Overuse of familiar phrases by individuals with Williams syndrome masks differences in language processing

Sederias, I., Krakovitch, A., Stojanovik, V. ORCID: https://orcid.org/0000-0001-6791-9968 and Zimmerer, V. C. (2024) Overuse of familiar phrases by individuals with Williams syndrome masks differences in language processing. Journal of Child Language. ISSN 1469-7602 doi: https://doi.org/10.1017/S0305000924000436 Available at https://centaur.reading.ac.uk/117311/

To link to this article DOI: http://dx.doi.org/10.1017/S0305000924000436

Publisher: Cambridge University Press

Journal of Child Language (2024), 1–15
doi:10.1017/S0305000924000436

CAMBRIDGE
UNIVERSITY PRESS

BRIEF RESEARCH REPORT

Overuse of familiar phrases by individuals with Williams syndrome masks differences in language processing

Ioana Sederias[1,2], Ariane Krakovitch[1,3], Vesna Stojanovik[4] and Vitor C. Zimmerer[1]

[1]Department of Language and Cognition, University College London, London, UK; [2]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK; [3]Hôpital Necker - Enfants Malades, Assistance Publique - Hôpitaux de Paris, Paris, France and [4]School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

Corresponding author: Vitor C. Zimmerer; Email: v.zimmerer@ucl.ac.uk

(Received 01 August 2023; revised 11 June 2024; accepted 04 July 2024)

Abstract

We investigated whether individuals with Williams Syndrome (WS) produce language with a bias towards statistical properties of word combinations rather than grammatical rules, resulting in an overuse of holistically stored, familiar phrases. We analysed continuous speech samples from English children with WS (n = 12), typically developing (TD) controls matched on chronological age (n = 15) and TD controls matched on language age (n = 14). Alongside word count, utterance length, grammatical complexity, and morphosyntactic errors, we measured familiarity of expressions by computing collocation strength of each word combination. The WS group produced stronger collocations than both control groups. Moreover, the WS group produced fewer complex sentences, shorter utterances, and more frequent function words than chronological-age matched controls. Language in WS may appear more typical than it is because familiar, holistically processed expressions mask grammatical and other difficulties.

Keywords: Williams syndrome; narrative language; grammar; usage-based linguistics; neuroconstructivism; collocation strength

Introduction

Williams syndrome (WS) is a rare neurodevelopmental disorder present in about 1 in 7,500 to 20,000 live births and caused by a micro-deletion on one copy of chromosome 7, which results in atypical physical, cognitive and behavioural phenotypes (Kozel et al., 2021; Royston et al., 2019; Tassabehji et al., 1999). Individuals with WS have been described as presenting with relatively good vocabulary and phonological skills and relatively spared grammar in the face of weaker pragmatic skills and moderate-severe deficits in nonverbal tasks including problem solving, spatial and number cognition and planning (Mervis & John, 2010). Reports have also highlighted relatively good performance on complex language structures such as passives, negations, and conditionals (Bellugi et al., 1988, 1994), inflections and derivations (Clahsen & Almazan, 1998) as well as increased use of narrative enrichment devices (Bellugi et al., 1994). Such observations made WS a popular example supporting the idea of an independent "language module" containing abstract grammatical representations (Pinker, 1999; Zukowski, 2005).

This assumption of a relative strength in language has been challenged by studies arguing that language in individuals with WS is either delayed or less developed than expected for their mental age. For example, individuals with WS produced more errors in the areas of lexical selection, word order, gender agreement and verb inflections, and showed poorer grammatical comprehension compared to neurotypical controls (Karmiloff-Smith et al., 1997; Volterra et al., 1996). Individuals with WS also performed similarly or less well than populations with comparable intellectual skills such as Down syndrome on processing wh-questions and passives (Joffe & Varlokosta, 2007). Furthermore, children with WS showed no verbal advantage over children with developmental language disorder on standardized tests or a narrative task (Stojanovik et al., 2004). Thomas et al. (2001) reported that participants with WS had difficulties generalizing past tense rules to novel verbs, often omitting obligatory inflections. Expressive language has been described as stylistically different, featuring atypical vocabulary, stereotyped phrases, idioms, overfamiliar language and excessive use of social evaluative devices including prosodic cues and dramatic narrative elements (Reilly et al., 2004; Thomas et al., 2006; Udwin & Yule, 1990), although these findings have not always been replicated (Crawford et al., 2008; Stojanovik & van Ewijk, 2008).

The neuroconstructivist explanation has been that rather than accessing a preserved language module, individuals with WS acquire language in a qualitatively different manner (Karmiloff-Smith et al., 2012; Levy & Eilam, 2013; Thomas et al., 2001). Pointing and categorization emerge late relative to lexical acquisition (Laing et al., 2002), and the developmental trajectory has been characterized by a stronger correlation between grammatical capacity and verbal working memory (Robinson et al., 2003). One proposal is that language processing in WS relies less on grammatical and lexical-semantic information and more on shallow acoustic and phonological features, with a bias towards imitation (Thomas & Karmiloff-Smith, 2003). An individual with WS may produce phrases or utterances because they heard them before and represent them as one unit, and not because they use abstract grammatical information to combine individual words and morphemes. This more shallow production may give the appearance of unaffected processing.

Such explanations call for the distinction between analytic and holistic (or gestalt) processing, which is rooted within usage-based linguistics (e.g., Jackendoff, 2003; Langacker, 1987; Tremblay & Baayen, 2010). Analytic processing uses abstract representations of phrasal structures to combine individual words and morphemes and is able to generate novel expressions. Holistic processing, on the other hand, involves learning and retrieving word combinations, such as single phrases, but also entire sentences, as a single unit (a formula). Imitation can be driven by holistic representations; however, holistic forms are also an essential part of everyday language use (van Lancker Sidtis & Rallon, 2004). Theories such as construction grammar (Goldberg, 2006, 2019) suggest that everyone uses both analytic and holistic processing, with the contribution of each changing depending on situational requirements. While proficient speakers have the capacity to analyze these utterances, in principle holistic phrases can be used without grammatical and lexical-semantic interpretation of their individual constituents.

It has been proposed that in typical development, children first primarily employ holistic processing, resulting in conservative and repetitive production of language formulas, and only later acquire more abstract grammatical representations that, along with lexical growth, enable more creative and flexible language (Bannard & Matthews, 2008; Lieven et al., 2009). Faster and more accurate processing of formulaic language in adults suggests that holistic knowledge remains relevant even after maturation (Conklin & Schmitt, 2012; Tremblay & Baayen, 2010). Formulaic language is also more likely preserved in people with neurological conditions such as aphasia or dementia (van Lancker Sidtis, 2012; Zimmerer et al., 2018, 2020). Because formulas can be long and morphologically rich, they give the impression of islands of intact grammatical knowledge, when instead they are likely either well-trained combinations or lexicalized multi-word sequences.

Since one predictor of holistic representations is frequency of (co-)occurrence of words in everyday language use, acquisition of formulaic phrases is supported by sensitivity to statistical patterns in language. In WS, this sensitivity has been demonstrated in artificial language learning studies. Using the word segmentation learning paradigm, Cashon et al. (2016) demonstrated that 20-month-old infants with WS could distinguish words from part-words after brief familiarization with statistically structured syllable sequences. Stojanovik et al. (2018) found that participants with WS prefer statistical representations to more abstract grammatical rules. They compared processing biases in artificial language learning performances between participants with WS, mental-age matched typically developing (TD) children, and chronological-age (CA) matched TD individuals. In a brief familiarization phase, participants listened to spoken syllable sequences generated by a simple Markov-grammar. It was explained that sequences were magic spells, and they were presented along with a cartoon magician. In the test phase, participants distinguished between correct and incorrect "spells" based on what they learned from the familiarization set. Participants with WS and younger, mental-age matched TD children preferred sequences that resembled exemplars from familiarization. CA-matched TD participants, on the other hand, accepted sequences that were grammatical, regardless of familiarity, demonstrating their ability to acquire more abstract grammatical knowledge. These results suggest that TD individuals switch from familiarity- to rule-based processing in their development while, in individuals with WS, the bias towards familiarity may remain for much longer.

Based on current evidence, one could hypothesize that natural language processing in WS would also be atypically biased towards co-occurrence of specific words and acquisition of holistically processed, formulaic language. Do individuals with WS produce more familiar language? In this current study we examined statistical properties of words and word combinations in narrative samples of individuals with WS, who we compared with CA and language-age matched (LA) TD controls to identify both delays and atypical trajectories.

We determined the usage-frequency of each word in a narrative sample, using the spoken section of the British National Corpus (BNC XML Edition, 2007) as reference. Higher frequency indicates that a word is more common in typical language use, which has been associated with ease of production. We analyzed word combinations by determining their collocation strength, again using the BNC as reference. Collocation strength shows how often words appear together, relative to how often each occurs in general. Collocation strength is therefore not merely a function of frequency. For example, *I go* is a more frequent bigram than *I tell* according to the BNC; however, the collocation strength of the latter is seven times as high, because when the words *I* and *tell* occur, they more likely appear together relative to appearing in other contexts. We extracted these variables using the Frequency in Language Analysis Tool (FLAT), a script which had previously been employed to study statistical properties of language production in adults with stroke aphasia (Zimmerer et al., 2018) and neurodegeneration (Zimmerer et al., 2020).

This study is a secondary analysis of language transcripts collected by Stojanovik et al. (2007), originally for their investigation of intonation. They found that the ability to produce and understand intonation of participants with WS was poorer than that of CA-matched controls, but mostly in line with language-age (LA)-matched controls. A subsequent study by Stojanovik and van Ewijk (2008) investigated lexical production. The authors report that individuals with WS did not differ from controls with regards to lexical diversity and the number of low frequency words produced.

We investigated additional features in order to present a broader profile and context and to address our questions regarding holistic vs. analytic processing. Our analysis includes two dimensions of language production: (1) complexity, which includes mean length of utterance (MLU), proportion of complex sentences, verb phrase complexity, and morphosyntactic errors, and (2) familiarity, which includes a more comprehensive measure of lexical frequency than the one used by Stojanovik and van Ewijk (2008), and the collocation strength of word combinations. While we predicted that CA-matched controls would produce the most complex and least familiar language, while individuals with WS would produce the least complex and most familiar language, our statistical analyses tested broader hypotheses, namely that groups would differ from another.

Methods

Participants

Twelve children (9 female, 3 male; mean age = 9.05 yrs) were recruited through the Williams Syndrome Foundation (UK). Ethnicity was not a recruitment criterion, however, all participants taking part in the study were white. Diagnosis was confirmed by a positive fluorescent in situ hybridization (FISH) test. Children's language skills were tested using the Test of Reception of Grammar (TROG-2; Bishop, 2003). The TROG-2 is a sentence-picture matching test, and sentences were read to the participants by the experimenter. The test contains a variety of unfamiliar sentences of increasing grammatical complexity, including subject-verb-object, subject-verb-adjunct, spatial prepositional phrases, sentences with pronouns, passive constructions and center-embedded clauses. Sentences and distractor images are designed in a way that the participant needs to interpret the grammatical structures to perform well. Non-verbal reasoning was assessed using Raven's Coloured Progressive Matrices (RCPM; Raven, 1984).

Children with WS were matched to two TD control groups (Table 1): 14 LA-matched controls (12 female, 2 male; mean age = 5.78 yrs) who did not differ on the TROG-2, $t(24) = -.297$, $p = .769$, but were significantly younger, $t(24) = 4.739$, $p < .001$. 15 CA-matched controls (13 female, 2 male; mean age = 9.91 yrs) did not differ in age from the WS group, $t(25) = -1.210$, $p = .237$, but scored significantly higher on the TROG-2, $t(25) = -14.389$, $p < .001$. RCPM raw scores differed significantly across groups, $F(2,38) = 53.253$, $p < .001$, with post-hoc tests identifying a significant difference between CA-matched controls and participants with WS, $p < .001$. LA-matched controls

Table 1. Means, standard deviations and ranges for chronological age, TROG-2 (language reception) and RCPM (non-verbal reasoning) for Williams Syndrome Group (WS), LA-matched and CA-matched TD controls

	Age (years)		TROG-2 raw scores		RCPM raw scores	
Groups	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
WS	9.05 (2.26)	6 – 13.42	7.58 (1.91)	4 – 14	14.33 (3.84)	8 – 21
LA	5.78 (1.19)	4.25 – 9.16	7.93 (3.00)	4 – 14	17.93 (5.41)	9 – 28
CA	9.91 (1.42)	8.08 – 12.7	19 (.93)	17 – 20	31.93 (4.77)	19 – 36

performed better than participants with WS, and that difference was close to the significance threshold, $p = .063$.

Procedure

Children had been asked to generate a story using the wordless picture book '*Frog, where are you?*' (Mayer, 1965). Samples had been orthographically transcribed using the Systematic Analysis of Language Transcripts (SALT; Miller & Chapman, 1985) and utterances had been segmented based on the conventions presented by Crystal et al. (1976). We manually annotated transcripts for features selected from the Northwestern Narrative Language Analysis (Thompson, 2013; see Appendix A for an example from this study). The features were sentence type (simple or complex; the latter defined by clause embedding or non-canonical word order), number and types of clause embedding, verb argument structure (number of arguments), and morphosyntactic errors. Annotators were blind to the participants' group membership. We calculated interrater reliability by computing intra-class correlations coefficients for each variable that was second rated for a subsample of 10 transcripts. Interrater reliability was satisfactory (sentence complexity: ICC (1,2) = .997; verb argument structure: ICC (1,2) = .945; grammatical errors: ICC (1,2) = .887). To investigate familiarity, usage-frequency was extracted for words and bigrams (two-word combinations) from the spoken subsection of the British National Corpus (BNC, 2007) using the Frequency in Language Analysis Tool (FLAT; Zimmerer et al., 2018, 2020). All words in each sample were included, and all bigrams except for ungrammatical combinations and words separated by a sentence or utterance boundary. Based on these variables, we calculated the following measures:

Complexity

Mean length of utterance in words (MLU-w)

The ratio of the number of word tokens divided by the number of utterances. MLU can be measured in words or morphemes; both variables correlate with another very strongly in a number of languages including English, which has relatively few inflectional markers (Ezeizabarrena & Garcia Fernandez, 2018; Parker & Brorson, 2005).

Word count

Total number of words produced.

Sentence complexity

The number of complex sentences (i.e., containing non-canonical word order and/or clause embedding) divided by the total number of sentences. We excluded nominal sentences, i.e., sentences without a finite verb.

Verb argument structure

The number of verb arguments in each sample divided by the number of verb tokens.

Morphosyntactic errors

The number of grammatically incorrect utterances divided by the number of utterances (including abandoned utterances and nominal sentences).

Familiarity

Lexical frequency

FLAT determined the average frequency of content words (words with a strong semantic representation, e.g., “table”, “blue”, or “swim”) and function words (words with a primarily grammatical function, e.g., “the”, “she”, “what”) separately based on the BNC. Averages for each participant were calculated based on types, i.e., each unique word was only entered once.

Bigram collocation strength

We followed the procedure from previous studies (e.g., Zimmerer et al., 2020). We calculated collocation strength for each bigram in a sample. For example, for the sentence *The boy went to sleep*, we analysed the collocation strength of *the boy*, *boy went*, *went to* and *to sleep.* We excluded ungrammatical bigrams from this part of the analysis (but counted these as morphosyntactic errors) and bigrams which crossed sentence or utterance boundaries. We also excluded immediate repetitions (e.g., the second *and then* in *and then and then he got up).* For quantifying collocation strength, we used t-scores (Gries, 2010), which, compared with the better known measure Mutual Information, does not inflate collocation strength when the frequency of the combination is low. We computed t-score averages for each participant based on bigram types, and only included bigrams with a frequency of one or more as t-scores for bigrams with a frequency of zero cannot be computed.

Proportion of bigrams in BNC

We computed the proportion of bigrams produced by the individual which occur in the BNC, i.e., have a frequency of one or more, as another measure of familiarity. This variable works in conjunction with collocation strength in order to describe word combinations produced by the participant.

Results

Analysis plan

Because of the novelty of this research, bidirectional hypotheses were tested. We compared means between each group for each independent variable. The main effect of group was inferred using one-way ANOVAs, followed by pairwise comparisons between groups. Bonferroni correction for pairwise comparisons (three groups) sets an adjusted significance threshold of p = .017. We mention, however, all pairwise differences with p < .1 to highlight the respective variables' potential for future work on the topic.

Group comparisons

See Table 2 for a summary of group performance containing group averages, standard deviations, and main effects of group. LA-matched controls produced fewer words than CA-matched controls, and significance was close to the adjusted threshold (p = .022), while the difference between CA-controls and speakers with WS was close to the unadjusted threshold (p = .052). With regards to complexity, MLU-w showed that CA-matched controls produced longer utterances than both LA-matched controls (p = .002) and individuals with WS (p < .001). LA-matched controls produced longer utterances than individuals with WS, but this difference too was not significant according to the adjusted threshold (p = .036). CA-matched controls also produced more complex sentences than both LA-matched controls (p = .006) and individuals with WS (p = .001). The difference between WS and LA-matched individuals was not significant (p = .488).

Table 2. Means and Standard deviations, Results from the nine ANOVAs for all linguistic variables for the Williams Syndrome Group (WS), Language-Age matched (LA) Group and Chronological-Age Matched Group (CA)

	WS	LA	CA	
VARIABLES	mean (SD)	mean (SD)	mean (SD)	Main effect of group (ANOVA)
Word count	222.42 (100.98)	213.99 (73.87)	287.07 (75.94)	F(2,38) = 3.36., p =.045, η^2 =.396
MLU-w	6.155 (1.42)	7.110 (1.50)	8.739 (1.84)	F(2,38) = 8.960, p <.001, η^2 =.320
Sentence complexity	.113 (.08)	.147 (.09)	.282 (.17)	F(2,38) = 7.077, p =.002, η^2 =.271
Verb argument structure	1.478 (.10)	1.527 (.11)	1.459 (.09)	F(2,38)=.585, p =.218, η^2 =.077
Morphosyntactic errors	.178 (.15)	.112 (.19)	.100 (.09)	F(2,38) = 1.052, p =.359, η^2 =.052
Frequency of content words (per million words)	4951.591 (1662.14)	3385.383 (1250.49)	3868.229 (1069.98)	F(2,38) = 4.668, p =.015, η^2=.468
Frequency of function words (per million words)	91793.481 (14677.44)	89437.555 (7271.66)	79519.53 (5829.25)	F(2,38) = 6.389, p =.004, η^2 =.403
Collocation strength	21.806 (5.52)	17.083 (3.49)	16.187 (3.04)	F(2,38) = 7.152, p =.002, η^2 =.273
Proportion of bigrams in BNC	.895 (.038)	.883 (.039)	.865 (.025)	F(2,38) = 2.651, p =.084, η^2 =.118

Table 3. Pearson correlations between TROG-2 and RCPM scores and all linguistic variables from the language production sample analysis. For all comparisons: df = 39.

	TROG -2	RCPM
RCPM	r = .83, p < .001	–
Word count	r = .34, p = .028	r = .39, p = .012
MLU-w	r = .61, p < .001	r = .74, p < .001
Sentence complexity	r = .47, p = .002	r = .65, p < .001
Verb argument structure	r = –.20, p = .201	r = –.14, p = .376
Morphosyntactic errors	r = –.25, p = .112	r = –.16, p = .332
Frequency of content words	r = –.08, p = .63	r = –.25, p = .118
Frequency of function words	r = –.56, p < .001	r = –.48, p = .001
Collocation strength	r = –.39, p = .011	r = –.37, p = .017
Proportion of bigrams in BNC	r = –.31, p = .048	r = –.34, p = .028

Groups did not differ on complexity of verb argument structure or the proportion of morphosyntactic errors.

Lexical familiarity effects were not significant for content words, though the difference between individuals with WS and LA-matched controls was notable (p = .058), as LA-matched controls produced less frequent content words. The effect was greater and significant for function words: individuals with WS produced more frequent function words than CA-controls (p < .001). The difference between CA- and LA-matched controls was close to the adjusted significance threshold (p = .024).

Collocation strength was the only variable on which individuals with WS were significantly different from both control groups. Word combinations were more strongly collocated in individuals with WS than in CA-matched (p = .003) and LA-matched (p = .005) participants. Control groups did not differ significantly on collocation strength.

Groups also differed in the proportion of bigrams in the BNC, but only at $p < .1$, driven by CA-speakers producing fewer combinations that occur in the corpus than individuals with WS, with the effect being above the adjusted significance threshold (p = .03).

Relationship between language production and standardized testing measures

Post-hoc, we investigated how properties for "Frog Story" narrations related to TROG-2 and RCPM scores (Table 3). Overall, individuals with higher TROG-2 and RCPM scores produced longer samples, longer utterances, more complex sentences, less frequent function words, weaker collocations and fewer bigrams which occur in the BNC. However, because TROG-2 and RCPM scores were strongly correlated, one cannot confidently separate these individual predictors.

Discussion

Analysis of spontaneous language production in a narrative task revealed substantial and significant differences between individuals with WS, CA-matched controls, and younger

LA-matched controls. Our data characterize language in individuals with WS as containing mostly grammatically correct, but short and syntactically simple utterances with a tendency to overuse familiar (strongly collocated) word combinations. In context of previous studies and usage-based theories of language, we regard the results as evidence for language in WS being more dependent on holistic representations, which is likely the result of a bias towards statistical processing of word co-occurrence patterns rather than application of abstract grammatical knowledge. However, before considering the implication of this finding, we provide context for the other results. We interpret as evidence for delay a pattern in which individuals with WS differ from CA-, but not LA-matched controls, while we regard differences between the WS and LA-matched groups as evidence for divergent developmental trajectories.

Considering language complexity, the CA-controls produced longer utterances, and proportionally more sentences with non-canonical structures and embedded clauses, than both LA-controls and individuals with WS. These results are in line with previous findings that showed delays in the language of people with WS with regards to both MLU (Levy & Eilam, 2013) and sentence complexity (Reilly et al., 2004; Stojanovik et al., 2004).

We found no significant group differences for usage-frequency of content words. This finding supports views that lexical difficulties play a relatively small role in WS, and is corroborated by previous results which suggest that lexical diversity is also not affected in WS (Stojanovik & van Ewijk, 2008). However, we consider that lexical effects may be diminished by lexical constraints of the task, since all children described the same content (the “Frog story”). Investigations of spontaneous conversations can address this limitation.

We did find strong effects of WS on the frequency of function words. These are a crucial aspect of grammatical knowledge. Function word frequency is rarely investigated, but it appears that it can be used to characterize language production. Previously, Mok et al. (2022) found that younger TD children produced significantly more frequent function words than older children. We found the same age difference in our comparison between CA- and LA-controls, and while it did not meet Bonferroni-adjusted criteria for statistical significance, this finding is worth highlighting for further investigations. Importantly, the difference between individuals with WS and CA-matched controls was greater and significant and may provide another way of capturing grammatical deficits in WS. Data from studies on reading suggests that less frequent function words are more demanding (Ong & Kliegl, 2008). However, we do not understand exactly what makes less frequent function words (e.g., *because*) more difficult than more frequent words (e.g., *and*). Less frequent function words may occur in more complex sentence structures and propositional representations. We suggest future research could break our binary distinction between content and function words into further categories. Future projects may look further into variables related to lexical frequency, such as grammatical function, phonological complexity, and age of acquisition.

Collocation strength is the only measure on which individuals with WS differed significantly from both control groups, after correction for multiple comparisons and with large effect sizes. When individuals with WS combined words, they did so in ways which are more common, rather than in rare or novel ways. Group comparisons suggest that this may not be an effect of cognitive delay (younger and older controls did not differ from one another), but rather a substantial deviation from the typical trajectory. As reviewed in the introduction, data from artificial grammar learning suggest a stronger bias towards familiarity of stimuli in individuals with WS. We provide evidence that it is present in spontaneous language production and suggest that this familiarity bias shapes language organization at the cognitive level in individuals with WS. High collocation strength is one indicator that a combination is processed as a holistic, formulaic unit, with fewer demands on abstract, grammatical processes. We propose that while TD children switch from predominantly holistic to more analytic language, enabling greater combinatorial creativity, individuals with WS rely on familiar and more fixed constructions for longer (if not through life), at the cost of generative capacities. Learning of formulas can be supported not only by statistical processing, but by processing of prosodic contour, found to be a relative strength in WS.

This explanation supports neuroconstructivist views, which propose that children with WS acquire language in a different way (Grant et al., 2002; Joffe & Varlokosta, 2007; Levy & Eilam, 2013; Thomas et al., 2001). The bias towards holistic processing may be present in other domains. For example, individuals with WS may process faces holistically ("globally") rather than as a combination of individual features (Annaz et al., 2009; Tager-Flusberg et al., 2003). More studies on the relationship between holistic language and processing in other domains could contribute to accomplishing a cognitive profile in WS.

Our study did not find a difference between the WS group and controls in the proportion of morphosyntactic errors, which contradicts previous results (Joffe & Varlokosta, 2007; Karmiloff-Smith et al., 1997). These contrasting results might be explained by the choice of the language elicitation task. Studies that indicate more erroneous language production in WS used tasks which constrained production to specific linguistic structures which were hypothesized to be difficult (Faitaki & Murphy, 2020). Our spontaneous narrative speech elicitation task did not constrain participants in such a way. Participants could have favoured selection of constructions they could produce with greater accuracy.

The relative lack of morphosyntactic errors in a narrative production would be a demonstration of how a reliance on familiar word combinations can mask possible language differences. Here, one could see parallels between WS and dementia. In early work on grammar in dementia, a lack of grammatical errors led to the conclusion that grammar was unimpaired (Kempler et al., 1987). Later studies found decrease in grammatical complexity, and finally an overreliance on formulaic language, also detectable using collocation strength measures (Bates et al., 1995; Zimmerer et al., 2020). Familiar language naturally is unlikely to strike the listener as unusual, which explains why early studies, which did not focus on familiarity of language, did not reveal atypical patterns.

Our work also parallels suggestions that language development in autistic individuals may rely more on holistic processing (or "gestalt" processing), resulting in acquisition and use of phrases and utterances which formally suggest complexity but may be unanalysed (Noens & Berckelaer-Onnes, 2005). Such bias in processing may support production of connected language where analytical understanding of language is less developed, but may underlie phenomena like echolalia and inaccurate pronoun use.

One important limitation of the current study is sample size, which limits the power of our statistical models, particularly since substantial individual differences have been reported in WS (Brock, 2007). Research on larger samples would also enable more complex models to investigate interactions between variables. This issue is common in WS research because the syndrome is rare. One alternative to studies involving larger samples can be reproductions using other smaller samples available to individual labs. Public availability of samples from individuals with WS can also aid research, as WS is not well-represented in public language corpora. For example, the CHILDES database (MacWhinney, 2000) currently only features two transcripts of a Spanish-speaking child with WS. Unfortunately, sharing our samples publicly was not covered in the original ethical approval. Future studies may also choose to elaborate on measures of lexical frequency and grammatical function. Content and function words are very large categories which each contain words with very different semantic, grammatical, and discourse functions.

Research of language in WS has seen a shift away from theories which propose that WS offers evidence that a language "module" can function independently of other cognitive deficits. Our work suggests that our understanding of WS can be supported by frameworks which regard language processing as a combination of two types of representations: more abstract and analytic grammatical frames, which enable more creative and flexible language use, and holistic, fixed representations which are acquired by statistical learning and are cognitively less demanding. A bias towards these holistic representations may be related to general cognitive deficits in WS.

Competing interest. The authors declare none.

References

Annaz, D., **Karmiloff-Smith, A.**, **Johnson, M. H.**, & **Thomas, M. S. C.** (2009). A cross-syndrome study of the development of holistic face recognition in children with autism, Down syndrome, and Williams syndrome. *Journal of Experimental Child Psychology*, **102**(4), 456–486. https://doi.org/10.1016/j.jecp.2008.11.005

Bannard, C., & **Matthews, D.** (2008). Stored word sequences in language learning: the effect of familiarity on children's repetition of four-word combinations: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, **19**(3), 241–248. https://doi.org/10.1111/j.1467-9280.2008.02075.x

Bates, E., **Harris, C.**, **Marchman, V.**, **Wulfeck, B.**, & **Kritchevsky, M.** (1995). Production of complex syntax in normal ageing and Alzheimer's disease. *Language and Cognitive Processes*, **10**(5), 487–539. https://doi.org/10.1080/01690969508407113

Bellugi, U., **Marks, S.**, **Bihrle, A. M.**, & **Sabo, H.** (1988). Dissociation between language and cognitive functions in Williams Syndrome. *Language development in exceptional circumstances*, 177–189.

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.

Bishop, D. (2003). *Test for the Reception of Grammar.*

The British National Corpus, version 2 (BNC XML Edition). (2007). Retrieved from http://www.natcorp.ox.ac.uk.

Brock, J. (2007). Language abilities in Williams syndrome: a critical review. *Development and Psychopathology*, **19**(1), 97–127. https://doi.org/10.1017/S095457940707006X

Cashon, C. H., **Ha, O.-R.**, **Graf Estes, K.**, **Saffran, J. R.**, & **Mervis, C. B.** (2016). Infants with Williams syndrome detect statistical regularities in continuous speech. *Cognition*, **154**, 165–168. https://doi.org/10.1016/j.cognition.2016.05.009

Clahsen, H., & **Almazan, M.** (1998). Syntax and morphology in Williams syndrome. *Cognition*, **68**(3), 167–198. https://doi.org/10.1016/s0010-0277(98)00049-3

Conklin, K., & **Schmitt, N.** (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, **32**, 45–61. https://doi.org/10.1017/s0267190512000074

Crawford, N. A., **Edelson, L. R.**, **Skwerer, D. P.**, & **Tager-Flusberg, H.** (2008). Expressive language style among adolescents and adults with Williams syndrome. *Applied Psycholinguistics*, **29**(4), 585–602. https://doi.org/10.1017/s0142716408080259

Crystal, D., **Fletcher, P.**, & **Garman, M.** (1976). *The grammatical analysis of language disability: A procedure for assessment and remediation* (Vol. **1**).

Ezeizabarrena, M. J., & **Garcia Fernandez, I.** (2018). Length of utterance, in morphemes or in words? MLU3-w, a reliable measure of language development in early Basque. *Frontiers in Psychology*, **8**, 2265. https://doi.org/10.3389/fpsyg.2017.02265

Faitaki, F., & **Murphy, V. A.** (2020). *Oral language elicitation tasks in applied linguistics research. The Routledge handbook of research methods in applied linguistics.* 360–369.

Goldberg, A. E. (2006). *Constructions at Work. The nature of generalization in language.* https://doi.org/10.1093/acprof:oso/9780199268511.001.0001

Goldberg, A. E. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions.* https://doi.org/10.2307/j.ctvc772nn

Grant, J., **Valian, V.**, & **Karmiloff-Smith, A.** (2002). A study of relative clauses in Williams syndrome. *Journal of Child Language*, **29**(2), 403–416. https://doi.org/10.1017/s030500090200510x

Gries, S. T. (2010). Useful statistics for corpus linguistics. A mosaic of corpus linguistics. *Selected Approaches*, **66**, 269–291.

Jackendoff, R. (2003). *Foundations of language: Brain, meaning, grammar, evolution.*

Joffe, V., & **Varlokosta, S.** (2007). Patterns of syntactic development in children with Williams syndrome and Down's syndrome: Evidence from passives and wh- questions. *Clinical Linguistics & Phonetics*, **21**(9), 705–727.

Karmiloff-Smith, A., **D'Souza, D.**, **Dekker, T. M.**, **Van Herwegen, J.**, **Xu, F.**, **Rodic, M.**, & **Ansari, D.** (2012). Genetic and environmental vulnerabilities in children with neurodevelopmental disorders. *Proceedings of the National Academy of Sciences of the United States of America*, **109** Suppl 2(supplement_2), 17261–17265. https://doi.org/10.1073/pnas.1121087109

Karmiloff-Smith, A., **Grant, J.**, **Berthoud, I.**, **Davies, M.**, **Howlin, P.**, & **Udwin, O.** (1997). Language and Williams syndrome: How intact is "intact"? *Child Development*, **68**(2), 246. https://doi.org/10.2307/1131848

Kempler, D., **Curtiss, S.**, & **Jackson, C.** (1987). Syntactic preservation in Alzheimer's disease. *Journal of Speech, Language, and Hearing Research: JSLHR*, **30**(3), 343–350. https://doi.org/10.1044/jshr.3003.343

Kozel, B. A., **Barak, B.**, **Kim, C. A.**, **Mervis, C. B.**, **Osborne, L. R.**, **Porter, M.**, & **Pober, B. R.** (2021). Williams syndrome. *Nature Reviews Disease Primers*, **7**(1), 42. https://doi.org/10.1038/s41572-021-00276-z

Laing, E., **Butterworth, G.**, **Ansari, D.**, **Gsödl, M.**, **Longhi, E.**, **Panagiotaki, G.**, **Paterson, S.**, & **Karmiloff-Smith, A.** (2002). Atypical development of language and social communication in toddlers with Williams syndrome. *Developmental Science*, **5**(2), 233–246. https://doi.org/10.1111/1467-7687.00225

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. **1**).

Levy, Y., & **Eilam, A.** (2013). Pathways to language: a naturalistic study of children with Williams syndrome and children with Down syndrome. *Journal of Child Language*, **40**(1), 106–138. https://doi.org/10.1017/S0305000912000475

Lieven, E., **Salomo, D.**, & **Tomasello, M.** (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, **20**(3). https://doi.org/10.1515/cogl.2009.022

MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk (third edition): Volume I: Transcription format and programs, Volume II: The database. *Computational Linguistics (Association for Computational Linguistics)*, **26**(4), 657–657. https://doi.org/10.1162/coli.2000.26.4.657

Mayer, M. (1965). *Frog, where are you?*

Mervis, C. B., & **John, A. E.** (2010). Cognitive and behavioral characteristics of children with Williams syndrome: implications for intervention approaches. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, **154**(2), 229–248. https://doi.org/10.1002/ajmg.c.30263

Miller, J. F., & **Chapman, R. S.** (1985). *Systematic Analysis of Language Transcripts.*

Mok, X. T. J., **Goh, S. L.**, **Saddy, J. D.**, **Varley, R.**, & **Zimmerer, V.** (2022). Language production and implicit statistical learning in typical development and children with acquired language disorders: an exploratory study. *Speech, Language and Hearing*, **25**(3), 349–363. https://doi.org/10.31219/osf.io/pc63b

Noens, I. L. J., & **Berckelaer-Onnes, I. A.** (2005). Captured by details: sense-making, language and communication in autism. *Journal of Communication Disorders*, **38**(2), 123–141. https://doi.org/10.1016/j.jcomdis.2004.06.002

Ong, J. K. Y., & **Kliegl, R.** (2008). Conditional co-occurrence probability acts like frequency in predicting fixation durations. *Journal of Eye Movement Research*, **2**(1). https://doi.org/10.16910/jemr.2.1.3

Parker, M. D., & **Brorson, K.** (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language*, **25**(3), 365–376. https://doi.org/10.1177/0142723705059114

Pinker, S. (1999). *Words and rules: the ingredients of language.* Weidenfeld & Nicholson.

Raven, J. C. (1984). *The coloured progressive matrices.*

Reilly, J., **Losh, M.**, **Bellugi, U.**, & **Wulfeck, B.** (2004). "Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and Language*, **88** (2), 229–247.

Robinson, B., **Mervis, C.**, & **Robinson, B.** (2003). The roles of verbal short-term memory and working memory in the acquisition of grammar by children with Williams syndrome. *Developmental Neuropsychology*, **23**(1), 13–31. https://doi.org/10.1207/s15326942dn231&2_2

Royston, R., **Waite, J.**, & **Howlin, P.** (2019). Williams syndrome: recent advances in our understanding of cognitive, social and psychological functioning. *Current Opinion in Psychiatry*, **32**(2), 60–66. https://doi.org/10.1097/YCO.0000000000000477

Stojanovik, V., **Perkins, M.**, & **Howard, S.** (2004). Williams syndrome and specific language impairment do not support claims for developmental double dissociations and innate modularity. *Journal of Neurolinguistics*, **17**(6), 403–424. https://doi.org/10.1016/j.jneuroling.2004.01.002

Stojanovik, V., **Setter, J.**, & **van Ewijk, L.** (2007). Intonation abilities of children with Williams syndrome: a preliminary investigation. *Journal of Speech, Language, and Hearing Research: JSLHR*, **50**(6), 1606–1617. https://doi.org/10.1044/1092-4388(2007/108)

Stojanovik, V., & **van Ewijk, L.** (2008). Do children with Williams syndrome have unusual vocabularies? *Journal of Neurolinguistics*, **21** (1), 18–34.

Stojanovik, V., **Zimmerer, V.**, **Setter, J.**, **Hudson, K.**, **Poyraz-Bilgin, I.**, & **Saddy, D.** (2018). Artificial grammar learning in Williams syndrome and in typical development: The role of rules, familiarity, and prosodic cues. *Applied Psycholinguistics*, **39**(2), 327–353. https://doi.org/10.1017/s0142716417000212

Tager-Flusberg, H., **Plesa-Skwerer, D.**, **Faja, S.**, & **Joseph, R. M.** (2003). People with Williams syndrome process faces holistically. *Cognition*, **89**(1), 11–24. https://doi.org/10.1016/s0010-0277(03)00049-0

Tassabehji, M., **Metcalfe, K.**, **Karmiloff-Smith, A.**, **Carette, M. J.**, **Grant, J.**, **Dennis, N.**, **Reardon, W.**, **Splitt, M.**, **Read, A. P.**, & **Donnai, D.** (1999). Williams syndrome: use of chromosomal microdeletions as a tool to dissect cognitive and physical phenotypes. *The American Journal of Human Genetics*, **64**(1), 118–125. https://doi.org/10.1086/302214

Thomas, M. S., **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. https://doi.org/10.1080/01690960500258528

Thomas, M. S., **Grant, J.**, **Barham, Z.**, **Gsödl, M.**, **Laing, E.**, **Lakusta, L.**, **Tyler, L. K.**, **Grice, S.**, **Paterson, S.**, & **Karmiloff-Smith, A.** (2001). Past tense formation in Williams syndrome. *Language and Cognitive Processes*, **16**(2–3), 143–176. https://doi.org/10.1080/01690960042000021

Thomas, M. S., & **Karmiloff-Smith, A.** (2003). Modeling language acquisition in atypical phenotypes. *Psychological Review*, **110**(4), 647. https://doi.org/10.1037/0033-295X.110.4.647

Thompson, C. K. (2013). Northwestern *narrative language analysis (NNLA) theory and methodology.*

Tremblay, A., & **Baayen, R. H.** (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In *Perspectives on formulaic language: Acquisition and communication*, 151–173.

Udwin, O., & **Yule, W.** (1990). Expressive language of children with Williams syndrome. *American Journal of Medical Genetics. Supplement*, **6**, 108–114. https://doi.org/10.1002/ajmg.1320370620

van Lancker Sidtis, D. (2012). Formulaic language and language disorders. *Annual Review of Applied Linguistics*, **32**, 62–80. https://doi.org/10.1017/s0267190512000104

van Lancker Sidtis, D., & **Rallon, G.** (2004). Tracking the incidence of formulaic expressions in everyday speech: methods for classification and verification. *Language & Communication*, **24**(3), 207–240. https://doi.org/10.1016/j.langcom.2004.02.003

Volterra, V., **Capirci, O.**, **Pezzini, G.**, **Sabbadini, L.**, & **Vicari, S.** (1996). Linguistic abilities in Italian children with Williams syndrome. *Cortex*, **32**(4), 663–677. https://doi.org/10.1016/s0010-9452(96)80037-2

Zimmerer, V. C., **Hardy, C. J. D.**, **Eastman, J.**, **Dutta, S.**, **Varnet, L.**, **Bond, R. L.**, **Russell, L.**, **Rohrer, J. D.**, **Warren, J. D.**, & **Varley, R. A.** (2020). Automated profiling of spontaneous speech in primary progressive aphasia and behavioral-variant frontotemporal dementia: An approach based on usage-frequency. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, **133**, 103–119. https://doi.org/10.1016/j.cortex.2020.08.027

Zimmerer, V. C., **Newman, L.**, **Thomson, R.**, **Coleman, M.**, & **Varley, R. A.** (2018). Automated analysis of language production in aphasia and right-hemisphere damage: frequency and collocation strength. *Aphasiology*, **32**(11), 1267–1283. https://doi.org/10.1080/02687038.2018.1497138

Zukowski, A. (2005). Knowledge of constraints on compounding in children and adolescents with Williams syndrome. *Journal of Speech, Language, and Hearing Research*, **48**(1), 79–92. https://doi.org/10.1044/1092-4388(2005/007)

Appendix A. Linguistic levels, features, codes and example of sample coded transcript

1. The boy was watch/ing out for the owl.

I: [s]

II: [ss][as][e0]

V:[ob2xy]

2. And he call/ed ‘frog, where are you’.

I: [s]

II: [ss][con][e0][wqj]

V: [cxy][copyp]

3. And a deer hold/ed hold[EWheld] him on his head.

I: [*s][g]

II: [ss][as][e0]

V: [ob2xy]

4. And he run/ed run[EW:ran].

I: [*s][g]

II: [ss][as][e0]

V: [ob1x]

5. And he push/ed him off the cliff.

I: [s]

II: [ss][as][e0]

V: [phob2xy]

6. ’splash’!

I: [ns]

II: -

V: -

7. And when the boy woke up he saw that the jar was empty.

I:[s]

II:[cs][as][e2][ac][cc]

V:[op2x][cxs’][copyp]

Levels	Linguistic Feature	Codes	Comments
First Line (I): Utterance Level	morphosyntactic errors	[s] no error [*s][m] semantic error [*s][g] grammatical error [*s][au] abandoned utterance [ns] non sentence	(example 1, 2 and 5) (example 3 and 4) a non sentence was not coded at levels II and V (example 5)
Second line (II): Sentence level	sentence complexity	[ss] simple sentnce [cs] complex sentence	We considered complex sentences those sentences that were either produced in non–canonical form such as passives, object wh–questions and object cleft sentences or contained at least one embedded clause (example 7).
	number of embeddings	[e0] no embeddings [e1] one embeddment [e2] two embeddings [en] n embeddings	
	types of embeddings	[cc] complement clause [ac] adjunct clause [rc] relative clause [ir] infinitval relative clause [sc] subject clause	(example 7) (example 7)
	sentence type	[ac] active, canonical [pa] passive [con] conjoined [wqs/o/j] subkect/object/ adjunct wh [yn] yes–no question	this was coded but not used in the analysis
Third Line (V): Verb Level*	verb type	[ob1] – intransitive [ob2]– obligatory two place [op2]–optional –two place [op3] optional three place [ob3] obligatory three place [cop] copula [c] complement	this was coded but not used in the analysis
	verb argument structure	[x] agent [y[theme\patient [z] goal [s] sentential complement [p] sentential phrase [j] adjunct	types of arguments a verb can take

Auxiliary verbs were not coded.

Cite this article: Sederias, I., Krakovitch, A., Stojanovik, V., & Zimmerer, V.C. (2024). Overuse of familiar phrases by individuals with Williams syndrome masks differences in language processing. *Journal of Child Language* 1–15, https://doi.org/10.1017/S0305000924000436