Expressive language profiles in a clinically screening sample of Mandarin-speaking preschool children with Autism Spectrum Disorder

Author Information

Li, Li[ad], Yi (Esther) Su[a*], Wenwen Hou[a], Muyu Zhou[a], Yixiang Xie[b], Xiaobing Zou[c], and Ming Li[d]

[a] Child Language Lab, School of Foreign Languages, Central South University, Changsha, China

[b] School of Psychology, South China Normal University, Guangzhou, China

[c] Child Development and Behavior Center, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

[d] Data Science Research Center, Duke Kunshan University, Suzhou, China

*Correspondence to Yi (Esther) SU: sy-esther@csu.edu.cn

Ethics statement: Studies involving human participants were reviewed and approved by the Institutional Review Board of the Third Affiliated Hospital of Sun Yat-sen University (IRB Number [2019]02-247-01). Written informed consent was obtained from the participants' legal guardians or from their next of kin.

Conflict of interest: The authors declare that this research was conducted without any commercial or financial relationships that could be construed as potential conflicts of interest.

Author contributions: LL, XZ, and ML designed the original data collection. YX and XZ coordinated the experiments and performed the clinical measurements. LL and ML created the models for data annotation. LL and YS analyzed the results and wrote the manuscript. WH and MZ edited and provided feedback regarding the manuscript. All authors contributed to the manuscript and approved the submitted version.

Funding Statement: This study was supported by the National Social Science Fund of China (Grant Number 20BYY086), and Science and Technology Program of Guangzhou City (Grant Number 202007030011).

Abstract

Purpose: This cross-sectional study aimed to depict expressive language profiles and clarify lexical-grammatical interrelationships in Mandarin-speaking preschoolers with Autism Spectrum Disorder (ASD) during the administration of the simplified Chinese Psychoeducational Profile-Third Edition (sCPEP-3) screening.

Methods: We collected naturalistic language samples from eighty-one (74 males) 2-7-year-old (*Mean age* = 55.6 months, *SD* = 15.17) Mandarin-speaking children with ASD in clinician-child interactions. The child participants were divided into five age subgroups with 12-month intervals according to their chronological age. Computer-assisted part-of-speech tagging, constituency analysis, and dependency analysis addressed the developmental trajectories of early lexical and grammatical growth in each age subgroup.

Results: Significant within-ASD differences were observed in content words, function words, and lexical categories. Nouns and verbs were the predominant lexical categories, while noun types overwhelmed verb types in children over three years old. The grammatical development of 5-to-6-year-old Mandarin-speaking children with ASD was better than 3-to-4-year-old children. The trends of syntactic structures, grammatical relations, and grammatical complexity in each age group were similar.

Conclusions: Mandarin-speaking preschoolers with ASD produce more lexicons with increasing age. They preserve the noun bias as a universal mechanism in early lexical learning. Moreover, their developmental trajectories of grammatical growth were comparable in each age subgroup. In

addition, their lexicons and grammar were synchronically developed during early language acquisition.

Key words: lexical categories, noun bias, grammatical profiles, sCPEP-3, Mandarin, children with ASD

Introduction

Language impairment has been excluded from the diagnostic criteria for autism spectrum disorder (ASD) in the DSM-5 (American Psychiatric Association, 2013), while delays in language development are part of parents' main concerns in seeking screening and intervention for young children with ASD (Robins et al., 2014). Early expressive language skills can predict long-term language outcomes in preschoolers with ASD (Mawhood et al., 2000). This study attempted to capture the expressive language profiles of 2-7-year-old Mandarin-speaking children with ASD in clinician-child interactions to support clinical decisions and individualized intervention plans.

Owing to social communication deficits in children with ASD, naturalistic language sampling is a critical measure for assessing expressive language variability (Bacon et al., 2019). However, studies that examined the heterogeneity of language abilities in children with ASD were mainly based on parent-child interactions (e.g., Song et al., 2022; Song & So, 2022). In this study, we focused on the lexical and grammatical characteristics of Mandarin-speaking preschoolers with ASD while administering the simplified Chinese Psychoeducational Profile-Third Edition (sCPEP-3) (Yu et al., 2019). Therefore, ours is the first study to investigate the expressive language profiles of 2-7-year-old Mandarin-speaking children with ASD in clinician-child interactions. In particular, we compared the lexical and grammatical subdomains of language based on computer-assisted Part-of-Speech (POS) tagging, Chinese constituency and dependency rules, and explored the lexical-grammatical interrelationships of this population.

Lexical development in expressive language of preschoolers with ASD

To date, research has adopted both standardized tests (e.g., Charman et al., 2003; Smith et al., 2007; Su et al., 2018) and naturalistic language samples (e.g., Song et al., 2022; Wittke et al., 2017) to reveal the heterogeneity of lexical production in children with ASD, contributing to our knowledge of lexical development in this group (Hart & Curtin, 2021). Despite unbalanced lexical development in children with ASD, few studies have addressed the developmental trajectories of lexical growth in Mandarin-speaking preschoolers with ASD.

In natural languages, lexical categories can be broadly classified into content words (CWs) and function words (FWs) (e.g., Hochmann et al., 2010; Radford et al., 2008). CWs, such as nouns and verbs, primarily carry lexical-semantic information and have many word types (i.e., the number of unique words that occur). By contrast, FWs, such as prepositions and conjunctions, primarily carry syntactic information and have fewer word types but a higher frequency of word tokens (i.e., the total number of words that occur) (Villena, 2002; Zaccarella et al., 2017). The content-function distinction is universal, although the particular lexical subcategories included in CWs and FWs could be language-specific (i.e., Mandarin has classifiers). Charman et al. (2003) reported that the percentages of common nouns, predicates, and closed-class terms in children with ASD were comparable to those in vocabulary-matched normative samples. Su et al. (2018) demonstrated that Mandarin-exposed preschoolers with ASD have lexical production similar to their total-vocabulary-matched typical development (TD) peers in the lexical categories of nouns, verbs, pronouns, classifiers, and question words. However, a critical gap in the literature is the lack of a thorough analysis of CWs and FWs of children with ASD in naturalistic language samples, which is one of the primary focuses of this study.

One major issue in early lexical composition was the relative importance of nouns and verbs. The debate has centered on whether a universally anchored noun bias helps children bootstrap into lexical learning (Gentner, 1982). It has been hypothesized that nouns are acquired faster than verbs because the concepts of nouns are perceptually more stable than those of verbs (Stolt et al., 2009). Cross-linguistic evidence for noun bias has been observed in Spanish, Dutch, French, Hebrew, Italian, Korean, and American English (e.g., Bassano, 2000; Bornstein et al., 2004). However, 22-month-old TD children exposed to verb-friendly languages such as Mandarin, Korean, and Japanese produced more verbs than nouns (Tardif, 1996; Tardif et al., 1997). Do young children with ASD exhibit noun bias during early lexical learning? Indeed, studies by Tager-Flusberg et al. (1990) and Charman et al. (2003) have suggested that such noun bias might operate in young English-exposed children with ASD, as they produce more nouns than other lexical categories. Moreover, Swensen et al. (2007) conducted research close to the beginning of the therapy. They found that 2-to-3-year-old English-exposed children with ASD showed similar noun bias as TD children in language comprehension. Given that these studies were conducted in the noun-friendly language of English, it would be interesting to investigate the noun/verb dominance of Mandarin-speaking preschoolers with ASD to explore whether noun bias is language-specific or universally inherent to young children across languages. To address this question, we explored the noun/verb dominance in Mandarin-speaking preschoolers with ASD.

Grammatical development in expressive language of preschoolers with ASD

Notably, recent work has mainly focused on the variability of structural language in children with ASD. Of particular interest in analyzing language samples collected from the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 1999), research has contributed to exploring the heterogeneous grammatical development measured by the Index of Productive

Syntax (IPSyn) (Scarborough, 1990) and Mean Length of Utterance (MLU) (Brown, 1973) in children with ASD. For example, Park et al. (2012) concluded that the overall grammatical development in 3-6-year-old children with ASD seemed to be delayed rather than unique. While their noun phrases, sentence structures, and MLU were below the expected level, the verb phrases and morphological structures developed atypically. In addition, Wittke et al. (2017) categorized children with ASD into three subgroups: normal language, grammatical impairment, and language impairment, according to the production of the MLU and the usage of Brown's 14 grammatical morphemes. Moreover, using the hierarchical phrase-based grammatical analysis, Leung and Li (2015) demonstrated that the flexibility and the number of different syntactic structures in Cantonese-speaking children with ASD were distinguishable from age-matched TD children. This work marked a call to characterize grammatical profiles of Mandarin-speaking preschoolers with ASD, particularly a systematic investigation of syntactic structures in this group.

Research has revealed a systematic pattern in children's morphosyntactic development across languages. For example, English-speaking children produce nonadult-like verbal forms when they are in two-word stages, and such verbal forms appear to be random when they reach three years of age in many languages, including Dutch, French, German, Icelandic and others (Legendre, 2006). However, Mandarin differs from other languages in that it lacks inflectional morphological markers for information orders (Li & Thompson, 1981). Unlike German-, Spanish-, and Arabic-speaking children, Mandarin-speaking children prefer "old-before-new" information structures (e.g., Narasimhan & Dimroth, 2008; Ceja Tel Toro et al., 2016; Semsem & Chen, 2019, See Chen et al., 2020). Despite this, Mandarin has a relatively straightforward structure that follows the canonical subject-verb-object (SVO) word order as in English (e.g., Xu et al., 2022),

while allowing the pervasive ellipsis of noun phrases in a sentence (Lee & Naigles, 2005). Thus, Mandarin has a distinct syntactic structure.

Until recently, MLU and the mean length of the five longest utterances (MLU5) (e.g., Wu, 2020) have been regarded as good indicators of grammatical complexity in Mandarin-speaking children. At the same time, only a few equivalent language assessment tools are available to measure the grammatical development of Mandarin. Nevertheless, from a theoretical perspective, Mandarin encodes grammatical information in a highly configurational rather than morphological way. In the current study, we sought to address the grammatical development of Mandarin-speaking preschoolers with ASD by focusing on constituency analysis and dependency grammar. Constituency analysis involves grouping smaller syntactic units (tokens) into more extensive constituents (phrases), and these constituents further group to form even more extensive syntactic constituents (larger phrases or clauses) (Poiret et al., 2021). Contrastingly, dependency grammar determines syntactic structures based on word-to-word connections and represents various grammatical relations (GRs). GRs are exemplified in traditional Chinese grammar by notions, such as subjects and direct/indirect objects, which can straightforwardly represent abstract syntactic information (Sun et al., 2019). Hence, this study provides new grounds for exploring grammatical development in Mandarin-speaking preschoolers with ASD.

The early lexical-grammatical relationship in children with ASD

The focus on lexical-grammatical relationships has been motivated by conflicting developmental linkages in early language acquisition (McGregor et al., 2012). One is an indispensable relationship between lexical and grammatical development in young TD children. For example, they must acquire a critical mass of lexicons that precedes their grammatical formation (Bates & Goodman, 1997; Marchman & Bates, 1994). However, this view contradicts

the modular dissociation view, which proposes that grammar is an autonomous and core representational domain in the linguistic system whose development is dissociated from the lexicon domain (Crain & Lillo-Martin, 1999). Thus, it is crucial to further distinguish between the disassociation and unification views of lexical and grammatical development in children's languages.

Similarly, studies on children with ASD have proposed two contrasting views: asynchronous and synchronous lexical-grammatical interrelationships in early language acquisition. The asynchronous view proposes that lexicon and grammar have independently developed principles. Of those supporting studies, Eigsti et al. (2007) demonstrated that 5-year-old children with ASD displayed syntactic impairments while they could produce similar word types and tokens as the vocabulary-matched TD children. Weismer et al. (2010) found that 30-month-old toddlers with ASD displayed a weaker lexical-grammatical association than vocabulary-matched 25-month-old late speakers. Moreover, Mandarin-acquiring preschoolers with ASD seem to have similar lexical and grammatical production as vocabulary-matched TD children, while their specific lexical-grammatical associations were less apparent with Putonghua Communicative Development Inventory (PCDIs; Su et al., 2018). In contrast, the synchronous view proposes that lexical and grammatical domains develop simultaneously during early language development (Dixon & Marchman, 2007). For example, Fusaroli et al. (2019) employed a longitudinal corpus of 2-5-year-old children with ASD over six visits and reported the coexistent growth of word types, tokens, and MLU with age. Given the inconclusive findings, it is interesting to elaborate on the developmental relationships of lexical and grammatical domains in Mandarin-speaking preschoolers with ASD to further understand "the inner architecture of the language faculty (e.g., the lexicon and grammar domains)" (Su & Naigles, 2022, p62).

Research gaps and current study

Research investigating the early language profiles of Mandarin-speaking preschoolers with ASD has primarily targeted parent-child interactions. Few research has examined the expressive language characteristics in clinician-child interactions. In contrast to caregivers, clinicians are less familiar with children. They are more likely to have systematic training on potential goals and motivations during interactions, which is beneficial for comprehensive evaluation (DeThorne & Channell, 2007). Concerns about language problems in children with ASD prompted the study to identify lexical and grammatical characteristics in a clinical screening context, to provide a framework for early screening and tailored interventions for this group.

This study aimed to answer the following questions: (1) What are the developmental trajectories of lexical growth, including CWs, FWs, and lexical subcategories, in Mandarin-speaking preschoolers with ASD? (2) Do Mandarin-speaking preschoolers with ASD show noun bias during early lexical learning? (3) What are the grammatical profiles, including syntactic structures, GRs, and grammatical complexity, of Mandarin-speaking preschoolers with ASD? (4) What is the lexical-grammatical relationship in the early language development of Mandarin-speaking preschoolers with ASD?

Methods

2.1 Participants

Participants in the current study were taken from the Big Data and Artificial Intelligence (AI) Assisted Early ASD Screening Project in Mainland China. This AI-assisted project recruited young Mandarin-speaking children with ASD at a local hospital's Child Development and Behavior Center in Southern China. Appropriate Institutional Review Board approval was obtained from the hospital, and written consent was obtained from the caregivers of the participants

before participation. The data collection for this project lasted for six months (from March 2020 to August 2020), and a subset of 119 language samples was available for transcription. Participants had to meet the following eligibility criteria: first, they were 2-7 years old and had Mandarin as their native language; second, they had to attend the sCPEP-3 assessments; and third, they had to be diagnosed with ASD based on the DSM-V without any prior intervention. Thirty-eight children were excluded because they did not complete the audio recordings or were not diagnosed with ASD. Participants with other known psychiatric disorders or severe neurological (intractable epilepsy) conditions were excluded from recruitment. Finally, eighty-one children (74 males) with ASD (mean age = 55.6 months, *SD* = 15.17) participated in this cross-sectional study. The participants were divided into five subgroups at 12-month intervals according to their chronological age (CA). The mean (M) and standard deviation (SD) of CA, developmental age of the standard scores of communication (DAC) in the sCPEP-3, and other demographic information by age group are presented in Table 1.

Table 1. Demographic information of participants

CA band		Sex ratio	Chronological Age		Developmental age of Communication	
(years; months)	*N (%)*	(M:F)	*M* (months)	*SD*	*M* (months)	*SD*
2;0-2;11	5 (6.17%)	3:2	30.0	3.16	23.6	12.55
3;0-3;11	24 (29.63%)	23:1	41.3	3.70	31.6	12.83
4;0-4;11	18 (22.22%)	8:1	52.8	2.78	39.6	4.08
5;0-5;11	18 (22.22%)	17:1	65.4	3.78	49.7	13.32
6;0-6;12	16 (19.75%)	15:1	77.1	4.17	46.2	16.51
Total	81(100%)	74:7	55.6	15.20	39.8	15.28

Measures

Simplified Chinese Psychoeducational Profile – Third Edition (sCPEP-3)

The sCPEP-3 is a standardized instrument with toys, games, objects, and pictures to measure children's behaviors and skills (Yu et al., 2019). Schopler et al. (2005) first developed the

third edition of the Psychoeducational Profile (PEP-3) from an earlier version of the PEP (Schopler & Reichler, 1979) for early evaluation and tailored interventions for children with ASD. The Heep Hong Society translated the PEP-3 into a simplified Chinese version (sCPEP-3) for Mainland China. The sCPEP-3 includes two main sections: a Performance Test and the caregiver report. In this study, we only analyzed the Performance part, composed of 10 subsets combined into three composites to assess the child's communication development, motor abilities, and presence of maladaptive behaviors (Schopler et al., 2005). The psychometric evaluation of the sCPEP-3 showed good internal consistency (ranging from .94 to .98), test-retest reliability (ranging from .73 to .98), and inter-rater reliability (ranging from .64 to .93). Notably, the sCPEP-3 is a promising comprehensive tool for characterizing the overall development of children with ASD, and it has been widely used in Chinese hospitals and training centers (Yu et al., 2019).

Clinician-child interaction

The Performance Test in the sCPEP-3 is a norm-referenced assessment based on the clinician's direct observation of a child's competence and performance. It is a semi-structured interview that requires limited verbal skills for children, and has 172 items with corresponding materials available for evaluation (Fu et al., 2010). This allows clinicians to adjust open questions or ongoing activities according to the children's performance. These characteristics enable children with ASD to complete screening-based evaluations in relevant scenarios during clinician-child interactions (Liu et al., 2022). Each time, the child was presented with a standardized set of assessment materials and was required to interact with the clinician. All participating children and clinicians were presented with similar materials.

Procedures

The Performance Test in sCPEP-3 was conducted as part of a clinical or research evaluation. All Performance Tests were administered and scored by qualified clinicians according to the examiner's manual guidelines. A parent was allowed to sit behind the child as a companion if needed, but he/she could not provide any hints that could interfere with the child's performance. The clinician starts with a set of open questions for the parent, including the child's medical history, developmental milestones, and family background while the child is playing in the room to warm up. The clinician then guides the child to sit at a desk to administer structured activities and answer questions promptly. Clinicians have attempted to elicit as many language productions of the children as possible during interactions. A research assistant audio-recorded language samples of clinician-child interaction during the Performance Test using a high-quality recording pen. The average time for each audio recording was approximately 90 minutes.

Transcription

The audio-recorded language samples of children with ASD were transcribed using the annotation software Praat (Boersma & Weenink, 2022). The transcribers attended one week of Praat annotation training until their correct rate reached at least 95% in practicing transcriptions. Two transcribers independently annotated the same audio file and then double-checked and corrected the miscoded characters and mismatched transcripts. If errors or disagreements occurred, the transcribers discussed them until agreement was reached. 85% of the annotated transcriptions were randomly selected to check their reliability and revised by the research assistant. All transcripts were saved as TextGrid files. We excluded stereotyped language (i.e., echolalia, singing, and idiosyncratic language) following the conventions for natural language sampling in children with ASD (Tager-Flusberg & Anderson, 1991).

Coding

Automatic coding was assisted by manual checking and fine-tuning adjustments according to the Mandarin context. In this study, tagging labels were defined on the transcripts with Python 3.4 by adhering to authorized linguistic data resources, including the POS tagging guidelines for the Penn Chinese Treebank (CTB) (3.0) (Xia, 2000), the Stanford Dependencies Chinese in the HanLP Documentation (Chang et al., 2009), and the Bracketing Guidelines for the Penn Chinese Treebank (3.0) (Xue et al., 2000). These guidelines were modeled following the Government and Binding (GB) theory introduced by Chomsky (1981). In the dependency grammar approach, tokens are the only nodes in the tree. The relationships between tokens are graphically represented as arrows in governor-dependent links (see Appendix 1). A dependent node has only one governor, except for the root node, which provides a tree structure (Poiret et al., 2021). Constituency parsing and dependency rules provide syntactic analysis, and their structural properties are represented in brackets (phrase-structure trees). The main criteria for tagging lie in the principles of contemporary linguistic theory. We follow the X-bar theory (Radford, 1988) for a definition of syntactic phrases such as Noun phrases (NP) and Verb phrases (VP); other syntactic structures also follow a two-level structure. This hierarchical model of early grammar reflects the adjacent and non-adjacent dependencies that govern simple and complex syntactic structures.

The 29 lexical tags (Appendix 2) included in this transcript were based on the 33 POS tags in the published Penn Chinese Treebanks (Xia, 2000) and adjusted based on Mandarin lexical categories. Following this convention, CWs include nouns, verbs, adjectives, and adverbs, whereas FWs include pronouns, prepositions, conjunctions, and determiners, including numerals and quantifiers (Fromkin et al., 2017). Moreover, interjections and onomatopoeias also belong to FWs in Mandarin (e.g., Huang & Liao, 2011). We annotated ten categories of syntactic structures (Appendix 3) based on 23 syntactic tags according to The Bracketing Guidelines for the Penn

Chinese Treebank (3.0) (Xue et al., 2000). Additionally, we annotated 17 categories of GRs (Appendix 4) based on the 45 named GRs in Mandarin Chinese, according to Stanford Dependencies Chinese in the HanLP Documentation (Chang et al., 2009).

Analyses

Following previous studies, we first reported the descriptive statistics of each measure of variance, including CWs, FWs, and lexical subcategories, as measured by types and tokens within the CA band. Repeated measures in the general linear model were applied to analyze within-subject factors. Considering age and lexical production interaction, we used lexical production as the dependent variable and the CA subgroups as independent variables. Box's Test of Equality of Covariance and Mauchly's test of sphericity were performed. Then, a one-way ANOVA was conducted to explore subgroup differences in each lexical category. Bonferroni's post hoc correction was performed to adjust for multiple comparisons within ASD subgroups. Noun-to-verb comparisons were performed for both types and tokens. We first examined the overall numbers of nouns and verbs using paired sample t-tests. Then we examined the noun-to-verb ratio by computing the production of nouns over the total sum of nouns and verbs. While a score greater than 0.5 means a noun-biased lexical production, and a score less than 0.5 means a verb-biased lexical production. We also investigated the type-token ratios (TTR) for nouns and verbs to explore the differences between nouns and verbs. We conducted Box-Whisker plots of syntactic structures and radar plots of GRs to investigate grammatical development. Canonical and Pearson's correlation analyses were conducted to explore the interrelationships between lexical growth and grammatical development. Data analyses were conducted using SPSS version 25.0.

Results

Comparison of lexical development in children with ASD

The first aim of this study was to examine the lexical profiles of Mandarin-speaking preschoolers with ASD. The results of the repeated-measures analyses in Table 2 revealed the

Table 2. Mean (*M*), *SD*, and post hoc analysis of lexical categories measured by types

age(y;m)	2;0-2;11A	3;0-3;11B	4;0-4;11C	5;0-5;11D	6;0-7;0E	*F*	*p*	η^2_p	Bonferroni's test
LCty	*M(SD)*	*M(SD)*	*M(SD)*	*M(SD)*	*M(SD)*				($p < .05$)
tot	148.40	173.00	241.17	293.83	250.25	4.90	**.001**	.21	A < D, B < D
	(71.84)	(95.92)	(104.58)	(112.63)	(87.71)				
CWs	107.00	128.63	177.78	222.06	186.38	4.65	**.002**	.20	B < D
	(50.12)	(75.05)	(82.01)	(89.97)	(72.26)				
FWs	41.40	44.38	63.39	71.78	63.88	5.23	**.001**	.22	B < D
	(22.70)	(21.96)	(24.61)	(23.90)	(16.98)				
nou	49.00	58.33	81.20	96.80	82.30	4.25	**.004**	.18	B < D
	(23.76)	(32.42)	(40.13)	(35.76)	(30.50)				
ver	40.20	49.38	67.61	85.83	70.94	4.54	**.002**	.19	A < D, B < D
	(16.25)	(30.29)	(30.53)	(36.20)	(29.09)				
adj	7.40	9.58	12.06	16.50	13.44	3.09	**.021**	.14	B < D
	(4.45)	(6.64)	(7.17)	(9.06)	(6.06)				
adv	10.40	11.33	16.89	22.94	19.69	3.98	**.006**	.17	B < D
	(8.79)	(8.64)	(10.81)	(11.89)	(11.16)				
pro	7.80	9.50	12.33	14.67	12.25	3.80	**.007**	.17	B < D
	(3.83)	(4.78)	(5.91)	(4.83)	(3.99)				
num	9.60	11.88	17.17	16.94	17.25	3.18	**.018**	.14	/
	(6.19)	(7.19)	(7.42)	(7.13)	(5.87)				
cla	1.80	2.29	4.39	5.72	4.94	5.65	**< .001**	.23	B < D, B < E
	(1.92)	(1.46)	(2.83)	(3.44)	(3.17)				
pre	.80	1.71	3.72	4.89	3.31	5.65	**< .001**	.23	A < D, B < D
	(1.79)	(1.78)	(2.30)	(3.32)	(2.60)				
con	.60	.29	.78	1.11	.88	2.12	.087	.10	/
	(.89)	(.75)	(.88)	(1.32)	(.81)				
par	9.80	8.88	12.56	15.28	13.25	5.22	**.001**	.22	B < D
	(6.18)	(5.03)	(4.50)	(5.86)	(2.50)				
int	10.40	8.79	11.44	11.78	11.00	1.17	.330	.06	/
	(7.70)	(4.66)	(5.62)	(5.08)	(3.81)				
ono	0.60	1.04	1.00	1.39	1.00	.630	.644	.03	/
	(.89)	(1.20)	(1.03)	(1.15)	(1.10)				

Abbreviations: age(y;m), CA(years; months); LCty, lexical categories measured by types; tot, total vocabulary; nou, nouns; ver, verbs; adj, adjectives; adv, adverbs; pro, pronouns; num, numerals; clas, classifiers; pre, preposition; con, conjunction; par, particles; int, interjections; ono, onomatopoeia.

Note: Bonferroni adjusted pairwise comparisons, all *p*s < .001.

main effects of age on the total sum of types, F (4, 80) = 4.90, p = .001, η^2_p = .21; types of CWs, F (4, 80) = 4.65, p = .002, η^2_p = .20; and types of FWs, F (4, 80) = 5.23, p = .001, η^2_p = .22. Likewise, we observed the main effects of age on the four lexical subcategories included in CWs (ps < .05) and on most of the eight lexical subcategories included in FWs (ps < .05), except for the conjunction, interjection, and onomatopoeia types (ps > .05). Bonferroni's post hoc tests showed significant group differences for the total sum of types and tokens, CWs, FWs, and ten of the 12 lexical subcategories (ps < .05). As expected, noun and verb types had dominant lexical production in each age group (nouns: 32.9% – 33.7%; verbs: 27.1% – 29.2%), while the types of classifiers, prepositions, conjunctions, and onomatopoeia only accounted for a small portion (i.e., less than 2%).

Table 3 presents the rapid growth of lexical development as measured by tokens in each age group. The results of repeated-measures analyses revealed the main effects of CA subgroups on the total sum of tokens, F (4, 90) = 4.61, p = .002, η^2_p = .20, tokens of CWs, F (4, 80) = 3.76, p = .008, η^2_p = .17, and tokens of FWs, F (4, 80) = 5.79, p < .001, η^2_p = .23. Similarly, we observed the main effects of the CA subgroups on eight out of the 12 lexical subcategories (ps < .05), except for adjective, numeral, interjection, and onomatopoeia tokens (ps > .05). Bonferroni's post hoc tests showed that group differences were significant for the total sum of tokens, CWs, FWs, and six of the 12 lexical subcategories (ps < .05). As expected, noun and verb tokens had dominant lexical production in each CA subgroup (nouns: 19.9% – 27.3%; verbs: 26.2% – 27.0%), while the token production of classifiers, prepositions, conjunctions, and onomatopoeia accounted for only a small portion (i.e., less than 2.6%). As expected, group comparisons revealed that 2-to-3-year-old children had the lowest lexical production, whereas 5-to-6-year-old children had the highest. Moreover, compared to 3-to-4-year-olds, children between five and six years old showed

345 significantly higher production in total vocabulary, CWs, FWs, and six of the ten lexical
346 subcategories, except for numerals and conjunctions.

347 Table 3. Mean (M), SD, and post hoc analysis of lexical categories measured by tokens

age(y;m)	2;0-2;11A	3;0-3;11B	4;0-4;11C	5;0-5;11D	6;0-7;0E	*F*	*p*	η^2_p	Bonferroni's test
LCto	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*				($p < .05$)
tot	326.00	449.88	704.94	859.72	633.94	4.61	**.002**	.20	A < D, B < D
	(199.92)	(302.11)	(409.26)	(411.43)	(311.92)				
CWs	211.60	286.42	434.22	516.94	406.81	3.76	**.008**	.17	B < D
	(124.19)	(197.04)	(253.28)	(251.50)	(208.14)				
FWs	114.40	163.46	270.72	342.78	227.13	5.79	**< .001**	.23	A < D, B < D
	(76.65)	(109.07)	(164.15)	(168.00)	(106.63)				
nou	89.00	105.33	163.39	171.06	143.56	2.95	**.025**	.13	/
	(47.38)	(63.89)	(102.49)	(72.97)	(70.17)				
ver	86.40	121.63	190.61	225.67	167.00	3.87	**.007**	.17	B < D
	(57.28)	(90.99)	(114.36)	(112.96)	(86.88)				
adj	14.60	28.21	30.83	44.33	34.81	1.84	.129	.09	/
	(10.46)	(28.79)	(19.51)	(26.52)	(26.36)				
adv	21.60	31.25	49.39	75.89	61.44	4.27	**.004**	.18	A < D, B < D
	(19.90)	(28.4)	(39.74)	(51.56)	(41.8)				
pro	24.80	58.67	107.39	131.89	71.50	5.54	**.001**	.23	A < C, A < D,
	(17.51)	(44.69)	(85.16)	(78.33)	(37.03)				A < E, B < D
num	33.20	34.42	49.89	54.22	44.81	2.38	.059	.11	/
	(25.09)	(22.21)	(28.85)	(23.52)	(18.22)				
cla	6.40	6.71	15.50	22.56	15.38	6.25	**< .001**	.25	B < C, B < D
	(8.20)	(5.99)	(10.74)	(15.37)	(11.25)				
pre	.80	3.46	6.44	11.00	5.44	5.09	**.001**	.21	A < D, A < C,
	(1.79)	(4.45)	(5.71)	(8.86)	(5.39)				B < D
con	.80	.33	1.67	2.67	2.31	3.58	**.010**	.16	/
	(1.10)	(0.82)	(2.14)	(3.13)	(2.7)				
par	22.80	36.38	58.61	84.22	53.00	4.73	**.002**	.20	A< D, B < D
	(18.24)	(34.60)	(41.90)	(50.46)	(31.7)				
int	24.40	22.00	29.61	33.67	33.19	1.14	.342	.06	/
	(26.76)	(13.60)	(24.63)	(23.87)	(18.75)				
ono	1.20	1.50	1.61	2.56	1.50	.820	.514	.04	/
	(2.17)	(1.96)	(1.82)	(2.94)	(2.00)				

Abbreviations: age(y;m), age(years; months); LCto, lexical categories measured by tokens; tot, total vocabulary; nou, nouns; ver, verbs; adj, adjectives; adv, adverbs; pro, pronouns; num, numerals; clas, classifiers; pre, preposition; con, conjunction; par, particles; int, interjections; ono, onomatopoeia.

Note: Bonferroni adjusted pairwise comparisons, all *p*s < .001.

348

Noun bias in early lexical acquisition

Table 4 displays overall noun-to-verb comparisons, revealing the noun/verb bias differences within CA subgroups. We observed that Mandarin-speaking preschoolers with ASD produced significantly more noun types than verb types ($p < .05$), except for children under three

Table 4. Nouns and verbs production

CA bands		Nouns-to-Verbs paired sample t-test			Nouns/(Nouns + Verbs) ratio	
		t	*p*	*MeD* (95%*CI*)	*M*	*SD*
total	types	7.45	**< .001**	10.9 (7.99 ~ 13.81)	**.53***	.09
	tokens	-4.63	**< .001**	5.94 (-39.30 ~ -15.66)	**.47***	.10
	N-TTR				**.58****	.12
	V-TTR				.45	.14
2;0-2;11	types	2.18	.095	8.8(-2.42 ~ 20.02)	.53	.07
	tokens	0.15	.888	2.6(-45.37 ~ 50.57)	.52	.08
	N-TTR				.55	.16
	V-TTR				.53	.13
3;0-3;11	types	3.67	**.001**	8.96(3.90 ~ 14.01)	.52	.14
	tokens	-1.58	.128	-16.29(-37.63 ~ 5.05)	.47	.14
	N-TTR				**.60****	.15
	V-TTR				.49	.18
4;0-4;11	types	3.33	**.004**	13.61(4.99 ~ 22.24)	**.54***	.06
	tokens	-1.90	.074	-27.22(-57.40 ~ 2.96)	.47	.09
	N-TTR				**.54****	.10
	V-TTR				.42	.14
5;0-5;11	types	4.31	**< .001**	10.94(5.59 ~ 16.30)	**.54***	.04
	tokens	-4.16	**.001**	-54.61 (-82.32 ~ -26.90)	**.45***	.07
	N-TTR				**.59****	.09
	V-TTR				.42	.11
6;0-7;0	types	3.26	**.005**	11.38 (3.93 ~ 18.82)	**.54***	.06
	tokens	-2.20	**.044**	-23.44 (-46.19 ~ -0.69)	.47	.09
	N-TTR				**.60****	.08
	V-TTR				.46	.10

Abbreviation: N-TTR, type/token ratio of nouns, V-TTR, type/token ratio of verbs, MeD, Mean Difference.
Two-tailed paired-sample t-test. The significant level is $p < 0.05$.
*Indicates that Nouns/(Nouns + Verbs) ratio is significantly different from 0.50, *p*s < .05.
**Indicates that TTR of nouns is significantly different from TTR of verbs, *p*s < .05

years old (p = .095). According to the noun-verb ratios of each age group and overall comparison measured by types, we categorized them as noun biased for ratios above 0.5. Similar to type production, we also compared noun and verb token productions. Interestingly, significantly more verb tokens than noun tokens were only found in Mandarin-speaking preschoolers with ASD over five years old (t = -4.63, p < .05). Moreover, the TTR for verbs and nouns was unequal, as the overall mean TTR for nouns (M = .58, SD = .12) was significantly higher than that for verbs (M = .45, SD = .14). Interestingly, we observed similar trends in TTR for verbs and nouns in all age subgroups except children between two and three years old. Thus, in general, 2-7-year-old Mandarin-speaking preschoolers with ASD showed a clear preponderance of noun types over verb types, though tokens showed a different tendency; their type-token ratios for nouns were much higher than those of verbs.

Grammatical development in Mandarin-speaking preschoolers with ASD

We explored the grammatical development of Mandarin-speaking preschoolers with ASD by analyzing the production of syntactic structures, GRs, and grammatical complexity. In the repeated-measures analyses, the production of syntactic structures based on constituency parsing was the independent variable, and the CA subgroups were the dependent variables. Significant main effects of age were observed for the NP, F (4, 80) = 4.63, p = .002, VP, F (4, 80) = 3.75, p = .008, ADVP, F (4, 80) = 4.11, p = .005, CLP, F (4, 80) = 6.08, p < .001, QP, F (4, 80) = 4.85, p = .002, PP, F (4, 80) = 5.60, p = .001, CP, F (4, 80) = 4.26, p = .004, DNP, F (4, 80) = 5.95, p < .001, IP, F (4, 80) = 3.44, p = .012. Post hoc analyses showed that 5-to-6-year-olds had the highest syntactic structures and were significantly higher than 3-to-4-year-olds. Moreover, Figure 1 revealed that Mandarin-speaking preschoolers with ASD produced more NP, VP, ADVP, and IP than other syntactic structures.

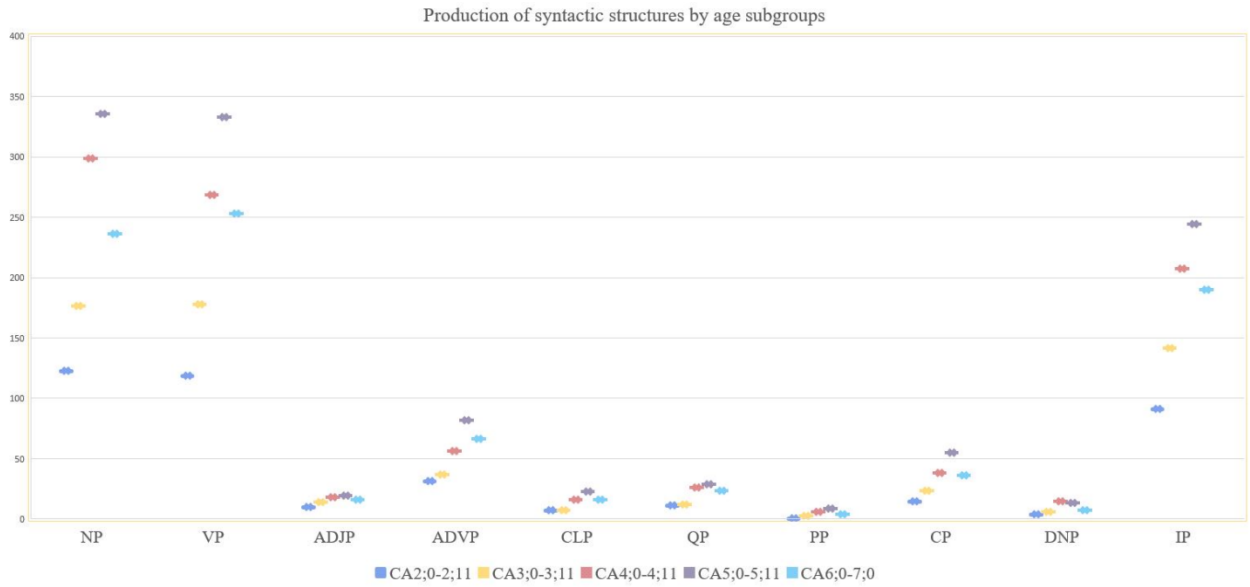


Figure 1. Box-Whisker plots for the mean of the syntactic structures

The developmental trajectories of the GRs were similar in each CA subgroup, as displayed by the radar plot in Figure 2. As expected, 5-to-6-year-old children with ASD outperformed the others in producing GRs. Generally, Mandarin-speaking preschoolers with ASD produced more nominal subjects, adverbial modifiers, and direct objects, which is consistent with the production of syntactic structures based on constituency parsing. Additionally, the one-way ANOVA analysis in Table 5 shows that the grammatical complexity of Mandarin-speaking preschoolers with ASD varied significantly in subgroups (MLU, F (4, 80) = 5.924, $p < .001$; MLU5, F (4, 80) = 4.98, $p = .001$). Generally, grammatical complexity increased with age, while 5-to-6-year-old children produced the longest MLU (M = 2.16, SD = .43) and MLU5 (M = 8.53, SD = 3.37), which was significantly longer than that of children under four years old ($p < .05$).

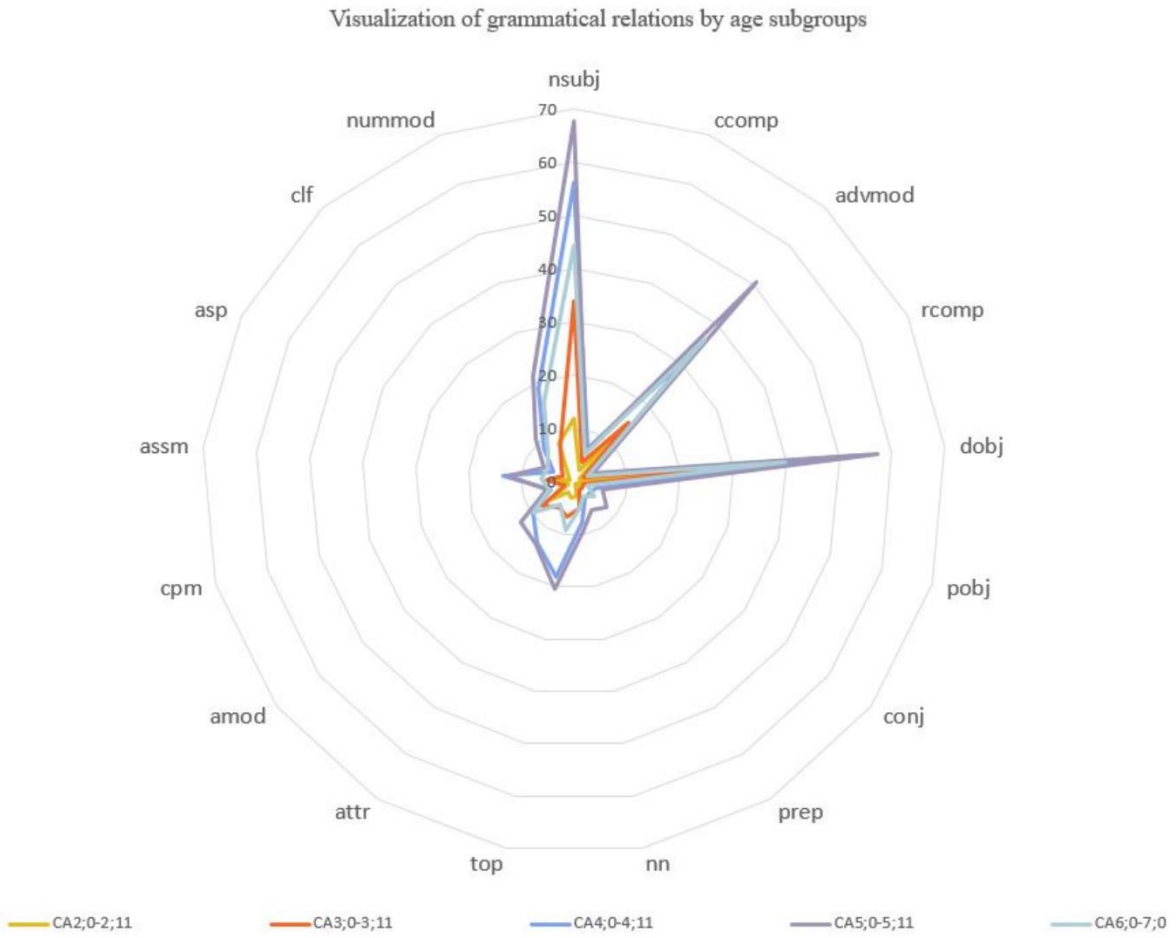


Figure 2. Radar plot for the means of GRs

Table 5. ANOVA analysis of MLU and MLU5 by age subgroups

Age (years; months)	2;0-2;11(A) *M(SD)*	3;0-3;11(B) *M(SD)*	4;0-4;11(C) *M(SD)*	5;0-5;11(D) *M(SD)*	6;0-7;0 (E) *M(SD)*	*F*	*p*	post hoc
MLU	1.48 (.31)	1.69 (.38)	1.99 (.36)	2.16 (.43)	1.91 (.31)	5.92	< .001	A < D,B < D
MLU5	4.44 (1.71)	5.44 (2.03)	7.06 (2.56)	8.53 (3.37)	6.94 (2.22)	4.98	.001	A < D,B < D

Note: The mean difference is significant at the 0.05 level. Adjustments for multiple comparisons were conducted by Bonferroni.

Associations between lexical acceleration and grammatical development

The results of the canonical correlation analysis (see Appendix 5) revealed strong associations between lexical production and grammatical development, including syntactic

structure (r = .998, p < .001), GRs (r = .985, p < .001), and grammatical complexity (r = .846, p < .001). The Pearson's correlation coefficient values for lexical categories and the ten syntactic structures ranged from .574 to .984 (ps < .05), the 17 GRs ranged from .194 to .455 (ps < .05), and grammatical complexity ranged from .276 to .863 (ps < .05).

To better understand the interrelationship between lexicons and grammar, we calculated the average mean of syntactic structures, GRs MLU, and MLU5 as the four independent variables of grammatical development. The scatterplot (Figure 3) revealed strong correlations between syntactic structures and types of CWs (r = .957, p < .001), types of FWs (r = .896, p < .001), tokens of CWs (r = .992, p < .001), and tokens of FWs (r = .963, p < .001). Likewise, Figure 4 illustrates the associations between the development of GRs and types of CWs (r = .927, p < .001), types of FWs (r = .867, p < .001), tokens of CWs (r = .948, p < .001), and tokens of FWs (r = .948, p < .001). They were stable across the dataset and did not rely on outliers for their significance. Figures 5 and 6 show that lexical categories and grammatical complexity are intertwined. As observed from the scatterplots, MLU (r = .808, p < .001) and MLU5 (r = .795, p < .001) were positively and linearly correlated with the type of FWs across the dataset. By contrast, the associations between grammatical complexity and CWs and the tokens of FWs were likely to be significant only because of a few outliers.

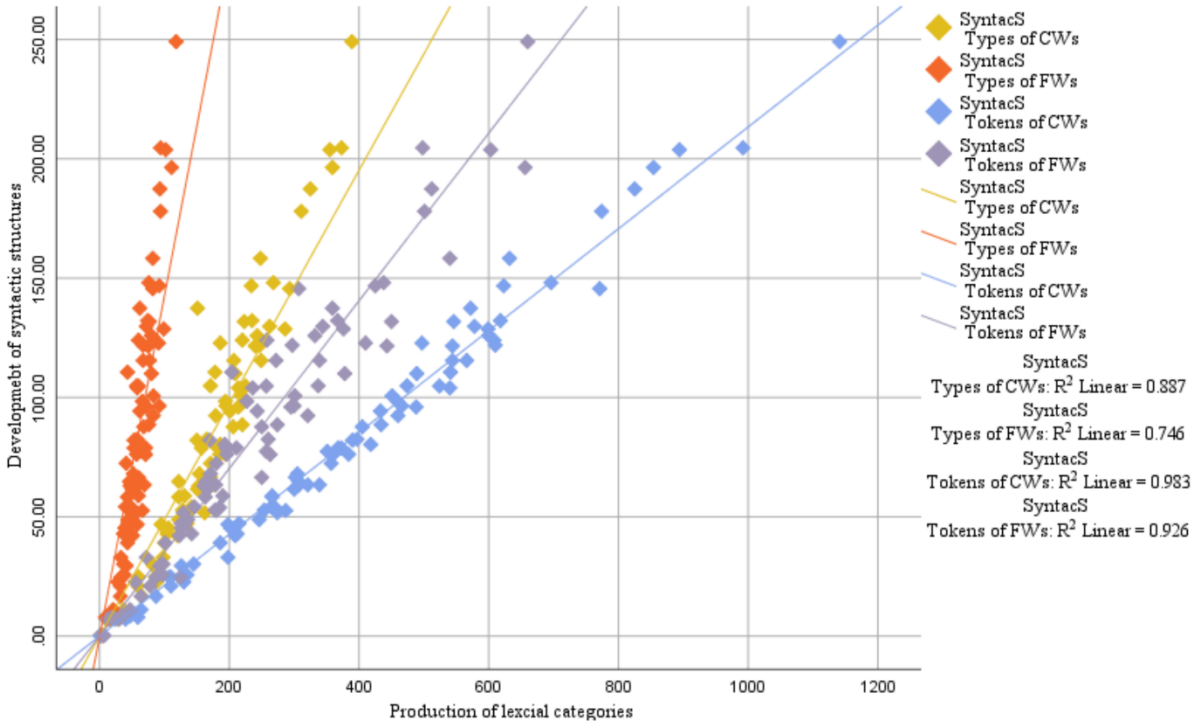


Figure 3. Correlations between vocabulary and syntactic structures

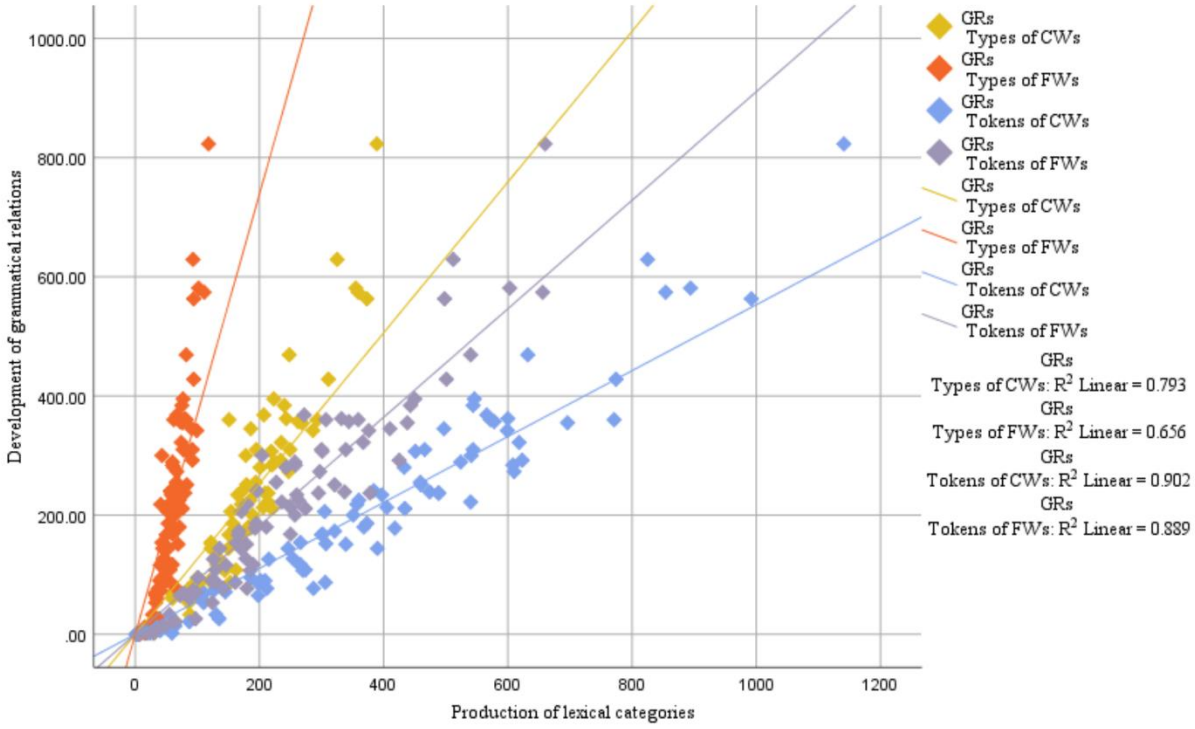
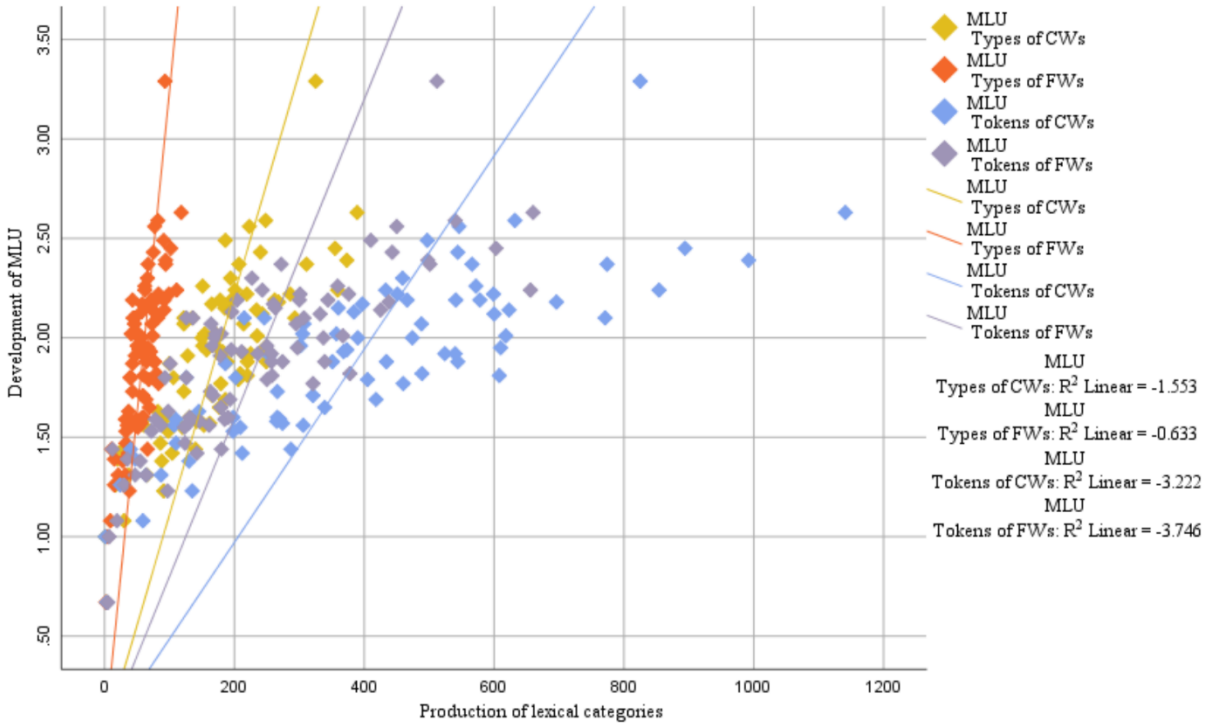


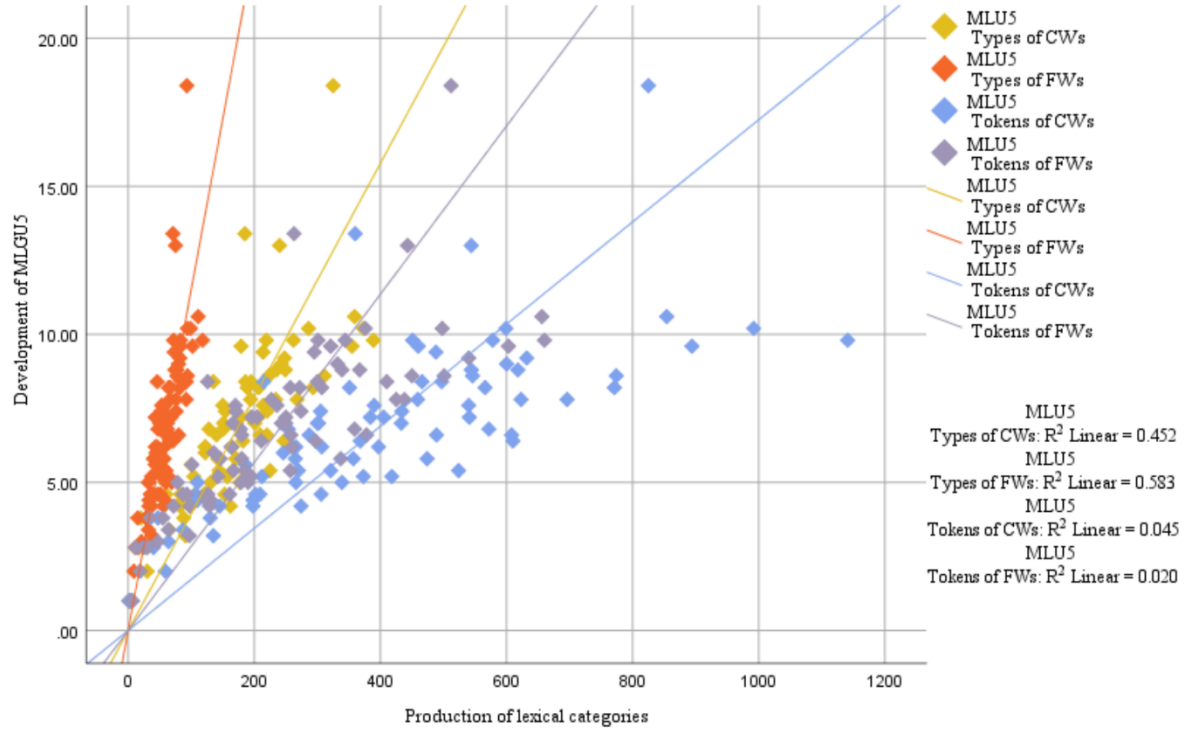
Figure 4. Correlations between vocabulary and GRs



421

422 Figure 5. Correlations between vocabulary and MLU

423



424

425 Figure 6. Correlations between vocabulary and MLU5

426

Discussion

This study performed within-ASD subgroup comparisons of language in 2-7-year-old Mandarin-speaking children with ASD using spontaneous language samples in clinician-child interactions to quantify the wide range of expressive language abilities and probe the lexical-grammatical interrelationships in ASD. Our findings confirmed that 5-to-6-year-old Mandarin-speaking children with ASD reach relatively stable language production. Their production of CWs, FWs, and lexical subcategories displayed significant differences compared to children under four. Specifically, noun bias is a universal lexical learning mechanism that may be preserved in Mandarin-speaking preschoolers with ASD. Additionally, the developmental trajectories of syntactic structures, GRs, and grammatical complexity were comparable in each CA subgroup. Moreover, grammatical development is intimately intertwined with lexical production in Mandarin-speaking preschoolers with ASD, supporting the hypothesis that different language domains develop synergistically and interactively during early language acquisition.

Lexical variables of Mandarin-speaking preschoolers with ASD

An overview of the subgroup findings highlights the distinctive features of lexical profiles in Mandarin-speaking preschoolers with ASD. Our results showed that their lexical production increased with age, echoing findings in English-speaking preschoolers with ASD via CDI (e.g., Charman et al., 2003) and parent-child interaction (e.g., Fusaroli et al., 2019). However, the trajectories of lexical development in Mandarin-speaking preschoolers with ASD are highly variable before they reach stability between five and seven of age. Reports on English-speaking children with ASD have revealed that their language development is characterized by enduring trait-like stability between 4-6 years old (Brinchmann et al., 2019; Pickles et al., 2014). It is also

consistent with studies on TD children showing that five-year-olds are typically versatile language users with large vocabularies (Vehkavuori et al., 2021).

Mandarin-speaking preschoolers with ASD demonstrate steady growth in CWs, FWs, and each lexical subcategory, except for interjection and onomatopoeia. Specifically, comparisons within the CA subgroups revealed that nouns and verbs are dominant lexical subcategories, while prepositions, conjunctions, and onomatopoeias occupy only a small portion of early lexical production. Similarly, Song et al. (2022) demonstrated that 3-8-year-old Chinese-speaking children with ASD produced relatively few prepositions, adjectives, adverbs, and pronouns (less than 10% of each component). The results confirmed that the growth of FWs was slow because there was a certain CWs predominance with nouns and verbs (Nóro et al., 2015). It is reasonable that the initial recognition and development of CWs and FWs differ in their unique acoustic, phonological, and distributional properties (i.e., word frequency) (Shi et al., 2006). Generally, if a child produces a small number of words, such as classifiers, prepositions, conjunctions, and onomatopoeias, these words increase slowly with age (Kim et al., 2000). It was assumed that most FWs belonged to the abstract grammatical category, which requires a combination of additional information for processing. Children with ASD may have weak central coherence (Bojda et al., 2021), which may account for their difficulties in learning words that require strong contextual support to decode meaning (Schafer et al., 2013).

Noun bias in early lexical learning

Mandarin-speaking preschool children with ASD use a relatively higher proportion of nouns than verbs in early language production. Indeed, Gentner's (1982) initial analysis of six languages (Mandarin Chinese, Japanese, Kaluli, German, English, and Turkish) proposes a universal noun-first predisposition. This finding is consistent with long-standing and cross-

linguistic findings that noun bias is not only limited to noun-friendly languages, such as English, Spanish, Dutch, French, Hebrew, and Italian (Bornstein et al., 2004), but also exists in verb-friendly languages, such as Korean (Au et al., 1994) and Japanese (Ogura et al., 2006). Moreover, our findings are consistent with previous research on children with ASD who exhibit noun bias in early lexical learning (Luyster et al., 2007; Rescorla & Safyer, 2013; Swensen et al., 2007). Mandarin-speaking preschool children with ASD appear to preserve the noun bias as a strategy for early lexical learning. Thus, noun bias is a universally shared cognitive factor in early lexical learning (Imai et al., 2008).

However, it is essential to note that noun dominance was insignificant in Mandarin-speaking children with ASD aged between two and three years. It aligns with the longitudinal research of Chen et al. (2016) that nouns were not overwhelmed at 18 months of age, but noun bias became apparent with age in Mandarin-exposed TD children. Likewise, Tardif (1996) demonstrated that 22-month-old Mandarin-exposed TD children produce more verbs than nouns. Moreover, children over five years of age produced significantly more verb tokens than noun tokens; however, they used only a small number of verbs more frequently (i.e., low type-token ratios for verbs in almost all age subgroups). The advantage of verb tokens at specific age periods in Mandarin-speaking preschoolers with ASD can be accounted for by the influence of parents' input (e.g., Tardif et al., 1997), and the subjects and/or objects are ellipses in the oral language of Mandarin (e.g., Xu et al., 2022). Nouns are often dropped in the surface structure, and verbs can occur alone as sentences in Mandarin; therefore, the frequently used verb tokens are repeated by minimally verbal children with ASD (Lee & Naigles, 2005). Further studies should explore the expressive language of young children with ASD in parent-child interactions to distinguish the influence of maternal input in early language acquisition.

Grammatical development in Mandarin-speaking preschoolers with ASD

The developmental trajectories of grammatical patterns are similar in each CA subgroup of Mandarin-speaking preschoolers with ASD. They produced more noun, verb, and adverb phrases than other syntactic structures. It differs from the findings in English-speaking children with ASD that they were only atypical in using verb phrases measured by IPSyn (Park et al., 2012). These controversial findings highlight the need for future studies to further explore grammatical development in children with ASD by collecting data from diverse samples. By contrast, nominal subjects, adverbial modifiers, and direct objects were produced more frequently than other GRs in Mandarin-speaking preschoolers with ASD. As previously reported, dependencies governing subject-noun/verb agreement and auxiliary/inflectional morpheme relations are acquired earlier than dependencies involving more abstract constituent relationships. These examples indicate congruent GRs within the nominal and/or verbal context, similar to the findings in 2-4-year-old French-speaking children (Le Normand & Thai-Van, 2022).

Crucially, MLU and MLU5 were weakly but positively correlated with the age of Mandarin-speaking preschoolers with ASD. The MLU reached its longest between five and six years old, which is consistent with the findings of recent studies. For instance, Kover et al. (2014) reported that 37-53-month-old children with ASD have an MLU of 2.35 in examiner-child play and an MLU of 2.10 in ADOS screening. In contrast, Song and So (2022) demonstrated that 4-8-year-old Chinese-speaking children with ASD have an average MLU of 2.65. The contrasting results of MLU in children with ASD may be due to the wide age ranges, varied autism severities of participants, and the discrepancy in sampling instruments in language assessment (Barokova & Tager-Flusberg, 2020). In our study, 4-6-year-old children with ASD had a much longer MLU5 than aged-matched TD children in previous studies (e.g., Zhou & Zhang, 2020). These

discrepancies could be attributed to the grouping methodology, as the age interval between the groups was six months in the previous study and 12 months in our study.

Synchronous development between lexicon and grammar in Mandarin-speaking preschoolers with ASD

Early grammatical development is closely tied to lexical acceleration in Mandarin-speaking preschool children with ASD. The association between lexicon and grammar in each CA subgroup confirmed the finding in Finish TD children that the lexical-grammatical interrelationship was not limited to 2-year-old children, which could be extended to younger or older age groups (Stolt et al., 2009). Likewise, Weismer et al. (2010) reported that the specific lexical-grammatical relationships of preschoolers with ASD were less apparent via CDI. Explanations for these different findings may include varied measurements and participants.

Specifically, the development of syntactic structures, GRs, and grammatical complexity (MLU and MLU5) were closely intertwined with lexical categories in Mandarin-speaking preschoolers with ASD. This is in line with the findings in English-speaking TD children that there is a close linkage between structural ability and FWs in early language development (Day & Elison, 2022). In addition, by analyzing spontaneous speech, Le Normand et al. (2013) also revealed that types of FWs were a more powerful predictor of grammatical complexity in 2-4-year-old French-speaking TD children. Theoretical interpretations of these early productions in Mandarin-speaking preschool children with ASD could be accounted for by the general (not language-specific) language learning features that the acquisition of grammar starts early and proceeds hand-in-hand with word learning (de la Cruz-Pavía et al., 2021). We challenged the assumption that children produce multi-word utterances before they have grammatical knowledge (Le Normand & Thai-Van, 2022).

Limitations

Although our study provides new insights into the lexical and grammatical profiles of Mandarin-speaking preschoolers with ASD before intervention, several limitations must be considered. First, the language samples were restricted to the sCPEP-3 clinical screening context, and future studies should collect language samples from different contexts. In addition, participants were limited to 2-7-year-old Mandarin-speaking children with ASD; longitudinal studies focusing on children with ASD under two years old should be conducted. Additionally, several clinicians contributed to the data, and the reciprocal effect of their individualized language and children's production should be considered. Moreover, we only reported lexical-grammatical relationships, and future studies should explore the predictive effect of specific lexical categories on grammatical development.

Conclusions

To summarize, ours is the first study focused on profiles of CWs, FWs, and lexical categories of 2-7-year-old Mandarin-speaking children with ASD in clinician-child interactions. Our findings support that noun bias is a universal inherent mechanism and that lexical and grammatical domains are synchronically developed in Mandarin-speaking preschoolers with ASD. This study provides empirical support to address the linguistic stages of children with ASD and points to intervention directions.

Acknowledgments: The authors thank the National Social Science Fund of China (Grant Number 20BYY086), and the Science and Technology Program of Guangzhou City (Grant Number 202007030011) for supporting this research. We also thank the Data Science Research Center of Duke Kunshan University for supporting data annotation that made this study possible. We especially thank the children and clinicians who participated in the study.

Data availability statement: The data supporting this study's findings are available from the corresponding author upon reasonable request. The data are not publicly available due to ethical restrictions.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. https://doi.org/10.1176/appi.books.9780890425596

Au, K. T., Dapretto, M., & Song, Y. (1994). Input vs. constraints: early word acquisition in Korean and English. *Journal of Memory and Language*, *33*, 567–82.

Bacon, E. C., Osuna, S., Courchesne, E., & Pierce, K. (2019). Naturalistic language sampling to characterize the language abilities of 3-year-olds with autism spectrum disorder. *Autism*, 23(3), 699–712. https://doi.org/10.1177/1362361318766241

Barokova, M., & Tager-Flusberg, H. (2020). Commentary: Measuring Language Change Through Natural Language Samples. *Journal of Autism and Developmental Disorders*, 20.

Bassano, D. (2000). Early development of nouns and verbs in French: Exploring the interface between lexicon and grammar. *Journal of Child Language*, 27(3), 521–559. https://doi.org/10.1017/S0305000900004396

Bates, E., & Goodman, J. C. (1997). On the Inseparability of Grammar and the Lexicon: Evidence from Acquisition, Aphasia and Real-time Processing. *Language and Cognitive Processes*, 12(5–6), 507–584. https://doi.org/10.1080/016909697386628

Boersma, P., & Weenink, D. (2022). Praat: doing phonetics by computer. Disponível em: https://www.fon.hum.uva.nl/praat/.

585 Bojda, A., Srebnicki, T., Konowlek, L., & Bryńska, A. (2021). Weak central coherence – construct
586 conception, development, research methods. *Psychiatria Polska*, 55(6), 1373–1386.
587 https://doi.org/10.12740/PP/OnlineFirst/120931
588 Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., Marie-Germaine, P.,
589 Ruel, J., Venuti, P., & Vyt, A. (2004). Cross-Linguistic Analysis of Vocabulary in Young
590 Children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child*
591 *Development*, 75(4).
592 Brinchmann, E. I., Braeken, J., & Lyster, S. H. (2019). Is there a direct relation between the
593 development of vocabulary and grammar? *Developmental Science*, 22(1), e12709.
594 https://doi.org/10.1111/desc.12709
595 Brown, R. (1973). Development of the first language in the human species. *American Psychologist*,
596 28(2), 97-106. https://doi.org/10.1037/h0034209
597 Ceja Tel Toro, Pablo, Jidong Chen, and Bhuvana Narasimhan. 2016. Information structure in
598 bilingual Spanish-English child speech. Paper presented at the 2016 International
599 Workshop on Language Processing and Production, San Diego, CA, USA, June 15–16.
600 Chang, P.-C., Tseng, H., Jurafsky, D., & Manning, C. D. (2009). Discriminative Reordering with
601 Chinese Grammatical Relations Features. *Proceedings of SSST-3*, 51–59.
602 Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in
603 preschool children with autism spectrum disorder using the MacArthur Communicative
604 Development Inventory (Infant Form). *Journal of Child Language*, 30(1), 213–236.
605 https://doi.org/10.1017/S0305000902005482
606 Chen, Y., Niu, J., & Zhu, L. (2016). Is there a noun bias for Mandarin-speaking infants: from the
607 perspective of vocabulary size. *Journal of Psychological Science*, 39(3), 600-605. Chen,

J., Narasimhan, B., Chan, A., Yang, W., & Yang, S. (2020). Information Structure and Word Order Preference in Child and Adult Speech of Mandarin Chinese. *Languages*, 5(2), https://doi.org/10.3390/languages5020014

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Crain, S., & Lillo-Martin, D. C. (1999). *An introduction to linguistic theory and language acquisition*. Blackwell Publishers.

Day, T. K. M., & Elison, J. T. (2022). A broadened estimate of syntactic and lexical ability from the MB-CDI. *Journal of Child Language*, 49(3), 615–632. https://doi.org/10.1017/S0305000921000283

de la Cruz-Pavía, I., Marino, C., & Gervain, J. (2021). Learning word order: Early beginnings. *Trends in Cognitive Sciences*, 25(9), 802–812. https://doi.org/10.1016/j.tics.2021.04.011

DeThorne, L. S., & Channell, R. W. (2007). Clinician–Child Interactions: Adjustments in Linguistic Complexity. *American Journal of Speech-Language Pathology*, 16(2), 119–127. https://doi.org/10.1044/1058-0360(2007/016)

Dixon, J. A., & Marchman, V. A. (2007). Grammar and the Lexicon: Developmental Ordering in Language Acquisition. *Child Development*, 78(1), 190–212. https://doi.org/10.1111/j.1467-8624.2007.00992.x

Eigsti, I.-M., Bennetto, L., & Dadlani, M. B. (2007). Beyond Pragmatics: Morphosyntactic Development in Autism. *Journal of Autism and Developmental Disorders*, 37(6), 1007–1023. https://doi.org/10.1007/s10803-006-0239-2

Fromkin, V., Rodman, R., & Hyams, N. (2017). *An Introduction to Language* (11th ed.). Cengage.

Fu, C.-P., Hsieh, C.-L., Tseng, M.-H., Chen, Y.-L., Huang, W.-T., Wu, P.-C., & Chiang, F.-M. (2010). Inter-rater reliability and smallest real difference of the Chinese Psychoeducational

631 Profile-third edition for children with Autism Spectrum Disorder. *Research in Autism*
632 *Spectrum Disorders*, 4(1), 89–94. https://doi.org/10.1016/j.rasd.2009.09.002Fusaroli, R.,
633 Weed, E., Fein, D., & Naigles, L. (2019). Hearing me hearing you: Reciprocal effects
634 between child and parent language in autism and typical development. *Cognition*, 183, 1–
635 18. https://doi.org/10.1016/j.cognition.2018.10.022
636 Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural
637 partitioning. *Language*, 2, 301–334.
638 Hart, C. M., & Curtin, S. (2021). Trajectories of Vocabulary Development in Children with Autism
639 Spectrum Disorder Across Multiple Measures. *Journal of Autism and Developmental*
640 *Disorders*. https://doi.org/10.1007/s10803-021-05379-w
641 Hochmann, J.-R., Endress, A. D., & Mehler, J. (2010). Word frequency as a cue for identifying
642 function words in infancy. *Cognition*, 115(3), 444–457.
643 https://doi.org/10.1016/j.cognition.2010.03.006
644 Huang, B., & Liao, X. (2011). *The Contemporary Chinese*. Higher Education Press.
645 Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., & Golinkoff, R. M. (2008). Novel Noun
646 and Verb Learning in Chinese-, English-, and Japanese-Speaking Children. *Child*
647 *Development*, 79(4), 979–1000. https://doi.org/0009-3920/2008/7904-0012
648 Kim, M., McGregor, K., & Thompson, C. (2000). Early lexical development in English-and
649 Korean-speaking children: language-general and language-specific patterns. *Journal of*
650 *Child Language*, 27:225-54.
651 Kover, S. T., Davidson, M. M., Sindberg, H. A., & Weismer, S. E. (2014). Use of the ADOS for
652 Assessing Spontaneous Expressive Language in Young Children With ASD: A

Comparison of Sampling Contexts. *Journal of Speech, Language, and Hearing Research*, 57(6), 2221–2233. https://doi.org/10.1044/2014_JSLHR-L-13-0330

Lee, J. N., & Naigles, L. R. (2005). The Input to Verb Learning in Mandarin Chinese: A Role for Syntactic Bootstrapping. *Developmental Psychology*, 41(3), 529–540. https://doi.org/10.1037/0012-1649.41.3.529Legendre, G. (2006). Early Child Grammars: Qualitative and Quantitative Analysis of Morphosyntactic Production. Cognitive Science, 30(5), 803–835. https://doi.org/10.1207/s15516709cog0000_85

Le Normand, M. T., Moreno-Torres, I., Parisse, C., & Dellatolas, G. (2013). How Do Children Acquire Early Grammar and Build Multiword Utterances? A Corpus Study of French Children Aged 2 to 4. *Child Development*, 84(2), 647–661. https://doi.org/10.1111/j.1467-8624.2012.01873.x

Le Normand, M.-T., & Thai-Van, H. (2022). The role of Function Words to build syntactic knowledge in French-speaking children. *Scientific Reports*, 12(1), 544. https://doi.org/10.1038/s41598-021-04536-6

Leung, M.-T., & Li, H.-L. (2015). Hierarchical phrase-based grammatical analysis of language samples from Cantonese-speaking children with and without autism. *Clinical Linguistics & Phonetics*, 29(8–10), 736–747. https://doi.org/10.3109/02699206.2015.1048529

Li, C., & Thompson, S. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkerley, CA: University of California Press.

Liu, X., Chen, J., Zhang, K., Wang, X., Wang, G., & Zhang, R. (2022). The evaluation of the cognitive and language abilities of autistic children with interactive game technology based on the PEP-3 scale. *Education and Information Technologies*. https://doi.org/10.1007/s10639-022-11114-4

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). *The Autism Diagnostic Observation Schedule-Generic (ADOS-G)*. Western Psychological Service.

Luyster, R., Qiu, S., Lopez, K., & Lord, C. (2007). Predicting outcomes of children referred for autism using the MacArthur-Bates Communicative Development Inventory. *Journal of Speech, Language, and Hearing Research: JSLHR*, 50(3), 667–681. https://doi.org/10.1044/1092-4388(2007/047)

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, 21(2), 339–366. https://doi.org/10.1017/S0305000900009302

Mawhood, L., Howlin, P., & Rutter, M. (2000). Autism and developmental receptive language disorder--a comparative follow-up in early adult life. I: Cognitive and language outcomes. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 41(5), 547–559. https://doi.org/10.1111/1469-7610.00642

McGregor, K. K., Berns, A. J., Owen, A. J., Michels, S. A., Duff, D., Bahnsen, A. J., & Lloyd, M. (2012). Associations Between Syntax and the Lexicon Among Children With or Without ASD and Language Impairment. *Journal of Autism and Developmental Disorders*, 42(1), 35–47. https://doi.org/10.1007/s10803-011-1210-4

Narasimhan, B., & Dimroth, C. (2008). Word order and information status in child language. *Cognition*, 107(1), 317–329. https://doi.org/10.1016/j.cognition.2007.07.010

Nóro, L. A., Silva, D. D. da, Wiethan, F. M., & Mota, H. B. (2015). Aquisição lexical inicial e verificação da hipótese do viés nominal. *Revista CEFAC*, 17(suppl 1), 52–59. https://doi.org/10.1590/1982-0216201517s14613

Ogura, T., Dale, P. S., Yamashita, Y., Murase, T., & Mahieu, A. (2006). The use of nouns and verbs by Japanese children and their caregivers in book-reading and toy-playing contexts. *Journal of Child Language*, 33(1), 1–29. https://doi.org/10.1017/S0305000905007270

Park, C. J., Yelland, G. W., Taffe, J. R., & Gray, K. M. (2012). Morphological and syntactic skills in language samples of preschool aged children with autism: Atypical development? *International Journal of Speech-Language Pathology*, 14(2), 95–108. https://doi.org/10.3109/17549507.2011.645555

Pickles, A., Anderson, D. K., & Lord, C. (2014). Heterogeneity and plasticity in the development of language: A 17-year follow-up of children referred early for possible autism. *Journal of Child Psychology and Psychiatry*, 55(12), 1354–1362. https://doi.org/10.1111/jcpp.12269

Poiret, R., Wong, T.-S., Lee, J., Gerdes, K., & Leung, H. (2021). Universal Dependencies for Mandarin Chinese. *Language Resources and Evaluation*. https://doi.org/10.1007/s10579-021-09564-2

Radford, A. (1988). *Transformational grammar*. Cambridge University Press.

Radford, A., Atkinson, M., Britain, D., Clahsen, H., & Spencer, A. (2008). *Linguistics An Introduction* (2nd ed.). Cambridge University Press.

Rescorla, L., & Safyer, P. (2013). Lexical composition in children with autism spectrum disorder (ASD). *Journal of Child Language*, 40(1), 47–68. https://doi.org/10.1017/S0305000912000232

Robins, D. L., Casagrande, K., Barton, M., Chen, C. M., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the modified checklist for Autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics*, 133(1), 37–45. https://doi.org/10.1542/peds.2013-1813

Scarborough, H. S. (1990). Index of Productive Syntax. Applied Psycholinguistics, 11(1), 1-22. https://doi.org/10.1017/s0142716400008262

Schafer, G., Williams, T. I., & Smith, P. T. (2013). Which Words are Hard for Autistic Children to Learn? *Mind & Language*, 28(5), 661–698. https://doi.org/10.1111/mila.12038

Schopler, E., & Reichler, R. J. (1979). Individual assessment and treatment for autistic and developmental disabled children: Psychoeducational Profile (Vol. 1). Baltimore: University Park Press.

Schopler, E., Lansing, M. D., Reichler, R. J., & Marcus, L. M. (2005). *Examiner's manual of psychoeducational profile* (3rd ed.). Pro-ed Incorporation.

Semsem, M., & Chen, J. 2019. The use of word order to mark information status in adult and child Arabic. *In The Proceedings of the 30th Western Conference on Linguistics*. Edited by T. Driscoll. Fresno: Department of Linguistics, California State University, pp. 173–78.

Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and Representation of Function Words in English-Learning Infants. *Infancy*, 10(2), 187–198. https://doi.org/10.1207/s15327078in1002_5

Smith, V., Mirenda, P., & Zaidman-Zait, A. (2007). Predictors of expressive vocabulary growth in children with autism. *Journal of Speech Language and Hearing Research*, 50(1), 149–160. https://doi.org/10.1044/1092-4388(2007/013)

Song, X.-K., Lee, C., & So, W.-C. (2022). Examining Phenotypical Heterogeneity in Language Abilities in Chinese-Speaking Children with Autism: A Naturalistic Sampling Approach. *Journal of Autism and Developmental Disorders*, 52(5), 1908–1919. https://doi.org/10.1007/s10803-021-05104-7

Song, X.-K., & So, W.-C. (2022). The influence of child-based factors and parental inputs on expressive language abilities in children with autism spectrum disorder. *Autism*, 26(6), 1477–1490. https://doi.org/10.1177/13623613211054597

Stolt, S., Haataja, L., Lapinleimu, H., & Lehtonen, L. (2009). Associations between lexicon and grammar at the end of the second year in Finnish children. *Journal of Child Language*, 36(4), 779–806. https://doi.org/10.1017/S0305000908009161

Su, Y. (Esther), & Naigles, L. R. (2022). Comprehension of core grammar in diverse samples of Mandarin-acquiring preschool children with ASD. *Evolutionary Linguistic Theory*, 4(1), 52–101. https://doi.org/10.1075/elt.00038.su

Su, Y., Naigles, L. R., & Su, L.-Y. (2018). Uneven Expressive Language Development in Mandarin.pdf. *Journal of Autism and Developmental Disorders*, 48, 3432–3448. https://doi.org/10.1007/s10803-018-3614-x

Sun, W., Chen, Y., Wan, X., & Liu, M. (2019). Parsing Chinese Sentences with Grammatical Relations. *Computational Linguistics*, 45(1), 95–136. https://doi.org/10.1162/coli_a_00343

Swensen, L. D., Kelley, E., Fein, D., & Naigles, L. R. (2007). Processes of Language Acquisition in Children With Autism: Evidence from Preferential Looking. *Child Development*, 78(2), 542–557. https://doi.org/10.1111/j.1467-8624.2007.01022.x

Tager-Flusberg, H., & Anderson, M. (1991). The Development of Contingent Discourse Ability in Autistic Children. *Journal of Child Psychology and Psychiatry*, 32(7), 1123–1134. https://doi.org/10.1111/j.1469-7610.1991.tb00353.x

Tager-Flusberg, H., Calkins, S., Nolin, T., Baumberger, T., Anderson, M., & Chadwick-Dias, A. (1990). A longitudinal study of language acquisition in autistic and down syndrome

765 children. *Journal of Autism and Developmental Disorders*, 20(1), 1–21.
766 https://doi.org/10.1007/BF02206853
767 Tardif, T. (1996). Nouns Are Not Always Learned Before Verbs: Evidence From Mandarin
768 Speakers' Early Vocabularies. *Developmental Psychology*, 32(3), 492–504.
769 Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus
770 verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, 24(3),
771 535–565. https://doi.org/10.1017/S030500099700319X
772 Vehkavuori, S.-M., Kamarainen, M., & Stolt, S. (2021). Early receptive and expressive lexicons
773 and language and pre-literacy skills at 5;0 years ? A longitudinal study. *Early Human*
774 *Development*, 156, 105345. https://doi.org/10.1016/j.earlhumdev.2021.105345
775 Villena, O. S. (2002). Materials for an introduction to language and linguistics. *Estudios*
776 *Filológicos*, 37. https://doi.org/10.4067/S0071-17132002003700022
777 Weismer, S. E., Lord, C., & Esler, A. (2010). Early language patterns of toddlers on the autism
778 spectrum compared to toddlers with developmental delay. *Journal of Autism and*
779 *Developmental Disorders*, 40(10),1259–1273. https://doi.org/10.1007/s10803-010-0983-1
780 Wittke, K., Mastergeorge, A. M., Ozonoff, S., Rogers, S. J., & Naigles, L. R. (2017). Grammatical
781 Language Impairment in Autism Spectrum Disorder: Exploring Language Phenotypes
782 Beyond Standardized Testing. *Frontiers in Psychology*, 8.
783 https://doi.org/10.3389/fpsyg.2017.00532
784 Wu, S.-Y. (2020). Mean length of utterance among Mandarin-speaking children with and without
785 DLD. *Child Language Teaching and Therapy*, 36(3), 165–179.
786 https://doi.org/10.1177/0265659020945366

Xia, F. (2000). The Part-Of-Speech Tagging Guidelines for the Penn Chinese Treebank (3.0) (No. 38; *IRCS Technical Reports Series*, p. 46).

Xu, Y. A., Naigles, L. R., & Su, Y. E. (2022). Early Word Order Usage in Preschool Mandarin-Speaking Typical Children and Children with Autism Spectrum Disorder: Influences of Caregiver Input? *Frontiers in Psychology*, 12, 766133. https://doi.org/10.3389/fpsyg.2021.766133

Xue, N., Xia, F., Huang, S., & Kroch, A. (2000). The Bracketing Guidelines for the Penn Chinese Treebank (3.0) (No. 39; IRCS Technical Reports Series, p. 193).

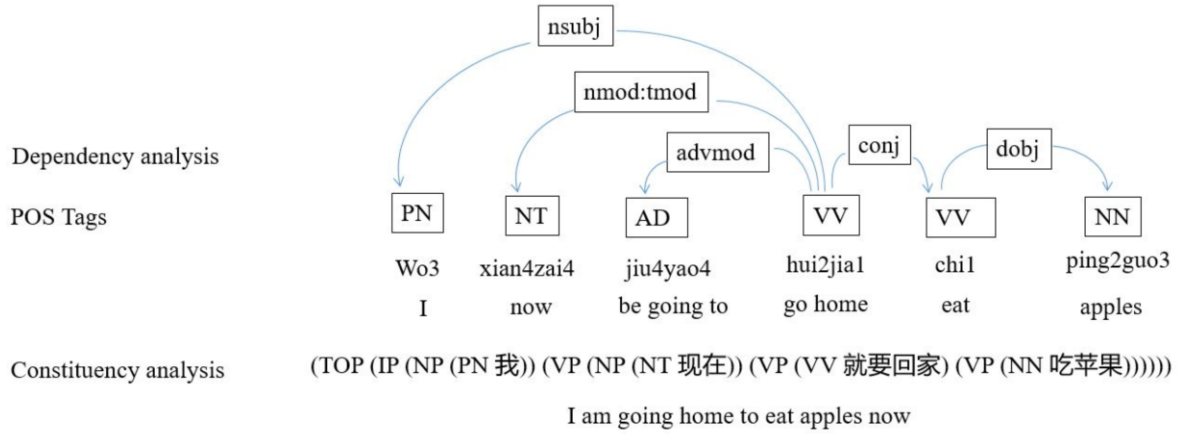
Yu, L., Zhu, X., Shek, D. T. L., Zou, X.-B., Deng, H.-Z., & Au Yeung, P. W. H. (2019). Validation of the Simplified Chinese Psychoeducational Profile Third Edition in Mainland China. *Journal of Autism and Developmental Disorders*, 49(4), 1599–1612. https://doi.org/10.1007/s10803-018-3827-z

Zaccarella, E., Schell, M., & Friederici, A. D. (2017). Reviewing the functional basis of the syntactic Merge mechanism for language: A coordinate-based activation likelihood estimation meta-analysis. *Neuroscience & Biobehavioral Reviews*, 80, 646–656. https://doi.org/10.1016/j.neubiorev.2017.06.011

Zhou, J., & Zhang, Y. (2020). Using the corpus-based language measuring system in Chine: a diagnostic assessment on language development of autistic children. *The Study of Chinese Characters*, 31: 222-239.

Appendices

Appendix 1. Syntactic tree from dependency and constituency analysis



(https://hanlp.hankcs.com/demos/con.html)

835 Appendix 2. POS tags for lexical categories

Lexical categories	Part-of-speech	Tags	Examples (Translation)
CWs			
noun	common nouns	NN	shu1 (book)
	proper nouns	NR	Bei2jing1 (Beijing)
	temporal nouns	NT	yi1yue4 (January)
	localizer	LC	qian2hou4 (around)
verb	copula	VC	shi4 (be)
	you3 as the main verb	VE	you3 (have)
	other verbs	VV	zou3 (walk)
adverb	adverbs	AD	reng2ran2 (still)
adjective	predicative adjectives	VA	xue3bai2 (snow white)
FWs			
pronoun	pronouns	PN	ni3 (you)
preposition	other prepositions	P	cong2(from)
	bei4 in short bei-construction	SB	bei4 xun4 le5 (bei/LB [scold] le/AS.) (be scolded)
	bei4 in long bei-construction	LB	ta1 bei4 wo3 xun4 le5. ([he] bei/LB [I] [scold] le/AS.) (He was scolded by me.)
	ba3 in ba-construction	BA	Ta1 ba3 ni3 pian4 le5. ([he] ba/BA [you][cheat] le/AS.) (He cheated you.)
conjunction	coordinating conjunction	CC	he2 (and)
	subordinating conjunction	CS	sui1ran2 (although)
numeral	cardinal number	CD	hao3xie1 (some)
	ordinal number	OD	di4yi1 (the first)
classifier	measure words	M	yi4 xiang1 shu1 ([one][box]/MW [books]) (one box of books)
particle	sentence-final particles	SP	Ta1 hao3 ba5? ([he][OK] ba/SP?) (Is he OK?)
	aspect markers	AS	zhe5/AS
	de5 for relative-clause etc., as a complementizer or nominalizers	DEC	Wo3 mai3 de5 shu1. ([I][buy] de/DEC [book].) (the book that I bought)
	associative de5, as a genitive markers and an associative markers	DEG	wo3 de5 shu1 ([I] de/DEG [book].) (my book)
	de5 in V-de constructions, and V-de-R, resultative de5	DER	pao2 de5 kuai4 ([run] de/DER [fast]) (run fast)
	de5 as the head of DVP, manner de5	DEV	man4 man4 de5 shuo1 ([slowly] de/DEV [speak]) (speak slowly)
	some particles	MSP	ta1 suo3 xu1yao4 de5 ([he] MSP [need]) (what he needs)
interjection	interjections	IJ	a(Ah)
onomatopoeia	onomatopoeias	ON	yu3 hua1hua5 de5 xia4 ([rain] huahua/ONO de/DEV[fall down].) (The rain pours down.)

836

Appendix 3. Tags for categories of syntactic structures

Tags	Syntactic structures	Examples (translation)
NP	noun phrase	zhe4 zhi1 mao1 (this cat)
VP	verb phrase	xi3 huan1 ma1ma5 (like Mom)
ADJP	adjective phrase	ke3 ai4 de5 mao1 (lovely cat)
ADVP	adverbial phrase headed by AD(adverb)	fei1 kuai4 de5 pao2 (run quickly)
CLP	classifier phrase	yi2 da4 xiang1 (a big box of)
CP	clause headed by C (complementizer)	wo3men5 yong4 guo4 de5 bei1zi5 (the cup we have used)
DNP	phrase formed by "XP+DEG"	he1 shui3 de5 bei1 zi5 (the cup that used for drinking water)
PP	preposition phrase	zai4 chuang2 xia4mian4 (under the bed)
QP	quantifier phrase	xu3duo1 ren2 (many people)
IP	simple clause headed by I(INFL)	jin3guan3 kan4zhe5 xiang4 zhe4yang4 (although it looks like this)

Appendix 4. Tags for categories of grammatical relations

Tags	Short description	Examples (translation)	Typed dependency
top	topic	jian4zhu4 shi4 zhu3yao4 huo2dong4 (construction is the main activity)	top (shi4, jian4zhu4)
nsubj	nominal subject	jia1 li3 you3 yi2 ge4 (there is one in my house)	nsubj (you3, jia1li3）
ccomp	clausal complement	wo3 xiang3 xian1 wan2 (I want to play it first)	ccomp (xiang3, wan2）
advmod	adverbial modifier	xiao2gou3 xian1 hui2jia1 (the puppy go home first)	advmod (hui2jia1, xian1）
rcomp	resultative complement	shou3 fang4 hao3 (put the hand in appropriate position)	rcomp (fang4, hao3）
dobj	direct object	liu4 xiao2gou3 (walk the dog)	dobj (liu4, xiao2gou3）
pobj	prepositional object	zai4 ce4suo3 (in the washroom)	pobj (zai4, ce4suo3）
nn	noun compound modifier	fu2wu4 zhong1xin1 (service center)	nn (zhong1xin1, fu2wu4)
clf	classifier modifier	shi2 ge4 ping2guo3 (ten apples)	clf (ping2guo3, ge4)
nummod	number modifier	shi2 ge4 ping2guo3 (ten apples)	nummod (ge4, shi2)
conj	links two conjuncts	jian3dao1 he2 bei1zi5 (scissors and cups)	conj (jian3dao1, bei1zi5)
prep	prepositional modifier	zai4 jia1li3 ti1 (kick it at home)	prep (ti1, zai4）
amod	adjectival modifier	kua4 shi4ji4 gong1cheng2 (cross-century projects)	amod (gong1cheng2, kua4 shi4ji4)
cpm	complementier	kai1fa1 Bei2jing1 de5 huo2dong4 (the activity to explore Beijing)	cpm (kai1fa1, de5)
assm	associative marker	wo3 de5 shu1 (my book)	assm (wo3, shu1)
asp	aspect marker	chi1 le5 ping2guo3 (ate apples)	asp (chi1, le5)
attr	attributive	zong3e2 wei2 er4shi2 mei3yuan2 (the total amount is twenty dollars)	attr (wei2, mei3yuan2)

Appendix 5. Canonical correlation analysis of lexical and grammatical development