in the paper by Bishop et al. (2016). For consistency in this thesis, and in accordance with the new consensus, from here on the term DLD will be used in place of SLI or any of the other names used in previous research to describe similar language difficulties.
Prevalence of DLD is about 7% of the population (Leonard, 2014). Although moderate to high heritability has been detected, research has so far shown that the influence of genes on DLD depends on a complex interaction of multiple genes and the environment, as is the case with other cognitive abilities (Bishop, 2009; Leonard, 2014). The most common difficulties that are noted are with grammar, morphology and phonology, semantics and pragmatics, leading to problems in reading, social and emotional areas (Leonard, 2014).

Since DLD is not a homogenous disorder, researchers endeavour to identify subgroups in pursuit of finding the appropriate intervention (Bishop, 2004). A classic categorisation was proposed by Rapin and Allen (1983) and is based on a classification of three main categories: (i) *mixed receptive/expressive* disorder, which leads to problems in language comprehension and production; (ii) *expressive* disorder, which refers to organisational aspects of speech; (iii) *higher order processing* disorders (Rapin, 1996). The third category refers to lexical difficulties and semantic pragmatic deficits, which make conversation difficult. Alternatively, one might take a data-driven approach. Botting and Conti-Ramsden (2004) used cluster analysis of results of psychometric clinical classifications of 242 seven-year-old children with DLD. They drew out six subgroups, of which five were straightforwardly related to the categorisations proposed by Rapin and Allen. However, a follow-up study found that children changed profiles when tested again. They were still diagnosed with DLD but did not necessarily show the same profiles as before, suggesting that DLD is a dynamic condition that changes with developmental time.

Continuing the quest to identify the most suitable interventions for DLD, some researchers focussed on the notion that an early auditory processing deficit could be a mechanistic cause of DLD. For example, Benasich and Tallal (2002) proposed that children who have difficulties in discriminating rapidly changing auditory cues early in development will subsequently grow to develop DLD. This view implies that difficulties originate in a phonological component of language acquisition preceding a semantic aspect, as most children will also develop difficulties in comprehension. The authors suggested that a deficit in rapid auditory temporal processing may be taken as a 'marker' for early identification of infants who will later develop a language delay, especially for those considered to be at high risk of developing DLD based on family history.

On the other hand, according to results of a study by Corriveau, Pasquini and Goswami (2007) the majority of children who have auditory processing difficulties and go on to develop DLD have difficulty in processing cues related to speech rhythm and syllable stress patterns. This would, in their view, explain potential difficulties that the children would have in the development of phonological representations. Yet an earlier study by Wright and colleagues (1997) suggested that children with DLD had difficulty in discriminating rapidly changing sound with similar frequencies. They proposed that children with DLD could benefit from having a diagnosis based on auditory testing and then having auditory training with a focus on their weaknesses.

In a review paper, Tallal and Gaab (2006) pointed to the presence of conflicting data from various behavioural studies as well as auditory event-related potentials (ERP) such as those found by Bishop and McArthur (2005) on the mechanistic cause of language delay. The study supports the suggestions that children with DLD may have had an auditory-processing deficit at an earlier stage in development that improves as they mature (Bishop & McArthur, 2005). Yet Tallal and Gaab (2006) proposed that training on rapid auditory sequencing and acoustically modified speech provided by intervention programmes such as ‘Fast ForWord’ would be beneficial for improving language abilities of children with DLD. However a meta-analysis study of the ‘Fast ForWord’ language intervention programme found no evidence of effectiveness of training on improving children’s reading or oral language (Strong, Torgerson, Torgerson & Hulme, 2011). Furthermore, having reviewed evidence from a number of studies Leonard (2014) concluded that, in general, indications that difficulties in auditory processing could be the mechanistic weakness leading to DLD are no stronger than those for other impairments that the children appear to have.

Word-Finding Difficulties. Children who have difficulty in the lexical retrieval of words they already understand are described as having word-finding difficulties (WFD). According to Dockrell, Messer, George and Wilson (1998), 23% of children in language support services had WFD, whereas 50% of learning disabled children were identified as having WFD. They added that children with dyslexia often exhibit WFD. The main characteristics of WFD include long delays in word retrieval, frequent circumlocution, use of substitutions, use of filler words such as ‘uh, er, uhm’, overuse of undefined words such as ‘stuff’ or ‘thing’ or repeating words (Dockrell et al., 1998;

German & Simon, 1991). Dockrell and colleagues point out children with these symptoms may become either quiet and non-communicative, or talk too much using fillers to compensate for the difficulty in finding the right word.

The most widely used tests to assess WFD were developed by German (1989) and they are the only ones that have U.S. national norms (Messer & Dockrell, 2006). These include The Test of Word-Finding, second edition (TWF-2; German, 1986, 2000), The Test of Adolescent/Adult Word-Finding (TAWF; German, 1990) and The Test of Word-Finding in Discourse (TWF-D; German, 1991). The TWF-2 and TAFW assess picture naming, different word classes, sentence completion, description naming and naming to categories. These tests also verify that the child actually comprehends the assessed words. TWF-D requires the child to make up stories about picture stimuli and the narratives are assessed for word-finding behaviours. Several non-standardised tests are used for research. The Word Naming Test (Weigel-Crump & Dennis, 1986) was originally used to test children with brain injuries by assessing picture naming, word definitions and rhyme prompts. The Boston Naming Test (Kaplan, Goodglas & Weintraub, 1976) assesses accuracy rather than speed of naming. The Rapid Automised Naming test (RAN; Wolf & Denckla, 2005) involves rapid naming of 50 stimuli, where five symbols in a given category are presented ten times in random order; the categories involve letters, numbers, colours or objects. The Rapid Automatic Switching test (RAS; Wolf, 1986), which was developed from the RAN to assess children with dyslexia, alternates between letters and numbers. Both the RAN and RAS assess speed of naming or reading so they measure automaticity, which is considered to rely on efficient lexical retrieval (Messer & Dockrell, 2006).

However, Tingley, Kyte and Johnson (2003) questioned whether difficulties in single word naming and frequent conversational disruptions are caused by the same construct, i.e., the “WFD”. They used total time from RAN (Wolf & Denckla, 2005), standard score from the Brief Test of TAWF (German, 1990) and total unique words from the Controlled Oral Word Association Task (FAS; Benton & Hamsher, 1978). Tingley and colleagues proposed that single-word performance did not translate into differences in levels of conversational fluency. Lahey and Edwards (1996) also questioned the concept of WFD and suggested the existence of different subgroups of children with naming difficulties. They suggested that some children have *expressive* only difficulties and that intervention should be focussed on phonological therapy; and some have both *receptive* and *expressive* difficulties so intervention should focus on semantic and phonological therapy (Lahey & Edwards, 1999).

Studies of cognitive processes behind lexical retrieval difficulties often investigate word accuracy (Newman & German, 2002), error patterns - for example semantic (Sheng & McGregor, 2010) versus phonological (Constable, Stackhouse & Wells, 1997) - and speed of retrieval by measuring reaction times (Dockrell, Messer & George, 2001). Both semantic errors (McGregor, 1997) and phonological errors (Dockrell et al., 2001) have been found to be common for children with WFD. However, McGregor (1994) pointed out that semantic errors could arise when a failure occurs in retrieving the correct word because of a difficulty in the phonological retrieval aspect and consequently a semantically similar word is activated instead, giving the appearance of a semantic error, despite the phonological cause. Hence error studies may not provide conclusive evidence for underlying cause of difficulty.

Studies that support the notion that WFD stems from problems in phonological representations include a study by German and Newman (2004) on lexical factors that affect lexical retrieval, which pointed to a deficit in the organisation of the phonological lexicon rather than the semantic component. Yet Messer, Dockrell and Murphy (2004) point to phonological awareness and decoding as a relative strength for children with WFD. In contrast another study by Sheng and McGregor (2010) supports the idea that WFD stems from problems in the lexical-semantic organisation for children with DLD.

In line with the idea that WFD stems from a problem with semantic representations, a study by Dockrell et al. (2001), which used a picture-naming task, reported that children with WFD were slower at naming than control groups. Interestingly, in the same study children with WFD seemed to be slower at naming pictures and colours but were not slower at naming numbers or letters. Pictures and colours are considered to have complex semantic representations whereas letters and numbers are considered to be overlearned and require minimal access to semantic representations during naming. Furthermore children with WFD had the longest latencies for high frequency objects, and low and high frequency actions, compared to chronologically age matched children. Taking into account that naming accuracy scores of children with WFD for high frequency objects were close to being significantly different from scores of children matched for their naming accuracy (as assessed by the British Abilities scale (BAS, Elliot, Smith & McCulloch, 1997), the authors suggested that this indicates that children have a specific difficulty in accessing representations even when they retrieve them accurately. The implication is that difficulties in retrieval could be due to children with WFD having less complex semantic representations than typically developing children, rather than the difficulties being due to slow access. This notion is corroborated by the notably low scores of children with WFD for semantic fluency (Dockrell et al., 2001). It was also found that children with WFD had more errors naming colours and letters than numbers.

Hence, in a 2006 review paper, Messer and Dockrell argued that WFD are caused by underlying cognitive processes that are heterogeneous in nature. Given that most models of lexical retrieval postulate that word production is a complex multi-stage process, Messer and Dockrell (2006) proposed that problems in word-finding could potentially stem from any of the different stages or locations in the process. They also noted that the same child could have a word-finding problem stemming from a different location for different words. The authors indicated that difficulties in word-finding may involve different combinations of deficits in semantic and phonological representations as well as in speed of processing. They recommended that research methods that include comparison of word production abilities of children with WFD with those of children who are matched for their language level, rather than only looking at the comparison with children matched with their chronological age, could be more useful in uncovering underlying causes of WFD (Dockrell et al., 2001; Messer & Dockrell, 2006).

More recently, event related potential (ERP) studies were used to explore lexical access because of their sensitivity to timing. Costa and colleagues used the idea of 'cumulative semantic interference effect' time of lexical retrieval as a key measure. In a 2009 study, they showed that upon seeing a picture, lexical retrieval takes approximately 200ms (Costa, Strijkers, Martin & Thierry, 2009). In another study a positive correlation was observed between mean amplitude of the P2 component and naming latencies (Strijkers, Costa & Thierry, 2010). Taking into account word frequency and cognate effects on ERP the authors argued that the P2 component is sensitive to lexical access rather than phonological retrieval. Hence the authors proposed that ERP studies can potentially be a valuable tool used to investigate underlying mechanisms of word production and thereby shed some light on causes of problems in lexical retrieval (Strijkers, Costa & Thierry, 2010).

Intervention Methods for Developmental Language Disorder and Word-Finding Difficulties

Efficient diagnosis and intervention is essential in the case of DLD because it affects decision making for the provision of educational and remedial funds as well as policies concerning health care. As Ebbels (2014b) put it 'The ultimate goal of intervention research is to establish which method is the most effective, for which areas of language, for which children, using which method of delivery' (p.8). Treatment, therapy or remediation are all words that are commonly used interchangeably to refer to intervention (Roulstone, Wren, Bakopoulou & Lindsay, 2012).

Although there are numerous intervention treatments (Ebbels, 2014b; Roulstone, Wren, Bakopoulou & Lindsay, 2012) they are generally based on two main ideas: (1) repetition, and (2) making sure that target words are used in a context that makes the meaning obvious to the child (Leonard, 2014). Common intervention approaches for DLD and WFD include the following five treatments (Leonard, 2014):

(i) *Imitation,* for example, the therapist shows a picture to the child along with a question acting as a prompt, and saying a word or a phrase, which should be exactly repeated by the child. Later the therapist will just show the picture along with the prompt;

(ii) *Modelling*, for example the child is asked to first observe and then give a new example using the same rule of combining, sequencing or inserting morphemes but not necessarily the exact words;

(iii) *Focused* stimulation is similar to modelling but the frequency of exposure is increased to particular semantic or morphosyntatic forms.

(iv) *Milieu teaching* relies on the child's surrounding to set the teaching where settings are arranged to increase the likelihood of the child communicating, but the therapist will let the child choose what to talk about;

(v) The *Conversational recasting* approach includes play activities where the adult lets the child speak while playing and then responds by saying the same phrase in the correct form. The idea is to do the treatment on phrases that the child is interested in (Leonard, 2014).

Other popular intervention programs include computer games based on the theory that children with DLD have difficulties in recognising some rapidly successive phonetic sound stimuli that affect the development of phonological representations (Tallal et al., 1996). Word webs are a particularly popular option for therapists to treat children with WFD. This method requires the therapist to use words that are either semantically (for example a word from a similar category) or phonologically (for example a word that rhymes) related to the target word to prompt the child to name a picture (Best, 2005; Best et al., 2015). Examples of semantic and a phonological word webs are shown in Figure 1.

(a) Semantic Therapy **(b) Phonological Therapy**

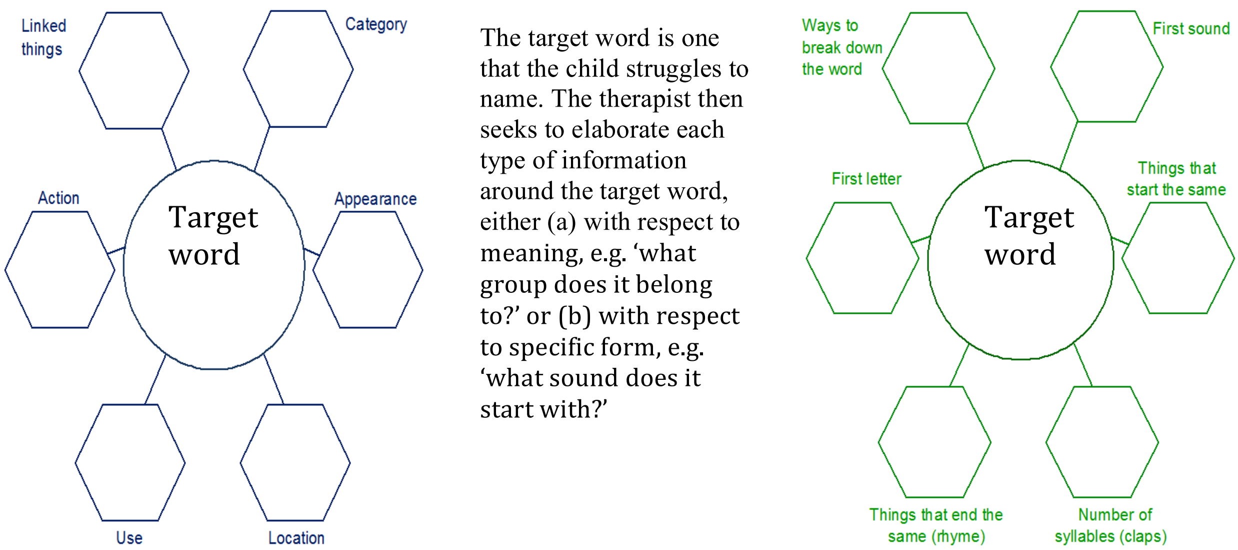


Figure 2.1. Word webs used for phonological and semantic therapy.

Reprinted from W. Best, A. Fedor, L. Hughes, A. Kapikian, J. Masterson, S. Roncoli, L. Fern-Pollak and M. S. C. Thomas (2015). Intervening to alleviate word-finding difficulties in children: case series data and a computational modelling foundation. *Cognitive Neuropsychology, 32*(3-4), 133-168. Copyright (2015), the authors.

DOI: 10.1080/02643294.2014.1003204

A summary of the therapy protocol is given in appendix 3 in Best et al. (2015).

Effective intervention. A study to explore the reasons for choice of intervention methods and their effectiveness conducted for the Better Communication Research Programme (BCRP) found that there was a lack of strong evidence of efficacy for most of the programmes used by speech and language therapists (Law et al., 2012). The BCRP targeted children whose primary difficulty is that of language as opposed to children who have a language difficulty as a by-product of other difficulties such as hearing impairment. The report suggested that the collection of further evidence is hindered by the fact that practitioners seem to be adapting and developing programmes according to local needs (Roulstone, Wren, Bakopoulou & Lindsay, 2012). However, stronger evidence supports the efficacy of specific intervention activities and principles practiced such as auditory discrimination activities and forced alternatives approaches (Warren & Yoder, 2004). Another survey, targeted towards therapists who work with children with WFD, raised the issue of co-occurring difficulties (Best, 2003). Phonological awareness was the most frequently mentioned learning difficulty associated with WFD. In the same survey, therapists generally commented that they tended to work on children's weaknesses rather than their strengths. Nevertheless 79% of therapists then ranked a semantic approach at the top of their list despite phonological problems being the most frequent deficit co-associated with WFD. Hence, further research to identify causes and appropriate interventions for the various cases of WFD was recommended by the author. So far research shows that both semantic and phonological therapies were found to improve WFD to a certain extent (Best, 2005; Bragard, Schelstraete, Snyers & James, 2012; Easton, Sheach & Easton, 1997), although they mainly improved retrieval of words used in therapy. The usual goal of therapy in the field of WFD is to use an appropriate intervention that has an effect which both generalises to words not included in the therapy and continues after therapy has ended (Green, 2020; Thomas et al., 2019). Few have met this benchmark reliably.

In a review paper that focussed mostly on the effectiveness of intervention for grammar in school aged children, Ebbels (2014b) found that most interventions were successful while therapy was going on but were not maintained after therapy had stopped. Interestingly another study (Tyler, Lewis, Haskill & Tolbert, 2003) showed better results when therapy *alternated* weekly between a focus on phonology and morphosyntax. The study by Ebbels (2014b), as well as the one by Law (2004), highlighted the lack of literature on effective intervention for children with receptive as well as expressive language difficulties (Boyle, McCartney, O'Hare & Law, 2010). Likewise in a more recent review paper, which looked at persistence and fade-out of educational intervention effects, the authors pointed out that a relatively small number of studies actually include long-run follow up assessments (Bailey, Duncan, Cunha, Foorman &Yeager, 2020). Bailey and Colleagus (2020) also concluded that persistence depends on the types of skills being taught (of which *building skills* being more persistent), constraints and opportunities within the institutional and social context as well as compatibility between the interventions and the environment.

Best (2005) conducted an intervention study for five children with WFD where the therapy used was a computerised aid that converts letter into sounds. Best identified four possible deficits based on the general model where conceptual information accesses semantic information, which in turn accesses phonological forms for production. Taking into account that it was a small study, the author proposed that outcomes of the therapy suggest that difficulties occur in accessing the phonological form for production and in storing phonological information for production. Best suggested that the therapy had the effect of 'strengthening the links between lexical semantic and phonological forms for production' (Best, 2005, p. 306) and 'helping storage of correct phonological form for production' (Best, 2005, p.306).

More recently, for the first part of the WORD project, a parallel group randomised control trial was run to investigate the effectiveness of intervention for WFD using word-webs intended to elaborate word knowledge (Best et al., 2017). Therapists used either phonologically or semantically based word-webs to prompt children diagnosed with WFD during one-to-one sessions (two examples of the word-webs are shown in Figure 1). For this part of the trial, therapeutic intervention was in the form of half hour sessions once a week for six weeks. Effectiveness of the intervention was measured by the children's ability to name 100 words by looking at pictures. Only 25 of the words were treated in the sessions, 25 words were named each session but not treated and 50 words were not previously seen. Results were compared between 11 children who had received the treatment and 9 children who were going to do the treatment at a later stage and whose results acted as a control group. Although the word-web intervention was found to be effective for treated items, it did not generalise to untreated words.

Similar results were found for the completed version of the same study, which compared the different outcomes of phonological and semantic intervention for both a group comparison and case series analyses (Best et al., 2021). In this part of the study, all 20 children diagnosed with WFD and aged between 6 and 8 years, were given both semantic and phonological intervention in the form of word-webs completing six-week interventions of each type of therapy with a six-week 'wash-out' period in between. The effectiveness of the intervention was again measured by the children's ability to name 100 items from pictures chosen specifically for the experiment. Details of the assessment procedure can be found in Best et al. (2017). Each child was assessed three times before the start of therapy and the average score was taken as the baseline score to be compared with the score at the end of therapy. The experiment was designed as a randomised controlled trial, with a wash-out period in between the two phases of intervention. A further objective of this study was to investigate whether one type of intervention (phonological versus semantic intervention) could be predicted to be more effective than the other according to the language profile presented by the child. For this reason the language profiles of children taking part in the study were categorised into three different sub-groups: (i) *Classic* WFD who were predicted to benefit from either or both semantic and phonological intervention, (ii) *semantic difficulties* who were predicted to benefit only from semantic intervention, and (iii) *phonological difficulties* (in the context of strong semantic processing) who were predicted to benefit only from phonological intervention and not from semantic intervention. Details of the experimental design and procedure are provided in Best et al. (2021) and explained in detail in the secondary data analyses in chapter 6. The strategy behind this intervention study was aligned with the notion that the causes of WFD may be heterogeneous in nature and were most likely to be due to problems in semantic or phonological representations, processing speed or even any combination of these cognitive processes (Messer & Dockrell, 2006). Different therapies were expected to work differently on each child based on the child's profile of four core language skills (word production, word comprehension, semantics and phonology) on which the children had been tested before intervention was started (details are in chapter 6). The combination of results from the different interventions, along with the results of the four core language skills, could in theory reveal the origin of the underlying cause of WFD in each case.

The Best et al. (2021) study found that WFD can be improved using both phonological and semantic word-webs but the semantic intervention was more effective than phonological intervention. Hence the authors recommended that resources for intervention should be directed towards semantic intervention rather than phonological intervention. In line with the heterogeneous character of WFD, children responded differently to the same intervention (Messer & Dockrell, 2006). Further analysis of the responses of the sub-groups of children to the two interventions in this study led the authors to reason that the division into sub-groups according to difficulties can lead to predictions of outcomes depending on which therapy is applied. Although in terms of generalisation, results were similar to those found in the first part of their study (Best et al., 2017), both phonologically and semantically based word-web interventions were found to be effective for treated items but neither one generalised to untreated words. A major limitation of the Best et al. (2021) study was that since difficulties were categorised, the size of intervention effects and the differential effect between phonological and semantic therapy could not be examined based on profile. Furthermore, within the design of this study there is no link to the developmental trajectories of these skills, in a form that can be simulated by developmental computational modelling.

Hence, endeavouring to augment the outcome of the Best et al. (2021) study, in Chapters 5 and 6 we present two sets of secondary data analyses. Both sets of data were taken from the WORD project, which comprised a series of studies to investigate the most effective methods of intervention for WFD (Best et al., 2015; Best et al., 2017; Best et al., 2021). Chapter 5 comprises an analysis of developmental trajectories of core abilities, which are word production, word comprehension, semantics and phonology. The secondary data analysis offered in Chapter 6 offers in-depth analyses employing regression models of the empirical data obtained in the Best et al. (2021) study. The regression analyses treat intervention effects as continuous variables, and seek to like these to cognitive profiles, thereby linking the developmental trajectories of these skills, in a form that can be simulated by developmental computational modelling. Another added value of the secondary analyses is in the notion that the size of intervention effects and the differential effect between phonological and semantic therapy can thus be examined based on profile. Although the regression analyses incorporate a continuous approach to explore effects of different interventions, results can however be compared to those derived in the Best et al. (2021) study where difficulties were categorised. Further details on the background of the intervention study are given at the end of chapter 3 following a review of computational modelling of intervention.

Chapter 3

Computational Modelling Literature

Computational Models of Language Development

Although there are various types of adult models of word production, there are relatively fewer developmental models of lexical retrieval. It is more difficult to investigate the production of language than comprehension because ‘it is difficult to control the input in experiments on production’ (Harley, 2001). Computational models that use artificial neural networks (ANN), referred to as connectionist modelling, have been used to simulate learning, and have therefore been notably useful in making advances in models of development of language and cognition (Li, Zhao & MacWhinney, 2007; Mayor & Plunkett, 2010; McMurray, Zhao, Kucker & Samuelson, 2013; Plaut & McClelland, 1993; Plunkett, Sinha, Møller & Stransby, 1992). Although it is essentially a theoretical top-down approach, when used to test empirical data, the models are usually constrained to try to accommodate empirical results thereby combining a bottom-up with a top-down approach (Harley, 2001).

One of the earlier connectionist models of vocabulary acquisition that used the idea of a connectionist autoassociator was the Plunkett, Sinha model (Plunkett, Sinha, Møller & Stransby, 1992). The model was made to form concepts and grow vocabulary by first training it to associate image representations and labels. The input and output representations had relatively low ecological validity. The image representations were clusters of random dots and the labels were individual (localist) units associated with each image. The network was tested to reproduce the image representations when it was only given the labels (comprehension) and then it was tested on whether it could identify the labels when it was only given the input images (production). The key empirical phenomena to which the model was addressed were the comprehension-production asymmetry, over/under extension errors and the non-linear trajectory of vocabulary growth.

Another example of a connectionist model is the Mayor and Plunkett model of taxonomic responding and fast mapping in early word learning (Mayor & Plunkett, 2010). Here taxonomic responding refers to the classification of objects according to taxonomic matches (dog-wolf) as opposed to thematic matches (dog-bone). Taxonomic matching is the way infants are believed to respond on learning a new word. The children's ability to generalise a newly learned label to other objects in the same category is referred to as fast mapping (Mayor & Plunkett, 2010). This model used self-organising maps for an emerging internal shift to categorical representation, rather than association, that accounted for fast mapping in early word learning and 'a rapid increase in the rate of acquisition of words observed in late infancy' (Mayor & Plunkett, 2010, p.1).

The Plaut and McClelland model of word and non-word reading (Plaut & McClelland, 1993) used a *parallel distributed processing* approach which can learn rule-based and exceptions in the naming of letter strings. Their network generalised relatively well because it developed componential attractors that reflected common sublexical correspondence between orthography (the conventional written spelling) and phonology for regular words, as well as less componential attractors that produced representations for irregular words. By contrast the McMurray, Kucker, Zhao and Samuelson model proposed that simple mechanisms like association and competition of concepts at a given time interact and guide choice of word, thereby accounting for the dynamic nature of word acquisition (McMurray, Zhao, Kucker & Samuelson, 2013). Their model was an associative connectionist system based on Hebbian Normalised Recurrence that included both internal representations and real time competition to enable fast learning by association.

The dynamic aspect of connectionist modelling is especially apt for exploring theories with a neuroconstructivist approach of development (Karmiloff-Smith, 1998; Karmiloff-Smith, 2009; Mareschal et al., 2007). Neuroconstructivists promote the study of developmental disorders from early stages of infancy, based on the idea that human intelligence is not a static state; it is a process that emerges from dynamic 'multidirectional interactions between genes, brain, cognition, behaviour, and environment' (Karmiloff-Smith, 2009, p. 56).

Computational Models of Developmental Language Disorders

Rumelhart and McClelland's Parallel Distributed Processing (PDP) theory (Rumelhart & McClelland, 1986a), especially the paper "on Learning Past Tenses of English Verbs" (Rumelhart & McClelland, 1986b), has been a particularly influential framework for artificial neural networks used to develop connectionist models of developmental disorders (Daniloff, 2002). Data from research in clinical treatments of delays in language development and previous models of development have been effectively explained in the context of parallel and distributed cognitive information processing (Norris & Hoffman, 2002). Artificial neural networks based on PDP theories have also been successful in predicting the performance of aphasic patients with naming disorders (Gangon & Martin, 2002). The researchers suggested that knowing what to expect through simulations that predict outcomes of language processing in patients with aphasia could be useful in developing appropriate rehabilitation treatments. In another study, ANN have been used to analyse outcomes of speech perception under different paradigms (Trentin, Brugnara, Bengio, Furlanello & De Mori, 2002). Results have been used for educational as well as clinical purposes.

It is the dynamic nature of PDP modelling which renders it suitable for constructing models of developmental disorders (McClelland, Rumelhart & Hinton, 1986). Learning takes place as small incremental changes to weights that connect input units to output units (sometimes through hidden units) when activations of output units are compared to target unit activations (McClelland, Rumelhart & Hinton, 1986). Accordingly, connectionist models are used to simulate typical and atypical cognitive development (Thomas & Karmiloff-Smith, 2002) aiming to seek explanations of underlying mechanisms that cause behaviours in developmental disorders, the implemented developmental process allowing disorders to be construed in terms of atypical constraints shaping that process. For example the question of whether a group with a certain behavioural impairment is found to be due to one underlying cause (homogeneous disorder group) or due to different underlying causes (heterogeneous disorder group) has been investigated using ANN (Thomas, 2003).

Connectionist models can in general be set to simulate typical and then atypical development by changing the start state of one (monogenic approach) or more (polygenic approach) of their free parameters, such as the number of hidden units and learning rate (Thomas et al., 2019). Researchers have also simulated atypical development by manipulating the training sets at the input, which is considered to effectively reflect the state of the environment (Thomas, 2016). For example Thomas, Forrester and Ronald (2013) used a model for the acquisition of the English past tense, to first simulate normal development, then they implemented manipulations on the training set as well as manipulations on the efficiency of the learning system to investigate the simulated effects of socioeconomic status on learning. Examples of ANN parameters that are usually considered to represent elements in the human brain include (i) a parameter for the number of connections which represents the number of axons and dendrites; (ii) a parameter for the number of hidden units which represents the number of neurons; (iii) a parameter for the learning rate which represents changes in synaptic strength in the brain (Bullinaria, 1994). For example Harm and Seidenberg (1999) took a monogenic approach to simulate developmental dyslexia by reducing the number of hidden units that were designated to learn mappings between orthography and phonology. An example of a polygenic approach is one taken by Thomas and Knowland (2014) in which they investigated the notion of resolving and persisting delays in language development using the English past tense as an example. Their model simulated language development of large populations where DLD were implemented by applying small variations in the values of a number of low-level neurocomputational parameters, such as those that represent learning capacity (intrinsic variation) or by varying the amount of information provided in the training set at the input (extrinsic variation). The idea behind simulating development of large populations is that a normal distribution is produced in which disorders are represented by the lower end of the continuous distribution (Thomas, 2016; Thomas et al., 2019).

Computational models of intervention for Developmental Language Disorders

Prior to simulating behavioural interventions, connectionist approaches were extensively used to predict effects of brain damage to neuropsychological patients, such as aphasics, to help decide on appropriate treatments for rehabilitation (Gagnon & Martin, 2002). Plaut and Shallice's (1993) preliminary connectionist account of deep dyslexia was inspired by Hinton and Sejnowski's (1986) model of mapping arbitrary representations of orthography onto semantics. A later study explored the effects of different types of training for patients of acquired dyslexia and results surprisingly suggested that training on atypical words from a category yielded better generalisation effects than training on typical words of the same category (Plaut, 1996). Inspired by these models, Harm and Seidenberg (1999) developed a model of reading acquisition focusing on the development of atypical phonological representations prior to learning to read, which was then used to explore the efficacy of intervention techniques for developmental dyslexia (Harm, McCandliss & Seidenberg, 2003).

In the first part of the study, referred to as "Remediating Phonology," five simulations were run: one simulation for typical development, and a second simulation with impairments made to the phonological component by lesioning a random 50% of its connections prior to training; three simulations were then run with the same constraints removed at different points of training: the first was alleviated at the beginning of training, the second after 10,000 reading trials and the third after 100,000 reading trials. The authors sought to investigate the potential of remediation if the impairments were alleviated at these different points in time. Results showed that the model performed as well as at the normal simulation when the impairment was removed at the early stage of training. However when the impairment was removed at the stage of 10,000 reading trials and beyond, the model's non-word reading did not improve. They suggested that interventions that target phonological representations alone would only be effective if they were introduced at a very early stage of learning.

The second part of the study "Simulating the Word Building Intervention", by contrast, involved a simulation of a Word Building intervention, which focuses on decoding skills and methodical letter to sound mappings (McCandliss, Beck, Sandak, & Perfetti, 2003). Remediation was simulated by additional lessons in decoding, implemented as before at the beginning of literacy training, at 10,000 word presentations and again at 100,000 word presentations. In this second part of the simulation, performance of the model on word accuracy improved after intervention was made at all the three stages of training. The outcomes of the simulations were similar to the results of empirical data of the same intervention implemented in another study for children aged between 7 and 10 years (McCandliss, Beck, Sandak, & Perfetti, 2003). Deductions about intervention possibilities were made from analysing the activations of hidden units under different paradigms. The theory was that analysis of hidden unit activations could shed some light on the mechanisms of internal representations of words (Harm et al., 2003). The strength of the second part of the study by Harm and colleagues lay in the availability of empirical data from the study by McCandliss et al. (2003) that could be compared with the results of simulations of the same intervention. The key findings of this model were that phonology-only interventions were less effective after the onset of literacy acquisition whereas interventions that address alphabetic decoding of letters to sounds can still be effective at the later stages of word development (Harm et al., 2003). These findings were consistent with the findings of the empirical study by McCandliss et al. (2003).

However, research using neurocomputational modelling is currently more advanced in exploring interventions for *acquired* deficits of language abilities than *developmental* difficulties. For example, Ueno, Saito, Rogers and Lambon Ralph (2011) have proposed the Lichtheim 2, a dual dorsal-ventral pathway neurocomputational model that has been drawn from prior model architectures to simulate normal word production, repetition and comprehension. By implementing constraints on the model architecture in the trained model, they simulated acquired aphasia profiles for those same language functions. Partial rehabilitation, usually observed in aphasic patients, was captured by retraining the model after damage with the same tasks as before damage. More recently another neurocomputational model, referred to as the connectivity-constrained cognition (C^3) model, has been brought forward by Chen, Lambon Ralph and Rogers (2017) to explore mechanisms behind semantic representations of words. The C^3 is a deep neural network model and works on the basis that semantic representations of words involve a modular domain-specific and a domain-general system, as well as a learning network. The model was used to simulate data from neuroimaging as well as naming accuracy for participants with typical word production and for those with disorders of semantic representation, which included semantic dementia, Herpes Simplex Viral Encephalitis, temporo-parietal tumor resection and category-specific visual agnosia.

Review of Computational Models of Intervention for Developmental Disorders. In a recent paper that focused mainly on computational modelling of interventions for developmental disorders of language and communication, Thomas and colleagues (2019) reviewed evidence from various connectionist models to assess the potential of identifying general principles of intervention by considering the mechanistic basis of the disorder. Their evaluation of interventions was mainly based on whether an intervention achieved (1) *generalisation* i.e. outcomes improve for items that were not included in the training set and (2) *persistence* i.e. outcomes keep on improving even after training has stopped; both of which have been proven difficult to achieve in real life interventions (Green, 2020; Thomas et al., 2019). There were four sections to their analyses which included (i) *Long term outcomes* in the case of potential compensatory outcomes if no intervention is applied and in the case of resolution of early delays; (ii) *Different methods to remediate atypical development in a single network*, namely: (a) where the disorder is due to inadequate stimulation at the early stage of development, (b) where improvements can be made to the training set of an atypical learning system, (c) where improvements can be made to the input and output representations of an atypical learning system, and (d) where interventions can be set to modify the computational properties of the learning system; (iii) *Interventions to encourage compensation via alternative pathways*; (iv) *Individual differences in response to intervention*.

In the case of long-term outcomes, Thomas and colleagues (2019) deduced that with repeated training and no particular intervention, some systems with deficits can show accurate outcomes on the trained items, but residual deficits would persist and show on generalisation or more difficult tasks. Analyses of factors that could influence resolution of early delays was mostly based on the study by Thomas and Knowland (2014) which used a population modelling approach to simulate 1000 individuals acquiring the English past tense. Having differentiated between a system with low *capacity* as one that has a decreased ability to learn complicated material, and a system with low *plasticity* as one that requires more practice to learn new information, Thomas and Knowland (2014) found that delays can be resolved if the limitations in processing is due to plasticity of the system, whereas delays would persist if the limitations were in the capacity of the system. For plasticity limitations, Thomas et al. (2019) recommended intervention in this case to be increased practice by providing more opportunities for the natural experience, for example, more frequent language communication at home. By contrast, capacity limitations may require specific interventions different from naturalistic language-based activities. However, they also pointed out that at the early stages of development, it is challenging to identify whether a delay is due to limitations in plasticity or capacity (Thomas et al., 2019).

Thomas and colleagues' (2019) evaluation of studies where the disorder is due to inadequate stimulation at the early stage of development showed that enrichment interventions can achieve remediation for some cognitive deficits if given at an early enough stage. The authors inferred that late intervention can work for more developed skills, but will be required to involve an approach that is more elaborate than enrichment.

In these cases interventions that include strategies that address the delayed mechanism in each particular case, for example, more feedback or perceptually exaggerated stimuli, would have an increased chance of achieving remediation (Thomas et al., 2019).

Computational simulations involving the reorganisation of training sets to better suit an atypical system have shown some improvement on generalisation (Fedor, Best, Masterson & Thomas, 2013; Thomas et al., 2019). Modifying the input or output representations can also help an atypical learning mechanism solve a problem represented in a simpler form (Thomas et al., 2019). However, again, timing of such interventions plays a major role in determining the outcome. For example Alireza, Fedor and Thomas (2017) used the same architecture of the Best et al. (2015) model to simulate typical and atypical word acquisition and then they introduced interventions that targeted either the weaknesses or the strengths of the learning system in the form of phonological or semantic representations (this study is described in more detail in chapter 8). The best outcome resulted when improvements were made on both semantic input and phonological outputs. Simulations also showed that interventions that targeted strengths were more effective if implemented at an early stage of training whereas interventions that remediated weaknesses were more effective later in training (Alireza et al., 2017; Thomas et al., 2019).

Simulations of interventions set to modify the computational properties of the learning system can be considered to simulate modification of biological mechanisms, for example, simulating the effects of pharmacological treatments on the system; or simulating the effects of change of behaviour such as diet, exercise or mindfulness training (Thomas et al., 2019). It was implied that combining evidence from cognitive neuroscience data, such as structural or functional MRI, with computational modelling data can lead the way to more successful inferences to be made on the mechanistic causes of deficits.

Computational modelling research on interventions to encourage compensation via alternative pathways or mechanisms to generate the same or similar outcome is less advanced (Thomas et al., 2019). The authors pointed out that the notion that therapists tend to employ implicit strategies of intervention for younger children and explicit strategies for older children to encourage compensatory strategies may indicate that meta-cognition is effective in prompting a reorganisation of mechanisms. They proposed that the latter concept should be included in future models of intervention that explore prompting of reorganisation of component mechanisms within the system.

Individual differences in response to intervention remain a challenge for therapists. This is the case since besides the mechanistic cause of the deficit, response to intervention tends to be influenced by other factors like attention skills, personality, motivation and engagement with the therapist (Ebbels, 2014b; Thomas et al., 2019). These skills have also proved to be difficult to simulate. Thomas and colleagues (2019) proposed that these factors could potentially be viewed to influence the plasticity of the mechanism, the amount of information given to the child in a setting and the dosage of the intervention.

The most prominent inferences made in terms of general principles of intervention in the Thomas et al., (2019) review are (i) different underlying deficits can produce the same behavioural deficit, which can then lead to different responses to the same intervention. (ii) Timing of interventions should be taken into account. Most interventions work better at the earlier stage of development, yet some techniques (e.g. meta-cognitive techniques) can be more effective at a later stage. (iii) The content of the intervention is important and should be developed to encourage the use of a different mechanism. Interventions that involve more practice on the same material, which would be adequate for a typically developing system to progress, will not usually work for an atypical system.

On the whole, when *generalisation* of intervention effects on items which were not included in the training sets were considered for the reviewed computational models, the intervention effects were very small (Thomas et al., 2019). The authors reasoned that since the mechanism of connectionist networks is error correction, there was no point including the results of trained items because they would improve with training anyway most of the time. However in reality generalisation has proved difficult to achieve and in general an intervention would still be considered appropriate for a child even if it improves the outcome of only trained items (e.g. to increase vocabulary).

In terms of models that considered *persistence* of intervention effects, Thomas and colleagues' (2019) evaluation found that effects of early intervention persisted in the long term only if the plasticity of the system was reduced during training (Davis, 2017; Thomas et al., 2019). In the case where plasticity remains constant throughout training, another strategy for sustaining the effects of intervention is to alter the training set of the system permanently. For example, this would be the equivalent of training parents to modify the language level when interacting with the child but this intervention would need to continue (Thomas et al., 2019). Hence the presence of plasticity across training can be advantageous in the sense that interventions can change the outcome if they are implemented at any time during development, yet their effect will dissipate if interventions are stopped (Thomas et al., 2019; Yang & Thomas, 2015). On the other hand the advantage of plasticity being reduced across training means that interventions can have a permanent effect but only if they are implemented at an early stage of training.

Background of Current Research

This research follows in the line of the study by Best et al. (2015), which extended the same principle applied in the Harm et al. (2003) study, of combining empirical data from a controlled intervention study with results from simulations of similar interventions in a more recent study. Best and colleagues used a connectionist model of word acquisition to investigate methods of intervention for WFD. The authors extended the idea of mapping phonological and semantic representations in self organising maps to simulate word acquisition in a similar way to the DevLex-II model (Li, Zhao & MacWhinney, 2007) and Mayor and Plunkett model (2010) of early word learning. Instead of using self-organising maps with a limiting two-dimensional factor for internal representations, the Best et al. model was based on separate autoassociators for phonology and semantics, trained with the backpropagation-learning algorithm (Rumelhart, Hinton & Williams, 1986a). Associations could then be learned between the internal representations of these phonological and semantic components to simulate comprehension and production, respectively. For the Best et al. (2015) study, the model was first set to simulate typical development of word acquisition. Parameters of the model were then adjusted to simulate errors and behaviours similar to empirical data for two children with WFD. These manipulations included (i) decreasing the number of hidden units; (ii) decreasing the number of connections between layers; or (iii) using a shallower sigmoid unit activation function for the artificial neurons in the model, The 'impairments' of the model were placed before training occurred with a view to simulate behaviour as the outcome of developmental deficits. After the impaired model was trained to simulate atypical development, interventions were incorporated in the simulations to test their effects. Interventions involved more training for the phonological component of the model, more training on elaboration of semantics for the semantic component, or more training for both components of the model on naming the target word. Results were thus considered to reflect the effects of intervention on WFD within the developmental trajectories approach.

The Best et al. (2015) model nonetheless had at least two major limitations. First, the model found it much harder to learn semantic hidden representations than phonological hidden representations. Second, training sets were abstract and a new one was generated for each simulation. This made it harder to compare results of the simulated interventions with empirical data from children.

Chapters 5 and 6, hence, present two sets of secondary data taken from the Best et al. (2021) study (which is the follow up of the Best et al., 2015 study). The idea was to build on data collected during the WORD project, to characterise the developmental trajectories of typical and atypical vocabulary development, and to evaluate the extent to which language profiles can predict response to interventions in children with WFD.

We are then in a position to turn to computational modelling of WFD and interventions. That work begins by attempts to address limitations of the Best et al. (2015) model, before we move on to simulating deficits and their remediation, and more widely, considering whether computational methods can be successful in helping narrow the gap between research and intervention.

Before we do that, however, we need to find out more about that gap. And to this end, the next chapter describes a questionnaire study carried out with speech and language therapists, which explained how they use research to guide their practice.

Chapter 4

Questionnaire Study

Background

This chapter presents a qualitative study using a questionnaire introduced to therapists, with the goal of characterising more clearly the challenges to be overcome in seeking to narrow the gap between research on effective methods of intervention for developmental language disorders (DLD) and clinical practice.

Research shows that although various theories of underlying causes of language deficits are emerging, intervention practices do not seem to be guided by those same theories (Law, Campbell, Roulstone, Adams & Boyle, 2008; Michie & Prestwich, 2010). The heterogeneous nature of DLD is undoubtedly a potential reason for the disparity between theories of causes of deficits and theories of intervention, as well as a variety of possible reasons which have already been discussed in Chapter 1. Several studies have shown that speech and language therapists (SLT) follow a large variation in procedures when deciding on which intervention to use for each child with DLD (Best, 2003; Law et al., 2008; Law et al., 2012; Roulstone, Wren, Bakopoulou & Lindsay, 2012).

A more recent study, by Morgan and colleagues (2019) on intervention for preschool children with DLD found that even when SLT endeavour to apply principles of evidence-based practice, in reality the application of these principles has not been implemented. The authors pointed out that the words “it depends” (p.962) were a recurring phrase used by therapists when describing how they decide on which intervention strategy to apply in each case (Morgan et al., 2019). The same study noted that factors believed to be influential on choice of therapy included the child’s needs, developmental level and the parents’ involvement with the therapy, as well as the resources available. The aim of the Morgan et al. (2019) study was to develop a framework for SLT to use when working with preschool children with speech and language disorders. The framework highlighted the need to focus on developing children's functional skills and to support adults who can in turn provide a supporting environment for these children.

One of the reasons for the persistence of the gap between theories of causes of deficits and theories of intervention could be that none of the studies included questions which explored underlying principles by which therapists currently make their choice of technique. It has been found that therapists often develop their own theories of causes of difficulties and adapt interventions accordingly (Law et al., 2008). In other words, their choice of intervention often depends on their implicitly held theories on what deficits the children have and how these should link to the therapies that particular therapist has available or the regimes that he or she prefers to use. Indeed, on answering preliminary questions posed by the author of the current study to SLT during informal interviews, some therapists said they were not always aware of the reason behind using specific interventions with different children and that their choices were sometimes guided by their instinct. In this respect instinct usually develops by experience; thus this phenomenon, along with the notion that therapists develop their own theories of causes of language deficits, brings to light the possibility that SLT have potentially accumulated knowledge which can be instrumental in developing principles of effective methods of intervention for DLD. However, developing the aforementioned principles is more likely to be achieved if the insight acquired by SLT is elucidated by targeted questions. The questionnaire used in the current study includes such targeted questions. The objective is to systematically link any implicit theories developed by SLT on causes of language deficits in clinical settings with research-driven theories based on behavioural studies and neuroscience. Inferences based on the responses of the questionnaire, were then integrated into the computational modelling part of the investigations of this study. The ultimate goal was to construct a more systematic theory of which language interventions work best for children with different strengths and weaknesses.

Method

The questions in this study were tailored to the example of word-finding difficulties (WFD) for the purpose of attaining questions targeted at implicit inferences made by SLT on underlying mechanisms that cause a particular language difficulty. The questionnaire was developed after having reviewed both a small survey of intervention methods from 24 SLT for children with WFD (Best, 2003); and a series of email responses to a pilot questionnaire asking: "what techniques do you typically use to support children with word-finding difficulties" which consisted of seven answers from SLT. Neither survey addressed questions that explore underlying principles by which therapists make their choice of technique. Only one email, which was addressed to a researcher with more specific questions relating to possible underlying mechanisms of WFD, gave a more informative response. This speech and language therapist argued that sometimes this individual found it more useful to work on a child's strength to find a compensatory skill rather than working on the child's productive vocabulary weakness. For example, if the therapist found the child becoming frustrated by not being able to access the phonological form of a word, then she would encourage the child to describe the word's semantic aspects. The same therapist mentioned the use of errorless learning in cases where she thought the child was sensitive to making mistakes. Errorless learning refers to training methods which avoid providing the opportunity for the learner to make a mistake (see Middleton and Schwartz (2012) research paper for an extensive review of errorless learning in cognitive rehabilitation). In the context of treatment for word training, errorless learning refers to oral repetition of naming pictures by the participant having been given the correct word from the start in order to strengthen the association between the correct name and the picture (Schuchard & Middleton, 2018).

Questionnaire Design

The introduction of the questionnaire comprised a few lines explaining in more detail the objective of the study and that, although questions were tailored to the example of WFD, participants were free to answer the questions with respect to behaviours more salient in their practice. The questionnaire was composed of 15 main questions some of which consisted of several parts that aimed to guide the respondent to think about the relevant aspects when giving his or her answers (the full questionnaire and responses are included in Appendices 1 and 2). The questions were generally in open form in order to be able to collect as much narrative data as possible. The questionnaire first collected information on the participant's background. This was then followed by questions directed at how and why therapists use particular interventions, starting out with an open question and then becoming more and more specific. The latter part of the questionnaire explored factors that influence effects of intervention. The outline of each question was as follows.

Question 1 was in four parts to establish background information, including years of experience of therapists and the age ranges of the children that they usually treat. Question 2 was open and in two parts: the first part asked therapists what intervention they would try first; the second part prompted participants into giving reasons for their choice of interventions. Questions 3 and 4 inquired about the use of variation. The string of questions on variation were included because it has been identified within the learning sciences as one of the important factors contributing to statistical learning, along with frequency of training and the introduction of novelty in familiar contexts (Borovsky &

Elman, 2006; Onnis et al., 2004; Thomas et al., 2019). Question 5 explored the idea of letting children make mistakes to learn from them by being given feedback versus the use of scaffolding to encourage errorless learning and boost confidence. A similar comparison was included in a study by McCandliss, Fiez, Protopapas, Conway and McClelland (2002) in which Japanese adults were given training to learn to identify the difference in pronunciation between the English [r] and [l] sounds *(e.g., rock-lock*) under conditions that included feedback or scaffolding. We will return to this example later in this chapter. Question 6 explored the use of breaking up words into syllables or similar techniques.

Moving on to other factors which influence the effects of intervention, Question 7 prompted the participant to describe the importance of the child's level of engagement and motivation. Question 8 looked at how much time therapists spent on working with the children at difficult versus easier tasks. The importance of parental contribution was explored in Question 9 as this factor has been highlighted in previous studies (Morgan et al., 2019). Question 10 explored the notion of working on a child's strength to find a compensatory skill versus working on improving a child's weakness. This concept was further investigated in a computational modelling study (Alireza, Fedor & Thomas, 2017) and is covered in more detail in Chapter 8. Question 11 tackled the issue of generalisation of behavioural gains beyond training items, which has typically been difficult to achieve (Best et al., 2021; Ebbels, 2014b; Thomas et al., 2019). Question 12 required participants to clarify how they estimated or determined the dosage requirement of therapy to be given to each child. Question 13 was open and prompted therapists to propose any other factor that they believed is influential on the outcome of behavioural intervention. Question 14 focused on the prediction of outcomes of intervention, based on the child's general intelligence (IQ). Finally, Question 15 was open for any further suggestions, feedback or comments from responders to the questionnaire.

The questionnaire was created on a Google Docs form so that participants could answer anonymously. The link of the questionnaire was sent by the author, by email, to 23 speech and language therapists (SLT), one specialist teacher and one primary school teacher. The latter were both experienced in their field and were asked to pass on the questionnaire to their colleagues. They were included for the purpose of increasing the diversity of the incoming data especially from the practical aspect of intervention. All the recipients of the email were asked to pass on the questionnaire to colleagues whom they thought would be interested in participating in the research. The link of the questionnaire, along with a small explanatory paragraph, was included in the July 2016 Royal College of Speech and Language Therapists (RCSLT) research newsletter after having obtained ethics permission.

Ethics

The questionnaire was granted ethical approval by the departmental ethics committee of the department of Psychological sciences of Birkbeck College, University of London. The accompanying short paragraph explained the purpose of the questionnaire and invited SLT who regularly practiced therapeutic interventions for children from different age groups to complete the questionnaire. The link to the questionnaire and the short paragraph were also posted on the RCSLT Research Facebook page in March 2018. The accompanying paragraph to the questionnaire is provided in Appendix 3.

Results

Participants. Ten responses were obtained, nine of which were from SLT and one response from a primary school teacher (for the sake of simplicity in this review all respondents will be referred to as therapists). Table 4.1 below presents the answers collected for Question 1 to give background information of the participants. The number of years of experience of the participants ranged from 3 to 24 years with a mean of 11.9 years. Six of the therapists gave the sessions in a health centre and the rest at school. The majority of children that were given the interventions ranged in age between 4 and 11 years, but some therapists taught children whose age ranged from 2 to 18 years.

Table 4.1 Participants' background information

Participant	Profession	Years of Experience	Place of therapy	Age Range of children in years
1	SLT	22	Health Centre	Under 5
2	SLT	14	Health Centre	Both under 5 and 4 to11
3	SLT	13	School	8 to13
4	SLT	4	School	4 to 16
5	SLT	7	Health Centre	Birth to 18
6	SLT	3	Health Centre	2.5 to 11
7	SLT	10	Health Centre	2 to 18
8	SLT	7	Health Centre	4 to 11
9	SLT	24	School	4 to 11
10	Teacher	15	School	4 to 11

SLT= Speech and Language Therapist

Responses: In the below, responses are summarised according to themes and cover in turn: intervention approaches and influencing factors, variation versus repetition, scaffolding, motivation and level of engagement, compensatory skills, generalisation, dosage, and other contributing factors including IQ.

Intervention Approaches and Influencing Factors. When asked what would be the first approach the therapist would select to improve a case of WFD only two therapists replied that they would assess the child first to try to find out whether the root of the child's difficulties was linked to a difficulty in semantics versus phonology. Three therapists replied that they would combine a semantic and a phonological approach from the beginning. They would start with a semantic approach explaining the meaning of the word followed by phonemic cues. Three therapists would use pictures, repetition or visual memory. Only one therapist took a phonological approach first and one therapist said that she used play therapy following the child's lead.

On answering the question "on what grounds would you decide what to try first?" three therapists said that depended on the language level of the child; three therapists said it depended on the engagement level of the child; two replied that it depended on the assessment done by the therapist; one replied that it depended on both language level and interest; and only one therapist replied that she would use pictures and sounds tailored to the child's knowledge of phonemes.

Most of the therapists agreed that children respond differently to the same approach because of the various kinds of comorbid difficulties they have in other areas, such as attention or other cognitive functions. One therapist remarked that even parental involvement in therapy seemed to make a difference on the child's response.

Variation versus Repetition. The majority of therapists said that they would use repetition to improve target behaviour; and that they would use variation in order to keep the child motivated and not become bored with repetition. In this context, respondents referred to variation of activities, for example using pictures on paper versus an iPad or physical items. Only two therapists mentioned the use of variation for the purpose of generalisation and one therapist interestingly pointed out that he/she used variation to prevent *over*generalisation. Although when specifically asked about the type of pictures used to prompt naming, for example for the word ‘dog’, a third of the therapists replied that they would only use pictures of dogs at the beginning and then they would bring in pictures of other animals. Most of the therapists would contrast the pictures with other animals from the start.

Scaffolding. Only two out of the ten respondents said they would allow the child to make a mistake before using cues. Most of the therapists would use a scaffolding approach in order to reduce the likelihood of the child making a mistake, thereby boosting the child’s confidence. Six out of ten therapists would break up the word into syllables to practice each syllable separately and combine it with practicing the word as a whole. They would use this method especially if they felt that the difficulty stemmed from a problem with processing phonology.

Motivation and Level of Engagement. Nine out of ten therapists agreed that motivation and level of engagement are very important. Having identified the child’s interests beforehand, activities and reinforcement rewards are then used accordingly. When asked about the ratio of working with more difficult versus less difficult tasks, therapists gave a wide variety of answers. Some therapists responded that they would start with easy tasks then move on to harder ones but that they would make sure to end with a successful answer. Some therapists would give two or three easy tasks and then one difficult task, others would give equal time to both kinds of tasks. Nine out of ten respondents agreed that parental involvement is a very important contribution to achieving faster progress of the intervention and that it helps in the generalisation process. Parental involvement has also been pointed out as one of the factors that therapists consider when deciding on dosage of therapy sessions.

Compensatory Skills. There was a large variation of explanations on answering the question of whether to persist on teaching skill X rather than encouraging a compensatory skill Y. Some therapists replied that it depended on the age or language level of the child. For example, one therapist replied that older and more motivated children were encouraged to use compensatory skills as well as word-finding techniques. Some therapists noted that it depended on the importance or relevance of skill X to the child, whereas others would not persist if they felt the child becoming stressed. One therapist simply replied that she would not accept substituting skill X with a compensatory skill Y and another replied that she would be happy to do so.

Generalisation. On encouraging generalisation of the target skills, most therapists would involve the parents and schoolteachers of the child giving them instructions to use the same strategies that the therapists used. Describing the target word was the technique that most therapists encouraged. One therapist would use iPhone pictures or advise watching National Geographic. Another would use group therapy and one therapist mentioned incidental teaching as a way to encourage generalisation of the target words. The general idea given by most therapists was to vary the context of usage for the tasked words, thereby enabling more repetition.

Dosage. When asked about how therapists decided on the 'dosage' or amount of therapy recommended for each child, their answers again were wide ranging. Some therapists thought that required dosage depended on the language level of the child; others would base their decision on whether the child has comorbid difficulties and on the nature and motivation of the child, as well as the level of parental involvement. Most therapists agreed that predicting the required amount of therapy depended on too many factors and that it is easier to predict once they assess the child's initial progress.

Unfortunately, the availability of therapists also plays a factor in deciding on the amount of sessions that would be assigned for each child.

Other Contributing Factors and IQ. Asked what else would influence their choice of intervention other than the suspected underlying cause of the child's difficulty, therapists replied that their personal expertise, parental level of involvement, institutional policy and availability of resources all play a part in their choice of intervention. However, when asked specifically about IQ most therapists agreed that the child's IQ is an important indicator for the child's response to intervention and prognosis.

Discussion

The purpose of the questionnaire was to investigate and understand the mechanisms underlying interventions used by SLT to support children with DLD. Ultimately the goal is that in doing so one would be able to answer questions such as: Which interventions will work for which deficit? At what stage of development should intervention be introduced? How should an intervention be delivered (for example, at school, in one-to-one settings or by teaching parents the skills rather than the child)? How flexibly can they be delivered? Who will they work for? And how long should the interventions last for in an ideal situation?

The strategies that seem to be most commonly used by therapists and teachers were repetition and scaffolding. When scaffolding, therapists would first explain the meaning of the word and then break it up into syllables thereby using both semantic and phonological approaches. In general, repetition was used to reinforce a correct answer and scaffolding was used to build confidence of the child by avoiding the chance of him/her making a mistake.

The use of scaffolding to encourage errorless learning is in line with the Hebbian learning model. The Hebbian learning rule states that "if one neuron takes part in firing another, the strength of the connection between them will be increased" (McClelland, Thomas, McCandliss & Fiez, 1999, p. 76) meaning that if a pattern of neural activity is generated by a stimulus then the likelihood of the same pattern being generated on subsequent presentations of the same stimulus will be high. Speech and language therapists largely use scaffolding to build confidence and, in terms of connectionist modelling, confidence can be modulated by measuring the learning rate of the system (Vallabha & McClelland, 2007).

A similar comparison between the use of scaffolding and the use of feedback was included in a study by McCandliss and colleagues (2002) in which 4 groups of 8 Japanese adults were given training to learn to identify the difference in pronunciation between the English phonemes [r] and [l] (as in *rock-lock*) under 4 different conditions. Two of the groups were trained with the standard or *fixed* pronunciation and two groups were trained with a scaffolding approach, referred to as *adaptive* to encourage errorless learning. Scaffolding was achieved by using an acoustically modified sound to exaggerate the difference between the [r] and [l] phonemes, which was then gradually returned to normal speech as the participants succeeded in learning the difference in pronunciation. Only one group in each of these two conditions were given feedback. Hence participants in one of the groups with the fixed training condition were given the opportunity to make mistakes and learn from them, whereas participants in the adaptive training were given more confidence to learn either with or without feedback.

Interestingly, in this study the group that showed the highest learning rate was the one which had fixed training with feedback. The next highest learning rate reported was for the group with the adaptive training and feedback, although there was no significant difference between this group's learning rate and that of the adaptive training group with no feedback. The fixed training group with no feedback showed little improvement.

These results bring into question whether the scaffolding approach, which according to the current study is used by most therapists, is more efficient than letting the child make mistakes to learn from them. Further research, whether in empirical form or connectionist modelling on this particular question, would shed more light on this important point. Connectionist modelling is appropriate to investigate the mechanisms of learning in this case since it is more feasible to simulate concrete aspects such as scaffolding or making a mistake, as opposed to having to simulate motivation or a child's engagement level.

Most therapists agreed that being able to motivate the child by having background knowledge of his or her interests, as well as parents' involvement in the therapy, are very important contributors to the child's progress. Although the questions on which intervention to use were specific, most of the therapists did not answer them directly explaining that the answer would depend on the combination of numerous factors like comorbid difficulties and motivation. They all agreed that IQ was an important predictor for the child's prognosis. This implied that a child's language profile could potentially predict his or her response to therapy.

Thus, on the whole, responses to the questionnaire presented a view in line with Bronfenbrenner's ecological systems theory (Bronfenbrenner, 1992), which indicates that outcomes of interventions for DLD rely on multiple interactive factors ranging from the *macrosystem* or the environment, for example the budget of policy makers for dosage, to aspects that are closer to the learning process, for example the child's motivation. Similarly, responses to the questionnaire are in keeping with the framework proposed by Thomas, Ansari and Knowland (2018) which was inspired by a combination of

Bronfenbrenner's theory and principles from Michie and colleagues' (2011) behaviour change wheel method (see Thomas, Ansari & Knowland, 2018, Figure 2, reproduced in Figure 4.1 below). In parallel with Thomas et al.'s review of the role of neuroscience in education, the questionnaire was mainly focused on elucidating theories of hidden mechanisms behind DLD that are in keeping with theories from the field of cognitive neuroscience. Likewise the questionnaire responses portrayed an overall viewpoint similar to the one illustrated in Figure 4.1 indicating that outcomes of learning interventions are constrained by the interaction of interdependent factors of various proximities from the learning process. The four main factors identified were 'Government', 'Society and family', 'School' and 'Child factors'. These factors in turn included more specific aspects such as 'education budget' under governmental factors or 'teaching materials' under school factors, which would affect learning outcomes, as indicated in the report (Thomas, Ansari & Knowland, 2018).

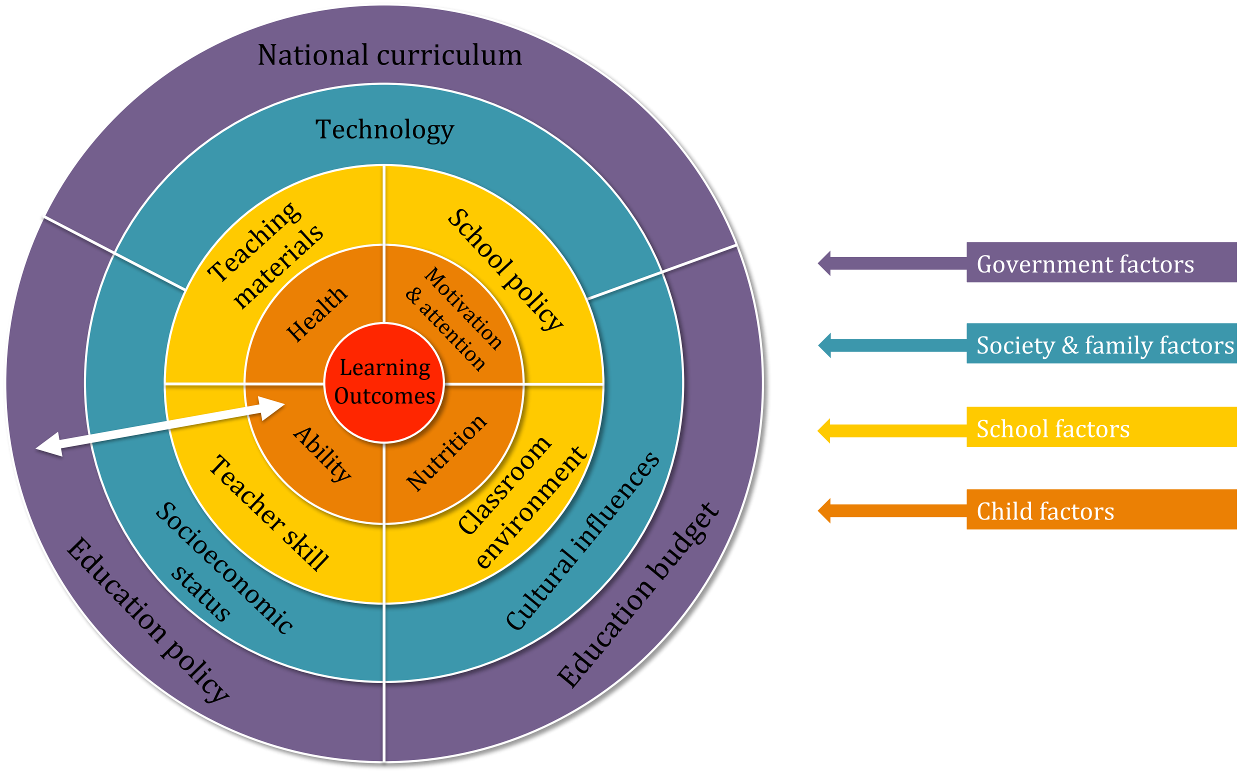


Figure 4.1. Proximal and distal factors that support and constrain change in learning outcomes, following the layered influences on behavioural change proposed by Michie et al., (2011), and the interactive relationships between an individual and their environment as proposed by Bronfenbrenner (1992) [reproduced from Thomas, Ansari & Knowland, 2018].

In general, the purpose of the questionnaire was to identify strategies that are most effective for the *different* cases of children with WFD. The number of participants that answered the questionnaire was not large enough to deduce conclusions on the effectiveness of specific interventions as a whole. Yet the range of variation in responses seems noteworthy and is in line with findings from past research on the various intervention strategies followed by SLT when treating different children with DLD (Best, 2003; Law et al., 2008; Law et al., 2012; Roulstone, Wren, Bakopoulou & Lindsay, 2012). Nonetheless, there was some information from these replies from which one can draw inferences on how therapists tailor their intervention to the specific cases of WFD.

Selecting topics that are related to the child's interests in order to encourage engagement of the child was a popular strategy, but this is a difficult aspect of intervention to capture in mechanistic terms, that is, how motivational states relate to the representations, processes, and facility for adaptive changes that mean naturalistic experience has not led to the typical development of the productive language system. Moreover, the lack of the child's engagement is unlikely to be the root cause of the child's learning difficulties, although it could prevent the intervention technique reaching the targeted cognitive mechanism. Hence, in a mechanistic model, it may be less important to simulate motivation and child engagement so long as the influence of dosage is understood. However, for the purpose of this study, other key aspects like repetition, scaffolding and co-morbid difficulties can be integrated in the simulation of language intervention part of this research. In any case, the broad range of answers to most questions in this questionnaire represents the need for more research into this field.

Although the questionnaire was well distributed, the number of responders was smaller than expected. However within this small group there were both diversities and similarities in their answers as well as relevant aspects that remained unmentioned. Diversity of approaches was immediately presented in their answers to question 2 when asked what would be the first approach the therapist would select to improve a case of WFD and why. They adhered (sometimes passionately) to their own implicit theories – rather than simply linking therapies to symptoms – but none seemed to mention evidence bases, research findings, or explicit (mechanistic) theories. Some therapists based their choice of intervention on engagement level and general abilities and others based their choice on the child's knowledge of particular phonemes. Interestingly none of the therapists based their choice of intervention on chronological age except when asked explicitly about strategies for compensatory skills.

Similarities in the SLT's answers were presented when most therapists viewed comorbidity as important. This would be hard to capture in a mechanistic model of an individual impaired processes; anyway comorbidity is less frequently the focus in mechanistic accounts which target specific causes of individual deficits. Most therapists also viewed variation as a means to avoid boredom, rather than a principled way to produce robust learning. Their responses included evidence of using semantic contrasts to support delineation of semantic categories. Scaffolding was used by the majority of therapists as a compositional approach, which was familiar from phonics training. This approach seemed well ingrained in practice, as did errorless learning. Likewise almost all therapists viewed motivation as important to maximise dosage of treatment, but never as a cause of deficit. Bringing children closer to their limits of performance was viewed as important, while maintaining motivation. Hence two constraints were being maximized here.

Diversity of approaches by the therapists was notably presented again on the subject of teaching compensatory skills. Therapists' responses to this question showed some strongly held implicit causal theories. Here was the first mention of chronological age influencing a strategy decision, supporting the view that explicit strategies require maturation (as summarised in the introduction of the Thomas et al., (2019) review paper). However the question of whether to work on a child's strength or weakness was also related by SLT to the child's functional requirements, and to the child's individual emotional response to the therapy situation. Interestingly almost all the therapists had a common point of view on the issue of generalisation. They viewed it as an issue largely to be addressed outside the therapeutic situation, advocating practicing techniques in the variety of situations the child experiences at home or at school, and being supported by parents and teachers. Furthermore the variety of answers given to the important question of how to decide on the amount of dosage show that in reality dosage was not allocated exclusively according to child-centred factors (ability level, comorbidity, motivation) and that it depended almost equally on external factors (parental involvement), and practical factors (SLT resource availability).

In this chapter responses to the questionnaire were analysed and reported per question or themed set of questions. In the context of a more formal framework for qualitative analysis such as thematic analysis, the method implemented here for the qualitative data analysis of questionnaire responses is similar to a 'theoretical' thematic analysis as opposed to the 'inductive' analysis method (Braun & Clarke, 2006). The process of deriving themes from the data using the theoretical thematic method is driven by a researcher's preconceived theories based on previous research, whereas the inductive analysis method tends to be more data-driven and does not rely on findings or themes that have been previously identified (Braun & Clarke, 2006). Since most of the questions in this study were targeted at specific aspects of intervention techniques employed by SLT in order to elucidate principles of effective methods of intervention, the answers in the questionnaire contained emerging themes that were reasonably drawn and analysed using the theoretical coding method.

The Framework Method (Gale, Heath, Cameron, Rashid & Redwood, 2013) is another qualitative data analysis method that could potentially be used to analyse qualitative data from the questionnaire responses and is especially useful for larger numbers of responses. The main feature of the Framework Method is that data is systematically organised into a matrix output where rows are cases (usually interviewees) and columns are codes of summarised data (usually themes). The matrix hence gives a useful summary of the data but keeps the views of each participant connected so the context is not lost (Gale, Heath, Cameron, Rashid & Redwood, 2013). Although the

Framework Method can be useful for summarising and comparing large quantities of qualitative data, it is at the same time not recommended for heterogeneous data and has the potential of losing significant interpretations (Gale, Heath, Cameron, Rashid & Redwood, 2013).

An alternative approach that could potentially collect a broader and richer set of qualitative data on the subject of intervention techniques is to use the questionnaire as a guide to interview SLT. The transcripts of the interviews can then be formally analysed by subjecting the data to all six phases of thematic analysis as described by Braun and Clarke (2006) using the inductive method. The six phases of thematic analysis are: (i) familiarising yourself with your data; (ii) generating initial codes; (iii) searching for themes; (iv) reviewing themes; (v) defining and naming themes; (vi) producing report (see Braun and Clarke (2006), p. 87). Employing a data-driven approach in this context may shed light on implicit theories developed by SLT on causes of language deficits which have not been previously identified.

Further research on effective interventions for WFD was pursued via computational modelling approach to test techniques that were described by SLT in the questionnaire responses. Chapter 8 of this thesis is an account of a simulation that focused specifically on the question of intervening to target strengths versus weaknesses to remediate problems of WFD.

However as mentioned earlier, the Fedor and Thomas computational model (Best et al., 2015) utilised in this study is a connectionist model based on the developmental trajectories approach, where the disorder is set in the context of how component skills develop for typically developing children. In order to establish the groundwork for this account, the next chapter presents developmental trajectories analyses on the relationship between four language skills, based on empirical data taken from both typically developing children and children with WFD. These analyses will establish the target empirical effects to be captured by the computational simulations.

Chapter 5

Secondary Analysis of Developmental Trajectories of Language Skills

Introduction

Empirical data used for this part of the analysis were taken with permission from the WORD project, a large study designed to investigate interventions for children with WFD (Best et al., 2021; Best et al., 2013). Developmental trajectories analyses were constructed for four core language skills for the purpose of simulating similar trajectories first for typical language development then atypical language development; and to then simulate various theory-based interventions to investigate their effectiveness in different cases of WFD. The trajectory analyses served to characterise the developmental characteristics of the language skills of the group of children with WFD compared to those with typical language development. Although there are limitations to a cross-sectional analysis in the case of WFD, because of the heterogeneous nature of the difficulties (Messer & Dockrell, 2006), one can at least check the validity of the most salient characteristics of WFD: that children with WFD are better at individual word comprehension than word production (Messer & Dockrell, 2006; German, 1986, 2000).

Hence comparisons of language skills between TD children and the same age children with WFD were expected to show deficits of language production for children with WFD but were not expected to show deficits in comprehension (Messer & Dockrell, 2006; German, 1986, 2000). Trajectory analyses could potentially reveal whether comprehension developed ahead of word production and whether these characteristics were consistent or were emergent effects across the age range included in this study. Since it was not clear how the different tasks were calibrated regarding how easy or difficult they were for the children of different ages, comparison of task performance levels alone would not be informative. It is the developmental relationship that was the main interest in this comparison of the two tasks, and the question of whether they developed at the same rate. However, the most important aspect of this comparison was to investigate whether the developmental relationship looked the same in the trajectories for TD children and for children with WFD. Correlations of production skills with phonological and semantic skills were also calculated to look for possible causes of the production deficits. Finally, trajectories of simple choice reaction times were included in the study in order to reveal or rule out whether the generally slower responses in naming from children with WFD (Dockrell, Messer, George, 2001) were due to motor difficulties, in the form of a general slowing of processing and motor responses, rather than an effect specific to language production.

The main results were therefore predicted as follows:

(i) The WFD group would show lower word production ability than the TD group but would show similar ability at comprehension.

(ii) The WFD group would show a bigger gap in the development of word production compared to comprehension (i.e., a bigger main effect of task).

(iii) The WFD group would show a slower developing word production ability compared to comprehension when compared to the TD group.

Method

A sample of 100 typically developing (TD) children, with age range from 4;0 years to 8;6, and a group of 24 children with WFD, with age range from 6;3 to 8;7, were tested on four core language tasks to give cross-sectional developmental trajectories on the tasks. One of the participants in the TD group had missing data for one of the core tasks, which was the Children's non-word repetition task.

Participants. Children included in the study either had English as the main language at home or had been in an English-speaking nursery since the age of 3 years and had at least one adult who spoke English at home. The most defining criteria for inclusion of children with WFD were (a) they showed a discrepancy between comprehension and production as measured by the Test of Word-Finding Second Edition (TWF-2, German, 2000), (b) their score had to be above the 7th percentile for BAS Pattern Construction Task, and (c) they did not have any other significant developmental diagnosis such as Autism, dyspraxia, ADHD, sensory or behavioural difficulties. Children who did not show word-finding difficulties in discourse when screened by clinicians involved in the project were excluded from the study.

Core Tasks. The four core language tasks were:

(i) *Picture Naming* (PN) to test word production [semantic input to phonological output (SP) component of simulation]. Scores were out of 72 correct. Materials for this confrontation-naming task were taken from Funnell, Hughes and Woodcock (2006) and consisted of 72 black and white line drawings of objects. The test was run by computer software and answers were recorded on a microphone. Naming latencies were also recorded. It was run in three blocks of 24 pictures and children were asked to give a single word for each picture.

(ii) *Word-Picture Verification Task* (WPVT) to test comprehension [phonological input to semantic output (PS) component of simulation]. The maximum score was 72. This test was devised specifically for the WORD project, where children were asked to match a spoken word against a picture but were not required to produce the word themselves. The single word comprehension task was developed using pictures from the PN task and 72 semantically related items were chosen. Pictures were presented one at time on a computer screen at the same time as a pre-recorded spoken word. The picture was presented with a matching verbal label on one occasion and on another occasion, it was presented with a semantically related word. The child was asked whether the spoken word corresponded to the picture. A score of 1 was given only if the child both accepted the correct name and rejected the semantically related one.

(iii) *Picture Judgement* (PJ) task to test semantic knowledge of children [semantic input to semantic output (SS) component of simulation]. The maximum score for this task was 20. This task required children to look at and point to pictures without the need to listen to or produce the phonological form of the words. A triad of pictures of objects from Funnell et al. (2006) and Druks and Masterson (2000) (as cited in Best et al., 2015) were presented on a computer screen where the target picture was at the top and two pictures were below. One of the two lower pictures was semantically associated to the target picture and the other item was from the same semantic category (for example a *tie* was presented as target with the associated item *shirt* and the distracting item *shorts*). The child was asked which of the two lower pictures fitted best with the top item.

(iv) *Children's Non-word Repetition Task* (CNRep), to test phonology [phonological input to phonological output (PP) component of simulation]. The maximum score was 40. This test is widely used to investigate children's ability to repeat unfamiliar phonological forms (Gathercole & Baddley, 1996). The test consisted of 40 non-words that increased in length and complexity with scores reported as standard scores and percentage correct.

In addition, reaction times were tested with a simple choice speed of reaction time task. For this task the sample of TD group consisted of only 52 children, out of the same sample of 100 TD children, who were chosen because their age range overlapped with the sample of 24 children with WFD. Hence the sample of TD group consisted of 52 children with age range from 6;3 to 8;6 and the WFD group sample consisted of 24 children with age range from 6;3 to 8;7. The children were assessed on a computerised choice reaction time task that was adapted from Powell, Stainthorp, Stuart, Garwood and Quinlan (2007) to measure general processing speed. The idea was to measure the response time, that is, the time taken for the child to press a key once a target picture appeared on the screen. The targets were two dinosaurs of two different colours. The child was asked to press the left control button as quickly as possible when the green dinosaur appeared and the right control button as soon as the orange dinosaur appeared. Full details of the procedures for taking the scores of the four core tasks and the reaction time task can be found in Appendix 1 in Best et al. (2015).

Results

In preparation for the computational modelling part of this research, which investigated the effectiveness of certain intervention techniques for difficulties in language development, it was important to first characterise the typical development trajectory for word acquisition. Hence, we begin the results section by characterising the TD trajectory of word acquisition for our sample of 100 TD children including confidence intervals of its gradient and a regression analysis to predict picture naming accuracy score based on age. A correlation matrix between the four tasks is then presented to bring to light any developmental relationships between the tasks for TD children. We then contrast developmental trajectories of the four core language tasks for our sample of TD children, including a direct comparison of the cross-sectional developmental trajectories for the two most relevant language tasks (measuring word production and comprehension) carried out by the same TD group using a repeated measures design. A similar sequence of results is then presented for the sample of children with WFD starting with the correlation matrix between the four core tasks; followed by a contrast of the overall pattern of developmental trajectories of the four core language tasks across age; and ending with a direct comparison of the cross-sectional developmental trajectories for the tasks measuring word production and comprehension, carried out by the same WFD group using a repeated measures design.

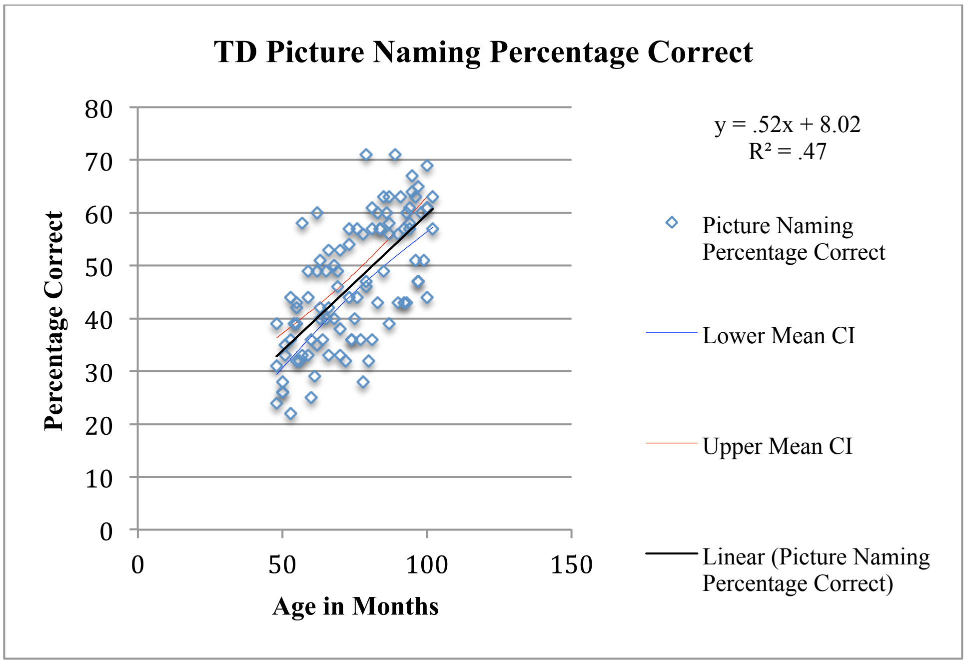
In the third step, between group comparisons are established for trajectories for each of the four tasks. Building on the previous results, a mixed design ANOVA with Group as a between-subjects factor and Task as a within-subjects factor is presented to show whether the disorder group showed the same relationship between the development of the two abilities, comprehension and word production, as the TD group. Finally, a comparison of reaction time trajectories between TD children and children with WFD is presented to illustrate whether there is a difference between the two groups across age, using log-log transformed data to accommodate the linear methods.

For all the following trajectory comparisons, since we planned to compare the intercept values when comparing the onset of any two trajectories, it was necessary to rescale the x-axis so that it measures the ‘age in months from the youngest age measured’ (Thomas et al., 2009). This ensured that group comparisons of the intercept occurred at the first part of the overlap of the trajectories. In this study, the youngest age measured was 75 months for the WFD group, so this number was subtracted from the age for each case and labelled Age_mya (Age_mya = Age-75).

Characterising the TD Trajectory of Word Production. A regression analysis was run using SPSS to derive an equation to predict PN percentage correct based on age. The generated model explained 47% of the variance in PN score (R^2 = .47, F(1, 98) =

85.59, $p < .001$). Unstandardised regression coefficients showed an intercept constant of 8.02 and a gradient of 0.52 giving the equation for *Predicted Picture Naming percentage correct as 0.52 x age in months + 8.02*.

Figure 5.1. Graph showing the developmental trajectory of PN task (word production) versus age for TD children including the upper and lower mean confidence intervals.



As can be seen in Table 5.1 and the graph in Figure 5.1, for our sample TD trajectory, the significant t-test and the upper and lower bounds of the 95% confidence interval above zero for age show a reliable improvement on the PN task (word production) as the TD children became older. The confidence intervals around the gradient allows us to additionally check whether the WFD group gradient falls within these limits.

Table 5.1. Coefficients and confidence intervals of the intercept and gradient of the developmental trajectory for correct percentage of PN task (word production) for TD children.

Constant & Predictor Variables	Unstandardised Regression Coefficients		Standardised Beta Coefficient	t	p value	95% Confidence Interval for B	
	B	Std Error				Lower Bound	Upper Bound
Constant	8.02	4.28		1.87	.064	-0.47	16.52
Age in Months	.52	0.06	.68	9.25	.000	0.41	0.63

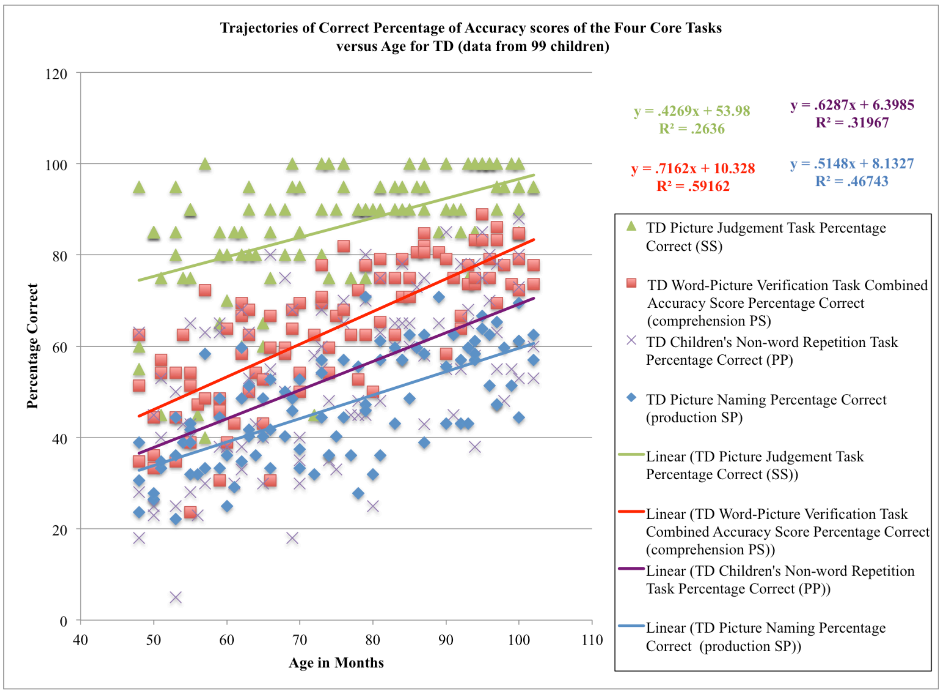
Correlations for the Four Core Tasks of TD Children. Table 5.2 presents a correlation matrix between scores of the four core tasks for TD children. Results showed a significant positive correlation between all the four core tasks indicating that the tasks were developing in tandem with each other. The highest value for Pearson correlation was between PN task and WPVT $r(100) = .772$, $p < .001$, showing that for our sample of TD children their abilities in word production and comprehension were highly correlated. Results of the PN task scores for TD children also showed a strong correlation with scores of the CNRep task, $r(99) = .684$, $p < .001$, indicating that for our sample of TD children their abilities in word production were highly correlated with their abilities in non-word repetition. Considering that the analysed data were cross-sectional these significant correlations between all the four tasks demonstrate that each skill normally improves with age for TD children and in this sample the children's ages spanned a range.

Table 5.2. Correlation matrix for the percentage correct of scores for the four core tasks, Picture naming task (PN), word-picture verification task (WPVT), picture judgement task (PJ), and children's non-word repetition task (CNRep) for TD children.

TD Correlations

		WPVT	PJ	CNRep
PN	Pearson Correlation	.772**	.519**	.684**
	Sig. (2-tailed)	<.001	<.001	<.001
	N	100	100	99
WPVT	Pearson Correlation		.475**	.604**
	Sig. (2-tailed)		<.001	<.001
	N		100	99
PJ	Pearson Correlation			.451**
	Sig. (2-tailed)			<.001
	N			99

Figure 5.2. Graph showing the developmental trajectories of the four core tasks for 99 TD children.



Trajectories of the Four Core Tasks for Typical Development. The overall pattern of developmental trajectories of the four core tasks depicted by the sample of 99 TD children is shown in Figure 5.2 (one case was omitted because there was no score for CNRep task). Scores for the PJ task are notably displayed near ceiling level on the graph, which brings into question whether this particular task was too easy for some TD children. The data furthermore suggest that comprehension (WPVT) develops ahead of word production (PN task) and from the gradients given in the relative equations (a gradient of 0.72 for comprehension and a gradient of 0.52 for production) the data could be interpreted as showing that comprehension develops at almost 40% faster rate than production. The difference in gradients between PN task and WPVT was analysed further using a repeated measures design and details of these results are in the next section.

Comparing Cross-sectional Developmental Trajectories for Two Tasks (Measuring Word Production and Comprehension) Carried out by the Same TD group of Children: [Task x Age]: A comparison of the cross-sectional trajectories for the sample of 100 TD children, for the PN task and WPVT, was made using a repeated measures design.

Results of a within subjects analysis showed a significant main effect of task where children were better at WPVT (comprehension) than at the PN task (word production) ($F(1,99) = 326.93$, $p < .001$, $\eta_p^2 = .77$). When the rescaled age was added as the covariate, results showed that there was an overall significant improvement of performance with age ($F(1,98) = 148.13$, $p < .001$, $\eta_p^2 = .60$). There was also a significant interaction between age and task ($F(1,98) = 12.36$, $p = .001$, $\eta_p^2 = .11$) indicating that WPVT develops at a significantly faster rate than PN for TD children. This interaction is reflected in the graph in Figure 5.3. The coefficients for the intercepts and gradients, confidence intervals and effect sizes for the trajectories of each task are provided in Table 5.3. Taken altogether, analysis of the trajectories within the group suggest that TD children perform better at WPVT than PN at onset and they continue improving at WPVT (comprehension) at almost 30% faster rate than their rate of improvement at PN (word production).

Figure 5.3. Comparison of trajectories of correct percentage of accuracy scores of the PN task (word production) and WPVT (comprehension) versus age for TD children.

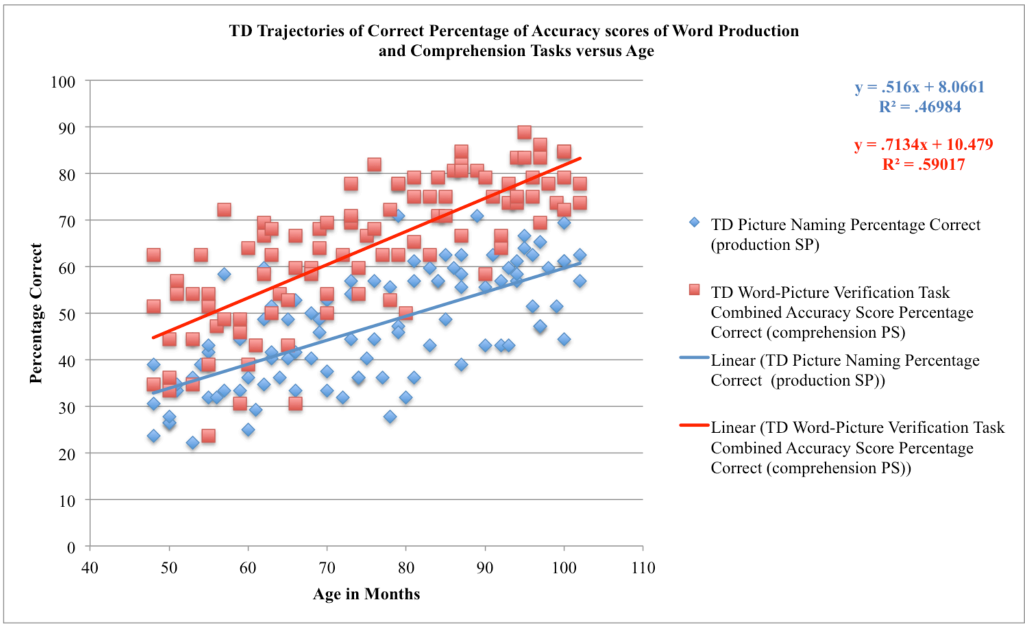


Table 5.3. Parameter Estimates for word production (PN task) and comprehension (WPVT) trajectories of TD group.

Dependent Variable	Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Partial Eta Squared
						Lower Bound	Upper Bound	
PN Percent Corr	Intercept	32.83	1.75	18.76	.000	29.36	36.30	.78
	TD_Age_mya	.52	.06	9.25	.000	.41	.63	.47
WPVT Percent Corr	Intercept	44.74	1.88	23.85	.000	41.02	48.46	.85
	TD_Age_mya	.72	.06	11.94	.000	.60	.83	.59

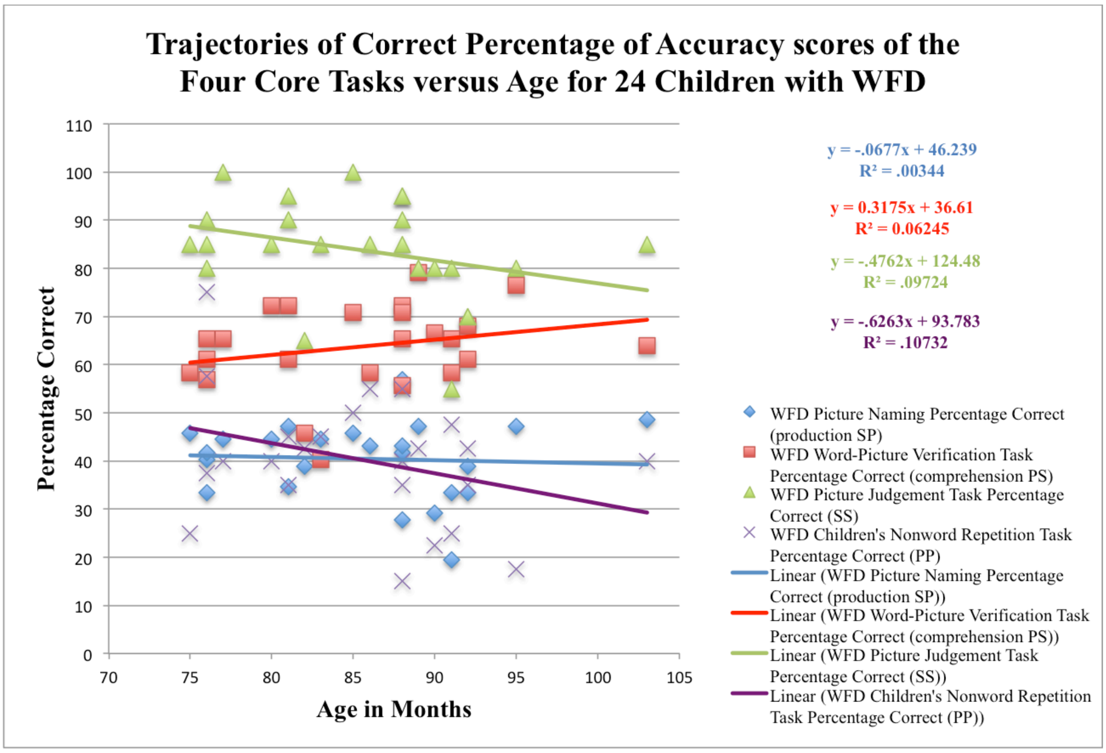
Correlations for the Four Core Tasks of Children with WFD. In contrast to the sample of TD children, for our sample of children with WFD results did not show any significant correlations between any two tasks (see Table 5.4). These preliminary results were not surprising for the case of children with WFD when one considers that the data were cross-sectional, hence for this sample, younger children could have had less severe deficits than the older ones, that is, the level of severity of language deficits in the children could be uncorrelated with their age.

Table 5.4. Correlation matrix for the percentage correct of scores for the four core tasks, Picture naming task (PN), word-picture verification task (WPVT), picture judgement task (PJ), and children's non-word repetition task (CNRep) for children with WFD.

WFD Correlations

		WPVT	PJ	CNRep
PN	Pearson Correlation	.330	.302	.226
	Sig. (2-tailed)	.115	.152	.288
	N	24	24	24
WPVT	Pearson Correlation		.174	-.046
	Sig. (2-tailed)		.415	.831
	N		24	24
PJ	Pearson Correlation			.014
	Sig. (2-tailed)			.949
	N			24

Figure 5.4. Graph showing the developmental trajectories of the four core tasks for 24 children with WFD**.**



Trajectories of the Four Core Tasks for WFD. The overall pattern of developmental trajectories of the four core tasks depicted by the sample of 24 children with WFD is shown in Figure 5.4. The only apparent similarity with the pattern of four tasks given by the sample of TD children is that the data suggest that comprehension (WPVT) develops ahead of word production (PN task) and some of the scores for the PJ task are shown to be close to ceiling. The difference in gradients between WPVT (a gradient .32 for comprehension) and PN task (a gradient of -.07 for word production) does not appear to be large on the graph. However, this difference was analysed further using a repeated measures design and details of these results are given in the next section.

Comparing Cross-sectional Developmental Trajectories for Two Tasks (Measuring Word Production and Comprehension) Carried out by the Same Group of Children with WFD [Task x Age]: A comparison of the cross-sectional trajectories for the sample of 24 children with WFD for the PN task and WPVT, was made using a repeated measures design.

Results of a within subjects analysis showed a significant main effect of task where children were better at WPVT (comprehension) than at the PN task (word production) ($F(1,23) = 132.33$, $p < .001$, $\eta_p^2 = .85$). When the rescaled age was added as the covariate, results showed that there was no significant improvement of performance with age ($F(1,22) = 0.39$, $p = .54$, $\eta_p^2 = .02$). The results also showed no interaction between age and task ($F(1,22) = 1.78$, $p = .195$, $\eta_p^2 = .08$) indicating that there was no difference in the rate of change between the two tasks versus age for the WFD group. This is reflected in the similar gradients for word production and comprehension in the graph in Figure 5.5. Parameter estimates, confidence intervals and effect sizes of the trajectories are provided in Table 5.5.

Figure 5.5. Comparison of trajectories of correct percentage of accuracy scores of the PN task (word production) and WPVT (comprehension) versus age for children with WFD.

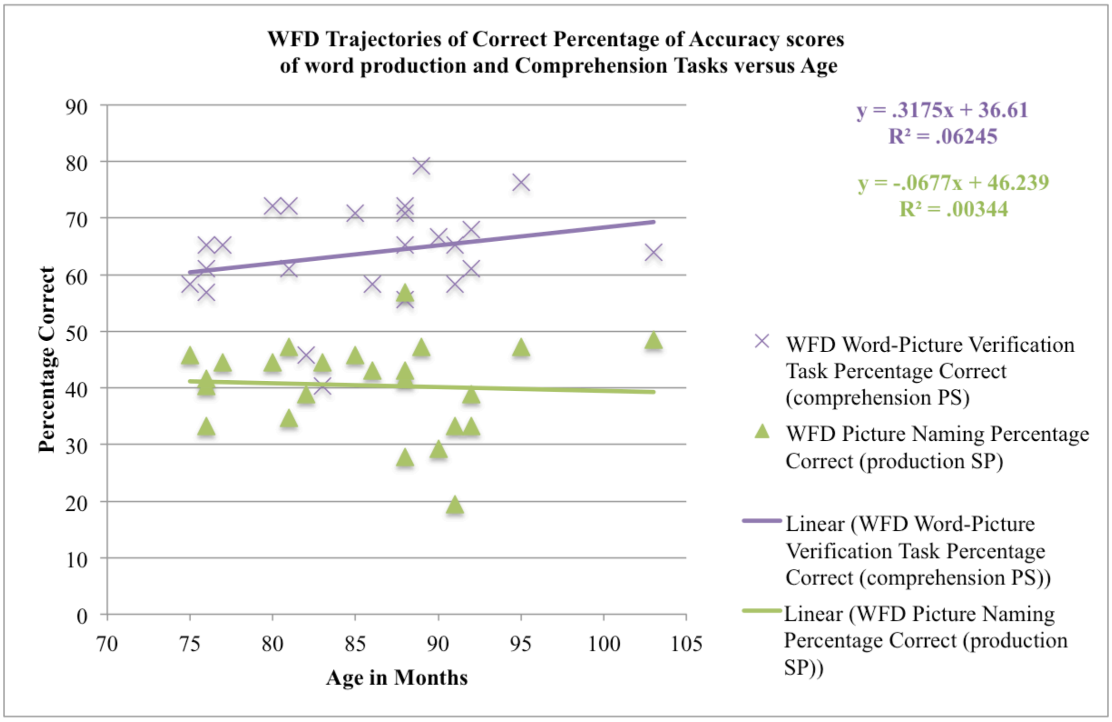


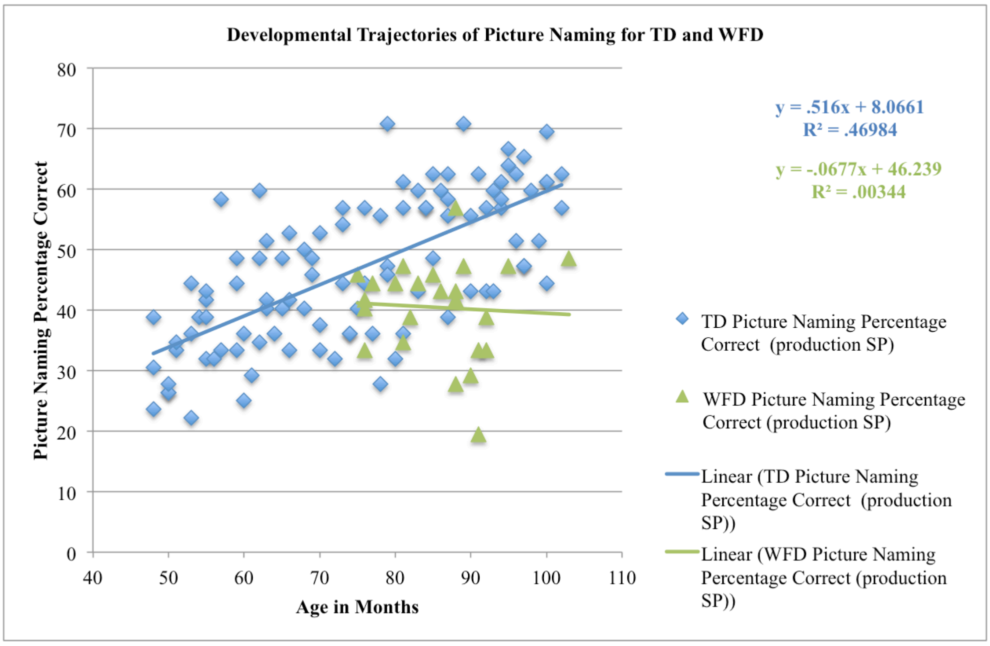
Table 5.5. Parameter Estimates for word production (PN task) and comprehension (WPVT) trajectories of WFD group.

Dependent Variable	Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound	Partial Eta Squared
PN Percent Corr For WFD	Intercept	41.03	3.11	13.18	.000	34.57	47.49	.89
	Age_mya	-.06	.25	-.25	.804	-.58	.45	.00
WPVT Percent Corr For WFD	Intercept	60.26	3.29	18.31	.000	53.43	67.08	.94
	Age_mya	.32	.26	1.24	.229	-.22	.87	.07

Comparison of Developmental Trajectories of Each of the Four Core Tasks between TD Children and Children with WFD. [Group x Age]: In this section we compare the developmental trajectories of each of the four core tasks between typical development and WFD by using SPSS to adapt the Analysis of Covariance function within the General Linear Model (ANCOVA).

Picture Naming Task. Tests of between subjects effects for the PN task between the TD (100 cases) and WFD (24 cases) groups produced an overall R^2 value of 0.45 showing that the model explained about 45% of the variance ($F(3,120) = 32.59$, $p < .001$, $\eta_p^2 = .45$]. Results showed no significant effect of group ($F(1,120) = 2.72$, $p = .102$, $\eta_p^2 = .02$). But from looking at Figure 5.6 one can see that there did not seem to be a big difference in the ability of word production between our sample of typically developing children and our sample of children with WFD at the youngest age measured for the latter group in this study. However, the graph shows that the difference in word production ability between the two groups increased with age. A non-significant result was shown for an interaction of level of performance according to age for the combined groups, giving a main effect of age as ($F(1,120) = 2.83$, $p = .095$, $\eta_p^2 = .02$). Again, the graph in Figure 5.6 shows that word production of typically developing children increased with age but for our sample of children with WFD word production seems to have decreased slightly with age. However, a significant result was observed for the Group x Age_mya interaction ($F(1,120) = 4.78$, $p = .031$, $\eta_p^2 = .04$), indicating that there was a generally bigger difference in word production skills between our sample of children with WFD and typically developing children at the older ages measured (up to 8 years and 7 months) than at the younger ages measured (at 6 years and 3 months) in this study. The gradient values show that for our samples the children with WFD became worse at picture naming as they became older by about 13% of the rate that typically developing children improved at picture naming. The intercept values show that at the youngest age measured for our samples the typically developing children performed better at the picture naming task than the children with WFD only by around 5.6%. Hence these results show that the cross-sectional trajectories for the two groups started off at the same level, but thereafter, the TD group improved with age, but the WFD group did not.

Figure 5.6. Graph showing the developmental trajectories of word production (PN task) for TD children (100 cases) and children with WFD (24 cases).



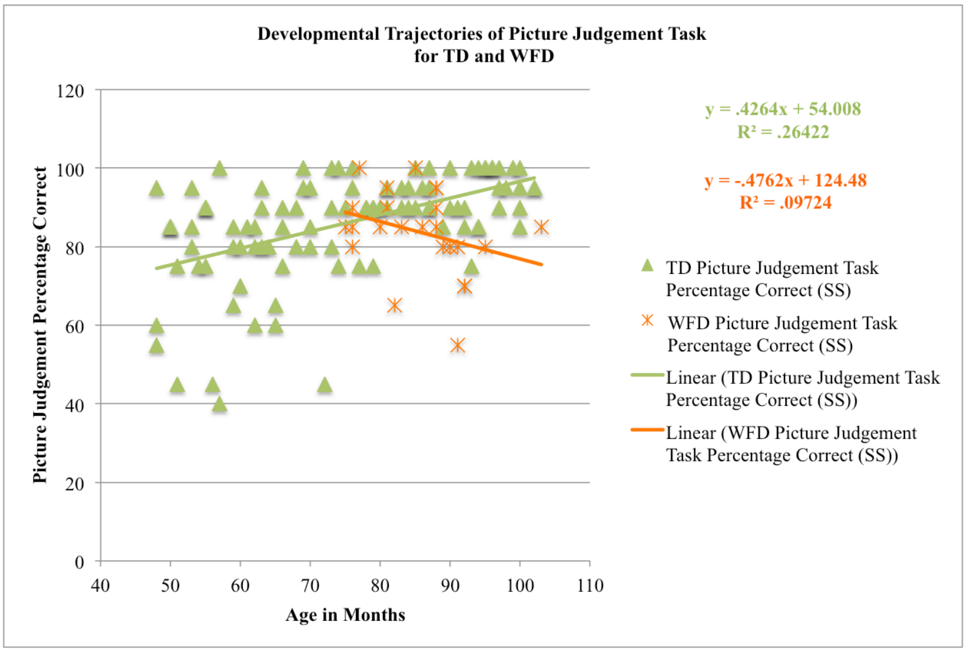
Word-Picture Verification Task. Tests of between subjects effects for the WPVT task between the TD (100 cases) and WFD (24 cases) groups produced an overall R^2 value of .55 showing that the model explained about 55% of the variance ($F(3,120)$ = 49.25, $p < .001$, η_p^2 =.55). Results showed no significant effect of group (F(1,120) = 0.98, p = .324, η_p^2 = .01). Similarly, to the picture naming trajectories, from looking at Figure 5.7, it is apparent that there was only a small difference in the ability of WPVT between our sample of TD children and our sample of children with WFD at the youngest age measured for the latter group in this study. However, in the case of WPVT results showed a significant effect of the combined groups versus age predicting an increased level of performance with age (F(1,120) = 13.00, p < .001, η_p^2 = .10). The graph in Figure 5.7 shows that both groups improved at WPVT with age. Interestingly, the non-significant Group x Age_mya interaction shows that there was no reliable difference in the rate of development of comprehension between the TD and WFD groups ($F(1,120) = 1.93$, $p = .168$, $\eta_p^2 = .02$). This suggests that the groups were indistinguishable in their development on comprehension measure. The difference in performance at the youngest age measured is shown by the intercept values where for our samples the TD children performed better than the children with WFD at WPVT by only around 4%. So statistically, on the comprehension measure, children with WFD did not differ from TD children, unlike the case with word production in the previous comparison where the WFD group were shown to be developing at a slower rate than that of TD children.

Figure 5.7. Graph showing the developmental trajectories of comprehension (WPVT) for TD children (100 cases) and children with WFD (24 cases).

Since the trajectory of scores for comprehension (WPVT) for children with WFD included in this study showed an improvement in scores, as children became older, a regression analysis was run using SPSS to check whether we could derive a reliable model to predict WPVT percentage correct based on age. The generated model explained only 7% of the variance in WPVT scores and did not show a reliable improvement of comprehension with age for children with WFD (R^2 = .07, F(1, 23) = 1.53, p = .229). Unstandardised regression coefficients showed an intercept constant of 36.00 and a gradient of 0.32.

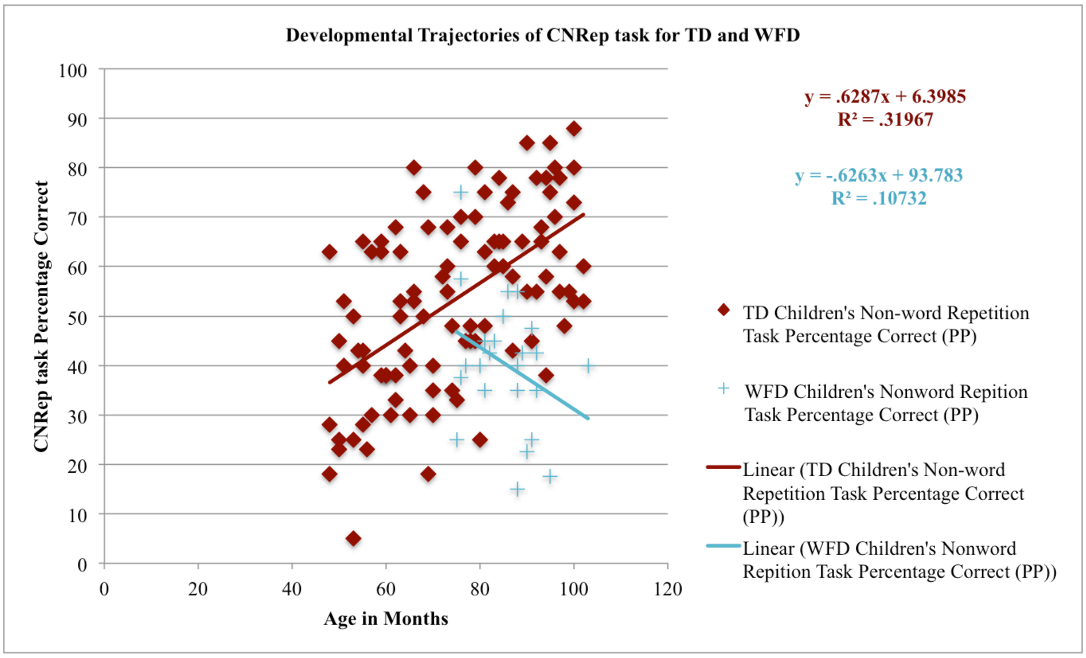
Picture Judgement Task. Tests of between subjects effects for the PJ task between the TD (100 cases) and WFD (24 cases) groups produced an overall R^2 value of .25 showing that the model explained 25% of the variance (F(3,120) = 13.04, p < .001, η_p^2 = .25). There was no significant overall effect of group (F(1,120) = 0.40, p = .526, η_p^2 = .003). The trajectories in Figure 5.8 show that at the youngest age measured for the WFD group there was no reliable difference in the scores of PJ task between the two groups. Likewise there was a non-significant result for the interaction between the overall performance of the combined groups versus age, giving a main effect of age as (F(1,120) = .02, p = .885, η_p^2 = .000). However, parallel to the trajectories of picture naming task, Figure 5.8 shows that the scores of TD children in the PJ task improved with age whereas the scores of those in the WFD group decreased with age. Results also showed a significant Group x Age_mya interaction (F(1,120) = 6.9, p = .010, η_p^2 = .06). In the case of our samples this indicates that the WFD group became worse at the PJ task at a faster rate than the TD group improved at this task. At the youngest age measured, the two groups had very similar scores where the children with WFD scored better than the TD group at the PJ task by about 3%. However, it was noted that the scores for the PJ task were close to ceiling for both groups of children, with mean scores over 80% correct, which could potentially lead to artefactual interactions and misleading results. There is no strong evidence here that the children with WFD overall had a semantic deficit.

Figure 5.8. Graph showing the developmental trajectories of semantics (PJ task) for TD children (100 cases) and children with WFD (24 cases).



Children's Non-word Repetition Task. Tests of between subjects effects for the CNRep task between the TD (99 cases) and WFD (24 cases) groups produced an overall R^2 value of .35 showing that the model explained 35% of the variance ($F(3,119) = 21.74$, $p < .001$, $\eta_p^2 = .35$). Results showed no significant effect of group ($F(1,119) = 1.34$, $p = .250$, $\eta_p^2 = .01$). This result is reflected in the graph of Figure 5.9, which shows that scores of the WFD group fell among the distribution of TD scores at the youngest age measured for children with WFD. However, the graph shows that scores of children with WFD decreased with age whereas scores of TD children increased with age. Results were non-significant for the main effect of age ($F(1,119) = 0.00$, $p = .994$, $\eta_p^2 = .000$). But there was a significant effect of Group x Age_mya interaction ($F(1,119) = 8.13$, $p = .005$, $\eta_p^2 = .06$) and this is illustrated in Figure 5.9 where scores of TD children improved with age whereas scores of children with WFD declined sharply with age. Parameter estimates showed that for our samples, the scores of CNRep for TD children improved at about the same rate as the scores of children with WFD decreased with age. At the youngest age measured, the TD group scored better than the WFD group at the CNRep task by about 7%. There is some evidence of phonological deficits in the WFD children, particularly for the older individuals.

Figure 5.9. Graph showing the developmental trajectories of phonology (CNRep task) for TD children (99 cases) and children with WFD (24 cases).



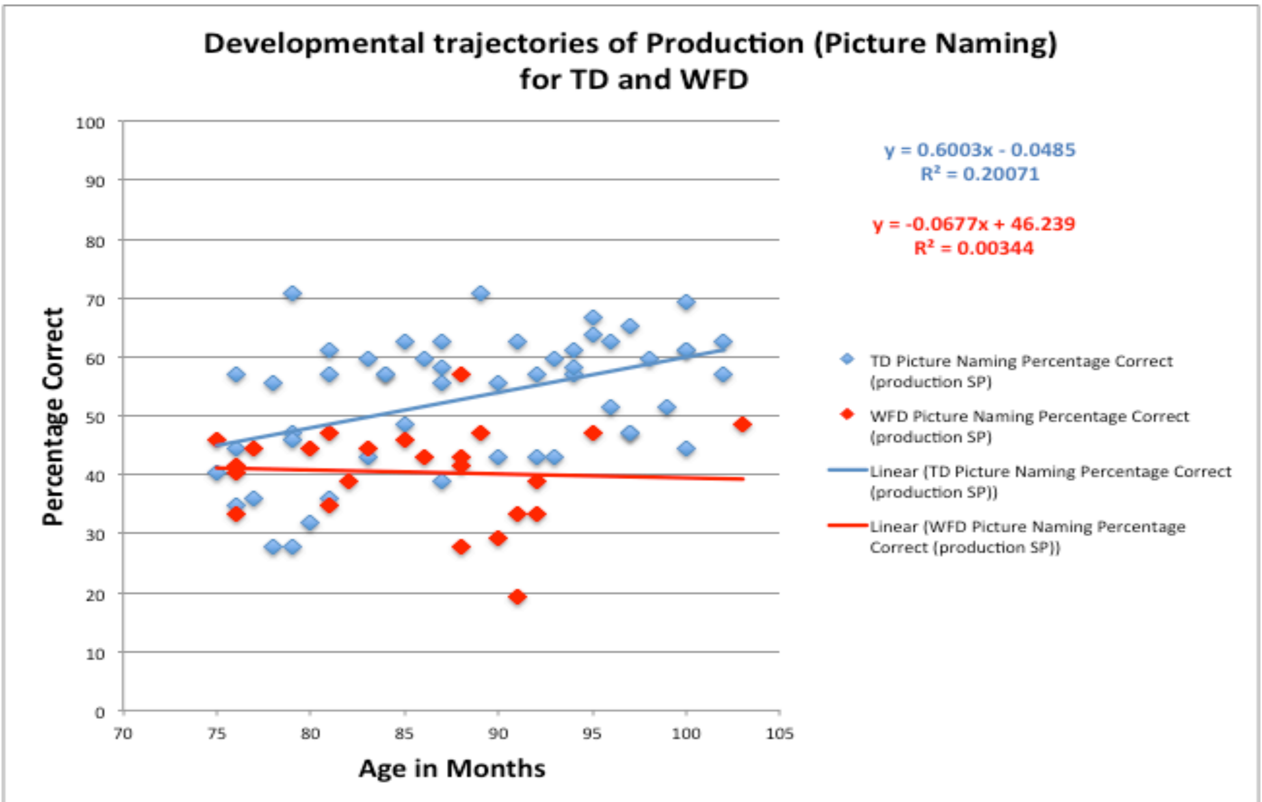
In these comparisons, TD trajectories included scores from children younger than those in the WFD group. To check the robustness of the between-group comparisons and ensure that non-overlapping portions of the trajectories were not skewing the results, the same analysis procedure for comparison of developmental trajectories of each of PN task, WPVT, PJ task and CNRep task between TD and WFD, as above, was repeated including only the TD children whose age range overlapped with those of the sample of children with WFD. In the latter case, the sample of TD children included 52 participants in all tasks (except in the case of CNRep, the sample size was 51 participants) with a minimum age of 6;3 and a maximum age of 8;6. The sample of children with WFD included 24 participants with age range between 6;3 and 8;7 in all four tasks. All the comparisons of trajectories that included only 52/51 TD children yielded the same qualitative pattern of results including all significant and non-significant group main effects and interactions as the previous analysis which included all the 100/99 TD children. Results of the trajectories analyses that included only the TD children whose age range overlapped with those of the sample of children with WFD are shown below.

Comparison of Developmental Trajectories of Each of the Four Core Tasks between TD children (52/51 cases) and children with WFD (24 cases). [Group x Age]

When we only compare trajectories that include the age range that overlaps for both TD and WFD groups then the sample of typically developing children includes 52 participants (except for the CNRep trajectory, 51 cases were included) with a minimum age of 6 years and 3 months and a maximum age of 8 years and 6 months. The sample of children with WFD includes 24 participants with age range between 6 years and 3 months and 8 years and 7 months. Following are comparisons of the developmental trajectories of each of the four core tasks between typical development and WFD that were analysed using SPSS to adapt the Analysis of Covariance function within the General Linear Model (ANCOVA).

Picture Naming Task. Tests of between subjects effects for the PN task between the TD (52 cases) and WFD (24) groups produced an overall R^2 value of .37 showing that the model explains only 37% of the variance ($F(3,72) = 14.36$, $p < .001$, $\eta_p^2 = .37$). Results showed no significant effect of group ($F(1,72) = 0.77$, $p = .383$, $\eta_p^2 = .01$). But from looking at figure 5.10 one can see that there did not seem to be a big difference in the ability of word production between our sample of TD children and our sample of children with WFD at the youngest age measured for the latter group in this study. However the graph indicates that the difference in word production ability between the two groups increased with age.

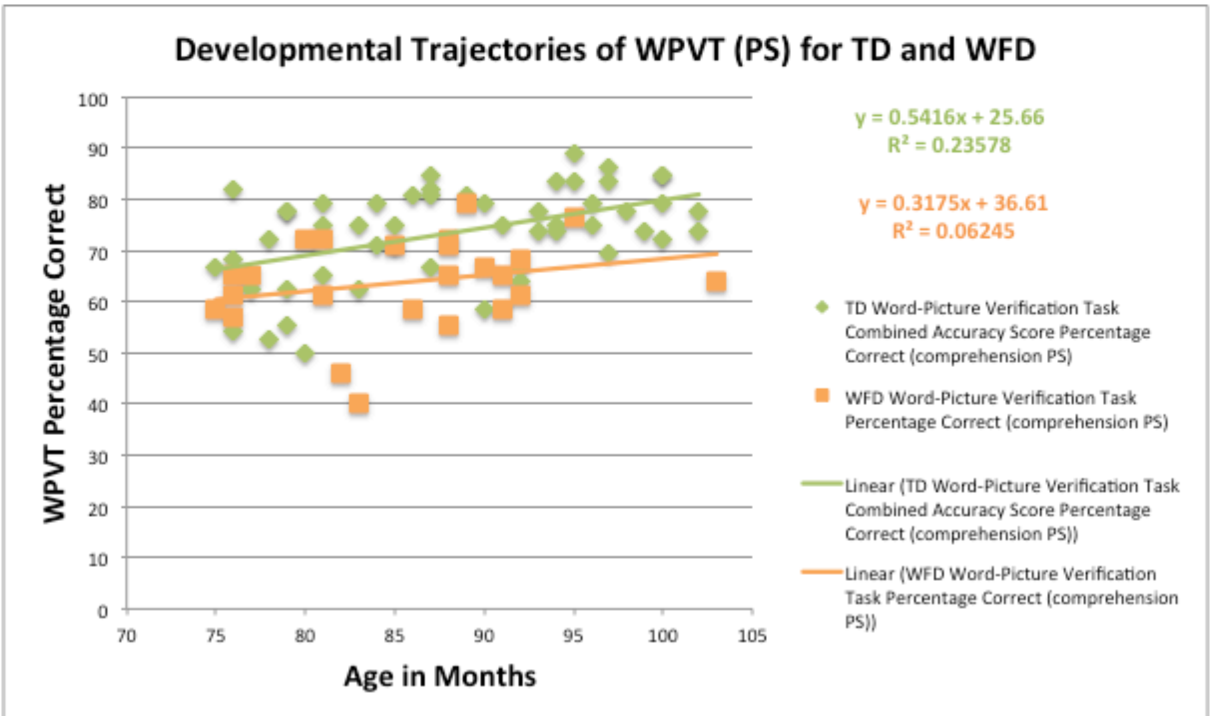
Figure 5.10. Graph showing the developmental trajectories of word production (PN task) for TD children (52 cases) and children with WFD (24 cases).



A non-significant result was shown for an interaction of level of performance according to age for the combined groups (F(1,72) = 2.72, p = .103, η_p^2 = .04). Again the graph in figure 5.10 shows that word production of typically developing children increased with age but for our sample of children with WFD word production seems to have decreased slightly with age. However a significant result (albeit not to a high power) was produced for the Group x Age_mya interaction (F(1,72) = 4.28, p = .042, η_p^2 = .06), indicating that word production skill of children with WFD was developing at a slower rate than that of typically developing children. The gradient values show that, for our samples, the children with WFD became worse at picture naming as they became older by about a tenth of the rate that typically developing children improved at picture naming. The intercept values show that at the youngest age measured, for our samples, the typically developing children performed better at the picture naming task than the children with WFD by around 4%.

Word-Picture Verification Task. Tests of between subjects effects for the WPVT task between the TD (52 cases) and WFD (24 cases) groups produced an overall R^2 value of .35 showing that the model explained about 35% of the variance ($F(3,72) = 12.97$, $p < .001$, $\eta_p^2 = .35$). Results showed no significant effect of group ($F(1,72) = 2.35$, $p = .130$, $\eta_p^2 = .03$).

Figure 5.11. Graph showing the developmental trajectories of comprehension (WPVT) for TD children (52 cases) and children with WFD (24 cases).

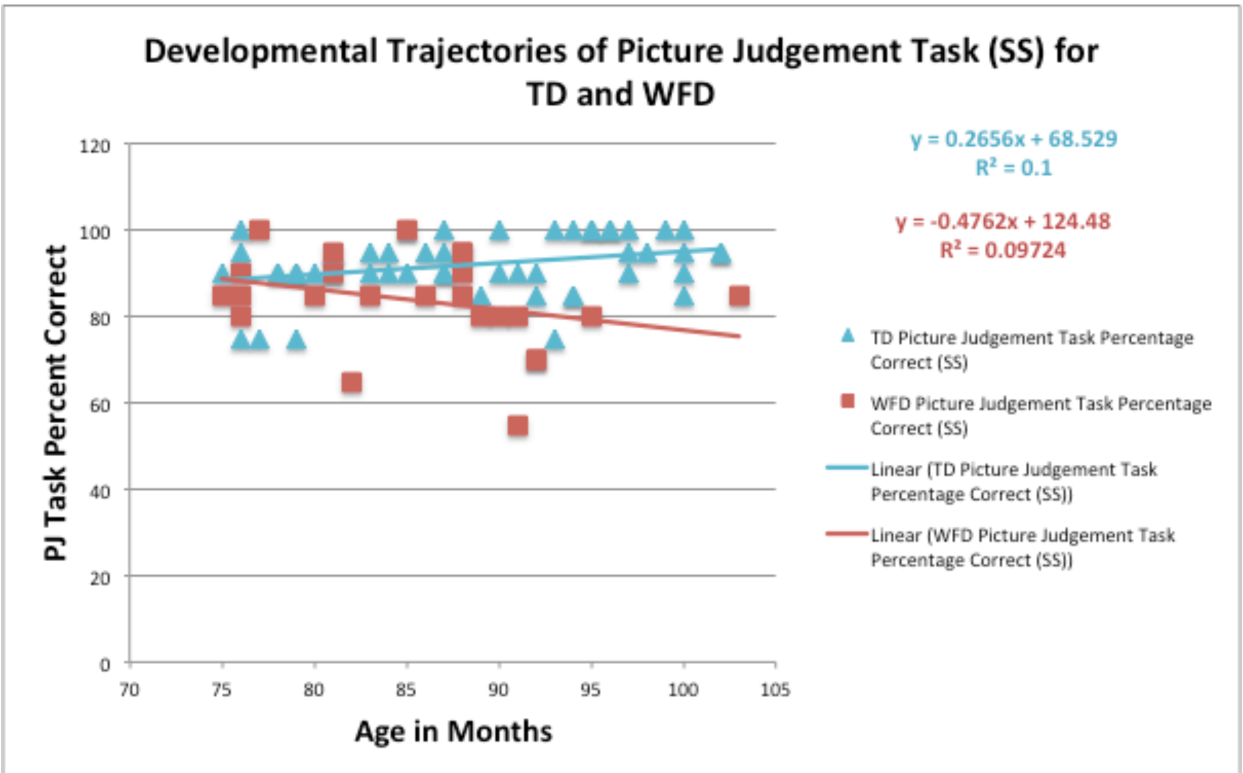


Similarly to the picture naming trajectories from looking at figure 5.11 it is apparent that there was only a small difference in the ability of WPVT between our sample of TD children and our sample of children with WFD at the youngest age measured for both groups in this study. But in this case the graph indicates that the difference in WPVT ability (or comprehension) between the two groups increased only slightly with age. Also in the case of WPVT results showed a significant effect of the combined groups versus age predicting an increased level of performance with age. ($F(1,72) = 9.14$, $p = .003$, $\eta_p^2 = .11$). The graph in figure 5.11 shows that both groups improved at WPVT with age. The non significant Group x Age_mya interaction shows that there was no reliable difference in the rate of development of comprehension between the TD and WFD groups ($F(1,72) = 0.62$, $p = .433$, $\eta_p^2 = .01$). Again this observation is reflected in the graph in figure 5.11 and implies that the groups were indistinguishable in their development on comprehension measure.

The difference in performance at the youngest age measured is shown by the intercept values where for our samples the TD children performed better than the children with WFD at WPVT by around 6% [66.278-60.426=5.852]. Hence, similar to the trajectories analysis which included the whole group of 100 TD children, this analysis shows that statistically, on the comprehension measure, children with WFD did not differ from TD children, unlike the case with word production in the previous comparison where the WFD group were shown to be developing at a slower rate than that of TD children.

Picture Judgement Task. Tests of between subjects effects for the PJ task between the TD (52 cases) and WFD (24 cases) groups produced an overall R^2 value of .26 showing that the model explained 26% of the variance (F(3,72) = 8.51, p < .001, η_p^2 = .26). There was no significant overall effect of group (F(1,72) = 0.01, p = .931, η_p^2 = .000).

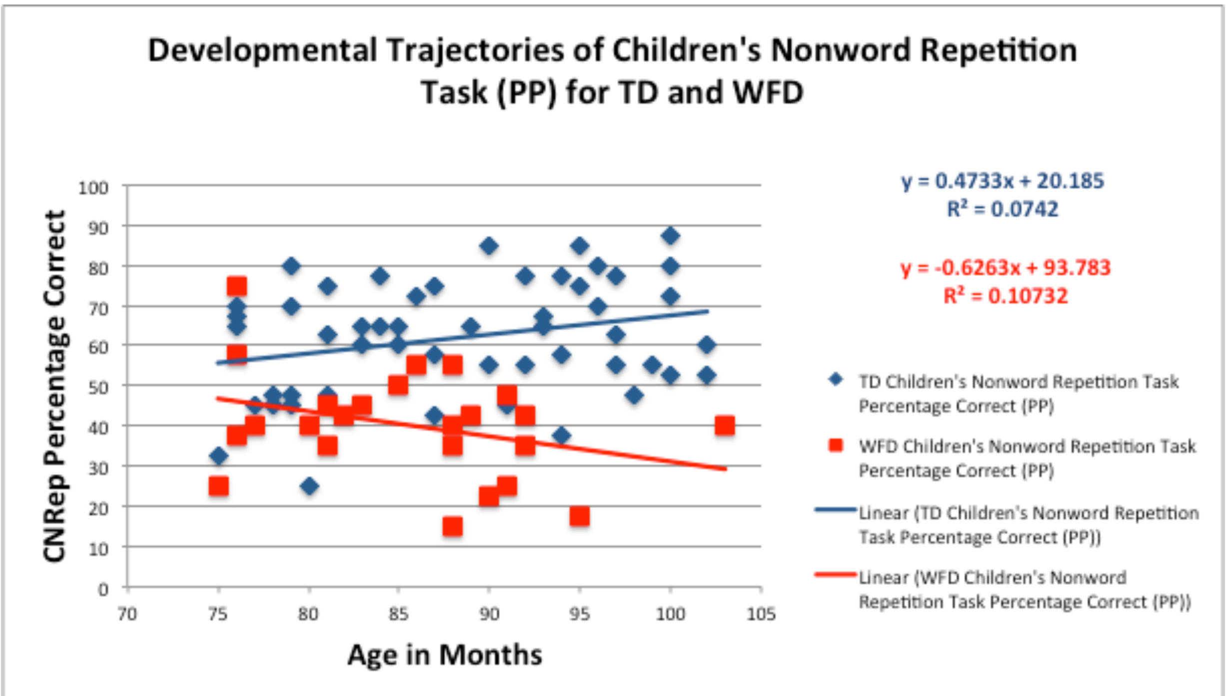
Figure 5.12. Graph showing the developmental trajectories of semantics (PJ tasks) for TD children (52 cases) and children with WFD (24 cases).



The trajectories in figure 5.12 show that at the youngest age measured for the WFD group there was no reliable difference in the scores of PJ task between the two groups. There is a non-significant result for the interaction between the overall performance of the combined groups versus age, where the main effect of age was given as (F(1,72) = 0.60, p = .442, η_p^2 = .01). This result may have been produced because the trajectories were developing in opposite directions and were cancelling each other's effects. Results also showed a significant group x Age_mya interaction (F(1,72) = 7.41, p = .008, η_p^2 = .09). In the case of our samples this indicates that the WFD group became worse at the PJ task at a faster rate than the TD group improved at this task. In the case of our samples, at the youngest age measured, the two groups had very similar scores where WFD group scored better than the TD group at the PJ task by only about 0.3% [88.77-88.45=.32]. However, it was noted that the scores for the PJ task were close to ceiling for both groups of children, which could lead to artefactual interactions and misleading results.

Children's Non-word Repetition Task. Tests of between subjects effects for the CNRep task between the TD (51 cases) and WFD (24 cases) groups produced an overall R^2 value of .40 showing that the model explained 40% of the variance ($F(3,71) = 16.05$, $p < .001$, $\eta_p^2 = .40$). Results showed no significant effect of group ($F(1,71) = 1.99$, $p = .163$, $\eta_p^2 = .03$). Similar to the results produced using the whole group of 99 TD cases, this result is reflected in the graph of figure 5.13 which shows that scores of the WFD group fell among the distribution of TD scores at the youngest age for children with WFD. However the graph shows that scores of children with WFD decreased with age whereas the scores of TD children increased with age at roughly the same rate thereby cancelling the effect for difference in group.

Figure 5.13. Graph showing the developmental trajectories of phonology (CNRep task) for TD children (51 cases) and children with WFD (24 cases).



Results were non significant for the main effect of age ($F(1,71) = 0.11$, $p = .744$, $\eta_p^2 = .002$). But there was a significant effect of group x Age_mya interaction ($F(1,71) = 5.53$, $p = .021$, $\eta^2 = .07$) and this can be seen in figure 5.13 where scores of TD children improved with age whereas scores of children with WFD decreased with age. Parameter estimates showed that for our samples the scores of CNRep for TD children improved at about 76% of the rate that the scores of children with WFD decreased with age [.47/-.63. =.75]. At the youngest age measured, the TD group scored better than the WFD group at the CNRep task by about 9 % [55.68-46.81=8.87].

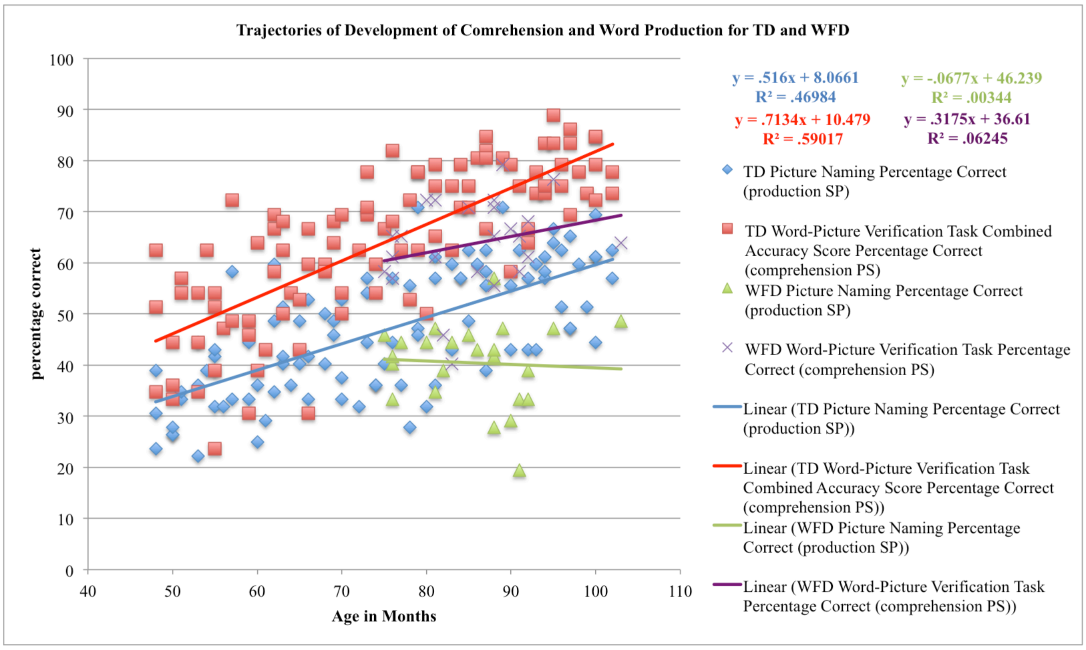
Analysing Mixed Design Linear Regression: Does the disorder group show the same relationship between the development of two abilities (comprehension and word production) as the TD group? [Task x Group]. In this step, scores from the TD sample of 100 children and the WFD sample of 24 children were analysed by running a mixed-design ANOVA with Group as a between-subjects factor and Task as a within-subjects factor.

Results showed that there was a significant main effect of task ($F(1,122) = 345.70$, $p < .001$, $\eta_p^2 = .74$). This is in line with the previous analysis, which revealed that both the TD and WFD groups showed an overall better performance at WPVT (comprehension) than PN task (word production). The rescaled age (Age_mya=Age-75) according to the youngest age measured for the disorder group was added as the covariate.

We must look at results of the 3-way interaction between Task x Group x Age_mya to answer the question of whether the disorder group followed the same developmental relationship between the two tasks (comprehension and word production) as the one followed by TD children. In this case the answer is yes, as shown by the non-significant interaction ($F(1,120) = 0.46$, $p = .500$, $\eta_p^2 = .004$). There was also no Task x Group interaction ($F(1,120) = 0.32$, $p = .572$, $\eta_p^2 = .003$) indicating that there was no difference in relative level of task performance of the two groups at the first point of overlap of the trajectories.

There was no main effect of group ($F(1,120) = 2.28$, $p = .134$, $\eta_p^2 = .02$). This is reflected in the graph in Figure 5.14 which shows that, at the youngest age that the disorder group was measured, the mean task performance of the two groups was comparable. However, from the same graph it is clear that the trajectories between the TD and WFD groups diverge with age. Overall age was a significant predictor of performance [Age_mya: $F(1,120) = 9.53$, $p = .003$, $\eta_p^2 = .07$) but the slower development of the WFD group was reflected in a reliable group by age interaction [Group x Age_mya: $F(1,120) = 4.19$, $p = .043$, $\eta_p^2 = .03$]. Hence, while the preceding comparison of individual trajectories showed that with increasing age, WFD fell behind TD children more so on production than comprehension, in line with their clinical diagnosis, in this combined analysis, which included the age range of the sample of 100 TD children in this study, the stronger pattern was overall slower development in both language abilities in the children with WFD.

Figure 5.14. Graph showing the comparison of developmental trajectories of word production (PN task) and comprehension (WPVT) between TD children (100 cases) and children with WFD (24 cases).



Just like in the previous section a second set of comparison analyses that included only the 52 TD children whose age range overlapped with the children with WFD were done to check that the previous results were not an artefact of comparing non-overlapping trajectories. With the exception of the group by age interaction, this set of comparisons of the trajectories that included only 52 TD children yielded the same qualitative pattern of results including all significant and non-significant group main effects and interactions as the previous analysis which included all the 100 TD children.

In contrast to the analysis which included the 100 cases of TD children, the group by age interaction did not show a reliable difference in the overall development of both language abilities between the TD group and the group of children with WFD when taking into account only the 52 cases of TD children whose age range overlapped with the sample of children with WFD [Group x Age_mya: $F(1,72) = 2.67$, $p = .107$, $\eta_p^2 = .04$]. Results of the comparison between the development of two abilities (comprehension and word production) between the two groups that included only the TD children whose age range overlapped with those of the sample of children with WFD are shown below.

Analysing Mixed Design Linear Regression: Does the disorder group show the same relationship between the development of two abilities (comprehension and word production) as the TD group (52 cases)? [Task x Group]. In this step, scores from the TD sample of 52 children and the WFD sample of 24 children were analysed by running a mixed-design ANOVA with Group as a between-subjects factor and Task as a within-subjects factor.

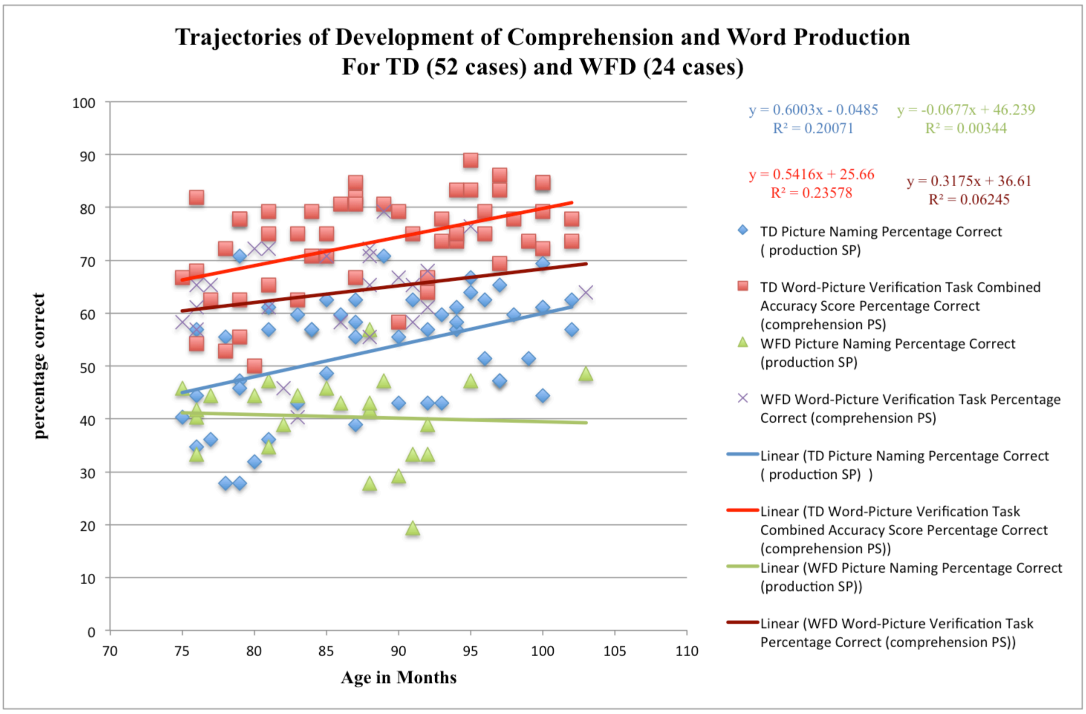
Results showed that there was a significant main effect of task ($F(1,74) = 509.18$, $p < .001$, $\eta_p^2 = .87$). This is in line with the previous analysis, which revealed that both the TD and WFD groups showed an overall better performance at WPVT (comprehension) than PN task (word production). The rescaled age (Age_mya = Age-75) according to the youngest age measured for the disorder group was added as the covariate.

We must look at results of the 3-way interaction between Task x Group x Age_mya to answer the question of whether the disorder group followed the same developmental relationship between the two tasks (comprehension and word production) as the one followed by TD children. In this case the answer is yes, as shown by the non-significant interaction ($F(1,72) = 2.76$, $p = .101$, $\eta_p^2 =.037$). There was also no Task x Group interaction ($F(1,72) = 0.32$, $p = .572$, $\eta_p^2 = .004$) indicating that there was no difference in relative level of task performance of the two groups at the first point of overlap of the trajectories.

There was no main effect of group ($F(1,72) = 1.73$, $p = .192$, $\eta_p^2 = .02$). This is reflected in the graph in Figure 5.15, which shows that, at the youngest age that the disorder group was measured, the mean task performance of the two groups was comparable. However, from the same graph it is clear that the trajectories between the TD and WFD groups for both word production and comprehension diverge with age although the developmental gap for the word production trajectory is more obvious. Overall age was still a significant predictor of performance [Age_mya: $F(1,72) = 6.49$, $p = .013$, $\eta_p^2 =.08$] although including only the 52 cases of TD children produced a lower level of significance than was shown for the analysis which included the 100 cases of typically developing children where $p = .003$.

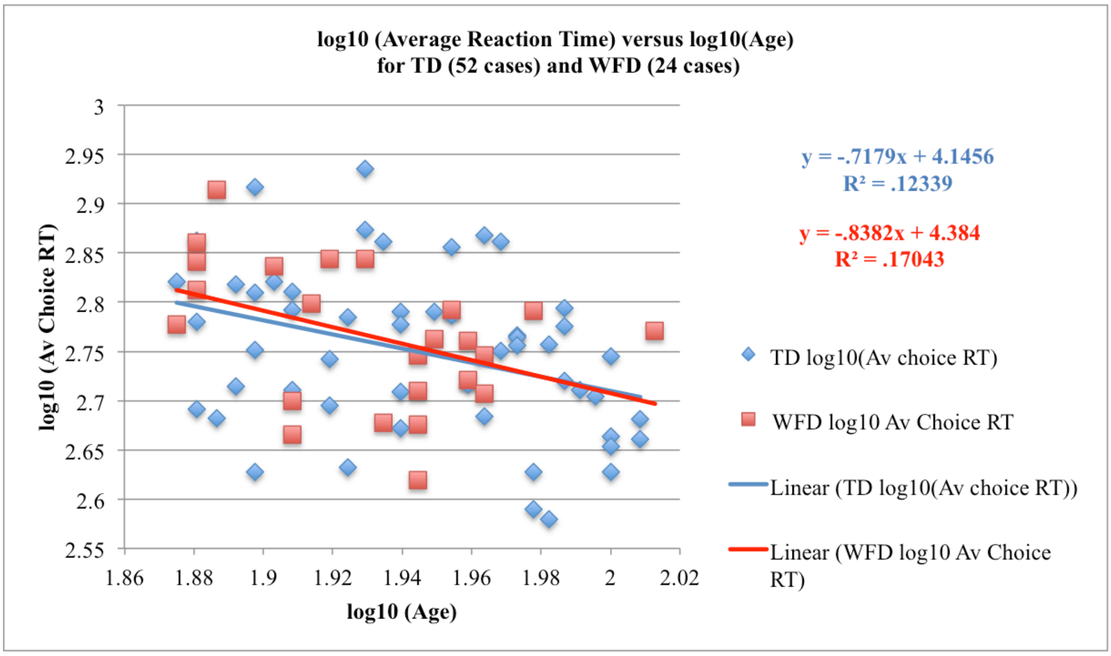
However in contrast to the analysis which included the 100 cases of TD children, the group by age interaction did not show a reliable difference in the overall development of both language abilities between the TD group and the group of children with WFD when taking into account only the 52 cases of TD children whose age range overlapped with the sample of children with WFD [Group x Age_mya: $F(1,72) = 2.67$, $p = .107$, $\eta_p^2 = .04$].

Figure 5.15 Graph showing the comparison of developmental trajectories of word production (PN task) and comprehension (WPVT) between TD children (52 cases) and children with WFD (24 cases).



Comparison of Reaction Time Trajectories between TD Children and Children with WFD [Group x Age]. To linearise the relationship between reaction time and age, both variables were log transformed. Tests of between subjects effects for reaction time between the TD and WFD groups then produced an overall R^2 value of .14 showing that the model explained 14% of the variance ($F(3,72) = 4.05$, $p = .010$, $\eta_p^2 = .14$). Results revealed that there was no overall effect of group ($F(1,72) = 0.08$, $p = .775$, $\eta_p^2 = .001$) indicating that the difference between the intercepts of the two groups was not reliably different (see Figure 5.16). Age predicted level of performance for the combined groups ($F(1,72) = 8.99$, $p = .004$, $\eta_p^2 = .11$) but there was no Group x Age interaction, which is also shown in Figure 5.16 where the gradients of the trajectories are indistinguishable and the intercepts are very close together ($F(1,72) = 0.03$, $p = .874$, $\eta_p^2 = .000$). Thus, where group disparities were observed in the language and semantic skills, these did not appear to originate in perceptual or motor differences as assessed by a simple response time task, but instead lay in higher-level processes.

Figure 5.16. Comparison of average reaction time between TD children and children with WFD.



Discussion

As expected, positive correlations were obtained between the scores of all four core tasks for the sample of TD children included in this study, since they were all expected to show a reliable improvement with age. This notion was also demonstrated in the graph depicting the developmental trajectories of the four core tasks for the TD children (see Figure 5.2). By contrast, no significant correlations were obtained between any of scores of the four core tasks for children with WFD. Thus, the correlation matrix for the scores of four core tasks for the WFD group did not provide any hints for follow up on potential causes of difficulties (i.e., phonological versus semantic causes).

As in previous literature, results of the trajectories analyses for TD children show comprehension to be developing ahead of word production (Reznick & Goldfield, 1992; Plunkett, Sinha, Møller, & Strandsby, 1992). For our sample data the rate of development of comprehension for TD children was 40% faster rate than that of production. Further analysis of the data did not show a reliable difference between the scores of TD children and children with WFD in all four tasks at the youngest age measured for children with WFD. This outcome suggests that it is more difficult to identify the potential of developing WFD at an early age. TD children showed a reliable improvement at all four tasks as they became older whereas children with WFD did not show a reliable improvement with age on any of the four core tasks (see Figure 5.4). However, because these are cross-sectional data, we cannot distinguish flat developmental trajectories from heterogeneous levels of deficit in the WFD group which are randomly associated with age during sampling. Only longitudinal follow-up could definitively resolve this question.

Additional analyses of the trajectories for PN and WPV tasks showed that TD children were better at comprehension than at word production and that they improved at both tasks with age. There is a caveat to take note of here that task difficulties were not calibrated, so the main effects of task were harder to interpret. However, the more definitive results showed that comprehension of TD children improved at a faster rate than their word production, as they became older. Analyses of the trajectories for PN and WPV tasks for children with WFD also showed that children were better at comprehension than at word production. This notion seems to be consistent with the literature on identification of children with WFD who are expected to have difficulty in word production that is greater than would be expected given the child's ability to comprehend words (Messer & Dockrell, 2006; German, 1986, 2000); but the same caveat applies again, that the task difficulties were not calibrated. Nevertheless, further analyses revealed that the children with WFD did not show a reliable difference in the rate of change between the two tasks as they became older. This indicates that this sample of children with WFD did not represent a pure picture of WFD. If the group of children with WFD had a 'pure' production problem, we would expect only a disparity (either main effect of Group, Group x Age interaction, or both) for word production (PN task), but not for semantics (PJ task), phonology (CNRep), or comprehension (WPVT). The results do not show that, but the caveats are: (a) there may be ceiling effects (as in the case of the PJ task), and (b) in a cross-sectional trajectory, severity will be randomly associated to age and therefore may attenuate ability-age relationships. For example, the sharp decline in scores of the CNRep task for children with WFD as age increased could be stemming from a sampling bias due to the fact that the data are cross-sectional, at the same time as indicating at least some of the (older) children exhibited evidence of a phonological deficit.

The notion that the sample of children with WFD in this study may not have reflected a true representation of WFD was also highlighted in the results of the 52 TD cases group of mixed design analysis, group by age interaction. Although the second set of comparison analyses that included only the 52 TD children whose age range overlapped with the children with WFD yielded the same qualitative pattern of most of the results including nearly all significant and non-significant group main effects and interactions as the previous analysis which included all the 100 TD children, there was one exception. In contrast to the analysis which included the 100 cases of TD children, the group by age interaction did not show a reliable difference in the overall development of both language abilities between the TD group and the group of children with WFD when taking into account only the 52 cases of TD children whose age range overlapped with the sample of children with WFD. The difference between the results of the analyses of the100 and 52 TD case groups could be due to the much smaller age range included in the analysis of the 52 TD cases (age range was 2 years and 3 months) compared to the age range covered in the analysis of the100 cases of TD children (age range was 4 years and 6 months). It could also be as mentioned earlier that the sample of children with WFD included in the study did not reflect a true representation of WFD.

There was no difference in reaction time between the TD children and children with WFD included in this study. Both groups showed that their reaction time improved with age. This implies that the generally slower response in naming from children with WFD (Dockrell, Messer, George, 2001) is unlikely to be stemming from difficulties in motor aspects of naming.

Results of the analyses of CNRep task scores for children with WFD were particularly surprising as they did not show a reliable correlation between any of the tasks for children with WFD. Had there been a reliable correlation between CNRep task and PN task for children with WFD, this would point to mainly phonological difficulties driving performance in production. On the other hand, a correlation between PN scores and, for example, PJ task scores, for children with WFD would indicate that difficulties could be heterogeneous in nature.

If the flat or even negative trajectory of word production in children with WFD were real, a possible explanations is that as they become older, there is a decline in confidence as children become aware of their difficulties. The most likely explanation is that the root of the difficulties is not homogeneous (Dockrell, Messer & George, 2001). Given that word-finding difficulty is argued to be a consequence of underlying cognitive causes that are heterogeneous in nature (Dockrell, Messer & George, 2001), the cross-sectional empirical data explored in this study offer limited opportunity for drawing inferences on the trajectories of language abilities for children with WFD. However, results of this study confirm that children with WFD showed word production deficits worse than comprehension, but that (in this cross-sectional analysis), this emerges over development.

Taken altogether our findings are consistent with two of our predictions. The first prediction was that the group of children with WFD would show lower word production ability than the TD group but would show similar ability at comprehension. This has been demonstrated to be the case but that it emerges over development. Comparisons of the trajectories of the PN task and WPVT for TD and WFD groups revealed that, for our samples, there was a significant difference in the rate of development of word production between the two groups but there was no reliable difference in the rate of development of comprehension between the two groups. This provides evidence supporting our third prediction: that the WFD group would show a slower developing PN ability compared to WPVT when compared to the TD group. Contrary to our second prediction, our sample of children with WFD did not show a bigger gap between the development of PN and WPVT abilities over time. Possible reasons for this discrepancy might be that the sample of 24 children is too small, cross-sectional data can lead to a sampling bias or it could be that our sample of children with WFD did not represent a pure picture of WFD.

Hence, so far, results in this study may be taken to present further evidence of the heterogeneous nature of the underlying causes of WFD. Ideally a longitudinal study should follow a cross-sectional one to investigate the effects of any developmental disorder (Thomas et al., 2009), but as this study was based on secondary data analysis, this was not an option in the current context. We now turn to look at another set of secondary data analyses which comprise regression analyses of intervention results from each of phonological and semantic interventions given to 20 of the children with WFD whose developmental trajectories of language skills were included in this chapter. The regression analyses were carried out in order to investigate whether the most effective intervention for each child can be predicted from their profile of core skills.

Chapter 6

Analysis of Results of Phonological and Semantic Interventions for WFD

Introduction

The main purpose of the analysis in this chapter was to explore whether the most effective intervention for each child with WFD can be predicted from their profile of core language skills. The previous chapter presented a comparison of cross-sectional developmental trajectories of typically developing and atypically developing language skills for children in each of four core language skills namely: word production, word comprehension, semantics and phonology. The purpose of constructing the trajectories was to establish the target empirical effects to be captured by computational simulations employing the modified Best et al. (2015) model (Best et al., 2021); and to then simulate various theory-based interventions to investigate their effectiveness in different cases of WFD. Empirical data used for the analysis in Chapter 5 were taken with permission from the WORD project, a large study designed to investigate causes and interventions for children with WFD (Best et al., 2017; Best et al., 2021; Best et al., 2013). This chapter presents a second set of secondary empirical data analyses comprising in depth regression analyses of intervention results from each of phonological and semantic interventions given to 20 children with WFD taken with permission from the same Best et al. (2021) study.

The Best et al. (2021) study used either phonologically or semantically based word-web interventions to examine the different outcomes of phonological and semantic intervention for both a group comparison and case series analyses. More details on the background of the study were given in Chapter 2. To recap, a further objective of the Best et al. (2021) study was to investigate whether one type of intervention (phonological versus semantic intervention) could be predicted to be more effective than the other according to the language profile presented by the child. The language profiles of children taking part in the study were therefore categorised into three different sub-groups: (i) *Classic* WFD (predicted to benefit from either or both semantic and phonological intervention), (ii) *semantic difficulties* (predicted to benefit only from semantic intervention), and (iii) *phonological difficulties* (in the context of strong semantic processing; predicted to benefit only from phonological intervention and not from semantic intervention). Results of the Best et al. (2021) showed that both semantic and phonological interventions were effective but semantic intervention was more effective than phonological intervention; that children responded differently to the same intervention; and that the division into sub-groups according to difficulties can lead to predictions of outcomes depending on which therapy is applied. Results furthermore showed that both phonological and semantic interventions were found to be effective for treated items but neither one generalised to untreated words.

The added value of the secondary data analyses in this chapter is in addressing one of the major limitations of the Best et al. (2021) study, which is, that the size of intervention effects and the differential effect between phonological and semantic therapy could not be examined based on profile because the difficulties were categorised. Regression analyses used in this study could potentially derive models with predictor variables that predict size of intervention effects according to the language profile of each child. A further limitation within the design of the Best et al. (2021) study was that there was no link to the developmental trajectories of these skills in a form that could be simulated by developmental computational modelling. This has been addressed in chapter 5 where developmental trajectories analyses were constructed for four core language skills. The aim was to simulate similar trajectories first for typical language development then atypical language development; and to then simulate various theory-based interventions to investigate their effectiveness in different cases of WFD. The trajectory analyses served to characterise the developmental characteristics of the language skills of the group of children with WFD compared to those with typical language development. These trajectories also served to explain some of the unexpected outcomes of regression analyses in this study.

A within-participant design was used to compare interventions, all 20 children were given both a semantic and a phonological intervention for the same amount of time with a washout period in between the two types of intervention; half the children received phonological intervention followed by semantic intervention; and the other half received semantic intervention followed by phonological intervention. A more detailed discussion of the research design will be presented in a later section.

The following analyses were carried out in pursuit of answering the question of whether the most effective intervention for each child with WFD can be predicted from their profile of core language skills, plus an extended list of questions such as: (i) Can we predict whether intervention per se works? (ii) Can we predict whether the semantic intervention works? (iii) Can we predict whether the phonological intervention works? (iv) Assuming that there is an intervention effect, does it generalise to words not included in the therapy? (v) Does the order of therapy influence the effectiveness of therapy? (i.e., phonological therapy followed by semantic therapy, versus semantic therapy followed by phonological therapy). But the key question targeted by using regression analyses is: (vi) Can we predict which intervention is more effective based on the profile of core skills of the child? The goal was to establish the extent to which the most effective intervention (phonological versus semantic) for each child can be predicted from their profile of core language skills.

Analyses in this chapter comprise t-tests, a one-way ANOVA and regression analyses of intervention results from each of phonological and semantic interventions given to 20 children with WFD that were taken with permission from the Best et al. (2021) study. The advantage provided by the regression analyses in our study is in treating intervention effects as continuous variables, endeavouring to like these to cognitive profiles, thus linking the developmental trajectories of these skills, in a form that can be simulated by developmental computational modelling. Furthermore the continuous approach of the secondary analyses in this study offers the opportunity to examine the size of intervention effects and the differential effect between phonological and semantic therapy based on profile. The strength of the study is in evaluating the extent to which language profiles can predict response to interventions in children with WFD and in providing a basis for developmental computational models to run simulations to investigate effects of intervention. The latter is an important aspect as computational methods have been identified as potentially useful tools in helping to narrow the putative gap between theories of underlying causes of language deficits and intervention practices (Law et al., 2007; Thomas et al., 2019).

Method

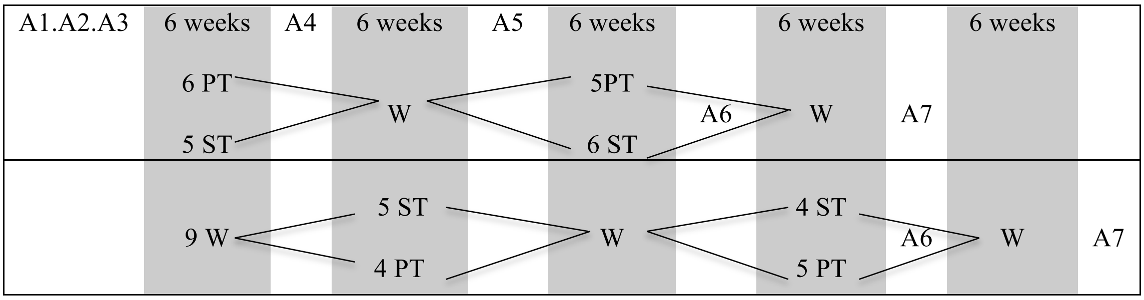
A group of 20 children with WFD, with age range from 6;3 to 8;7 selected from 10 mainstream primary schools, were first tested on the four core language skills: word production, word comprehension, semantics and phonology. The tasks, along with cross-sectional developmental trajectories, were analysed in the previous chapter. A detailed description of the tasks is provided in Chapter 5 and a more detailed description of the procedures for taking the scores of the four core tasks can be found in Appendix 1 in Best et al. (2015). For the purposes of this part of the research, the scores of the core tasks provided a profile of language abilities for each child, which was to be used in regression analyses to explore the effect of different interventions (phonological versus semantic) on the different language profiles. The interventions were already described in chapter 2 and more detail is provided below. Details of the participants included in the study were already given in Chapter 5 but are provided again here.

Participants. Children included in the study either had English as the main language at home or had been in an English-speaking nursery since the age of 3 years and had at least one adult who spoke English at home. The main criteria for inclusion of children with WFD in the intervention study were: (a) they showed a discrepancy between comprehension and production as measured by the Test of Word-Finding Second Edition (TWF-2, German, 2000); (b) their score had to be above the 7^{th} percentile for BAS Pattern Construction Task; (c) they did not have any other significant developmental diagnosis such as Autism, dyspraxia, ADHD, sensory or behavioural difficulties. In addition to test results, children who did not show word-finding difficulties in discourse when screened by clinicians involved in the project were excluded from the study. The children's scores from five language tests were also recorded: the Clinical Evaluation of Language Fundamentals (CELF-4) Core Language standard score; the British Ability Scale (BAS) Non-Verbal t-score; the Test of Word-Finding TWF-2 naming standard score; TWF-2 comprehension raw score; and the Test for Reception of Grammar TROG standard score. Since the scores from the five language tests were already available, they were included in the current work as potential predictor variables to explore potential links between the most effective intervention and a child's language profile. This was another aspect that was different from the Best et al. (2021) study. Including the scores of the language tests as predictors meant that a lot more information was contributed towards identifying the language profile of a particular child. In the Best et al. (2021) study the children were categorised into three subgroups only according to their performance on a semantic (picture judgement) and a phonological task (CNRep).

Design. The effectiveness of the intervention was measured by the children's ability to name 100 items from pictures chosen specifically for the experiment from Funnell, Hughes, and Woodcock (2006) and Druks and Masterson (2000). The 100 words were divided into four sets of 25 words. Twenty-five words were only used for phonological therapy (PT); 25 words were only used in semantic therapy (ST); 25 words were named at the beginning of every session to control for the effect of repetition, referred to as naming control (NC) set; and 25 words were not previously seen and a different set of 25 words was presented on each assessment of the 100-word set, referred to as unseen (US) set. Details of the assessment procedure can be found in Best et al. (2017). The sets of 25 words were matched for pre-intervention naming accuracy, spoken frequency, imageability, visual complexity, number of phonological neighbours and word-length in phonemes. Each child was assessed three times on the whole set of 100 words before the start of therapy and the average score was taken as the baseline score to be compared with at the end of therapy. The children were assessed at the end of each of six-week block of therapy and at the end of the six weeks of the wash out period between therapies. They were then subsequently tested again at the end of a follow-up period six weeks after the end of the second phase of therapy to check whether the effect of therapy had been maintained six weeks after therapy had ended.

The experiment was designed as a randomised control trial with a washout period in between the two phases of intervention. Therapy took place for 30 minutes once a week for six weeks. The first three assessments were referred to as A1, A2 and A3 in the order that they were taken. After the A3 assessment, 11 of the children undertook therapy for six weeks and the remainder began a six-week wait period. Six of the 11 children had phonological therapy and five had semantic therapy. All 20 children (those who had therapy as well as those who were in a wait period) were assessed at the end of the six weeks with assessment A4. There was then a wash out period of 6 weeks for those who had therapy, while therapy commenced for the children who were in the wait period (five of the 9 children had semantic therapy and four had phonological therapy), at the end of which assessment A5 was taken. This was followed by a washout period of six weeks for the subgroup of nine children who had just finished therapy; and each child in the subgroup of 11 children (who had previously been in the washout period) undertook the other intervention (i.e. phonological if he had already completed the semantic intervention and vice versa). Assessment A6 was given to the 11 children at the end of their six weeks of their second therapy. In the next six weeks each child in the subgroup of 9 children were given the intervention that they had not previously taken; and the subgroup of 11 children went on a washout period (or follow up period) of six weeks. This was a repeated crossover design after which Assessment A6 was given to the subgroup of 9 children who had just finished therapy. By the time assessment A6 was taken, all 20 children had been through both semantic and phonological therapy (for both groups assessment A6 refers to scores taken just as their second therapy ended). For each group six weeks after assessment A6, the children were assessed again with assessment A7 to check whether the effects of intervention have been maintained six weeks after their last therapy ended. Therefore the children's scores of assessment A6 were referred to as 'end of therapy scores' and their scores of assessment A7 were referred to as 'scores at maintenance'. Figure 6.1 below shows the number of children in each phase of therapy and wait period and clarifies the crossover design.

Figure 6.1. Study design showing the number of children in each phase of therapy and wait periods.



A = Assessment
PT = Phonological Therapy
ST = Semantic Therapy
W = Wait period

Procedure. Both phonological and semantic interventions used word webs as the therapy technique and were given on a 1:1 basis. An example of each word-web used is given in Figure 2.1 in Chapter 2. Sessions took place once a week and lasted 45 minutes. There were approximately 10 minutes of assessment first, five minutes of activity and then 30 minutes of intervention. In each case as can be seen from the word-webs in Figure 2.1, a picture of the target word was shown in the middle of the web. In both the semantic and phonological therapy, the child was asked for features related to the target word in the order that they appear on the hexagons around the target picture, the goal in each case being to enrich the relevant mental representation. In the case of the semantic therapy, the features were: category, appearance, location, use, action, and linked words. In each case, only the features that were appropriate to the target word were used (e.g., an animal may not have a 'use'). In the case of phonological therapy the child was also asked for the features in the set order that they appeared in the hexagons around the target word (see Figure 2.1b). The child was asked for first sound, things that start with the same sound, number of syllables, things that rhyme, first letter and ways to break down the word. In both therapies, a choice was provided if the child was unable to name a feature. After all the features had been named, they were reviewed and the child was encouraged to say and repeat the word. A more detailed description of the intervention therapies can be found in Best et al. (2021).

Analysis Strategy. The present study investigated the effects of the completed intervention on the word sets (as opposed to the effects at different stages in between therapies) by analysing the mean differences between the average scores of the three pre-therapy assessments (A1, A2 and A3) and assessment A7 taken at maintenance (this was taken 6 weeks after the last therapy session). Five paired-samples t-tests were conducted to explore the effects of intervention for WFD. In each case the t-test compared the means of the average scores of the children with WFD before any therapy was started with the means of their scores at maintenance (A7) as the most robust indicator of the effectiveness of the intervention. Paired samples t-tests were chosen because all the children went through both types of therapy (semantic and phonological) but each type of therapy focused on a different set of 25 words. The first paired samples t-test was conducted to test whether there was any effect of intervention at maintenance by comparing the means of the total set of 100 words from pre-therapy to maintenance. Subsequently four t-tests were conducted on each of the four 25-word sets separately. The t-test for the ST set was to test whether there was an effect of semantic therapy and the t-test on the PT set was to test whether there was an effect of phonological therapy. The t-test for the NC set was to test whether there was an improvement of scores due to repetition of names only without an elaborate therapy session. The t-test for the US set evaluated whether there was transfer from words used during the interventions to produce an improvement for words that have not been practiced previously.

A one-way repeated measures ANOVA was then run to determine whether the improvements due to phonological therapy were significantly better than improvements due to semantic therapy. At the same time a check was made on whether there was a significant difference in improvement of scores across the four different sets of words. The hypothesis was that there would be a significant difference between the sets of words used in therapy (both ST and PT) and the other two sets.

Pearson's Correlations. Pearson's correlations were first utilised to assess which predictors would be helpful in constructing the regression models to predict intervention effect sizes based on the child's language profile of the four core tasks and the five language tests taken pre-therapy. Based on past research, the most prominent correlations were expected to be as follows: CNRep task scores would correlate negatively with scores of PT, showing that children with worse phonological skills would benefit from phonological intervention; and PJ task scores would correlate negatively with ST showing that children with worse semantic skills would benefit from semantic intervention (Best et al., 2021).

Regression Analyses. Exploratory multiple linear regression analyses were run to reveal reliable predictors and effect sizes of semantic and phonological interventions combined and separately. Multiple linear regression analyses were run for three different outcomes, namely: (i) the difference in total set (all 100 words) between maintenance (at assessment A7) and pre-therapy; (ii) the difference in ST set (25 words) between maintenance (at assessment A7) and pre-therapy; (iii) the difference in PT set (25 words) between maintenance (at Assessment A7) and pre-therapy. The regression analyses were first run while exploring the inclusion of scores on: age, gender, order of therapy and each of the four core tasks. Since the models produced with data from these predictors were not statistically significant, we then explored the inclusion of data from the five language tests that the children with WFD had taken before starting therapy, namely: the Clinical Evaluation of Language Fundamentals (CELF-4) Core Language standard score, the British Ability Scale (BAS) Non-Verbal t-score, the Test of Word-Finding (TWF-2) naming standard score, the TWF-2 comprehension raw score and the Test for Reception of Grammar (TROG) standard score as predictors.

The first model was run with the outcome as the difference in total set between maintenance and pre-therapy and was expected to include all four core language skills as significant predictors. This is because each of the core language skills was specifically chosen to assess a certain aspect of the putative mechanisms involved in the process of word production as mentioned in Chapter 5. The picture naming (PN) task was to test word production [semantic input to phonological output (SP) component of simulation]; Word-Picture Verification Task (WPVT) was to test comprehension [phonological input to semantic output (PS) component of simulation]; Picture Judgement (PJ) task was to test semantic knowledge of children [semantic input to semantic output (SS) component of simulation]; and the Children's Non-word Repetition Task (CNRep), was to test phonology [phonological input to phonological output (PP) component of simulation].

The second model was run with the outcome as the difference in ST set (25 words) between maintenance and pre-therapy and was expected to include PJ accuracy score as the most reliable predictor since the PJ task was assumed to be testing associative semantics of children. Hence poor accuracy scores on the PJ task were expected to predict bigger intervention effects on the basis that intervention should be targeting the weak skill of the child.

The third model on the other hand was run with the outcome as the difference in PT set (25 words) between maintenance and pre-therapy and was therefore expected to include CNRep raw score as a significant predictor since this task tested the phonological ability of children.

In theory, we could either (i) expect children with weaker semantic skills to benefit more from semantic therapy and those with weaker phonological skills to benefit more from phonological therapy on the basis of targeting intervention on the weak skills of the child; or (ii) expect children with weaker semantic skills to benefit more from phonological therapy and children with weaker phonological skills to benefit more from semantic therapy on the basis that children would show better improvements on WFD if intervention was targeted on the child's strengths.

All the regression models that were identified and included in this study were derived using the enter method after having tried both the forward and backward stepwise methods. In all cases the models were deemed as best fit because within the trials the selected models explained the largest variance with the least amount of predictors.

Results

T-tests and ANOVA. The first paired samples t-test conducted to test whether there was any effect of intervention at maintenance showed a significant difference in scores between the average score of the total set of 100 words taken pre-therapy ($M = 53.05$, $SD = 7.24$) and the score of the total set of 100 words taken at maintenance ($M = 65.75$, $SD = 10.04$) where $t(19) = -7.95$, $p < .001$.

The paired samples t-test conducted to test whether there was a difference between the average score of the ST set taken pre-therapy ($M = 13.26$, $SD = 1.86$) and the score of the ST set taken at maintenance ($M = 17.30$, $SD = 3.11$) showed a significant difference in scores, where $t(19) = -6.76$, $p < .001$. Likewise paired samples t-test conducted to test whether there was a difference between the average score of the PT set taken pre-therapy ($M = 13.28$, $SD = 1.86$) and the score of the PT set taken at maintenance ($M = 17.85$, $SD = 3.78$) showed a significant difference in scores, where $t(19) = -7.22$, $p < .001$. Again the paired samples t-test conducted to test whether there was a difference between the average score of the NC set taken pre-therapy (M = 13.17, SD = 1.82) and the score of the NC set taken at maintenance (M = 15.85, SD = 2.66) showed a significant difference in scores, where $t(19) = -5.65$, $p < .001$. Finally, the paired samples t-test conducted to test whether there was a difference between the average score of the US set taken pre-therapy (M = 13.36, SD = 1.84) and the score of the US set taken at maintenance (M = 14.75, SD = 2.51) showed a significant difference in scores, where $t(19) = -2.84$, $p = .011$.

A one way within subjects ANOVA indicated that there was a significant difference between at least one set of improvements in scores and another, [$F(3,57) = 9.79$, $p < .001$, $\eta_p^2 = .34$]. Table 6.1 shows the results of pairwise comparisons of the difference in scores between the assessment taken at maintenance and the average pre-therapy assessments for each of the four word sets. These results show that improvements in scores due to phonological intervention did not produce significantly better results than improvements due to semantic therapy at maintenance. Since multiple comparisons were being made the Bonferroni method was used to adjust the significance level for the individual tests (Howitt & Cramer, 1997). Four groups of mean results were compared leading to six comparisons to be made. The significance level for each test was then adjusted by dividing the overall significance level (.05) by the number of comparisons (6), .05/6 = .008. Hence the critical value for *p* was taken at the .008 level for a comparison to be accepted as significant. In this case only two comparisons have been shown to be significant: Improvement in scores of the PT and ST sets were both significantly better than the improvement in scores of the US set.

Table 6.1. Pairwise comparisons of the difference in scores between the assessment taken at maintenance and the average score of pre-therapy assessments for each of the word sets.

Difference in Set (i) between pretherapy and maintenance	Difference in Set (j) between pretherapy and maintenance	Mean Difference (i-j)	Std. Error	Sig.	95% Confidence Interval for Difference in mean Lower Bound	Upper Bound
ST	PT	-.53	.73	.478	-2.06	1.001
	NC	1.36	.70	.068	-.11	2.83
	US	2.59*	.65	.001	1.22	3.95
PT	ST	.53	.73	.478	-1.00	2.06
	NC	1.89	.67	.01	.48	3.30
	US	3.12*	.52	.000	2.02	4.21
NC	ST	-1.36	.70	.068	-2.83	.11
	PT	-1.89	.67	.011	-3.30	-.48
	US	1.23	.48	.020	.22	2.23
US	ST	-2.59*	.65	.001	-3.95	-1.22
	PT	-3.12*	.52	.000	-4.21	-2.02
	NC	-1.23	.48	.020	-2.23	-.22

Pearson's Correlations. Table 6.2 shows Pearson's correlations between three key results of the therapy which are: Difference in scores of total set of 100 words between maintenance and pre-therapy, difference in scores of ST set between maintenance and pre-therapy, difference in scores of PT set between maintenance and pre-therapy; the four core skills and the five language tests. Statistical significance was accepted at the $p < .05$ level (2-tailed). In this case, the only statistically significant correlation for improvement of assessment results after therapy was found to be between 'difference in scores of PT set between maintenance and pre-therapy' and 'TWF-2 comprehension Raw score': $r(20)=.518$, $p=.019$. This shows that we could expect TWF-2 comprehension raw score to be included in the model predicting the effects of phonological therapy. The prediction was thereby, the greater the level of single word comprehension of a child, the greater the benefit of the phonological intervention.

A significant correlation was found between PN accuracy score and TWF-2 naming standard score: $r(20) = .481$, $p = .032$ which supports the view that these were consistent measures of naming ability. Another significant correlation between a core skill and a language test was found between CNRep raw score and TROG standard score: $r(20) = .494$, $p = .027$, indicating a relationship between phonological working memory and receptive grammar. It was noted that surprisingly there was no significant correlation between WPVT and TWF-2 comprehension raw score, even though both tests were supposed to measure comprehension.

Table 6.2. Pearson's Correlations matrix for the Difference in total score, the four core tasks and the five language tests: Picture naming task (PN) accuracy score, word-picture verification task (WPVT) accuracy score, picture judgement task (PJ) accuracy score, children's non-word repetition task (CNRep) Raw score, the Clinical Evaluation of Language Fundamentals (CELF-4) Core Language standard score, the British Ability Scale (BAS) Non-Verbal t-score, the Test of Word-Finding (TWF-2) naming standard score, the TWF-2 comprehension raw score and the Test for Reception of Grammar (TROG) standard score.

		Diff in total set at Maint	Diff in ST Set at Maint	Diff in PT set at Maint	PN Acc. score	WP VT Acc score	PJ Acc. score	CNRep Raw score	CELF-4 Core stand. score	BAS Non-V t- score	TWF-2 naming stand. score	TWF-2 Comp Raw score	TROG stand. score
Diff in total set at maint	Pearson Correl.		.635**	.791**	-.204	-.202	.192	.300	.042	-.075	-.069	.235	-.008
	Sig. (2 tailed)		.003	<.001	.388	.394	.417	.198	.861	.753	.772	.319	.973
Diff in ST set at Maint	Pearson Corr.	.635**		.295	-.314	-.092	-.022	.423	-.014	-.057	-.111	-.119	.290
	Sig. (2 tailed)	.003		.206	.178	.701	.928	.063	.954	.810	.641	.617	.216
Diff in PT set at Maint.	Pearson Correl.	.791**	.295		.116	-.050	.299	.189	.306	-.120	.225	.518*	-.023
	Sig. (2 tailed)	<.001	.206		.626	.834	.200	.425	.189	.613	.339	.019	.925
PN Acc. score	Pearson Correl.	-.204	-.314	.116		.319	.296	.357	.290	.090	.481*	.115	.057
	Sig. (2 tailed)	.388	.178	.626		.170	.204	.122	.214	.707	.032	.630	.812
WPVT Acc. score	Pearson Correl.	-.202	-.092	-.050	.319		.052	.025	-.116	.033	.316	.210	-.113
	Sig. (2 tailed)	.394	.701	.834	.170		.827	.915	.627	.891	.175	.373	.635
PJ Acc. score	Pearson Correl.	.192	-.022	.299	.296	.052		.061	.385	.377	.400	.444	.190
	Sig. (2 tailed)	.417	.928	.200	.204	.827		.799	.093	.101	.081	.050	.423
CNRep Raw score	Pearson Correl.	.300	.423	.189	.357	.025	.061		.361	-.046	.132	-.029	.494*
	Sig. (2 tailed)	.198	.063	.425	.122	.915	.799		.118	.847	.578	.902	.027
CELF-4 Core stand. score	Pearson Correl.	.042	-.014	.306	.290	-.116	.385	.361		.308	.481*	.344	.550*
	Sig. (2 tailed)	.861	.954	.189	.214	.627	.093	.118		.187	.032	.137	.012
BAS Non-V t-score	Pearson Correl.	-.075	-.057	-.120	.090	.033	.377	-.046	.308		.214	.100	-.049
	Sig. (2 tailed)	.753	.810	.613	.707	.891	.101	.847	.187		.365	.674	.836
TWF-2 nam. stand. score	Pearson Correl.	-.069	-.111	.225	.481*	.316	.400	.132	.481*	.214		.522*	.247
	Sig. (2 tailed)	.772	.641	.339	.032	.175	.081	.578	.032	.365		.018	.293
TWF-2 comp Raw score	Pearson Correl.	.235	-.119	.518*	.115	.210	.444	-.029	.344	.100	.522*		.177
	Sig. (2 tailed)	.319	.617	.019	.630	.373	.050	.902	.137	.674	.018		.454
TROG stand. score	Pearson Correl.	-.008	.290	-.023	.057	-.113	.190	.494*	.550*	-.049	.247	.177	
	Sig. (2 tailed)	.973	.216	.925	.812	.635	.423	.027	.012	.836	.293	.454	

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).
All correlations are for comparisons of the mean scores of 20 individuals.
Diff in total set at Maintenance = Difference in scores of total set of 100 words between maintenance and pre-therapy.
Diff in ST set at Maintenance = Difference in scores of semantic therapy set between maintenance and pre-therapy.
Diff in PT set at Maintenance = Difference in scores of phonological set between maintenance and pretherapy.
PN Acc score = Picture naming accuracy score.
WPVT = word-picture verification task accuracy score.
PJ Acc. score = picture judgment accuracy score.

CELF- 4 Core = Core language standard score.
BAS Non-V t-score = Non-verbal t-score.
TWF-2 nam. stand. score = TWF-2 naming standard score.
TWF-2 comp Raw score = TWF-2 comprehension raw score.
TROG stand. score = TROG standard score.

Regression Analyses. Exploratory regression analyses were initially run to investigate whether the effect of both semantic and phonological therapies on the total set of 100 words can be predicted by the scores of any of the four core tasks, which are PN, WPVT, PJ and CNRep, order of therapy (semantic versus phonological), age or gender. The total set of 100 words included the 25 words treated with ST, 25 words treated with PT, 25 words of the NC set and the 25 words of the US set.

None of the obtained models, which included various combinations of the predictors, were significant. Even when we explored the inclusion of the scores from the five language tests as predictors we could not obtain a statistically reliable model for predicting effects of both interventions with the available data.

Reducing the number of predictors was therefore considered the way forward. Age was not expected to be a good predictor of the effects of therapy for the sample of children with WFD included in this study since the trajectories analysis described in chapter 5, for the same group of children, showed that there was no significant improvement of performance with age in any of the developmental trajectories of the four core tasks. Hence age was left out from the subsequent regression analysis. Although gender was included in the preliminary exploratory regression analysis in case there were any obvious differences in response to intervention between boys and girls, this was subsequently left out on the basis that the number of cases available was not large enough to be divided according to type of therapy as well as gender.

The regression analysis was then run for model (T) with the difference in total set between maintenance and pre-therapy as the outcome and the following five predictors: The scores on the four core tasks (PN, WPVT, PJ, and CNRep), and order of therapy. These were the five predictors considered to be the most salient to this study and so were the ones initially included in subsequent regression analyses.

Table 6.3. Regression coefficients for the difference in total set between maintenance and pre-therapy as the criterion (model (T)).

Constant & Predictor Variables	Unstandardised Regression Coefficients		Standardised Beta Coefficient	t	p value
	B	Std Error			
Constant	3.48	17.12		0.20	.842
PN Acc score	-0.57	0.34	-.45	-1.71	.110
WPVT Acc score	-0.09	0.27	-.08	-0.32	.753
PJ Acc score	1.04	0.70	.36	1.49	.159
CNRep Raw score	0.73	0.37	.47	1.95	.072
Order of Therapy	2.87	3.25	.21	0.88	.392

$R^2 = .33$, adjusted $R^2 = .09$, $F(5,14) = 1.37$, $p = .295$

Regression analysis which included the above mentioned five predictors where the difference in total set between maintenance and pre-therapy was the criterion still resulted in a non significant model (T) as can be seen from the figures in Table 6.3. The non significant model explained 33% of the variance with an adjusted R squared value of 9%. Non of the predictors produced significant coefficients but the CNRep Raw score had the largest standardised beta coefficient (.47) and a *p* value of .072 which is close to the normally accepted critical value of *p* at .05. This indicates that the CNRep Raw score could potentially be a reliable predictor for the effects of both semantic and phonological intervention.

Table 6.3 also shows that the beta coefficients for the PN and WPVT accuracy scores are both negative whereas the beta coefficients for the PJ accuracy score and CNRep raw score are both positive. This implies that for the present study, interventions showed bigger effects for children who were worse at picture naming and comprehension but better at semantics and phonology.

Model (ST) was obtained when regression analyses were run to investigate how the results of semantic intervention were affected by the different profiles of core language skills and order of therapy. The regression was run with the difference in ST set between maintenance and pre-therapy as the criterion and the inclusion of the following five predictors was explored: The scores on the four core tasks (PN, WPVT, PJ, and CNRep) and order of therapy.

Table 6.4. Regression coefficients for difference in semantic set between maintenance and pre-therapy as the criterion. (model (ST)).

Constant & Predictor Variables	Unstandardised Regression Coefficients		Standardised Beta Coefficient	t	p value
	B	Std Error			
Constant	2.07	5.78		.36	.725
PN Acc score	-.29	.11	-.62	-2.59	.021
WPVT Acc score	.04	.09	.09	.42	.683
PJ Acc score	.16	.24	.15	.69	.503
CNRep Raw score	.38	.13	.65	2.99	.010
Order of Therapy	.60	1.10	.12	.55	.592

R^2 = .45, adjusted R^2 = .26, F(5,14) = 2.33, p = .098

As can be seen from Table 6.4, although model (ST) as a whole was not significant (p = .098), two of the predictors (PN accuracy score and CNRep raw score) showed significant coefficients at (p = .021 and p = .010 respectively).

Hence the regression analysis was run again for model (STa) where only PN accuracy score and CNRep raw score were included as predictors. A significant best-fit model(STa) was obtained, using the enter method, when including as predictors CNRep raw score and PN accuracy score, with both predictors showing significant coefficients. Results for model (STa) are shown in Table 6.5.

Table 6.5. Regression coefficients for difference in semantic set between maintenance and pre-therapy as the criterion. (model (STa)).

Constant & Predictor Variables	Unstandardised Regression Coefficients		Standardised Beta Coefficient	t	p value
	B	Std Error			
Constant	6.00	2.68		2.24	.039
Pic Naming Acc score	-0.25	0.09	-.53	-2.71	.015
CNRep Raw score	0.36	0.12	.61	3.12	.006

$R^2 = .43$, adjusted $R^2 = .36$, $F(2,17) = 6.31$, $p = .009$

The 95% Confidence intervals for the unstandardised B value of CNRep (lower bound at 0.12 and upper bound at 0.60) and for Picture naming accuracy score (lower bound at -0.45 and upper bound at -0.06) do not cross zero but are large relative to the value of B, indicating that the model was not considered as a good model. The Durbin-Watson test value is 3.28 indicating the possibility of a high negative correlation between adjacent residuals thereby casting doubt on the assumption that errors in regression are independent. The value of Cook's distance for case number 12 is 1.45 indicating that it has more influence on the model than the rest of the cases. The standardised residual for this case is 1.86, which also reveals that it is possibly an outlier. The rest of the standardised residual values are within the acceptable range. The scores of case number 12 were checked and it was found that the score for assessment A1 (score of 31 out of 100) was much lower than the average score of assessment A1 of the other 19 cases (average score of 52 out of 100).

Hence after excluding case number 12, the regression was run again with the difference in ST set between maintenance and pre-therapy as the criterion for model (STb). Again, the best fit model included CNRep raw score and PN accuracy score as predictors, however, only CNRep raw score was shown to be a significant variable in this model as can be seen from the results in Table 6.6 below. These results show that the score of case number 12 had indeed influenced the results by a large margin. Including scores of case number 12 resulted in a model that suggested that both PN accuracy scores and CNRep raw scores would be good predictors of the outcome of semantic intervention. Similarly to model (T) which looked at predictors for effect of both semantic and phonological therapies on the total set of 100 words, the implication was that semantic intervention showed bigger effects for children who were worse at picture naming but better at CNRep. The directions of these effects were the same for both models (STa) and (STb) (i.e. whether scores of case 12 were included or excluded). However on excluding the scores of case number 12, the best fit model for predicting the outcome of semantic intervention where only CNRep was a significant predictor, explained a larger percentage of the variance (variance of model including scores of case number 12, R^2=43%; variance of model excluding scores of case number 12, R^2=49%). These results indicate that for the present study, semantic intervention showed bigger effects for children who were better at CNRep.

Table 6.6. Regression coefficients for difference in semantic set between maintenance and pre-therapy as the criterion, (model (STb)) excluding case number 12.

Constant & Predictor Variables	Unstandardised Regression Coefficients		Standardised Beta Coefficient	t	p value
	B	Std Error			
Constant	.63	2.88		.22	.830
PN Acc score	-.09	0.10	-.17	-.90	.381
CNRep Raw score	.37	0.10	.72	3.89	.001

R^2 = .49, adjusted R^2 = .42, F(2,16) = 7.58, p = .005

For model (PT) exploratory regression analyses were run to investigate how the results of phonological intervention were predicted by the different profiles with the difference in PT set between maintenance and pre-therapy as the outcome variable and initially exploring the inclusion of the same five predictors as for model (ST) above.

Table 6.7. Regression coefficients for difference in phonological set between maintenance and pre-therapy as the criterion (model (PT)).

Constant & Predictor Variables	Unstandardised Regression Coefficients		Standardised Beta Coefficient	t	p value
	B	Std Error			
Constant	-4.71	7.25		-.65	.526
PN Acc score	-.04	.14	-.07	-.26	.800
WPVT Acc score	-.01	.11	-.03	-.12	.904
PJ Acc score	.47	.30	.41	1.59	.135
CNRep Raw score	.14	.16	.23	.92	.375
Order of Therapy	1.95	1.38	.35	1.41	.179

$R^2 = .23$, adjusted $R^2 = -.04$, $F(5,14) = .851$, $p = .536$

As can be seen from Table 6.7, model (PT) was not significant and non of the obtained beta coefficients showed near significance p values. Hence taking into consideration that Pearson's Correlation matrix (shown in Table 6.2) revealed a statistically significant correlation between 'difference in scores of PT set between maintenance and pre-therapy' and 'TWF-2 comprehension Raw score': $r(20)=.518$, $p=.019$, another regression analysis was run for model (PT) which only included TWF-2 comprehension raw score as predictor (referred to as model (PTa)).

A significant best-fit model (PTa) was obtained when including TWF-2 comprehension raw score as predictor, showing a significant coefficient. Results for the obtained model (PTa) are shown in Table 6.8. These results indicate that for the present study, phonological intervention showed bigger effects for children who were better at comprehension.

Table 6.8. Regression coefficients for difference in phonological set between maintenance and pre-therapy as the criterion (model (PTa)).

Constant & Predictor Variables	Unstandardised Regression Coefficients		Standardised Beta Coefficient	t	p value
	B	Std Error			
Constant	-38.39	16.75		-2.29	.034
TWF-2 comp Raw score	0.65	0.26	.52	2.57	.019

R^2 = .27, adjusted R^2 = .23, F(1,18) = 6.59, p = .019

Discussion

This study used t-tests, a one way ANOVA, Pearson's correlations and regression analyses of secondary data (taken with permission from the Best et al. (2021) study) to explore the effects of phonological and semantic interventions, given to 20 children with WFD, based on a child's language profile.

Results of the t-tests show that, in general, the semantic intervention and the phonological intervention were both successful and had improved word-finding skills by a similar amount. These results are in line with past research (Best, 2005; Best et al., 2021; Bragard, Schelstraete, Snyers, James, 2012; Easton, Sheach & Easton, 1997). Results of the one-way within subjects ANOVA indicated that improvement in scores of the PT set was not significantly better than improvement of scores of the ST set at maintenance. For the group of children in this study improvement in scores of the PT and ST sets were both significantly better than improvement in scores of the US set indicating that intervention effects did not generalise significantly to words that have not been used in the therapy. These results are again compatible with past research (Best, 2005; Best et al., 2021; Bragard, Schelstraete, Snyers, James, 2012; Easton, Sheach & Easton, 1997; Ebbels, 2014b). It was noted that improvement in scores of the ST and PT sets were not shown to be significantly better than improvement in scores of NC set. This brings into question of whether improvement in scores were caused by repetition of naming rather than the different types of intervention effects.

Pearson's correlations did not show significant correlations between intervention effects and any of the core skills. Based on past literature, significant negative correlations were expected to at least show between phonological therapy and CNRep task scores (assuming the difficulty is stemming from weak phonological skills), and between PJ task scores and semantic therapy (assuming the difficulty is stemming from weak semantic skills). One reason for the absence of a correlation between scores of the PJ task and semantic intervention effects could be because of the ceiling effect of the PJ task score that has been noted in the trajectories analyses in chapter 5. On the other hand the lack of correlation effects between CNRep score and PT is consistent with the results discussed in chapter 5 on trajectory analyses. In chapter 5 we mentioned that there was no significant correlation between CNRep and any of the core tasks, and that had there been a reliable correlation between CNRep task and PN task for children with WFD, this would point to mainly phonological difficulties driving performance in production. Assuming that the core skills included in this study do measure the relevant skills that they have been selected to measure (PN measures production; WPVT measures comprehension; PJ measures semantics; and CNRep measures phonology), then the lack of correlation effects between intervention types and core skills could be explained by the notion that WFD difficulties are caused by a combination of problems in semantic representations, phonological representations and processing speed (see Messer & Dockrell, 2006). However, a statistically significant correlation between improvement in scores of the PT set and the TWF-2 comprehension raw scores was observed. This correlation was also reflected in the regression analyses where TWF-2 raw score was shown to be a significant predictor in a model that predicts the effects of phonological intervention in this study. The correlation between scores of PT set and the comprehension subscale of TWF-2 brings into question whether WPVT is indeed measuring comprehension since no correlation was observed between TWF-2 comprehension raw score and WPVT score. The latter point brings to light the difficulty researchers face in selecting tasks for the purpose of measuring intervention effects on specific skills. The fact that the data being analysed in this study is cross-sectional renders task selection for profiling interventions even more difficult.

The score results for the group of children in this study showed a significant correlation between their phonological skills (CNRep raw score) and their grammatical structure skills (TROG standard score). The link in this relationship may be explained by a strength in working memory which would serve the phonological as well as the receptive grammar aspects.

Regression models were run to investigate predictors of both phonological and semantic interventions, including each intervention separately, long term effects of intervention and generalisation effects. Ideally, in order to be able to answer the question of which intervention works best from a child's profile, one would need a larger number of participants than was available for this study, since, in general, it is recommended to have at least 10 or 15 cases of data for each predictor in a regression model (Field, 2009). The limited participant numbers available here reflects the pragmatic difficulty of recruiting a relatively homogeneous sample with respect to age and symptoms, and carrying out an extended within-participant intervention study to compare multiple intervention types.

The effect of both phonological and semantic therapies, as well as generalisation effects, should in theory be predicted from the model for the regression of scores for the difference in total set between maintenance and pre-therapy (model (T)), since this includes scores for unseen words as well as controls. Although the best fit model was not overall significant, the enter method was used to identify the model that explained the maximum amount of variance with the fewest predictors and this included five predictors out of the pool of 12 available predictors. This is an indication that one would need at least 75 cases to make better predictions (Field, 2009). Nonetheless the most influential variable for this regression model was CNRep raw score showing that individuals with better phonological skills would benefit more from both interventions.

CNRep raw score was also the most influential variable in the model obtained for predicting the effects of semantic intervention alone. The outcomes from the regression analyses investigating the effects of semantic intervention on the ST set between maintenance and pretherapy (model (STa)) show that the effects of semantic intervention could be reliably predicted from scores on two of the core tasks namely: CNRep raw score and PN accuracy score as predictors. When the regression was run again, after excluding a potential outlier (model (STb)), only the CNRep raw score was shown to be a reliable predictor. In both these models the regression coefficients for PN accuracy score were negative indicating that bigger effect of semantic intervention depends on worse PN but better CNRep scores. In terms of strengths and weaknesses this outcome indicates that for individuals who had difficulty in picture naming skills (indicated by a negative coefficient for PN) but stronger phonological skills (a reliably positive coefficient for CNRep), then WFD could be stemming from a weakness in semantic representation as opposed to a weakness in phonological representation. In this case the improvement due to semantic therapy would be the result of intervention targeted at the weakness of the child.

The outcome of regression analyses that investigated the effects of phonological intervention on the PT set between maintenance and pretherapy (model (PTa)), included only the TWF-2 comprehension raw score as a reliable predictor. This outcome is in line with the results of the Pearson correlation matrix showing a significant correlation between scores of TWF-2 comprehension raw score and the difference in scores of PT set between maintenance and pre-therapy. Hence this indicates that individuals with better comprehension could potentially show better improvement effects in the long term due to phonological intervention. In terms of strengths and weaknesses, this outcome could be an indication that children whose comprehension abilities were already strong, benefitted more from intervention targeted at phonological difficulties (a weakness in this context).

On examining the results for all the models, by looking in particular, and where possible, from the perspective of interventions targeted at strengths or weaknesses of a child, the following pattern emerges:

(i) Model (T) predicted that intervention targeting both weaknesses and strengths benefitted those who had stronger phonological skills in the long term.

(ii) Model (STb) predicted that intervention targeting a weakness (semantics) benefitted children who had stronger phonological skills in the long term.

(iii) Model (PTa) predicted that intervention targeting a weakness (phonology) benefitted children who had stronger comprehension skills in the long term.

First, we note from model (T) that intervention targeting both weaknesses and strengths is predicted to benefit those who have stronger phonological skills. Second we note that models (STc) and (PTb) predict that intervention would be effective when it is targeted at the weakness of the child.

On the other hand one could further speculate from model (T) that assuming a child has stronger phonological skills (the most influential variable for this regression model was CNRep raw score) and is given both interventions, then the stronger the phonological skills of a child the stronger is the predicted effect of both interventions.

Order of therapy was included as a predictor in the trials of model construction and in theory this effect could have been picked up. However the low number of participants and the crossover design, which included the washout period, meant that there were not enough participants (less than 10) that started and finished the same therapies in the same order at the same time (see figure 6.1) to do a comparison of the different types of intervention effects.

The low number of participants could in general be the reason that some of the language tests such as CELF-4 and (BAS) Non-Verbal t-score did not show correlations with scores of core language tasks or any predictive powers regarding intervention effects. For example if the majority of children in the group had a CELF-4 score that shows difficulties were due to accessibility of semantics (the receptive type) then scores would be expected to correlate with those of ST; and if the majority of children in the group had a CELF-4 score that shows difficulties were due to phonology (the expressive type) then scores would be expected to correlate with those of PT. The absence of predictive powers could also be due to the heterogeneous nature of the difficulties, which would generate a considerable variety in the children's profiles making it more difficult to derive predictions from a small number of participant data.

Each of the regression models given above has at least one case with a residual value around 2. Given that the study was made on 20 cases this means that 5% of the residuals for each regression model have a value bordering on the unacceptable level of error, showing that none of them have particularly good predictions. This may explain the absence of the expected core skills scores as variables in some of the regression models; for example PJ task score was expected to be included in model (STb) as a significant predictor variable for the difference in ST set score between maintenance and pre-therapy.

However, for the current available dataset, here is the general picture that emerges: (i) Both semantic and phonological intervention for WFD can work on the words that have been used in the therapy session. (ii) Semantic intervention can be effective for children whose WFD is due to semantic difficulty. (iii) Phonological intervention can be effective for children whose WFD is due to phonological difficulty. (iv) Intervention effects in this study did not generalise effectively to words not included in the therapy. Findings (i) to (iv) are in line with past research (Best, 2005; Best et al., 2021; Bragard, Schelstraete, Snyers, James, 2012; Easton, Sheach & Easton, 1997). (v) Effects of order of therapy could not be elucidated in this particular study due to the relatively small number of cases. However if this type of analysis were to be applied on a larger data set there is potential in investigating whether order of therapy has influence on the effectiveness of therapy depending on the root cause of the difficulty, i.e. whether the child's weaker skills lay in phonological or semantic abilities. (vi) The finding that there was no significant difference between the word sets used in both semantic and phonological therapy and the naming controls brings into question whether improvements were due to the therapy or were due to repetition of the words. Hence it could be that the items and the 'intervention' were perfectly confounded. These findings suggest that further research on larger data sets is needed in order to find out whether we could potentially predict which intervention is more effective based on the profile of core language skills of the child.

The findings in this study are hence mostly in line with those from the Best et al. (2021) study. Both studies are in agreement of points (i) to (iv) above: Both semantic and phonological intervention for WFD can work on the words that have been used in the therapy session; and each therapy type works if the weakness that is the underlying cause of the WFD is the target of the therapy. However, Best and colleagues (2021) suggested that semantic intervention was more effective than phonological intervention at the group level. By contrast in our study we did not find that semantic intervention was more effective than phonological intervention in the long term effects. In their discussion the authors contemplated recommending the use of semantic word-webs, rather than phonological word-webs as a general rule, on the basis of their group level results. However in the clinical practice implications section they suggested that to optimise outcomes of intervention results therapists should take account of their findings for case series and sub-grouping profiles. On that basis, ideally the recommendation is that therapists should assess each child's profile in detail. Since this approach is often not feasible they suggested that therapists should at least observe responses to each cue closely for a few items on which WFD are displayed and to use this as a basis to decide on which intervention approach to follow with each child. Clinical practice recommendations based on our results are in line with Best et al.'s suggestion of tailoring interventions according to each child's language profile. However our regression analyses pointed out that it is better to target a child's weakness rather than his strength first.

We now shift to consider the possible mechanistic basis for differential effects of behavioural interventions, and turn to connectionist computational modelling as the next step of this investigation. Hence, Chapter 8 provides an account of the Alireza, Fedor and Thomas (2017) study, in which intervention simulations were run on the modified Best et al. (2015) model to investigate the question of whether therapy should focus on a weak skill to try to remediate a weakness or focus on the child's strengths to try to find a compensatory skill to work with instead. The following chapter will first present a description of the development work completed on the design of the Best et al. (2015) model and the process involved in building appropriate training sets to be used in the connectionist modelling investigations.

Chapter 7

Building Training Sets for Computational Modelling

This chapter presents a description of the further development work completed on the Best et al. (2015) model, initially the building of appropriate training sets and comparing alternative architectures. The detailed background of the Best et al. (2015) connectionist model and the associated research were given in Chapter 3. Here, we briefly recap: Best and colleagues used a connectionist model of word acquisition to investigate methods of intervention for WFD by extending the idea of mapping phonological and semantic representations in self organising maps to simulate word acquisition similarly to the DevLex-II model (Li, Zhao & MacWhinney, 2007) and Mayor and Plunkett model (2010) of early word learning. The Best et al. model was based on an autoassociator with a backpropagation learning algorithm (Rumelhart, Hinton & Williams, 1986a) that simulated learning associations between phonological and semantic representations to develop internal distributed representations of each. The model was based on the developmental trajectories approach where the disorder was set in the context of how the specific skills evolved for typically developing children. It was a simple connectionist model that was first set to simulate the mechanisms of vocabulary acquisition for typical development. It was then modified to display atypical development and then a series of simulations were run to explore the effectiveness of intervention on different components of the theoretical cognitive processes. As mentioned earlier, two major limitations of the Best et al. (2015) model were initially found to be: (i) the model found it much harder to learn semantic hidden representations than phonological hidden representations; (ii) a new abstract training set was generated for each simulation, which made it difficult to compare results of the simulated interventions to previous non-intervention simulations and to the empirical data from children. Thus one of the main aims of this project was to develop a more realistic vocabulary corpus that could be used for further simulations, thereby advancing the associator model towards a more ecologically valid version. Another advantage of this study was the notion that phonological and semantic interventions could be simulated and assessed both as separate interventions and then as a combined intervention strategy.

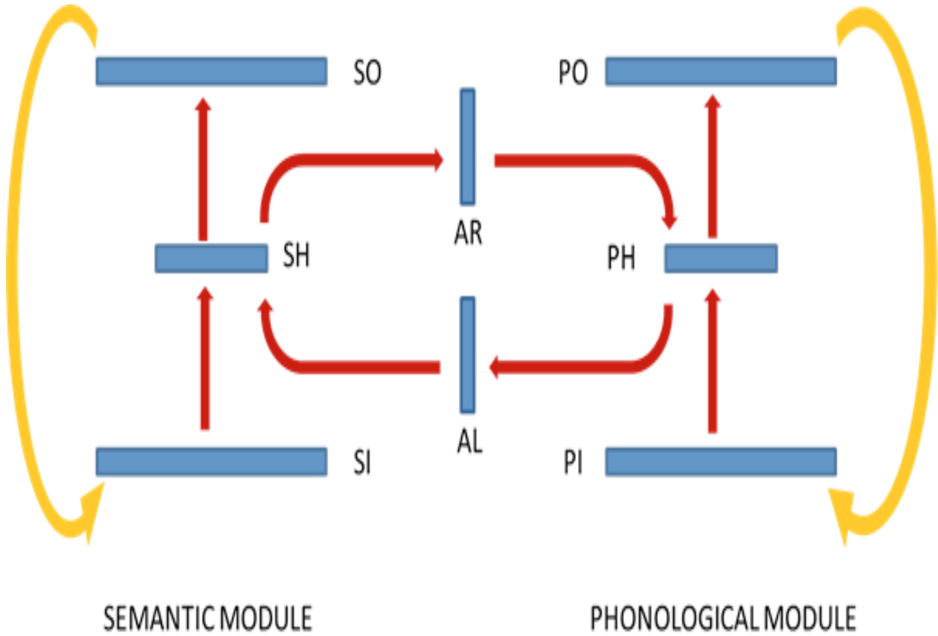
This chapter will first present a brief description of the Best et al. (2015) model, its assumptions and constraints to motivate the precise area of further model development. It then presents the steps taken to build up more appropriate training sets to run simulations of word acquisition. This is followed by a description of two alternative architectures for models of word acquisition and their evaluation in pursuit of choosing the most pertinent model for our purposes, specifically, contrasting semantic, phonological, or combined interventions to address WFD. Each of the two model architectures was initially tested using two small word sets, the Small Real and Small Confusable training sets of 57 words each, which are described in a later section. The aim of the preliminary testing was to simulate typically developing word acquisition with trajectories similar to the trajectories shown in Figures 5.1 and 5.2 (chapter 5), while at the same time producing error profiles similar to the ones obtained from the same cross-sectional, empirical data from TD children (Best et al. (2021)). The small word sets were a convenient tool used to investigate the changes in the models' outcome relative to incremental changes in their parameters where the simulations were completed in a reasonable amount of time. However the small number of word sets could not be used in simulations for interventions because the outcome would not contain enough data for analysis. Hence once the better model architecture was identified for the purpose of this study, then the larger word sets of

397 words, described below, were used to run simulations of interventions, which took a relatively long time to complete.

Background of Computational Model

The Best et al. (2015) model was made up of two autoassociator modules, one for semantic representations and one for phonological representations, each with a hidden layer of 500 hidden units. The two hidden layers were connected to each other via two pathways, one to permit a mapping of semantic to phonological representations, the other to permit a mapping of phonological to semantic representations. Each pathway contained a hidden layer of 500 hidden units. The size of the semantic input and output layers was 57 units; the size of the phonological input and output layers was 171 units. The model architecture is shown in Figure 7.1. The model was programmed using Matlab.

Figure 7.1. Architecture of the associator model of vocabulary acquisition constructed by Best et al. (2015). A module for semantic representation is connected to a module for phonological representation by two associative pathways each with 500 hidden units. The size of both the semantic and phonological hidden layers is 500 units. The size of the semantic input and output layers is 57 units, the size of the phonological input and output layers is 171 units.



Blue bars = layers of units
SI = Semantic Input, SH = Semantic Hidden, SO = Semantic Output
PI = Phonological Input, PH = Phonological Hidden, PO = Phonological output
AL = Associative Layer to the left (from the phonological to the semantic module
AR = Associative Layer to the right (from the semantic to the phonological module)
Red arrows = learning weights between units
Yellow arrows = recurrent weights which were not trained

The Best et al. (2015) Associator model was a successful proof of principle that developmental trajectories of vocabulary acquisition could be altered by internal computational constraints and that training interventions could improve outcomes. However it had the following limitations: its behaviour was determined by the greater difficulty of learning semantic representations compared to phonological representations; the training set of 100 items was too small; a new abstract training set was generated for each simulation, which made it difficult to compare results of the simulated interventions to previous non-intervention simulations; while semantics used prototypes or family resemblance around categories and phonology employ articulatory features, phonological representations were very sparse within phonological space (few vectors in a large space) and there was no notion of phonological neighbourhood, thereby limiting opportunities for phonological errors. This is one possible reason the model found it harder to learn semantic hidden representations than phonological hidden representations. Hence initially, the aim was to increase the training set to an ecologically valid lexicon of 500 monosyllabic words instead of 100 abstract words.

Training Set Design

Semantic features have been proposed to be the building blocks of semantic representation (McRae, Cree, Seidenberg & McNorgan, 2005; Vinson & Vigliocco, 2008). One of the objectives of designing a new semantic feature training set was to scale up the model by including more psychologically constrained and ecologically valid training items, instead of the abstract simulations that generated semantic representations. Semantic feature lists are not considered as literally static lists present in the mind of participants; the assumption is that “participants construct a holistic simulation of the target category” (Barsalou, 2003, p. 1184). For example, in the instructions of their study, Vinson and Vigliocco (2008) asked participants particularly not to use dictionary style definitions of words, nor word associations as features. Vinson and Vigliocco (2008) provide an example of the list of features given by participants to define the word ‘a dog’: “pet, animal, has fur, barks, 4 legs, friendly, has a tail, mammal,…” (p.190). It is these reflected semantic feature representations that connectionist models seek to capture in the activations of hidden units (McRae, et al., 2005).

Three sources of semantic feature sets were identified as potentially useful for the current work. Two sets were collected from adult participants (McRae et al., 2005; Vinson & Vigliocco, 2008). The third feature set was collected from an online algorithm that used WordNet (Miller, Beckwith, Fellbaum, Gross & Miller, 1990) to generate semantic features for English monosyllables (Harm, 2002). WordNet is an online lexical database, which is based on adult vocabulary, in which the lexical information of words are organised in terms of word meanings rather than word forms. It is designed to be used under computational programming. English nouns, verbs, adjectives, and adverbs are organised into sets of synonyms, each representing one underlying lexical concept (Miller, 1995). The fact that WordNet is organised by semantic relations renders it a useful tool for computational linguistics and natural language processing. WordNet has been used for many language applications such as word sense disambiguation, information retrieval and text categorisation (Jing,1998). The semantic feature sets derived from human participants such as the first two sets mentioned earlier (from McRae et al. (2005) and Vinson & Vigliocco (2008)) would typically have fewer features compared to feature sets derived by computational programming using WordNet. This is because for the first two sets, adults were asked what the salient and diagnostic features are for a large set of concepts (e.g., zebra = four legs, stripes, lives in Africa, eats grass). From this, a complete feature set was generated, and each concept was defined by the subset of features it contained. There were typically very few features for each concept from the full set of features, a so-called sparse representation. WordNet is an automated computational system for representing concepts and their relations to each other, and it is possible to generate vectors which represent the similarity of each concept to the others based on their position in the tree-structure knowledge base.

In view of the fact that the real sets were obtained from adult participants and because of a lack of a semantic feature set from children, the MRC Psycholinguistic Database (Coltheart, 1981) was initially used to generate the top 500 monosyllabic words with the lowest Age of Acquisition to represent a vocabulary set appropriate for child vocabulary acquisition. Training sets for connectionist models of language acquisition also take into account the frequency of children's exposure to different language elements, which then alters how often items are included in the training set. In this case, word frequency for vocabulary appropriate to children was derived from The Children's Printed Word Database (CPWD) (Masterson, Stuart, Dixon, Lovejoy, 2010). The Children's Printed Word Database is an online database of the vocabulary in reading materials used by 5-9 year old children in the UK (Masterson et al., 2010). The strategy followed was to choose words from the Vinson and Vigliocco (2008) and the Harm (2002) feature sets that were most appropriate for a developing vocabulary, also noting the frequency of chosen words from the CPWD (Masterson et al., 2010).

Three training sets were constructed for the purpose of running simulations to test the effects of intervention on WFD. Two of the training sets were derived from a combination of word lists obtained from Vinson and Vigliocco (2008) and the CPWD (Masterson et al., 2010). They each included 397 words, but one comprised a matrix with analogue values for semantic features (henceforth referred to as 'V&V 397 analogue semantics') and the other comprised binary values for semantic features (referred to as 'V&V binary semantics'). The third set was made up of 1444 words and was derived from a combination of word lists from the online algorithm that used WordNet (Miller et al., 1990) to generate semantic features for English monosyllables

(Harm, 2002) and the CPWD (Masterson et al, 2010) which included a matrix with binary values for semantic features (referred to as 'Harm 1444 binary semantics').

A further two smaller training sets were used during the design and testing of the productive vocabulary model (PVM) model. Each set was made up of the same list of 57 items created by Small, Hart, Nguyen and Gordon (1996) in their connectionist model of acquired aphasia and was based on pictures taken from an aphasia naming test battery. Small et al. created a bespoke semantic feature set for these items that differentiated them as animals, vegetables, tools, and vehicles. Phonological vectors were created for the Small et al. set using a localist scheme (one unit per phoneme in each word position of onset, coda, and nucleus). This was called the 'Small 57 real' set. However, this created a phonological space that was sparsely populated. A final test set generated a small artificial language from a more restricted set of locally encoded phonemes, producing greater opportunity for interference, neighbourhood effects, or confusion in learning phonological structure. This was called the 'Small 57 confusable'. Following is a description of the background, design principles and steps taken in building the five above mentioned training sets. Note, although the set of feature norms from McRae et al. (2005) was available, it was not utilised because it would have needed a considerable amount of time to reorganise the data in a manner that could be used as input for the neural network employed for the purpose of this project.

The V&V 397 Analogue and Binary Semantics Training Sets. The set of feature norms collected by Vinson and Vigliocco (2008) from 280 participants was obtained and training to learn the whole set was explored. There were originally 456 words in total in the Vinson and Vigliocco (2008) list and a total of 1029 features collected from 280 participants. The feature set contained 240 nouns, of which 169 referred to objects and 71 referred to events, along with 216 verbs that referred to events. A matrix of 1029 features by 456 words was obtained. It had a range of 0 to 20, where values corresponded to the number of participants from the adult studies who had generated a given feature for a given word. The matrix was transposed to give 456 words by 1029 features and scores for the features were normalised relative to the words to give values between 0 and 1, using the minimum-maximum feature scaling method [1]. The matrix with activation values between 0 and 1 was then used as input for the connectionist model.

However, this matrix comprised analogue semantics values which can be harder for ANN's to learn if they use sigmoid activation functions, since such values fall within the narrow dynamic range (versus saturated at 0 and 1 for large negative net inputs or large positive net inputs) and therefore require more careful calibration of weight values during training. Hence another version of the semantics matrix was created with a binary set where any value greater than 0 was set to be 1, i.e., if any adult listed a given semantic feature as relating to a word in the Vinson and Vigliocco (2008) list, that feature was set to 1 in the semantic code for a word.

The list of 397 words was derived by choosing words that were found in both the Vinson and Vigliocco (2008) semantic features list and the CPWD (Masterson et al., 2010). The Vinson and Vigliocco (2008) list included some of the action words twice, once as a noun and the other as a verb. Both words were included in the V&V 397 word training set and were disambiguated by a context vector (see below). The frequency was recorded to be the same for both words because the CPWD list only

[1] Feature scaling is used to bring all values into the range 0 to 1 using the following formula: X' = (X – Xmin) / (Xmax – Xmin). Here X is the original value, X' is the normalized value, Xmax and Xmin are the maximum and minimum values of the features respectively.

included the words once without indicating whether the word referred to a verb or noun. The Vinson and Vigliocco (2008) list only included verbs in their present form. For the purpose of keeping the training set as large as possible the training set did not exclude verbs that were listed in the CPWD in their past or present participle tense only, while they were included only in their present form in the Vinson and Vigliocco list. Similarly, the training set did not exclude nouns that were included in the CPWD only in their plural form whereas the Vinson and Vigliocco list included them only in their singular form. An exception was made for the word 'Trade'. 'Trade' was not in the CPWD but 'Traders' was included with a frequency of 3. Although the Vinson and Vigliocco list included 'trade' twice, in this case it was decided to include 'trade' once in the training set with a frequency of 3. There were originally 456 words in the Vinson and Vigliocco (2008) list but 55 words were taken out because they were not included in the CPWD list. The MRC website was used to transcribe the words into their phoneme symbols. The transcriptions for 'Miaow ' and 'oink' were not available on the MRC dataset hence they were also excluded. Since they were each included twice in the Vinson and Vigliocco list, four more words were taken out of the list. Thus, the semantic feature training set had 397 words in total for which the phoneme symbol transcriptions were obtained.

Matlab was then used to make a cluster analysis of the 397-word semantic training set in order to arrange the words into 20 semantic categories. The categorisation of words was made in preparation for classifying naming errors as semantically related or unrelated as this would be needed to differentiate between semantic and phonological errors at a later stage when simulating word acquisition.

An appropriate phonological pattern generator that would convert the strings of phonemes into binary patterns that could be used as input for the phonological module of the autoassociator model was identified. This contained 46 legitimate codes for the phonemes in the English language based on 19 binary articulatory features (Thomas & Karmiloff-Smith, 2003). Phonological representations employed a left-justified slot-based CCCVVCCC-CCCVVCC-CCCVVCCC syllabic scheme, accommodating monosyllabic, bisyllabic and multi-syllabic words. With 24 available slots, each containing a 19-bit phoneme vector, phonological inputs and outputs were therefore encoded across 24 x 19 = 456 bits. An exception was made for the word helicopter, which has been encoded as HE-LI-COPTR, with a very complex final consonant cluster. Frequencies were taken from the CPWD, with different transformations considered during initial piloting (raw counts per million, log10 transformed raw counts, an intermediate categorical version with 10 levels). To resolve ambiguous homophones (e.g., noun versus verb usage of the same word: same phonology different semantics), a 4-bit phonological context vector was added. This coded verb, object-noun, event-noun.

A point that should be taken into account is that derivation of the words from the Masterson set improved ecological validity; however, the V&V 397 semantics training set were adult derived. Within the embodied account of semantics (Barsalou, 2003), features emerge from invariances in the sensori-motor experiences of a concept across physical and social contexts. In some sense, adult-derived features represent the ultimate target of the acquisition of concepts. However, the child's level of cognitive development may not have reached a level where they can realistically possess all the adult semantic features as primitives. Therefore, the adult-derived feature set must be viewed as a limiting simplification.

The Harm 1444 Binary Semantics Training Set. The list of 1444 words was derived by choosing words that were present in both an online algorithm that used

WordNet (Miller et al., 1990) to generate semantic features for English monosyllables (Harm, 2002), and the CPWD. Again, the strategy was to choose words from the Harm (2002) feature set that were most appropriate for a developing vocabulary, noting the frequency of chosen words from the CPWD. The original list that has been obtained from Harm and Seidenberg (2004) was a list of 6103 monosyllabic words that were encoded using 2442 features. Each word had a list of numbers on a row because each feature was encoded with a number. Another list, termed the key list, which defined the semantic feature for each number in the previous list was also obtained. Only nouns and verbs in the present tense were included in the list in order to be able to make the simulations as close to behavioural experiments as possible. Abreviations were omitted. Semantic feature lists for a total of 1444 words with 1879 semantic features were eventually compiled from the Harm and Seidenberg (2004) 6103-word list. Of the original 2442 semantic features, 1879 were selected which pertained to semantic rather than syntactic distinctions, since semantic but not syntactic distinctions are observed to characterise children's naming errors in WFD (Dockrell, Messer & George, 2001). Since features were referred to as numbers, a script was developed in Matlab that converted the feature lists for the 1444 word list compiled from the Harm & Seidenberg (2004) into binary information that could be used as a semantic training set.

Similarly to the V&V 397 word sets, the MRC website (Coltheart, 1981) was used to transcribe the words into their phoneme symbols; and the same phonological pattern generator that converted the strings of phonemes into binary patterns was used in order to produce the set of phonological representations to be included in the training set. A 4-bit phonological context vector was also added to resolve ambiguities between homophones (verb, object-noun, event-noun and idiosyncratic ambiguities, e.g., pair/pear).

The 57 Word Training Set. A smaller realistic training set was needed during the design and testing of the PVM to pilot learning algorithms and automated error classification in a realistic amount of time. Small et al. (1996) designed a set of 57 items, based on pictures taken from an aphasia naming test battery, and created hand crafted semantic representations over 77 features (see appendix 5). Their training set was then used in an autoassociative model of acquired semantic deficits (also used in modified form in Thomas and de Wet (1999) and Richardson and Thomas (2008) in further models of acquired deficits using autoassociative and self-organising maps, respectively. This training set was used as a starting point: the semantic features were revised and redundant features were dropped, phonological representations of the names of the items were added (mainly animals, tools, and vehicles) over the same phonological vector as the V&V 397 set. Frequency estimates were derived from Kucera-Francis database via the MRC Psycholinguistic database (again, either raw, log10 transformed, or intermediate categorical version). This created the 'Small Real' training set of 57 items.

While the Small 57 Real set provided ample opportunity for semantic errors, the phonological space of 57 names in a space defined by 456 phonological features (24x19) was again too sparse for much phonological confusability: thus, there was little scope for phonological errors or generalisation of phonological knowledge to repeat/autoassociate non-words. Hence the Small 57 set of semantic features was then linked to an artificial language of monosyllabic words, to create the 'Small Confusable' set. The artificial language was encoded using the same phonological features, but just a single syllable (8x19 features). The artificial language was based on Thomas (1997) comprising sets of rhymes and a few non-rhyme items.

Phonological strings were arbitrarily associated with semantic representations once to create the training set, and the same mappings retained thereafter. The Kucera-Francis frequencies of the original words were retained. A denser phonological neighbourhood was expected to create greater scope of phonological errors, as well as higher non-word repetition accuracy. Comparison of Small Real and Confusable training sets allowed consideration of the impact of phonological neighbourhood density on error patterns.

List of Non-words. In order to test auto-association of the V&V 397 and Harm 1444 training sets in the phonological component, two non-word repetition sets were combined to generate the generalisation set. There was a 20-item set with non-words between two and three syllables in length taken from the Children's Non-word Repetition Test from Gathercole et al. (1994). The other source was a 100-item set of monosyllabic non-words created by Seidenberg, Plaut, Petersen, McClelland and McRae (1994) and used in the Plaut and McClelland (1993) Model. These were converted into phoneme strings which were then organised into syllable slots in the same format as the above mentioned training sets. Given the 397 set covered a relatively small proportion of phonological space, particularly for multi-syllabic words, only low levels of generalised auto-association were expected.

In order to test the auto-association of the Small 57 Real and Confusable item sets, a generalisation set of 25 non-word items was created with overlapping phonemes to the training set. The expectation here was that the sparse phonological space (i.e. large dimensionality compared to number of training items) found in the Small 57 Real set will produce low levels of non-word generalisation whereas the denser phonological space of Small 57 Confusable will produce higher rates of generalisation.

Preliminary Simulations

An initial neural network architecture was constructed using OXlearn, a Matlab based simulation software (Ruh & Westermann, 2009), for the purpose of preliminary testing the semantic module separately. The whole set of feature norms collected by Vinson and Vigliocco (2008), which contained 456 words and 1029 semantic features was initially used for the trials in order to assess modifications that could be implemented on the semantic module of the associator model of Best et al. (2015). More importantly the pilot simulations were run in order to find out whether the V&V training sets were learnable. The difference between using the analogue and binary versions of the V&V semantic features vectors was also investigated.

Method

The normalised semantic vector matrix of the Vinson and Vigliocco (2008) word list was used as the input for an autoassociator connectionist model, where the number of training patterns corresponded to the number of 456 words, and the normalised scores for each feature corresponded to the input and target activation values for each word. The model was a three-layer autoassociator with a backpropagation learning algorithm and sigmoidal activation functions which had 1029 input units and 1029 output units. Explorations for simulation of typical development involved changing the number of hidden units, the value of the learning rate and the value for the momentum systematically, and inspecting the performance of the network in terms of the mean square error and percentage of correctness according to a specified criterion level as follows.

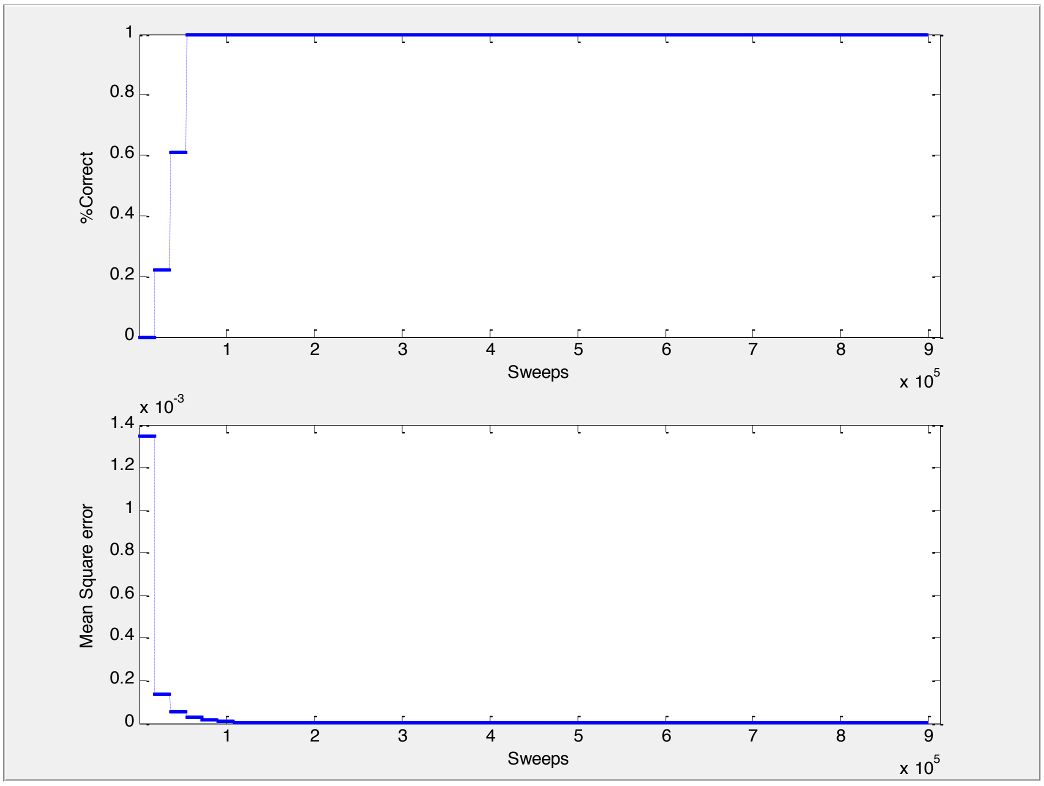
At first, the model was given 500 hidden units and various values of learning rate ranging from 0.01 to 1 and momentum values between of 0.1 and 0.5 were tested. The aim was for learning to be 100% correct at a criterion level of <.1 deviation. This means that for each of the output units the deviation of the output activation from target is evaluated and that each sweep (one sweep means one pattern) will be deemed correct only if none of the output units deviates by more than 0.1 from the target value (Westermann, 2011).

Results

Eventually 100% correct learning was achieved at a criterion level of <.1 deviation after 912000 sweeps or 2000 epochs (where one epoch means one pass through all the training patterns), when the number of hidden units was changed to 100 at a learning rate of 0.1 and momentum of 0.3. Figure 7.2 shows the performance of the model with these parameters in terms of percentage of correctness and the mean square error. This was an indication that the Vinson and Vigliocco (2008) set of feature norms was learnable and could be used as a training set.

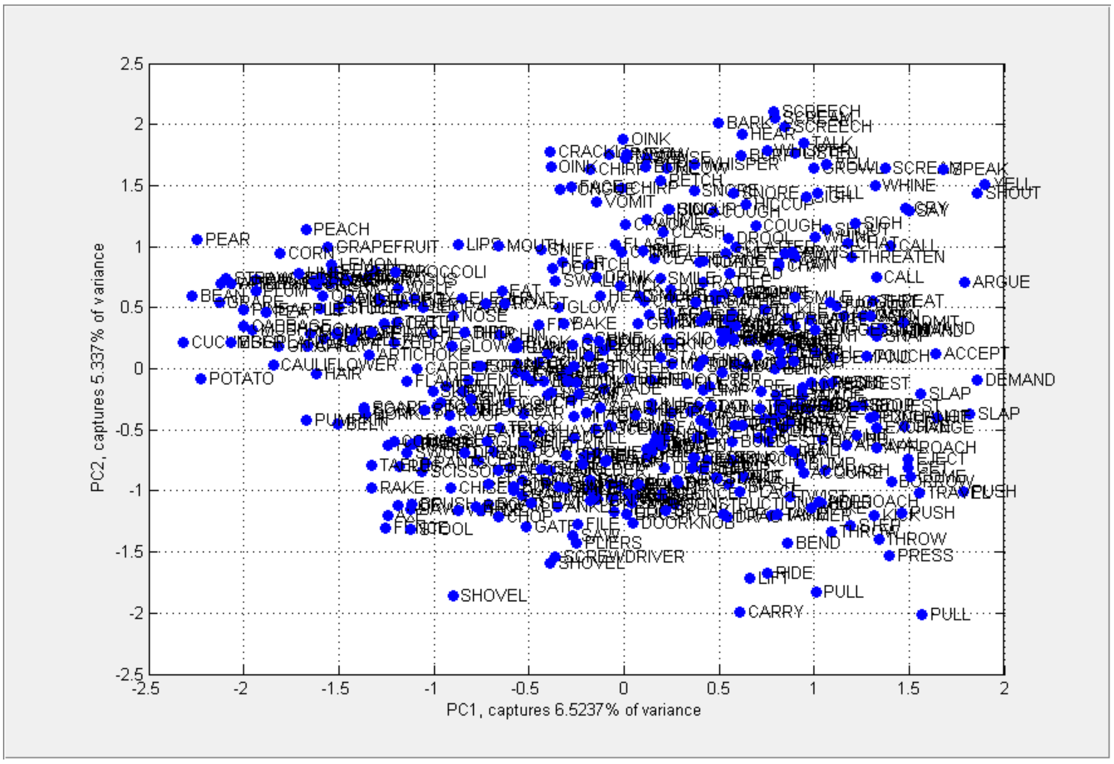
Interestingly, increasing the number of hidden units seemed to increase the number of sweeps that the model needed to learn the words. This effect is consistent with the findings by Bullinaria when Discriminant Analysis was used to investigate subspaces of hidden unit activations for pronouncing different versions of ‘i’ (Bullinaria, 1997). This point would be taken into account when designing methods to simulate trajectories of atypical development in the model.

Figure 7.2 Sample of plot of results obtained for preliminary training on the 456 words from Vinson and Vigliocco (2008) feature set. For this result the simple autoassociator had 100 hidden units with a learning rate of 0.1 and a momentum value of 0.3; simulations were run for 912000 sweeps (equivalent to 2000 epochs).



In the same investigation, once simulations of typical development were captured, Principal Component Analysis (PCA) was applied to hidden layer activations at the end state of typical development. The training set are plotted according to their values on the first two principal components, shown in Figure 7.3. Although the PCA is crowded with words, it is possible to discern clustering of words with similar meanings such as fruits and vegetables are close to each other, illustrating that the model has captured emerging semantic structure despite the large, sparse representations of meaning.

Figure 7.3. Principal component analysis plot showing the variance captured by the first two principal components of hidden unit activations after 2000 epochs of training. The first principal component captured 6.52% of the variance and the second captured 5.34% of the variance.



Next, the effect of transforming activation values of the input and target units to binary numbers was investigated. It was found that the model needed fewer epochs to reach the same accuracy level when the values of the feature set were turned to binary scores.

Cluster analysis using the Ward method was used to assess whether the emerging categorisation structure of the simulations changed if the activation values of the input and target units were converted to binary scores. The Ward method measures the incremental sum of squared Euclidean distance between two points. It is a hierarchical clustering method ‘where the sum of squares starts out at zero (since every point is its own cluster and then grows as the clusters merge and Ward’s method keeps this growth as small as possible' (https://jbhender.github.io/Stats506/F18/GP/Group10.html#ward-linkage-method). After several trials the most appropriate number of categories for the V&V 397 analogue semantic word set was found to be 20 categories. When the Ward clustering method was used on the binary scores of the V&V 397 word set instead of the analogue scores, the resulting 20 categories were made up of lists similar to the ones resulting from analogue scores (appendix 4 provides the categories lists resulting from using the Ward clustering method on each of analogue and binary scores for comparison). Hence binary scores were considered a better option to be used in training sets since they produced faster results. Appendix 4 also presents the categories of the words provided in the Vinson and Vigliocco (2008) study for comparison.

Model Architecture

The next step was to incorporate both the semantic and phonological modules into the modified Best et al. (2015) associator model, permitting the same PCA procedures for simulations of typical and atypical development and intervention simulations. Two models whose architectures were based on the Best et al. (2015) model were developed and tested with the Small 57 word sets and their results were compared.

Method

The two models were similar in that they each had a separate semantic and a phonological module that comprised 3-layer autoassociators. But the difference between the two architectures was that the first model, named the Linked Internal Representations Architecture (LIRA), was based on the DevLex-II model (Li, Zhao & MacWhinney, 2007), which used self organising maps and had two separate pathways linking the semantic and the phonological modules, one for naming and one for comprehension (see Figure 7.4). The LIRA model hence learnt associations between semantic and phonological codes when they were at different stages of development. While the second model, named the Shared Associative Layer Architecture (SALA), was based on the Plunkett, Sinha, Møller and Stransby (1992) model, which had only one layer between the semantic and phonology systems that was shared between naming and comprehension and had predetermined semantic and phonological codes (see Figure 7.5). The neural network for both the LIRA and SALA models was constructed using PlaNet 5.7 simulator software running on the Linux operating system and interfaced with c-shell scripting.

Figure 7.4 The Linked Internal Representations Architecture (LIRA), based on the DevLex-II model (Li, Zhao & MacWhinney, 2007).

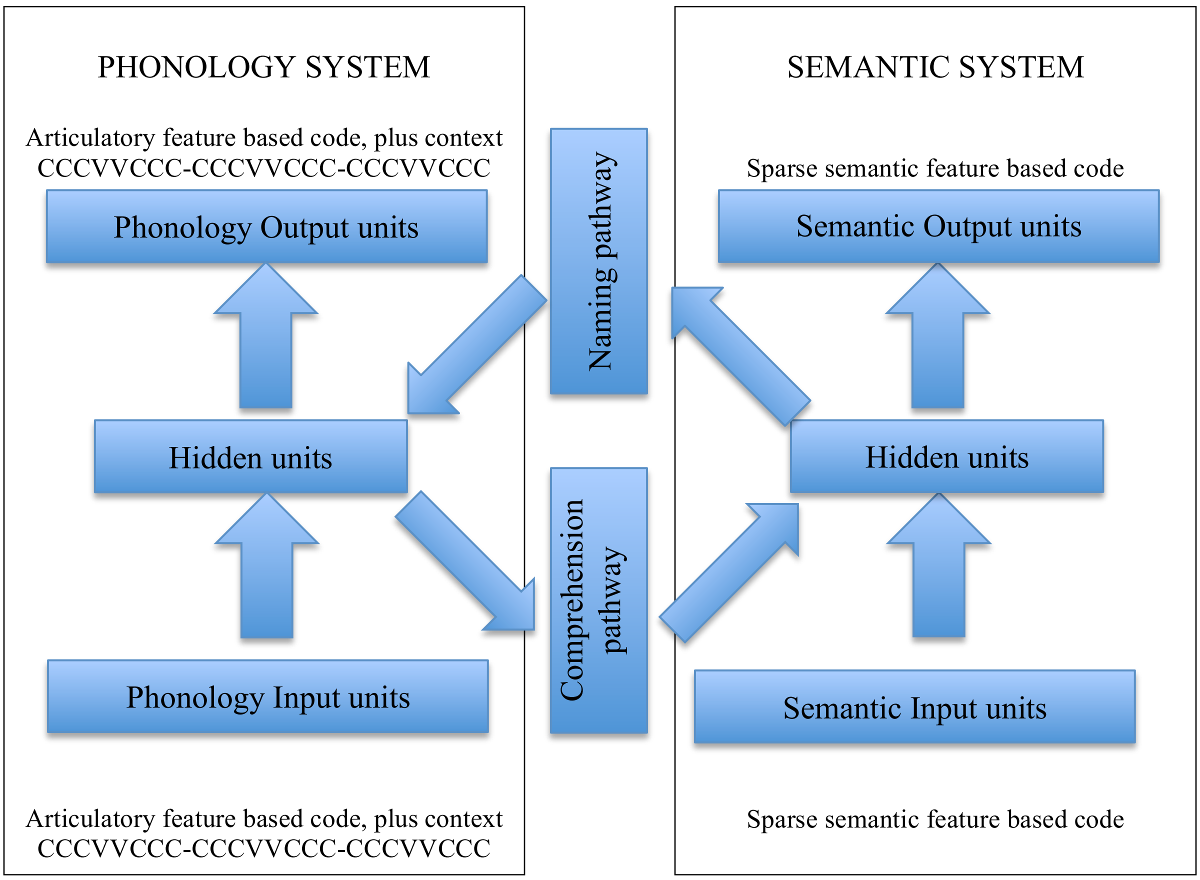
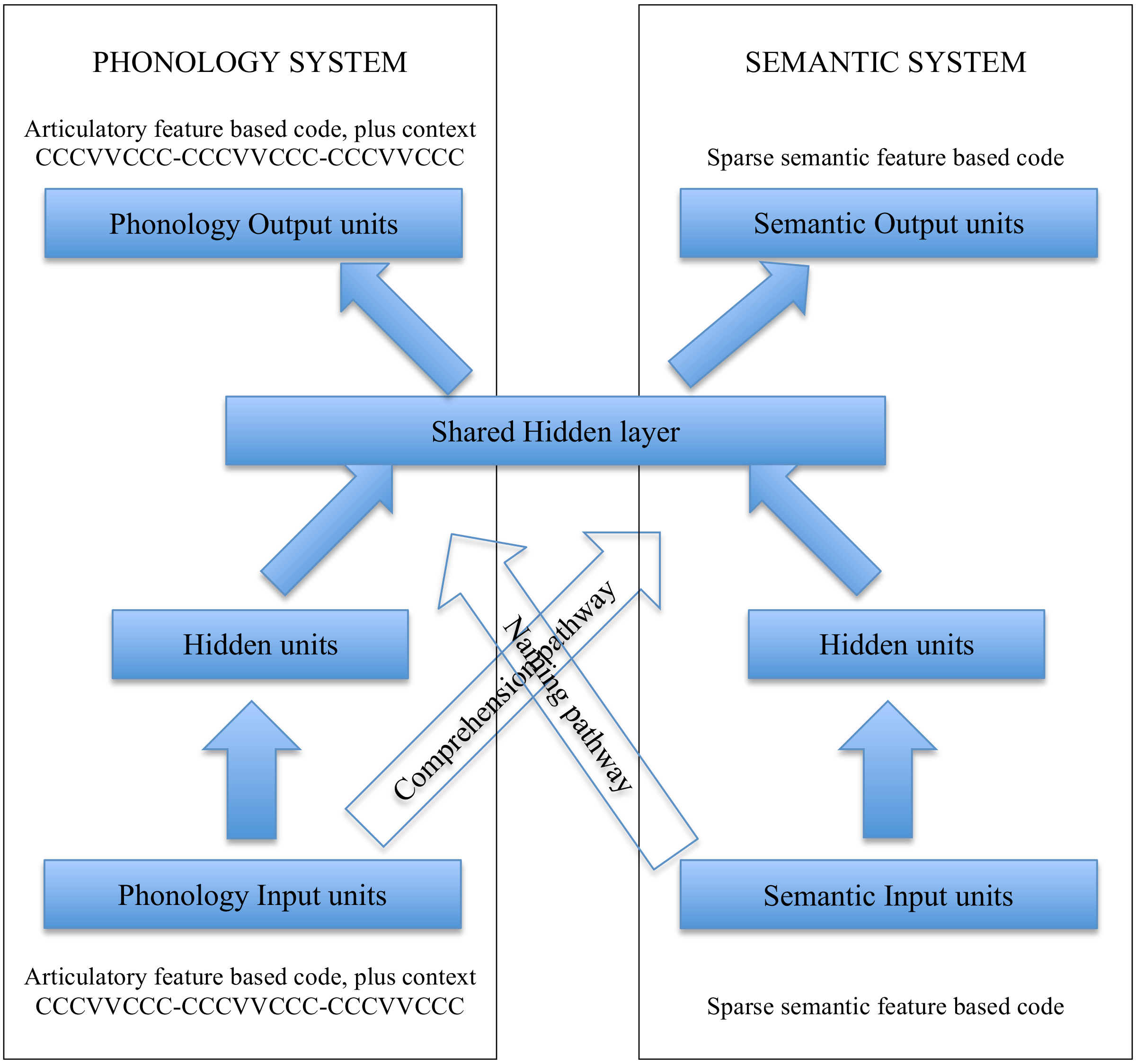


Figure 7.5 Shared Associative Layer Architecture (SALA), based on the Plunkett, Sinha, Møller and Stransby (1992) model.



The number of input and output units for the semantic and phonological modules were set at the beginning of the simulations according to the number of features in each of the training sets. During testing and trial simulations a number of parameters could be adjusted to investigate their effects. Four global parameters that were adjusted during the investigations for each model were: the learning rate of the system; momentum of the system; a parameter for the backpropagation error measure for RMS or cross-entropy (Chan, Glynn & Kroese, 2011), and a 'careoff' parameter which reduces the importance of errors for zero output units and helps to train sparse binary outputs. The careoff parameter is a scaling factor that reduces weight change in response to errors for output units that should be off compared to those that should be on. This helps the system learn sparse binary representations[2].

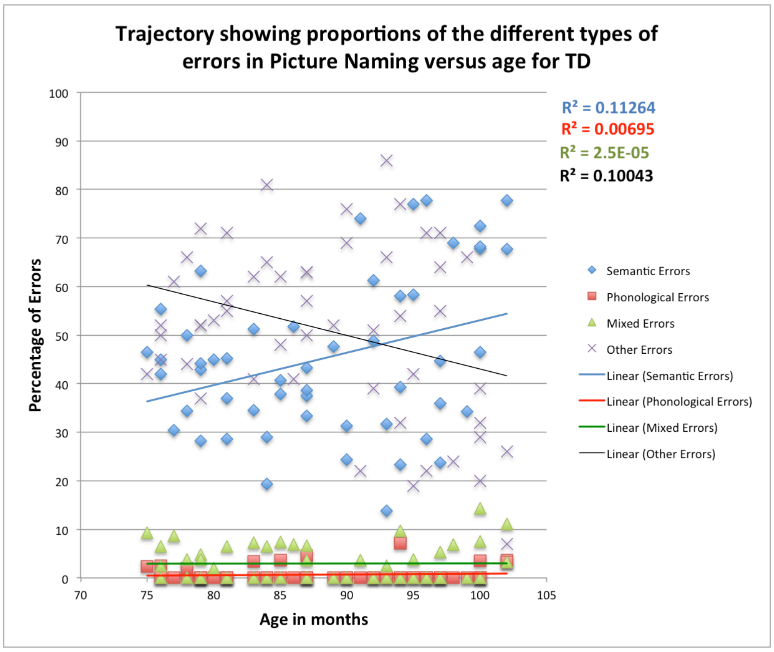
Other parameters that could be adjusted separately for each of the four components (phonology (P), semantic (S), naming (SP) and comprehension (PS)) were the number of hidden units; the learning rates; the value for the momentum; the temperature of the sigmoid activation function, the noise level and the level of initial sparseness of connectivity. The change in the processing units of artificial neural networks activation level, according to the net excitation and inhibition levels they receive, is determined by the units' activation function. The models here used a sigmoid activation transfer function. To simulate deficits, the units were made to respond less sensitively to changes in input by using a shallow sigmoid function by reducing a parameter known as "temperature". The lower temperature therefore made it more difficult for the units to discriminate between small changes in the signals they receive. Noise level was added as Gaussian noise with a standard deviation of .15 to the net input of units in the components and noise was added in the pathways with a standard deviation of .05, to provide a stochastic basis for naming errors in normal functioning. The parameter for the level of sparseness controlled the proportion of connections in any layer that are removed prior to training.

[2] The difficulty of training with large sparse binary representations, arises because the majority of the time in the training set, any given output unit is off. Therefore, gradient descent algorithms can massively reduce global error simply by turning all output units off. This is a local minimum it can then structure to escape from.

A series of trials were conducted to test the effect of changing the values of the various parameters at the start state of the simulations on both the LIRA and SALA model architectures. The trials were conducted, using the Small 57 Real and Small 57 Confusable training sets. More than 200 trials were run before eventually identifying a model with parameter values and training set that were appropriate to produce trajectories for word acquisition that were similar to those of empirical data from typically developing children (see details below). Once a set of parameter values, model and training set were identified, simulations were run another two times under the same conditions with different random seeds to show the robustness of the patterns.

The main aim of the trials was to first find values of the above mentioned model parameters that would enable the simulation of typically developing word acquisition with trajectories similar to the trajectories shown in Figures 5.1 and 5.2 (chapter 5) which were obtained from cross-sectional, empirical data from TD children (Best et al., 2021). At the same time as producing trajectories of TD word acquisition, another goal was to set the parameters of the model to produce error patterns (semantic, phonological or mixed errors) similar to the ones obtained from the same cross-sectional, empirical data from TD children Best et al. (2021) as shown below in Figure 7.6. In the Best et al. (2015) study, errors in the empirical data were classified as semantic if the mistake was of a coordinate, superordinate, functional, circumlocution or a visual attribute type; and errors were classified as phonological if the mistake was of a phonological non-word or a real word that shared at least 50% phonemes with the target word; errors were classified as mixed if they were both semantically and phonologically related (see appendix 2 in Best et al. (2015)). The main characteristics of the error trajectories in the empirical data were increasing proportions of semantic errors, low and consistent proportions of phonological and mixed errors, and high but reducing proportion of other errors (e.g, perceptual, metalinguistic/metacognitive or unrelated errors). These were taken as the principal qualitative patterns within the error data. A simplified classification of errors was used for the simulation of word acquisition in the trials of this study, where errors were deemed semantic if the word was from the same semantic category as the target word and were deemed phonological if the word shared at least 50% phonemes with the target word and was either a non-word or was not semantically related.

Figure 7.6 Trajectory showing proportions of semantic, phonological, mixed and other error types for picture naming versus age taken from empirical data from 52 TD children (Best et al., 2021).



Semantic error, e.g. coconut → *pineapple*
Phonological error, e.g. squirrel → /grirel/
Mixed error, e.g. Saw → *sword*
Other error, e.g. no response

Results

It was found that simulations using Small 57 Confusable training set with the LIRA architecture and the parameter settings in List A, gave a reasonable typical development simulation in terms of comprehension and production asymmetry when compared to the trajectories of word acquisition versus age from empirical data, but not in terms of type of errors. The outcome of this simulation showed comprehension ahead of production and each eventually reached 100% (see Figure 7.7a). These were the main properties that needed to be captured in the simulated trajectories. However these simulations produced more phonological errors than semantic errors (see Figure 7.7b) unlike the target empirical data, which showed more semantic errors than phonological errors (Figure 7.6).

Parameter settings List A:

Phonological (p) hidden units = (20)

Semantic (s) hidden units = (20)

Semantic-Phonological (sp) hidden units = (40)

Phonological-Semantic (ps) hidden units = (40)

global eta = (1)

global alpha = (1)

careoff = (1)

sparseness p = (0)

sparseness s = (0)

sparseness sp = (0)

sparseness ps = (0)

noise p = (0)

noise s = (0)

noise sp = (0)

noise ps = (0)

vanilla backprop or cross entropy = (1)

learning rate p = (0.04)

learning rate s = (0.06)

learning rate ps = (0.3)

learning rate sp = (0.2)

momentum p = (0.06)

momentum s = (0.04)

momentum ps = (0.05)

momentum sp = (0.05)

temperature p = (1.5)

temperature s = (1.5)

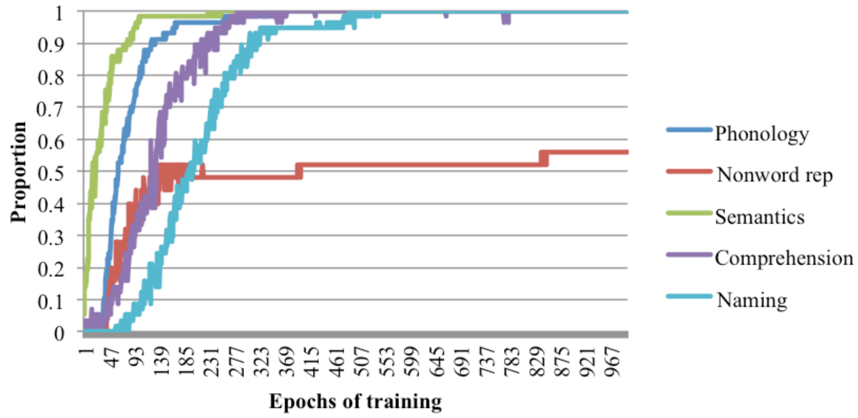
temperature sp = (1)

temperature ps = (1)

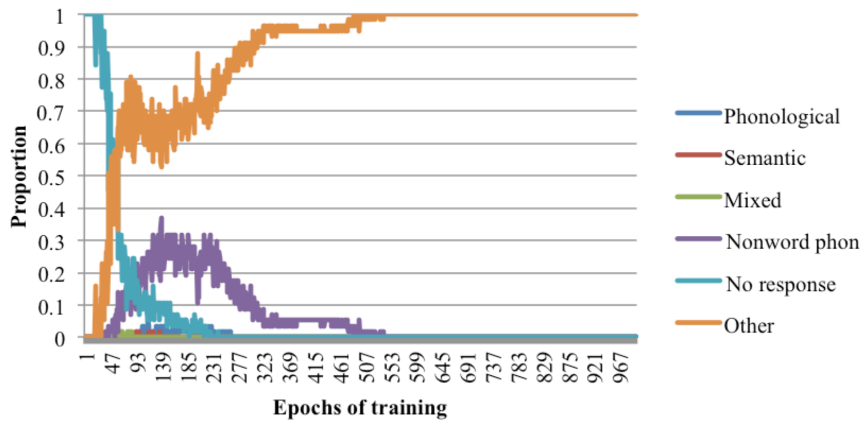
This combination of parameters consisted of symmetrical hidden units, high levels of connectivity and low noise were preferable. Higher temperatures worked better in the autoassociators and differential learning rates were required in the components to coordinate the trajectories to fit with typical development. The global eta and alpha were set to 1 which are high values normally for learning rate and momentum, but in any pathway, the effective learning rate is much reduced by the pathway specific parameters. The learning rate with the highest value was that of the comprehension pathway (ps = 0.3) followed by the learning rate for the naming pathway (sp = 0.2). The learning rate of phonology had the lowest value (p = 0.04) and semantics had a similarly low value for the learning rate (s = 0.06). Sparseness had values of 0 meaning that none of the connections were removed when the network was initialised. Noise had a value of 0 therefore no Gaussian noise was added to the net input of each unit.

Figure 7.7 Simulation results using the Small 57 Confusable training set with the LIRA architecture and parameter settings in list A.

(a) Task accuracy score:



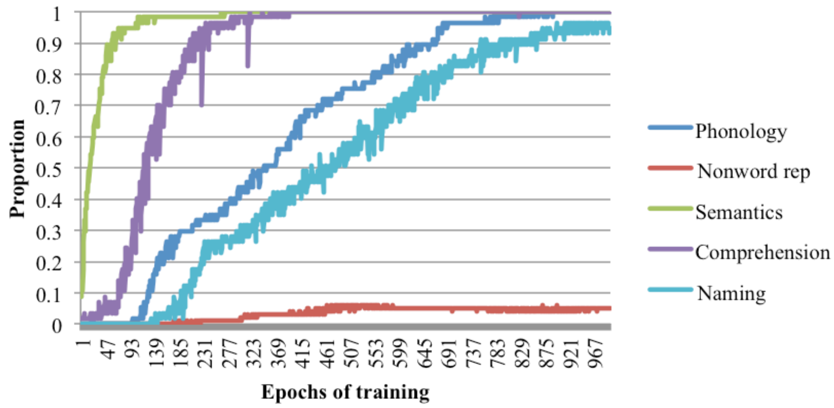
(b) Naming error proportions



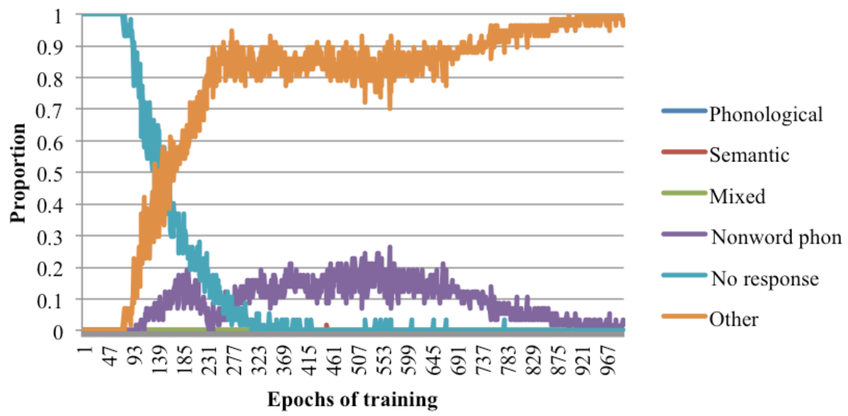
Comparison of training sets. To compare outcomes of simulations using the different training sets, the same parameters in list A were used to run a simulation with the Small 57 Real training set and the LIRA architecture (see Figures 7.8a and 7.8b). The outcome of the simulation using the Small 57 Real training set showed that naming was worse than for the Small 57 Confusable set and was only 95% correct after1000 epochs, whereas the Small 57 Confusable had reached ceiling after 480 epochs ; the outcomes of semantics and comprehension were similar for both sets; phonology reached ceiling much later for the Small 57 Real training set (after almost 800 epochs) than the Confusable set (after 160 epochs); and the Non-word repetition again was much worse for the Small 57 Real set which was less than 10% correct after 1000 epochs, whereas the outcome for the Small 57 Confusable set was over 50% correct at 1000 epochs. These differences were expected because of the denser phonological neighbourhood in the Small 57 Confusable set. In terms of differences in naming error proportions, the order of type of naming errors was similar in the outcomes of simulations for both Small 57 Confusable and Small 57 Real. i.e. unlike expectations, there were not more semantic errors than phonological errors for the Small 57 Real training set.

Figure 7.8 Simulation results using the Small 57 Real training set with the LIRA architecture and parameter settings in list A.

(a) Task accuracy score



(b) Naming error proportions



Comparison of Architectures. To compare outcomes of simulations using the different model architectures a simulation was run for the Small 57 Confusable with the SALA architecture using the same parameter settings as those used in List A for the LIRA architecture wherever possible (see Figures 7.9a and 7.9b). The parameter settings for the simulation using the SALA architecture are shown below in Parameter List B.

Parameter settings List B:

Phonological hidden units = (20)

Semantic hidden units = (20)

Association hidden units = (40)

global eta = (1)

global alpha = (1)

careoff = (1)

sparseness p = (0)

sparseness s = (0)

sparseness a = (0)

prune_epoch = (500)

foreach prob (0) # prune_probability

foreach pthresh (0.5) # prune_threshold = (0.5)

noise p = (0)

noise s = (0)

noise a = (0)

vanilla backprop or cross entropy = (1)

learning rate p = (0.04)

learning rate s = (0.06)

learning rate a = (0.2)

momentum p = (0.06)

momentum s = (0.04)

momentum a = (0.05)

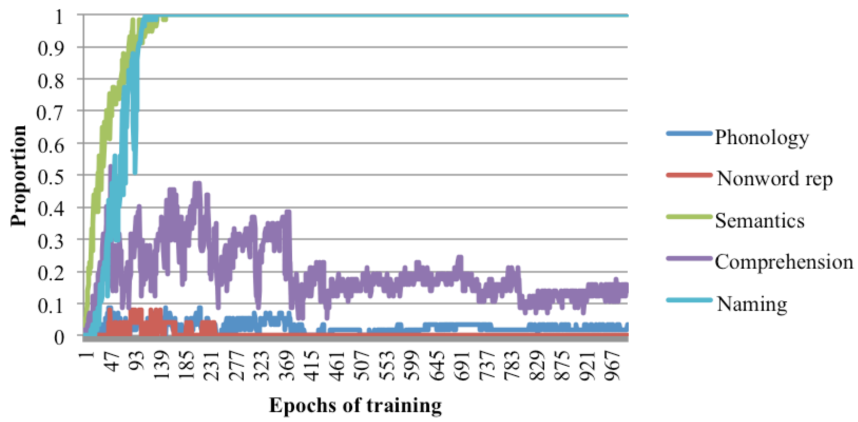
temperature p = (1.5)

temperature s = (1.5)

temperature a = (1)

Figure 7.9 Simulation results using the Small 57 Confusable training set with the SALA architecture and parameter settings in list B.

(a) Task accuracy score



(b) Naming error proportions

In this simulation, semantics and naming were both learnt quickly and efficiently by the 150th epoch. Comprehension was much more variable at the beginning although it generally improved but by the 356th epoch it started to worsen again and did not reach ceiling by 1000th epoch. Phonology stayed very low and was not learnt with these parameters. Non-word repetition stayed very low and did not improve with these parameters. Phonological and semantic errors were not showing on the naming error proportions graph. It was most likely the need to encode both mappings for naming and for comprehension within a single shared hidden layer that led to the poorer performance of the SALA architecture

Error Proportions. Comparisons of simulations using the training sets and architectures described above showed that the outcome simulations most similar to the target cross-sectional empirical data of vocabulary acquisition was obtained when using Small 57 Confusable training set with the LIRA architecture and parameter settings in list A. However these parameter settings did not enable the model to simulate error profiles similar to those obtained from the empirical data. Hence further simulations were run to investigate the effect of changing parameters while keeping the same model architecture (LIRA) and training set (Small 57 Confusable) in an attempt to produce error profiles closer to the target.

Simulations of word production with parameter settings in List A produced more phonological errors than semantic errors whereas the empirical data showed more semantic errors than phonological errors. Hence noise was added by 0.1, 0.2 and 0.3 to explore whether this would have an effect of making the produced names jump into adjacent semantic categories but still be phonologically well formed. In each case the proportion of phonological errors was reduced slightly but the proportion of semantic errors did not increase (see Figure 7.10).

The number of semantic hidden units was then reduced to 5 and noise was added to the semantic module by 0.1 and 0.2: The outcome showed that the number of semantic errors *did increase* but, even before adding noise, reducing the number of semantic hidden units to five had the effect of slowing down semantics and comprehension by about the same amount and slowed down naming even more and none of these three tasks reached ceiling within 1000 epochs (see Figures 7.11 and 7.12). Adding noise rendered the performance of the model worse on those same three tasks of comprehension, semantics and naming, again preventing them from reaching ceiling by 1000 epochs. Phonology was slightly affected and reached ceiling later when the number of semantic hidden units was reduced.

Figure 7.10. Naming error proportions for simulation results using the Small 57 Confusable training set with the LIRA architecture and parameter settings in list A except for a change in one parameter: Ns= 0.3, noise was added in the semantic component.

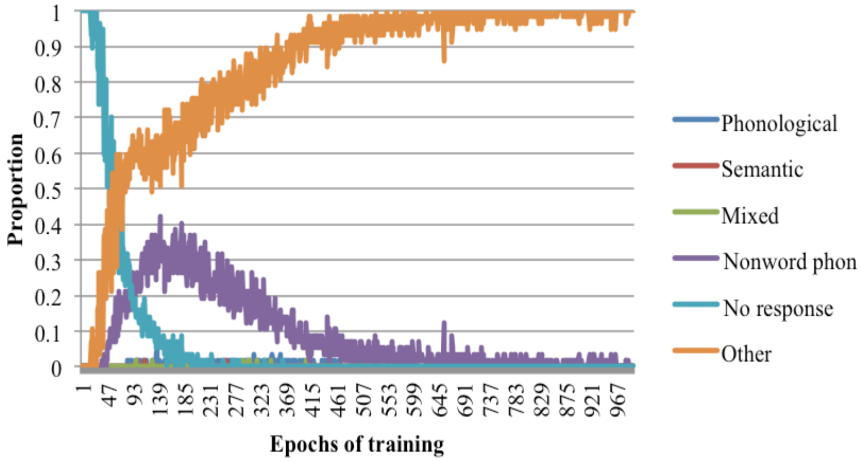
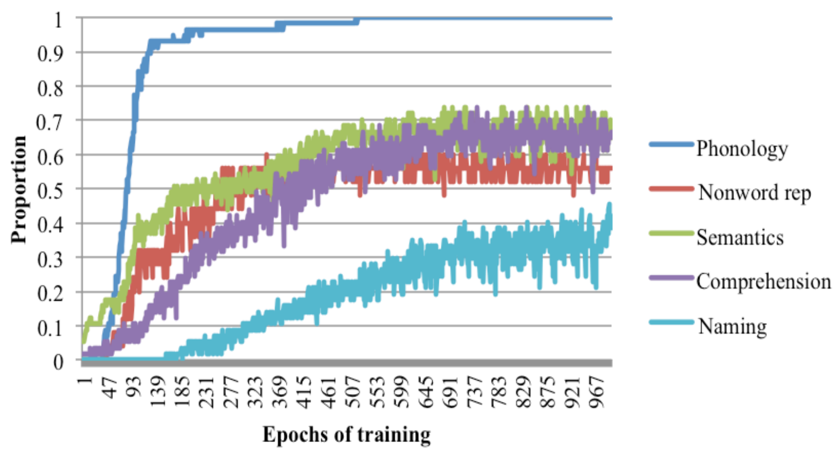


Figure 7.11. Simulation results using the Small 57 Confusable training set with the LIRA architecture and parameter settings in list A except for a change in one parameter: S = 5, number of semantic hidden units was reduced to 5.

(a) Task accuracy score



(b) Naming error proportions

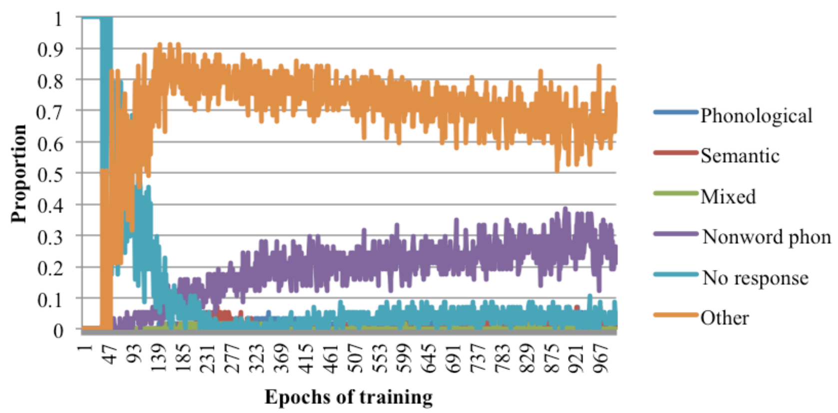
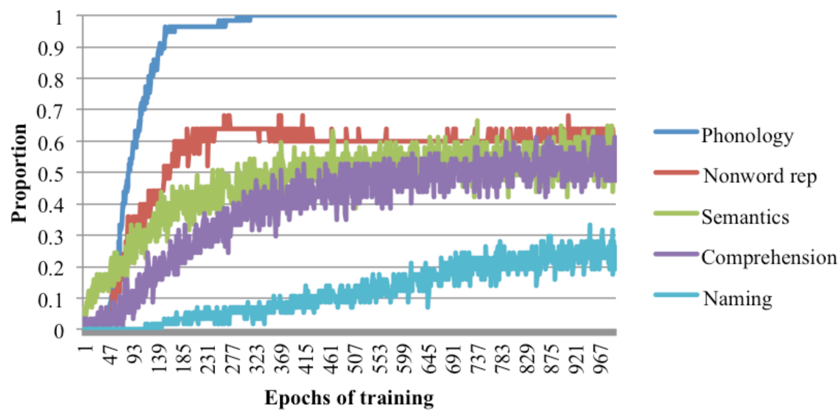
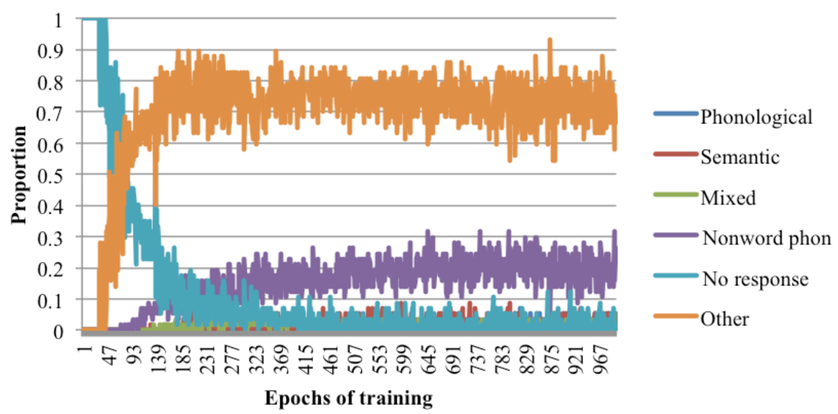


Figure 7.12. Simulation results using the Small 57 Confusable training set with the LIRA architecture and parameter settings in list A except for a change in two parameter settings: S = 5, number of semantic hidden units was reduced to 5; Ns = 0.2, noise was added in the semantic component.

(a) Task accuracy score



(b) Naming error proportions



Discussion

This chapter started with a description of the Best et al. (2015) computational model and then focused on addressing some of its limitations. Since one of the main limitations of the Best et al. model was its abstract training set, the next section presented a description of the design principles, initial aims and assumptions in building more appropriate training sets. This was followed by a comparison of two model architectures that were developed and preliminarily tested for the purpose of further research on simulating interventions for developmental language disorders.

The development of five semantic feature sets was presented in this chapter. They were all based on real semantic features given by adult participants. Three sets were developed in preparation to be used for simulations of interventions for developmental language disorders. These were the V&V 397 analogue and binary sets and the Harm 1444 binary set which only included words that were also found in the CPWD, in order to have the training sets as close to an appropriate developing vocabulary as possible. The Small 57 Real and Small 57 Confusable sets were used for preliminary testing of model architectures to find the most pertinent architecture for initially simulating word acquisition of TD children.

A neural network for the semantic module of the modified Best et al.(2015) associator model was constructed using OXlearn in order to investigate whether the semantic features vector obtained from Vinson and Vigliocco (2008) was learnable at least via auto-association. It was determined that the analogue version of the semantic features vector was learnable and that when the simulation was run with the binary version of the semantic features vector, fewer epochs were needed to reach the same accuracy level. Clustering analysis using the Ward method produced similar categories for the analogue and binary versions of the semantic features vectors. Hence binary scores were considered a better option to be used in training sets since they produced faster results.

Two models of productive vocabulary, the LIRA architecture based on the DevLex-II model (Li, Zhao & MacWhinney, 2007), and the SALA architecture based on the Plunkett et al., (1992) model, were then compared using the Small 57 Real and Confusable word sets. The simulation of word production using the combination of parameter settings in List A in the architecture of the LIRA model and the Small 57 Confusable training set produced the preferred outcome in terms of comprehension and production asymmetry when compared to the typical trajectories of word acquisition versus age from empirical data, but not in terms of type of errors. The combination of parameters included symmetrical hidden units, high levels of connectivity, low noise, relatively high temperatures and different levels of learning rates in the components.

These parameters led the model to produce more phonological errors than semantic errors unlike the target empirical data, which showed more semantic errors than phonological errors. The different error types (semantic versus phonological) resulting from simulations using the Small 57 Real and Small 57 Confusable training sets were similar in proportion, although a denser phonological neighbourhood, present in the Small 57 Confusable set, was expected to produce more phonological errors.

Alternative parameter settings were investigated in an attempt to produce error proportions more similar to the trajectories of word acquisition obtained from the Best et al. (2021) empirical data while maintaining the target comprehension and production asymmetry. Reducing the number of semantic hidden units and adding noise to the semantic module increased the number of semantic errors, however these changes also affected the performance of the model on the three tasks of comprehension, semantics and naming, and the target simulation of comprehension and production asymmetry was not maintained. The model was not sufficient, within the parameter spaces explored, to capture the high level of semantic errors observed in the empirical data.

Interestingly, it was noted that the preliminary results of error profiles which show 'other' as the majority of error types could be seen in line with the theory behind Dell's (1986) localist model on semantic error production. Dell's (1986) model accounts for the existence of multiple relations between the target word and the substituted word (or error) (e.g. Let's stop for Let's start) (p. 318) suggesting that phonological relations are mostly present in semantic errors. Dell (1986) furthermore explains the notion of the more common presence of phonological errors as opposed to semantic errors (Garrett, 1975) in speech as: "Speech errors in the theory are seen not so much as malfunctions but rather as the consequence of the need for language production to be productive, coupled with some reasonable assumptions about the way that linguistic knowledge is represented and retrieved" (p.319).

Our connectionist model was trained on the backpropagation learning algorithm (Rumelhart et al., 1986b) which works on the principle of continuously adjusting the weights of the connections in the network so that the difference between the actual output and the target output is minimized. The fact that the activations of the output units come after the mapping between the phonological and semantic pathways could explain why it becomes difficult to separate errors into the two separate categories (i.e. due to difficulties in phonological or semantic representations). Nevertheless, the preliminary results in our study show that there is potential for much more research to be done to reveal parameter settings that could enable the model to simulate the development of word production and error profiles similar to empirical data from TD children.

Moreover, the V&V 397 training sets and the LIRA associator model have the potential to be used for further research as follows: the values of hidden unit weights can be dumped at appropriate intervals of training. Then Principal component analysis (PCA) can be applied to the hidden layer activations at each of the chosen intervals in order to investigate how the network forms semantic categories from the feature sets during training. Cluster analysis can also be applied at the same intervals as PCA, and the emergence of category structure can be investigated (Rogers & McClelland, 2006). Atypical developmental constraints can be placed at the start state of the model and simulations can be run to capture atypical development. Impairments and analyses of simulations of interventions can be based on those used in a former study by Fedor and Colleagues (Fedor, Best, Masterson and Thomas, 2013) where a 2x2 design of problem domain and deficit type was implemented. Developmental deficits can be explored by reducing the number of connections (analogous to axons and dendrites), reducing or increasing the number of hidden units (analogous to neurons), changing the learning rate or inducing shallow sigmoid activations (synaptic strength) (Bullinaria, 1994). Once an atypical profile is captured by the simulations, the same procedure of dumping weights to use PCA can be made for the atypical learning trajectory.

The same test can be run on the phonological module separately using the same PCA procedure as for the semantic module. Cluster analysis and PCA can then be used to study the formation of typical and atypical category boundaries in each of the semantic and phonological components before investigating their combined effect. This may help uncover what problems occur when the model is trying to learn mappings between the codes developed in each component during training simulations of atypical development. Ambiguity either in the sending semantic codes or the receiving phonological codes would impair the learning of semantics-phonology mappings.

Preliminary investigations in this study established that the simulation of word production using the architecture of the LIRA model and the Small 57 Confusable training set with the combination of parameter settings in List A produced the outcome most similar to the trajectories of word acquisition versus age from empirical data in terms of comprehension and production asymmetry, even though not in terms of type of errors. We then used the LIRA model architecture with the larger training sets for subsequent simulations to investigate effects of intervention for WFD.

The first area of focus in this investigation was to compare interventions that work on improving strengths to try to find a compensatory skill to work with instead of trying to remediate the weakness. For example, would working on the semantic aspect, generally believed to be a relative strength for children with WFD, produce better results than working on repetition of naming, which is the weakness itself ? The next chapter provides an account of the Alireza, Fedor and Thomas (2017) study, in which intervention simulations were run on the modified Best et al. (2015) model developed and piloted here, to explore the notion of intervening to target strengths versus weaknesses to remediate problems of WFD.

Chapter 8

Intervention Simulations of Strengths Versus Weaknesses

This chapter describes the work subsequently published in Alireza, Fedor and Thomas (2017), which presents the use of a computational model of development based on an ANN, to investigate the mechanisms underlying language disorders and to explore the effects of specific types of behavioural intervention. The model builds on the training sets and architectures explored in the previous chapter and utilises computational modelling as a means to narrow the gap between theories of deficits and clinical practice, because it can be used to determine and test causes of developmental deficits and then evaluate possible interventions (Thomas, Fedor, Davis, Yang, Alireza, Charman, Masterson & Best, 2019).

A principal question that therapists must ask themselves when choosing an intervention programme for a child with DLD is whether to target remediating an area of weakness or whether to build on the child's strengths (Bishop, Nation & Patterson, 2014), perhaps establishing compensatory pathways or strategies (Leonard, 2014). In the case of WFD, the question may be whether to work on elaborating the semantic aspect of words (e.g., McGregor & Leonard, 1989) or whether to focus on the phonological component of word-finding (e.g., Best, 2005). Both semantic and phonological therapies have been found to improve WFD to a certain extent (Best, 2005; Best et al., 2015; Bragard, Schelstraete, Snyers, James, 2012; Easton, Sheach & Easton, 1997), although in these studies the treatment mainly improved retrieval of words used in therapy and the effect did not generalise to words not included in the therapy sessions.

One explanation for the tendency of therapists to work less on areas of weaknesses and more often on areas of strengths is to raise the child's confidence in the field where he or she is struggling. Moreover, from a theoretical perspective, there are at least three ways that improving a strength could remediate a behavioural impairment. First, the 'strength' could represent an alternative cognitive system or pathway to deliver a similar behavioural outcome. Improving a strength could then be depicted as encouraging a compensatory strategy. Second, the target behaviour may be delivered by an interactive system in which multiple sources of knowledge combine to drive behaviour. This stronger input from one source might make up for a weaker input from another. Third, the target behaviour may require mappings to be learned between input and output representations (such as the link between semantic representations and phonological representations in naming). In the face of a weakness in the input representations, improving the structure of the (otherwise typically developing) output representations might facilitate learning those mappings (or, correspondingly, strengthening typically developing inputs may make up for a weakness in output representations). The current study used computational modelling to investigate the third option. In this regard, the goal of the model was to force more detailed specification of the theoretical proposal that intervening to buttress a strength can compensate for a developmental weakness. We begin by briefly recapping previous developmental models of naming and models of interventions that motivated this work.

Models of Word Production

Most models of adult lexical retrieval depict it as a multi-stage process with distinct components and the consensus is that the process comprises three main areas: 'Conceptualization', 'formulation' and 'encoding for articulation' (Friedmann, Biran & Dotan, 2013). 'Conceptualization' involves deciding what to say and is sometimes referred to as the message level of representation. 'Formulation' involves two parts selecting the word that we want to say and then selecting the appropriate syntax as required to put into a sentence. Finally, the process of 'encoding' involves turning the word into a phonological sound using the appropriate syntactic form. Connectionist models of word production tend to focus on this last stage, which involves retrieving the phonological forms given a semantic concept of the desired lexical item. Developmental models have tended to simulate the learning of mappings between pre-specified semantic and phonological codes (e.g., Plunkett et al., 1992) or between semantic and phonological representations emerging in self-organising maps (e.g., Li, Zhao & MacWhinney, 2007; Mayor & Plunkett, 2010).

Computational Modelling of Interventions

The Best et al. (2015) model of naming was based on an autoassociator with a backpropagation learning algorithm (Rumelhart, Hinton & Williams, 1986) that simulated learning associations between phonological and semantic representations. The hidden unit representations of the semantic component of the model were mapped to the hidden units of the phonological component, via an intermediate layer of hidden units, to provide a pathway for the development of naming. A reverse pathway simulated comprehension. Thus, naming behaviour began to emerge while the semantic and phonological representations were themselves still developing. Atypical naming development, which was observed in a sample of children with WFD, was captured by changing the start state of some parameters; for example, by reducing the number of hidden units or the learning rate in the semantic component, the phonological component, or the pathway between them.

The model used in this study was an extension of the one described in the Best et al., (2015) paper, which utilised the LIRA architecture discussed previously (see Figure 7.4). The model architecture was adapted to incorporate the more ecologically valid semantics, phonology, and training corpus constructed in the previous chapter. Part of this project was to consider how different deficits to the model may produce different patterns of deficits, and to see what kinds of interventions best remediate different patterns of deficit.

The Best et al. (2015) model used an abstract representation of semantics and phonology and a training set of only 100 items. The model in this study used the V&V 397 binary semantics training set, described in the previous chapter, which is made up of more realistic semantic and phonological representations. The typically developing model was designed in such a way that it reflected salient properties of vocabulary growth, including a comprehension-production asymmetry and a vocabulary explosion / exponential growth in vocabulary size (e.g., McMurray, 2007). Three assumptions were included in the extended version of the Best al. (2015) model that were not in the original version. First, phonological representations were required to be more accurate than semantic outputs to drive a behavioural response. The reasoning behind this assumption was that phonological output needs to drive motor assemblies, while semantic comprehension only requires that the output fall in the correct attractor basin for the category (Hinton & Shallice, 1991) thereby generating the production-comprehension asymmetry. Second, a sensitive period was implemented in the development of the components but not the pathways in the model, through pruning[3] of network connectivity after a given point in development. This followed the

[3] Pruning for this model started after an onset epoch. Thereafter, for each epoch, every connection was tested to evaluate its magnitude. If it was less than the threshold, it had a certain probability of being removed permanently. Pruning in small connections is an optimisation procedure. Therefore in this case, it had three parameters, onset, threshold and probability.

approach used by Thomas et al. (2019) in their population-level modelling of interventions. The availability of rich connections created the potential for early training to create enriched lower-level representations. This pruning was not executed on pathways, under the view that sensitive periods are characteristic of lower but not higher cognitive systems (Takesian & Hensch, 2013). This allowed the effect of timing of intervention to be subsequently evaluated. Third, by using the learning rate parameter, plasticity was set higher in the pathways than the components; thereby the development of semantic and phonological representations would be the limiting factor on the development of naming. The logic was that interventions targeting phonology and semantics would have no scope to improve naming performance if the semantic and phonological representations were to quickly reach ceiling before naming had developed. The effects of both the second and third assumptions were evaluated by also running the model in their absence.

Overall Simulation Design

A typically developing model was created, followed by a version that exhibited a developmental impairment in naming generated by restricting the number of hidden units in the naming pathway. Five different interventions to remediate the naming deficit were then compared, which were of three types:

(i) Remediating the weakness where the model was provided with additional training on the naming pathway;

(ii) Improving the strength where the model was provided with additional training to improve the semantic representations, the phonological representations, or both at the same time;

(iii) Both types (i) and (ii) were combined into an intervention that sought to simultaneously improve strength and remediate weakness.

The immediate effect of intervention was explored, in terms of potentially accelerated vocabulary growth, and the eventual outcome, in terms of the largest vocabulary size achieved following each type of intervention. Finally, simulations of the five types of intervention were performed on a typically developing system as well as a delayed system to explore whether the interventions would have the potential to enhance the performance of a typically developing system or just the performance of systems exhibiting delayed development.

Model Architecture and Simulation Detail

Architecture: The architecture of the vocabulary development model used was similar to the linked internal representations architecture (LIRA) shown in Figure 7.4 in the previous chapter. It had a semantic component, a phonological component and four linked back-propagation networks. The semantics component was made up of a 3-layer autoassociator with 1029 input and output units and 45 hidden units. The phonology component was an autoassociator with 456 input and output units and 60 hidden units. The naming pathway linked the semantic hidden units with the phonological hidden units via an intermediate layer of 175 units (a value established by piloting). Naming constituted activating semantic inputs and measuring phonological outputs. The comprehension pathway ran in the other direction and also contained 175 units. In the atypical model, the number of hidden units in the naming pathway was reduced to 90 prior to training.

Additional parameters: The learning rate in the semantic component was .015 and in the phonological component was .025, values chosen to match their developmental rate. In the pathways, the learning rate was .15, per the assumption outlined previously. Sigmoid activation functions had a temperature of 1.5 in the components and 1 in the pathways, again to support faster development in the components. In the components, after epoch 75, any connection weights with an absolute magnitude of less than .5 had a 5% chance of being permanently removed each epoch, to produce developmental reductions in plasticity. Initial weights were given random values via a Gaussian distribution with mean 0 and standard deviation 0.5, which are default values. Gaussian noise with a standard deviation of .15 was added to the net input of units in the components, and noise with a standard deviation of .05 in the pathways, to provide a stochastic basis for naming errors in normal functioning. Continuous activation values on the phonological output were converted to responses by finding the nearest legal phoneme in each slot (i.e., comparing the output to each possible phoneme and seeing which it was closest to) and assessing whether the full phoneme string was then the correct name. The average root mean square error was used as the measure of distance between the activation vector for each phoneme and the nearest legal phoneme code. If it exceeded 0.03 (a value calibrated against the level of noise), that phoneme was coded as no response. A nearest neighbour technique was likewise used to assess the accuracy of semantic outputs.

Training Set: The V&V binary semantics training set, described in chapter 7, was employed for the simulations in this study. The training set was made up of 397 words, each had a separate set of phonological representations and a set of semantic representations. The total number of available phonological features was 456 features. There were a total number of 1029 semantic features available.

Training schedule: Networks were trained for 1000 epochs, with random presentation and pattern update. Training of autoassociators and pathways was interleaved. Weights were updated using the backpropagation algorithm with the cross-entropy error measure.

Simulation of interventions: For atypical networks, the intervention was limited in duration. It began at 100 epochs of training and lasted for 100 epochs. For the main condition, at this point TD models had acquired a productive vocabulary size of 67 words, while atypical models had a vocabulary size of 36 words. For the intervention, one or more components or pathways were trained with 5 times the frequency of the rest of the system. The extra training could be on the semantics-phonology naming pathway, the semantics component alone, the phonological component alone, both semantics and phonological components, or all of these combined.

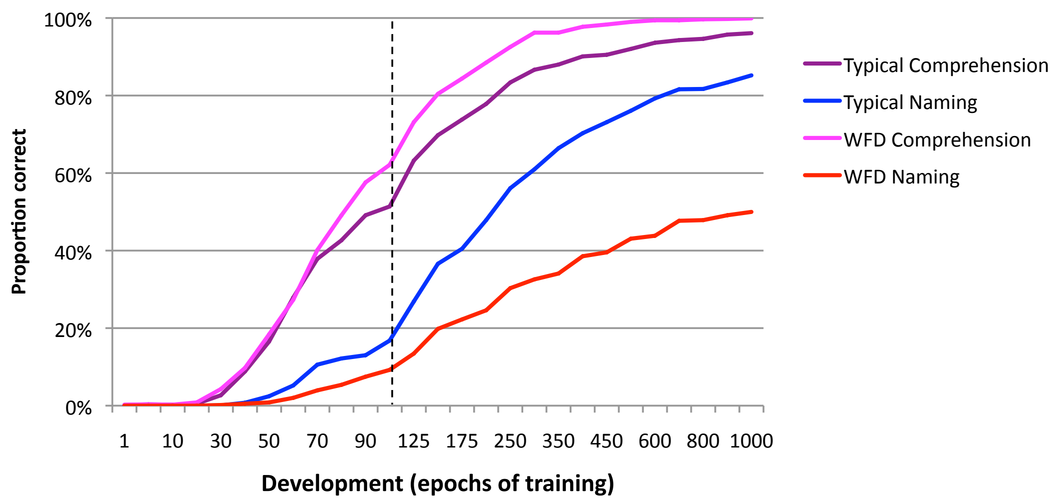
Conditions: The effect of applying intervention at different points in training was investigated. Interventions were compared at 100, 250 and 750 epochs to test the importance of timing. To test the effect of plasticity assumptions in the model, the first variant removed connectivity pruning from the components. The second variant removed the higher plasticity of the pathways, setting their value to .025.

Replications: All conditions were replicated three times with different random seeds. The full design took approximately 100 days of simulation time. Simulation time was relatively long because of the different conditions of interventions, which included simulating the different types of intervention, different timing of intervention and different plasticity conditions. Results graphs are shown in the results section averaged over three replications. The corresponding results graphs of each of the three replications are shown in appendix 6. The effects of timing and plasticity are shown in Tables 8.1 and 8.2.

Results

The developmental trajectories for naming and comprehension in the typical and atypical models are shown in Figure 8.1. For the typical model naming lagged behind comprehension presenting the expected comprehension-production asymmetry. The model was consistent with a vocabulary explosion as naming itself showed an accelerating rate of development. The atypical model with WFD trajectory presented delayed development in naming but showed good development of comprehension. The slightly better development of comprehension than the typical model was seen as a stochastic difference. The dotted line represents where the intervention would take place and, by then, the size of the vocabulary for the simulated children with WFD was about half that of the typically developing model.

Figure 8.1. Developmental trajectories of naming and comprehension for TD and WFD model simulations. The dotting line depicts the point at which the intervention was applied.



The graph in Figure 8.2 shows the effects of remediating weakness (referred to as Naming), which took the form of more practice on naming compared to the impaired model without intervention. The trajectories demonstrate that the intervention produced accelerated development while it operated, but the advantage did not last and there was no gain in final productive vocabulary level. The green shade is where the intervention took place.

Figure 8.2. The effect of intervention to remediate weakness on naming accuracy. Period of intervention is marked by the shaded region.

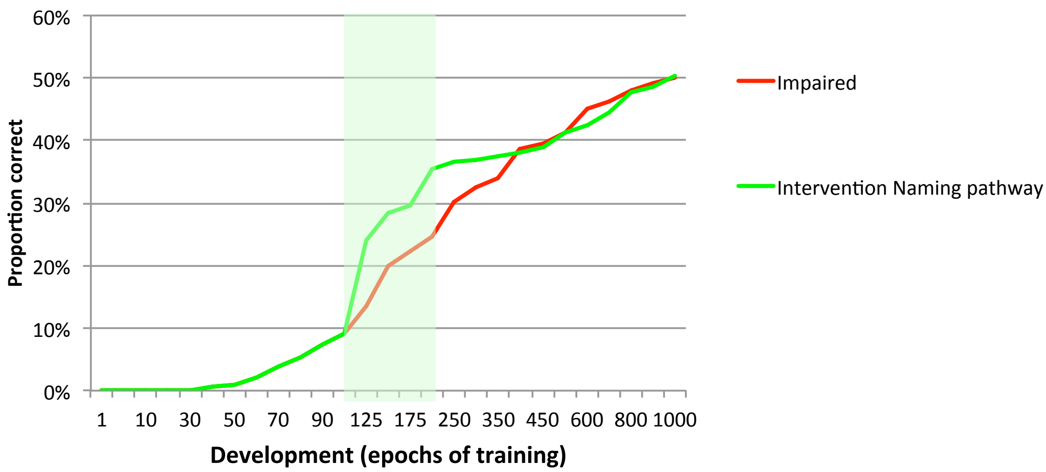
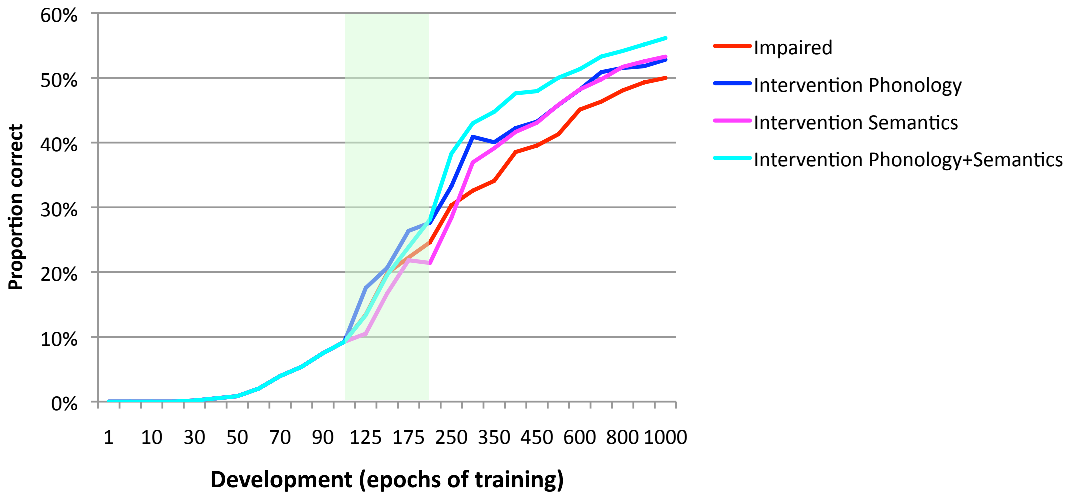
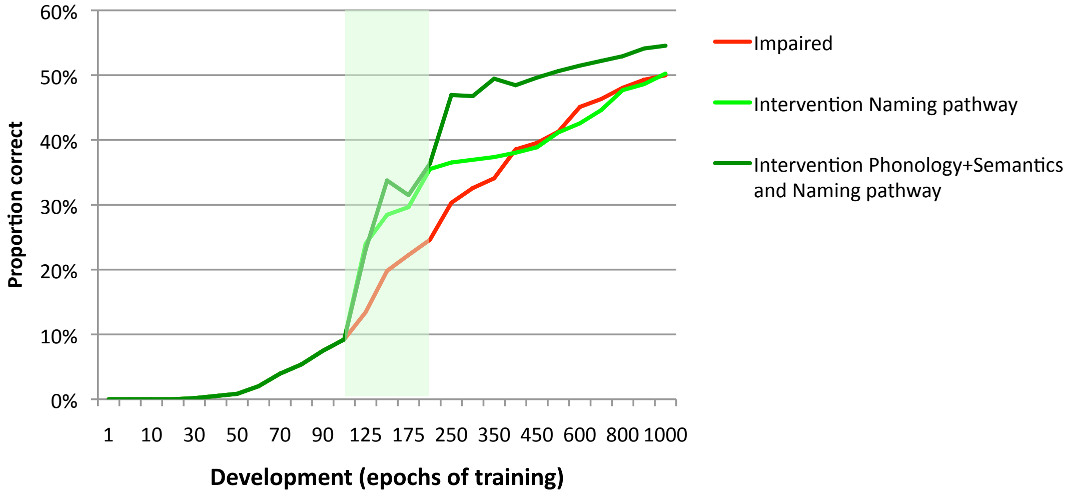


Figure 8.3. The effect on naming accuracy of intervention to improve strengths. Period of intervention is marked by the shaded region.



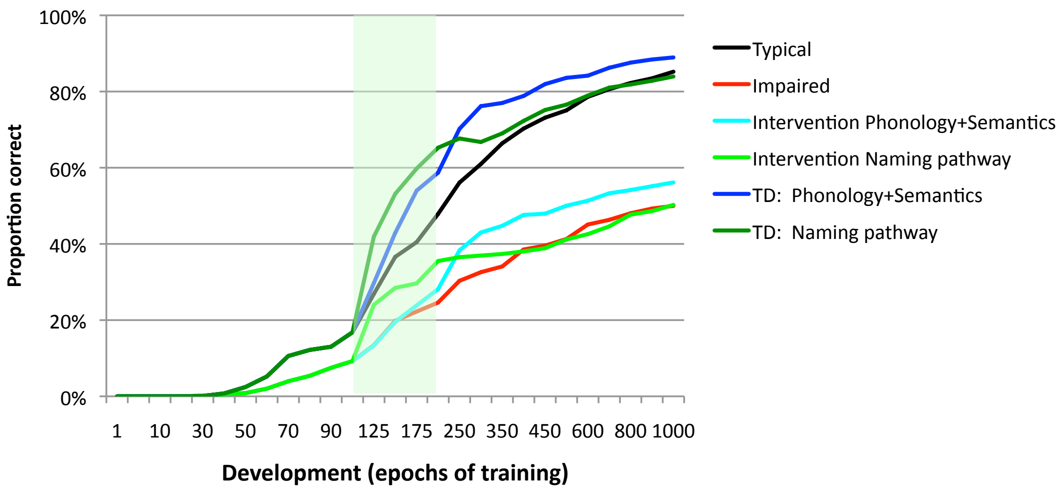
The graph in Figure 8.3 shows the effects on naming accuracy of intervention to improve strengths where the period of intervention is marked by the shaded region. Intervention took the form of extra training on the otherwise typically developing semantic and phonological representations (referred to as S+P), which were the inputs and outputs of the naming pathway. These interventions produced improvement in naming accuracy although at a slower rate than interventions on remediating weakness during the intervention itself. However, interventions to improve strengths led to an average improvement in naming accuracy by 5% (equivalent to about 10 words out of the training set of 397 words) which is a better outcome in the final productive vocabulary level than the outcome for interventions to remediate weakness.

Figure 8.4. The effects of both interventions to remediate weakness and improve strengths.



The graph in Figure 8.4 presents the combined effect of interventions to remediate weakness and improve strengths (Both). The combined interventions showed the initial immediate improvement from the remediation intervention as well as the long-term improvement at the final level due to the strengths intervention.

Figure 8.5. Comparison of the effects of interventions to remediate weakness and improve strengths on TD and WFD models.



The graph in Figure 8.5 shows that the interventions had the same effects on the development of a TD model as on the development of a model of a child with WFD. Development of naming was accelerated with extra practice on naming but the final level was not raised; extra elaboration of semantic and phonological representations increased the final productive vocabulary size for both TD and WFD networks.

Results of applying intervention at different points in training and testing the assumption of plasticity are shown in Tables 8.1 and 8.2 below. The tables show the final levels obtained from three replications of interventions testing weakness (Naming), strengths (S+P) and both weakness and strength (Both) and their average outcome.

Table 8.1: Naming accuracy at the end of training for typical (TYP), atypical (ATYP), and atypical intervened networks showing effects of timing of intervention on final level of performance: (a) Phonological + Semantic intervention, (b) Naming intervention, (c) Combined intervention at 100, 250, and 750 epochs on the atypical network. Each intervention lasted 100 epochs.

(a) Phonological + Semantic intervention

	TYP	ATYP	100	250	750
R1	85%	50%	57%	53%	50%
R2	85%	51%	57%	54%	52%
R3	87%	51%	54%	55%	52%
Avg	86%	51%	56%	54%	51%

(b) Naming intervention

	TYP	ATYP	100	250	750
R1	85%	50%	50%	50%	53%
R2	85%	51%	50%	51%	55%
R3	87%	51%	51%	51%	55%
Avg	86%	51%	50%	51%	54%

(c) Combined intervention

	TYP	ATYP	100	250	750
R1	85%	50%	52%	54%	52%
R2	85%	51%	57%	55%	55%
R3	87%	51%	55%	55%	55%
Avg	86%	51%	55%	55%	54%

From Table 8.1 we can see that the strengths intervention (S+P) became less effective the later in development it was applied. The weakness intervention (Naming) showed the opposite pattern, increasing in size the later it was applied. The combined intervention (Both) showed a uniform effect across development. Within each condition, the three replications demonstrated a common profile.

Table 8.2: Naming accuracy at the end of training for typical (TYP), atypical (ATYP), and atypical intervened networks: (a) the base condition; (b) removing plasticity reduction in the semantic and phonological components; (c) removing greater plasticity in pathways. S+P = strengths intervention. Naming = weakness intervention. Both = combined. Three replications and average are shown.

(a) Naming accuracy at the end of training

	TYP	ATYP	Intervention		
			S + P	Naming	Both
R1	84%	49%	57%	50%	52%
R2	86%	52%	57%	50%	57%
R3	86%	49%	54%	51%	55%
Avg	85%	50%	56%	50%	55%

(b) Without plasticity reduction in S and P components

	TYP	ATYP	Intervention		
			S + P	Naming	Both
R1	94%	49%	54%	48%	51%
R2	95%	52%	55%	51%	53%
R3	95%	51%	55%	50%	53%
Avg	95%	51%	55%	50%	53%

(c) Equalized plasticity in pathways and components

	TYP	ATYP	Intervention		
			S + P	Naming	Both
R1	84%	55%	61%	54%	60%
R2	81%	55%	64%	53%	60%
R3	82%	54%	60%	50%	59%
Avg	82%	55%	61%	52%	60%

Table 8.2(a) contains the data for the base condition; 8.2(b) demonstrates that when plasticity reductions in the components were removed as a model assumption, the same pattern of results was produced. Table 8.2(c) shows that without the assumption of greater plasticity in the pathways, the same pattern also held. None of these assumptions appeared to be key to the behaviour of the model.

Discussion

In this study, an artificial neural network model that simulated typical and atypical vocabulary development was used to compare the effects of intervention that was targeted to remediate a weakness versus one that was targeted to bolster a strength. Remediating weakness required more practice on naming. This produced immediate improvements in the naming pathway but did not raise the final level of vocabulary attained by the system. Intervention had progressed learning faster along the *same atypical trajectory*, because the reduced computational capacity of the naming pathway had constrained the ceiling level of performance of the model. On the other hand, improving strengths, which involved greater elaboration of semantic and phonological representations, lifted the ultimate level of proficiency, producing long-term gains in the size of the productive vocabulary although initially at a slower pace. This is because the intervention had improved either the semantic representation or the phonological representation, slowly nudging the model towards a more typical trajectory and long-term gains in vocabulary level. The largest gains were produced when intervention improved both semantic and phonological representations. These gains were produced because semantic and phonological representations became more defined (that is, less confusable) as a result of additional training. The target behaviour of the model required mappings to be learned between representations. Intervention improved the structure of the representations serving to make learning the mappings easier for a restricted pathway. Hence higher accuracy could be achieved by a pathway with limited capacity. The best outcome occurred when intervention on weakness and strengths were combined leading to both immediate gains during the intervention period and a long-term improvement in the ultimate vocabulary size that could be attained.

Investigations of introducing the interventions at different points in time showed that improving strengths produced greatest gains early in development. This is because the benefit of improved input and output representations relies on those supporting a subsequent phase of naming development. As has been shown in models of reading development, fixing input or output representations after input-output representations has been learned has diminishing effects (Harm, McCandliss & Seidenberg, 2003). By contrast, remediating weakness produced the greatest gains late in development. This is because gains in forcing the naming pathway dissipate across development, and late interventions have less time to dissipate. Combining both interventions on weakness and strengths delivered the best outcome as a uniform effect across development. Results that showed greater gains were obtained from introducing the strengths intervention (elaboration of semantic representation, phonological representation or both) early in development are also compatible with the view that perceptual categorisation occurs early in development and this in turn affects perceived similarity of objects (Westerman and Mareschal, 2014).

The model was designed to have sensitive periods in the components so that lower plasticity in the components than the pathways would be a limiting factor in naming development. However, these assumptions were found not to be crucial, as lowering the plasticity of the component did not affect the outcome.

Linking Back to the Empirical Data. Overall, this model demonstrated how intervention that is targeted at a strength can be a contributing factor to improving behavioural outcomes in developmental disorders, an effect that has not previously been shown in a mechanistic computational model of development.

These simulated intervention effects which present modest improvements in the putative semantic and phonological representations are in line with those observed empirically for children with WFD as shown in the regression analysis of the secondary data taken from Best et al., (2021) in chapter 6 in this study as well as previous research (Best, 2005; McGregor & Leonard, 1989). Although the regression analyses in chapter 6 looked at the differences in scores of naming of the children with WFD as a whole group, their individual scores and sub-grouping information were also available for analysis. In the Best et al. (2021) study, children were assigned to three subgroups according to their performance on two tasks, one that tapped into their phonological skills and another assessing their semantic skills. Each child was thus identified to have a language profile which fell in one of the following subgroups: (i) Primarily semantic difficulties; (ii) primarily phonological difficulties in the context of strong semantic processing; and (iii) classic WFD, which referred to children diagnosed with WFD who did not show a more dominant type of difficulty in terms of phonological versus semantic skills.

Taking one example of a child with predominantly phonological difficulties and another example of a child with predominantly semantic difficulties Figures 8.6 and 8.7 demonstrate the effects of therapy on the naming scores of each child on each of the four subgroups of 25 words (semantic therapy, phonological therapy, naming controls and unseen). A detailed account of the procedure for the intervention was given in chapter 6. In each of these two cases the first therapy was targeted at the child's weakness (the child whose scores are in Figure 8.6 had primarily phonological difficulties and the child whose scores are in Figure 8.7 had primarily semantic difficulties) and the second therapy was targeted at the child's strength. In each case the therapy was only implemented on the subgroup of 25 words assigned to the named therapy. In both cases, immediately after the therapy targeting the weakness of the child, the scores of the subgroup of words improved. Naming scores improved by 15% in the case of a phonological difficulty (Figure 8.6) and by 15% in the case of a semantic difficulty (Figure 8.7), but six weeks later the scores had declined considerably in both cases. However, the important point to note here is that in each case after the second therapy had finished it was found that scores of the subgroup of words that were included in the first therapy only (which was targeted at the child's weakness) and had not been included in the second therapy (which was targeted at the child's strength in each case) also improved.

A similar trend was observed in the case of classic WFD subgroup as shown by the two examples in Figures 8.8 and 8.9. In these two cases although no dominant difficulty was identified but in each case after the first therapy was applied scores improved markedly for the word set of the relative therapy. Naming scores improved by 30% in the case of phonological therapy given first (Figure 8.8) and by 20% in the case of semantic therapy given first (Figure 8.9). The important point to note here again is that in each case after the second therapy had finished it was found that scores of the subgroup of words that were included in the first therapy only and had not been included in the second therapy also improved. Hence, similar to the simulations in this study, the line profiles in Figures 8.6 to 8.9 demonstrate the importance of targeting intervention on both the weaknesses and strengths of the child to maintain any progress made.

Figure 8.6. Naming accuracy scores of a child with WFD with primarily phonological difficulties showing changes in score after each therapy.

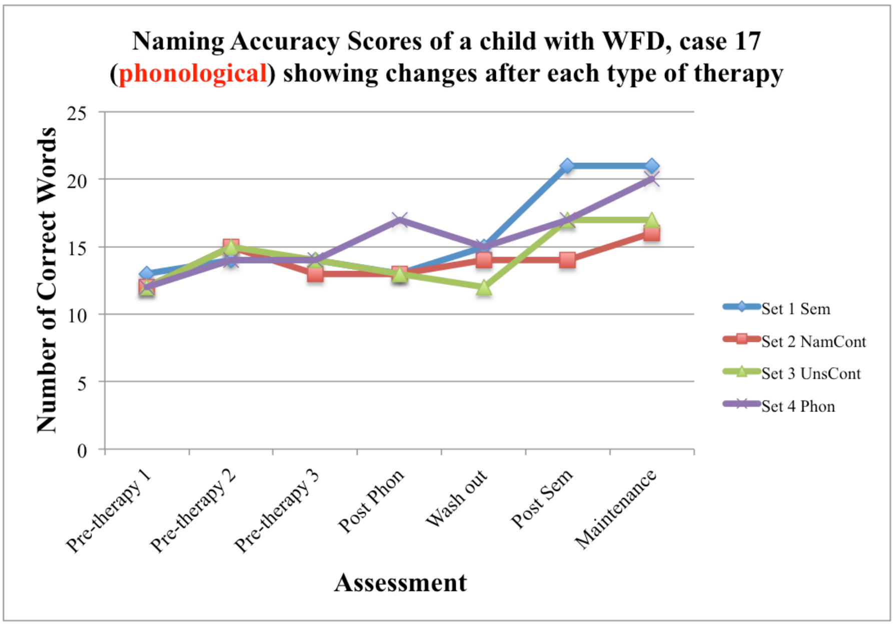


Figure 8.7. Naming accuracy scores of a child with WFD with primarily semantic difficulties showing changes in score after each therapy.

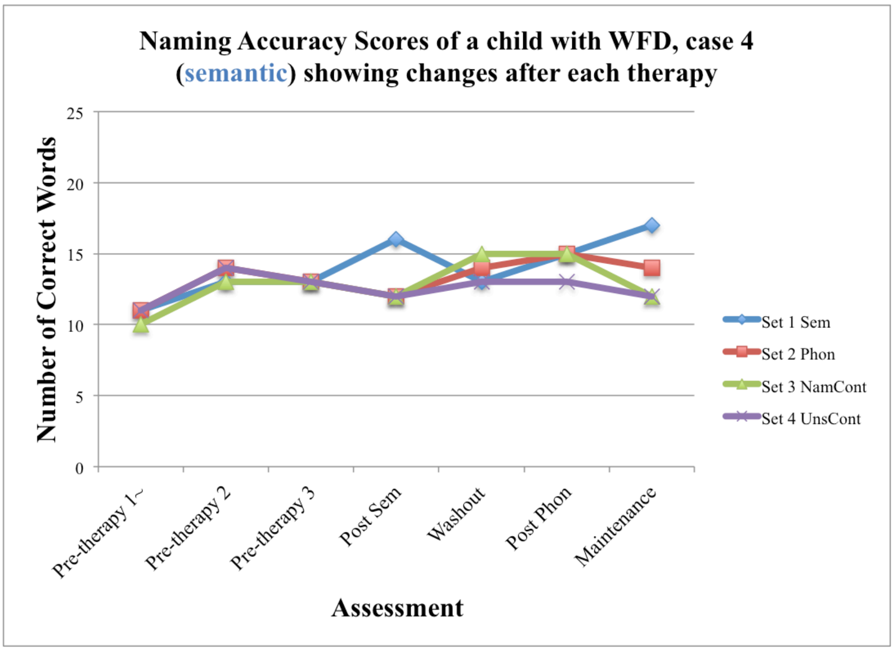


Figure 8.8. Naming accuracy scores of a child with WFD in the classic WFD subgroup showing changes in score after each therapy (Phonological therapy first).

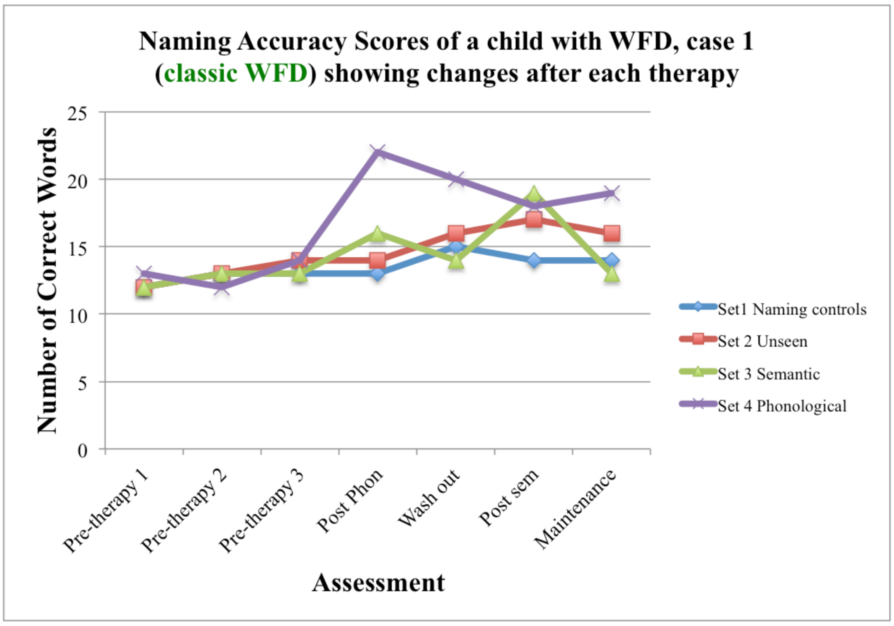
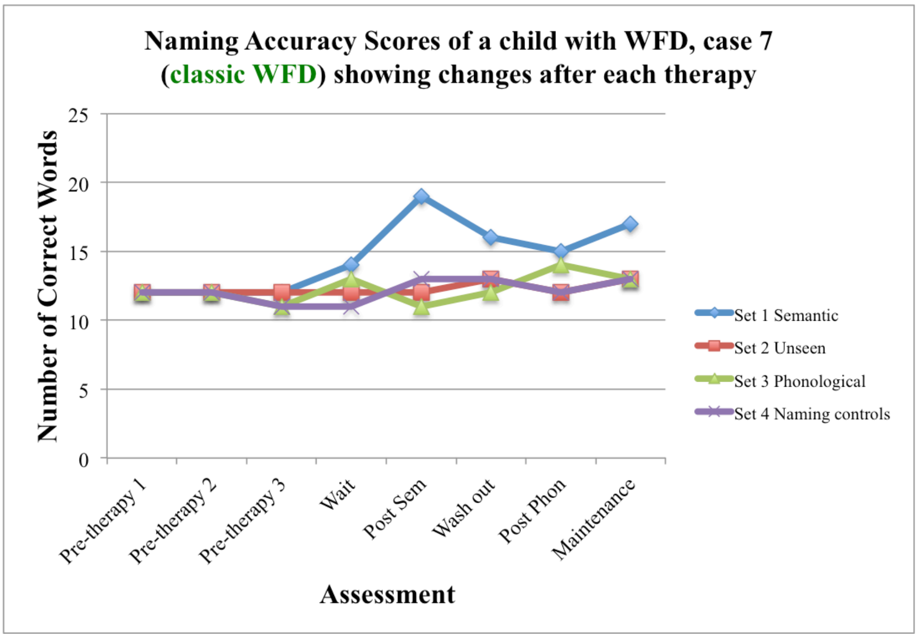


Figure 8.9. Naming accuracy scores of a child with WFD in the classic WFD subgroup showing changes in score after each therapy (semantic therapy first).



The intervention profiles of all 20 children, arranged by subgroup, are available in Appendix 7. The figures for naming accuracy showed that out of the 7 cases where intervention targeted weakness first then strengths, 3 cases were compatible with the notion of improvement in the first set of words (weakness) immediately after the second therapy (strength) ended (cases 4, 5 and 17); one case showed improvement on the first therapy set of words (weakness) 6 weeks after the second (strength) therapy ended (case 8); and 2 cases showed a relatively small drop in the improved scores of the first therapy (weakness) immediately after the second therapy ended (strength) (cases 2 and 16). The remaining case (case 6) does not show improvement on the word set of the weakness therapy but does show improvement in the NC and US sets both of which were not used in the therapy. Two cases were given therapy targeted at a strength first and then a weakness (cases 11 and 12) and only one of them showed no drop in improvement on the first set of words (strength) after the second therapy (on weakness) ended.

The figures show that out of the 11 cases in the classic WFD subgroup, 6 cases improved for a second time in the first set of words either immediately after the second therapy (case 9) or 6 weeks later (1, 7, 3, 10, 14). In 4 other cases it was noted that where there was no improvement in the first set after therapy was finished on the second set, but at least the drop in gain was relatively small or no drop was shown (cases 15, 19, 20, 18).

The detailed scores of the 7 assessments, on naming accuracy of 100 items, given to each of 20 children at different stages of the therapy were provided in Appendix 1 of the Best et al., (2021) paper. Table 8.3, below, presents an overview of the percentage in improvement in total scores between average baseline scores before therapy and scores after the first therapy, and between average baseline scores before therapy and scores at maintenance, for children in each of the different subgroups of the 20 cases. The table, furthermore, presents the results of the subgroups of children according to whether intervention was targeted at a weakness or strength first for the children whose predominant difficulties were identified. In the case of classic WFD the table shows whether semantic or phonological intervention was given first. All but one case show that improvement in scores is maintained six weeks after the last therapy ended. These results are consistent with the notion of targeting intervention on both weaknesses and strengths of the child to maintain improvement effects.

Table 8.3. Percentage of improvement shown on average for children in the different subgroups between their average baseline scores before therapy and their scores after the first therapy, and between their average baseline scores before therapy and their scores at maintenance. All comparisons were done on the total scores of the assessments on naming accuracy of 100 items for each child.

Subgroup	Number of cases	Average % of improvement in total scores shown by children after first therapy	Average % of improvement in total scores shown by children between maintenance and pretherapy
Semantic Difficulties: Strength intervened on first	1	36	38
Semantic Difficulties: Weakness intervened on first	5	13.8	28.2
Phonological Difficulties: Strength intervened on first	1	18	-4
Phonological Difficulties: Weakness intervened on first	2	1	24
Classic WFD: Semantic Therapy first	4	21	24
Classic WFD: Phonological Therapy first	7	11	20

Returning to theoretical issues, we identified in the introduction at least two more methods by which improving strengths might improve behavioural outcomes. One method is to encourage compensatory mechanisms through intervention. In this method the ‘strength’ could represent an alternative cognitive system or pathway to deliver a similar behavioural outcome and intervention can encourage this. Another method is bolstering convergent sources of information in interactive systems. Investigation of either would involve a more complex architecture than that used here, one that in principle provided the scope for compensatory pathways or convergent interaction of information sources, such as the Seidenberg and McClelland 'triangle' model of reading (Harm & Seidenberg, 1999; Plaut et al., 1996; Seidenberg & McClelland, 1989).

Model Mechanism in relation to WFD. The neural network model used in this study is primarily a connectionist cognitive model based on the notion that processing is represented by patterns of activation across simple processing units, which are connected together forming complex networks (Thomas & McClelland, 2023). Knowledge is thus considered to be stored in the strength of connections between processing units (Thomas & McClelland, 2023). In this study WFD was simulated by reducing the number of hidden units in the naming pathway (see Figure 7.4 in chapter 7) thereby reducing the capacity of a single system. WFD is thought to be heterogeneous in its cause but in this study only one possible cause was examined. In this respect effects of the model in terms of the neurocognitive mechanisms underlying WFD are in line with the neuroanatomically-constrained dual-pathway language model as proposed by Ueno, Saito, Rogers and Lambon Ralph (2011). In their paper the authors specifically proposed 'that single-word comprehension, production (speaking/naming) and repetition are supported by the interactive contributions of the dorsal and ventral language pathways' (Ueno et al., 2011, p. 385). Their findings proposed that the ventral pathway was associated with semantic input and output and the dorsal pathway was associated with acoustic – phonology input and motor – phonology output as illustrated in Figure 8.10. From this perspective our model could map onto their neuroanatomically –constrained model where the naming pathway from our model represents the ventral pathway (primary auditory ⇔ mid-STG ⇔ anterior-STG ⇔ opercularis-triangularis ⇔ insular-motor cortex: supported by the middle longitudinal fasciculus and the extreme capsule) and incorporating the ventral anterior temporal region (vATL) via a connection to the aSTG. The comprehension pathway of our model could then represent the Dorsal pathway (primary auditory ⇔ inferior supramarginal gyrus ⇔ insular-motor cortex: supported by the arcuate fasciculus) of the Ueno et al. model (Ueno et al. 2011, p. 386). Reducing the number of hidden units in the naming pathway of our model could thus be simulating a reduced number of neurons in the regions connecting vATL and IMC or it could be simulating a reduction in the white matter of the axons connecting the regions in the same ventral pathway. A possible explanation for the simulated effects of intervention could be that improving the structure of the semantic and phonological representations and increasing the number of practice repetitions has an effect of increasing myelination (white matter in the brain) (Fields, 2010) allowing higher accuracy to be achieved by the same pathway.

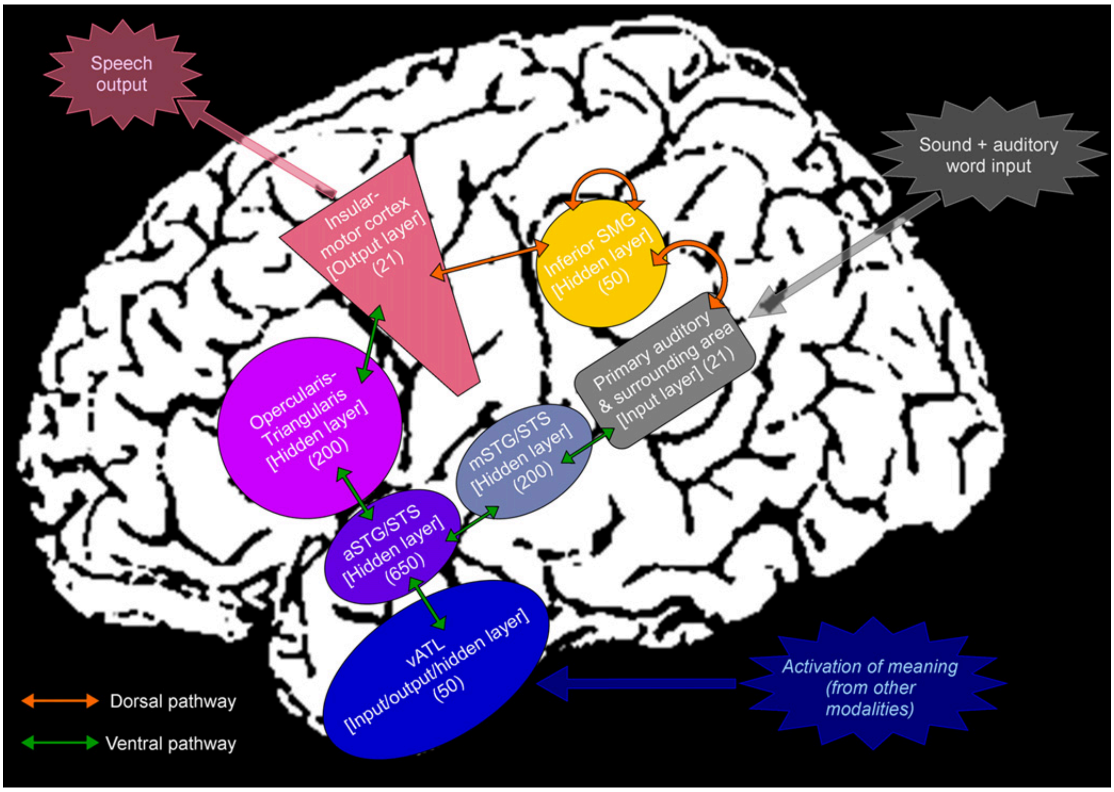


Figure 8.10. Lichtheim 2—the Implemented Neuroanatomically-Constrained Dual-Pathway Language Model [reproduced from Ueno, Saito, Rogers & Lambon Ralph, 2011]

The limited improvements in vocabulary of the model in this study and the notion that the deficit was not eliminated (it was reduced from 35% to 29%) are in line with the proposals of Thomas et al. (2019): based on a neurocomputational analysis of intervention effects, Thomas and colleagues argued that, some exceptions permitting, where deficits arise through neurocomputational constraints in developing systems, behavioural interventions alone are unlikely to completely ameliorate deficits. Of course, a major limitation to the model is that it simulates an isolated, single system. However, the outcomes in this study still demonstrate that computational modelling can make a useful contribution towards closing the gap between theories of deficits and clinical practice, as argued in this thesis. We now turn to summarise the overall arguments and findings of the thesis.

Chapter 9

Discussion

The main purpose of this research was to investigate the possibility of narrowing the gap between theories of deficits and clinical practice for developmental disorders. The focus was on the question of 'what works' in terms of intervention for developmental disorders of language and cognition. The strategy was to first investigate mechanisms behind one type of disorder, and then explore the most suitable methods of intervention, with a view to develop general principles for effective methods of intervention for developmental disorders. Taking word-finding difficulties (WFD) as an example disorder to be investigated, our study employed a unique approach of investigating ways to uncover the most appropriate interventions, which had several advantages. The research consisted of a combination of three methods of investigation: (i) Qualitative, in which a detailed questionnaire was distributed to therapists to try to elicit mechanistic causes of deficits from the reasoning behind their instinctive use of certain interventions in the different cases of developmental language disorder (DLD). (ii) Empirical, in which two sets of secondary data were analysed in search of more detailed explanations for intervention effects using regression models to achieve a continuous approach analysis. This approach seeks to assess predictors of the size of intervention effects across different intervention methods. (iii) Computational, in which developmental trajectories of core language skills were simulated for typically developing (TD) children, children with WFD, and change following intervention for WFD, including a realistic semantic training set. The computational modelling part of this research focused in particular on the question of whether intervention should be targeted at a weak skill that could potentially be the direct cause of the relevant difficulty or be targeted at another strong skill that could indirectly compensate for the difficulty that is manifested by the child (Alireza et al., 2017). In the context of WFD, the present study explored the question of whether intervention should be targeted towards improving access of semantic or phonological representations of a word when a corresponding picture is presented to the child. Moreover, the key question investigated in this study was whether the most effective intervention for each child can be predicted from their profile of core skills.

In terms of whether intervention should be targeted at strengths or weaknesses of the child, my computational modelling investigations indicated that the best long-term outcome was shown when intervention targeted both areas. Those addressing weakness had fast but transient effects, propelling the system more quickly along the same atypical trajectory; those addressing strengths had slower but longer lasting effects able to raise the final level of attainment. Analyses of the questionnaire results and empirical data were in line with the computational modelling results. However, these investigations also illustrated that, provided the number of participants is large enough, more research on the subject can potentially attain a set of principles by which one could predict the most effective interventions according to a child's language profile.

Overview of Qualitative Results

Overall findings from the questionnaire responses were consistent with previous research showing a variation in intervention strategies followed by speech and language therapists (SLT) when treating different children with DLD (Best, 2003; Law et al., 2008; Law et al., 2012; Roulstone, Wren, Bakopoulou & Lindsay, 2012). Moreover, the general picture emerging from the analysis indicates that outcomes of interventions for DLD rely on the interaction of multiple factors ranging from the environment, for example the budget of policy makers for dosage, to aspects that are relevant to the learning process, for example the child's motivation. This notion is in line with the framework proposed by Thomas, Ansari and Knowland (2018), which was inspired by a combination of Bronfenbrenner's ecological systems theory and principles from Michie and colleagues' (2011) behaviour change wheel method (see Figure 4.1). In their framework they identified four main factors as 'Government', 'Society and family', 'School' and 'Child factors', whose interdependent interactions constrained the outcome of learning interventions (Thomas, Ansari & Knowland, 2018). Their report included more specific aspects that come under each factor, for example, 'education budget' under governmental factors and 'teaching materials' under school factors, which would affect learning outcomes (Thomas, Ansari & Knowland, 2018).

Although the questionnaire responses were diverse for most questions, there were a few questions on topics which drew notably similar answers. This similarity draws attention to potentially important aspects of intervention such as common strategies, motivation and IQ effects, suggesting that more research should focus on them. For example, responses of this study revealed that the most common strategies used by SLT and teachers to support children with DLD were repetition and scaffolding. Repetition was generally used to reinforce a correct answer and scaffolding was used to avoid the chance of the child making a mistake, thereby building his or her confidence. The use of scaffolding to encourage errorless learning was likened to Hebbian learning (McClelland, Thomas, McCandliss & Fiez, 1999) and in terms of computational modelling, confidence can be modulated by measuring the learning rate of the system (Vallabha & McClelland, 2007). However, a study by McCandliss et al. (2002), described in chapter 4, in which a similar comparison between the use of scaffolding and the use of feedback was made, showed better results with the use of feedback than with the use of a type of acoustic scaffolding. This discrepancy in results brings into question whether the scaffolding approach, shown to be used by a majority of therapists, is indeed the most efficient approach and highlights the need for a more extensive evidence based research to support decision making on this question. Other popular strategies described by therapists as important contributors to the child's progress were being able to motivate the child by having background knowledge of his or her interests, and parents' involvement in the therapy. Interestingly, none of the therapists mentioned IQ in their responses, except when asked directly whether they considered it an important predictor for the child's prognosis. They all agreed that IQ was indeed important for predicting the child's prognosis, which implied that a child's language profile could potentially predict his or her response to therapy.

Overview of Empirical Data Analysis

Chapter 5 described the developmental trajectories constructed for four core language skills for typical and atypical language development using secondary empirical data taken with permission from Best et al. (2021). The trajectory analyses served to characterise the developmental characteristics of the language skills of the group of children with WFD compared to those with typical language development. The main purpose of constructing the trajectories was to seek to simulate similar trajectories for typical language development, then atypical language development, so that the effectiveness of various theory-based interventions could be investigated for different simulated cases of WFD. The four core tasks were: (i) *Picture Naming* (PN) to test word production; (ii) *Word-Picture Verification Task* (WPVT) to test comprehension;

(iii) *Picture Judgement* (PJ) task to test semantic knowledge of children; (iv) *Children's Non-word Repetition Task (*CNRep), to test phonology.

As expected from previous literature, results of the trajectories analyses for our sample data of TD children showed comprehension to be developing ahead of word production (Reznick & Goldfield, 1992; Plunkett, Sinha, Møller, & Strandsby, 1992). On comparing results of the trajectories analyses with the three predictions outlined in the Introduction of chapter 5, we found that two out of three predictions matched our expectations. Consistent with our first prediction, analyses of trajectories for the four core language tasks of our sample of cross-sectional data from children with WFD showed word production deficits worse than comprehension; however, this was shown to be an emerging effect over development. Evidence supporting our third prediction, that the WFD group would show a slower developing PN ability compared to WPVT when compared to the TD group, was also found on comparisons of the trajectories of the PN task and WPVT for TD and WFD groups. However, contrary to our second prediction, our sample of children with WFD did not show a bigger gap between the development of PN and WPVT abilities over time. This finding is less surprising if we consider that cross-sectional data can lead to a sampling bias, or it could be that our sample of 24 children with WFD was too small and/or did not represent a pure picture of WFD.

Chapter 6 presented secondary data analysis of intervention results (taken from the Best et al. (2021) study) to explore the relative effects of phonological and semantic interventions, given to 20 children with WFD in a within-participants design, based on each child's language profile. Consistent with previous studies, the results of this analysis confirm that the semantic intervention and the phonological intervention were both successful and had improved word-finding skills by a similar amount.

However, neither of them generalised to words not included in the therapy (Best, 2005; Best et al., 2021; Bragard, Schelstraete, Snyers, James, 2012; Easton, Sheach & Easton, 1997; Ebbels, 2014b). No significant correlations were shown between any of the four core skills (PN test of production, WPVT test of comprehension, the PJ test of semantics and CNRep test of phonology) and intervention effects. Hence Pearson's correlations did not point to supporting evidence for either semantic or phonological difficulties as principal limiting factors in performance in production that might be released by intervention. The most likely reason for the lack of correlation effects between intervention types and core skills is the small sample size of only 20 children, especially when taking into account the different language profiles of the participants and the heterogeneous nature of WFD (Best et al., 2021; Messer & Dockrell, 2006). Other reasons for the lack of correlation effects could be due to limitation in the predictor variables, for example ceiling effects in the case of the PJ task; or due to tasks not being sensitive enough to measure the underlying competencies, for example no correlation was observed between TWF-2 comprehension raw score and WPVT score, both putative measures of the ability to comprehend a single spoken word; or it could be explained by the notion that WFD difficulties are caused by a combination of problems in semantic representations, phonological representations and processing speed (see Messer & Dockrell, 2006). Nonetheless, a statistically significant correlation between improvement in scores of the phonological therapy (PT) word set and the TWF-2 comprehension raw scores was observed. This could be an indication that individuals with better comprehension benefitted from phonological therapy. On the other hand the significant correlation between the PT word set and TWF-2 could be an indication that TWF-2 was a reliable tool. This notion in turn suggests the need to assess the psychometric properties of all the core skills assessments in this study in order to check that they are indeed valid, reliable measures of the skills for which they are assumed to be measuring.

Long-term effects of each type of intervention (semantic versus phonological) and generalisation effects were investigated using regression models, which were detailed in chapter 6. Although for the first model, the regression analyses did not derive a significant model for identifying good predictors for the effects of both interventions between maintenance and pre-therapy, CNRep was found to be the most influential variable, showing that individuals with better phonological skills would benefit more from both interventions. Second, a model run to explore reliable predictors for the effects of semantic intervention on the semantic therapy (ST) word set between maintenance and pretherapy showed CNRep raw score to be a reliable predictor with a negative coefficient for PN, indicating that bigger effects of semantic intervention depended on worse PN but better CNRep. Third, a significant best-fit model to explore predictors for the effects of phonological intervention on the PT set between maintenance and pretherapy showed only TWF-2 comprehension raw score as a significant predictor, implying that phonological intervention could be more effective for children with better comprehension skills.

Regression results were discussed in detail at the end of chapter 6 and were examined from the perspective of intervention targeted at weaknesses versus strengths. Taken altogether, results of the regression analyses from this perspective indicated that interventions should be targeted at both weaknesses and strengths to gain best results (Alireza et al., 2017).

Results of the empirical data analysis were in line with previous research which found that most interventions were successful while therapy was going on but were not maintained after therapy had stopped (Ebbels, 2014b; Thomas et al., 2019). Likewise, in keeping with past literature and taking into account that the number of participants was relatively small, the heterogeneous character of WFD (Messer & Dockrell, 2006) could have been the reason for the predominant lack of predictive power between the core skills and effects of intervention.

Overview of Computational Modelling

The objective of the third strand of the project was to leverage computational modelling methods to explore the implications of targeting interventions towards weaknesses versus strengths in language development. One of the main advantages in this study was in developing a more realistic vocabulary corpus than the one employed in the Best et al. (2015) model in order to make a more ecologically valid version of the model. Chapter 7 first presented the development of five semantic feature sets, all based on real semantic features given by adult participants. This was followed by exploratory work comprising a comparison of two model architectures for an artificial neural network which were developed and preliminarily tested for the purpose of simulating interventions for developmental language disorders. The idea was to use a computational model to simulate typical, then atypical vocabulary development, followed by a series of simulated behavioural interventions to compare the outcome of targeting more practice on naming (weakness) versus more practice to improve semantic and phonological representations (strength) to remediate problems of WFD (Alireza et al., 2017). This study was described in chapter 8 and included investigations on the effects of timing of these interventions and the effects of altering assumptions on the developmental plasticity of the system.

Simulations to remediate weakness by additional practice in naming produced immediate improvements in naming. However, the level of improvement was not maintained after training had stopped and the outcome level of naming was similar to that of the atypical trajectory which had no intervention applied. By contrast, simulation of intervention targeting strengths, which took the form of more frequent training on the semantic component, the phonological component or both at the same time, produced improvements at a slower rate than intervention targeted at naming (weakness) but in this case, some improvement was maintained after intervention was stopped. The best outcome was when intervention was targeted at both weakness and strength. Investigations of timing of interventions showed that training on the weakness was more effective when applied at a later point in training, and intervention on strengths was more effective when applied at earlier points in training. The application of both interventions showed a uniform effect across development. Simulations to explore the effects of plasticity assumptions indicated that the same pattern held when plasticity reductions in the components were removed and when equalised plasticity in the pathways and components was in place.

Taken together, results of targeting intervention at naming at different points in time of training are in line with the outcome from models of reading development such as the Harm, McCandliss and Seidenberg (2003) model. The target behaviour of the model in the present study was to learn mappings between semantic and phonological representations. Intervention applied on weakness meant that more practice was made on naming using mappings of representations that had already been constrained, similar to the Harm et al., (2003) study which showed a diminishing effect of fixing input or output representations after input-output representations had been learned. Likewise in this study, when intervention was applied later in training, the outcome appeared better because there was less time for the effect to diminish between the end of the intervention and measuring the outcome. It follows that intervention targeted at more practice on defining semantic and phonological representations was more effective when applied earlier in the training session, so that improved input-output representations could be exploited, and the benefits were maintained because mappings were based on improved representations which supported more correct naming. These results, which show that introducing the strengths intervention (elaboration of semantic representation, phonological representation or both) early in development produce better outcomes, are also in line with the notion that perceptual categorisation occurs early in development and this in turn affects perceived similarity of objects (Westermann & Mareschal, 2014). The simulated effects of improving semantic and phonological representations that produce a better outcome are compatible with the empirical data results from the Best et al. (2021) study and other previous studies (Best, 2005; McGregor & Leonard, 1989).

Intervention on Weaknesses versus Strengths

At the end of chapter 8, we re-examined the effects of intervention on the individual cases of the empirical data drawn from the Best et al. (2021) study but now through the lens of intervention targeting weaknesses versus strengths (and in cases where a particular weakness was not identified, in terms of first and second therapy). For most cases when intervention targeted a weakness first, considerable gain was made on the words used in the therapy, but six weeks later there was a drop in scores. However, after the second therapy targeted a strength, an improvement in the scores of the set of words used only in the first therapy was noted. A similar trend was observed for some of the cases in the classic WFD subgroup in terms of first therapy versus second therapy (where weaknesses and strengths were not identified). The important point to note here is that in each case, after the second therapy had ended, it was found that scores of the subgroup of words that were included in the first therapy only and had not been included in the second therapy also improved. Chapter 8 presented four examples of graphs demonstrating the changes in naming accuracy scores after each type of therapy (cases 17, 4, 1 and 7 in Figures 8.6 to 8.9). Appendix 7 provides the graphs of all 20 cases arranged by subgroups. The point to note from these trend lines is that although in chapter 6 paired samples t-tests showed a significant improvement in mean scores of each of the four word sets between the average pre-therapy score and scores taken at maintenance, the figures depicting the change in naming accuracy scores show scores fluctuating according to the type and order of therapy rather than a steady improvement at different rates for the subgroups of scores in each case. In general, these findings are consistent with the notion of targeting intervention on both weaknesses and strengths of the child to maintain any progress made.

Implications for Narrowing the Gap Between Theories of Causes of Deficits and Clinical Practice

The focus of the three strands in this study has, in due course, been narrowed down to comparisons of targeting intervention on a child's weakness versus strength for the purpose of deriving preliminary practical implications for intervention on developmental disorders of language. From this respect, taking the results of the empirical data and computational modelling together suggest that alternating therapy between targeting weaknesses and strengths could maintain gains in scores for longer periods. This idea is also based on the conclusions from two studies. The first is the Tyler et al. (2003) study, which suggested that better results were obtained when therapy *alternated* weekly between a focus on phonology and morphosyntax. The second is the Bailey et al. (2020) study, which suggested that persistence depends on the types of skills being taught (e.g. *building skills* are more persistent), constraints and opportunities within the institutional and social context as well as compatibility between the interventions and the environment.

Furthermore, the computational model indicated that introducing intervention on elaboration of semantic and phonological representations early in development and introducing intervention on naming repetition at a later stage produced a better outcome. This would suggest that the most effective intervention for WFD could be in two stages. In the first stage, intervention on elaboration of semantic representations is suggested to alternate with elaboration on phonological representations, to enhance input and output representations in the language system. In the second stage, intervention on repetition of naming, to take advantage of the enhanced representations of meanings and word sounds, could potentially produce gains for longer periods. In line with the responses from SLT to the questionnaire as well as previous research, the recommendation would include parents' and teacher's need to be aware of the words which are being used in therapy in order to provide more opportunity for practice (Morgan et al., 2019; Thomas, Ansari & Knowland, 2018; Thomas et al., 2019).

This research has shown an added value in combining computational modelling with qualitative and quantitative empirical data for future studies on effective interventions for DLD. Similar to the example of targeting intervention to weaknesses or strengths, a suitable strategy may be to first use the questionnaire responses to draw out potential theories and build models on mechanistic causes of DLD based on the expectations of clinicians. Computational modelling can then be used to carry out preliminary investigations on effects of certain types of intervention, regimes of intervention, and timing effects, and make any necessary adjustments to the theory which can then be followed by empirical research. Evaluation of empirical data using regression analysis, as was done in this project, has more than one advantage. One is in treating intervention effects as continuous variables, endeavouring to link these to cognitive profiles, and thereby establishing developmental trajectories and relationships between skills, in a form that can be simulated by developmental computational modelling. A second advantage of the continuous approach of the regression analyses is that it offers the opportunity to examine the size of intervention effects and the differential effect between different types of intervention (e.g., in this study the differential effect between phonological and semantic therapy) based on profiles. A third advantage is that the regression analyses used in this study could potentially derive models with predictor variables that predict size of intervention effects according to the language profile of each child.

Limitations of the Study

A predominant limitation of this study is in the number of participants in both the empirical data analysis and the questionnaire responses. The apparent heterogeneous nature of WFD (Messer & Dockrell, 2006) magnifies the need for larger numbers of participants in the empirical data analysis in order to allow for larger subgroupings of participants. Time and resource limitations are pragmatic reasons why such intervention studies are often small scale. Nevertheless, larger participant numbers would in turn drive more reliable inferences to be made according to the effectiveness of interventions on the different subgroups. This is especially true for regression analysis as has been mentioned earlier, it is recommended to have at least 10 or 15 cases of data for each predictor in a regression model (Field, 2009). For example, by this rule of thumb, if four predictors were expected to be included in the regression equation, one would require at least 60 cases to make better predictions. The heterogeneous nature of WFD is most likely the very same reason for the variability in the majority of responses of the questionnaire. Analysis of a larger number of questionnaire responses may potentially derive patterns for therapists’ choices of intervention based on children’s behaviour that could not be derived from 10 participants.

Another limiting factor of this study is that the empirical data on which the core-skills trajectories analysis was based were cross-sectional. However, the intervention study part of this research was effectively longitudinal since assessments were taken at multiple points for the same individual and each child’s profile was known at the beginning of the study. The main goal in this study was to establish the extent to which the most effective intervention (phonological versus semantic) for each child can be predicted from their profile of core language skills. This would ideally need to be based on predictions from developmental trajectories of longitudinal data analysis rather than cross-sectional data (Thomas et al., 2009). The small number of 25 words in each of the four word sets used for assessments of intervention has also been identified as a limiting factor. Moreover, some of the results in this study bring to light that task selection for profiling interventions is rendered even more difficult when the data being analysed are cross-sectional (e.g., ceiling effect of PJ task and lack of correlation between TWF-2 comp and WPVT). Core skills would therefore need to be modified to tap into targeted language skills for future research.

Although there are advantages to employing computational modelling in this study to simulate trajectories from empirical data and to explore the effects of specific types of intervention, the single neural network is not complex enough to capture behaviour resulting from interactions within a real cognitive system (Thomas et al., 2019). Simplification is both a strength of formal modelling, because it provides mechanistic transparency to support theory construction, and a weakness, inasmuch as the phenomena under study are often complex. For example, behaviour of an actual cognitive system would be affected by sensorimotor components, emotional components, social components, executive function components, metacognition and motivation (Thomas et al., 2019). Effects of intervention would also depend to a large extent on outside factors which are difficult to measure, such as parent's involvement in reinforcing the therapy and the child's willingness to work on the target.

Future Directions

Inferences drawn from the three strands of this study suggest that there is potential in making progress towards finding principles of intervention by pursuing further research on the possibility of predicting the most effective intervention for each child with WFD from their profile of core language skills. Furthermore, in terms of computational modelling, the V&V 397 training sets and the LIRA associator model have the potential to be used for further research using the method of PCA and cluster analysis on emergent category structures (Rogers & McClelland, 2006) as described in chapter 7. This type of preliminary investigation may help uncover what mechanistic problems can occur when the model is trying to learn mappings between the codes developed in each component (semantic and phonological) during training simulations of atypical development.

The next step could be to run a study, which includes a relatively large number of participants, to accommodate for the heterogeneous nature of WFD, that is *also* longitudinal (Thomas et al., 2009). The longitudinal component of the study would be needed to track developmental trajectories of core language skills and to assess longer-term outcomes of intervention. Comparison of computational modelling simulations to empirical data has shown to be an instrumental aspect of this study. Thus, taking into account that researchers would almost always face the issue of a limited number of participants as a source for WFD empirical data, it is suggested that combining the two methods of empirical data analysis and computational simulations of intervention as demonstrated in this study, (whenever possible) may be a beneficial strategy for moving forward.

References

Alireza, H., Fedor, A., & Thomas, M. S. C. (2017). Simulating behavioural interventions for developmental deficits: When improving strengths produces better outcomes than remediating weaknesses. In G. Gunzelmann, A. Howes, T. Tenbrink & E. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, London, UK, 26-29 July 2017.

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yaeger, D. S. (2020). Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions. *Psychological Science in the Public Interest, 21(2),* 55-97. https://doi.org/10.1177%2F1529100620915848

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London: Series B, 358*, 1177-1187.

Benasich, A. A., & Tallal, P. (2002). Infant Discrimination of Rapid Auditory Cues Predicts Later Language Impairment. *Behavioural Brain Research, 136,* 31–49.

Benton, A. L., & Hamsher, K. (1978*). Multilingual Aphasia Examination.* Iowa City: University of Iowa Hospitals.

Best, W. (2003). Finding the right approach: How do Speech and Language Therapists tackle word-finding problems in children? *Royal College of speech and Language Therapists Bulletin*, September 2003, 5-6.

Best, W. (2005). Investigation of a new intervention for children with word-finding problems. *International Journal of Language and Communication Disorders, 40*(3), 279-318.

Best, W., Fedor, A., Hughes, L., Kapikian, A., Masterson, J., Roncoli, S., Fern-Pollak, L., & Thomas, M. S. C. (2015). Intervening to alleviate word-finding difficulties in children: case series data and a computational modelling foundation. *Cognitive Neuropsychology, 32*(3-4), 133-168, DOI: 10.1080/02643294.2014.1003204

Best, W., Hughes, L. M., Masterson, J., Thomas, M., Fedor, A., Roncoli, S., … Kapikian, A. (2017). Intervention for children with word-finding difficulties: a parallel group randomised control trial. *International Journal of Speech-Language Pathology*, 1–12. http://doi.org/10.1080/17549507.2017.1348541

Best, W., Hughes, L., Masterson, J., Thomas, M. S. C., Howard, D., Kapikian, A. & Shobbrook, K. (2021). Understanding differing outcomes from semantic and phonological interventions with children with word-finding difficulties: a group and case series study. *Cortex, 134*, 145-161. https://doi.org/10.1016/j.cortex.2020.09.030

Best, W., Masterson, J., Thomas, M. S. C., Hughes, L., Fedor, A., Roncoli, S., & Kapikian, A. (2013, June 24– 25). The WORD Project: A case series study on intervention for word-finding difficulties. Presented at the Child Language Seminar 2013, Manchester.

Bishop, D. V. M. (2004). Specific Language Impairment: Diagnostic Dilemmas. In L. Verhoeven, & H. van Balkom (Ed.), *Classification of Developmental Language Disorders Theoretical Issues and Clinical Implications* (pp. 309-326). Mahwah, NJ: Erlbaum.

Bishop, D. V. M. (2009). Genes, cognition and communication: insights from neurodevelopmental disorders. *The Year in Cognitive Neuroscience: Annals of the New York Academy of Sciences, 1156,* 1–18.

Bishop, D. V. M. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language and Communication Disorders, 49,* 381–415.

Bishop, D. V. M., & McArthur, G. M. (2005). Individual Differences in Auditory Processing in Specific Language Impairment: A Follow-Up Study Using Event-Related Potentials and Behavioural Thresholds. *Cortex, 41*(3)*,* 327-341.

Bishop, D. V. M., Nation, K., & Patterson, K. (2014). When Words Fail Us: Insights into Language Processing from Developmental and Acquired Disorders. *Philosophical Transactions of The Royal Society B,* 369.

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & The CATALISE Consortium. (2016). CATALISE: a multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. PLOS One, 11(7). http://dx.doi.org/10.1371/journal.pone.0158753

Borovsky, A. & Elman, J. L. (2006). Language input and semantic categories: a relation between cognition and early word learning. *Journal of Child Language*, *33(4)*, 759-790.

Botting, N., & Conti-Ramsden, G. (2004). Characteristics of Children With Specific Language Impairment. In L. Verhoeven, & H. van Balkom (Ed.), *Classification of Developmental Language Disorders Theoretical Issues and Clinical Implications* (pp. 23-38). Mahwah, NJ: Erlbaum.

Boyle, J., McCartney, E., O'Hare, A., & Law, J. (2010). Intervention for mixed receptive-expressive language impairment: A review. *Developmental Medicine and Child Neurology 52*, 994–99.

Bragard, A., Schelstraete, M-A., Snyers, P., & James, D. G. H. (2012). Word-Finding intervention for children with specific language impairment: A multiple single-case study. *Language, Speech, and Hearing Services in Schools, 43,* 222-234.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3* (2)*,* 77-101.

https://doi.org/10.1191/1478088706qp063oa

Bronfenbrenner, U. (1992). Ecological systems theory. In R. Vasta (Ed.), Six Theories of Child Development. *Annals of Child Development*, (Vol. 6, pp. 187-249). London, UK: Kingsley (Kindle edition).

Bullinaria, J. A. (1994). Representation, Learning, Generalization and Damage in Neural Network Models of Reading Aloud. *Edinburgh University Technical Report.*

Bullinaria, J.A. (1997). Analysing the Internal Representations of Trained Neural Networks. In: A. Browne (Ed.), *Neural Network Analysis, Architectures and Applications,* 3-26. Bristol: Institute of Physics Publishing.

Chan, J., Glynn, P., & Kroese, D. (2011). A comparison of cross-entropy and variance minimization strategies. *Journal of Applied Probability, 48*(A), 183-194. doi:10.1239/jap/1318940464

Chen, L., Lambon Ralph, M. A., & Rogers, T. T. (2017). A unified model of human semantic knowledge and its disorders. *Nature Human Behaviour*, *1*, 0039.

Coady, J. A. (2013). Rapid Naming by Children With and Without Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *56*, 604–617.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology, 33* (4), 497- 505. https://websites.psychology.uwa.edu.au/school/mrcdatabase/uwa_mrc.htm

Constable, A., Stackhouse, J., & Wells, B. (1997). Developmental word-finding difficulties and phonological processing: the case of the missing handcuffs. *Applied Psycholinguistics, 18,* 507-536.

Corriveau, K., Pasquini, E., & Goswami, U. (2007). Basic Auditory Processing Skills and Specific Language Impairment: A New Look at an Old Hypothesis. *Journal of Speech, Language, and Hearing Research, 50,* 647-666.

Costa, A., Strijkers, K., Martin, C., & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences, 106*(50)*,* 21442-21446.

Daniloff, R. G. (2002). *Connectionist approaches to clinical problems in speech and language.* Mahwah, NJ: Erlbaum.

Davis, R. (2017). *An empirical and computational investigation of variable outcomes in autism spectrum disorder*. Unpublished PhD Thesis. University of London, UK.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,*(3), 283-321.

Dockrell, J. E., Messer, D., & George, R. (2001). Patterns of naming objects and actions in children with word-finding difficulties. *Language and Cognitive Processes, 16*, 261-286.

Dockrell, J. E., Messer, D., George, R., & Wilson, G. (1998). Notes and Discussion: Children with word-finding difficulties-prevalence, presentation and naming problems. *International Journal of Language & Communication Disorders, 33*(4), 445-454.

Druks, J. & Masterson, J. (2000). *Object and Action Naming Battery*. Hove: Psychology Press.

Easton, C., Sheach, S. & Easton, S. (1997). Teaching vocabulary to children with word-finding difficulties using a combined semantic and phonological approach: an efficacy study. *Child Language Teaching and Therapy, 13*(2), 33-

50.

Ebbels, S. (2014a). Introducing the SLI debate. *International Journal of Language and Communication Disorders,* 49, 377-380, http://onlinelibrary.wiley.com/doi/10.1111/1460-6984.12119/full

Ebbels, S. (2014b). Effectiveness of intervention for grammar in school-aged children with primary language impairments: A review of the evidence. *Child Language and Teaching Therapy*, *30*(1), 7-40.

Ebbels, S. (2016, November 17). *International consensus on diagnosis for children with problems with language development- Summary*. *New Terminology For Language Impairments- SLI to become DLD*. https://www.afasic.org.uk/2016/11/new-terminology-for-language-impairments-sli-to-become-dld/ (downloaded 21 March 2020).

Elliott, C. D., Smith, P., & McCulloch, K. (1997). The British Ability Scales II. Windsor, NFER-Nelson.

Fedor, A., Best, W., Masterson, J., & Thomas, M. S. C. (2013). Towards identifying principles for clinical intervention in developmental language disorders from a neurocomputational perspective. *DNL Tech report 2013-1.*

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd Ed.). London: Sage Publications.

Fields, R. D. (2010). Change in the Brain's white matter. *Science, 330* (6005), 768-769. https://www.science.org/doi/abs/10.1126/science.1199139

Friedmann, N., Biran, M., & Dotan, D. (2013). Lexical retrieval and its breakdown in aphasia and developmental language impairment. In C. Boeckx & K. K.

Grohmann (Eds.), *The Cambridge Handbook of Biolinguistics.* Cambridge: University Press.

Fromkin, V. A. (Ed.) (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

Funnell, E., Hughes, D., & Woodcock, J. (2006). Age of acquisition for naming and knowing: A new hypothesis. *The Quarterly Journal of Experimental Psychology, 59*(2), 268-295.

Gagnon, D. A., & Martin, N. (2002). Diagnosis, prognosis, and remediation of acquired naming disorders from a connectionist perspective. In R. G. Daniloff (Ed.), *Connectionist approaches to clinical problems in speech and language* (pp. 147-188). Mahwah, NJ: Erlbaum.

Gale, N. K., Heath, G., Cameron, E., Rashid, S. & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *Medical Research Methodology 13*, 117. http://www.biomedcentral.com/1471-2288/13/117

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 133-177). New York: Academic Press.

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. http://doi.org/10.1080/09658219408258940

Gathercole, S., & Baddeley, A. (1996). *The children's test of non-word repetition.* London: Psychological Corporation.

German, D. J. (1986). *National College of Education, Test of Word-Finding (TWF).* Allen, TX: DLM Teaching Resources.

German, D. J. (1989). A diagnostic model and a test to assess word-finding skills in children. *British Journal of Disorders of Communication, 24*, 21-39.

German, D. J. (1990). *National College of Education, Test of Adolescent/Adult Word-Finding (TAWF).* Allen, TX: DLM Teaching Resources.

German, D. J. (1991). *National College of Education, Test of Word-Finding in Discourse (TWF-D).*Allen, TX: DLM Teaching Resources.

German, D. J. (2000). *Test of Word-Finding* (2nd ed.) (TWF-2). Austin, TX: ProEd.

German, D. J., & Newman, R. S. (2004). The impact of lexical factors on children's word-finding errors. *Journal of Speech, Language, and Hearing Research, 47,* 624-636.

German, D. J., & Simon, E. (1991). Analysis of children's word-finding skills in discourse. *Journal of Speech and Hearing Research, 34,* 309–316.

Green, C. S. (2020). Interventions to Do Real-World Good: Generalization and Persistence. *Association for Psychological Science, 21* (2), 43-49.

Harley, T. (2001). *The Psychology of Language* (2nd Ed.). New York: Psychology Press.

Harm, M. W. (2002). Building Large Scale Distributed Semantic Feature Sets With WordNet. *Carnegie Mellon University/CNBC Tech Report.*

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*(3), 491-528.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labour between visual and phonological processes. *Psychological Review, 111 (3),* 662-720.

Harm, M. W., McCandliss, B. D., & Seidenberg, M. S. (2003). Modelling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading 7,* 155-182. https://doi.org/10.1207/S1532799XSSR0702_3

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in boltzmann machines. In D. E. Rumellhart & J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition* (Vol. 1) (pp.282-317). Cambridge, MA: MIT Press.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98*(1), 74-95.

Howitt, D & Cramer, D (2011). *Introduction to Statistics in Psychology.* (5th ed.). Harlow, England: Prentice Hall.

Jing, H. (1998). Usage of WordNet in Natural Language Generation. In *Proceedings of the Joint 17th International Conference on Computational Linguistics 36th Annual Meeting of the Association for Computational Linguistics (COLINGACL'98) workshop on Usage of WordNet in Natural Language Processing Systems,* p. 128-134.

Kaplan, E., Goodglass, H., & Weintraub, S. (1976). *Boston Naming Test.* Boston: Boston University School of Medicine.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences, 2,* 389-398.

Karmiloff-Smith, A. (2009). Nativism Versus Neuroconstructivism: Rethinking the Study of Developmental Disorders. *Developmental Psychology, 45*(1), 56-63.

Lahey, M., & Edwards, J. (1996). Why do children with specific language impairment name pictures more slowly than their peers? *Journal of Speech and Hearing Research, 39,* 1081–1098.

Lahey, M., & Edwards, J. (1999). Naming errors of children with specific language impairment. *Journal of Speech Language and Hearing Research, 42,* 195–205.

Law, J. (2004). The close association between classification and intervention for children with primary language impairments. In L. Verhoeven, & H. van Balkom (Ed.), *Classification of Developmental Language Disorders Theoretical Issues and Clinical Implications* (pp. 401-418). Mahwah, NJ: Erlbaum.

Law, J., Campbell, C., Roulstone, S., Adams, C., & Boyle, J. (2008). Mapping practice onto theory: The speech and language practitioner's construction of receptive language impairment. *International Journal of Language and Communication Disorders*, *43*(3), 245–263.

Law, J., Lee, W., Lindsay, Roulstone, S., Wren, Y., Zeng, B., & Lindsay, G. (2012). 'What Works': Interventions for children and young people with speech, language and communication needs. *Better communication research programme: 2nd interim report*. Technical report, UK Department for Education.

Leonard, L. B. (2014). *Children with Specific Language Impairment.* Cambridge, MA: MIT Press.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 42,* 1-22.

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Science 3,* 223–232.

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences USA*, *98*(23), 13464-13471. Retrieved from http://www.pnas.org/cgi/content/full/98/23/13464

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science: A Multidisciplinary Journal, 31,* 581-612.

Mareschal, D., Johnson, M. H., Sirois, S., Spratling, M., Thomas, M. & Westermann, G. (2007*). Neuroconstructivism, Vol. I: How the brain constructs cognition.* Oxford, UK: Oxford University Press.

Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2010). Children's printed word database: Continuities and changes over time in children's early reading vocabulary. *British Journal of Psychology, 101,* 221–242.

Mayor, J., & Plunkett, K. (2010). A neuro-computational account of taxonomic responding and fast mapping in early word learning. *Psychological Review, 117*(1), 1–31. doi:10.1037/a0018130

McCandliss, B. D., Beck, I., Sandak, R., & Perfetti, C. (2003). Focusing attention on decoding for children with poor reading skills: A study of the Word Building intervention. *Scientific Studies of Reading, 7*(1),75-105.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the r – l contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive Affective and Behavioural Neuroscience,* 2(2), 89-108.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumellhart & J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition* (Vol. 1) (pp.3-44). Cambridge, MA: MIT Press.

McClelland, J. L., Thomas, A. G., McCandliss, B. D., & Fiez, J. A. (1999). Understanding failures of learning: Hebbian learning, competition for representational space, and some preliminary experimental data. *Progress in*

Brain Research, 121, 75-80.

McGregor, K. K. (1994). The use of phonological information in a word-finding treatment for children. *Journal of Speech and Hearing Research, 37,* 1381–1393.

McGregor, K. K. (1997). The nature of word-finding errors of preschoolers with and without word-finding deficits. *Journal of Speech and Hearing Research, 40*, 1232–1244.

McGregor, K. K., & Leonard, L. B. (1989). Facilitating Word-Finding Skills of Language-Impaired Children. *Journal of Speech and Hearing Disorders, 54,* 141-147.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science, 317*, 631.

McMurray, B., Zhao, L., Kucker, S. C., & Samuelson, L. K. (2013). Pushing the envelope of associative learning: Internal representations and dynamic competition transform association into development. In L. Gogate & G. Hollich (Eds.), *Theoretical and computational models of word learning: Trends in psychology and artificial intelligence* (pp. 49 80). Hershey, PA: Information Science Reference. doi:10.4018/978-1 4666-2973- 8.ch003

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behaviour Research Methods, Instruments, & Computers*, *37*(4), 547-559.

Meringer, R., & Mayer, K. (1895). Versprechen und Verlesen, Goschenscher-Verlag (Reprinted 1978, with introductory essay by A. Cutler and D.A. Fay, Benjamins), https://pure.mpg.de/rest/items/item_2359139_4/component/file_2359138/content

Messer, D. & Dockrell, J. (2006). Children's word-finding difficulties: Descriptions, and explanations. *Journal of Speech, Language and Hearing Research, 49, 309-324*.

Messer, D., Dockrell, J. E., & Murphy, N. (2004). The relationship between naming and literacy in children with word-finding difficulties. *Journal of Educational Psychology*, *96*, 462-470.

Michie, S., & Prestwich, A. (2010). Are interventions theory-based? Development of a theory coding scheme. *Health Psychology*, *29(1)*, 1-8. doi: 10.1037/a0016939.

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6:42.

Middleton, E. L., & Schwartz, M. F. (2012). Errorless learning in cognitive rehabilitation: A critical review. *Neuropsychological Rehabilitation, 22*(2), 138–168.

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM 38* (11): 39-41.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography, 3,* 235-312.

Miller, G.A. (1991). *The Science of Words,* Scientific American Library

Morgan, L., Marshall, J., Harding, S., Powell, G., Wren, Y., Coad, J., & Roulstone, S. (2019). 'It depends': Characterizing speech and language therapy for preschool children with developmental speech and language disorders. *International Journal of Language and Communication Disorders, 54*(6), 954-970.

Newman, R. S., & German, D. J. (2002). Effects of lexical factors on lexical access among typical language-learning children and children with word-finding difficulties. *Language & Speech, 43,* 285–317.

Norris, J. A., & Hoffman, P. R. (2002). Language development and late talkers. In R. G. Daniloff (Ed.), *Connectionist approaches to clinical problems in speech and language* (pp. 1-110). Mahwah, NJ: Erlbaum.

Onnis, L., Monaghan, P., Christiansen, M., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. *Proceedings of the 26th Annual Conference of the Cognitive Science Society, 26,* 1047-1052. Retrieved from https://escholarship.org/uc/item/4558716n

Plaut, D. C. (1996). Relearning after Damage in Connectionist Networks: Toward a Theory of Rehabilitation, *Brain and Language, 52,* 25–82.

Plaut, D. C., & McClelland, J. L. (1993). Generalization with componential attractors: Word and non-word reading in an attractor network. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 824–829). Hillsdale, NJ: Erlbaum.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology, *Cognitive Neuropsychology*, *10*(5), 377-500.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115. https://doi.org/10.1037/0033-295X.103.1.56

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition, 48,* 21-69.

Plunkett, K., Sinha, C., Møller, M. F., & Strandsby, O. (1992). Symbol Grounding or the Emergence of Symbols? Vocabulary Growth in Children and a Connectionist Net. *Connection Science, 4,* 293-312.

Powell, D., Stainthorp, R., Stuart, M., Garwood, H., & Quinlan, P. (2007). An experimental comparison between rival theories of rapid automatised naming performance and its relationship to reading. *Journal of Experimental Child Psychology, 98* (1), 46– 68.

Rapin, I. (1996). Practitioner Review: Developmental Language Disorders: A Clinical Update. *Journal of Child Psychology and Psychiatry, 37*(6), 643-655.

Rapin, I., & Allen, D. (1983). Developmental language disorders: Nosologic considerations. In U. Kirk (Ed.), *Neuropsychology of language, reading, and spelling* (pp. 155-184). New York: Academic Press.

Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology,* 28(3), 406-413.

Richardson, F., & Thomas, M. S. C. (2008). Critical periods and catastrophic interference effects in the development of self-organising feature maps. *Developmental Science*, *11*(3), 371-389.

Rogers, T. T., & McClelland, J. L. (2006). *Semantic Cognition: A Parallel Distributed Processing Approach.* Cambridge, MA: The MIT Press.

Roulstone, S., Wren, Y., Bakopoulou, I. & Lindsay, G. (2012). Interventions for children with speech, language and communication needs: An exploration of current practice. *Child Language Teaching and Therapy, 28*(3), 325–341.

Ruh, S., & Westermann, G. (2009). OXlearn: A new MATLAB-based simulation tool for connectionist models. *Behaviour Research Methods, 41 (4),* 1138-1143. doi:10.3758/BRM.41.4.1138

Rumelhart, D. E., & McClelland, J. L. (1986a). *Parallel Distributed Processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986b). On learning the past tense of English verbs. In D. E. Rumellhart & J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition* (Vol. 2) (pp.216-271). Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, J. (1986a). Learning internal representations by error propagation. In D. E. Rumellhart & J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition* (Vol. 1) (pp. 318-362). Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986b). Learning representations by back-propagating errors. *Nature 323* (6088): 533–536.

Schuchard, J. & Middleton, E. L. (2018). The roles of retrieval practice versus errorless learning in strengthening lexical access in aphasia. *Journal of Speech, Language, and Hearing Research 61,* 1700-1717.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.

Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L., & McRae, K. (1994). Non-word pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1177–1196.

Sheng, L., & McGregor, K. K. (2010). Lexical–Semantic Organization in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research 146*(53), 146–159.

Small, S. L., Hart, J., Nguyen, T., & Gordon B. (1996). Distributed representations of semantic knowledge in the brain: Computational experiments using feature based codes. In J. Reggia, E. Ruppin & R. S. Berndt (Eds.), *Neural modelling of brain and cognitive disorders.* World Scientific.

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex 20*, 912-928. doi:10.1093/cercor/bhp153

Strong, G. K., Torgerson, C. J., Torgerson, D., & Hulme, C. (2011). A systematic meta-analytic review of evidence for the effectiveness of the 'Fast For Word' language intervention program. *Journal of Child Psychology and Psychiatry 52*(3), 224-235. doi:10.1111/j.1469-7610.2010.02329.x

Takesian, A. E., & Hensch, T. K. (2013). Balancing plasticity/stability across brain development. *Progress in Brain Research*, *207*, 3–34.

Tallal, P., & Gaab, N. (2006). Dynamic Auditory Processing, Musical Experience and Language Development. *Trends in Neurosciences 29* (7), 382-390.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., & Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science, 271*, 81–84.

Thomas, M. S. C. (1997). Connectionist networks and knowledge representation: The case of bilingual lexical processing. Unpublished D.Phil. thesis, University of Oxford.

Thomas, M. S. C. (2003). Multiple causality in developmental disorders: Methodological implications from computational modelling. *Developmental Science, 6*(5), 537-556.

Thomas, M. S. C. (2016). Understanding delay in developmental disorders. *Child Development Perspectives*, *10(2)*, 73-80.

Thomas, M. S. C., & de Wet, N. M. (1998). Stochastic double dissociations in distributed models of semantic memory. In: G. Humphreys & D. Heinke (Eds.), *Proceedings of the 5th Neural Computation and Psychology Workshop*. Springer.

Thomas, M. S. C., & McClelland, J. L. (2023). Connectionist models of cognition. In R. Sun (Ed.), *The Cambridge Handbook of Computational Cognitive Sciences*. Cambridge University press.

Thomas, M. S. C., & Karmiloff-Smith, A. (2002). Modelling typical and atypical cognitive development. In U. Goswami (Ed.), *Handbook of Childhood Development* (pp. 575-599). Blackwells Publishers.

Thomas, M. S. C., & Karmiloff-Smith, A. (2003). Modelling Language Acquisition in Atypical Phenotypes. *Psychological Review, 110* (4), 647-682.

Thomas, M. S. C., & Knowland, V. C. P. (2014). Modelling mechanisms of persisting and resolving delay in language development. *Journal of Speech, Language, and Hearing Research*, *57(2)*, 467-483.

Thomas, M. S. C., Annaz, D., Ansari, D., Serif, G., Jarrold, C., & Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language, and Hearing Research, 52,* 336-358.

Thomas, M. S. C., Ansari, D., & Knowland, V. C. P. (2018). Annual Research Review: educational neuroscience: progress and prospects. *Journal of Child Psychology and Psychiatry,* ISSN 0021-9630.

Thomas, M. S. C., Fedor, A., Davis, R., Yang, J., Alireza, H., Charman, T., Masterson, J., & Best, W. (2019). Computational modelling of interventions for developmental disorders. *Psychological Review*, *126*(5), 693-726. Advance online publication. http://dx.doi.org/10.1037/rev0000151

Thomas, M. S. C., Forrester, N. A., & Ronald, A. (2013). Modelling socio-economic status effects on language development. *Developmental Psychology*, *49 (12)*, 2325-43. doi: 10.1037/a0032301

Tingley, S. J., Kyte, C. S., & Johnson, C. J. (2003). Single-word and conversational measures of word-finding proficiency. *American Journal of Speech-Language Pathology, 12,* 359-368.

Trentin, E., Brugnara, F., Bengio, Y., Furlanello, C., & De Mori, R. (2002). Statistical and neural network models for speech recognition. In R. G. Daniloff (Ed.), *Connectionist approaches to clinical problems in speech and language* (pp.213- 264). Mahwah, NJ: Erlbaum.

Tyler, A., Lewis, K. E., Haskill, A., & Tolbert, L. C. (2003). Outcomes of different speech and language goal attack strategies. *Journal of Speech Language and Hearing Research 46*, 1077-94.

Ueno, T., Saito, S., Rogers, T. T., & Lambon Ralph, M. A. (2011). Lichtheim 2: Synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, *72*, 385–396. http://dx.doi.org/10.1016/j.neuron.2011.09 .013

Vallabha, G. K., & McClelland, J. L. (2007). Success and failure of new speech category learning in adulthood: Consequences of learned Hebbian attractors in topographic maps. *Cognitive Affective and Behavioural Neuroscience,* 7 (1), 53-73.

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods, 40*(1), 183-190.

Warren, S. F., & Yoder, P. J. (2004). Early Intervention for Young Children With Language Impairments. In L. Verhoeven, & H. van Balkom (Ed.), *Classification of Developmental Language Disorders Theoretical Issues and Clinical Implications* (pp. 367-381). Mahwah, NJ: Erlbaum.

Weigel-Crump, C. A., & Dennis, M.D. (1986). Development of word-finding. *Brain and Language, 27,* 1-23.

Westermann, G., & Mareschal, D. (2014). From perceptual to language-mediated categorization. *Philosophical Transactions of The Royal Society B,* 369. http://dx.doi.org/10.1098/rstb.2012.0391

Wolf, M. (1986). Rapid alternating stimulus naming in the developmental dyslexias. *Brain and Language, 27,* 360–379.

Wolf, M., & Denckla, M. B. (2005). *RAN/RAS: Rapid Automatized Naming and Rapid Alternating Stimulus Tests.* Austin, TX: Pro-Ed.

Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M., & Merzenich, M. M. (1997). Deficits in Auditory Temporal and Spectral Resolution in Language-Impaired Children. *Nature 387*, 176-178.

Yang, J., & Thomas, M. S. C. (2015). Simulating intervention to support compensatory strategies in an artificial neural network model of atypical language development. In G. Airenti and M. Cruciani (Eds.), *Proceedings of EuroAsianPacific Joint Conference on Cognitive Science*, Torino, Italy, September 25-27th, 2015.

Appendix 1

The Questionnaire as seen in Google Docs Form

Intervener's Toolbox Questionnaire V7

We are currently engaged in a research project seeking to narrow the gap between a mechanistic, cognitive neuroscience-based understanding of language intervention, and the lessons learned by clinicians about what types of intervention are most effective for different children. Intervention is considered by some therapists to be a process of discovery and so we believe that therapists are in the best position to advance our field by sharing what they discover during their therapy sessions.

We would greatly appreciate your help in deepening our understanding of the practical, experience-based guidelines and rules of thumb that clinicians themselves have developed which they use to shape their own intervention practices. From these, we hope to draw firmer links between the implicit causal theories of language deficits grounded in clinical experience, and research-driven theories grounded in experimental behavioural studies and neuroscience. The ultimate goal is a more systematic theory of which language interventions work best for children with different strengths and weaknesses.

The below is a questionnaire aimed at therapists or teachers who regularly practice therapeutic interventions for children from different age groups who show developmental language problems. We have tailored the questions to the example of children with word-finding difficulties, but if there are other example behaviours that are more salient to your practice, please feel free to answer with respect to these.

Many thanks for your help – we greatly appreciate your time.

Hala Alireza
Michael Thomas

Consent:

Responses submitted on the google form will be completely anonymous even to the researchers. Results are normally presented in terms of groups of individuals. If any individual data are presented, the data will be totally anonymous, without any means of identifying the individuals involved.
A responder may withdraw consent for the study at any time without giving any reason and may decline to answer particular questions or request their data to be deleted.

By completing the questionnaire you agree that the details of the study have been explained to you and that you willingly consent to take part.

Please base your answers on 'one to one' intervention sessions rather than 'group' sessions.

1- Background: (a) What is your title/job role ?

1. *Mark only one oval.*

   - Speech and Language therapist
   - Occupational therapist
   - Researcher
   - Teacher
   - Teacher Assisstant
   - Special Needs Teacher
   - Clinical Psychologist
   - Educational Psychologist
   - Psychologist

2. If other please specify:

   ____

3. - (b) How many years of experience do you have in this field ?

   ____

4. - (c) Where do the intervention sessions take place?

   Mark only one oval.

   - School
   - Health Centre
   - Home

5. If other please specify

   ____

6. - (d) What is the range of age of the children you work with?

Mark only one oval.

- under 5 years old
- 4 to 11 years old
- 11 to 18 years old
- Other: ____________________

2 - In every situation, there is often more than one approach and a range of techniques that can be used to treat a language difficulty.

7. (a) In the case of word finding difficulties what would be the first or default approach you would select to improve the target behaviour X? For example what intervention would you use to support a child who had retrieval difficulties for the word 'dog'?

8. (b) On what grounds would you decide what to try first? What would the choice of intervention plan depend on and (if applicable) how would the characteristics of the child influence the choice of therapy technique? How often do children with similar language difficulties respond differently to the same approach or technique?

3 - Say you are trying to improve a target behaviour (e.g., saying the word 'dog') using some technique. We are interested in how much you reinforce the behaviour by repetition, and how much you use variation of items or variation of techniques (and why)? If you use variation, what would you vary (e.g., items – pictures of different types of dogs; techniques – naming from pictures, modelling your naming)? If you use repetition, do you select particular more representative items to work with (e.g., prototypical dog, say a Labrador rather than a Chihuahua)?

9.

4 - For our example of vocabulary retrieval difficulties if you were using pictures to prompt naming, would you use only pictures of dogs, or contrast with different animals as well?

10.

5 - How do you approach the possibility of a child making a mistake? In general do you let the child make a mistake so that they learn from it or would you support/scaffold the child's behaviour in a way to reduce the possibility of mistakes (e.g., context, cuing)? What type of scaffolding would you use?

11.

6- How would you balance practising a behaviour as a whole (e.g., saying the word 'dog') versus breaking it up into component skills, practising the components, and then putting them together into the whole (e.g., breaking into syllables practising component phonemes or onset vs. rime, of 'dog')? Why would breaking up or simplifying help?

12.

7 - How important is the child's level of engagement in the intervention task, and the child's general motivation? How would you manipulate either engagement in the task or general motivation?

13.

8 - Sometimes, working on weaknesses can be uncomfortable for the child. How much time would you apportion working on more difficult versus less difficult tasks in a session?

14.

9 - How do parental motivation and resources contribute to intervention success?

15.

10 - For behaviour X, to what extent would you persist in improving X, compared to encouraging a compensatory strong skill Y that can be used as an alternative to X? (e.g., seeking a substitute word or phrase for the target word 'dog')

16.

11 - How would you support generalisation of the target skills (e.g., naming) beyond the items used in the intervention session? What different techniques might support generalisation?

17.

12 - How much does the 'dosage' required to ameliorate a behavioural difficulty vary across children, where dosage is the time per session or total number of sessions? i.e. How do you gauge a dosage? Can this be predicted early on during therapy or only assessed retrospectively based on rate of progress? What characteristics of the child do you think influence the required dosage?

18.

13- Given what you think is the underlying cause of the child's difficulties, are there any other factors that might also contribute to your choice of intervention tool (e.g., personal expertise with technique; institutional policy; availability of materials; cost efficiency; anticipated contact time with child)?

19.

14- How important is the child's general intelligence (IQ) in predicting the child's response to intervention?

20.

15 - Further comments: We would greatly appreciate hearing your views on any other factors that you feel are relevant to intervention choices and outcome success, or views on our research project.

21.

Thank you again for taking the time to complete this questionnaire!

This content is neither created nor endorsed by Google.

Google Forms

Appendix 2

Responses to the Questionnaire

H

QUESTIONS RESPONSES 10

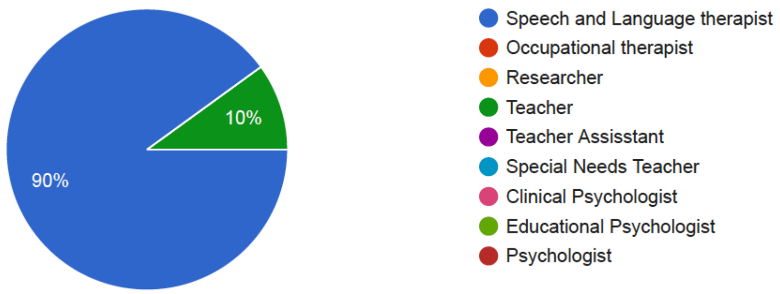
10 responses

SUMMARY INDIVIDUAL

Accepting responses

1- Background: (a) What is your title/job role ?

10 responses



If other please specify:

0 responses

No responses yet for this question.

- (b) How many years of experience do you have in this field ?

10 responses

https://docs.google.com/forms/d/1g7F3tAxLVPqST0flYvlr6d7VBndmZ671j3kKgENBRWQ/edit#responses 1/14

22 years

14

13

4

7

3

10

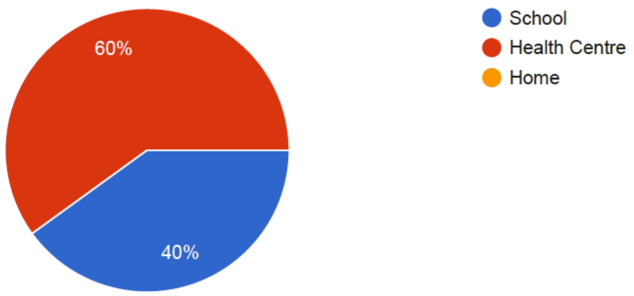
7 years

24

15 years

- (c) Where do the intervention sessions take place?

10 responses



If other please specify

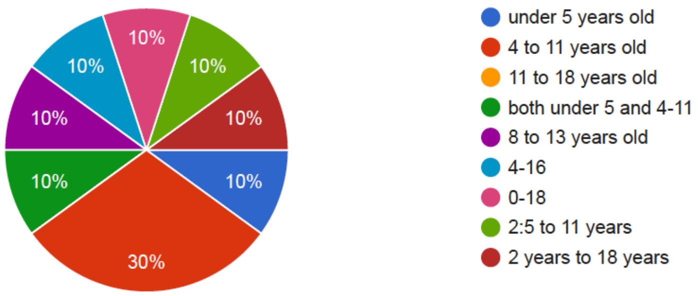
2 responses

school, HC and early childhood centres

And autism center

- (d) What is the range of age of the children you work with?

10 responses



2 - In every situation, there is often more than one approach and a range of techniques that can be used to treat a language difficulty.

(a) In the case of word finding difficulties what would be the first or default approach you would select to improve the target behaviour X? For example what intervention would you use to support a child who had retrieval difficulties for the word 'dog'?

10 responses

For under fives I would use a mainly semantic based approach where I talked about the qualities of a dog, eg , category, what does it do, where does it live. I may also talk about the number of syllables and possibly the first sound and perhaps a 'word that sounds nearly the same' (rhyming word)

Wordaware type approach, multiple linguistic/phonological details

It would depend on the child's language profile - i.e. the other language difficulties that they have and the naming errors that they made. For example, if they had very weak semantics and called the dog another animal's name then this would be an area to address, whereas if they made a lot of phonological errors, then this area would be a good candidate to address.

Semantic links and phonemic cues

Use circumlocution strategies: describe the target item as much as possible. If the client is literate I can use orthographic cues or ask the client to try to imagine how the word is written

Visual memory

Naming pictures and repititions

I would assess the child first and see whether he/she has semantic or phonological type of word finding difficultly and that would give me a guide regarding the needed intervention

Using play -following child lead

Picturrs

(b) On what grounds would you decide what to try first? What would the choice of intervention plan depend on and (if applicable) how would the characteristics of the child influence the choice of therapy technique? How often do children with similar language difficulties respond differently to the same approach or technique?

10 responses

The age, language level, attention level and my hypothesis regarding the underlying cause of the difficulty. Children can respond differently because of motivation, self -awareness and ability to problem-solve.

Depends on the child's level of engagement and maturity.

See above. Assessment would guide intervention. I also use dynamic assessment during the initial evaluation and this can be useful in guiding the first approach to try. With regard to children with similar language difficulties responding differently to the same approach or technique - yes, this definitely happens. However, sometimes there are other factors contributing to this e.g. regularity of home practice, child's motivation to achieve their targets etc and this makes it difficult to quantify the answer to the above question.

Other areas of need e.g. OT, other language needs, attention and listening skills. Children with overall similar difficulties respond similarly. Children with needs in other areas will respond differently.

Depending on the degree of impairment and response to dynamic assessment during which I test different prompts and see which ones would be more effective

Low IQ, artistic child and ADHD

Least known items to more known

The intervention approach selected would be tailored to the child. This would be done by first understanding the child's deficits. Then, relying on evedince based approaches that yielded to positive in similar cases to my client. On top of the that, the child's parents also has to agree on the approach suggested.
Children with the same language deficits respond differently to the same approach because of differences in other factors such as attention, cognition, behavior, and even parental involvement in therapy.

Depending on the child char.,and level ,his interest

Pictures and sounds according to child's knowledge of phonemes

3 - Say you are trying to improve a target behaviour (e.g., saying the word 'dog') using some technique. We are interested in how much you reinforce the behaviour by repetition, and how much you use variation of items or variation of techniques (and why)? If you use variation, what would you vary (e.g., items – pictures of different types of dogs; techniques – naming from pictures, modelling your naming)? If you use repetition, do you select particular more representative items to work with (e.g., prototypical dog, say a Labrador rather than a Chihuahua)?

10 responses

I would use lots of repetition of the word , particularly if the child's underlying deficit were more in the phonological area and particularly if I was targeting longer words. I wouldn't use a huge amount of variation as the child understand the word, ie knows what a do is . The problem is with access/storage.

a lot of repetition, variation in presentation to stop disengagement/boredom, vary pictures to prevent overgeneralisation but have pertinent salient core features.

For word finding, I often use an eclectic approach e.g. I have a resource that contains both semantic and phonological prompts. I would definitely vary the items and the techniques as variation helps to keep the child motivated and engaged in the sessions. I would vary the difficulty of the tasks - using easier targets at the start and after very difficult tasks. I would only introduce more difficult tasks once the child has experienced some success with easier tasks. An example of an 'easier' target might be an activity involving pictures that the child has seen before, with a variety of cues if the child cannot recall the target word (e.g semantic lead in/phonological cue). An example of a harder task would be recalling the name within a structured game (e.g. a barrier game where other language demands are high) or in a conversational task. Harder tasks also involve fewer cues and more prompts for the child to self-cue using the strategies taught. I monitor how the child is responding and use this to guide future therapy sessions. I would use more representative items during the initial stages of therapy at least.

repetition will be key in a 6 week block. Variation may come in activities, different pictures, paper vs physical items vs ipad games. Always use standard pictures that come up on google or Widgit

Reinforce every correct response at the acquisition stage of learning. Use a variety of stimuli such as different pictures to promote generalization

I may repeat the trail if the child is closed to the correct or not produced correctly

No

Repetition is needed for the behavior to be acquired and fixed sechduled of reinforcement would aid in the acquisition as well. Multiple examples of the the traget word is also needed to generalize the skill.

Yes . I use all variations you mentions

Same picture of dog initially but towards end when child is more confident a variation

4 - For our example of vocabulary retrieval difficulties if you were using pictures to prompt naming, would you use only pictures of dogs, or contrast with different animals as well?

10 responses

Contrast (2)

If I were using assorting task I would use pictures of other animals/objects.

initially only dogs, once established introduce other animals to contrast and probe retrieval

Other animals as well.

Will contrast after I ensure mastery of different pictures of different dogs

Contrast for other animals

At the beginning stage, I would only use picture of dogs. Once, the child has mastered the skill and is able to retrieve the word with different pictures than the one I have used. Then, I will mix the picture of the dog with other picture of animals.
This also would depend on how many words I am targeting at the same time as the word "dog" .

With animals,hous,as story

Pictures with the same phonemes

5 - How do you approach the possibility of a child making a mistake? In general do you let the child make a mistake so that they learn from it or would you support/scaffold the child's behaviour in a way to reduce the possibility of mistakes (e.g., context, cuing)? What type of scaffolding would you use?

10 responses

Again , depends on the child. Some children are demotivated by making mistakes so I might scaffold , eg use initial sound cues and forced choices to help them access a word.

scaffold to reduce mistakes, context, semantic and phonemic cues, forced alternative

It depends on the child's stage of therapy and their level of success using self-cuing strategies. For example, if they have already learned a lot of self-cueing strategies then I would let them try out one self-cuing strategy, evaluate whether it was the word they wanted and then try another one. If their self-monitoring was not at that level yet and if they were not using strategies and were likely to just give a random guess, then I would scaffold their answer. Scaffolds include contextual support (e.g. lead ins, context/semantic clues etc) and phonological supports. I also use visual supports e.g. the aforementioned prompt sheet.

Both allow them to make a mistake and scaffold. Using cueing and visuals.

At the beginning, I would choose errorless teaching to build the chain for fhe contingency of (stimulus- correct response- reinforcement) with the appropriate level of prompts that will result in a correct response

Yes

Reduce

Errorless teaching should be used to help the child acquire the skill without frustration. The child will be taught the skill using with the needed cues. These cues will be gradually faded as the child master the skill.

Starting training from less error teaching using prompts to reach normal level of mistake

First make a mistake then correct child using a positive approach

6- How would you balance practising a behaviour as a whole (e.g., saying the word 'dog') versus breaking it up into component skills, practising the components, and then putting them together into the whole (e.g., breaking into syllables practising component phonemes or onset vs. rime, of 'dog')? Why would breaking up or simplifying help?

9 responses

If children have difficulties in the phonological area I find it help them to break down longer words (more than one syllable). For these children , sometimes all I need do is clap out the rhythm of the word and they can access it. Rhyming may also give them another hook, or bereaking down the word into other known words, eg anne-ten-eye

Would start with whole, then break up then back to whole. this would ensure multiple links with the target word.

The balance would depend on the child. If they made a lot of phonological errors then I would spend more time raising their awareness of initial sound/syllables etc.

Breaking up could be a type of phonetic cuing that will involve both hemispheres of the brain in the process of retrieving the word

Work on breaking up the behavior first and do some drilling and then saying the whole word at the end for three times

No

The goal I have to achieve has to be clearly defined. I would not break the word "dog" into syllables because the aim is word finding difficulty and not articulation. If the prompt would be phonological then the level would depend on the child's performance and the level of prompt that helped him to remember the word.

It depends

Practice using different phases of phonemes becoming more difficult as child becomes more confident

7 - How important is the child's level of engagement in the intervention task, and the child's general motivation? How would you manipulate either engagement in the task or general motivation?

10 responses

Obviously engagement is key and I would use a variety of techniques including intrinsic and extrinsic rewards. I would try to make the session fun by using motivation materials, eg photos from a favourite film/messy activities for some children.

very important, try to identify activities enjoyed and use these with appropriate vocab targets

Both are very important. An example of facilitating engagement in the task would be by using the child's interests e.g. if you know they love football, using a football game that also involves naming of the targets. An example of an activity to increase general motivation is involving the child in their own goal setting and self evaluation.

Very. Offer rewards, base activities around their interests

Choosing reinforcement that the child likes and will actually motivate him to work for, it could be an edible, tangible or an activity. Whatever the child would be motivated to make an effort to earn

By using the reinforcement

Very important

Motivation is very important for the child to perform the task and work to do it right. This would be done by assessing the child's reinforcers beforehand in order to use them during the tasks.

For reaching the target it nessary tokeep child enggage and keep him
Work and end sessions successavely

Have different activities all related to the subject

8 - Sometimes, working on weaknesses can be uncomfortable for the child. How much time would you apportion working on more difficult versus less difficult tasks in a session?

10 responses

Depends on the child's personality and motivation.

2:1 ratio less difficult to more difficult ending with success

It would depend on how uncomfortable the child was with the more difficult tasks. If a child very uncomfortable with the harder tasks I would make sure that the tasks incorporated an element of their motivators (preferred activities). I would mix the harder and easier targets in the activities and then include a short activity with easier items. I would also mention to the child that some might be tricky, but not to worry - all children find some things easier and some things harder, and would remind them of this during the session. I would also have a child-friendly version of the therapy plan at hand (and possibly even involve them with deciding the order of activities). I would make sure to praise the effort esp. in difficult tasks and use a child-friendly effort-self-evaluation tool at the end of the task. I would also use a visual chart to show them their progress and particularly point out progress with the activities that are challenging for them. If the child was still very uncomfortable with the more difficult tasks then I would make them shorter, with more scaffolds until they became more confident and then incrementally increase the difficulty level e.g. reducing adult prompts.

50/50

2-3 easy trials then 1 difficult trial - based on the behavior momentum principle

By alternating the weakness and difficult tasks at the same session or even at the same period of time for applying the goals

Start with less then to more at the same session

If the task was difficult for the child, then it should be mixed with easy tasks.
It's different for each child. Some children get bored from easy tasks and liked to be challenged.

With less motvating child or child with low skills i start with less difficult task to let him feel confedince anf sucess then make variation in difficulty

Starting with 15 mins and increasing as child becomes more confident

9 - How do parental motivation and resources contribute to intervention success?

10 responses

I think intervention can still be effective for school aged children without much parental involvement if you have a good collaborative relationship with school. However parental involvement often makes the difference between a child exceeding a target/achieving a target.

crucial as working in isolation will achieve little

Parental motivation and home practice definitely helps intervention success and generalisation of the skill set. The only time when this would not be the case is if the parent gives negative feedback to the child if they do not succeed or does not recognise the child's small successes. However, in my experience, parental involvement is almost always a positive influence on therapy success.

School staff more important

When parents are involved it results in more practice of the newly learn d skills thus faster learning and generalization

The therapy goals are achieved very fast most of the time

Very improtant

Parent motivation and involvement is vital in intervention. They can aid in making the child progress faster in the targeted goals and generalize these goals outside of the therapy room. Parents will also learn ways to communicate with their children in better ways.

Very important

Very important

10 - For behaviour X, to what extent would you persist in improving X, compared to encouraging a compensatory strong skill Y that can be used as an alternative to X? (e.g., seeking a substitute word or phrase for the target word 'dog')

9 responses

Depends on the age/language ability of the child. For older , motivated children I would work on compensatory strategies alongsige word finding techniques. Sometime describing a word is enough to support access.

would depend how functionally important it was for the child

I would encourage both for two reasons. Firstly, the child needs to be able to communicate their message successfully and if this can be achieved by Y, then that is good progress. Secondly, the compensatory strategy can often be a stepping stone to the target (i.e. used as a form of self cuing). For example, "my pet (Y)...my dog (X)."

If behavior x is important, I could use differential reinforcement to teach x while still reinforcing Y

I might accept clause approximation not the substitution

Ill be happg with an alternative behaviour

It depends child's performance in the skill taught. How long have I been working on this skill? Is it helping the child? Have I used other strategies?

However, If the child's best performance is for example to describe a dog instead of recalling the word "dog", I would not stop him from doing so.

I use the strong skill in training target goal often

If child becomes stressed then changing tactics becomes helpful . If child is showing small signs of improvement then persist with initial tactic

11 - How would you support generalisation of the target skills (e.g., naming) beyond the

items used in the intervention session? What different techniques might support generalisation?

9 responses

For older children I would make the techniques, eg description , explicit and let them know they can use these anytime. I'd have a reminder on their desk and would training and encourage teaching assistants and teachers to prompt them to use the techniques.

stickers with 'speak to me about' for other environments, homework activities both paper and practical.

In the session, increase task demands e.g. telling a story involving the targets, conversational practice, and outside the session involving parents, teacher and other professionals working with the child (e.g. giving a short checklist of things to do when the child experiences word finding difficulties, doing a joint session with the parent/teacher, classroom visits. asking the parent/teacher to keep a log of word finding difficulties/strategies the child use to get through the difficulties for short periods during the day.)

Handing over to school workers/teachers to use in everyday situations

Use iphone pictures, watch an episode on National Geographic

Floortime technique and group therapy

Parent traning to target same tasks in natural environment per

1- Naming the items outside of the therapy room (different settings)
2- different people asking the child about the the targets

Incidential teaching

12 - How much does the 'dosage' required to ameliorate a behavioural difficulty vary across children, where dosage is the time per session or total number of sessions? i.e. How do you gauge a dosage? Can this be predicted early on during therapy or only assessed retrospectively based on rate of progress? What characteristics of the child do you think influence the required dosage?

8 responses

Often the children who have difficulties which are specific to language require less time to learn the techniques then those who have a more generalised learning difficultly. Also those who have awareness of their needs and are motivated to change move faster. I would try some dynamic assessment initially to help support my theory about how long the therapy would take.

Depends of clinical pressures however 8-10 1/2 hr sessions with break for consolidation. Depending on response to this initial period of input will influence further therapy. Many factors within the child and their environment influence progress.

Yes, dosage varies across children a lot. Amount of sessions required is often influenced by the child's language profile (e.g. receptive/expressive language difficulties, severity of difficulties), presence of other co-morbid difficulties e.g. ASD, ADHD, age of the child, child's motivation/engagement, and parental motivation/engagement to name just a few factors. These factors help predict how long therapy will take to an extent i.e. you could make a reasonable estimate based on these factors. However, rate of progress during the sessions (response to therapy) as well as generalisation to other environments are also strong predictive factors of dosage required. Your dosage estimate will likely be more accurate if you include these factors too.

The best is to refer to evidence-based literature available on the dosage appropriate for teaching a given skill

Based on the evaluation and the severity of the delay

Usually the dosage of an approach is mentioned in the literature.

Itvdepends

Depends on the child's SEN level - varying degree of difficulty e.g. Is it a speech difficulty , or behaviour related or some spectrum of autism

13- Given what you think is the underlying cause of the child's difficulties, are there any other factors that might also contribute to your choice of intervention tool (e.g., personal expertise with technique; institutional policy; availability of materials; cost efficiency; anticipated contact time with child)?

10 responses

not really. I fee the departmental approach is based in research and is appropriate. If I had less time , then I'd adapt the number of words tI'd work on , not the technique.

Yes, all of the above mentioned factors

Yes. For example, for stammering I would not use the Lidcombe programme if I was unable to see the child on a weekly basis and if the parent was unable to attend the sessions as this is what the programme requires in order to work. I would not use a tool just because I had expertise with the technique unless I thought that it was the most appropriate technique for the child. I try to use the evidence base to guide my intervention so would avoid using any intervention that has been shown to have an unsound evidence base (e.g. I would not use non-speech oro-motor exercises for speech sound disorders). I am fortunate that in my setting, institutional factors do not impact on therapy.

Time effective and used it before. Colleague recommendation

Personal expertise, child's and parental level of engagement , available resources, policies in the facility that I work at, feasibility

Structured and unstructured environment

Personal expertize

Parental involvement is a factor because certain approaches mandate commitment from the parents to carryover the goals in a specific way.

Some of what you mention

School resources , parents ability to maintain support out of school , specialist help within school all play a part

14- How important is the child's general intelligence (IQ) in predicting the child's response to intervention?

10 responses

Can be , as explain above, but it's not always the case.

this is important regarding rate of progress and ability to self-manage compensatory strategies for independent use

Overall IQ is made up of both PIQ (non-verbal) and verbal IQ. Verbal IQ is important as this score is based on language skills and language skills are what SLTs target. PIQ is not so important in predicting oral language based interventions.

Very - cognition relates to ability to learn new skills

Highly related, sometimes a child with a low IQ would platou and not progress

Give us the clear image and prognosis for responding to the intervention

70% important

The child's IQ level can limit the child's progress in therapy and it would also affect the type of intervention selected for them.

Important in planing therapy and running session and to avoid child depress and escape from therapy (behavior problem)

Depends on the child's difficulty within the class setting but general intelligence does play a part

15 - Further comments: We would greatly appreciate hearing your views on any other factors that you feel are relevant to intervention choices and outcome success, or views on our research project.

5 responses

I had not considered the above factors regarding my management prior to completing this survey. Thanks you for the opportunity to reflect.

I think that the motivation for the research project is great. Good luck with the project!

This questionnaire is rather difficult to follow and time consuming. Easier to give a scale for "to what extent" questions.

No more

Maybe a multiple choice questionnaire would be easier and quicker to answer . Some questions are repetitive .

Thank you again for taking the time to complete this questionnaire!

Appendix 3

Paragraph Accompanying RCSLT Research Newsletter

Language interventions questionnaire

A team at Birkbeck, University of London, is carrying out a research project seeking to narrow the gap between a cognitive neuroscience-based understanding of language intervention and the lessons learned by clinicians about what types of intervention are most effective for different children. SLTs who regularly practice therapeutic interventions for children from different age groups are invited to complete the questionnaire by 31 July: http://tinyurl.com/hz37znd Contact: halaalireza@gmail.com

Appendix 4.

Table A. Words grouped by the cluster analysis of the 'V&V 397 analogue semantics' list using linkage command by the 'Ward' method : Cutoff was at 20:
V = action Verb, N = action Noun, O = object

Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7	Node 8	Node 9	Node 10
Aggressive	Water & cooking related	Various mostly verbs	Eye related	Body sound	Furniture & house	Body parts	Leg & foot related	Commu-nication	Movement
PUNCH(V)	FILL(V)	ESCAPE(N)	LOOK(V)	HICCUP(N)	GATE	TOUCH(N)	SHOE	CALL(V)	ARRIVAL(N)
HIT(V)	SINK(V)	PLACE(V)	BLINK(N)	BURP(N)	FENCE	FINGER	STAGGER(V)	TELL(V)	RUN(V)
POUND(V)	BURN(V)	ACHE(N)	WINK(V)	COUGH(V)	CEILING	KNEE	LEG	TALK(V)	SHAKE(V)
SLAP(V)	BOIL(V)	EXCHANGE(V)	NOTICE(V)	BREATHE(V)	SOFA	ELBOW	KICK(V)	WHISPER(V)	THROW(N)
DROP(V)	FRY(V)	SUGGEST(V)	WATCH(V)	SIGH(N)	ROOF	WRIST	WANDER(V)	ARGUE(V)	CREEP(V)
FALL(V)	FLAME(N)	PAINT(V)	SEE(V)	COUGH(N)	CARPET	FEEL(V)	SOCK	WRITE(V)	SKID(V)
CUT(V)	SHIP	SMOKE(V)	BLINK(V)	SIGH(V)	STOOL	HAIR	MARCH(V)	SAY(V)	TAKE(V)
DESTROY(V)	ROAST(V)	MURDER(V)	WINK(N)	HICCUP(V)	SEAT	ITCH(N)	LIMP(V)	CHAT(V)	JOG(V)
BREAK(V)	RAFT	ESCAPE(V)	EYE	BURP(V)	CURTAIN	THUMB	PAW	WHISPER(N)	CARRY(V)
CRASH(V)	WASH(V)	TEACH(V)	WINDOW		RUG	TREMBLE(V)	FLOOR	CHAT(N)	ASCEND(V)
CRASH(N)	EMPTY(V)	WANT(V)			TABLE	TICKLE(V)	STAND(V)	CALL(N)	SLIDE(V)
BUMP(V)	FLAME(V)	RETURN(N)			DOOR	SHOULDER	TOE	DRAW(V)	RISE(V)
DESTRUCTION(N)	BAKE(V)	BLEED(V)			CHAIR	LEAD(V)	STEP(V)	SPEAK(V)	STIR(V)
HIT(N)	STEAM(V)	READ(V)			WALL	ANKLE		CALL(V)	WALK(V)
SLAP(N)	SPRAY(V)	LOSE(V)				NECK			BOUNCE(V)
SMASH(V)	BOAT	MIX(V)				ARM			PUSH(N)
	DIVE(V)	GIVE(V)				HAND			DRAG(V)
	POUR(V)	BOMB(V)				TOUCH(V)			PUSH(V)
	DRINK(V)	BELT				TREMBLE(N)			TWIST(V)
	GRILL(V)	MURDER(N)				EAR			LEAVE(V)
	FISH	PLEAD(V)				ITCH(V)			APPROACH(N)
	WADE(V)	ENTER(V)							BEND(V)
	SWIM(V)	ACHE(V)							PRESS(V)
		TRADE(V)							LIFT(V)
		STAY(V)							SWERVE(V)
		ADMIT(V)							THROW(V)
		PUT(V)							GO(V)
		SING(V)							APPROACH(V)
		DEMAND(N)							PULL(N)

Table A continued									
Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7	Node 8	Node 9	Node 10
Aggressive	Water & cooking related	Various mostly verbs	Eye related	Body sound	Furniture & house	Body parts	Leg & foot related	Commu-nication	Movement
		SWORD							HOLD(V)
		FIND(V)							MOVE(V)
		SEND(V)							COME(V)
		ENTRY(N)							ARRIVE(V)
		BUY(V)							HOP(V)
		LICK(V)							PULL(V)
		INVITE(V)							TRAVEL(V)
		BORROW(V)							
		PENCIL							
		PLEA(N)							
		LEND(V)							
		STOP(V)							
		WARN(V)							
		RETURN(V)							
		GUN							
		GREET(V)							
		PEN							
		RECEIVE(V)							
		SELL(V)							
		SIT(V)							
		BOOK							
		DIE(V)							
		BOX							
		HALT(V)							
		SUGGESTION(N)							
		ASK(V)							
		PAY(V)							
		FOLLOW(V)							
		KILL(V)							
		DEMAND(V)							
		GET(V)							
		EXCHANGE(N)							
		CHASE(V)							
		BOMB							

Table A continued

Node 11	Node 12	Node 13	Node 14	Node 15	Node 16	Node 17	Node 18	Node 19	Node 20
Fruits	Transport	Head related	Noise	Tools	Vegetables	Animals	Light related	Flight related	Clothing
CHERRY	BUS	CHIN	CRACKLE(V)	KNIFE	CELERY	SHEEP	GLOW(V)	CHIRP(V)	PANTS
ORANGE	BICYCLE	GRIN(V)	CHIME(N)	TOOTHBRUSH	MUSHROOM	CAMEL	FLICKER(N)	FLY(V)	HAT
APPLE	DRIVE(V)	RAISIN	CLATTER(V)	COMB	ONION	HORSE	SPARKLE(V)	DUCK	COAT
CORN	TRUCK	SNEEZE(V)	CLASH(N)	RAKE(N)	BEAN	FOX	FLICKER(V)	FEATHER	BLOUSE
GRAPE	CAR	DROOL(V)	CLATTER(N)	BRUSH(V)	CUCUMBER	DONKEY	SHINE(N)	HELICOPTER	VEST
LEMON	VAN	YAWN(V)	SCREAM(N)	SHAVE(V)	PEA	RABBIT	GLOW(N)	SWAN	GLOVE
PEAR	RIDE(V)	FROWN(V)	CLANG(N)	DRILL(V)	CABBAGE	BEAR	SPARKLE(N)	BIRD	SUIT
PLUM	TRAIN	SPOON	YELL(V)	FIX(V)	CARROT	MOUSE	SHINE(V)	WING	DRESS
STRAWBERRY	TRICYCLE	FROWN(N)	KNOCK(N)	SAW(N)	POTATO	LION	FLASH(V)	CHIRP(N)	SKIRT
LIME		FORK	GROWL(N)	BUILD(V)	LETTUCE	WOLF			MITTEN
PEACH		HEAD	SNORE(V)	CHOP(V)	PUMPKIN	ELEPHANT			SWEATER
BANANA		EAT(V)	WHINE(N)	CONSTRUCTION(N)	SPINACH	FUR			SHIRT
		TEETH	RATTLE(V)	SHIELD		GIRAFFE			SCARF
		SENSE(V)	YELL(N)	HAMMER		ZEBRA			
		SMILE(N)	CRACKLE(N)	BRUSH		GOAT			
		SMELL(V)	SCREECH(V)	SCISSORS		PIG			
		LIPS	SHOUT(N)	SAW(V)		DOG			
		PEPPER	CHIME(V)	DRILL		CAT			
		SNIFF(V)	SCREECH(N)	HOE(V)		COW			
		TASTE(V)	BARK(V)	AXE		TAIL			
		CRY(V)	SNORE(N)	HAMMER(V)		LEOPARD			
		FACE	KNOCK(V)	MAKE(V)		TIGER			
		BEAK	HEAR(V)	CONSTRUCT(V)					
		YAWN(N)	LISTEN(V)	REPAIR(V)					
		SPIT(V)	SHOUT(V)	BROOM					
		SMILE(V)	SCREAM(V)	RAKE(V)					
		NOSE	GROWL(V)	PLIERS					
		CRY(N)	CLANG(V)	REPAIR(N)					
		SNEEZE(N)	SNAP(V)	FILE					
		COOK(V)	WHINE(V)	HOE					
		SWALLOW(V)							
		MOUTH							
		TONGUE							

Table B. Words grouped by the cluster analysis of the 'V&V 397 binary semantics' list using linkage command by the 'Ward' method:
Cut off was at 20:

V = action Verb, N = action Noun, O = object

Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7	Node 8	Node 9	Node 10
Aggressive Movement	Cooking	Action	Head	Body Motion	House	Sense	Clothing	Tools	Body Parts
PUNCH(V)	FILL(V)	ESCAPE(N)	CHIN	HICCUP(N)	GATE	TOUCH(N)	SHOE	KNIFE	FINGER
HIT(V)	PAINT(V)	PLACE(V)	GRIN(V)	BURP(N)	FENCE	ACHE(N)	PANTS	TOOTHBRUSH	KNEE
POUND(V)	BOIL(V)	ARRIVAL(N)	DROOL(V)	SMOKE(V)	CEILING	LOOK(V)	HAT	COMB(O)	ELBOW
SLAP(V)	FRY(V)	RUN(V)	FROWN(V)	COUGH(V)	ROOF	FEEL(V)	FEATHER	RAKE(N)	WRIST
KICK(V)	SHAVE(V)	SINK(V)	FROWN(N)	BREATHE(V)	SHIELD	BLEED(V)	COAT	BRUSH(V)	LEG
HIT(N)	ROAST(V)	SHAKE(V)	HEAD	SNEEZE(V)	CURTAIN	NOTICE(V)	BLOUSE	DRILL(V)	THUMB
SLAP(N)	SPOON	THROW(N)	TEETH	YAWN(V)	FLOOR	ITCH(N)	VEST	SAW(N)	SHOULDER
STEP(V)	MIX(V)	CREEP(V)	SENSE(V)	SIGH(N)	DOOR	ACHE(V)	HAIR	CHOP(V)	ANKLE
	FORK(O)	BICYCLE	SMILE(V)	COUGH(N)	BOX	WATCH(V)	BELT	CUT(V)	NECK
	WASH(V)	ESCAPE(V)	LIPS	SNORE(V)	WINDOW	FIND(V)	SOCK	HAMMER(O)	ARM
	EMPTY(V)	STAGGER(V)	LICK(V)	SIGH(V)	WALL	SEE(V)	CARPET	SWORD	HAND
	EAT(V)	SKID(V)	TASTE(V)	HICCUP(V)		DIE(V)	GLOVE	BRUSH(O)	TAIL
	BAKE(V)	BLINK(N)	FACE	BURP(V)		TOUCH(V)	SUIT	SCISSORS	TOE
	STIR(V)	DROP(V)	BEAK	SMELL(V)		ITCH(V)	DRESS	SAW(V)	
	STEAM(V)	TAKE(V)	SPIT(V)	SNIFF(V)			FUR	DRILL(O)	
	SPRAY(V)	FALL(V)	SMILE(V)	YAWN(N)			RUG	HOE(V)	
	POUR(V)	JOG(V)	NOSE	SNORE(N)			SKIRT	AXE(O)	
	DRINK(V)	FLY(V)	FISH	SNEEZE(N)			MITTEN	HAMMER(V)	
	GRILL(V)	CARRY(V)	EYE				SWEATER	BROOM	
	COOK(V)	WANT(V)	EAR				SHIRT	RAKE(V)	
		RETURN (N)	SWALLOW(V)				SCARF	PLIERS	
		ASCEND(V)	MOUTH					FILE(O)	
		WINK(V)	TONGUE					HOE(O)	
		DRIVE(V)							
		SLIDE(V)							
		LOSE(V)							
		WANDER(V)							
		RATTLE(V)							
		RISE(V)							

Table B continued

Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7	Node 8	Node 9	Node 10
Aggressive Movement	Cooking	Action	Head	Body Motion	House	Sense	Clothing	Tools	Body Parts
		ENTER(V)							
		WALK(V)							
		BOUNCE(V)							
		PUSH(N)							
		STAY(V)							
		DRAG(V)							
		PUSH(V)							
		PUT(V)							
		TWIST(V)							
		MARCH(V)							
		ENTRY(N)							
		LEAVE(V)							
		TREMBLE(V)							
		APPROACH(N)							
		DIVE(V)							
		BUMP(V)							
		BLINK(V)							
		BEND(V)							
		PRESS(V)							
		STOP(V)							
		RETURN(V)							
		TICKLE(V)							
		LIFT(V)							
		SWERVE(V)							
		THROW(V)							
		GO(V)							
		LIMP(V)							
		LEAD(V)							
		PAW							
		GREET(V)							
		RECEIVE(V)							
		APPROACH(V)							
		WING							
		WINK(N)							

Table B continued									
Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7	Node 8	Node 9	Node 10
Aggressive Movement	Cooking	Action	Head	Body Motion	House	Sense	Clothing	Tools	Body Parts
		PULL(N)							
		HOLD(V)							
		SIT(V)							
		STAND(V)							
		MOVE(V)							
		COME(V)							
		HALT(V)							
		ARRIVE(V)							
		RIDE(V)							
		STEP(V)							
		TREMBLE(N)							
		TRICYCLE							
		WADE(V)							
		HOP(V)							
		PULL(V)							
		SWIM(V)							
		CHASE(V)							
		TRAVEL(V)							

Table B continued

Node 11	Node 12	Node 13	Node 14	Node 15	Node 16	Node 17	Node 18	Node 19	Node20
Fruits & Vegetables	Transport	Exchanges	Noise	Aggressive Action	Various	Animals	Furniture	Speech	Light
CHERRY	BUS	EXCHANGE(V)	CRACKLE(V)	MURDER(V)	FIX(V)	SHEEP	SOFA	CHIRP(V)	BURN(V)
CELERY	SHIP	SUGGEST(V)	CHIME(N)	BOMB(V)	READ(V)	CAMEL	STOOL	CALL(V)	GLOW(V)
MUSHROOM	RAFT	TEACH(V)	CLATTER(V)	MURDER(N)	BUILD(V)	HORSE	SEAT	TELL(V)	FLICKER(N)
ORANGE	HELICOPTER	GIVE(V)	CLASH(N)	DESTROY(V)	CONSTRUCTION(N)	FOX	TABLE	TALK(V)	FLAME(N)
ONION	TRUCK	PLEAD(V)	CLATTER(N)	BREAK(V)	WRITE(V)	DUCK	CHAIR	WHISPER(V)	SPARKLE(V)
RAISIN	BOAT	TRADE(V)	CLANG(N)	CRASH(V)	PENCIL	DONKEY		SCREAM(N)	FLICKER(V)
BEAN	CAR	DEMAND(N)	KNOCK(N)	CRASH(N)	MAKE(V)	RABBIT		YELL(V)	SHINE(N)
CUCUMBER	VAN	SEND(V)	CRACKLE(N)	DESTRUCTION(N)	PEN	BEAR		ARGUE(V)	GLOW(N)
APPLE	TRAIN	BUY(V)	CHIME(V)	GUN	CONSTRUCT(V)	MOUSE		GROWL(N)	FLAME(V)
PEA		INVITE(V)	KNOCK(V)	KILL(V)	REPAIR(N)	SWAN		WHINE(N)	SPARKLE(N)
CORN		BORROW(V)	CLANG(V)	SMASH(V)	BOOK	LION		SAY(V)	SHINE(V)
GRAPE		PLEA(N)	SNAP(V)	BOMB	DRAW(V)	WOLF		CHAT(V)	FLASH(V)
CABBAGE		LEND(V)			REPAIR(N)	ELEPHANT		YELL(N)	
PEPPER		SELL(V)				BIRD		WHISPER(N)	
LEMON		SUGGESTION(N)				GIRAFFE		SCREECH(V)	
CARROT		ASK(V)				ZEBRA		ADMIT(V)	
POTATO		PAY(V)				GOAT		SHOUT(N)	
PEAR		DEMAND(V)				PIG		SING(V)	
LETTUCE		GET(V)				DOG		CHAT(N)	
PLUM		EXCHANGE(N)				CAT		CRY(V)	
PUMPKIN						COW		SCREECH(N)	
STRAWBERRY						LEOPARD		CALL(N)	
SPINACH						TIGER		WARN(V)	
LIME								BARK(V)	
PEACH								HEAR(V)	
BANANA								LISTEN(V)	
								CHIRP(N)	
								SHOUT(V)	
								SCREAM(V)	
								GROWL(V)	
								CRY(N)	
								SPEAK(V)	
								WHINE(V)	

Table C. Vinson and Vigliocco (2008) semantic fields:

Object Nouns						
Animals	Fruit & Vegetables	Tools	Vehicles	Body Parts	Clothing	Miscellaneous Artifacts
SHEEP	CHERRY	KNIFE	BUS	CHIN	SHOE	GATE
CAMEL	CELERY	TOOTHBRUSH	BICYCLE	FINGER	PANTS	FENCE
HORSE	MUSHROOM	COMB	AIRPLANE	KNEE	HAT	CEILING
FOX	ORANGE	RAKE	SHIP	ELBOW	COAT	SOFA
DUCK	ONION	DRILL	RAFT	FEATHER	BLOUSE	FORK
DONKEY	RAISIN	SPOON	TRUCK	WRIST	VEST	BOMB(V)
RABBIT	BEAN	SHIELD	BOAT	LEG	BELT	ROOF
BEAR	CUCUMBER	SWORD	CAR	HAIR	SOCK	CARPET
MOUSE	APPLE	BRUSH	VAN	HEAD	GLOVE	STOOL
SWAN	PEA	SCISSORS	TRAIN	TEETH	SUIT	SEAT
LION	PINEAPPLE	SAW	TRICYCLE	THUMB	DRESS	CURTAIN
WOLF	CORN	HOE		LIPS	SKIRT	RUG
ELEPHANT	GRAPE	PENCIL		FACE	MITTEN	TABLE
BIRD	CABBAGE	AXE		BEAK	SWEATER	FLOOR
GIRAFFE	PEPPER	HAMMER		FUR	SHIRT	DOOR
ZEBRA	LEMON	GUN		SHOULDER	SCARF	CHAIR
GOAT	CARROT	PEN		PAW		BOOK
PIG	POTATO	BROOM		WING		BOX
DOG	PEAR	PLIERS		ANKLE		WINDOW
CAT	LETTUCE	FILE		NECK		WALL
COW	PLUM			ARM		
FISH	PUMPKIN			NOSE		
LEOPARD	STRAWBERRY			EYE		
TIGER	SPINACH			HAND		
	LIME			TAIL		
	PEACH			TOE		
	BANANA			EAR		
				MOUTH		
				TONGUE		

Table C continued
V=Verbs, N= Nouns

Action Words (Verbs and Nouns)							
Body Actions	Body Sense	Change of Location	Change of State	Noises	Communication	Construction	Contact
HICCUP(N)	LOOK(V)	PLACE(V)	FILL(V)	CRACKLE(V)	SUGGEST(V)	PAINT(V)	PUNCH(V)
TOUCH(N)	FEEL(V)	DROP(V)	SHAKE(V)	CHIRP(V) (animal)	CALL(V)	FIX(V)	HIT(V)
ACHE(N)	NOTICE(V)	CARRY(V)	MIX(V)	CHIME(N)	TELL(V)	BUILD(V)	SLAP(V)
BURP(N)	ACHE(V)	LOSE(V)	EMPTY(V)	CLATTER(V)	TALK(V)	CONSTRUCTION(N)	KNOCK(N)
SMOKE(V)	SENSE(V)	PUSH(N)	STIR(V)	CLASH(N)	TEACH(V)	CONSTRUCT(V)	CRASH(V)
THROW(N)	SMELL(V)	DRAG(V)	TWIST(V)	CLATTER(N)	WHISPER(V)	REPAIR	CRASH(N)
COUGH(V)	SNIFF(V)	PUSH(V)	FIND(V)	CLANG(N)	SCREAM(N)	DRAW(V)	BUMP(V)
BLINK(N)	TASTE(V)	PUT(V)	SPRAY(V)	GROWL(N)	READ(V)	REPAIR(N)	PRESS(V)
BREATHE(V)	SEE(V)	SEND(V)	POUR(V)	SIGH(V)	YELL(V)	MAKE(V)	HIT(N)
GRIN(V)	HEAR(V)	LIFT(V)	BEND(V)	RATTLE(V)	ARGUE(V)		SLAP(N)
SHAVE(V)	LISTEN(V)	PULL(N)		CRACKLE(N)	SIGH(N)		
SNEEZE(V)	TOUCH(V)	MOVE(V)		SCREECH(V)	WRITE(V)		
DROOL(V)		PULL(V)		SING(V)	WHINE(N)		
BLEED(V)				CHIME(V)	SAY(V)		
YAWN(V)				SCREECH(N)	CHAT(V)		
COUGH(N)				BARK(V)(animal)	YELL(N)		
FROWN(V)				CHIRP(N)(animal)	PLEAD(V)		
WINK(V)				GROWL(V)(animal)	WHISPER(N)		
KICK(V)				CLANG(V)	ADMIT(V)		
SNORE(V)				SNAP(V)	SHOUT(N)		
FROWN(N)					DEMAND(N)		
WASH(V)					INVITE(V)		
EAT(V)					CHAT(N)		
HICCUP(V)					PLEA(N)		
ITCH(N)					CALL(N)		
STAY(V)					WARN(V)		

Table C continued							
Action Words (Verbs and Nouns)							
Body Actions	Body Sense	Change of Location	Change of State	Noises	Communication	Construction	Contact
BURP(V)					GREET(V)		
SMILE(N)					SUGGESTION(N)		
WATCH(V)					ASK(V)		
LICK(V)					SHOUT(V)		
CRY(V)					SCREAM(V)		
TREMBLE(V)					CRY(N)		
YAWN(N)					DEMAND(V)		
BLINK(V)					SPEAK(V)		
SPIT(V)					WHINE(V)		
DRINK(V)							
SMILE(V)							
TICKLE(V)							
SNORE(N)							
KNOCK(V)							
THROW(V)							
WINK(N)							
HOLD(V)							
SIT(V)							
STAND(V)							
DIE(V)							
TREMBLE(N)							
SNEEZE(N)							
ITCH(V)							
SWALLOW(V)							

Table C continued

Action Words (Verbs and Nouns)						
Cooking	Destruction	Exchange	Heat/Light Emission	Motion Direction	Motion manner	Tool Action
BOIL(V)	MURDER(V)	EXCHANGE(V)	BURN(V)	ESCAPE(N)	RUN(V)	BRUSH(V)
FRY(V)	CHOP(V)	TAKE(V)	GLOW(V)	ARRIVAL(N)	CREEP(V)	POUND(V)
ROAST(V)	MURDER(N)	WANT(V)	FLICKER(N)	SINK(V)	STAGGER(V)	SAW(V)
BAKE(V)	DESTROY(V)	GIVE(V)	FLAME(N)	ESCAPE(V)	SKID(V)	CUT(V)
STEAM(V)	BREAK(V)	BUY(V)	SPARKLE(V)	FALL(V)	JOG(V)	HAMMER(V)
GRILL(V)	DESTRUCTION(N)	BORROW(V)	FLICKER(V)	RETURN(N)	FLY(V)	DRILL(V)
COOK(V)	KILL(V)	LEND(V)	SHINE(N)	ASCEND(V)	DRIVE(V)	RAKE(V)
	SMASH(V)	RECEIVE(V)	GLOW(N)	ASCENT(N)	SLIDE(V)	HOE(V)
	BOMB	SELL	FLAME(V)	RISE(V)	WANDER(V)	
		PAY(V)	SPARKLE(N)	ENTER(V)	WALK(V)	
		GET(V)	SHINE(V)	ENTRY(N)	BOUNCE(V)	
		EXCHANGE(N)	FLASH(V)	LEAVE(V)	MARCH(V)	
				APPROACH(N)	DIVE(V)	
				APPROACH(V)	PEDAL(V)	
				RETURN(V)	STOP(V)	
				GO(V)	SWERVE(V)	
				LEAD(V)	LIMP(V)	
				APPROACH(N)	HALT(V)	
				COME(V)	RIDE(V)	
				ARRIVE(V)	WADE(V)	
				FOLLOW(V)	HOP(V)	
					SWIM(V)	
					STEP(V)	
					CHASE(V)	
					TRAVEL(V)	

Appendix 5

The 57 word training set semantic features list

Times feature used			47	0	0	0	8	1	1	0
			Limbs							
			1	2	3	4	5	6	7	8
Word	**Number of features**	**Frequency**	**Limbs-0**	**limbs-1**	**limbs-2**	**limbs-3**	**limbs-4**	**limbs-5**	**limbs-6**	**Limbs>6**
apple	15	1	1	0	0	0	0	0	0	0
banana	13	1	1	0	0	0	0	0	0	0
carrot	14	1	1	0	0	0	0	0	0	0
celery	14	1	1	0	0	0	0	0	0	0
cucumber	15	1	1	0	0	0	0	0	0	0
lettuce	14	1	1	0	0	0	0	0	0	0
lime	14	1	1	0	0	0	0	0	0	0
nut	15	1	1	0	0	0	0	0	0	0
potato	15	1	1	0	0	0	0	0	0	0
pumpkin	13	1	1	0	0	0	0	0	0	0
watermelon	16	1	1	0	0	0	0	0	0	0
cow	17	1	0	0	0	0	1	0	0	0
dog	19	1	0	0	0	0	1	0	0	0
dolphin	16	1	1	0	0	0	0	0	0	0
eagle	16	1	0	0	0	0	1	0	0	0
elephant	18	1	0	0	0	0	0	1	0	0
fox	17	1	0	0	0	0	1	0	0	0
grasshopper	13	1	0	0	0	0	0	0	1	0
horse	18	1	0	0	0	0	1	0	0	0
pig	17	1	0	0	0	0	1	0	0	0
squirrel	16	1	0	0	0	0	1	0	0	0
tiger	17	1	0	0	0	0	1	0	0	0
egg	15	1	1	0	0	0	0	0	0	0
bread	13	1	1	0	0	0	0	0	0	0
butter	12	1	1	0	0	0	0	0	0	0
cheese	12	1	1	0	0	0	0	0	0	0
milk	10	1	1	0	0	0	0	0	0	0
aeroplane	17	1	1	0	0	0	0	0	0	0
bicycle	13	1	1	0	0	0	0	0	0	0
bus	17	1	1	0	0	0	0	0	0	0
canoe	11	1	1	0	0	0	0	0	0	0
car	17	1	1	0	0	0	0	0	0	0
helicopter	17	1	1	0	0	0	0	0	0	0
motorcycle	15	1	1	0	0	0	0	0	0	0
sailboat	13	1	1	0	0	0	0	0	0	0
train	16	1	1	0	0	0	0	0	0	0
truck	17	1	1	0	0	0	0	0	0	0
blender	15	1	1	0	0	0	0	0	0	0
bowl	12	1	1	0	0	0	0	0	0	0
broom	14	1	1	0	0	0	0	0	0	0
cup	12	1	1	0	0	0	0	0	0	0
fork	12	1	1	0	0	0	0	0	0	0
pan	12	1	1	0	0	0	0	0	0	0
plate	12	1	1	0	0	0	0	0	0	0
pot	12	1	1	0	0	0	0	0	0	0
spoon	11	1	1	0	0	0	0	0	0	0
toaster	12	1	1	0	0	0	0	0	0	0
chisel	14	1	1	0	0	0	0	0	0	0
drill	13	1	1	0	0	0	0	0	0	0
hammer	15	1	1	0	0	0	0	0	0	0
nails	12	1	1	0	0	0	0	0	0	0
pliers	12	1	1	0	0	0	0	0	0	0
sandpaper	10	1	1	0	0	0	0	0	0	0
saw	12	1	1	0	0	0	0	0	0	0
screwdriver	13	1	1	0	0	0	0	0	0	0
shovel	12	1	1	0	0	0	0	0	0	0
wrench	10	1	1	0	0	0	0	0	0	0

Times feature used	1	12	4	5	7	5	3	3	12	5	1	16
	Colour											
	1	2	3	4	5	6	7	8	9	10	11	12
Word	**blue**	**brown**	**red**	**yellow**	**green**	**white**	**black**	**orange**	**grey**	**silver**	**pink**	**variable-colour**
apple	0	0	0	0	1	0	0	0	0	0	0	0
banana	0	0	0	1	0	0	0	0	0	0	0	0
carrot	0	0	0	0	0	0	0	1	0	0	0	0
celery	0	0	0	0	1	0	0	0	0	0	0	0
cucumber	0	0	0	0	1	1	0	0	0	0	0	0
lettuce	0	0	0	0	1	0	0	0	0	0	0	0
lime	0	0	0	0	1	0	0	0	0	0	0	0
nut	0	1	0	0	0	0	0	0	0	0	0	0
potato	0	1	0	0	0	1	0	0	0	0	0	0
pumpkin	0	0	0	0	0	0	0	1	0	0	0	0
watermelon	0	0	1	0	1	0	0	0	0	0	0	0
cow	0	1	0	0	0	0	0	0	0	0	0	0
dog	0	1	0	0	0	0	0	0	0	0	0	0
dolphin	1	0	0	0	0	0	0	0	0	0	0	0
eagle	0	1	0	0	0	0	0	0	0	0	0	0
elephant	0	0	0	0	0	0	0	0	1	0	0	0
fox	0	1	1	0	0	0	0	0	0	0	0	0
grasshopper	0	0	0	0	1	0	0	0	0	0	0	0
horse	0	0	0	0	0	0	1	0	0	0	0	0
pig	0	0	0	0	0	0	0	0	0	0	1	0
squirrel	0	0	1	0	0	0	0	0	1	0	0	0
tiger	0	0	0	1	0	0	1	0	0	0	0	0
egg	0	0	0	0	0	1	0	1	0	0	0	0
bread	0	1	0	0	0	1	0	0	0	0	0	0
butter	0	0	0	1	0	0	0	0	0	0	0	0
cheese	0	0	0	1	0	0	0	0	0	0	0	0
milk	0	0	0	0	0	1	0	0	0	0	0	0
aeroplane	0	0	0	0	0	0	0	0	1	0	0	1
bicycle	0	0	0	0	0	0	0	0	0	0	0	1
bus	0	0	1	1	0	0	0	0	0	0	0	0
canoe	0	0	0	0	0	0	0	0	0	0	0	1
car	0	0	0	0	0	0	0	0	0	0	0	1
helicopter	0	0	0	0	0	0	0	0	1	0	0	1
motorcycle	0	0	0	0	0	0	0	0	0	0	0	1
sailboat	0	0	0	0	0	0	0	0	0	0	0	1
train	0	0	0	0	0	0	0	0	0	0	0	1
truck	0	0	0	0	0	0	0	0	0	0	0	1
blender	0	0	0	0	0	0	0	0	0	0	0	1
bowl	0	0	0	0	0	0	0	0	0	0	0	1
broom	0	1	0	0	0	0	0	0	0	0	0	0
cup	0	0	0	0	0	0	0	0	0	0	0	1
fork	0	0	0	0	0	0	0	0	0	1	0	0
pan	0	0	0	0	0	0	0	0	0	1	0	0
plate	0	0	0	0	0	0	0	0	0	0	0	1
pot	0	0	0	0	0	0	0	0	0	1	0	0
spoon	0	0	0	0	0	0	0	0	0	1	0	0
toaster	0	0	0	0	0	0	0	0	0	1	0	0
chisel	0	1	0	0	0	0	0	0	1	0	0	0
drill	0	0	0	0	0	0	0	0	0	0	0	1
hammer	0	1	0	0	0	0	0	0	1	0	0	0
nails	0	0	0	0	0	0	0	0	1	0	0	0
pliers	0	0	0	0	0	0	0	0	1	0	0	1
sandpaper	0	1	0	0	0	0	1	0	0	0	0	0
saw	0	1	0	0	0	0	0	0	1	0	0	0
screwdriver	0	0	0	0	0	0	0	0	1	0	0	1
shovel	0	0	0	0	0	0	0	0	1	0	0	0
wrench	0	0	0	0	0	0	0	0	1	0	0	0

Times feature used	**29**	**10**	**17**	**35**	**14**
	size			**Cross-section**	
	1	**2**	**3**	**1**	**2**
Word	**size<1**	**size1->2**	**size>2**	**XS-circular**	**XS-rectangle**
apple	1	0	0	1	0
banana	1	0	0	1	0
carrot	1	0	0	1	0
celery	1	0	0	1	0
cucumber	0	1	0	1	0
lettuce	1	0	0	1	0
lime	1	0	0	1	0
nut	1	0	0	1	0
potato	1	0	0	1	0
pumpkin	0	1	0	1	0
watermelon	0	1	0	1	0
cow	0	0	1	1	0
dog	0	1	0	1	0
dolphin	0	0	1	1	0
eagle	0	1	0	1	0
elephant	0	0	1	1	0
fox	0	1	0	1	0
grasshopper	1	0	0	1	0
horse	0	0	1	1	0
pig	0	0	1	1	0
squirrel	1	0	0	1	0
tiger	0	0	1	1	0
egg	1	0	0	1	0
bread	1	0	0	0	1
butter	1	0	0	0	1
cheese	1	0	0	0	1
milk	0	0	0	0	0
aeroplane	0	0	1	1	0
bicycle	0	0	1	0	0
bus	0	0	1	0	1
canoe	0	0	1	1	0
car	0	0	1	0	1
helicopter	0	0	1	1	0
motorcycle	0	0	1	0	1
sailboat	0	0	1	1	0
train	0	0	1	0	1
truck	0	0	1	0	1
blender	1	0	0	1	0
bowl	1	0	0	0	1
broom	0	0	1	1	0
cup	1	0	0	1	0
fork	1	0	0	0	0
pan	0	1	0	0	1
plate	1	0	0	0	0
pot	1	0	0	0	1
spoon	1	0	0	0	0
toaster	1	0	0	0	1
chisel	1	0	0	0	1
drill	1	0	0	1	0
hammer	0	1	0	1	0
nails	1	0	0	1	0
pliers	1	0	0	0	0
sandpaper	1	0	0	0	0
saw	0	1	0	0	0
screwdriver	1	0	0	1	0
shovel	0	1	0	1	0
wrench	1	0	0	0	1

Times feature used	12	12	2	8	12	2	2	8
	Form							
	1	2	3	4	5	6	7	8
Word	**cylindrical**	**round**	**square**	**triangular**	**rectangular**	**runny**	**winged**	**4-legged**
apple	0	1	0	0	0	0	0	0
banana	1	0	0	0	0	0	0	0
carrot	1	0	0	0	0	0	0	0
celery	1	0	0	0	0	0	0	0
cucumber	1	0	0	0	0	0	0	0
lettuce	0	1	0	0	0	0	0	0
lime	0	1	0	0	0	0	0	0
nut	0	1	0	0	0	0	0	0
potato	0	1	0	0	0	0	0	0
pumpkin	0	1	0	0	0	0	0	0
watermelon	1	0	0	0	0	0	0	0
cow	0	0	0	0	0	0	0	1
dog	0	0	0	0	0	0	0	1
dolphin	1	0	0	0	0	0	0	0
eagle	0	0	0	0	0	0	1	0
elephant	0	0	0	0	0	0	0	1
fox	0	0	0	0	0	0	0	1
grasshopper	0	0	0	0	0	0	1	0
horse	0	0	0	0	0	0	0	1
pig	0	0	0	0	0	0	0	1
squirrel	0	0	0	0	0	0	0	1
tiger	0	0	0	0	0	0	0	1
egg	0	1	0	0	0	1	0	0
bread	0	0	0	0	1	0	0	0
butter	0	0	1	0	0	0	0	0
cheese	0	0	1	0	0	0	0	0
milk	0	0	0	0	0	1	0	0
aeroplane	1	0	0	0	0	0	0	0
bicycle	0	0	0	0	1	0	0	0
bus	0	0	0	0	1	0	0	0
canoe	1	0	0	0	0	0	0	0
Car	0	0	0	0	1	0	0	0
helicopter	0	1	0	0	0	0	0	0
motorcycle	0	0	0	0	1	0	0	0
sailboat	0	0	0	1	0	0	0	0
train	0	0	0	0	1	0	0	0
truck	0	0	0	0	1	0	0	0
blender	1	0	0	0	0	0	0	0
bowl	0	1	0	0	0	0	0	0
broom	0	0	0	1	0	0	0	0
cup	1	0	0	0	0	0	0	0
fork	0	0	0	1	0	0	0	0
pan	0	1	0	0	0	0	0	0
plate	0	1	0	0	0	0	0	0
pot	0	1	0	0	0	0	0	0
spoon	0	0	0	1	0	0	0	0
toaster	0	0	0	0	1	0	0	0
chisel	0	0	0	0	1	0	0	0
drill	1	0	0	0	0	0	0	0
hammer	0	0	0	1	0	0	0	0
nails	0	0	0	1	0	0	0	0
pliers	0	0	0	1	0	0	0	0
sandpaper	0	0	0	0	1	0	0	0
saw	0	0	0	0	1	0	0	0
screwdriver	1	0	0	0	0	0	0	0
shovel	0	0	0	1	0	0	0	0
wrench	0	0	0	0	1	0	0	0

Times feature used	5	23	1	7	13	12	15	12
	Composition							
	1	2	3	4	5	6	7	8
Word	**wood**	**metal**	**liquid**	**glass**	**other_manmade**	**plastic**	**from_animal**	**from_plant**
apple	0	0	0	0	0	0	0	1
banana	0	0	0	0	0	0	0	1
carrot	0	0	0	0	0	0	0	1
celery	0	0	0	0	0	0	0	1
cucumber	0	0	0	0	0	0	0	1
lettuce	0	0	0	0	0	0	0	1
lime	0	0	0	0	0	0	0	1
nut	0	0	0	0	0	0	0	1
potato	0	0	0	0	0	0	0	1
pumpkin	0	0	0	0	0	0	0	1
watermelon	0	0	0	0	0	0	0	1
cow	0	0	0	0	0	0	1	0
dog	0	0	0	0	0	0	1	0
dolphin	0	0	0	0	0	0	1	0
eagle	0	0	0	0	0	0	1	0
elephant	0	0	0	0	0	0	1	0
fox	0	0	0	0	0	0	1	0
grasshopper	0	0	0	0	0	0	1	0
horse	0	0	0	0	0	0	1	0
pig	0	0	0	0	0	0	1	0
squirrel	0	0	0	0	0	0	1	0
tiger	0	0	0	0	0	0	1	0
egg	0	0	0	0	0	0	1	0
bread	0	0	0	0	0	0	0	1
butter	0	0	0	0	0	0	1	0
cheese	0	0	0	0	0	0	1	0
milk	0	0	1	0	0	0	1	0
aeroplane	0	1	0	1	1	1	0	0
bicycle	0	1	0	0	1	0	0	0
bus	0	1	0	1	1	1	0	0
canoe	0	0	0	0	1	0	0	0
car	0	1	0	1	1	1	0	0
helicopter	0	1	0	1	1	1	0	0
motorcycle	0	1	0	0	1	1	0	0
sailboat	1	0	0	0	1	0	0	0
train	0	1	0	1	1	1	0	0
truck	0	1	0	1	1	1	0	0
blender	0	1	0	0	0	1	0	0
bowl	0	0	0	1	0	0	0	0
broom	1	0	0	0	0	1	0	0
cup	0	0	0	0	1	0	0	0
fork	0	1	0	0	0	0	0	0
pan	0	1	0	0	0	0	0	0
plate	0	0	0	0	1	0	0	0
pot	0	1	0	0	0	0	0	0
spoon	0	1	0	0	0	0	0	0
toaster	0	1	0	0	0	0	0	0
chisel	1	1	0	0	0	0	0	0
drill	0	1	0	0	0	1	0	0
hammer	1	1	0	0	0	0	0	0
nails	0	1	0	0	0	0	0	0
pliers	0	1	0	0	0	1	0	0
sandpaper	0	0	0	0	1	0	0	0
saw	1	1	0	0	0	0	0	0
screwdriver	0	1	0	0	0	1	0	0
shovel	0	1	0	0	0	0	0	0
wrench	0	1	0	0	0	0	0	0

Times feature used	**8**	**39**	**1**	**6**	**1**	**2**	**9**	**18**	**6**	**26**
	Texture							**Noise**		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**1**	**2**	**3**
Word	**rough**	**smooth**	**feathered**	**hairy**	**leafy**	**wet**	**sharp**	**makes_own**	**no_input**	**can't_make_ noise**
apple	0	1	0	0	0	0	0	0	0	1
banana	0	1	0	0	0	0	0	0	0	1
carrot	1	0	0	0	0	0	0	0	0	1
celery	0	1	0	0	0	0	0	0	0	1
cucumber	0	1	0	0	0	0	0	0	0	1
lettuce	0	0	0	0	1	0	0	0	0	1
lime	1	0	0	0	0	0	0	0	0	1
nut	0	1	0	0	0	0	0	0	1	1
potato	1	0	0	0	0	0	0	0	0	1
pumpkin	0	1	0	0	0	0	0	0	0	1
watermelon	0	1	0	0	0	0	0	0	0	1
cow	0	0	0	1	0	0	0	1	0	0
dog	0	0	0	1	0	0	0	1	0	0
dolphin	0	1	0	0	0	0	0	1	0	0
eagle	0	0	1	0	0	0	0	1	0	0
elephant	1	0	0	0	0	0	0	1	0	0
fox	0	0	0	1	0	0	0	1	0	0
grasshopper	0	0	0	0	0	0	0	1	0	0
horse	0	0	0	1	0	0	0	1	0	0
pig	1	0	0	0	0	0	0	1	0	0
squirrel	0	0	0	1	0	0	0	1	0	0
tiger	0	0	0	1	0	0	0	1	0	0
egg	0	1	0	0	0	1	0	0	0	1
bread	1	0	0	0	0	0	0	0	0	1
butter	0	1	0	0	0	0	0	0	0	1
cheese	0	1	0	0	0	0	0	0	0	1
milk	0	0	0	0	0	1	0	0	0	1
aeroplane	0	1	0	0	0	0	0	1	0	0
bicycle	0	1	0	0	0	0	0	0	0	0
bus	0	1	0	0	0	0	0	1	0	0
canoe	0	1	0	0	0	0	0	0	0	0
car	0	1	0	0	0	0	0	1	0	0
helicopter	0	1	0	0	0	0	0	1	0	0
motorcycle	0	1	0	0	0	0	0	1	0	0
sailboat	0	1	0	0	0	0	0	0	0	0
train	0	1	0	0	0	0	0	1	0	0
truck	0	1	0	0	0	0	0	1	0	0
blender	0	1	0	0	0	0	1	0	1	1
bowl	0	1	0	0	0	0	0	0	0	1
broom	1	0	0	0	0	0	0	0	0	1
cup	0	1	0	0	0	0	0	0	0	1
fork	0	1	0	0	0	0	1	0	0	1
pan	0	1	0	0	0	0	0	0	0	1
plate	0	1	0	0	0	0	0	0	1	1
pot	0	1	0	0	0	0	0	0	0	1
spoon	0	1	0	0	0	0	0	0	0	1
toaster	0	1	0	0	0	0	0	0	0	1
chisel	0	1	0	0	0	0	1	0	1	0
drill	0	1	0	0	0	0	1	0	1	0
hammer	0	1	0	0	0	0	0	0	1	0
nails	0	1	0	0	0	0	1	0	0	0
pliers	0	1	0	0	0	0	1	0	0	0
sandpaper	1	0	0	0	0	0	0	0	0	0
saw	0	1	0	0	0	0	1	0	0	0
screwdriver	0	1	0	0	0	0	1	0	0	0
shovel	0	1	0	0	0	0	1	0	0	0
wrench	0	1	0	0	0	0	0	0	0	0

Times feature used	1	18	16	6	11	13	9	1	10	10
	Function									
	1	2	3	4	5	6	7	8	9	10
Word	pet	food	makes_ waste	esthetic	reproduces	Transpor-tation	cooking/ eating	cleaning	warmth	repairs
apple	0	1	0	0	0	0	0	0	0	0
banana	0	1	0	0	0	0	0	0	0	0
carrot	0	1	0	0	0	0	0	0	0	0
celery	0	1	0	0	0	0	0	0	0	0
cucumber	0	1	0	0	0	0	0	0	0	0
lettuce	0	1	0	0	0	0	0	0	0	0
lime	0	1	0	0	0	0	0	0	0	0
nut	0	1	0	0	0	0	0	0	0	0
potato	0	1	0	0	0	0	0	0	0	0
pumpkin	0	1	0	0	0	0	0	0	0	0
watermelon	0	1	0	1	0	0	0	0	0	0
cow	0	1	1	0	1	1	0	0	1	0
dog	1	0	1	0	1	0	0	0	1	0
dolphin	0	0	1	1	1	0	0	0	1	0
eagle	0	0	1	1	1	0	0	0	1	0
elephant	0	0	1	1	1	1	0	0	1	0
fox	0	0	1	0	1	0	0	0	1	0
grasshopper	0	0	0	0	1	0	0	0	0	0
horse	0	0	1	0	1	1	0	0	1	0
pig	0	1	1	0	1	0	0	0	1	0
squirrel	0	0	0	0	1	0	0	0	1	0
tiger	0	0	1	1	1	0	0	0	1	0
egg	0	1	0	0	0	0	0	0	0	0
bread	0	1	0	0	0	0	0	0	0	0
butter	0	1	0	0	0	0	0	0	0	0
cheese	0	1	0	0	0	0	0	0	0	0
milk	0	1	0	0	0	0	0	0	0	0
aeroplane	0	0	1	0	0	1	0	0	0	0
bicycle	0	0	0	0	0	1	0	0	0	0
bus	0	0	1	0	0	1	0	0	0	0
canoe	0	0	0	0	0	1	0	0	0	0
car	0	0	1	0	0	1	0	0	0	0
helicopter	0	0	1	0	0	1	0	0	0	0
motorcycle	0	0	1	0	0	1	0	0	0	0
sailboat	0	0	0	1	0	1	0	0	0	0
train	0	0	1	0	0	1	0	0	0	0
truck	0	0	1	0	0	1	0	0	0	0
blender	0	0	0	0	0	0	1	0	0	0
bowl	0	0	0	0	0	0	1	0	0	0
broom	0	0	0	0	0	0	0	1	0	0
cup	0	0	0	0	0	0	1	0	0	0
fork	0	0	0	0	0	0	1	0	0	0
pan	0	0	0	0	0	0	1	0	0	0
plate	0	0	0	0	0	0	1	0	0	0
pot	0	0	0	0	0	0	1	0	0	0
spoon	0	0	0	0	0	0	1	0	0	0
toaster	0	0	0	0	0	0	1	0	0	0
chisel	0	0	0	0	0	0	0	0	0	1
drill	0	0	0	0	0	0	0	0	0	1
hammer	0	0	0	0	0	0	0	0	0	1
nails	0	0	0	0	0	0	0	0	0	1
pliers	0	0	0	0	0	0	0	0	0	1
sandpaper	0	0	0	0	0	0	0	0	0	1
saw	0	0	0	0	0	0	0	0	0	1
Screw-driver	0	0	0	0	0	0	0	0	0	1
shovel	0	0	0	0	0	0	0	0	0	1
wrench	0	0	0	0	0	0	0	0	0	1

Times feature used	1	1	26	13	10	8	9	2	9	3	3
	Location										
	1	2	3	4	5	6	7	8	9	10	11
Word	bedroom	living-room	kitchen	basement	yard	road	woods	jungle	farm	sea	air
apple	0	0	1	0	1	0	1	0	0	0	0
banana	0	0	1	0	0	0	0	0	0	0	0
carrot	0	0	1	0	1	0	0	0	0	0	0
celery	0	0	1	0	1	0	0	0	0	0	0
cucumber	0	0	1	0	1	0	0	0	0	0	0
lettuce	0	0	1	0	1	0	0	0	0	0	0
lime	0	0	1	0	0	0	1	0	0	0	0
nut	0	0	1	0	0	0	1	0	0	0	0
potato	0	0	1	0	1	0	0	0	0	0	0
pumpkin	0	0	1	0	0	0	0	0	0	0	0
watermelon	1	0	1	0	0	0	0	0	0	0	0
cow	0	0	0	0	0	0	0	0	1	0	0
dog	0	1	0	0	1	0	1	0	1	0	0
dolphin	0	0	0	0	0	0	0	0	0	1	0
eagle	0	0	0	0	0	0	0	0	0	0	1
elephant	0	0	0	0	0	0	1	1	0	0	0
fox	0	0	0	0	0	0	1	0	1	0	0
grasshopper	0	0	0	0	1	0	1	0	0	0	0
horse	0	0	0	0	0	1	1	0	1	0	0
pig	0	0	1	0	0	0	0	0	1	0	0
squirrel	0	0	0	0	0	1	1	0	0	0	0
tiger	0	0	0	0	0	0	0	1	0	0	0
egg	0	0	1	0	0	0	0	0	0	0	0
bread	0	0	1	0	0	0	0	0	0	0	0
butter	0	0	1	0	0	0	0	0	0	0	0
cheese	0	0	1	0	0	0	0	0	0	0	0
milk	0	0	1	0	0	0	0	0	0	0	0
aeroplane	0	0	0	0	0	0	0	0	0	0	1
bicycle	0	0	0	1	1	1	0	0	0	0	0
bus	0	0	0	0	0	1	0	0	0	0	0
canoe	0	0	0	0	0	0	0	0	0	1	0
car	0	0	0	1	0	1	0	0	0	0	0
helicopter	0	0	0	0	0	0	0	0	0	0	1
motorcycle	0	0	0	0	0	1	0	0	0	0	0
sailboat	0	0	0	0	0	0	0	0	0	1	0
train	0	0	0	0	0	1	0	0	0	0	0
truck	0	0	0	0	0	1	0	0	1	0	0
blender	0	0	1	0	0	0	0	0	0	0	0
bowl	0	0	1	0	0	0	0	0	0	0	0
broom	0	0	0	1	1	0	0	0	0	0	0
cup	0	0	1	0	0	0	0	0	0	0	0
fork	0	0	1	0	0	0	0	0	0	0	0
pan	0	0	1	0	0	0	0	0	0	0	0
plate	0	0	1	0	0	0	0	0	0	0	0
pot	0	0	1	0	0	0	0	0	0	0	0
spoon	0	0	1	0	0	0	0	0	0	0	0
toaster	0	0	1	0	0	0	0	0	0	0	0
chisel	0	0	0	1	0	0	0	0	0	0	0
drill	0	0	0	1	0	0	0	0	0	0	0
hammer	0	0	0	1	0	0	0	0	1	0	0
nails	0	0	0	1	0	0	0	0	1	0	0
pliers	0	0	0	1	0	0	0	0	0	0	0
sandpaper	0	0	0	1	0	0	0	0	0	0	0
saw	0	0	0	1	0	0	0	0	0	0	0
screwdriver	0	0	0	1	0	0	0	0	0	0	0
shovel	0	0	0	1	0	0	0	0	1	0	0
wrench	0	0	0	1	0	0	0	0	0	0	0

Times feature used	**11**	**7**	**21**	**8**	**22**
	Movement				
	1	**2**	**3**	**4**	**5**
Word	**self**	**propelled**	**not_moveable**	**with_fuel**	**grows**
apple	0	0	1	0	1
banana	0	0	1	0	1
carrot	0	0	1	0	1
celery	0	0	1	0	1
cucumber	0	0	1	0	1
lettuce	0	0	1	0	1
lime	0	0	1	0	1
nut	0	0	1	0	1
potato	0	0	1	0	1
pumpkin	0	0	1	0	1
watermelon	0	0	1	0	1
cow	1	0	0	0	1
dog	1	0	0	0	1
dolphin	1	0	0	0	1
eagle	1	0	0	0	1
elephant	1	0	0	0	1
fox	1	0	0	0	1
grasshopper	1	0	0	0	1
horse	1	0	0	0	1
pig	1	0	0	0	1
squirrel	1	0	0	0	1
tiger	1	0	0	0	1
egg	0	0	1	0	0
bread	0	0	1	0	0
butter	0	0	1	0	0
cheese	0	0	1	0	0
milk	0	0	0	0	0
aeroplane	0	0	0	1	0
bicycle	0	1	0	0	0
bus	0	0	0	1	0
canoe	0	1	0	0	0
car	0	0	0	1	0
helicopter	0	0	0	1	0
motorcycle	0	0	0	1	0
sailboat	0	1	0	0	0
train	0	0	0	1	0
truck	0	0	0	1	0
blender	0	0	0	1	0
bowl	0	0	1	0	0
broom	0	1	0	0	0
cup	0	0	1	0	0
fork	0	1	0	0	0
pan	0	0	1	0	0
plate	0	0	1	0	0
pot	0	0	1	0	0
spoon	0	1	0	0	0
toaster	0	0	1	0	0
chisel	0	0	0	0	0
drill	0	0	0	0	0
hammer	0	1	0	0	0
nails	0	0	0	0	0
pliers	0	0	0	0	0
sandpaper	0	0	0	0	0
saw	0	0	0	0	0
screwdriver	0	0	0	0	0
shovel	0	0	0	0	0
wrench	0	0	0	0	0

Appendix 6

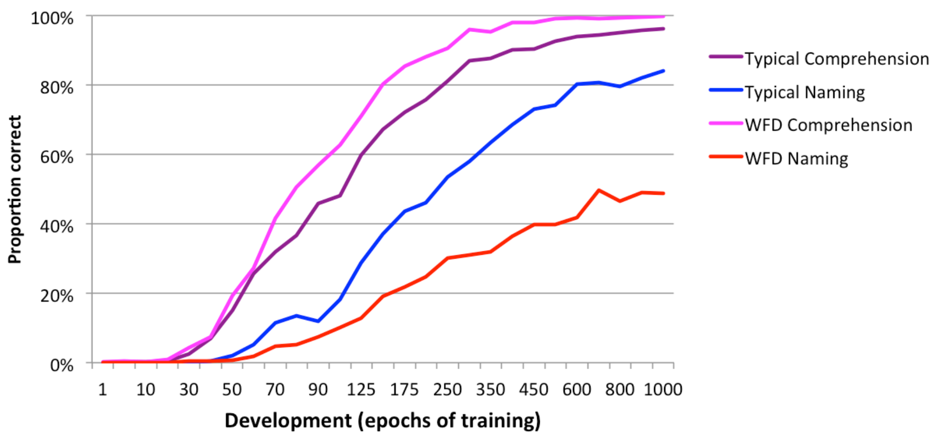
Replications of Graphs in Chapter 8

The graphs shown in Figures 8.1, 8.2, 8.3, 8.4 and 8.5 in chapter 8 were averaged over three replications with different random seeds for each condition.

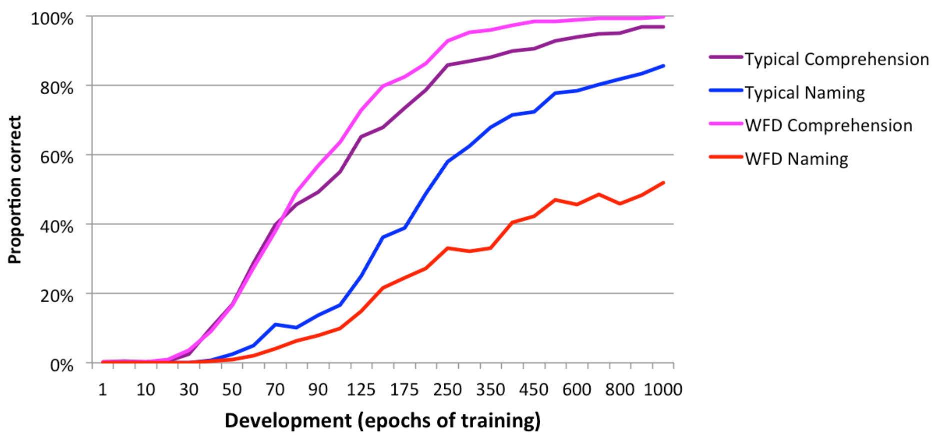
The Figures below present the results graphs of each of the three replications which correspond to Figures 8.1, 8.2, 8.3, 8.4 and 8.5.

Replications for Figure 8.1. Developmental trajectories of naming and comprehension for TD and WFD model simulations.

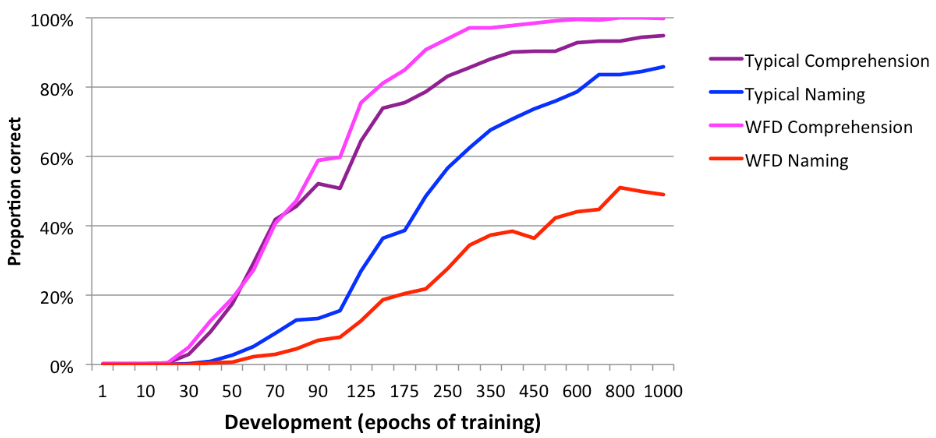
Replication 1



Replication 2



Replication 3



Replications for Figure 8.2. The effect of intervention to remediate weakness on naming accuracy. Intervention began at 100 epochs of training and lasted for 100 epochs.

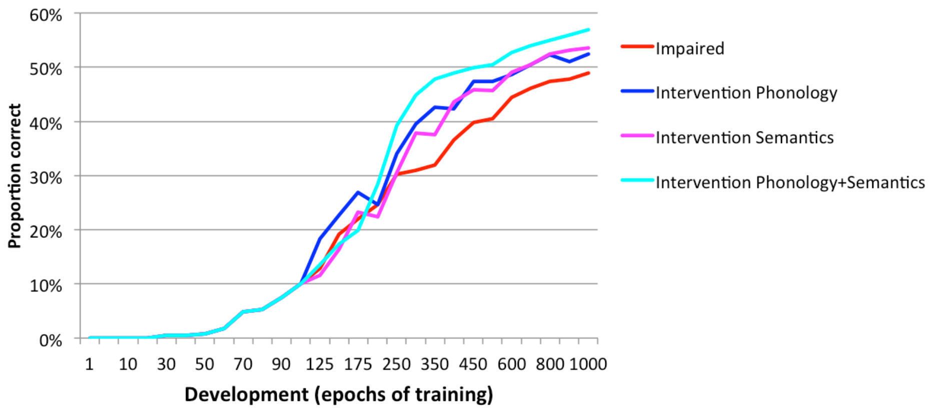
Replication 1

Replication 2

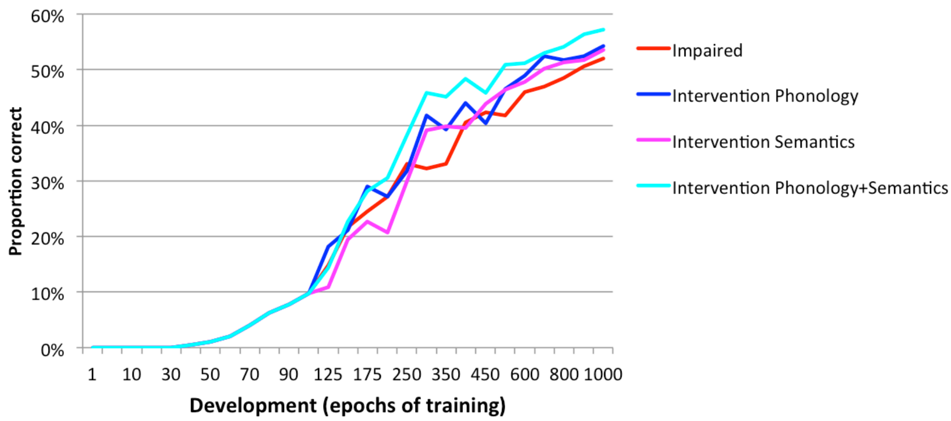
Replication 3

Replications for Figure 8.3. The effect on naming accuracy of intervention to improve strengths. Intervention began at 100 epochs of training and lasted for 100 epochs.

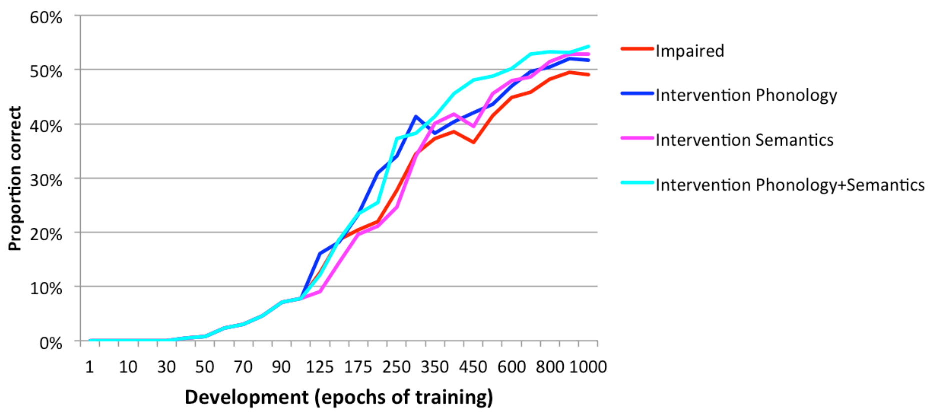
Replication 1



Replication 2

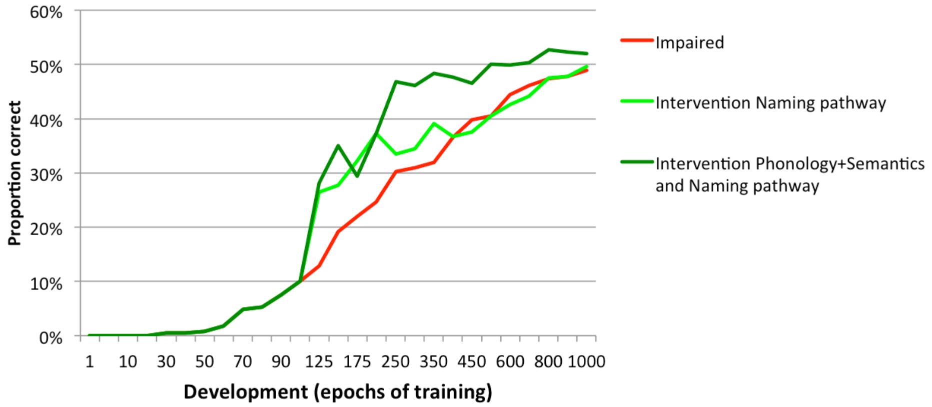


Replication 3

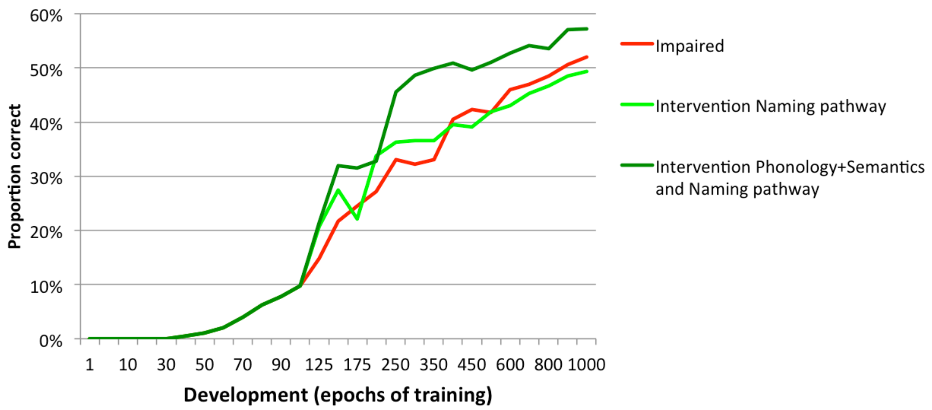


Replications for Figure 8.4. The effects of both interventions to remediate weakness and improve strengths. Intervention began at 100 epochs of training and lasted for 100 epochs.

Replication 1



Replication 2



Replication 3

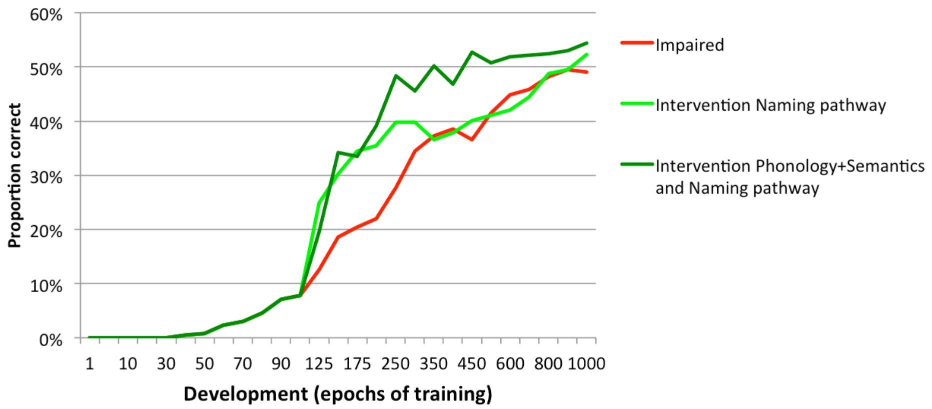
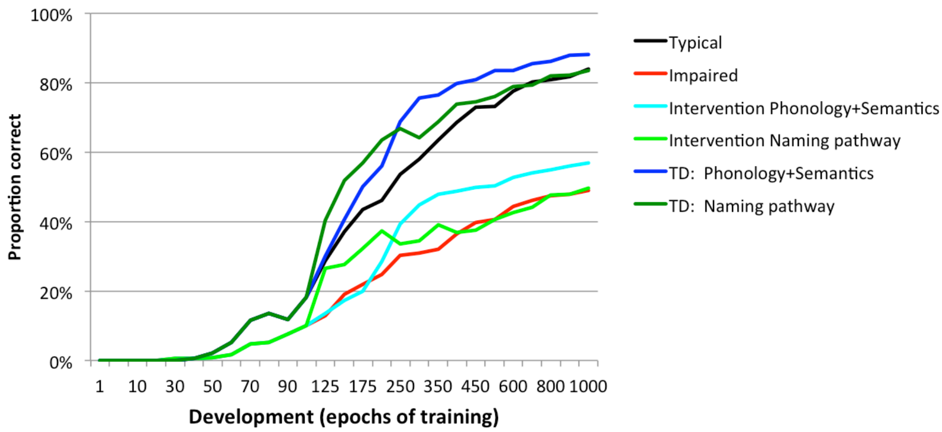
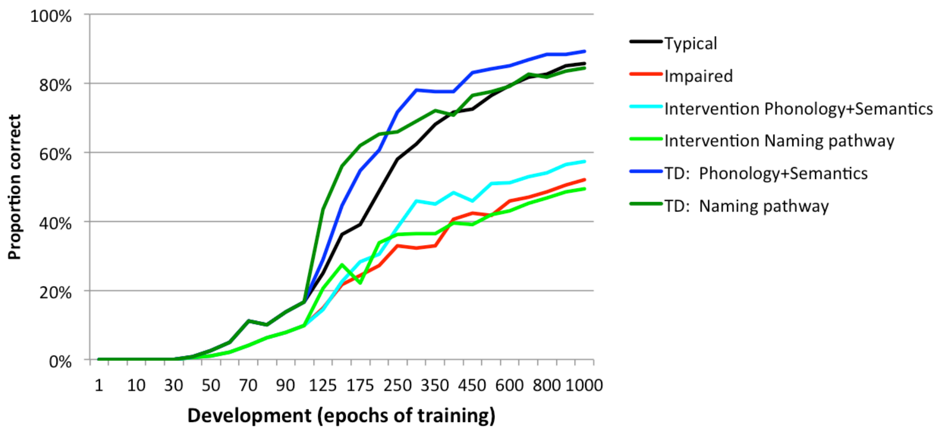


Figure 8.5. Comparison of the effects of interventions to remediate weakness and improve strengths on TD and WFD models. Intervention began at 100 epochs of training and lasted for 100 epochs.

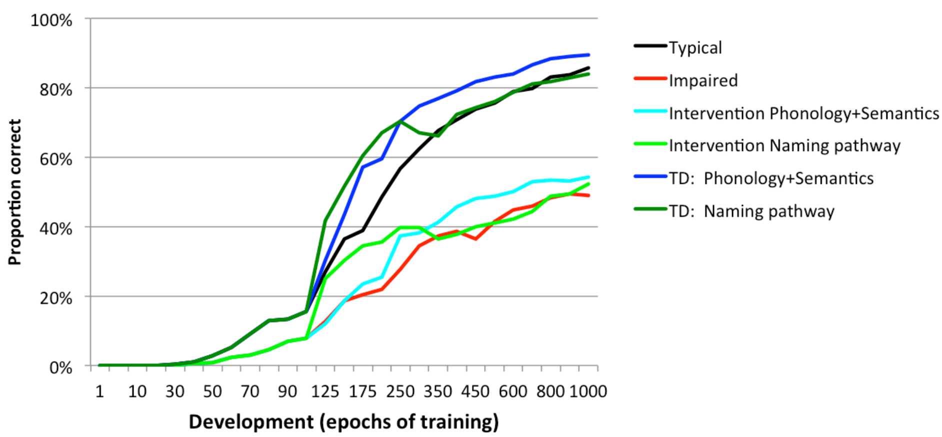
Replication 1



Replication 2



Replication 3

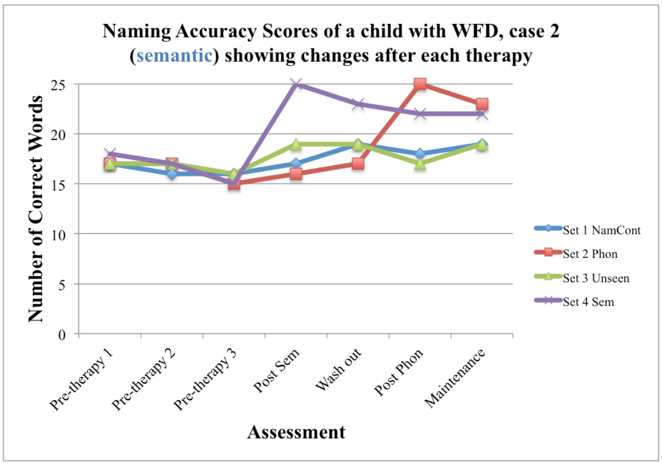


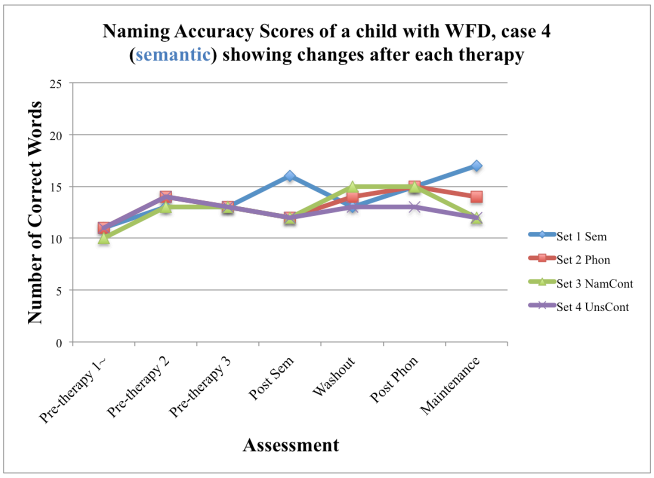
Appendix 7

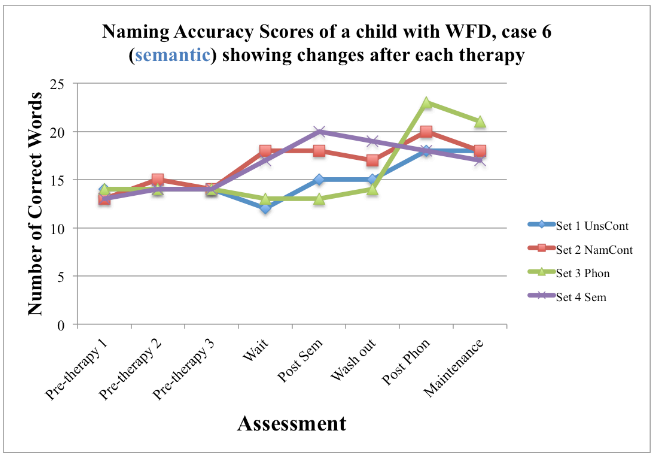
Effects of Therapy on the Naming Scores of Each of the Four Subgroups of 25 Words for 20 Individual Cases

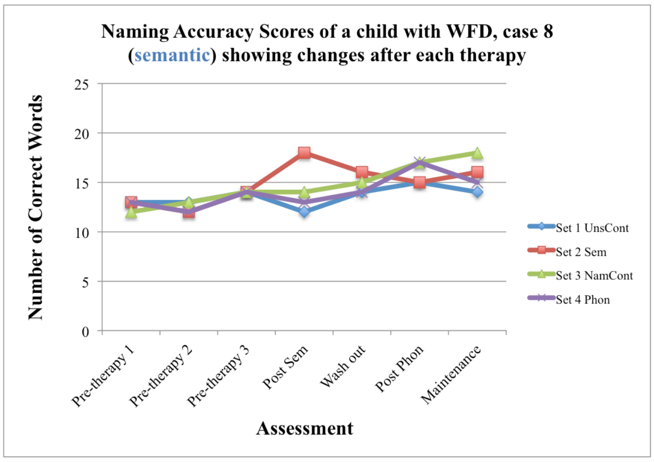
(Sem = semantic set, Phon = phonological set, NamCont = naming control set, Unseen = unseen set). For experimental design, see Chapter 6.

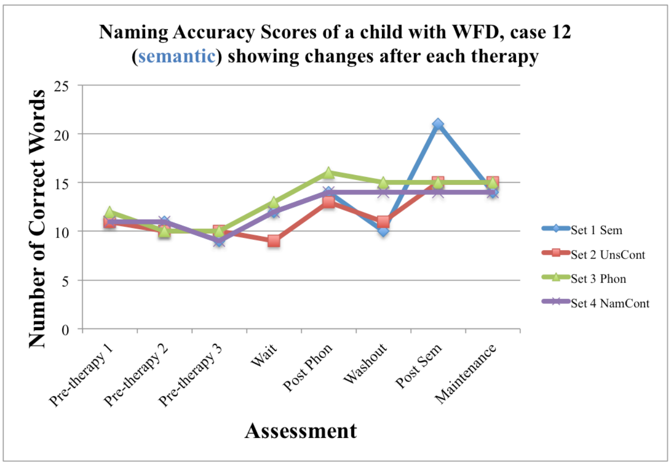
Subtype 1: Primarily Semantic difficulties

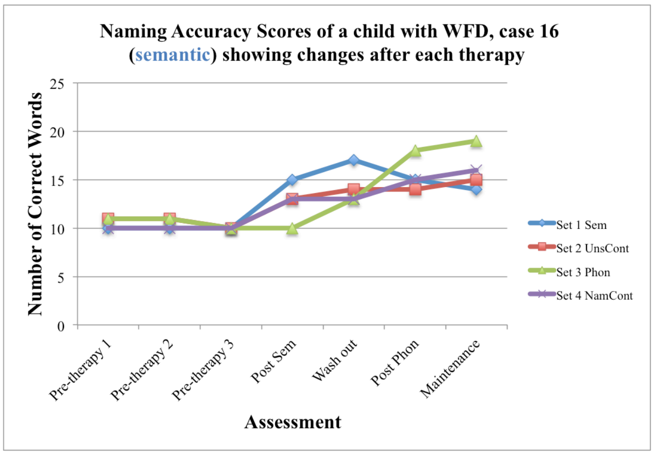




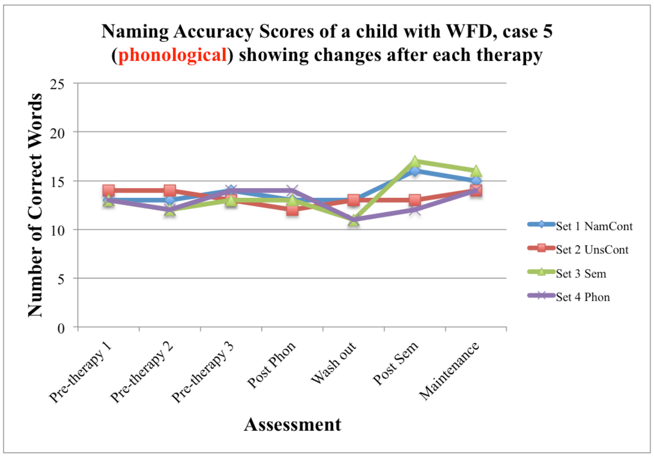

Naming Accuracy Scores of a child with WFD, case 6 (semantic) showing changes after each therapy
Number of Correct Words
25
20
15
10
5
0
Pre-therapy 1
Pre-therapy 2
Pre-therapy 3
Wait
Post Sem
Wash out
Post Phon
Maintenance
Assessment
Set 1 UnsCont
Set 2 NamCont
Set 3 Phon
Set 4 Sem

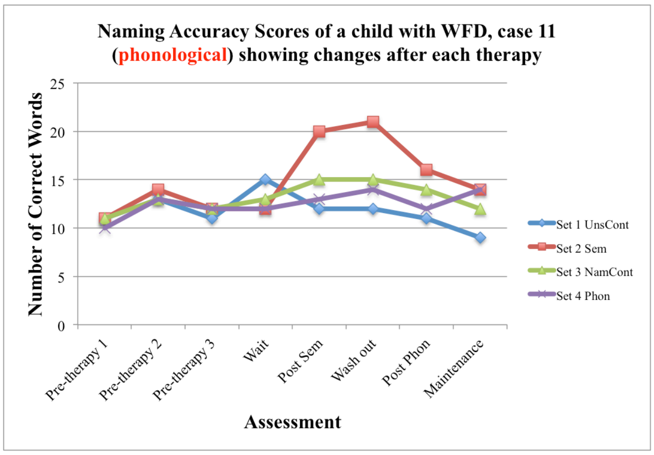

Naming Accuracy Scores of a child with WFD, case 8 (semantic) showing changes after each therapy
Number of Correct Words
25
20
15
10
5
0
Pre-therapy 1
Pre-therapy 2
Pre-therapy 3
Post Sem
Wash out
Post Phon
Maintenance
Assessment
Set 1 UnsCont
Set 2 Sem
Set 3 NamCont
Set 4 Phon

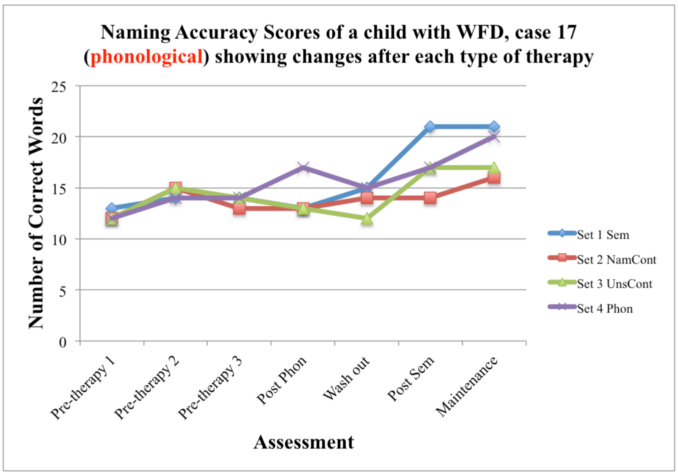

Naming Accuracy Scores of a child with WFD, case 12 (semantic) showing changes after each therapy
Number of Correct Words
25
20
15
10
5
0
Pre-therapy 1
Pre-therapy 2
Pre-therapy 3
Wait
Post Phon
Washout
Post Sem
Maintenance
Assessment
Set 1 Sem
Set 2 UnsCont
Set 3 Phon
Set 4 NamCont



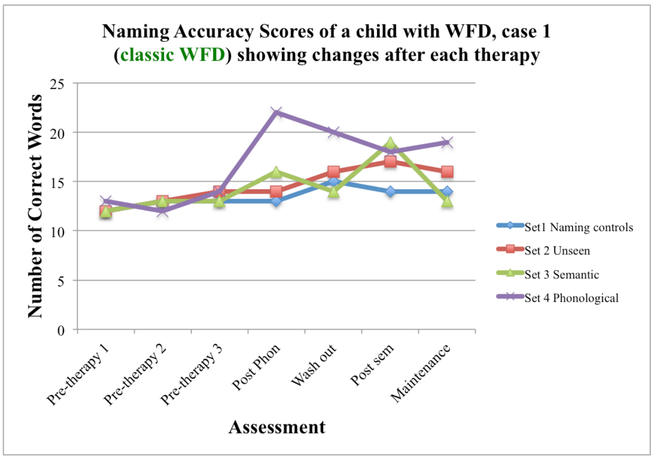
Subtype 2: Primarily Phonological difficulties

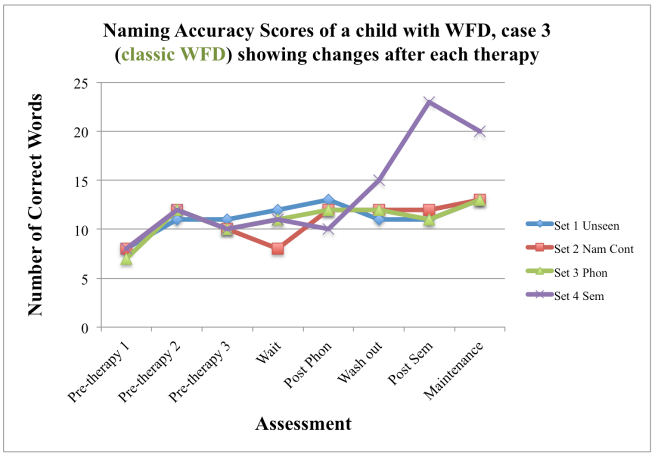


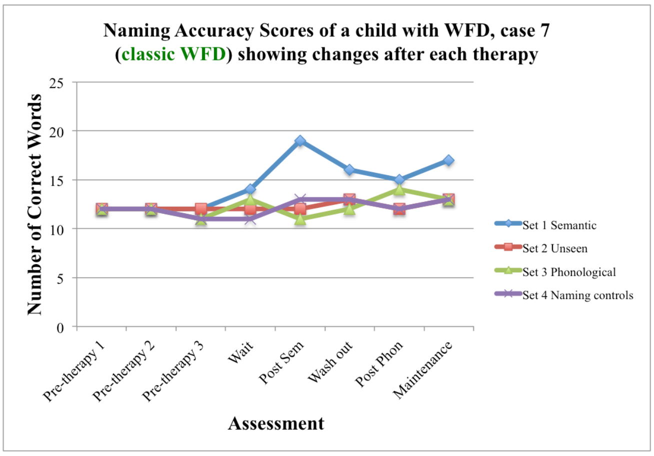


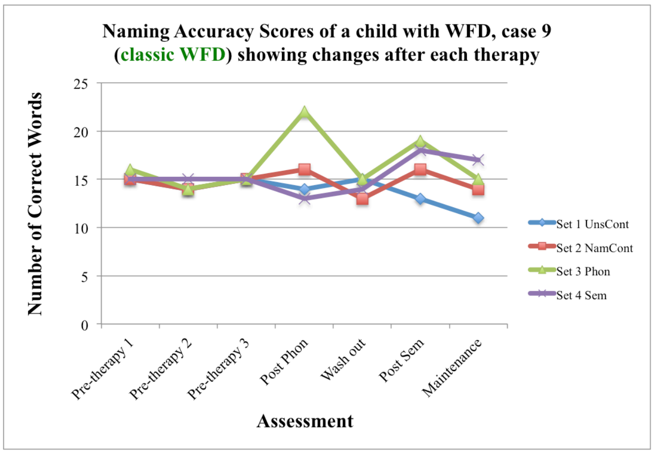


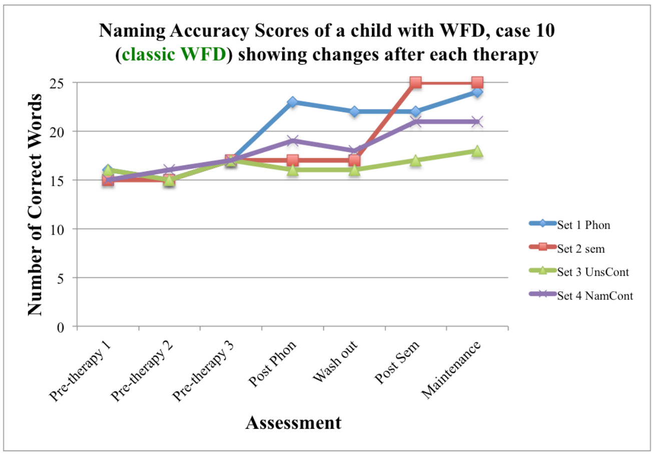
Subtype 3: Classic Word-Finding Difficulties

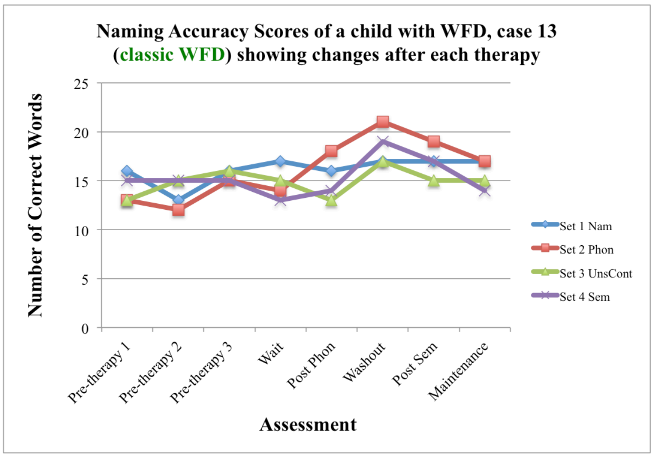


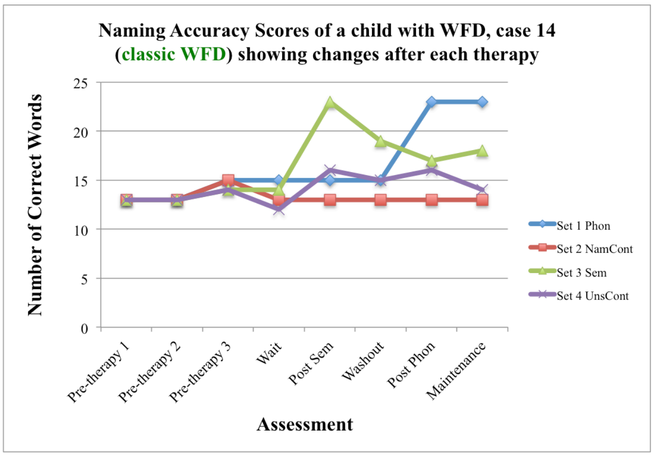


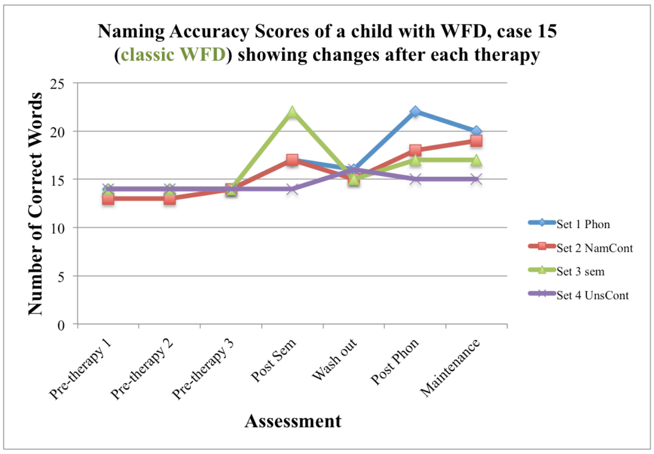

Naming Accuracy Scores of a child with WFD, case 7
(classic WFD) showing changes after each therapy
Number of Correct Words
25
20
15
10
5
0
Pre-therapy 1
Pre-therapy 2
Pre-therapy 3
Wait
Post Sem
Wash out
Post Phon
Maintenance
Assessment
Set 1 Semantic
Set 2 Unseen
Set 3 Phonological
Set 4 Naming controls

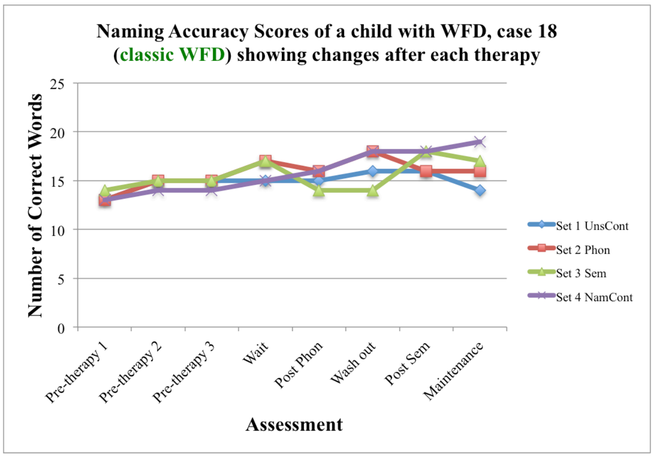

Naming Accuracy Scores of a child with WFD, case 9
(classic WFD) showing changes after each therapy
Number of Correct Words
25
20
15
10
5
0
Pre-therapy 1
Pre-therapy 2
Pre-therapy 3
Post Phon
Wash out
Post Sem
Maintenance
Assessment
Set 1 UnsCont
Set 2 NamCont
Set 3 Phon
Set 4 Sem

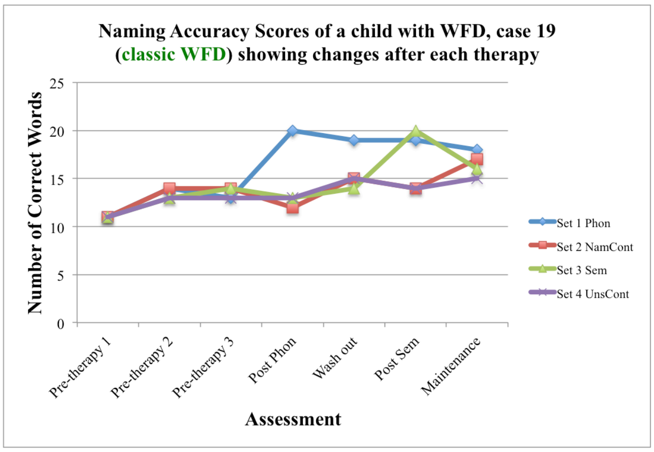

Naming Accuracy Scores of a child with WFD, case 10
(classic WFD) showing changes after each therapy
Number of Correct Words
25
20
15
10
5
0
Pre-therapy 1
Pre-therapy 2
Pre-therapy 3
Post Phon
Wash out
Post Sem
Maintenance
Assessment
Set 1 Phon
Set 2 sem
Set 3 UnsCont
Set 4 NamCont

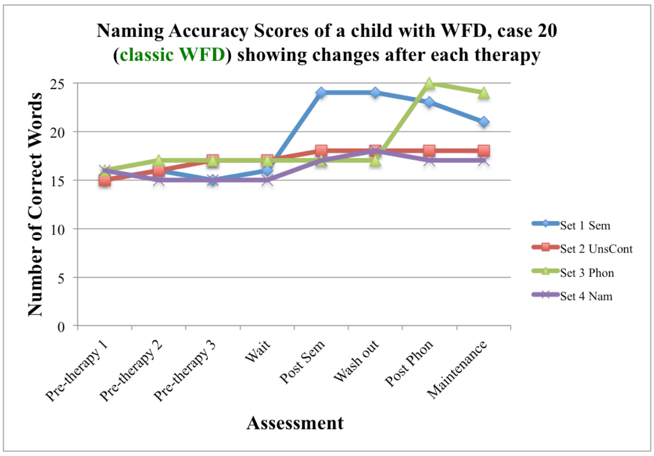

Naming Accuracy Scores of a child with WFD, case 13
(classic WFD) showing changes after each therapy
Number of Correct Words
25
20
15
10
5
0
Pre-therapy 1
Pre-therapy 2
Pre-therapy 3
Wait
Post Phon
Washout
Post Sem
Maintenance
Assessment
Set 1 Nam
Set 2 Phon
Set 3 UnsCont
Set 4 Sem


Naming Accuracy Scores of a child with WFD, case 14
(classic WFD) showing changes after each therapy
Number of Correct Words
25
20
15
10
5
0
Pre-therapy 1
Pre-therapy 2
Pre-therapy 3
Wait
Post Sem
Washout
Post Phon
Maintenance
Assessment
Set 1 Phon
Set 2 NamCont
Set 3 Sem
Set 4 UnsCont


Naming Accuracy Scores of a child with WFD, case 15
(classic WFD) showing changes after each therapy
Number of Correct Words
25
20
15
10
5
0
Pre-therapy 1
Pre-therapy 2
Pre-therapy 3
Post Sem
Wash out
Post Phon
Maintenance
Assessment
Set 1 Phon
Set 2 NamCont
Set 3 sem
Set 4 UnsCont


Naming Accuracy Scores of a child with WFD, case 18
(classic WFD) showing changes after each therapy
Number of Correct Words
25
20
15
10
5
0
Pre-therapy 1
Pre-therapy 2
Pre-therapy 3
Wait
Post Phon
Wash out
Post Sem
Maintenance
Assessment
Set 1 UnsCont
Set 2 Phon
Set 3 Sem
Set 4 NamCont


Naming Accuracy Scores of a child with WFD, case 19
(classic WFD) showing changes after each therapy
Number of Correct Words
25
20
15
10
5
0
Pre-therapy 1
Pre-therapy 2
Pre-therapy 3
Post Phon
Wash out
Post Sem
Maintenance
Assessment
Set 1 Phon
Set 2 NamCont
Set 3 Sem
Set 4 UnsCont


Naming Accuracy Scores of a child with WFD, case 20
(classic WFD) showing changes after each therapy
Number of Correct Words
25
20
15
10
5
0
Pre-therapy 1
Pre-therapy 2
Pre-therapy 3
Wait
Post Sem
Wash out
Post Phon
Maintenance
Assessment
Set 1 Sem
Set 2 UnsCont
Set 3 Phon
Set 4 Nam