2017

S. Meeuwsen
3895351
Utrecht University
Master's Thesis
Supervisor: Dr. E. Tribushinina
13/02/2017

[LEXICAL ABILITIES IN WILLIAMS SYNDROME]

Abstract

This study aims to illuminate the organisation of the mental lexicon in individuals with Williams Syndrome (WS). Previous research of WS lexical abilities has focused mainly on analysis on a quantitative level, but included observations of unusual performance that motivate a qualitative investigation. This study explored WS performance using data from 64 Dutch-speaking WS individuals for a picture vocabulary task, a naming task and a word-definition task. The findings from the research support Temple et al.'s (2002) findings regarding a lack of semantic precision in lexical comprehension and suggest this may extend to the productive domain as well. In addition, a specific weakness regarding adjective comprehension has been observed, as well as remarkable productive creativity when participants attempted to circumvent word-finding problems. Based on these findings, qualitative research could well be the next chapter of research into the fascinating language profile of these individuals.

Acknowledgements

Writing a thesis is often a long and intensive process, and I can safely say this thesis was no exception to that rule. The past year and a half have been tough, and the question "So, did you finish your thesis yet?" was one I dreaded for a long time. Yet, here I am, writing the last words. I could not have made it this far without the help of a number of people, whom I would like to thank at this stage.

First of all, I would like to thank my supervisor, Elena Tribushinina, for all the support and guidance you offered. I know you must have had your doubts about whether this thesis would ever come to an end, especially after all the times I said I really was going to finish it now, but did not manage to meet the deadlines I had set. It hasn't been easy, but I have learnt a great deal from you. Thank you!

Secondly, a great thank you to Maarten Frens and Jos van der Geest at the Erasmus MC, for providing me with the data that I based this thesis on, for your hospitality (I really enjoyed the Fridays I spent working with the data in Rotterdam!) and for supporting me throughout the process, both morally and with statistics whenever it came up. Who would've thought a random Ingress chat would lead to this? And in addition, I am incredibly grateful to the WS study participants and their parents, as well as everyone involved in the study at the Erasmus MC.

A big shout out goes to the people that I spent countless hours at libraries and other university buildings with, as we were all working on our theses: Rose-Ann, Marloes, Maaike, the *afstudeergroep* crew – you rock. The same applies for the people at Studiemeesters: Sietske, Felicia, Roline, the other coaches, the students there that made the coffee breaks fun and the rest of the day bearable ;-).

Then there's my friends – especially those that understood to check "Can I ask about your thesis or is it off-limits today?" – for supporting me, and some of the ENL communities, from VDM/Utrecht to Europe, the US and some of the global bunch, who always provided me with ample distraction (and then some). And of course my parents and in-laws, who always tried to be there for me in whichever way they could, whether that meant "Don't mention the T-word" or cheering me on.

I can safely say I've saved the best for last: the one person without whom this thesis could not have happened as I would have rage-quit many, many months ago, and the one that was there for me every single day and that never, not even for a moment, stopped believing in me: my husband, Dennis. You sat with me so many times to help me come up with a battle plan, make a schedule, figure out what to do next and turn the massive iceberg of tasks into tiny little ice cubes that I could get through. You always managed to get my chin up, even when I was ready to give up and admit defeat. The thought of the proud look in your eyes with every small victory, even if that victory was as small as "I went to uni today" is what kept me going throughout. We're the dream team, I couldn't have done it without you. And to think we'll be parents soon!

THIS THESIS IS DEDICATED TO OUR CHILD, WHO IS STILL SAFELY TUCKED AWAY AS I AM WRITING THIS. I CAN'T WAIT TO FINALLY MEET YOU AND HOLD YOU IN MY ARMS.

Contents

1. Introduction

Williams Syndrome (WS) is a rare developmental disorder caused by a microdeletion on chromosome 7q11.23, and is characterised by a range of physical and cognitive features. Individuals with WS are often described to have a very typical facial appearance, which has been referred to as 'elfin' for its resemblance of elves (Bellugi, Wang, & Jernigan, 1994; Bennett, LaVeck, & Sells, 1978; Brock, 2007). Prevalence has been estimated as low as 1:50.000 (Bellugi et al., 1994), but was usually found to be estimated as 1 in 20.000 live births (Mervis, 1999). In a more recent epidemiological study, this was corrected to 1 in 7.500 (Strømme, Bjømstad, & Ramstad, 2002). According to the Dutch WS patient organisation, approximately 10 to 15 children with WS are born in the Netherlands per year (NVWBS, 2016).

WS is associated with an intellectual disability that ranges from mild to severe; previous studies have described a rather sizable degree of variability in WS scores on IQ measures, with some individuals attaining very low scores and others reaching scores attesting an average level of intelligence (Mervis et al., 2000; Udwin, Yule, & Martin, 1987). However, in a study of 62 adults with WS, 87% was found to have an IQ score of 50-69, indicating mild intellectual disability (Howlin, Davies, & Udwin, 1998). Within the cognitive profile, several studies including that of Howlin et al. described a significant discrepancy between verbal and performance IQ. This is not necessarily only a consequence of relative verbal strength, but may also be impacted by a distorted representation of general cognitive ability due to a notable weakness in the visuospatial domain (Brock, 2007; Farran & Jarrold, 2003).

The verbal advantage reported in WS inspired researchers to propose it as a prime example of unimpaired language in individuals with cognitive impairments, supporting the modularity hypothesis; the belief that language forms a separate module in the brain, independent of general cognition (Fodor, 1983). As will be evident from the literature review in chapter 2, these views have since been rejected by many studies in which WS individuals were found to perform well below MA-level on tasks of a multitude of linguistic abilities. Nonetheless, the relative strength of some parts of language in comparison to the general cognitive level did spark an interest for this rare condition in psychological and linguistic research.

The linguistic abilities of individuals with Williams Syndrome are still very much subject of debate, with some researchers finding particular strengths in areas such as receptive vocabulary, morphosyntax and semantic fluency, while other studies were unable to replicate those findings. WS individuals are often matched with groups with other types of learning impairment such as Down Syndrome, and are found to outperform their IQ-matched counterparts on various tests of grammatical production and comprehension (Brock, 2007). However, these results are also criticised as they may be confounded due to language being a particular weakness in Down Syndrome (Temple, Almazan, & Sherwood, 2002). Conversely, WS groups do not outperform MA-matched TD controls on these measures, indicating their grammatical abilities would be on par with, but not exceeding their general cognitive level.

For this study, the focus will be on the lexicon of individuals with WS. Within this particular domain, both areas of great strength and areas of relative weakness have been described in the literature, as will be discussed in greater detail in chapter 2. WS research has, to this day, almost exclusively focused on examining language skills based on how WS groups perform on psychometric tasks in comparison to control groups, thus investigating how they compare to other individuals on predefined measures of specific linguistic abilities. This study, however, will take a very different approach by examining the information contained in their language. This is a novel method for the line of research, and will therefore reveal more about the workings of the mechanisms relating to the WS mental lexicon.

2. Lexical development in WS

The outline of this chapter will be as follows: first of all, the linguistic subfield that is the focus of this study will be introduced, after which some of the core literature on the subject of lexical abilities in WS will be presented. The relevance of the current study will then be further motivated in the context of these studies. This will form the basis for the research questions and hypotheses that will be presented at the end of the chapter.

2.1 The mental lexicon

The study of the mental lexicon concerns itself with the storage and retrieval of words in the mind. As such, the mental lexicon can be considered the 'dictionary' of the brain (Aitchinson, 2012). The storage of a word requires specific information; Jackendoff (2013) described knowledge of a word as "*a long-term memory association of phonological, syntactic, and semantic features*" (Jackendoff, 2013, p. 130). Thus, associated abilities concern both comprehension and production of words, and rely on the knowledge of word meaning in order to do so. In this study, the lexical semantics domain of language will be examined, which also involves word meanings and relations between words, or semantic categories (for instance 'animals' or 'musical instruments'). Lexical abilities can be elicited through tasks that require specific kinds of operationalisation of these abilities, or through the examination of spontaneous speech samples. In this study, the focus will be on a limited number of tasks, which do not represent the ability as a whole, merely an application thereof.

2.2 Lexical abilities in WS

Among the first and most cited studies of linguistic abilities in Williams Syndrome, are those by Bellugi and her colleagues (Bellugi, Bihrle, Jernigan, Trauner, & Doherty, 1990; Bellugi et al., 1994). Bellugi et al. (1990) reported WS participants to perform considerably better than chronological age (CA) and IQ-matched participants with Down Syndrome for both the Peabody Picture Vocabulary Task (PPVT) (Dunn & Dunn, 1997), a test of receptive vocabulary where participants must select one out of four visual stimuli to match the target word, and Semantic Fluency, in which a participant is asked to name as many members of a certain category in a limited time span. Bellugi et al. are often cited for their findings regarding

Semantic Fluency, as they reported the production of 'unusual, low-frequency, non-prototypical[1] category members' on this test, with examples including "*unicorn, tyrandon, brontosaurus, yak, ibex, water buffalo, sea lion, sabre-toothed tiger* and *vulture*" (Bellugi et al., 1990, p. 116).

Furthermore, Bellugi et al. (1990) reported performance above mental age (MA) level for the PPVT, which has led to numerous further examinations and the general consideration of receptive vocabulary in WS as a relative strength. Finally, they addressed one of the key debates in the field in those days: the question of whether language in cognitively impaired individuals represents a delayed development, as was commonly thought at that time, or a deviant picture altogether. The results of this study led the authors to believe that the latter is actually the case; they argued that their findings indicate deviant language abilities in WS, when compared to typically developing (TD) controls (Bellugi et al., 1990).

Based on these and other studies at the same research institute, a description of the neuropsychological profile of WS was composed, in which Bellugi et al. (1994) sought to incorporate several levels of functioning. Ultimately, Bellugi et al. proposed that WS may well be a prime example of the modularity of certain, but not all, aspects of linguistic ability. On the one hand, semantic knowledge in WS was said to be partially preserved with performance on measures of receptive vocabulary being above MA-level, but also partially deviant in the aforementioned unusual responses given in Semantic Fluency tasks. On the other hand, their results indicated that grammar was mostly unimpaired, leading them to propose that semantic knowledge may be dependent on general cognitive ability, with semantics requiring a degree of understanding of the world around a person, whereas syntax would be an independent system, as it does not require any such worldly comprehension. These potentially ground-breaking findings and theories formed the ground for a great number of studies of individuals with WS, a few of which will be discussed hereafter (Bellugi et al., 1994).

[1] The classification 'non-prototypical category members' denotes that while these items are member of the target category, they are not central to the category (i.e. they are not 'prototypical'), based on Rosch's (1973) prototype theory. This model of concepts considers categories to be structured in such a way that there are central or typical members of a category, such as *chair* for the category *furniture*, but also less typical or peripheral members such as *lamp* for this same category (Saeed, 2009, p.37).

Volterra, Capirci, Pezzini, Sabbadini and Vicari (1996) tested these claims of a disharmonic linguistic profile in a different group. While Bellugi et al.'s (1990) study involved English-speaking adolescents, Volterra et al.'s participants were young, Italian-speaking children. Furthermore, they chose to incorporate a TD control group, with participants' CA matching the MA of the WS group. They employed not only the Semantic Fluency test and the PPVT, but also the Boston Naming Test to measure lexical production abilities. Moreover, contrary to Bellugi et al.'s observation that WS groups perform above MA-level on the PPVT, that same test did not produce a significant difference between WS groups and TD controls in this study. Comparison of Semantic Fluency test results for both groups did not reach statistical significance either, and no production of rare or unusual items was found. An additional complicating discovery for the already complex semantic ability profile of WS came from the Naming task, where WS children's poor performance was well below the level expected based on their MA. In conclusion, Volterra et al.'s young participants' performance did not reveal any of the anomalous, within-domain strengths reported by Bellugi, and added weak naming to the general profile (Volterra et al., 1996).

Rossen, Klima, Bellugi, Bihrle and Jones (1996) also studied the lexical semantics domain in WS, and similar to Bellugi et al. (1990), tested adolescents with WS and IQ- and CA-matched Down Syndrome counterparts. They described a 'remarkable wealth of word knowledge' (p. 377) in WS based on both PPVT and Semantic Fluency tasks, in which WS groups clearly outperformed the Down Syndrome counterparts who were both CA- and IQ-matched. Furthermore, Rossen et al. also reported more low-frequency words by the WS participants on the Semantic Fluency task, but only towards the end of the task. This study supported earlier findings, and also provided a possible explanation: word choice was said to be 'unexpected' at times (p. 387), and the authors suggested this may be attributed to a potential lack of specificity regarding lexical processing (Rossen et al., 1996).

Following these broader studies, several researchers have attempted to zoom in and investigate particular aspects in more detail. These will be presented in the following sections, starting with receptive vocabulary, which is considered a relative strength in WS individuals.

2.3 Area of strength: receptive vocabulary

A vast majority of studies of receptive vocabulary in WS reported either performance exceeding MA-based expectations or WS outperforming control groups selected using various criteria (Bellugi et al., 1990; Clahsen, Ring, & Temple, 2004; Temple et al., 2002; Vicari et al., 2004). Consequently, further research into the nature of this apparent strength was pursued; these studies will be discussed in this section.

In two strongly related studies, Temple et al. (2002) and Clahsen et al. (2004) confirmed the general strength in the common receptive vocabulary tests, but also found evidence indicating certain limitations to these abilities. They applied self-made tests similar to the PPVT, but with distractor items sharing semantic features with the correct response. For Temple et al.'s study, the items were presented in a context with 23 distractor items from the same category. The weaker performance found on the latter test led Temple et al. to argue that, despite fair performance on tasks of general receptive vocabulary, WS individuals reveal impairments when the task demands more detailed lexical ability (Temple et al., 2002). However, they did not take into account the possibility that the number of distractor items played a role in these findings.

In another study, Clahsen et al. (2004) replicated these findings and ruled out the suggestion of number of items causing the deviating findings compared to a standardised test such as the British Picture Vocabulary Scale (BPVS) (Lloyd, Dunn, Dunn, Whetton & Burley, 1997), a test that is highly similar to the PPVT. They found their WS participants still outperformed MA-matched controls on tasks where specific semantic knowledge was not required, because the distractor items were only remotely related to the target items, even when the number of distractors was increased from three to eleven or to 23. However, when the distractor items were retrieved from the same semantic class, WS groups suddenly performed significantly worse than MA-matched counterparts. This led Clahsen et al. to conclude that the effect is caused by the nature of the distractors, and cannot be contributed to a lack of inhibition that may be triggered by the high number of distractors to choose from (Clahsen et al., 2004). These findings may be interpreted to indicate a weakness regarding vocabulary depth, or the measure of how well participants know words, while vocabulary breadth, or the number of words known, seems to be a relative strength in WS (Vermeer, 2001).

In conclusion, the general impression of strong performance on these receptive vocabulary tasks has already been questioned and appears to be limited to certain tasks and contexts. The aforementioned studies all focussed on performance on the tests as a whole, though, and did not investigate whether any potential performance patterns occurred for particular item categories. For instance, verbs are often considered to be more difficult to acquire than nouns (Gentner & Boroditsky, 2001; Maguire, Hirsh-Pasek, & Golinkoff, 2006), and thus, one might expect to see a pattern where participants perform worse on verbs than on nouns. Yet, this line of research has not been pursued for WS groups, leaving many questions open as to whether any patterns can be discerned in their performance. This calls for further, more detailed investigation to determine the specific abilities unique to this group.

2.4 Unusual vocabulary

Secondly, our attention turns to WS performance on the Semantic Fluency tasks described in §2.2, which represent an operationalisation of lexical production and semantic ability. Findings concerning overproduction of low-frequency items for these tasks led researchers to hypothesise unusual vocabulary to be characteristic of WS. In this section, several studies are presented that investigated the suggestion of unusual vocabulary in WS.

The reported unusual production on the Semantic Fluency task inspired Jarrold, Hartley, Philips & Baddeley (2000) to investigate the nature of the items produced in this test. They tested a group of 13 WS participants, who were matched based on BPVS scores to 39 individuals with moderate learning difficulties, and analysed performance on the Semantic Fluency task for the number and nature of the items produced, as well as the order of production. Receptive vocabulary was chosen as a matching criterion because this would ensure that differences in typicality would indicate a deviance in the structure, rather than the extent, of the lexical knowledge. They found no significant group difference between the WS group and the group with moderate learning difficulties. Both groups produced the same number of novel items, and were found to produce the same responses; typicality of responses was therefore also similar for both groups. No unusual responses were reported for either of the groups. The only striking feature of the WS group was their high number of repetitions, or instances in which a response was given more than once by the same participant, in comparison to the moderate learning difficulties group, which Jarrold et al. attributed to the executive demands of the Semantic Fluency task. However, as the groups were matched based on receptive vocabulary scores, which is a relative strength in WS, they may have had a disadvantage in terms of lower general intelligence compared to the moderate learning difficulties group. This cannot be ruled out as a confounding factor, and may have impacted the findings of this study; perhaps Jarrold et al. found no unusual WS performance because of the matching criterion chosen (Jarrold et al., 2000).

Lukács, Pléh and Racsmány (2004) also sought to investigate the reported atypicality on Semantic Fluency tests, and employed a number of categories to examine the frequency of items produced by both WS and control participants matched on verbal age. Like Jarrold et al. (2000), they found no difference between the groups for this measure. Interestingly, also in line with the previous study, they did find that WS subjects produced more items that did not belong to the target category (i.e. non-category or other category members) and more repetitions (Lukács et al., 2004).

Another study worth mentioning in this context, is Stojanovik and Van Ewijk (2008). They did not use the Semantic Fluency task but rather sought to examine the suggestion of unusual vocabulary in WS in Frog Story narratives, a task in which participants are asked to tell a story by means of a picture book about a boy who goes looking for his frog (Mayer, 1965). To prevent findings being confounded by the choice of matching criterion as was suggested for Jarrold et al. (2000), Stojanovik and Van Ewijk chose to employ a design featuring multiple control groups: a control group matched for receptive vocabulary, one matched for performance on a non-verbal task and one matched for chronological age. They found no difference in production of low-frequency items, nor in general number of different words used between WS and the control groups. As such, even with an alternative research design, the reported findings of deviating vocabulary that Bellugi et al. (1990) described, could not be replicated in any of these studies. This led the authors to question whether there really is anything extraordinary about the performance of WS participants on this task (Stojanovik & van Ewijk, 2008).

Finally, a study that explored the supposedly deviant semantic organisation that would underlie the overproduction of low frequency-items as discussed in section 2.4, is reported in Thomas et al. (2006). While their study was published well after some of the studies cited in this section that found no evidence whatsoever for unusual vocabularies in WS, Thomas et al. still described overproduction of low frequency-items as characteristic for WS, and sought to investigate whether this is a result of a deviation in the way the lexicon operates (the so-called 'intra-lexicon hypothesis'), or related to problems outside the mental lexicon (the 'extra-lexicon hypothesis'). To test these hypotheses, they employed a speeded naming task and a comprehension task, seeking to test the effect of word frequency on both TD and WS groups.

A base naming speed was established through numeral and letter naming tasks, as these categories were highly familiar to all participants. This was then compared to results for object- and action-naming, both of which included high- and low-frequency items to examine frequency effect. Thomas et al. confirmed previous reports of weak naming ability in WS groups, but found no insensitivity to word frequency. Such an insensitivity would have attested to deviating lexical development, and thus have supported the intra-lexicon hypothesis. However, as their findings did not indicate the root of the problem to be found within the lexicon, they concluded that the higher production of low-frequency items by WS must be caused by a factor outside the lexicon, supporting the extra-lexicon hypothesis (Thomas et al., 2006).

However, it should be kept in mind that this study was based on the presupposition that WS individuals use more uncommon words than typically developing peers; a belief that has been criticised recurrently, as discussed in section 2.4. As Thomas et al. (2006) did confirm the previously reported naming weakness, the current study will further explore WS performance on naming tasks, but frequency effects will not be the focal point, following Thomas et al.'s findings. As such, for the current study, attention will focus on other measures where deviating or superior performance in WS has been reported more frequently.

2.5 Area of weakness: naming

The third focus area is a skill that was not actually part of the initial studies by Bellugi et al. (1990), but that was reported to be relatively weak by later studies: naming ability. Naming tasks are an operationalisation of lexical production, as they rely on word retrieval to correctly identify the item a participant is presented with. Contrary to their performance on receptive vocabulary, WS participants achieved results similar to their MA-matched peers (Bello, Capirci & Volterra, 2004) or even significantly worse results on these lexical production tasks in most studies (Lukács et al., 2004; Temple et al., 2002; Vicari et al., 2004; Volterra et al., 1996). The study by Thomas et al. (2006) that was presented in the previous section also supported these findings, as they found the WS group to perform worse than TD controls on naming tasks, and to also have a lower overall naming speed. In this section, the results from a number of key studies involving naming tasks will be presented, in order to establish a hiatus in current research that will be filled by the study at hand.

Temple et al. (2002) investigated whether any areas of linguistic deficit could be found for WS participants, and stated no detailed analysis of lexical naming tasks had been reported to date. They found all four of their participants to have naming ages that were significantly lower than their receptive vocabulary ages, with differences varying between 2;0 years and 3;9 years. For three out of four participants, naming scores were also well below what might be expected based on their MA level. While they reported their task to have too few items to allow for a detailed error analysis, they did describe an atypical response, where three out of four participants identified a component of the object shown in the picture, rather than the object in its entirety.

Furthermore, in the same study, Temple et al. (2002) also reported a different experiment involving a naming task. In this experiment, they tested naming ability for specific semantic categories; these categories were ‘animals’, ‘indoor objects’, ‘foods’ and ‘clothes’. Their scores were compared to TD controls matched for full-scale MA, and two out of four participants were found to score significantly lower overall. Temple et al. did consider the matching criterion a potential confounding factor, as performance IQ is often lower than verbal IQ in WS individuals. Therefore, the TD controls may have had a lower verbal IQ than their WS counterparts in some instances. When this was corrected for one WS participant, that participant was found to have a significant impairment for three out of four item categories. Like in their first experiment, here, Temple et al. also found errors indicating WS participants had named a component of the item rather than the whole object. As Temple et al. found atypical errors in both naming tasks, this makes their proposal of doing an error analysis for a naming task very interesting, but this has not yet been done on a larger scale for WS. This provides an excellent approach from which to examine naming ability in WS, to investigate what this reveals about WS lexical production.

2.6 Definitions

Finally, performance on definitions tasks such as the WISC-R Vocabulary subtask (Wechsler, 1974) is examined. These tasks represent a different operationalisation of lexical production than naming tasks, as they measure vocabulary depth, or how well a participant knows certain words, while the other tasks focused on vocabulary width, or the number of words known to participants (Vermeer, 2001). An important remark about these tasks is that performance relies not only on linguistic, but also on metalinguistic ability, as defining a word also requires the participant to know what constitutes a definition. Consequently, these tasks have been suggested to underestimate word knowledge even in TD groups (Nippold, 1995; Wehren, De Lisi, & Arnold, 1981).

As also discussed in §2.2, Bellugi et al. (1990) and Rossen et al. (1996) reported that WS participants' responses, while often incorrect, did reveal that they possessed some knowledge of target items and knew more than their Down Syndrome counterparts in these studies, even if WS group scores did not reveal this knowledge. This finding was also supported by Temple et al. (2002), who argued the strict scoring criteria of the task could easily lead to "*formal failure […] when the child provides associative or circumlocutory responses*" (p. 473). Temple et al.'s statement serves to illustrate the previous argument regarding metalinguistic demands: responses are awarded fewer or no points when the child's response does not comply with the expectations set for the form and structure of a definition.

Following these reports, performance on definitions tasks was examined in further detail by Purser, Thomas, Snoxall, Mareschal and Karmiloff-Smith (2010). They investigated the actual amount of information in the WS lexicon, questioning if vocabulary level may, in fact, overestimate this. As such, their study also added to the receptive vocabulary studies discussed in section 2.3. They employed two paradigms: the first one was a definitions task in which participants were tested for knowledge of the names of 21 animals, and was used to examine lexicosemantic processes. However, performance on this test was expected to be confounded by metacognitive demands, and for this reason, a categorisation task was added. In this task, knowledge of the same 21 animals was tested through questions such as 'Which are birds?' and 'Which can sting?'. The authors also included a developmental component in their study by analysing the relationship between verbal MA, which the authors defined as performance on the PPVT, and number of correct responses. They found that WS performance for the definitions task was similar to their verbal MA for younger participants, but as their receptive vocabulary knowledge grew and thus, verbal MA increased, performance on the definitions task diverged from this trajectory. Performance on the categorisation task, on the other hand, showed a developmental trajectory comparable to that of the TD controls, but was weaker than predicted by verbal MA throughout the entire course of development. This difference in performance between the two tasks was taken to indicate that the two tasks test different abilities, and led Purser et al. to conclude that metacognitive aspects may confound results for the definitions task.

However, other accounts for weak performance on definitions tasks must also be considered. For instance, Temple et al. (2002) argued that associative and circumlocutory responses would be failed under the scoring criteria, so the possibility of WS participants choosing such an approach needs to be investigated. Additionally, Temple et al. also described excellent WS performance when a task did not require 'fine-grained semantic specifications', but merely a superficial level of word-retrieval. As discussed in section 2.3, this can be interpreted to indicate shallowness of word knowledge in WS; this level of knowledge would suffice to link a word to the correct picture as is required for receptive vocabulary tasks, but would cause participants to fall short on a definitions task, which demands more detailed word knowledge. This would be supported by reports of WS participants' definitions containing some correct information about the target item, yet lacking sufficient diagnostic features of the target. This may also account for problems in definitions tasks, as these require participants to go beyond a shallow degree of knowledge.

Whether the weak WS group performance on definitions tasks is caused by the task's metacognitive demands, by a shallow level of word knowledge in WS or by another cause altogether (or a combination thereof) is still up for debate, calling for further investigation into WS performance on these tasks. Even if the reported difficulty were to be caused by the task's metacognitive demands, WS responses may still offer an interesting perspective into the organisation of the WS mental lexicon. This will be discussed in more detail in chapter 6.

2.7 Aim & Research questions

The preceding sections revealed that the bulk of studies so far focused primarily on general, quantitative analyses and performance on standardised measures, thus presenting a broad outline of strengths and weaknesses between domains, yet neglecting to expand the scope beyond the superficial level. Therefore, in this study the central question will be: "What does performance on a range of lexical tasks reveal about the organisation of the mental lexicon in individuals with WS?". To reach an answer to this question, a range of sub-questions will be answered in three studies, which will be discussed in the subsequent chapters.

In these chapters, knowledge of words and word meanings will be elucidated on three levels: first, lexical comprehension ability will be investigated through a picture-selection task; second, our attention turns to displays of vocabulary knowledge on a productive level or more specifically, the naming ability; and finally, extending on the productive level, the ability to define words will be examined. In each chapter, the central question, rationale and research methodology for that theme will be presented.

The decision to employ a different approach for each theme is motivated by the richness of the data for each field; ranging from multiple-choice tasks for receptive vocabulary to entire sentence clauses for definition tasks, the amount of information contained in each answer varies greatly per task. In the next chapter, methodological characteristics of the entire study will be listed, followed by sub-studies examining each theme respectively.

3. Methods

This study will examine data that was gathered for a large study of Williams Syndrome at the Erasmus MC Department of Neuroscience in 2001. This study sought to establish a connection between the neurological and behavioural phenotype of Williams Syndrome with its genotype. Part of this study consisted of several neuropsychological and linguistic tests, including receptive vocabulary, naming, semantic fluency, narratives and sentence repetition (van Hagen et al., 2007). The data from these tasks was only analysed on a general quantitative level for a master's thesis (Franken, 2003) but was never used for any published research. For the purpose of this study, a subset of the data was selected for containing most information regarding lexical performance. As part of the methodology is identical for all tasks discussed, the general characteristics will be presented here, before presenting the sub-studies that comprise this study in the subsequent chapters.

3.1 Participants

The participants were 64 Dutch L1-speakers diagnosed with Williams Syndrome (age range 6;1 – 40;11, mean CA = 17;11), of which 55 were known to have the gene-deletion associated with WS, which was confirmed using a fluorescence in-situ hybridisation (FISH) test. Participation was voluntary, and recruitment was coordinated through the Dutch WS patient organisation, the Erasmus Medical Centre in Rotterdam and the VU Medical Centre in Amsterdam. Participants did not receive any financial compensation other than travel costs.

To allow for comparison between age groups, the participants were divided into three groups according to chronological age: children (CA <12), adolescents (CA 12-21) and adults (CA >21). Further group demographics are presented in Table 1.

Table 1: Participant group information

Group	*N*	*Range*	*Mean (SD)*
1 (children)	20	6;1 – 11;8	8;7 (2;0)
2 (adolescents)	21	12;0 – 20;5	15;9 (2;7)
3 (adults)	23	21;4 – 40;11	27;11 (6;3)

3.2 Procedure

All participants were asked to visit the Erasmus MC for the first test session. As taking all tests in a single day would be too demanding for the participants, a few tasks (including those discussed in the study at hand) were administered during a second test session. Researchers chose to have this second session take place in the participants' home environment, so they would not need to come to Rotterdam twice. In order to facilitate this, participants and/or their caretakers were asked to set up a quiet room with a large table. To prevent any influencing by the parents, all tests were administered 1-on-1. The test sessions took about two hours, including a short break.

4. Study 1: Lexical comprehension

4.1 Rationale

Receptive vocabulary is arguably the linguistic domain addressed most frequently in research towards a potential linguistic advantage or strength in WS individuals. As described more elaborately in sections 2.2 and 2.3, a range of studies have investigated the early claims of superior performance on tests of this ability. Temple et al. (2002) and Clahsen et al. (2004) put these initial reports into perspective by showing that the apparent advantage is only expressed in tasks that do not rely on WS participants' fine-grained semantic knowledge. Moreover, Volterra et al. (1996) added to these early claims by suggesting the advantage may be limited to older participants, after finding their young children did not perform above expectations. This may, however, simply be a consequence of language development at young ages as this also occurs in TD children. Further research is thus needed to address the many questions that currently remain unanswered.

In this chapter, data from a picture selection task is examined on two levels to investigate whether WS participants' performance concurs with expectations based on the order of acquisition of words. First of all, the influence of syntactic category (verb, noun, adjective) is studied, as this has often been cited as a factor in the order of acquisition, where nouns are considered conceptually less complex and are thus often acquired before words belonging to other syntactic categories (Gentner, 1982). Secondly, the learnability of these items will also be scored based on a measure of conceptual abstractness. Do these individuals perform worse on items that are considered more difficult according to these abstractness scores, or do they show a different pattern than predicted by TD word judgements? Before inquiring further into this question, the results from analysis on a quantitative level will be discussed and the relevant constructs will be explained in greater detail.

4.1.1 Previous analysis test scores

As mentioned earlier, the data utilised in this study has been reviewed on a general quantitative level in an unpublished study (Franken, 2003). While the findings from this study do not answer the questions discussed in this study, they are of great value in showing whether the WS participants in this test group resemble those in previous WS research in terms of scores on particular tests. As such, that study's findings will be reviewed in each chapter to establish the reliability of the current data set as a representation of the WS population.

For the PPVT, Franken (2003) found that participants' corrected scores[2] were significantly higher than predicted by their general MA, which supports earlier findings such as those by Bellugi et al. (1990), Clahsen, Ring and Temple (2004) and others (for a more extensive discussion of this topic, please refer to section 2.3 of this thesis).

[2] The test battery employed for this study contained a shortened IQ measure, calculated over the Vocabulary and Block Design subtests of the WISC-R IQ test. However, as WS are reported to below well beneath MA level on the Block Design test, Franken (2003) believed this may have distorted the general MA calculated. Therefore, test scores were corrected for the difference found between the calculated MA and the test age found on the other verbal tests.

4.1.2 Learnability measure: SICI

In order to gain a measure of difficulty of each particular item, one must first revisit a conceptual framework underpinning language development. In response to a multitude of separate theories developed to account for the acquisition of different word classes, Hollich, Hirsh-Pasek and Golinkoff (2000) attempted to develop a single, unified theory of word learning: the Emergentist Coalition Model. Central to this model is the belief that children discover the referent of a new word using several different cues, on perceptual, linguistic and social levels, but that stronger reliance on one kind of cue over another evolves throughout development. While this model was originally intended to account for trends witnessed in noun acquisition, it was later extended beyond the borders of word class in an article by Maguire, Hirsh-Pasek and Golinkoff (2006). Their extension of the model posits that the learning of all lexical items, thus including not only nouns but also verbs and adjectives, is supported by the aforementioned perceptual, social and linguistic cues. Maguire et al. further extend this model to a single conceptual framework on which all words can be classified. This continuum, bearing the acronym SICI, is based on four major factors that are believed to impact the ease of word learning: shape, individuation, concreteness and imageability (SICI).

The SICI continuum focuses on conceptual abstractness, which is not the only factor influencing order of acquisition. However, the applicability of the SICI continuum as a measure of learnability on a level transcending word category is what motivates the decision to utilise SICI for the current study, as the test will involve nouns, verbs and adjectives, rather than limiting the scope to only one of these categories.

SICI scores for these four aspects together provide a continuum predicting word learning difficulty. In this continuum, Maguire et al. (2006) define 'shape' as the shape of an object for nouns, or the general shape of an action for verbs. According to this criterion, items such as *bag* and *walking* will be acquired early on in the process, as they have a very distinct shape, while items such as *idea* and *thinking* will be much harder to acquire.

The second factor, 'individuation', is defined by Maguire et al. (2006) as the degree in which the concept a word refers to can be observed and inspected in the world. This is based on the conviction that word learning is influenced by the complexity of the word's referent; items for which the referent can readily be discerned in a scene, for example *bottle* or *car*, are considered highly individuable. Verbs, on the other hand, will always be less individuable according to Maguire et al., as they believe that correct interpretation of these items would require a greater degree of understanding of the linguistic context than is the case for nouns.

The last two factors are closely related and are both sensory measures: 'concreteness' is a measure of sensory perception, or the ability to see, hear or touch something, and entails the ease with which a sensory mental image of the referent of a word can be produced. Maguire et al. (2006) state that the two terms are often used interchangeably, and that while imageability also includes categories such as emotions, which would not be part of concreteness, ratings for both measures are still highly correlated. Furthermore, Maguire et al. also report that imageability has proven strongly predictive of order of word acquisition, more so than the noun-verb dichotomy, referring to two studies by Bird et al. (2001) and Gillette et al. (1999) that support their claims.

4.1.3 Hypotheses

The question central to this chapter is: "What does a detailed analysis of WS performance on the PPVT reveal about lexical comprehension abilities in these individuals and the WS mental lexicon?" As this is still a fairly broad question, it is important to first distinguish the sub-questions encompassed therein.

First, word class will be analysed as a factor. Maguire et al. (2008), among others, reported that on the whole, verbs are 'more conceptually abstract' (p. 375), and thus, are often more difficult to acquire. This view is also supported by Gentner and Boroditsky (2001), who described the main issue in learning verbs, and even more so in closed grammatical classes such as prepositions, is the reliance on linguistic knowledge; these word classes require more contextual understanding than do concrete nouns. Therefore, the expected outcome of this part of the study is that participants will perform better on nouns than on verbs.

While adjectives have not been discussed as explicitly in the literature as have nouns and verbs, the argumentation presented by Gentner and Boroditsky (2001) can easily be extended to this class: as an adjective is a modifier of a noun, comprehension of the meaning of the adjective requires some degree of understanding of the noun it modifies. This is further illustrated by Tribushinina et al. (2014), who argued that an adjective can be interpreted in very different ways, depending on the noun it is combined with, and by Mintz & Gleitman (2002), who referred to the 'whole-object bias', or the bias found in toddlers to associate new labels with whole objects, rather than only a part or feature thereof. As adjectives are used to describe a property of an object, this bias would make adjectives more difficult to acquire. Additionally, Mintz & Gleitman considered adjectives to be more complex for two reasons: first of all, adjectives represent several conceptual classes, for which the corresponding lexical categorisation is an arbitrary one that varies from one language to another (Mintz & Gleitman, 2002). Consequently, participants are predicted to perform worse on adjectives than on nouns and verbs.

Secondly, the learnability measure will be utilised to see if WS performance follows the pattern predicted by the SICI ratings. Ultimately, this boils down to "Does WS receptive vocabulary development deviate from TD development?", as SICI ratings are expected to be an accurate predictor of WS performance. Some of the previously-discussed studies, such as Thomas et al. (2006), claimed delayed, rather than deviating language ability in WS, leading to the prediction that WS performance and SICI ratings will be highly correlated. This would indicate that WS participants show the same sensitivity to the four SICI criteria as do TD participants, although this would then requisite further investigation with a TD control group to confirm this assumption.

Third, the chronological age factor is considered. Findings such as those by Volterra et al. (1996) suggest the reported relative strength of receptive vocabulary is only exposed at a later age, as they did not find any advantage when testing young children with WS. This would suggest that the initial delay is caught up on, at least to a degree, as WS individuals grow older. As such, the difference between children and adolescents is expected to be much larger than the difference between adolescents and adults. Age group will also be examined by incorporating word class, to discover more about the sensitivity to this factor.

4.2 Materials & Procedure

In this section, the test used will be presented, and some background will be provided on the way in which SICI scores were attained for all test items.

4.2.1 PPVT

The standardised test of receptive vocabulary used for this part of the study was the Peabody Picture Vocabulary Test-III (Dunn & Dunn, 1997), a measure of receptive vocabulary. In this test, participants heard a word and were asked to identify which out of four pictures was the one corresponding to the target word. One such item is presented in Figure 1, where the participant was asked to point to the item representing the word *mowing*.

Figure 1: A sample item from the PPVT (Dunn & Dunn, 1997)

Items were presented in sets of twelve, with a total of seventeen sets or 204 items, with increasing difficulty. The start set was determined based on a participant's age, but if they made more than one mistake in that start set, they needed to go back to an earlier set. Furthermore, the test was discontinued in the event of eight or more errors in a single set. As such, participants would usually not complete all 204 items. An example of the PPVT answer sheet is included in Appendix A.

4.2.2 Survey

In an attempt to establish item difficulty, the learnability of each item was established through scores for each of the four SICI measures. Those scores are not readily available, so consequently, a panel of TD adult native speakers of Dutch was asked to rate each item for shape, individuation, concreteness and imageability. The respondents were not specifically selected for demographics such as age or level of education, as a larger number of respondents was deemed more valuable than a particular background. All respondents participated voluntarily and anonymously, but could enter their e-mail address to take part in a raffle for a book voucher.

In the survey, respondents were presented with items in sets of twelve, while the explanation of each of the criteria was also constantly visible to them. They were asked to rate each item on each of the four measures using a five-point scale. The survey is included in Appendix B. The length of the survey led to many respondents not managing to complete it in its entirety, but each item was rated by at least twenty respondents. Both sum scores (all four factors combined) and separate scores for each of the SICI factors were correlated with the scores per item for the WS participants to test for a relation between the SICI and the difficulty WS participants experience with these items. Prior to this, however, the scores for each of the SICI criteria were correlated with each other to see whether any real difference could be found between scores for one criterion versus another.

4.2.3 Analysis

For the PPVT, participants' responses were listed for each individual item as either correct (1) or incorrect (0) and error percentage was calculated, thus revealing group performance per item. Syntactic category was incorporated, distinguishing verbs, nouns and adjectives. Furthermore, to study the relation between learnability scores and WS performance, error percentages per item were correlated with learnability scores as defined through the SICI continuum that was introduced in section 4.1.1. Finally, an analysis of variance was performed to test for an effect of age group on general performance, as well as performance specific to syntactic category.

4.3 Results

In order to examine the receptive vocabulary abilities of the WS participants, their performance on the PPVT was first of all examined on a word class level, and was correlated to the learnability scores gathered through the SICI-continuum survey.

4.3.1 Syntactic category

As discussed in the hypotheses, performance was expected to be as follows: *noun* > *verb* > *adjective*, based on theories of word learning that argue that concrete nouns are the easiest to acquire. A paired-samples t-test was conducted to compare performance scores for nouns, verbs and adjectives. There was a significant difference in the scores for nouns (M = .68, SD = .07) and adjectives (M = .51, SD = .14); $t(63)$ = 10.79, p < .001, as well as in the scores for verbs (M = .69, SD = .10) and adjectives (M = .51, SD = .14); $t(63)$ = 8.34, p < .001, which is in accordance with the predictions. However, contrary to expectations, no significant difference was found between nouns (M = .68, SD = .07) and verbs (M = .69, SD = .10); $t(63)$ = -.60, p = .548. The analysis was repeated for each age group to test whether the effect persisted in all ages and this, too, yielded significant differences for noun vs. adjective and verb vs. adjective in all groups, as presented in Table 2 below.

Table 2: Paired-samples t-test results for syntactic category for all three age groups

Age group	*Cat. 1*	*M*	*SD*	*Cat. 2*	*M*	*SD*	*df*	*t*	*p*
Children (N = 20)	Nouns	.66	.07	Adjectives	.51	.14	19	5.39	.000
	Nouns	.66	.07	Verbs	.66	.12	19	0.34	.973
	Verbs	.66	.12	Adjectives	.51	.14	19	3.76	.000
Adolescents (N = 21)	Nouns	.69	.07	Adjectives	.48	.15	20	7.68	.000
	Nouns	.69	.07	Verbs	.72	.08	20	-1.07	.300
	Verbs	.72	.08	Adjectives	.48	.15	20	6.15	.000
Adults (N = 23)	Nouns	.69	.05	Adjectives	.52	.12	22	5.82	.000
	Nouns	.69	.05	Verbs	.69	.10	22	-.135	.894
	Verbs	.69	.10	Adjectives	.52	.12	22	6.15	.000
All groups (N = 64)	Nouns	.68	.07	Adjectives	.51	.14	63	10.79	.000
	Nouns	.68	.07	Verbs	.69	.10	63	-.60	.548
	Verbs	.69	.10	Adjectives	.51	.14	63	8.34	.000

These tables show the most striking result for this analysis: for all three age groups, highly significant differences are found between nouns and adjectives and verbs and adjectives with $p < .001$ for all cases, which means that the specific difficulty with adjectives is persistent throughout all age groups. This is further illustrated by the box plot in Figure 2 below.

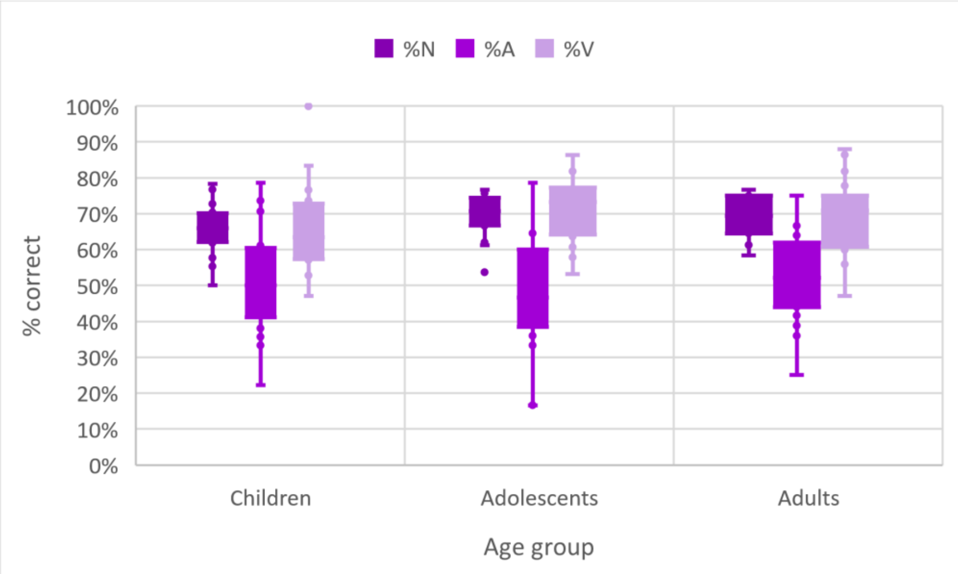

Figure 2: Box plot representing PPVT performance per syntactic category

Another interesting result to remark upon is that a much higher variance in results from all participants is found for adjectives (*SD* = .14) than for nouns (*SD* = .07) and verbs (*SD* = .10), meaning that adjective performance varied greatly between participants, whereas this is not reflected in the other two word categories. As such, while all groups perform significantly worse on adjectives than on nouns and verbs, there is quite some difference between the best and worst results for this category. These findings will be elaborated on in section 4.4.

4.3.2 Across word classes

First of all, the four SICI dimensions were correlated to one another and to the sum score of all four criteria. Descriptive statistics of the data are presented in

Table 3.

Table 3: Descriptive statistics for the SICI criterion analysis (N = 204)

	Mean	*SD*
Shape	3.39	.85
Individuation	3.35	.70

	Mean	*SD*
Concreteness	3.44	.84
Imageability	4.00	.63
SumScore	14.17	2.91
Error percentage	.355	.251

Statistical analysis was carried out by calculating the Pearson correlation coefficient of the SICI-survey results, and revealed a strong and highly significant ($p < .001$) positive correlation between all individual criterion scores and the total scores, indicating that each of the separate scores proved highly predictive of the others at that same level; correlation values are shown in Table 4. This shows that survey respondents found the items to have similar values for all four scales, indicating the dimensions all measure the same construct.

Table 4: Pearson correlation values for total SICI scores and per-criterion scores for all PPVT items

	Sum score	*Shape*	*Individuation*	*Concreteness*	*Imageability*
Sum Score	1	.967	.984	.963	.944
Shape	.967	1	.936	.899	.886
Individuation	.984	.936	1	.943	.923
Concreteness	.963	.899	.943	1	.859
Imageability	.944	.886	.923	.859	1

Overall, per-criterion scores were found to test the same construct, thus obviating the need for a more detailed examination of individual criteria. As such, the sum learnability scores is considered for the remainder of this study.

The second part of this study involved testing whether WS participants' results confirmed the SICI scores' predictions of word difficulty. As the SICI continuum predicts order of acquisition based on only one factor, abstractness, one would expect a significant correlation between performance (as measured by error percentage per item) and SICI score, where items with higher SICI scores are done correctly more frequently. To test this hypothesis, the performance was calculated for each PPVT item and correlated with the SICI score for that respective word; this showed a weak to moderate correlation between performance and SICI score, $r = .43$, $p < .001$. This means that, on the whole, items that received lower SICI scores, also showed lower performance scores for participants.

This analysis was then extended to the age group level, to test how SICI scores relate to performance in each age group. Descriptive statistics for the SICI age group analysis are presented in Table 5.

Table 5: Descriptive statistics for the SICI age group analysis

	N items	*Mean*	*SD*
SICI score	204	14.17	2.91
% Error children	168	.433	.294
% Error adolescents	192	.324	.282
% Error adults	204	.328	.288

Pearson correlation coefficients were then calculated for SICI scores and error percentages per age group. Based on the results of this analysis, SICI score shows a weak negative relation to child error percentage, $r = .312$, $p < .001$, and a weak to moderate relation to both adolescent error percentage ($r = .391$, $p < .001$) and adult error percentage ($r = .417$, $p < .001$). This indicates that SICI scores are more strongly related to error percentages in older participants than in young ones.

However, the analysis presented so far does not yet answer the question whether the SICI scores can, in fact, be considered an accurate predictor of WS participants' performance on an item. To gain more insight into the predictive value of the SICI score, a regression analysis was executed. The regression model with the error percentage as its dependent variable and the SICI score as its independent variable proved to be significant, $F(1, 202) = 45.356$, $p <.01$; as such, the model can be employed to predict the error percentage, but the prediction is relatively weak: 18% of performance can be predicted based on SICI score ($R^2 = 0.183$). A scatter plot of items with their respective error percentage and SICI score is presented in Figure 3.

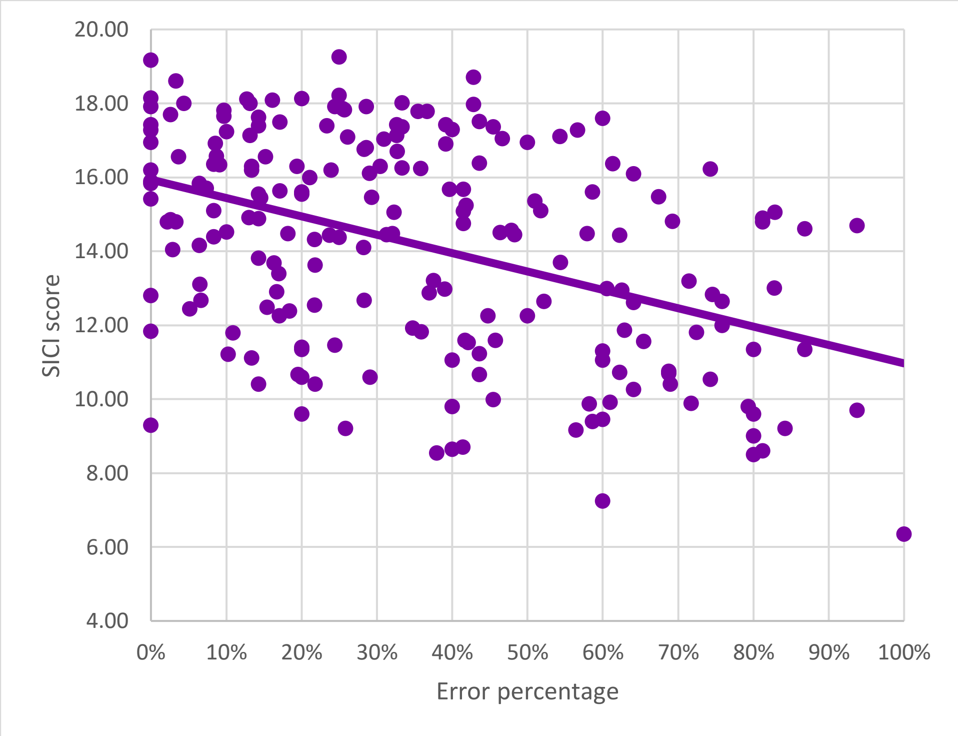

Figure 3: Distribution of PPVT items for error percentage vs. SICI score

The scatter plot in Figure 3 visualises the results of the Pearson correlation analysis; there is a moderate correlation between the two variables. As also discussed before, SICI only measures one factor thought to influence lexical acquisition, and as such, the moderate correlation found concurs with the expectations.

4.4 Discussion

This chapter sought to investigate the receptive vocabulary abilities of individuals with WS, quite possibly the linguistic sub-domain that has been studied most in the past in this group, by examining WS performance on the PPVT on a more detailed level. Specifically, word class and learnability were analysed for their impact on performance scores.

For word class, nouns were predicted to be the easiest to acquire, and thus to yield the lowest error percentage for participants; verbs were expected to be significantly harder than nouns, and finally, adjectives were thought to be the most difficult and thus, to show worst scores. Interestingly, adjective performance was found to be significantly weaker than the other categories in all three age groups, suggesting that this may represent a particular area of weakness in WS participants. It is therefore prudent to elaborate on previous research into adjectives in WS; Clahsen & Temple (2003) and Clahsen et al. (2004) did describe issues with adjectival inflection, but found WS participants to have no issues with the lexical retrieval of adjectives. As was also discussed in §4.1.3, adjectives are much harder to acquire because of factors including the cross-linguistic variability of their application and the whole-object bias (Mintz & Gleitman, 2002). However, this would sooner lead to expect that adjectives represented a particular weakness only in younger participants, but no age effect was found; adjectives were problematic in all age groups.

Furthermore, contrary to the expectations, no significant difference was found between nouns and verbs. The most probable explanation for this seems to be that the models discussed referred to concrete nouns as easy to acquire, but also stipulated that some verbs may in fact be easier to acquire than highly abstract nouns: "Yet, when we consider the range of nouns, from *cup* to *justice*, and verbs, from *running* to *being*, these categories [noun and verb] begin to blur." (Maguire et al., 2008, p. 375).

The aim of the next sub-question was to investigate to what degree WS performance was related to word learnability, as defined by a measure of abstractness, the SICI score. A significant weak to moderate correlation was found between SICI scores and WS performance, and this relation was weakest in children and strongest in adults. As SICI predicts learnability based on only one factor, conceptual abstractness, one would not expect it to fully predict performance scores. As such, these results indicate that conceptual abstractness partially accounts for the WS PPVT performance pattern.

A critical remark on these results would concern the reliability of the SICI survey results. As mentioned, the number of items proved to be a major problem for many of the participants, so it would be recommendable to reduce the items rated to a subset of the total PPVT. Furthermore, the inclusion of a TD control group to examine whether the respondents' rankings of each item concur with TD performance would allow for a better validation of this approach.

Putting methodological concerns aside, however, this study does yield some very interesting results. WS performance does not seem to diverge from what would be expected based on abstractness ratings, so no evidence was found for a divergent, rather than delayed, developmental trajectory in WS. Furthermore, WS participants showed a significant impairment regarding adjective comprehension, which persisted throughout all age groups. These findings will be related to the greater picture in the general discussion in chapter 7.

5. Study 2: Lexical production

5.1 Rationale

Lexical production, like lexical comprehension, has also received a lot of attention in linguistic research of WS. Contrary to its comprehension counterpart, though, the productive domain is so significant because performance is relatively weak. As discussed in greater detail in section 2.5, following findings from multiple studies of WS performing much worse on naming tasks than MA-matched controls, Temple et al. (2002) suggested anomia may be part of the WS profile.

This apparent incredible contrast with the receptive ability gave rise to many questions. For instance, when bearing in mind that the strength on receptive vocabulary was less prominent when the test included semantically related distractors (Temple et al., 2002; Clahsen et al., 2004), the WS lexical strength has been hypothesised earlier in this study to indicate a broad, but shallow, vocabulary. In this chapter, rather than examining the WS performance on the quantitative level as was seen in previous research, data from a naming task will be examined on a qualitative level: if responses were scored as incorrect, what were they? Was there any kind of semantic relation between the target response and the given response? And what does this reveal about the WS mental lexicon? These questions will be addressed after the methodological context of the study has been established.

5.1.1 Quantitative analysis

Similar to the previous chapter, for this chapter, the quantitative analysis of the participants' test scores that was done by Franken (2003) will also be incorporated here; therein also investigating whether these results can be considered to fit in with the findings of earlier studies.

Franken (2003) found WS performance on the naming task used here, the Boston Naming Test (BNT) to be at MA level, which can be considered surprising, as most earlier studies (Volterra et al. (1996) among others) found performance to be below MA level (also see section 2.5 of this thesis). Franken considers a plausible account for this finding to lie in the decision by Volterra et al. to match the control group's CA with the MA of the WS participants, which does not take regular variance in IQ into consideration, Franken argues. This may indeed explain the difference in these findings, though another aspect to consider is the age span of participants in both studies; Volterra et al.'s study concerned young children, whereas this study has a larger age range among its participants. The influence of age will also, once again, be topic of interest for this part of the study.

5.1.2 Qualitative analysis

Contrary to previous studies involving the BNT as a measure of linguistic ability in WS and following recommendations by Temple et al. (2002), for this study, an error analysis will be performed to learn more about the types of errors most frequent in WS performance on this task and how this relates to the error types found in TD groups, as well as what this reveals about the WS lexicon. Temple et al. described the atypicality of the errors produced by their WS participants, and specifically remarked that participants identified a part of an object, rather than the object in its entirety.

5.1.3 Hypotheses

The central question for this chapter is: "What does error analysis of naming task performance reveal about specific weaknesses within the domain of lexical production in WS individuals?". To reach an answer this question, a number of sub-questions need to be discussed. These will be presented hereafter.

The first sub-question is "What error categories are most prominent in WS performance, and how does this compare to the distribution pattern found in TD groups?". In their study of elderly speakers, Mariën et al. (1998) found that one third of all errors made by participants in their study were errors of a semantic nature: responses that were semantically related to the target item. As such, these errors were the most common by far. Storms et al. (2004) found a mostly similar distribution of error types in their study of TD primary school children, with some noteworthy differences: semantic errors once again accounted for about a third of the mistakes, but fewer cases of adequate circumlocutions were reported, versus more inadequate circumlocutions, which may be attributed to the lower age of their participants. As such, Storms et al.'s participants may be most comparable to the WS participants in this study; with a CA distribution of 6-12 (mean: 8.8), these participants are closer to the CA of the WS group, and can be assumed to be closer to their MA as well. As this sub-question seeks to investigate if and how WS performance patterns diverge from those found in TD groups, the most prevalent categories found by Storms et al. can serve as the null-hypothesis for this topic.

Based on findings by Temple et al. (2002) and Clahsen et al. (2004), who reported WS individuals' performance on receptive vocabulary tasks to deteriorate when semantically close distractors were involved, one may expect a high number of semantic errors here. Approximate knowledge of a word can be thought to suffice when picking the correct image in a picture selection task, but may cause them to erroneously identify another member of the same semantic category in a task setting where they are not provided with options to choose from.

Furthermore, the influence of age group will once again be investigated to answer the sub-question: "What influence does age group have on the prevalence of error categories, and what does this disclose about the development of lexical abilities in WS individuals?". As this study focuses on a qualitative analysis of the errors found, rather than comparing overall performance scores, this sub-question will not investigate the general development of lexical abilities. Instead, the error distribution patterns will be compared for age groups to examine the differences and similarities in the types of erroneous responses they are found to produce.

5.2 Materials & Procedure

5.2.1 BNT

The task used for this study is the Boston Naming Test (BNT) (Kaplan, Goodglass, Weintraub, Segal & van Loon-Vervoorn, 2001), a picture-naming task consisting of a total of 60 pictures of animals and objects. For each picture, participants are asked to identify what they see on this picture. The pictures are sketched outlines of objects and animals, and one such example is presented below in Figure 4.

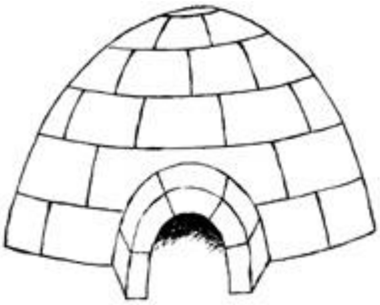

Figure 4: The image corresponding to target item 'iglo' (igloo) for the Boston Naming Test (Kaplan et al., 2001)

5.2.2 Error analysis

In order to retrieve meaningful information from an error analysis, errors need to be categorised for their type. These error categories have been retrieved from a normative study of the BNT in Dutch-speaking elderly people by Mariën, Mampaey, Vervaet and Saerens (1998), which sought to analyse participants' performance on both a quantitative and qualitative level, and were also applied in a normative BNT study featuring TD school children by Storms, Saerens and De Deyn (2004). In Table 6, the error categories are presented with a description of the type of error that belongs to that category, as well as an example of such an error.

Table 6: Error categories BNT as defined by Mariën et al. (1998).
Examples featured have been retrieved from the data of this study.

Name	*Description*	*Example*
1. Phonemic or literal errors	Alterations of the target word through addition, omission, substitution or transposition of the constitutive phonemes.	*ocpus* for *octopus* ('octopus'), *adicon* for *accordeon* ('accordion')
2. Verbal morphological errors	Erroneous lexical items that highly resemble the target words with respect to their phonological structure.	*makreel* ('mackerel') for *kameel* ('camel'), *orkaan* ('hurricane') for *vulkaan* ('volcano')
3. Verbal semantic errors	Erroneous words that semantically relate to the target word.	*vliegtuig* ('aeroplane') for *helikopter* ('helicopter'), *nijlpaard*

Name	*Description*	*Example*
		(‘hippopotamus’) for *neushoorn* (‘rhinoceros’)
4. Verbal or unrelated errors	Erroneous words that share no visual, phonological or conceptual characteristics with the target item.	*zaadje* (‘seed’) for *kameel* (‘camel’)
5. Aborted words	Phonologically or morphologically incomplete responses.	*monica* for *mondharmonica* (‘harmonica’)
6. Non-words or neologisms	Responses that do not belong to the lexicon.	*dalling* for *harp* (‘harp’)
7. Portmanteau words or semantic neologisms	Neologistic combination of lexical elements that have a meaning on their own.	*stofschoonmaker* (‘dust cleaner’) for *bezem/veger* (‘broom’), *ijspakker* (‘ice grabber’) for *tang* (‘tongs’)
8. Perseverations or perservative errors	Recurrent responses that lack any resemblance with the target item presented.	N/A
9. Delayed responses	Chronologically wrong answers that relate to an item presented before and for which the right word was still lacking.	N/A
10. “don’t know” responses	Absence of any response or the expression for not knowing the right target word.	N/A
11. Non-specific utterances	These include empty words, interjections and generalisations, and have no determining quality.	“*Hoe heet ‘t ook al weer*” (‘what’s it called again?’), “*geen banaan*” (‘not a banana’)
12. Adequate circumlocutions	Conceptually correct descriptions of the target item.	“*boom met stekels*” (‘tree with spines’) for *cactus* (‘cactus’), “*kun je rondjes mee tekenen*” (‘you can draw circles with it’) for *passer* (‘compass’)
13. Inadequate circumlocutions	Semantically inadequate or conceptually incomplete descriptions.	“*soort paardje*” (‘a kind of horse’) for *zeepaardje* (‘seahorse’), “*dokters hebben het*” (‘doctors have it’) for *stethoscoop* (‘stethoscope’)

Name	*Description*	*Example*
14. Foreign words	Replacement of the target word by a correct alternative from another language	The English word 'compass' for *passer* ('compass')
15. Visual misperceptions	Misidentification of the target item in the picture.	*slang* ('snake') for *krakeling* ('pretzel'), *vuurwerk* ('fireworks') for *vulkaan* ('volcano')

5.2.3 Procedure

The productive vocabulary abilities were examined through a qualitative analysis of a naming task, the BNT. A total of 909 errors were classified through a neurolinguistic taxonomy consisting of 15 categories (cf. Table 6). Most responses clearly belonged to one category and could thus be coded with ease. Ambiguous cases, responses that could fit into more than one category, were discussed until a consensus could be reached on the most fitting category for the response. Error type 13, which denotes inadequate circumlocutions, was especially complex: responses such as those provided as examples in Table 6 were very obvious examples of inadequate circumlocutions, but responses that contained no meaningful information about the target item were also encountered. For instance, one participant responded "*die heeft mijn zus*" ('my sister has those') for target item *stelten* ('stilts'). Whether this was to be considered an inadequate circumlocution or a non-specific utterance was subject of debate; this was then clarified to be an example of a non-specific utterance after consulting Prof. Dr. Mariën, the lead author of the study from which the error analysis categories were retrieved.

As the test administrator did not systematically record 'don't know'-responses as such, on some occasions no response was recorded while others had *"weet niet"* ('don't know') listed, statistics for this particular error type proved unrepresentative of the number of responses of this kind. It was therefore decided to omit this category from the results presented below in Table 7; therein allowing for comparison with the error distribution found in TD children by Storms et al. (2004). As Storms et al.'s results did include this category, their results needed to be corrected for the exclusion of the 'don't know'-response category. In the subsequent sections of this chapter, the corrected scores will serve as data for comparison to those of our WS participants.

5.3 Results

The application of the classification system introduced in Table 6 to the participants' responses, yielded the data put forward in Table 7. Categories that accounted for less than 1% of the errors found throughout the entire WS group were combined and included as 'other categories'.

Table 7: Error percentages per error type, both in total and for every age group. The final column lists the corrected scores reported by Storms et al. (2004) for their TD children.

Error type	*Group 1*	*Group 2*	*Group 3*	*Total*	*TD*
1. Phonemic or literal errors	3.8%	7.1%	2.2%	4.2%	2.5%
2. Verbal morphological errors	6.0%	6.4%	3.3%	5.1%	0.9%
3. Verbal semantic errors	44.9%	41.8%	44.5%	43.8%	48.4%
4. Verbal or unrelated errors	4.5%	0.7%	0.3%	1.7%	3.5%
7. Portmanteaus/semantic neologisms	4.2%	3.5%	5.0%	4.3%	8.9%
11. Non-specific utterances	0.8%	1.1%	2.2%	1.4%	0.5%
12. Adequate circumlocutions	6.4%	7.4%	10.2%	8.4%	8.6%
13. Inadequate circumlocutions	11.3%	22.3%	22.4%	19.0%	12.2%
15. Visual misperceptions	14.7%	7.8%	8.0%	9.9%	11.3%
Other categories (frequency <1.0%)	3.4%	1.8%	1.9%	2.3%	3.3%

The results in Table 7 show that errors indicating semantic paraphasia (#3) constituted the largest error category, representing 43.8% of all errors analysed. The items with the highest number of semantic errors were *kano* ('canoe') and *bever* ('beaver'). Errors in this category were, very frequently, the naming of another member of the same semantic category. Semantic categories that proved especially prone to this type of error included (but were not limited to) animals, with *nijlpaard* ('hippopotamus') for target word *neushoorn* ('rhinoceros'), *flamingo* ('flamingo') for target *pelikaan* ('pelican') and *paard* ('horse') for target word *eenhoorn* ('unicorn'); and musical instruments, with examples including *piano*

(‘piano’) for target item *accordeon* (‘accordion’) and *fluit* (‘flute’) for target item *mondharmonica* (‘harmonica’).

A specific type of semantic error that was encountered frequently, was the substitution of the target word by its superordinate, such as *vogel* (‘bird’) for *pelikaan* (‘pelican’); *dier* (‘animal’) for *bever* (‘beaver’), *neushoorn* (‘rhinoceros’) and *kameel* (‘camel’); *vis* (‘fish’) for *octopus* (‘octopus’) and *zeepaardje* (‘seahorse’); *instrument* (‘instrument’) or *muziekinstrument* (‘musical instrument’) for *mondharmonica* (‘harmonica’), *harp* (‘harp’) and *accordeon* (‘accordion’). In all these cases, participants correctly identified the semantic category to which the target belonged, but proved unable to name the specific item. The use of superordinate names was not incorporated in the original test design, but rather attracted attention during the process of error analysis. As such, these cases were marked specifically as instances of superordinate use, and were found to account for 8% of all errors, while the entire category of semantic errors accounted for nearly 44%. It was thus found that between 1 in 5 and 1 in 6 semantic errors described occurrences of superordinate use.

Second most frequent were the inadequate circumlocutions (#13), accounting for 19.0% of all errors. The items with most occurrences of this error type were *stelten* (‘stilts’, 13 times) and *stethoscoop* (‘stethoscope’, 12 times). These responses often revealed some knowledge of the target item, often describing functional properties of the target. For instance, for target *stelten* (‘stilts’), most inadequate circumlocutions did contain a reference to walking, but participants did not provide sufficient information for their response to be a conceptually complete description of the target.

The third most frequent type of error found were the visual misperceptions; these accounted for nearly 10% of all errors found. However, these are mistakes of a non-linguistic nature, and as such, are outside the scope of this study.[3]

[3] While not linguistic, a potentially interesting trend was found in these visual misperceptions: participants frequently focused on the wrong aspect of the picture. This confirms reports by Temple et al. (2002), who described a similar observation for their WS participants in naming tasks, and may thus be interesting to researchers investigating the visuospatial abilities of WS, which are known to be a particularly weak area.

5.3.1 Comparison to TD scores

To test for an association between the TD scores reported by Storms et al. (2004) and the WS participants in this study, a Chi-square test for independence was calculated. This revealed a significant interaction for group and error frequency for error categories 1, 2, 4, 7, 11 and 13. χ^2 values are presented in Table 8.

Table 8: Results for a Chi-square test of independence comparing the occurrence of each type of error in Storms et al.'s (2004) group of TD children and the WS group examined in this study.

Error type	*df*	χ^2	*p*
1	1	7.5	.006
2	1	93.7	< .001
3	1	2.5	.116
4	1	7.8	.005
7	1	19.0	< .001
11	1	10.3	.001
12	1	.1	.822
13	1	23.7	< .001
15	1	1.3	.248

As the results in Table 8 show, WS performance differed significantly from the TD children for most error types, excluding error types 3 (semantic errors), 12 (adequate circumlocutions) and 15 (visual misperceptions). WS participants made significantly more errors belonging to types 1 (phonemic errors), 2 (verbal morphological errors), 11 (non-specific utterances) and 13 (inadequate circumlocutions), while TD errors were more often categorised as types 4 (verbal or unrelated errors), 7 (portmanteaus or semantic neologisms) and 15 (visual misperceptions). These results will be discussed more extensively after evaluating the effect of age group on error type prevalence.

5.3.2 Age effects

To answer the sub-question regarding the influence of age group on the prevalence of error categories, an analysis of variance was conducted to compare the effect of age group on the prevalence of each error type. Characteristics for this dataset are presented in Table 9.

Table 9: Descriptive statistics for the BNT age group analysis

Error type	*Age group*	*N*	*Mean*	*SD*
1. Phonemic or literal errors	Children	20	0.042	0.100
	Adolescents	21	0.063	0.080
	Adults	23	0.018	0.033
	Total	64	0.040	0.076
2. Verbal morphological errors	Children	20	0.066	0.054
	Adolescents	21	0.071	0.082
	Adults	23	0.039	0.065
	Total	64	0.058	0.068
3. Verbal semantic errors	Children	20	0.386	0.172
	Adolescents	21	0.370	0.189
	Adults	23	0.416	0.220
	Total	64	0.391	0.194
4. Verbal or unrelated errors	Children	20	0.041	0.067
	Adolescents	21	0.006	0.021
	Adults	23	0.001	0.007
	Total	64	0.015	0.043
7. Portmanteaus or semantic neologisms	Children	20	0.035	0.054
	Adolescents	21	0.035	0.058
	Adults	23	0.052	0.067
	Total	64	0.041	0.060
11. Non-specific utterances	Children	20	0.014	0.064
	Adolescents	21	0.008	0.022
	Adults	23	0.016	0.031
	Total	64	0.013	0.042
12. Adequate circumlocutions	Children	20	0.063	0.082
	Adolescents	21	0.070	0.078
	Adults	23	0.095	0.111
	Total	64	0.077	0.092
13. Inadequate circumlocutions	Children	20	0.085	0.099
	Adolescents	21	0.204	0.129
	Adults	23	0.201	0.151
	Total	64	0.166	0.139
15. Visual misperceptions	Children	20	0.140	0.124
	Adolescents	21	0.076	0.108
	Adults	23	0.075	0.080
	Total	64	0.096	0.107
Other categories (frequency <1.0%)	Children	20	0.057	0.074
	Adolescents	21	0.025	0.039
	Adults	23	0.033	0.052
	Total	64	0.038	0.057

The results showed that for most error types, age group had no significant effect. However, a significant effect was found for age group on verbal or unrelated errors, $F(2, 61) = 6.17$, $p = .004$. Post hoc analysis using a Tukey test revealed the effect to be significant specifically when comparing children and adolescents ($p = .017$) and children and adults ($p = .005$), but not adolescents and adults ($p = .923$). Another significant effect was found for age group on inadequate circumlocutions, $F(2, 61) = 6.17$, $p = .005$. Similar to the previous category, a post hoc Tukey was done here as well, and also revealed a significant age effect between children and adolescents ($p = .012$) and children and adults ($p = .013$), but not adolescents and adults ($p = .996$). Potential explanations for these, as well as other trends will be discussed in the next section.

5.4 Discussion

For this chapter, the aim was to discover what can be learnt about the organisation of the WS mental lexicon from an error analysis of WS responses for a naming task. All erroneous responses were coded according to a coding scheme presented by Mariën et al. (1998) in a study of elderly speakers of Dutch, which was also employed by Storms et al. (2004) in a study of primary school children. The latter study was used to compare the WS error distribution pattern to TD children, and this yielded some interesting results.

WS performance was found to diverge from the distribution pattern reported by Storms et al. (2004) for primary school children for some, but not all error types. WS participants produced more errors in which one or more phonemes are replaced or omitted, and were also found to more frequently produce words that highly resemble the target item in terms of its morphological structure. This raises questions about the storage of phonological information for the items in their mental lexicon: mistakes such as responding with '*ocpus*' for target *octopus* ('octopus'), where the second syllable is omitted, seem to indicate the participant retrieved the lexical entry belonging to the correct concept, but had stored an inaccurate phonological representation of the target word. In addition, WS individuals produced fewer semantic neologisms and portmanteau words than the TD children, while producing more inadequate circumlocutions. This could indicate that, when unable to come up with the correct name for the target, TD children made up a name for the concept they are presented with, while WS participants attempted to describe the concept, rather than inventing a name for it. Finally, the WS distribution of errors was found to be on par with the TD children as far as the most prominent error type, semantic errors, was concerned. It can therefore be concluded that the hypothesised shallow nature of the WS vocabulary did not lead to them producing more semantic errors than the TD children in the study by Storms et al. (2004).

The prominence of semantic errors is interesting, as they reveal some degree of knowledge of the target item, even if the participant was not, in fact, able to produce the correct word that corresponds to the image. As reported in section 5.3, participants were often found to name the semantic category the item belonged to by responding with its superordinate, identifying items such as 'harmonica' and 'harp' simply by calling it a musical instrument, or identifying items such as 'beaver' and 'camel' only as animals. While this particular finding cannot be compared to TD performance, as neither of the TD studies employing the same error analysis approach reported any such results, it does spark interest. Relating back to the suggestion of shallow vocabulary knowledge, these findings may be explained to show that the lexical entry for an item does not contain sufficiently detailed information for the participant to be able to select the correct entry out of a number of related words for the item in question, and thus only identifies the semantic category the target belongs to, such as 'animal', 'fish' or 'musical instrument'. However, as semantic errors in general were equally prominent in the TD group, further research is required to determine if and how WS semantic errors diverge from those found in TD, and in turn, whether this signifies a divergent organisation of the mental lexicon in WS.

Proceeding to the second topic of research, the influence of CA on the error distribution patterns, some remarkable trends were observed for verbal errors, which was observed more frequently in children, and inadequate circumlocutions, for which a higher prevalence was reported in adolescents and adults. Combining these two significant age effects, an interesting image is found: children are very prolific in coming up with single-word erroneous responses, that may in fact be meant to describe the target but simply lack context to be recognised as such, while as they grow older, they gain the ability to provide a longer, if still inadequate, description of the target. Therefore, these findings may attest to WS participants attaining a different solution in the event of a word-finding problem as their general cognitive development advances with age. This will be discussed in greater detail in chapter 7.

On the whole, the errors produced by participants often revealed they had some knowledge of the target. As said, many of the semantic errors (#3) involved naming other members of the same semantic category or identifying a superordinate, but the participants also showed remarkable creativity in formulating mostly inadequate circumlocutions (#13) to describe the object they saw in the picture, coming up with circumlocutions like "*paard met wortel op zijn hoofd*" ('horse with carrot on its head') for target word *eenhoorn* ('unicorn') and "*stoel voor oude vrouwen*" ('chair for elderly women') for target *rolstoel* ('wheelchair'). What all of these have in common, is that through these, the participants clearly display that despite being either unfamiliar with the particular object (which may be the case for 'unicorn') or not knowing what it is called (which seems to be true for 'wheelchair'), they are able to formulate a response based on associations and features of the object they recognise, and become more and more proficient in solving this word finding problem as they grow older, by coming up with descriptions of whatever it is they are unable to name exactly.

These findings concur with the expectations based on both the TD data and WS literature, but also raise even more questions that cannot be answered by the evidence currently available. This gives rise to some critical observations regarding the current study, as well as calling for further research into the matter.

First of all, the incidence of particular the incidence of superordinates, which accounted for between 1 in 5 and 1 in 6 cases of semantic paraphasia, cannot be compared to control group scores, as this was not specifically tested or discussed in either of the TD studies employing the same classification. However, this was more of a *post-hoc* finding in this study, it was therefore not originally part of the test design. Future research into the use of superordinates by both WS and TD groups would therefore need to adjust the experimental design to investigate this particular topic. Based on this study, no conclusions can be drawn on whether this is typical of WS performance, or may occur to the same degree in other populations.

In addition, findings regarding inadequate circumlocutions addressed the same issues regarding control groups, and, similar to the previously-discussed use of superordinates, observations regarding inadequate circumlocutions also call for further investigation. WS participants were found more prolific regarding the production of inadequate circumlocutions than the TD participants described in Storms et al. (2004), and further analysis revealed this particular error type to occur more frequently in adolescents and adults with WS than in WS children. However, no further analysis of the nature of each circumlocution was pursued as this did not fit the current study's scope. Furthermore, to the author's knowledge, no such studies have been executed into the nature of inadequate circumlocutions in TD participants either.

While the task at hand was a naming task, and thus, target responses were single words, the production of circumlocutive responses as described in this chapter exhibited a degree of lexicosemantic abilities, as some of these responses were reminiscent of word definitions instead. It will therefore be especially interesting to examine the responses given by these individuals on a definitions task, which is what the following chapter will investigate.

6. Study 3: Semantic abilities

6.1 Rationale

In the previous chapters, both production and comprehension of individual words have been examined. For this chapter, however, focus will be on a skill that elicits not only lexical, but also semantic abilities: defining words. Charkova (2003) accurately pointed out what makes definition tasks so fascinating and complex, explaining what is required to formulate a word definition: "[…] *requires cognitive maturity and semantic knowledge as well as familiarity with the conventional syntactic form.*" (p. 506). This immediately introduces an issue described for the use of word-definition tasks as a measure of linguistic ability in WS populations: potential interference of metacognitive demands. Before discussing this, however, the results for this test as described by Franken (2003) will be presented.

6.1.1 Quantitative analysis

Franken predicted the WS individuals to perform below MA level, which would correspond with previous findings for said test in WS in other languages. This prediction proved accurate: in accordance with earlier studies, WS participants performed well below what might be expected based on their general MA. Franken's results contributed to the general impression of weak WS performance on word-definition tasks.

The generally weak performance on these tasks has led several researchers to question the applicability of word-definition tasks as a measure of linguistic ability in WS, as discussed in §2.6. This was tested by Purser et al. (2011), who compared WS and TD groups on two tasks: one was a standard word-definition, the other was a categorisation task. They found the word-definition task underestimated WS lexicosemantic knowledge and argued the task's metalinguistic demands obscured performance. However, even when taking these objections into account, the task does not need to be disregarded altogether; even without knowledge of formal definition structure, responses on such a task may still contain a great wealth of information. Semantic ability, as elicited by a word-definition task, thus warrants an alternative, qualitative approach to illuminate the organisation of the mental lexicon.

6.1.2 Qualitative analysis

Following our findings of the previous chapter, where participants would sometimes give descriptions or definitions of target items rather than naming them (which was required by the naming task they performed), and where it was remarked that even the inadequate descriptions or circumlocutions would often contain accurate characteristics of the target item (yet not sufficient to allow correct identification of the target), this chapter will delve deeper into the definitions provided by participants to investigate what information lies therein. As such, we will seek to describe what WS participants' responses on a definitions task reveal about the mechanisms underlying their semantic abilities. Responses will therefore be analysed on a qualitative level, investigating the nature of the features identified by participants.

Qualitative analysis of definitions has been done in a number of studies in the past decades. Understanding and application of formal definition structure (i.e. the way in which definitions are presented in dictionaries, featuring a superordinate category and several characteristics), was reported to be found in children "about or older than age 10 and adults" (Charkova, 2003, p. 507).

In one such qualitative study, Al-Issa (1969) tested children aged 5-10 and classified features into three categories: concrete or descriptive, functional, and categorical or abstract. Al-Issa found that functional features were produced most frequently by younger children and decreased with age, while the opposite was observed for abstract features, which increased with age and were the most frequently produced feature type in the oldest children in the study. However, Al-Issa did not report on whether feature types were combined in definitions, or which types of features often coincided in a single definition.

This question was investigated by Wehren et al. (1981); they reported that as the age of the children increased, so did the number of definitions containing more than one feature. Furthermore, they found an emphasis on functional features in the definitions given by the youngest children, while older children and adults would more often include both functional and perceptual features of the target object, and were also described to include superordinate terms in their definition. These were all studies of TD groups, though, raising questions regarding definitions in special populations such as the WS group in this study.

6.1.3 Hypotheses

In this chapter, responses from a definitions task are analysed for the features contained therein to answer the following central question: "What does qualitative analysis of the definitions provided in a word-definition task reveal about the lexicosemantic abilities of WS individuals and the WS mental lexicon?". this chapter will seek to answer the sub-questions discussed hereafter.

First, the distribution over different categories will be investigated and distribution patterns will be compared for all three age groups to answer the sub-question: "What types of features are encountered most frequently in definitions produced by WS participants, and how does this develop with age?". The studies discussed in §6.1.2 reported on the developmental trajectories witnessed in TD groups, where younger children were found to emphasise function, while perceptual features and superordinate terms were observed more frequently in definitions by older children and adults. WS performance will be examined to discover what tendencies are found there, as this will also shed light on how WS performance diverges from the TD performance as reported by Al-Issa (1969) and Wehren et al. (1981). As formal definition structure was attributed to older children and adults (Charkova, 2003), WS participants are not expected to produce definitions that follow this structure.

Furthermore, the findings of the previous chapter will be kept in mind, where it was suggested that Temple et al.'s (2002) suggestion of an impoverished fine-grained semantic ability may be extended to lexical production, as this could indicate WS vocabulary to be relatively shallow. For a definitions task like this one, a lack of vocabulary depth could be found to lead to participants identifying fewer features as the items become more difficult.

Finally, to give direction to future research into WS performance on word-definition tasks, general observations will also be presented.

6.2 Materials & Procedure

6.2.1 Tasks

The word-definition task examined in this study is the vocabulary subtask of the Wechsler Preschool and Primary Scale of Intelligence Revised edition (WPPSI-R) (Wechsler, 1990). In this task, participants were asked to provide verbal definitions for target items, one at a time. Depending on the quality of the response, they were awarded 2, 1 or 0 points per item. The participant was first shown three images for which they were asked to identify what they saw in the picture, similar to the naming test discussed in the previous chapter. A stop rule was applied in the examination: in the event that a participant answered four consecutive items incorrectly (i.e. was awarded 0 points), the task was discontinued.

6.2.2 Feature analysis

To analyse the responses given by the participants, the responses were first divided into individual features of the target items. For instance, if the item to be defined had been ‘donkey’, and the participant had responded with “That’s an animal with four legs that works the land”, the response would be analysed as containing three pieces of information on the target item, or three features: it is an animal; it has four legs; and finally, it works the land. All responses were analysed in this manner, and those features were then coded for the type of information they contained.

For the analysis of the features, the aforementioned studies by Al-Issa (1969) and Purser et al. (2011) were used as a guideline. Purser et al. first distinguished between correct and incorrect features, which was also the procedure for this study. While Purser et al. chose to categorise features as salient and diagnostic, these categories have been excluded from the study at hand as they are very subjective, and thus hard to define. Instead, the decision was made to adopt the categories from Al-Issa, with some slight adjustments: the category ‘function’ was adopted as is, but rather than ‘descriptive’ or ‘concrete’, the label ‘perceptual’ was used for these features, in accordance with the categories Purser et al. used.

Additionally, Al-Issa and Purser et al. used ‘abstract’ as a category, and for this study, that category was divided into two separate categories; the first one, ‘taxonomic features’, was included to investigate reports of increasing use of superordinate terms as participants grow older. The second category, ‘contextual features’, was added to accommodate the creativity that was reported for the inadequate circumlocutions in the naming task in chapter 5. This decision was further supported by the difference in item type between the WPPSI-R and the task in Purser et al.’s study, as the latter involved only animals as items, whereas this study involved a whole range of semantic categories. As such, ‘abstract’ seemed a little too broad a category for the current study’s purposes. The categories were further tested and defined throughout the process of analysing and coding the data.

For each response, individual features of the target item were identified and coded as belonging to one of four categories. These categories are described in more detail in Table 10, below. To illustrate the categories, examples of all kinds of features are presented. These examples are fictional and must not be considered a representation of actual data; examples from actual data will be presented in the results in section 6.3.

Table 10: Description of feature categories and examples of features belonging to these categories.

Category	Description	Example
Perceptual features	These include any features that can be readily observed from a mental image of the target item. These are expected to be mostly visible characteristics (1a), but there may also a number of audible features (1b).	(1) a. It has four legs b. It brays
Taxonomic features	These include hypernyms (or superordinates) (2a), which were also discussed in chapter 5, but also hyponyms (or subordinates) and synonyms (2b).	(2) a. It is an animal. b. It is an ass.
Functional features	These include any features that are related to the function or purpose of the target item (3a-b).	(3) a. It can carry supplies. b. It can work the land.
Contextual features	These include features related to the context or situation in which the target item occurs (4a-b).	(4) a. It lives on a farm. b. They have them at the zoo.

6.2.3 Data selection

The data acquired from the task contained nouns, verbs and adjectives. However, the items were not distributed evenly over all three word categories; the WPPSI-R consists of 17 nouns, three verbs and two adjectives. As a result, the tasks contained very little data for verbs and adjectives; this was deemed too small a sample to properly study the definition of items belonging to those syntactic categories. In addition, the studies of word-definition tasks in TD groups were also restricted to the definition of nouns; consequently, the decision was made to exclude the other syntactic categories so a more accurate comparison between this study and other previous reports could be made.

Furthermore, only 15 participants were considered to have a high enough level of cognitive ability to do the WISC-RN instead of the WPPSI-R (which was done by 49 participants). Here, too, sample size was a concern, so this study only focuses on the WPPSI-R responses.

6.3 Results

Responses by 49 participants were examined for no more than 17 items, with the total number of items per participant depending on their general performance and the application of the stop rule discussed in section 6.2.1. Two participants did not produce a single correct feature, and were therefore excluded from further analysis. Characteristics of the entire data set are presented in Table 11 below.

Table 11: Descriptive statistics for the data analysed for the definitions task (N = 47)

Category	*Age group*	*N*	*Mean*	*SD*
Total features produced	Children	18	23.5	9.17
	Adolescents	15	30.13	10.01
	Adults	14	31.64	13.85
	Total	47	28.04	11.37
Correct features produced	Children	18	17.28	7.23
	Adolescents	15	24.53	9.97
	Adults	14	26.07	13.20
	Total	47	22.21	10.71
Perceptual features	Children	18	3.67	3.09
	Adolescents	15	5.93	6.15
	Adults	14	3.93	4.01
	Total	47	4.47	4.54
Taxonomic features	Children	18	1.06	0.94
	Adolescents	15	1.8	1.61
	Adults	14	3.5	1.91
	Total	47	2.02	1.79
Functional features	Children	18	8.78	3.54
	Adolescents	15	11.87	4.75
	Adults	14	12.21	5.21
	Total	47	10.79	4.66
Contextual features	Children	18	3.78	2.82
	Adolescents	15	4.93	2.49
	Adults	14	6.43	3.92
	Total	47	4.94	3.22

The numbers in Table 11 show a great degree of variance for each characteristic; this is especially so for perceptual features, where one participant identified 23 perceptual features, but the average over all 47 participants was only 4.47 perceptual features per participant. Following these findings, the dataset was analysed for outliers for each of the feature types. Correcting for these outliers resulted in the dataset in Table 12.

Table 12: Descriptives for the corrected data, after excluding outliers per feature type

Category	*Age group*	*N*	*Mean*	*SD*
Perceptual features	Children	15	2.47	1.302
	Adolescents	13	3.92	2.929
	Adults	14	3.93	4.009
	Total	42	3.4	2.947
Taxonomic features	Children	18	1.06	0.938
	Adolescents	14	1.5	1.16
	Adults	13	3.15	1.463
	Total	45	1.8	1.455
Functional features	Children	18	8.78	3.541
	Adolescents	15	11.87	4.749
	Adults	13	11	2.646
	Total	46	10.41	3.936
Contextual features	Children	18	3.78	2.819
	Adolescents	15	4.93	2.492
	Adults	13	5.69	2.898
	Total	46	4.7	2.796

Even after correcting for outliers, the data still reveal a great degree of variation: some participants were found to be incredibly prolific and produce multiple features for each item, while others only produced one, perhaps two features per item. As the number of items also varied between participants, the average number of features per item was calculated, as well as the average number of correct features per item. These statistics were then correlated to test if the participants that were more prolific, also produced more correct features; this way, it was made sure that the impression of productivity was not the result of participants naming an incredibly high number of incorrect features. Based on the results of the study, the average number of features is strongly related to the average number of correct features produced; $r = .87$, $p < .001$.

Relating these findings to the sub-question regarding performance when vocabulary depth is tested, it can be said that these results do not unequivocally support the impression of a broad, yet shallow vocabulary. This will be elaborated on in section 6.4.

6.3.1 Age effects

The first sub-question concerned itself with whether an age effect could be observed on the frequency with which feature categories were found. To this end, an analysis of variance was conducted to test the effect of age group on the prevalence of each feature type. This showed that the effect of age group on the number of taxonomic features produced was significant, $F(2, 42) = 12.665$, $p < .001$. A post hoc Tukey test revealed the effect to apply for children and adults ($p < .001$) and adolescents and adults ($p = .002$), but not children and adolescents ($p = .544$). This indicates that adults produced significantly more taxonomic features than children and adolescents. In this respect, WS participants did not diverge from the performance reported by Wehren et al. (1981) for TD groups; findings from that study also indicated use of superordinate features to be characteristic of older children and adults.

Following the findings reported by Al-Issa (1969) and Wehren et al. (1981), a significant age effect was also anticipated for the number of functional features produced, but this effect fell short of statistical significance; $F(2, 43) = 2.958$, $p = .063$. Additionally, the effect expected was a decrease of these features, but the means for each age group in Table 12 indicate the adolescents were actually found to produce more functional features than their younger counterparts in the child group.

6.3.2 Associative approach

In this section, the focus will be on the explorative analysis of the data. Participants' responses on the definitions task revealed that, even if they did not quite grasp the standard structure of a definition (where one might expect diagnostic and salient features before other features), they did have some idea of what the target item referred to. This was shown through the inclusion of many correct features; as Table 11 showed, on average, participants produced 22 correct features. This translates to an average of 1.6 per item answered.

These correct features showed an emphasis on functional, rather than perceptual features, as presented in Table 12: nearly half of the features were of a functional nature. For target item 'mes' (knife) for instance, participants often described how it could be used to cut things, such as bread or vegetables; target item 'fiets' (bicycle) elicited responses on how this could be used for cycling and how you could sit on it and go places. Only 20% of all features were perceptual; some examples include: for 'mes' (knife), participants described its sharp and (in their definition, where it seems they thought of a bread knife) serrated blade and its handle; for 'fiets' (bicycle), they described features such as the handlebar, pedals, a metal frame, two wheels and a saddle. Whether WS performance deviates from TD in this regard cannot be determined at this stage, as none of the TD studies reported discussed the definitions their participants gave into such detail.

This focus on function supports a trend witnessed throughout the results: the seemingly strong associative ability that appeared to reveal itself in most responses for this test. This is further supported by the prominence of contextual features: this category was the second most prominent and accounted for 10% of all features. In these cases, rather than defining items based on their perceptual features, participants described the associations they had with the item in particular. Responses indicating this line of reasoning occurred for *paraplu* ('umbrella'), where some participants were found to describe the situation as a whole, rather than only its function, such as the following (translated) response: "if you're going somewhere, you bring an umbrella in case it's going to rain.". A similar approach was found for *brief* ('letter'), for which one participant gave the following (translated) response: "If the mail man visits, you have to put it in a mailbox.", and a number of other responses involved the option of sending a letter to someone that is ill or celebrating their birthday.

Interestingly, some items seemed to elicit participants' fantasy: target item *kasteel* ('castle') for example resulted in many participants describing it as a place with kings and queens, but also knights and dragons. Such responses did confirm the expectations regarding definition structure: participants did not adhere to the formal definition structure of identifying the superordinate category 'castle' belongs to, then naming diagnostic features thereof; instead, they were found to reply based on their associations with the target. This was also seen for target item *held* ('hero'), where a participant simply replied by mentioning 'Superman', rather than defining what a hero is or does.

The least prominent were taxonomic features. However, while this feature was found rarely overall, it was very prominent for two particular items: target item *mot* ('moth') was already discussed before as almost exclusively eliciting responses on how this is an animal, and target item *feestdag* ('holiday') almost exclusively resulted in responses featuring only specific holidays such as Christmas and Queen's Day.

6.4 Discussion

In this chapter, responses from a word-definition task were examined to learn more about the types of features used and to discern what this tells about the organisation of the mental lexicon in WS individuals. First of all, a strong emphasis on function has become apparent: while Purser et al. (2011) found WS participants to produce more perceptual than abstract features, this trend was not found in the current study. As Purser et al.'s participants had a CA range of 12;0-44;11, they would be most comparable to the adolescents and adults in this study, but these groups did not show a trend towards perceptual features either. The most likely explanation for this deviating result is that Purser et al.'s study involved only animals as target items, which may elicit fewer functional features than the items in this study, most of which were objects.

The age effect described in previous TD studies, where fewer functional features were reported in older participant groups, was not found here either. This could indicate that WS participants do not develop beyond the infantile, functional level; perhaps WS individuals' ceiling level in this regard is comparable to that of the younger children in the studies by Al-Issa (1969) and Wehren et al. (1981). However, another account could be a different perspective on word definition altogether, where the functional features identified by adolescent and adult WS participants might be more complex than those found in their younger WS counterparts and in young TD children, but without shifting attention to other feature levels as was reported in older TD children. On the other hand, another feature type did show the same pattern of development as was reported for TD participants: WS adults were found to produce significantly more taxonomic features than the children and adolescents in this study. Ultimately, while the attention does not move away from the functional features as was reported in TD study, older WS participants do add more taxonomic features to their definitions.

Considering the hypothesised shallow, yet broad vocabulary in WS individuals, the high variability in the number of features produced was unexpected. While this cannot be compared to TD as no data is available for the number or quality of features produced by TD participants for this task, this study's findings do show that performance on a task that involves vocabulary depth, such as a word-definition task, is not characterised by a consistently low number of features produced by participants. Instead, the results give the impression of a high discrepancy in performance levels between participants, which could not be linked to CA. It would be fascinating to examine performance on an even more detailed level by developing a scale of per-feature quality, assessing the depth of each feature, to examine whether the quality of the features shows the same pattern of variability as was found in this study for the feature quantity. Assessing TD groups along the same scale would allow conclusions on how the depth of the vocabulary, and consequently, the organisation of the mental lexicon, in WS diverges from that found in TD individuals.

In addition to the impression of an associative approach discussed above, the question posed in 6.1.3 still needs to be answered: how do these findings reflect on the assumed weakness regarding fine-grained semantic knowledge, put forward by Temple et al. (2002)? The predicted pattern of features becoming less specific as items become more difficult, was indeed found in some participants: for target item *microscoop* ('microscope'), several participants identified it as something one can look through, without detailing what it is used for exactly or in what kind of environment it is used. However, this did not only occur for the most difficult items; for both *schoen* ('shoe') and *hoed* ('hat'), some participants only stated (translated) "you can put it on", without giving further defining characteristics of either item. These responses give the impression that participants were aware of the semantic category the items belonged to, but could not give more detailed descriptions. As this could also indicate a lack of understanding of what is required to formulate a proper definition, responses of those participants were examined to see if other ones followed the same pattern. This was the case for some, but certainly not all participants that gave inaccurate replies for either piece of clothing. As such, the evidence on the lack of depth in WS vocabulary is inconclusive; the general impression left by WS performance on a word-definition task is that of shallow definitions and little attention for detail, but this did not apply to all participants. Furthermore, whether this is to be attributed to the organisation of the mental lexicon, or to metalinguistic and cognitive factors, cannot be determined at this stage.

In conclusion, WS participants did diverge from their TD counterparts in that the emphasis on function, which was found to disappear with age in TD participants, remained quite constant in all WS age groups. On the other hand, older WS participants did produce more taxonomic features, an age effect that was also reported in previous TD studies. WS participants were found to rely strongly on association in their performance on definitions tasks, and thus, this seems to be the drive behind their semantic abilities. While further investigation is required to determine whether this is part of the WS phenotype or simply a consequence of the metalinguistic demands of the test, these results do not rule out the assumption of the issues with fine-grained semantic ability extending beyond the receptive domain, and inspire further research into the WS mental lexicon by examining the qualitative nature of the features produced on an even more detailed level.

7. General discussion

The aim of this thesis was to investigate what performance on measures of lexical and lexicosemantic abilities revealed about the organisation of the mental lexicon in individuals with Williams Syndrome. This was investigated on both a quantitative and a qualitative level; for lexical comprehension, performance on the PPVT (a picture-selection task) was examined for effects of syntactic category, age group and learnability measures. The study of lexical production investigated the nature of the errors made for the BNT (a naming task), how the errors were distributed over the categories and how WS compared to TD counterparts in this regard and whether age group had a significant effect on the prevalence of particular error categories. Finally, for semantic abilities, a word-definition task was analysed for types of features produced, as well as age group effects in this respect.

Before connecting the results from the three sub-studies, the most remarkable findings are summarised. In the first sub-study, adjectives appeared a particular area of weakness in WS performance. While this could be expected for the youngest participants, as adjectives are predicted to be acquired later than nouns and verbs, what was most striking was that this effect persisted in all age groups. Previous studies of adjectives in WS only reported issues with adjectival inflection, but found no impairment in lexical retrieval and production, making this a fascinating finding that calls for further examination of the comprehension and production of this syntactic category in WS.

In the second sub-study, great creativity was observed in the production of circumlocutions as a solution to word-finding problems, and WS participants were found to produce these circumlocutions more frequently than their TD counterparts in previous research by Storms et al. (2004). On the other hand, WS participants produced fewer portmanteaus and semantic neologisms, suggesting WS groups favoured describing the target over inventing a novel word based on the target item’s characteristics. In addition, WS participants produced more errors of the phonemic and morphological types, which led to hypothesise an impairment in the storage and retrieval of phonological information for lexical entries. The high prevalence of semantic errors in the WS data was interpreted to support the impression of a broad, but shallow nature of vocabulary in WS participants.

Finally, the third sub-study revealed some interesting differences between feature type distribution as found in WS data, compared to what was found in TD data in earlier studies. Strong emphasis was observed on functional characteristics in the definitions given by WS participants, which, contrary to reports from TD studies, did not decrease for older participants. While this could imply that perhaps, WS development reaches its ceiling level at the functional stage that is witnessed in young TD children, another finding would argue against this: TD studies found an increase in the number of taxonomic features as participants grew older, and this effect was also found in the WS data, where adults produced a far higher number of those features than the other groups did. Furthermore, an associative approach was observed, as WS individuals were found to define words using their personal associations with the concepts, rather than merely identifying general characteristics thereof. Whether this approach is unique to WS or can also be observed in TD groups is, however, unclear, as none of the TD word-definition studies reported on any such findings.

A central focus of all studies was the weakness of fine-grained semantic ability Temple et al. (2002) found in WS on receptive vocabulary tasks; this was interpreted to indicate a broad, yet shallow vocabulary in these individuals. For lexical production, the erroneous responses analysed did indicate participants had some knowledge of the target item, and would often name other members of the same semantic category or provide circumlocutions that, in themselves, were truthful, but were too inaccurate to be considered adequate. This may indicate that the underlying lexical entries for items lack detailed info, resulting in the selection of the wrong member of a semantic category, or providing a circumlocution lacking specificity. For the word-definition task, a high degree of variability was observed in the number of features produced, which led to suggest that if vocabulary depth is indeed a weakness in WS, the degree to which this is an issue, varies between members of the WS population. These findings give reason to believe Temple et al.'s finding may not only apply to the receptive vocabulary sub-domain as they suggested, but rather be characteristic of lexical abilities in WS in a broader sense. This will be elaborated on in the following sub-sections of this chapter.

Based on the findings of all three sub-studies, the general presentation of lexical abilities in WS can be summarised through a number of main points, which will be discussed here. First of all, as discussed above, findings of this study support the impression of WS participants relying on approximate knowledge in their lexical abilities. While this suffices in receptive vocabulary tasks such as the PPVT, where distractors are not strongly related on a semantic level, it is more problematic on a production level, which is more demanding as participants are not offered options to choose from, but rather have to rely on own vocabulary. As a result, in lexical production, WS individuals are frequently found to provide responses in which they confuse the target with a member of the same semantic category or in which they describe some general characteristics of the target's category, without providing diagnostic features that distinguish the target from the other category members. In other words: they have and show some knowledge of the target, but do not quite reach the level of precision required especially in the production of lexical items.

Secondly, this study has led to suggest that WS participants may rely strongly on an associative approach to lexical knowledge, as they were often found to describe situations and contexts in which objects appeared, rather than focusing on the object itself. While no such performance was reported on in TD studies, at this time, it cannot be said if this is unique to the WS phenotype. When relating this to the general cognitive profile reported for WS, one could imagine that the weak visuospatial abilities considered characteristic of these individuals, result in less attention for perceptual details of items. In this study, it seemed that these perceptual features were replaced by contextual or situational information. At times, their language showed a great degree of creativity or improvisation, especially when trying to circumvent a hiatus in their knowledge (such as not knowing the name for an object). Rather than forming a semantic neologism when unable to name something, as was found more frequently in TD children, WS participants came up with descriptions and circumlocutions of the target in question. This may in fact have contributed to the initial linguistic research on WS reporting individuals with exceptionally well-developed language skills considering their cognitive ability, as the creative production could be thought to lead attention away from verbal limitations. In short: what these individuals seem to lack in semantic and lexical precision, they conceal and, to some degree, compensate for through resourcefulness and creativity in circumventing these problems.

Previous linguistic research of WS often concentrated on psychometric test scores and would usually involve fewer participants, providing a general impression of the linguistic ability based on these tests, but leaving at least as many questions unanswered regarding the finer mechanisms of their language skills. This study has sought to take the next step in WS language research by looking beyond test scores and zooming in on the wealth of knowledge contained inside the responses on those psychometric measures through qualitative analyses of these tests.

The ambitious intentions behind this research have led to some limitations that need to be taken into account when considering the findings of this study. This study explored the language of WS individuals and did not feature a TD control group; instead, TD norm data was employed when available, and conclusions from comparable TD research were reported to allow for a comparison between WS performance and TD performance in previous studies. Naturally, this does limit the ability to compare the groups, as methodological differences may confound findings.

Another point of improvement for this study is that the data was not collected by the author; instead, the filled out test forms were analysed. It would be preferable to have the tests administered by (one of) the author(s), as this would allow including further impressions from the test administration process, which may not be visible in the test transcripts.

Taking these limitations into consideration, this study contributes to the current state of WS research by providing findings that suggest previous theories of weak semantic precision on receptive vocabulary tasks may in fact play a part in the entire lexical domain, and as such, would hope to inspire further research in that direction. An especially interesting topic related to fine-grained semantic ability would be a more detailed qualitative examination of the circumlocutions produced by participants on the BNT, as the nature of these responses should reveal more about the approach chosen by WS individuals when the target name cannot be retrieved.

The second recommended line of further research, which can also be pursued for the previous topic, is to investigate whether the qualitative findings from this study are typical for the WS phenotype. To this end, a test design featuring a WS group as well as MA-matched TD controls and MA-matched peers with a different developmental disorder (e.g. Down Syndrome) would be ideal, comparing performance for all groups on the same production task (such as the BNT and the WISC-R Vocabulary task). This would rule out confounding factors such as differences in target items and diverging interpretation of error or feature category, and would thus allow for strongly supported conclusions on what can be considered typical for WS.

Finally, the unexpected findings regarding a weakness specific to the syntactic category of adjectives would also inspire further investigation. As previous research on adjective performance (cf. Clahsen & Temple, 2003) in WS reported no issues with the lexical retrieval of said items, and only found issues regarding adjectival inflection, this does raise questions on how the apparent impairment can be thought to manifest differently in varying task designs. Taking into consideration that the performance pattern persisted in all age groups, a study focusing specifically on the comprehension of different types of adjectives, should feature a broad sample of WS participants. This would allow localisation of the specific adjectival category forming a problem. Considering the visuospatial weakness reported for WS, they may be found to find adjectives describing, for instance, shape and size properties of objects especially difficult to comprehend.

The language domain in individuals with Williams Syndrome has inspired linguistic research for nearly three decades, and for good reason. The complex profile of strengths and weaknesses witnessed in these individuals is incredibly fascinating, especially when comparing WS to other disorders with a similar degree of cognitive impairment. With every study, a little more is learnt about the language mechanisms, and more questions are raised to be investigated in further studies. Time will tell how much more can be learnt about the workings of language from the studies of these intriguing people.

References

Aitchinson, J. (2012). *Words in the Mind: An Introduction to the Mental Lexicon* (4th ed.). John Wiley & Sons, Incorporated. Retrieved from http://www.uunl.eblib.com.proxy.library.uu.nl/patron/FullRecord.aspx?p=821664

Al-Issa, I. (1969). The development of word definition in children. *The Journal of Genetic Psychology*, *114*(1), 25–28.

Bello, A., Capirci, O., & Volterra, V. (2004). Lexical production in children with Williams syndrome: Spontaneous use of gesture in a naming task. *Neuropsychologia*, *42*(2), 201–213.

Bellugi, U., Bihrle, A., Jernigan, T. L., Trauner, D., & Doherty, S. (1990). Neuropsychological, neurological, and neuroanatomical profile of Williams syndrome. *American Journal of Medical Genetics*, *37*(S6), 115–125.

Bellugi, U., Wang, P. P., & Jernigan, T. L. (1994). Williams syndrome: An unusual neuropsychological profile. *Atypical Cognitive Deficits in Developmental Disorders: Implications for Brain Function*, *23*, 23–56.

Bennett, F. C., LaVeck, B., & Sells, C. J. (1978). The williams elfin facies syndrome: the psychological profile as an aid in syndrome identification. *Pediatrics*, *61*(2), 303–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/634688

Brock, J. (2007). Language abilities in Williams syndrome: A critical review. *Development and Psychopathology*, *19*(1), 97–127.

Charkova, K. D. (2003). Early foreign language education and metalinguistic development: A study of monolingual, bilingual and trilingual children on noun definition tasks. *Annual Review of Language Acquisition*, *3*(1), 51–88. http://doi.org/10.1075/arla.3.04cha

Clahsen, H., Ring, M., & Temple, C. M. (2004). Lexical and morphological skills in English-speaking children with Williams Syndrome. In S. Bartke & J. Siegmüller (Eds.), *Williams syndrome across languages* (Vol. 36, pp. 221–244). Amsterdam: John Benjamins Publishing Company.

Clahsen, H., & Temple, C. M. (2003). Words and rules in children with Williams syndrome. In *Language competence across populations: Toward a definition of specific language impairment* (pp. 323–352).

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test*. American Guidance Service Circle Pines, MN.

Farran, E. K., & Jarrold, C. (2003). Visuospatial Cognition in Williams Syndrome: Reviewing and Accounting for the Strengths and Weaknesses in Performance. *Developmental Neuropsychology*, *23*(1–2), 173–200. http://doi.org/10.1080/87565641.2003.9651891

Fodor, J. A. (1983). *The modularity of the mind: An essay on faculty psychology*. MIT press.

Franken, C. (2003). *Het verbale profiel bij Nederlands-sprekende personen met het Williams-syndroom*.

Gentner, D. (1982). Why Nouns Are Learned before Verbs: Linguistic Relativity Versus Natural Partitioning. *Language Development*, *2*, 67–88.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge, UK: Cambridge University Press.

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). II. The Emergentist Coalition Model. *Monographs of the Society for Research in Child Development*, *65*(3), 17–29.

Howlin, P., Davies, M., & Udwin, O. (1998). Cognitive functioning in adults with Williams syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *39*(2), 183–189. http://doi.org/10.1111/1469-7610.00312

Jackendoff, R. (2013). Foundations of Language. In *Brain, Meaning, Grammar, Evolution* (Vol. 53, pp. 1689–1699). http://doi.org/10.1017/CBO9781107415324.004

Jarrold, C., Hartley, S. J., Phillips, C., & Baddeley, A. D. (2000). Word fluency in Williams syndrome: Evidence for unusual semantic organisation? *Cognitive Neuropsychiatry*, *5*(4), 293–319.

Kaplan, E., Goodglass, H., Weintraub, S., Segal, O., & van Loon-Vervoorn, A. (2001). *Boston naming test*. Pro-ed.

Lloyd, M., Dunn, L., Dunn, M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale*. Windsor: NFER-Nelson.

Lukács, Á., Pléh, C., & Racsmány, M. (2004). Language in Hungarian children with Williams syndrome. In S. Bartke & J. Siegmüller (Eds.), *Williams syndrome across languages* (Vol. 36, pp. 187–220). Amsterdam: John Benjamins Publishing Company.

Maguire, M. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2006). A Unified Theory of Word Learning: Putting Verb Acquisition in Context. In *Action Meets Word* (pp. 364–

391). Oxford University Press.

Marien, P., Mampaey, E., Vervaet, A., Saerens, J., & De Deyn, P. P. (1998). Normative Data for the Boston Naming Test in Native Dutch-Speaking Belgian Elderly. *Brain and Language*, *65*(3), 447–467.

Mayer, M. (1965). *Frog, where are you?* New York: Dial Books.

Mervis, C. B. (1999). The Williams syndrome cognitive profile: Strengths, weaknesses, and interrelations among auditory short term memory, language, and visuospatial constructive cognition. In E. Winograd, R. Fivush, & W. Hirst (Eds.), *Ecological Approaches to Cognition: Essays in Honor of Ulric Neisser* (pp. 193–227). Mahwah, NJ: Erlbaum.

Mervis, C. B., Robinson, B. F., Bertrand, J., Morris, C. A., Klein-Tasman, B. P., & Armstrong, S. C. (2000). The Williams Syndrome Cognitive Profile. *Brain and Cognition*, *44*(3), 604–628. http://doi.org/10.1006/brcg.2000.1232

Mintz, T. H., & Gleitman, L. R. (2002). Adjectives really do modify nouns: The incremental and restricted nature of early adjective acquisition. *Cognition*, *84*(3), 267–293. http://doi.org/10.1016/S0010-0277(02)00047-1

Nippold, M. A. (1995). Norms for word definition..pdf. *Language, Speech, and Hearing Services in Schools*, *26*(4), 320–325.

NVWBS. (2016). Wat is het syndroom van Williams-Beuren? Retrieved January 5, 2017, from http://www.williams-syndroom.nl/informatie-over-het-syndroom-van-williams-beuren/williams-syndroom

Purser, H. R. M., Thomas, M. S. C., Snoxall, S., Mareschal, D., & Karmiloff-Smith, A. (2010). Definitions versus categorization: assessing the development of lexico-semantic knowledge in Williams syndrome. *International Journal of Language & Communication Disorders*, *46*(3), 361–373.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*, 328–350.

Rossen, M., Klima, E. S., Bellugi, U., Bihrle, A., & Jones, W. (1996). 17. Language and cognition: evidence from Williams syndrome. In J. H. Beitchman, N. J. Cohen, M. M. Konstantarea, & R. Tannock (Eds.), *Language, learning, and behavior disorders: Developmental, biological, and clinical perspectives* (p. 367). Cambridge University Press.

Saeed, J. I. (2009). *Semantics* (3rd ed.). UK: Wiley-Blackwell.

Stojanovik, V., & van Ewijk, L. (2008). Do children with Williams syndrome have unusual vocabularies? *Journal of*

Neurolinguistics, *21*(1), 18–34.

Storms, G., Saerens, J., & De Deyn, P. P. (2004). Normative data for the Boston Naming Test in native Dutch-speaking Belgian children and the relation with intelligence. *Brain and Language*, *91*(3), 274–281.

Strømme, P., Bjømstad, P. G., & Ramstad, K. (2002). Prevalence Estimation of Williams Syndrome. *Journal of Child Neurology*, *17*(4), 269–271. http://doi.org/10.1177/088307380201700406

Temple, C. M., Almazan, M., & Sherwood, S. (2002). Lexical skills in Williams syndrome: A cognitive neuropsychological analysis. *Journal of Neurolinguistics*, *15*(6), 463–495.

Thomas, M. S. C., Dockrell, J. E., Messer, D., Parmigiani, C., Ansari, D., & Karmiloff-Smith, A. (2006). Speeded naming, frequency and the development of the lexicon in Williams syndrome. *Language and Cognitive Processes*, *21*(6), 721–759.

Tribushinina, E., van den Bergh, H., Ravid, D., Aksu-Koç, A., Kilani-Schoch, M., Korecky-Kröll, K., … Dressler, W. U. (2014). Development of adjective frequencies across semantic classes: A growth curve analysis of child speech and child-directed speech. *Language, Interaction and Acquisition/Langage, Interaction et Acquisition*, *5*(2), 185–226.

Udwin, O., Yule, W., & Martin, N. (1987). Cognitive abilities and behavioural characteristics of children with idiopathic infantile hypercalcaemia. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *28*(2), 297–309. http://doi.org/10.1111/j.1469-7610.1987.tb00212.x

van Hagen, J. M., van der Geest, J. N., van der Giessen, R. S., Lagers-van Haselen, G. C., Eussen, H. J. F. M. M., Gille, J. J. P., … De Zeeuw, C. I. (2007). Contribution of CYLN2 and GTF2IRD1 to neurological and cognitive symptoms in Williams Syndrome. *Neurobiology of Disease*, *26*(1), 112–124.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, *22*(2), 217–234. Retrieved from https://www.cambridge.org/core/journals/applied-psycholinguistics/article/breadth-and-depth-of-vocabulary-in-relation-to-l1l2-acquisition-and-frequency-of-input/6A7CD827022283582903ED690EF4EE06

Vicari, S., Bates, E., Caselli, M. C., Pasqualetti, P., Gagliardi, C., Tonucci, F., & Volterra, V. (2004). Neuropsychological profile of

Italians with Williams syndrome: An example of a dissociation between language and cognition? *Journal of the International Neuropsychological Society*, *10*(6), 862–876.

Volterra, V., Capirci, O., Pezzini, G., Sabbadini, L., & Vicari, S. (1996). Linguistic Abilities in Italian Children With Williams Syndrome. *Cortex*, *32*(4), 663–677.

Wechsler, D. (1974). *WISC-R, Wechsler intelligence scale for children, revised*. Psychological Corporation.

Wechsler, D. (1990). *Wechsler Preschool and Primary Scale of Intelligence: Manual containing British amendments to the administration and scoring criteria sections (rev.)*. London: The Psychological Corporation.

Wehren, A., De Lisi, R., & Arnold, M. (1981). The development of noun definition. *Journal of Child Language*, *8*(1), 165–175.

Appendix A: PPVT answer form

Below is the first page of the PPVT answer form (Dunn & Dunn, 1997). It shows the first six sets, with the start age indication for each set listed at the top, and the start and stop rules listed underneath:

- Start rule: 0 or 1 error in a set;
- Stop rule: 8 or more errors in a set.

Start leeftijd 2.6 - 3 — SET 1

Item	Woord	Plaatje	Antw.	Fout
1	bus	(4)		F
2	drinken	(3)		F
3	hand	(1)		F
4	klimmen	(1)		F
5	sleutel	(4)		F
6	lezen	(1)		F
7	kast	(2)		F
8	springen	(3)		F
9	lamp	(4)		F
10	helikopter	(2)		F
11	ruiken	(2)		F
12	vlieg	(3)		F
		Aantal fouten		

Start leeftijd 4 — SET 2

Item	Woord	Plaatje	Antw.	Fout
13	graven	(2)		F
14	koe	(1)		F
15	trommel	(3)		F
16	veer	(1)		F
17	verven	(3)		F
18	kooi	(2)		F
19	knie	(1)		F
20	inpakken	(4)		F
21	hek	(3)		F
22	elleboog	(4)		F
23	vuilnis	(2)		F
24	trainen	(4)		F
		Aantal fouten		

Start-regel: 0 of 1 fout in een set — *Stop-regel: 8 of meer fouten in een set*

Start leeftijd 5 — SET 3

Item	Woord	Plaatje	Antw.	Fout
25	leeg	(1)		F
26	schouder	(3)		F
27	vierkant	(4)		F
28	meten	(4)		F
29	stekelvarken	(1)		F
30	pijl	(2)		F
31	schillen	(3)		F
32	fontein	(2)		F
33	botsing	(2)		F
34	pinguïn	(1)		F
35	versierd	(4)		F
36	nest	(3)		F
		Aantal fouten		

SET 4

Item	Woord	Plaatje	Antw.	Fout
37	kasteel	(2)		F
38	zagen	(4)		F
39	cactus	(3)		F
40	boerderij	(1)		F
41	gaan	(2)		F
42	harp	(1)		F
43	astronaut	(3)		F
44	wasbeer	(4)		F
45	jongleren	(4)		F
46	envelop	(2)		F
47	scheuren	(3)		F
48	klauw	(1)		F
		Aantal fouten		

Start-regel: 0 of 1 fout in een set — *Stop-regel: 8 of meer fouten in een set*

Start leeftijd 6 - 7 — SET 5

Item	Woord	Plaatje	Antw.	Fout
49	parachute	(3)		F
50	bezorgen	(1)		F
51	rechthoek	(1)		F
52	duiken	(2)		F
53	camper	(4)		F
54	roos	(2)		F
55	schrijven	(1)		F
56	harig	(4)		F
57	boren	(2)		F
58	haak	(3)		F
59	groep	(3)		F
60	druppelen	(4)		F
		Aantal fouten		

SET 6

Item	Woord	Plaatje	Antw.	Fout
61	voertuig	(4)		F
62	ovaal	(1)		F
63	bagage	(2)		F
64	prijsuitreiking	(3)		F
65	brandkraan	(4)		F
66	moeras	(3)		F
67	rekenmachine	(2)		F
68	verkeerslicht	(1)		F
69	selderij	(1)		F
70	wereldbol	(2)		F
71	groente	(3)		F
72	lijst	(1)		F
		Aantal fouten		

Start-regel: 0 of 1 fout in een set — *Stop-regel: 8 of meer fouten in een set*

Appendix B: SICI survey

The text below contains the introduction and explanation for the SICI survey as it was presented to respondents, as well as the first item with the details on the SICI criteria. As the survey was aimed at Dutch L1-speakers, the text is in Dutch.

In deze vragenlijst krijg je telkens een woord te zien waarbij gevraagd wordt deze te classificeren met behulp van de zogenaamde SICI-criteria. Het SICI-continuum wordt gebruikt om de volgorde van verwerving van woorden te voorspellen, en pas ik in mijn onderzoek toe om te voorspellen welke woorden meer of juist minder problemen op zullen leveren voor de groep waar ik onderzoek naar doe (mensen met het Williams-syndroom). De criteria worden hieronder uitgelegd; deze uitleg wordt bij elk woord kort herhaald. Meer uitleg staat in het artikel (pp. 17-22 bevatten beschrijvingen van de vier criteria) waar ik deze schaal op baseer. Dit artikel kun je hier vinden: http://tinyurl.com/SICIuitleg

Shape **In welke mate heeft datgene waar het woord naar verwijst een duidelijk identificeerbare vorm? In het geval van werkwoorden wordt hier verwezen naar de beweging van de persoon die de actie uitvoert: in welke mate hoort er een duidelijk herkenbare beweging bij een werkwoord?** ***(1 = heel moeilijk identificeerbare of geheel onidentificeerbare vorm, 5 = zeer gemakkelijk identificeerbare vorm)***
Individuation **In welke mate kun je datgene waar het woord naar verwijst, direct afleiden en onderscheiden in de wereld om je heen? Eigen namen hebben bijv. een zeer hoge score hier ("Jan" verwijst voor een kind duidelijk naar die persoon in de wereld), grammaticale elementen scoren juist erg laag (een woord als "en" heeft geen bijbehorend element in de wereld, en vereist dus talige kennis voordat het verworven kan worden). Ook kun je denken aan het verschil tussen 'suikerpot' (hoge score, want specifiek voorwerp in de wereld) en 'suiker' (lage score, want verwijs je naar een korrel, een theelepel, een heel pak suiker?).** ***(1 = heel moeilijk of helemaal niet af te leiden door observatie v/d wereld om je heen, 5 = zeer gemakkelijk af te leiden door observatie v/d wereld om je heen)***
Concreteness **In welke mate kun je datgene waar het woord naar verwijst, zien, horen of aanraken? Woorden als 'appeltaart' en 'huilen' zullen hoog scoren, terwijl 'emotie' en 'nadenken' juist een lage score krijgen.** ***(1 = heel moeilijk of helemaal niet zichtbaar, hoorbaar of tastbaar, 5 = zeer gemakkelijk zichtbaar, hoorbaar of tastbaar)***

Imageability
In welke mate kun je je een voorstelling maken (mental image) van datgene waar het woord naar verwijst? In welke mate roept het een duidelijk beeld bij je op? Hier kun je bijvoorbeeld denken aan 'liefde', wat op Concreteness laag zou scoren omdat het niet direct waarneembaar is, maar op Imageability dan weer hoger kan scoren omdat het bijvoorbeeld een beeld van twee kussende mensen op kan roepen.
(1 = heel moeilijk of helemaal geen voorstelling van te maken, 5 = zeer gemakkelijk of zeer duidelijke voorstelling van te maken)

1. bus

Shape
In welke mate heeft datgene waar het woord naar verwijst een duidelijk identificeerbare vorm, of hoort er een duidelijk herkenbare beweging bij (voor werkwoorden)?
(1 = heel moeilijk identificeerbare of geheel onidentificeerbare vorm, 5 = zeer gemakkelijk identificeerbare vorm)

Individuation
In welke mate kun je datgene waar het woord naar verwijst, direct afleiden en onderscheiden in de wereld om je heen?
(1 = heel moeilijk of helemaal niet af te leiden door observatie v/d wereld om je heen, 5 = zeer gemakkelijk af te leiden door observatie v/d wereld om je heen)

Concreteness
In welke mate kun je datgene waar het woord naar verwijst, zien, horen of aanraken?
(1 = heel moeilijk of helemaal niet zichtbaar, hoorbaar of tastbaar, 5 = zeer gemakkelijk zichtbaar, hoorbaar of tastbaar)

Imageability
In welke mate kun je je een voorstelling maken (mental image) van datgene waar het woord naar verwijst? In welke mate roept het een duidelijk beeld bij je op?
(1 = heel moeilijk of helemaal geen voorstelling van te maken, 5 = zeer gemakkelijk of zeer duidelijke voorstelling van te maken)

Appendix C: BNT answer form

Below is the first page of the BNT answer sheet (Kaplan et al., 2001).

BOSTON NAMING TEST

Naam: Patientnummer:

Geb. datum: Testdatum:

		Correct	Tijd	Stim. Cue	Phon. Cue
1	bed (een meubelstuk)	✓			
2	boom (iets, dat buiten groeit)	✓			
3	potlood (gebruik je om mee te schrijven)	✓			
4	huis (een soort gebouw)	✓			
5	fluitje (gebruik je om op te blazen)	✓			
6	schaar (gebruik je om te knippen)	✓			
7	kam (gebruik je om je haar goed te doen)	✓			
8	bloem (groeit in de tuin)	✓			
9	zaag (gebruikt een timmerman)	✓			
10	tandenborstel (doe je in je mond)	✓			
11	helicopter (gebruik je om door de lucht te reizen)	✓			
12	bezem/veger (gebruik je bij het schoonmaken)	✓			
13	octopus (een zeedier)	✓			
14	paddestoel (kan je eten)	✓			
15	hanger (kan je in de kast vinden)	✓			
16	rolstoel (kan je in het ziekenhuis vinden)	✓			
17	kameel (een dier)	✓	dromedaris (later gecorrigeerd)		
18	masker (onderdeel van een kostuum)	✓			
19	krakeling (iets om te eten)	✓			
20	bank (kan je op zitten)	✓			
21	tennisracket (gebruik je bij het sporten)	✓			

21 + 26 = 47

Afdeling Kinderneurologie AZR/Sophia Kinderziekenhuis

Appendix D: WPPSI-R Vocabulary answer form

Below is the answer sheet for the WPPSI-R Vocabulary subtask (Wechsler, 1990).

12. Woordenschat

item	antwoord	score
tekeningen		1 of 0
1. kat		
2. boom		
3. sleutel		
verbale items		2, 1 of 0
4. mes		
5. schoen		
6. fiets		
7. hoed		
8. paraplu		
9. brief		
10. plaatje		
11. gloeien		
12. kasteel		
13. happen		
14. schommel		
15. feestdag		
16. held		
17. publiek		
18. mot		
19. dubbel		
20. beleefd		
21. moed		
22. gokken		
23. microscoop		
24. oudheid		
25. lastpost		
	TOTAAL (max = 47)	