Brain & Language 266 (2025) 105571

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Semantic overreliance as a suboptimal compensation for syntactic impairments in children with Developmental Language Disorder☆

Jueyao Lin [a,b], Xiaocong Chen [a,b], Xunan Huang [c], Patrick Chun Man Wong [d], Angel Wing Shan Chan [b,e], Michael T. Ullman [f], Caicai Zhang [a,b,e,*]

[a] Neurocognition of Language, Music and Learning (NLML) Lab, Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Hong Kong, China
[b] Research Centre for Language, Cognition and Neuroscience, Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Hong Kong, China
[c] School of Life Science and Technology, University of Electronic Science and Technology of China, Sichuan Province, China
[d] Department of Linguistics and Modern Languages, Chinese University of Hong Kong, Hong Kong, China
[e] The Hong Kong Polytechnic University - Peking University Research Centre On Chinese Linguistics, Hong Kong, China
[f] Department of Neuroscience, Georgetown University, USA

ARTICLE INFO

Keywords:
Developmental Language Disorder
N400
P600
N400-P600 tradeoff
Response dominance
Classifier-noun agreement

ABSTRACT

The neurocognitive dynamics of semantic-syntactic interplay are not well understood in children with and without Developmental Language Disorder (DLD). This study examined the N400, P600 and their interplay in Cantonese-speaking children with DLD and age-matched typically developing (TD) children, by manipulating semantic and syntactic violations in Chinese classifier-noun agreement. Behaviorally, children with DLD demonstrated overall lower accuracy in grammaticality judgment. The N400 and P600 analyses respectively confirmed robust semantic processing but attenuated syntactic processing in the DLD group. Crucially, the N400-P600 interplay analyses revealed that TD children prioritized syntactic processing over semantic processing for outright syntactic violations, as indicated by less N400-P600 dependence and robust P600 dominance, whereas children with DLD relied on semantic processing and showed reduced P600 dominance. These results underscore a challenge to prioritize syntactic processing and (suboptimal) compensatory reliance on semantic processing in children with DLD, compatible with the predictions of the Procedural circuit Deficit Hypothesis.

1. Introduction

The acquisition of language is pivotal in childhood development. In this context, Developmental Language Disorder (DLD)—previously known as Specific Language Impairment (SLI)—emerges as a significant concern (Bishop et al., 2017). DLD is a neurodevelopmental disorder which entails difficulties in understanding, speaking, and learning the intricacies of language; it impacts approximately 5–11 % of children globally (Leonard, 2014; McGregor, 2020; T'sou et al., 2006; Wu et al., 2023). Notably, this disorder is not attributable to hearing impairment, nonverbal intellectual disabilities, or other overt biomedical conditions (Bishop et al., 2017; Leonard, 2014). Children with DLD exhibit significant deficits in a broad range of language areas, among which morphosyntax is one of the significantly affected areas (Leonard, 2014; Sheng et al., 2023). For example, English-speaking children with DLD commonly struggle with verb tense and agreement (Conti-Ramsden & Durkin, 2012; Haebig et al., 2017; Leonard, 2014; Purdy et al., 2014). Beyond linguistic deficits, non-linguistic deficits in domain-general procedural memory, working memory, rapid temporal processing, speed of processing, and processing limitations are also observed in children with DLD (Leonard, 2014).

However, the neurocognitive substrates of this common childhood language disorder remain controversial, particularly in real-time language processing. Real-time processing of naturally spoken sentences is closely linked to everyday functioning and offers an important testing ground for current theories on the neurocognitive mechanisms of DLD. Real-time language processing can reveal specific areas of difficulty, such as syntactic, semantics, and their real-time interplay in children with DLD, as compared to children with typical development (TD). Several studies have employed electroencephalography (EEG), which

☆ This article is part of a special issue entitled: 'DLD in Chinese' published in Brain and Language.

* Room HHB710, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University (Hung Hom Bay Campus), Hung Hom, Hong Kong Special Administrative Region.

E-mail address: caicai.zhang@polyu.edu.hk (C. Zhang).

https://doi.org/10.1016/j.bandl.2025.105571
Received 22 May 2024; Received in revised form 20 March 2025; Accepted 21 March 2025
Available online 29 March 2025

has excellent temporal resolution, to investigate semantic violation processing (e.g., N400) and syntactic violation processing (e.g., Early Left-Anterior Negativity/ELAN or P600) in children with DLD (Fonteneau & van der Lely, 2008; Haebig et al., 2017; Purdy et al., 2014; Weber-Fox et al., 2010). These studies have generated important insights into the underlying neural deficits of DLD. However, these studies often examined semantic and syntactic processing separately, and little attention has been paid to the neurocognitive dynamics of semantic-syntactic interplay, which is the focus of the current study.

1.1. ERP studies on semantic and syntactic processing in individuals with DLD

Several ERP studies have examined syntactic processing during sentence judgment tasks in children or adolescents with DLD and reported syntactic processing deficits (Courteau et al., 2023; Fonteneau & van der Lely, 2008; Haebig et al., 2017; Purdy et al., 2014; Weber-Fox et al., 2010). In comparison to their TD counterparts, adolescents and children with DLD demonstrated diminished P600 effects in processing third-person singular tense agreement (Haebig et al., 2017; Weber-Fox et al., 2010), regular and irregular number agreement (Courteau et al., 2023), and long-distance finiteness errors (Purdy et al., 2014), highlighting a consistent pattern of syntactic processing challenges in the DLD population. While significant group differences in the Anterior Negativity were generally not observed (Purdy et al., 2014; Weber-Fox et al., 2010), an exception is noted in a specific subgroup with grammatical impairments (i.e., G-SLI), where an Early Left Anterior Negativity (ELAN) anomaly was reported in response to syntactic violations in question sentences (Fonteneau & van der Lely, 2008). These findings provided support for syntactic deficits in DLD.

In contrast to the commonly observed syntactic deficits marked by attenuated P600, results of semantic processing abilities presented interesting variations. These studies investigated N400 responses in children with DLD primarily by tasks featuring semantic incongruity. Some of them suggested that semantic processing is relatively intact in individuals with DLD, as indicated by their comparable N400 responses to matched controls with TD in these studies (Courteau et al., 2023; Fonteneau & van der Lely, 2008; Haebig et al., 2017; Weber-Fox et al., 2010). Interestingly, there are also studies reporting atypical N400 patterns in the DLD population, such as enhanced N400 effect for incongruous items (Neville et al., 1993) or the lack of N400 effect due to elevated N400 responses in the congruous conditions (Sabisch et al., 2006). Furthermore, Popescu et al. (2009) found that children with DLD showed no N400 differences between incorrect and correct conditions before the intervention, but after intervention, they showed reduced N400 magnitudes in the correct baseline condition but not in the incorrect condition, leading to a stronger N400 effect in turn. Helenius et al. (2009) found that adults with DLD history showed comparable N400 responses to the control group in the baseline, but their N400 effects were not modulated by repetition (the first versus the second time of the word repetition) or lexicality (word versus nonword). Altogether, these findings indicate that atypical N400 in DLD, if any, tends to manifest in a form of over activation or inflexible activation (e.g., lack of modulation by other factors). These patterns of anomaly stand in contrast to that of the P600, where atypical P600 in DLD is often presented as attenuation.

It is worth noting that the previous studies on DLD typically analyzed the N400 and P600 components separately and solely focused on one of the functions, either semantic integration or syntactic reanalysis, to correspondingly probe the specific neural deficits in DLD. However, the involvement of these functions during language processing might not be as clear-cut as previously thought. For example, the P600 may be elicited with some semantic violations (Kim & Osterhout, 2005; Nakano et al., 2010), while the N400 may be elicited with syntactic violations (e.g., in second language (L2) learners) (Tanner et al., 2013).

Importantly, recent evidence indicates a real-time interplay between the N400 and P600 during online language comprehension, which likely reflects the brain's allocation of cognitive resources between semantic and syntactic processing (Coderre & Cohn, 2023; Kim et al., 2018; Tanner et al., 2013). Across individuals, a negative correlation between the N400 and P600 responses (i.e., a tradeoff) has been reported, where larger magnitudes of N400 effects were associated with smaller magnitudes of P600 effects (Kim et al., 2018; Tanner et al., 2013). Furthermore, by comparing the magnitude of N400 and P600 effects within an individual, differences in the relative response dominance between the two components have been noted, ranging from N400-dominant to P600-dominant patterns, which likely reflect automatic prioritization of semantic or syntactic processing (Coderre & Cohn, 2023; Delogu et al., 2019; Kim et al., 2018; Tanner et al., 2013). Such differences in response dominance can be at least partly explained by individual differences in behavioral accuracy and working memory capacities. For example, among L2 speakers, early-stage L2 learners with lower accuracy in language tasks exhibited a more pronounced N400 effect coupled with a less pronounced P600 effect in response to morphosyntactic violations (Tanner et al., 2013). In contrast, experienced L2 speakers with higher accuracy tended to show a P600-dominant trend, indicative of stronger syntactic processing and structural repair (Hahne & Friederici, 2001; Tanner et al., 2013). This discrepancy could suggest an over-reliance on semantic processing over syntactic processing in early-stage L2 learners. Even among native speakers, individual differences have been noted. Kim et al. (2018) showed that some individuals showed more P600-dominant effects to semantic anomaly, whereas others demonstrated N400-dominant effects to the same stimuli, whereby the degree of P600 dominance can be explained by larger verbal working memory capacities. Delogu et al. (2019) found that the most accurate subjects demonstrated reliable P600 effects compared to the baseline condition, while the least accurate subjects had no P600 effects.

Taking inspiration from the above work on real-time interplay between the N400 and P600 in neurotypical individuals who are first language (L1) and L2 learners, we aim to investigate the neurocognitive dynamics of semantic-syntactic interplay in children with DLD, apart from potential group differences in semantic or syntactic processing per se. This approach serves a dual purpose: it fills a critical empirical gap to identify the N400-P600 interplay in children with and without DLD, and it enables us to examine potential compensatory mechanisms in children with DLD.

1.2. The declarative/procedural model and procedural circuit deficit hypothesis

The Declarative/Procedural (DP) model (Ullman, 2016) and the Procedural circuit Deficit Hypothesis (PDH) of DLD (Ullman et al., 2020; Ullman & Pierpont, 2005) offer an appropriate framework to study and interpret the N400 and P600 responses and their interplay in children with and without DLD. These models generate theoretical predictions that align with the observed N400-P600 tradeoff, providing insights into potential compensatory mechanisms in children with DLD. The DP model assumes that language processing is subserved by two competitive long-term memory systems: declarative memory and procedural memory. In typical development, these two systems have specialized, though partly overlapping, roles in language: declarative memory primarily supports semantic processing (indexed by N400), while procedural memory primarily supports syntactic processing (indexed by ELAN or P600). This dynamic, competitive relationship between declarative memory and procedural memory—often described as a 'see-saw' effect, where increased reliance on one system may reduce engagement of the other system—might partly explain the N400-P600 interplay observed in the aforementioned studies of real-time language processing in L2 and L1 neurotypical individuals (Coderre & Cohn, 2023; Kim et al., 2018; Tanner et al., 2013).

Extending this framework to atypical development, the PDH applies the DP model to characterize linguistic and non-linguistic deficits in children with DLD (Ullman et al., 2020; Ullman & Pierpont, 2005). The PDH posits that DLD is primarily associated with deficits in procedural memory, while declarative memory remains relatively intact. Given the competitive relationship between the two memory systems, the PDH suggests that children with DLD may rely more heavily on declarative memory as a compensatory mechanism for procedural memory deficits. In the context of real-time language processing, this implies that children with DLD may exhibit an atypical pattern of N400-P600 tradeoff, with a relatively comparable N400, attenuated P600, and an over-reliance on N400 during syntactic processing. Therefore, the DP model and PDH framework could be used to predict not only deficits in single ERP components related to semantic and syntactic processing (e.g., N400 and P600, respectively), but also the typical and atypical interplay between these ERP components.

In light of this theoretical framework, the current study investigated the neurocognitive dynamics of semantic-syntactic interplay in Chinese children with DLD compared to their TD peers. Moving beyond the traditional approach of examining N400 and P600 in isolation, this study adopts a comprehensive approach to analyze the real-time interaction between N400 and P600. By focusing on the N400-P600 interplay, we aim to deepen our understanding of DLD, providing new insights into individual differences in language processing propensities and compensatory mechanisms across neurotypical and language-impaired populations.

1.3. *The current study*

As the majority of existing DLD studies are centered on English and other Indo-European languages, Chinese-speaking children with DLD are less well-studied (Sheng et al., 2023). In particular, there is a gap in studies on real-time language processing. Therefore, we aim to examine real-time semantic processing, syntactic processing and their interplay, via classifier-noun agreement, in an ERP study on Cantonese children with DLD in the current study.

Chinese classifier is at the interface of syntactic and semantic processing. Syntactically, classifier is mandatory in a noun phrase when a numeral or demonstrative precedes a noun, linking the numeral or demonstrative to the noun (Matthews & Yip, 2011). For example, *jat1 bun2 siu2 syut3*[1] means 'one classifier$_{\text{bun2}}$ novel.' The current study mainly focuses on sortal classifiers, which is a type of classifiers that characterizes semantic features, such as shape, function and animacy, of the noun. The classifier-noun agreement is largely governed by semantic features. For example, *tiu4* is a shape classifier that denotes long and flexible shape of the noun (e.g., *jat1 tiu4 jyu2* 'one classifier$_{\text{tiu4}}$ fish'), and it would be inappropriate to say *jat1 tiu4 siu2 syut3* '*one classifier$_{\text{tiu4}}$ novel.' Nonetheless, in some cases, a noun could be associated with different classifiers, each emphasizing different features of the noun (e.g., *jat1 tiu4 jyu2* and *jat1 zek3 jyu2*, where *tiu4* emphasizes the shape of the fish, while *zek3* emphasizes animacy). It is worth mentioning that *go3* is a default classifier, which is often used when the specific classifier of a noun is unknown to the speaker, or when the noun lacks a specific classifier (Chan, 2019). *Go3* is less semantically specified and mainly used to quantify the number of objects (e.g., *jat1 go3 jan4* 'one classifier$_{\text{go3}}$ person', *jat1 go3 caang2* 'one classifier$_{\text{go3}}$ orange'). It can also be used to quantify abstract concepts (e.g., *jat1 go3 nam2 faat3* 'one classifier$_{\text{go3}}$ idea') or used in informal speech where specificity is less critical.

Several studies have examined the acquisition of classifiers in Cantonese children, which appears to reveal a complex interplay of the development of syntactic and semantic knowledge (Mak, 1991; Stokes & So, 1997). For example, Mak (1991) divided 4- to 8-year-old children into five groups based on their accuracy rate in an elicitation task of Cantonese sortal classifiers. The author found that the rate of omissions was low in groups 1–3, which may suggest that early in the development, children have acquired syntactic knowledge that classifier is mandatory in a noun phrase. Intriguingly, the rate of omission increased in group 4, before dropping again in group 5. Stokes and So (1997) further proposed a stage model of classifier acquisition based on the results of Mak (1991). In an early stage (21–50 % accuracy), the children mainly used a "fill space" syntactic strategy. In the intermediate stage (51–80 % accuracy), a semantic strategy appeared to dominate over the syntactic strategy, where the children would rather omit the classifier than use a wrong or default classifier, leading a temporary increase of omission errors. In the advanced stage (>80 % accuracy), the syntactic strategy prevailed again, and omission errors became rare. Children would use the default classifier when a specific classifier is unknown, like typical adults. The researchers also reported that Cantonese children with DLD exhibited a delay in acquiring classifiers – they were at earlier stages with lower accuracy where either a syntactic strategy was rigorously applied (e.g., incorrect classifiers are applied to fill the space), or the semantic strategy started to take over (e.g., omission errors as a result of avoidance of over-generalization). In a recent study (Cheung, 2008), it reported similar findings that Chinese children with DLD used a narrower range of specific classifiers in spontaneous speech contexts and might require stronger contextual cues to learn and apply specific classifiers correctly compared to younger and language-matched TD children.

ERP studies on neurotypical adults confirmed the engagement of both semantic and syntactic processing in classifier-noun agreement. When the noun is combined with an incorrect classifier, which is semantically incongruent with the noun (e.g., **jat1 bun2 toi2* '*one classifier$_{\text{bun2}}$ table'), the N400 effect is observed (Chan, 2019; Zhang et al., 2012). When the classifier is omitted (e.g., one ∅ table), which is an apparent syntactic violation, the P600 effect tends to be observed (Chan, 2019; Qian & Garnsey, 2016). However, as far as we are aware, no studies have directly examined the dynamic interplay between semantic and syntactic processing (i.e., the N400-P600 tradeoff and response dominance) reflecting automatic processing prioritization and compared the differences between children with DLD and TD. The unique nature of Chinese classifiers, situated at the interface of semantic features and syntactic rules, makes them particularly relevant for investigating the N400-P600 interplay in children with DLD.

To address the above issues, the current study employed a combined behavioral and ERP approach to investigate semantic processing (N400 effect), syntactic processing (P600 effect) and their interplay (see details below), as they subserve classifier-noun agreement processing, in a group of Cantonese-speaking school-age children with DLD and another group of age-matched children with TD. Four types of classifier-noun relationship in spoken sentences were manipulated: (1) correct classifier, (2) incorrect classifier, (3) omission of classifiers, and (4) default classifier (i.e., *go3*). Sentences in the correct condition contained a specific classifier which was *semantically congruent* with the ensuing noun and served as the baseline. Sentences in the incorrect condition contained a specific classifier which was *semantically incongruent* with the accompanied noun. Sentences in the omission condition had no classifier, which was an outright syntactic violation. Sentences in the default classifier condition contained a default classifier *go3* in the place of a correct, specific classifier, leading to a mild semantic anomaly. The children were asked to perform a grammaticality judgment task on these sentences.

As informed by previous studies (Chan, 2019; Frankowsky et al., 2022; Qian & Garnsey, 2016), in children with TD, we predicted that the incorrect condition would mainly elicit a N400 effect relative to the baseline correct condition; the omission condition would mainly elicit a P600 effect; the default condition with *go3* might mainly elicit a mild N400. Note that we do not expect each condition to only elicit a N400 or P600 effect. Indeed, both N400 and P600 are likely to be elicited in a

[1] The pronunciation is denoted with *Jyutping*, which is a romanization system for Cantonese developed by the Linguistic Society of Hong Kong in 1993.

condition, but their relative dominance is likely to be modulated by the severity of semantic and syntactic violations across conditions (Grey et al., 2017; Kim et al., 2018; Tanner et al., 2013). For children with DLD, we predicted that they would show comparable semantic processing (i.e., comparable N400) to children with TD, but they would have deficits in syntactic processing (i.e., attenuated P600), based on previous reports of syntactic impairments in DLD (Courteau et al., 2023; Fonteneau & van der Lely, 2008; Haebig et al., 2017; Purdy et al., 2014; Weber-Fox et al., 2010).

Importantly, we also analyzed the N400-P600 interplay using two measures to capture different aspects of the interplay and compared them across the conditions and the two groups. The first measure was the N400-P600 tradeoff. Previous studies reported the tradeoff as a negative correlation between the magnitude of N400 and P600 effects (Kim et al., 2018; Tanner et al., 2013); accordingly, the current study quantified the tradeoff as the regression slope between the magnitudes of the N400 and P600 effects across individuals of a group under a specific condition (for details, see Section 2.5). Given that we examined a tradeoff, a negative slope was anticipated, and a more negative slope meant a sharper decrease in the magnitude of P600 effects predicted by an increase in the magnitude of N400 effects. Based on the DP model and the findings from neurotypical adults (Chan, 2019; Frankowsky et al., 2022; Kim et al., 2018; Qian & Garnsey, 2016), we predicted that at least in children with TD, they would show dynamic adaptation to different features of violations across the conditions (i.e., the more semantically dominant incorrect condition vs. the more syntactically dominant omission condition), thereby exhibiting different degrees of N400-P600 tradeoff slopes. For example, in the omission condition with outright syntactic violations (where syntactic processing should be prioritized over semantic processing), less tight N400-P600 coupling would be expected: a strong P600 effect would be expected relatively independently from the magnitude of the preceding N400 effect. We expected the patterns to deviate in children with DLD, especially in the omission condition. The PDH postulates that children with DLD may rely on semantic processing as a compensation for the syntactic challenges. Thus, we expected that, compared to TD children, children with DLD would show a relatively stronger N400-P600 tradeoff (i.e., steeper slopes) in the outright syntactic violation (i.e., the omission condition).

The second measure was the relative dominance of P600 over N400 effects, which was calculated using the response dominance index (RDI, for calculation details see *2.5 Statistical analysis*; Grey et al., 2017). While the tradeoff slopes quantified the degree of correlation between the N400 and P600 effects across individuals of a group, the RDI probed the primary processing employed in comprehension, whether semantic or syntactic, based on the relative strength of N400 and P600 effects within an individual. Therefore, analyzing both the tradeoff slope and RDI was crucial for a holistic understanding of the N400-P600 interplay, allowing us to discern the balance between semantic and syntactic processing and identify the dominant processing under various linguistic conditions. Based on previous studies that reported syntactic impairments and intact semantic processing in DLD (Courteau et al., 2023; Fonteneau & van der Lely, 2008; Haebig et al., 2017; Purdy et al., 2014; Weber-Fox et al., 2010) together with the PDH framework, we predicted that children with DLD might have lower RDI, i.e., being less P600-dominant and more N400-dominant, especially in the omission condition. Compared to the slope measure, the RDI measure served as a more direct index to quantify individual differences in the relative dominance of semantic or syntactic processing and identify the compensatory semantic overreliance in the DLD group by using the TD group as a baseline. We further explored the relationship between the P600-dominance (RDI) and individual differences in grammatical scores from a standardized language assessment, to probe whether the RDI may serve as a general neural marker of the children's grammatical abilities. Collectively, based on the PDH framework, we hypothesized an overreliance on semantic processing in children with DLD, which potentially compensates for their syntactic deficits.

2. Methods

2.1. Participants

Twenty[2] school-aged Chinese children with DLD (age range: 83–132 months, mean = 103.35, SD = 13.41; 8 girls) and twenty age-matched Chinese TD children (age range: 80–135 months, mean = 102.5, SD = 16.34; 10 girls) participated in this study. All the participants were native Cantonese speakers with normal non-verbal intelligence, normal hearing, and no other known neurodevelopmental disorders such as autism spectrum disorder. One DLD child with Attention Deficit and/ Hyperactivity Disorder (ADHD) was included,[3] following the CATALISE diagnostic criteria that recognizes that DLD may co-occur with ADHD (Bishop et al., 2017). All the participants were assessed with the Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS) (T'sou et al., 2006), which is a standardized norm-referenced spoken language assessment tool. HKCOLAS, combined with the exclusion of biomedical conditions, is used for diagnosing DLD in Cantonese-speaking children in Hong Kong. It consists of six subtests (Grammar, Textual Comprehension, Word Definition, Lexical-Semantic Relations, Narrative, and Expressive Nominal Vocabulary). Children scoring 1.25 SD below age means in at least two or more subtests were included in the DLD group, as per the standard diagnostic criteria. Children with TD scored above −1.25 SD across all the subtests, except one child who scored −1.3 SD in textual comprehension.[4] The demographic details can be seen in Table 1. All the participants and their caregivers provided informed written consent in compliance with the experimental protocol approved by the Human Subjects Ethics Sub-committee of The Hong Kong Polytechnic University. They were provided a small monetary compensation

Table 1
Participant characteristics.

	Children with TD n = 20 (Female = 10)		Children with DLD n = 20 (Female = 8)	
	Mean	SD	Mean	SD
Chronological age in months	102.50	16.34	103.35	13.41
HKCOLAS standard scores[1]				
Grammar	0.77	0.75	−0.73	1.20
Textual comprehension	0.35	0.98	−0.79	1.23
Word definition	0.80	0.93	−0.59	1.20
Lexical-semantic relations	0.04	0.78	−1.55	0.62
Narrative	0.24	0.91	−1.81	1.17
Expressive nominal vocabulary	0.7	0.77	−1.52	1.22

Note [1]*:* Standard scores are age-normed scores that adjust for age-related language development in each subtest.

[2] We conducted *post hoc* power analysis on the key ERP effects. Given the known limitations of traditional *post hoc* power analysis, we adopted a simulation-based approach with 1000 iterations to provide a more robust estimate of the observed power, using the R package *simr* (Green & MacLeod, 2016). The results indicated 81.5% power for the N400 effect, 100% power for the P600 effect, and 99.1% power for the N400-P600 tradeoff. Also note that the sample size of this study (*N*=20 for each group) is compatible with prior EEG studies investigating the N400 and P600 components in DLD, which typically ranged from 15 to 20 participants per group (Courteau et al., 2023; Haebig et al., 2017; Purdy et al., 2014; Sabisch et al., 2006; Sabisch et al., 2009; Weber-Fox et al., 2010).

[3] This DLD child was also the only participant who reported to take medication during the time of the experiment. Exclusion of the data from this participant did not change the overall patterns of the results presented in the paper. Thus, we kept the data of this participant.

[4] This TD child's standard scores in the other subtests are: 1.4 (grammar), −0.1 (word definition), −0.9 (lexical-semantic relations), 0.2 (narrative), and 0.1 (expressive nominal vocabulary) respectively.

for their participation in the study. The participants were recruited via poster promotion and referral from the Speech Therapy Unit at the Hong Kong Polytechnic University.

2.2. Stimuli

The stimuli were naturally spoken short sentences with six to seven syllables. Each of the four conditions consisted of 60 sentences, except for the default classifier condition (i.e., 58 sentences). Note that a "correct" response was expected for each of the 60 sentences in the correct condition, whereas "incorrect" responses were expected for sentences in the other conditions. After a careful reevaluation of sentence acceptance, two out of 60 sentences in the default classifier condition were deemed to be acceptable and therefore excluded from the analysis, thus resulting in 60 "correct" and 178 "incorrect" sentences, or 238 target sentences in total. Accordingly, 122 filler sentences, including 120 "correct" sentences and two "incorrect" sentences, were added to balance the proportion of grammatically correct and incorrect sentences. Each sentence had seven syllables, except for the 60 sentences in the omission condition, which each had six syllables due to the omission of the classifier. Accordingly, all the 7-syllable sentences were normalized to 1.7 s, and all the 6-syllable sentences were normalized to 1.4 s. All the sentences were simple SVO sentences, such as *go4 go1/gin3 dou2/loeng5/zo6/saan1* "the elder brother/saw/two/classifier$_{zo6}$/ mountain" (see Table 2 for examples and supplementary materials for the full list of sentences). The words in the sentences are early developing and highly familiar to children. The average length of the object nouns was 491 ms. Note that across the four conditions, the numeral and the noun were fixed, while the classifier-noun agreement was manipulated. All the sentences were recorded from a Cantonese native male speaker in a quiet room sampled at the rate of 44,100 Hz with 16 bits per sample.

2.3. Procedure

Children were individually tested in a sound-attenuated chamber for EEG recording. Seated comfortably, they listened to sentences through earphones and performed a grammaticality judgment task. During a trial, the spoken sentence was presented binaurally, and the children were instructed to judge the well-formedness of the sentence only after hearing the whole sentence. They were asked to click the left button of the mouse if they considered the sentence to be grammatical, and to click the right button if ungrammatical. There was no time limit for the response and the next trial began only after participants clicked the mouse button. There were 360 trials in total, which were divided into four blocks with equal proportion of sentences from each condition. The trials in each block were presented in a randomized order. Children were given sufficient time to rest between blocks.

Table 2
Stimulus examples.

Conditions	Example sentences
Correct classifier(baseline)	go4go1/gin3dou2/loeng5/**zo6**/saan1 哥哥/見到/兩/座/山 The elder brother /saw/ two/ **CL**/ mountains
Incorrect(semantic anomaly)	baa4baa1/gin3dou2/loeng5/**do2**/saan1 *爸爸/見到/兩/朵/山 The father /saw /two/$\mathbf{CL_{do2}}$/ mountains
Omission(outright syntactic anomaly)	dai6dai6/gin3dou2/loeng5/saan1 *弟弟/見到/兩/山 The younger brother/ saw/ two/ ∅ /mountains
Default(mild semantic anomaly)	ze4ze1/gin3dou2/loeng5/**go3**/saan1 *姐姐/見到/兩/個/山 The elder sister/ saw /two /$\mathbf{CL_{go3}}$/ mountains

Notes. The ERP analysis is time-locked to the target noun (i.e., *saan1* "mountain") , where the congruence or incongruence of the classifier-noun agreement is detectable (for details see Section 2.4 EEG recording and data analysis). CL = classifier.

2.4. EEG recording and data analysis

EEG signals throughout the experiment were collected, using small-sized (50–53 cm) 64-channel EEG caps with sintered Ag/AgCl electrodes in the standard 10–20 system for EEG electrode placement (Neuroscan, Quick-Cap). The sampling rate was 1,000 Hz. The online reference electrode was located between Cz and CPz. External electrodes placed above and below left eye were used to measure the vertical electrooculogram (VEOG) and electrodes were placed on the left and right eye's outer canthi to measure the horizontal electrooculogram (HEOG). The impedance of the electrodes was maintained below 5kΩ during the experiment.

Signal processing and analysis were performed with EEGLAB (Delorme & Makeig, 2004) and ERPLAB toolbox (Lopez-Calderon & Luck, 2014) in MATLAB. The data were first down sampled to 500 Hz and re-referenced offline to the mean of the left and right mastoid electrodes. Noisy intervals in the data that contained irregular or large muscular artifacts and extreme movements were removed manually by the first author via visual inspection. The remaining data were then filtered with an Infinite impulse response (IIR) Butterworth (Db/oct = 12, Db/dec = 40, Order 2) bandpass filter between 0.1 Hz and 30 Hz. Independent component analysis (ICA) was conducted through AMICA (Palmer et al., 2012) on the continuous data to remove components of eye blinks and eye movements.[5] The ICA-corrected data were epoched from −200 ms before the onset of the noun to 1500 ms after its onset. The epoch was time-locked to the onset of the target noun (e.g., mountains), where semantic or syntactic anomaly of classifier-noun agreement can be detected. The data was then baseline corrected using the −200 to 0 ms time window before the target noun. Artifact detection was conducted by a moving window peak-to-peak analysis (Luck, 2014), and trials with detected artifacts were excluded from the following analysis (overall acceptance rate is 93.7 %; 93.3 % and 94.1 % for the TD and DLD group, respectively).[6] The moving window peak-to-peak analysis period was set within 200 ms ∼ 1200 ms, with a moving window full width of 200 ms, and a window step of 100 ms, and the voltage threshold of 150 μV, performed on each of the selected centro-posterior and parieto-occipital electrodes (CP1, CPz, CP2, P1, Pz, P2, PO3, POz, and PO4). As the current study focused on the N400 and P600, the selection of the above electrodes for the later analyses was based on previous studies (Chan, 2019; Kim et al., 2018; Pijnacker et al., 2017; Purdy et al., 2014; Qian & Garnsey, 2016). The time window for N400 and P600 was selected as 250 ∼ 500 ms and 750 ∼ 1000 ms respectively, which were based on the typical windows suggested in previous ERP studies on Chinese classifiers and previous ERP studies on children with DLD using morphosyntactic tasks of spoken sentences (Fonteneau & van der Lely, 2008; Haebig et al., 2017; Purdy et al., 2014; Weber-Fox et al., 2010). To compare the N400 and P600 effects between the three anomaly conditions (incorrect, omission, and default), difference waves of each selected electrode were also generated by subtracting the correct condition from the incorrect, omission and default

[5] For better ICA decomposition, the ICA was applied on the same dataset but with a high-pass filter at 1.5 Hz, following the recommended practice (Winkler et al., 2015). The ICA weights were then imported back to the filtered dataset for the steps mentioned in the main text. The ICA computation was conducted with the resources provided by the NSG Portal (Sivagnanam et al., 2013, https://www.nsgportal.org).

[6] The average numbers of the remaining trials in each condition for the two groups were presented as follows. TD group: Correct (M = 56.2, SD = 5.4), Incorrect (M = 56.7, SD = 4.7), Omission (M = 56.8, SD = 4.5), Default (M = 54.5, SD = 5.1); DLD group: Correct (M = 57.0, SD = 3.1), Incorrect (M = 56.8, SD = 3.8), Omission (M = 57.3, SD = 3.2), Default (M = 55.1, SD = 3.3).

conditions respectively (i.e., incorrect-correct, omission-correct, and default-correct). The mean amplitudes of N400 and P600 were extracted from each selected time window of the waveforms or difference waves. We also used these mean amplitude measures to examine the N400-P600 tradeoff and compute the RDI (see details later).

In order to visualize the N400-P600 tradeoff, we plotted the magnitude of the N400 effect relative to the correct baseline (i.e., incorrect-correct, omission-correct, and default-correct) on the x-axis, against the magnitude of P600 effect. For the N400, which typically exhibits a more negative amplitude with a stronger semantic anomaly, we inverted the amplitude values to represent the effect magnitude in the positive direction. This allows for a more direct and intuitive comparison where an increase along the x-axis corresponds to a stronger N400 effect. The P600, where a more positive amplitude reflects stronger syntactic integration or reanalysis efforts (Delogu et al., 2019; Kaan et al., 2000; Kim et al., 2018; Tanner, 2019), is represented in its original positive direction along the y-axis. This presentation strategy ensures that data visualization is congruent with the cognitive processes being represented, with increases in both dimensions reflecting increased processing efforts. The resulting scatterplots show different degrees of N400-P600 tradeoff across conditions and groups. Steeper slopes indicate stronger tradeoffs whereby the P600 effect is heavily decreased with increasing N400 effect, and shallower slopes indicate weaker tradeoffs where the P600 effect is less dependent on the N400 effect.

2.5. Statistical analysis

The behavioral accuracy of the grammaticality judgment task was operationalized as the proportion of correct responses to the total number of trials for each participant within each condition. This yielded a continuous dependent variable ranging from 0 to 1, where 0 indicated no correct responses and 1 indicated perfect accuracy. The accuracy data were analyzed using the mixed-effect model, which included group (DLD vs TD), condition (Correct, Incorrect, Omission, and Default) and their interaction as the fixed effects, age as the controlled covariate, and the by-participant random intercept. Reaction times were analyzed while trials with RTs shorter than 200 ms and longer than 10000 ms were excluded from the RT analysis.

For the ERP data, the N400 and P600 mean amplitude measures were also analyzed in the same way as the behavioral data using the linear mixed-effects model (see details later). To further examine the N400-P600 tradeoff, we built a mixed-effects model and used the P600 mean amplitude of the difference waves for each selected electrode (i.e., the P600 effect magnitude) from the three anomaly conditions as the dependent variable. For the fixed effects, we added the corresponding inverted N400 mean amplitude of difference waves for each selected electrode (i.e., the N400 effect magnitude), together with group and condition and their two-way and three-way interactions with the N400 effect magnitude. By-participant and by-electrode random intercepts were included. Slope estimations and pairwise comparisons were then conducted to probe the N400-P600 dependence (i.e., the tradeoff) differences across conditions and between groups (see details later).

To assess the relative dominance between N400 and P600, we computed the RDI of each selected electrode using the N400 and P600 mean magnitude of difference waves (the N400 and P600 effects) for the three anomaly conditions respectively. The RDI was calculated using the following formula from (Grey et al., 2017):

$$(P600_{effect\ magnitude} - N400_{effect\ magnitude})/\sqrt{2}$$

Depending on whether the calculation uses ($\mathrm{P600_{effect\ magnitude}} - \mathrm{N400_{effect\ magnitude}}$) or ($\mathrm{N400_{effect\ magnitude}} - \mathrm{P600_{effect\ magnitude}}$), the resulting RDI would reflect either relative response dominance of P600 or N400, respectively. The formula above operationalized the relative dominance of P600 as the perpendicular distance from the individual data point to the equal effect sizes line (i.e., the Y axis represented $\mathrm{P600_{effect\ magnitude}}$, the X axis represents $\mathrm{N400_{effect\ magnitude}}$, and the line is $y = x$). The more positive the RDI is, the more P600-dominant the response is. If not specified, the RDI mentioned in the current study refers to the RDI of P600. Similar to the ERP analysis, we also constructed a linear mixed-effects model, with RDI from each selected electrode as the dependent variable. Group, condition, and their interaction were included as the fixed effects while age was controlled as a covariate, together with by-electrode and by-participant random intercepts. To test our prediction that the children' grammar performance positively predicts their processing dominance (i.e., N400-dominance vs P600-dominance), a linear mixed effect model was constructed with RDI as the dependent variable, the children's age-normed scores[7] from HKCOLAS's subtest of grammar as the fixed effect, and the age-normed score of Expressive Nominal Vocabulary was controlled. The random intercepts were participants and electrodes.

All the mentioned analysis was conducted with R version 4.3.2 (R Core Team, 2023). The mixed-effects models were conducted with the Package *lme4* (Bates et al., 2015). To obtain the significance of the main effects and the interaction effects from the mixed-effect models when it involved four-level or three-level categorical factor (i.e., condition), we used the anova function in R, and p-values were estimated based on the Satterthwaite's approximation method using the Package *lmerTest* (Kuznetsova et al., 2017). Post-hoc analysis and tradeoff slope estimations were conducted with the Package *emmeans* (Lenth, 2023). To correct the p-values for post-hoc multiple comparisons, Tukey's HSD adjustment was applied. Table of results for all the mixed effects model and correspond post-hoc analyses can be found in the Supplementary Materials in order.

3. Results

3.1. Behavioral results

3.1.1. Accuracy

The accuracy of the two groups of children in each condition in the grammaticality judgment task is displayed in Fig. 1a. The accuracy data were analyzed using a linear mixed-effects model with the group (DLD, TD) as a between-subjects factor and condition (Correct, Incorrect, Omission, General) as a within-subjects factor, and age was controlled. The analysis revealed a significant main effect of group ($F(1, 37) = 21.02$, $p < 0.001$), where the TD group had significantly higher accuracy compared to the DLD group, confirming that the DLD group had consistent difficulties with classifiers. A significant main effect of condition was also observed ($F(3, 114) = 18.84$, $p < 0.001$), suggesting that different conditions were associated with varying levels of accuracy. Post-hoc comparisons revealed that the correct condition elicited significantly lower accuracy than all the other three conditions (all $ps < 0.001$), which had similar accuracies and were not significantly different from each other (all $ps > 0.05$). The group by condition interaction was not significant ($F(3, 114) = 1.53$, $p = 0.21$).

3.1.2. Reaction time

The reaction times (RTs) for the grammaticality judgment task (see Fig. 1b) were analyzed using a linear mixed-effects model with the group as a between-subjects factor and condition (Correct, Incorrect, Omission, Default) as a within-subjects factor, and age was controlled. RTs were measured from the offset of each sentence to provide a fair comparison across conditions with varying sentence lengths. The analysis

[7] Age-normed scores adjust for a child's age-related language development, providing a benchmark to compare language abilities across children of different ages. In our study, we used these scores from the HKCOLAS, a norm-referenced diagnostic tool for DLD, to ensure that our analysis accurately reflects language performance while accounting for age differences.

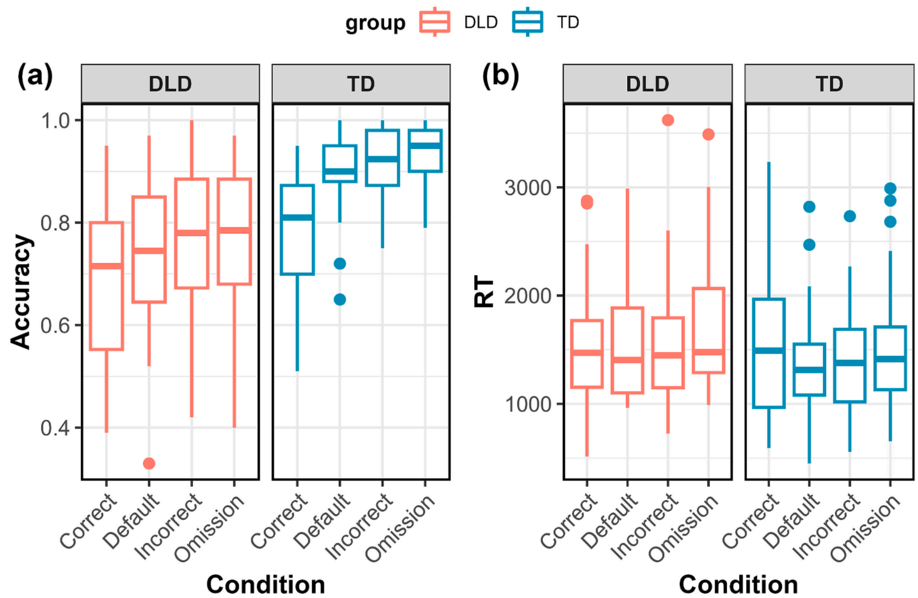

Fig. 1. Behavioral Results. (a) accuracy and (b) RT by groups across conditions. Median, interquartile ranges, and outliers are shown.

indicated neither significant main effect of group ($F(1, 37) = 0.90$, $p = 0.35$) nor significant group by condition interaction ($F(3, 114) = 0.44$, $p = 0.73$). There was a significant main effect of condition ($F(3, 114) = 2.74$, $p < 0.05$). However, post-hoc comparisons only revealed that the omission condition showed marginally significantly longer RTs compared to the default condition ($t(114) = -2.53$, $p = 0.06$), and the incorrect condition (t (114) = -2.43, $p = 0.08$), and there were no other RT differences between the conditions approaching significance after p-value correction.

3.2. ERP results

The following section will detail the outcomes of the ERP analyses, specifically focusing on the amplitudes of the N400 and P600 components, their interplay in terms of the N400-P600 tradeoff as well as the P600 dominance (i.e., RDI) and its relation to the children's grammatical performance. The ERP waveforms of the four conditions in the two groups of participants based on the average of the selected electrodes can be found in Supplementary Materials Fig. S1.

3.2.1. N400 amplitude

To the best of our knowledge, this is the first ERP study on the neural correlates of classifier-noun agreement processing conducted on Chinese children. Thus, we first analyzed the N400 mean amplitude in the TD group to confirm whether those classifier anomaly conditions induced the predicted N400 effects (see Fig. 2a). To this end, a linear mixed-effects model with the N400 mean amplitude as the dependent variable, condition as the fixed effect, age as the controlled variable, and with by-electrode and by-participant random intercepts. The analysis revealed a significant main effect of condition ($F(3, 689) = 62.88$, $p < 0.001$). Age did not significantly affect the N400 mean amplitude ($F(1, 18) = 0.36$, $p = 0.56$), suggesting that the elicited N400 effects were

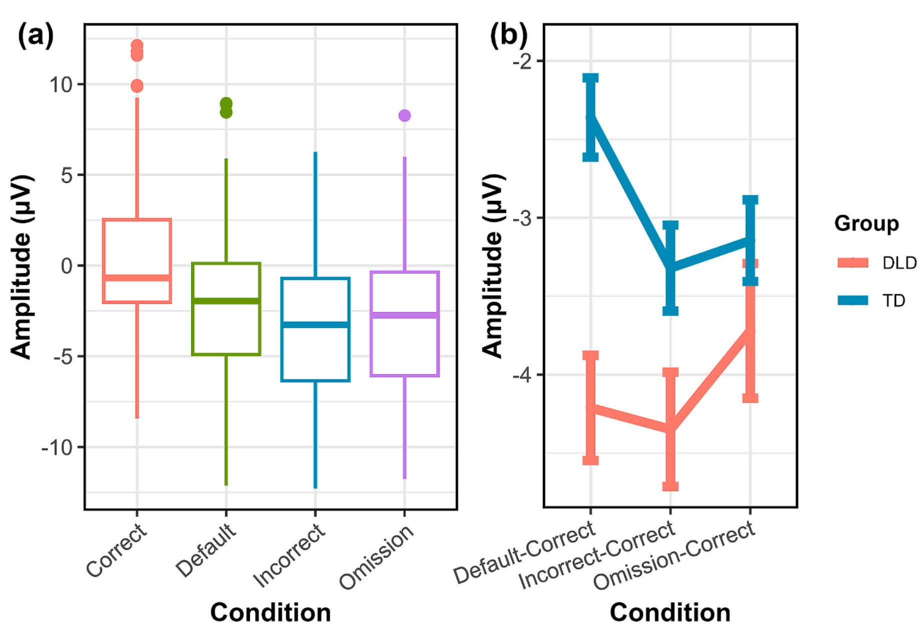

Fig. 2. N400 results. (a) the N400 amplitude for TD children across conditions. Median, interquartile ranges, and outliers are shown. (b) the N400 effect by groups across condition contrasts. The error bars represent one standard error of the mean (SEM).

robust across the age range tested.

Post-hoc multiple comparisons indicated that the incorrect, omission, and default conditions all induced robust N400 amplitude changes (i.e., more negative N400) relative to the correct condition as the baseline: default vs. correct, $t(689) = 8.66$, $p < 0.001$; incorrect vs. correct, $t(689) = 12.18$, $p < 0.001$; omission vs. correct, $t(689) = 11.53$, $p < 0.001$. Moreover, the default condition elicited smaller (less negative) N400 effect than the incorrect ($t(689) = 3.52$, $p < 0.01$) and omission condition ($t(689) = 2.88$, $p < 0.05$), while the latter two were not significantly different from each other. These findings confirmed our predictions that, at least in the TD group, the incorrect condition with semantic anomaly triggered largest N400 effects, whereas the default condition with mild semantic anomaly triggered a smaller but significant N400 effect. It is intriguing to note that the omission condition elicited comparable N400 amplitude to the incorrect condition.

Secondly, we compared the TD and DLD groups on the magnitude of N400 effects of the three classifier anomaly conditions using the difference waves (incorrect-correct, omission-correct, default-correct). Fig. 2b displays the N400 effect amplitudes of the difference waves for the two groups. The linear mixed-effects model included group, condition, and their interaction as fixed effects, age as the controlled variable, and the by-electrode and by-participant random intercepts. Although there was a numeric trend that the children with DLD exhibited stronger N400 effects in general ($M_{\text{difference}} = -1.09$), the main effect of group did not reach significance ($F(1,37) = 1.08$, $p = 0.31$). A significant main effect of condition was found, $F(2, 1028) = 3.61$, $p < 0.05$, indicating reliable differences in the N400 effects across the three conditions. The group by condition interaction was also significant, $F(2, 1028) = 4.69$, $p < 0.01$, suggesting that the N400 effects varied across conditions and between the TD and DLD groups.

Post-hoc between-group comparisons for each of the three anomaly conditions did not reveal any significant differences between the two groups: default, $t(41.1) = -1.66$, $p = 0.10$; incorrect, $t(41.1) = -0.90$, $p = 0.37$; and omission, $t(41.1) = -0.48$, $p = 0.63$. That is, under each condition, the N400 effect was comparable between the two groups. Post-hoc within-group comparisons revealed that in the TD group, the magnitude of the N400 effect was significantly smaller in the default condition compared to that in the incorrect ($t(1028) = 3.2$, $p < 0.01$) and omission condition ($t(1028) = 2.6$, $p < .05$). This pattern of smaller N400 effect in the mild semantic anomaly condition (i.e., default condition) has already been reported above. In contrast, in the DLD group, no pairwise comparisons involving the default condition reached significance while the contrast between incorrect and omission was marginally significant ($t(1028) = -2.10$, $p = 0.09$). That is, for the children with DLD, all anomaly conditions elicited relatively comparable N400 effects and there were no clear distinctions between the default condition and other anomaly conditions.

3.2.2. P600 amplitude

Similar to the analysis of the N400 amplitude above, we first compared the P600 mean amplitude across the four conditions in the TD group (see Fig. 3a). We built a linear mixed-effects model incorporating condition as the fixed effect and age as the controlled variable, together with by-electrode and by-participant random intercepts. It revealed a significant main effect of condition ($F(3, 689) = 3.02$, $p < 0.05$). Post-hoc analyses identified a significantly heightened P600 response in the omission condition compared to the baseline correct condition ($t(689) = -2.93$, $p < 0.05$). No other pairwise comparisons were significant. This pattern was compatible with our prediction and confirmed the validity of the omission condition as a salient syntactic anomaly, as intended. These results demonstrated that the P600 amplitude was particularly sensitive to classifier omission, possibly due to the syntactic integration difficulty or syntactic repair processes in this condition.

Then we compared the two groups in the magnitude of the P600 effects of the three classifier anomaly conditions using the difference waves (incorrect-correct, omission-correct, default-correct). Fig. 3b displays the P600 effect magnitudes of the difference waves for the two groups. We constructed a linear mixed-effects model incorporating group, condition, and their interaction as the fixed effects, along with age as the control covariate and the by-electrode and by-participant random intercepts. The TD group exhibited a numerically larger P600 effect compared to children with DLD ($M_{\text{difference}} = -2.34$, $t(37) = -1.25$, $p = 0.22$), but the main effect of the group did not reach statistical significance. Importantly, a significant interaction between condition and group was observed ($F(2, 1028) = 13.04$, $p < 0.001$).

Post-hoc between-group comparisons for each anomaly condition highlighted a significant group difference in the omission condition ($\beta = -3.91$, $SE = 1.9$, $t(39.2) = -2.06$, $p < 0.05$), where the TD group had a significantly larger P600 effect than the DLD group. No group difference was found in the other two anomaly conditions ($ps > 0.05$). Post-hoc

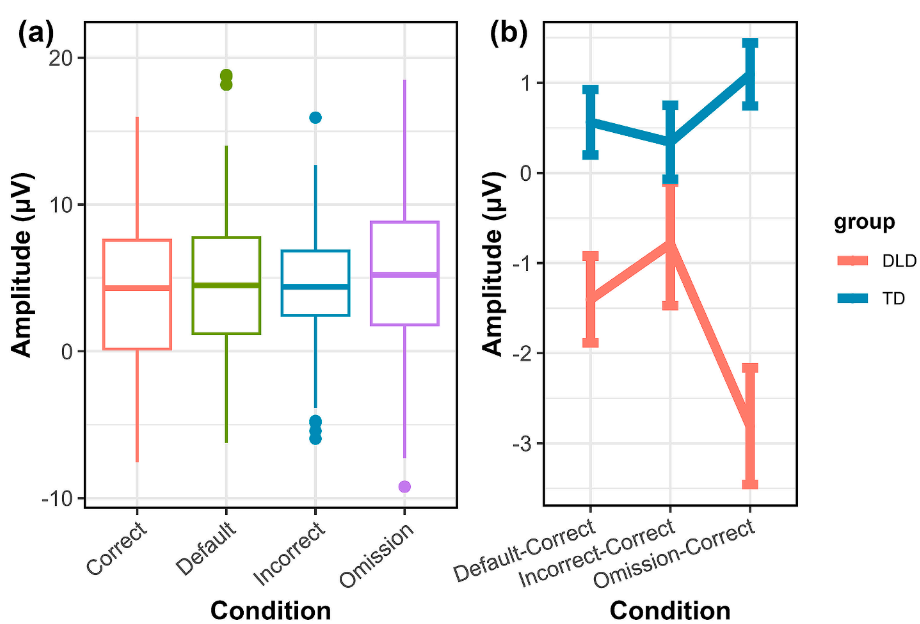

Fig. 3. P600 results. (a) the P600 amplitude for TD children across conditions. Median, interquartile ranges, and outliers are shown. (b) the P600 effect by groups across condition contrasts. The error bars represent one SEM.

within-group pairwise comparisons revealed no significant differences between the three anomaly conditions in the TD group ($ps > 0.05$), albeit the P600 effect magnitude was numerically larger and significantly above zero in the omission condition as mentioned above. On the other hand, the omission condition elicited significantly smaller P600 effects compared to the other two conditions ($ps < 0.01$) in the DLD group.

These results showed that, unlike TD children, those with DLD exhibited an attenuated P600 effect, particularly for syntactic omissions. This indicated a potential disruption in the neural mechanisms underlying syntactic processing among children with DLD and possibly reflected a deficit of their syntactic integration or repair processes.

3.2.3. The N400-P600 tradeoff

The linear mixed-effects model for the tradeoff analysis revealed that the three-way interaction between the N400 magnitude, group, and conditions was significant ($F(2, 1026.99) = 10.14$, $p < 0.001$). This interaction was particularly informative as it indicated that the dependence between semantic processing (as measured by N400 effect magnitude) and syntactic processing (as indicated by P600 effect magnitude) did not operate uniformly across conditions or groups.

To unravel the complexities of this three-way interaction, we estimated the regression slopes, which quantified the degree of changes in P600 effects in accordance with the N400 effect changes under each condition across individuals within each group. The “emtrends” function from the “emmeans” package (Lenth, 2023) in R was used for this purpose. A more negative slope (e.g., −1 vs −2) suggests a steeper tradeoff, indicating a stronger dependence between the N400 and P600 effects (e.g., a more substantial decrease in the P600 effect associated with an increase in the N400 effect).

All the confidence intervals of the estimated slopes did not include zero (see Supplementary Materials Table S16), indicating that under all the conditions within each group, the N400-P600 tradeoff was significant. For the DLD group, the most pronounced tradeoff was observed in the incorrect-correct condition ($M = -1.00$, $SE = 0.06$), while the TD group exhibited a similar pattern but with a less steep slope ($M = -0.86$, $SE = 0.07$). The omission-correct condition had the least steep slope in the TD group ($M = -0.31$, $SE = 0.07$), suggesting a weaker tradeoff in the omission condition, i.e., less dependence between the N400 and P600 effect.

Subsequent analyses were conducted to compare the slopes between the DLD and TD groups for each condition respectively (see Supplementary Materials Table S17). The results revealed a significant group difference in the omission-correct condition (Slope Difference = -0.56, $t(1043) = -6.17$, $p < 0.0001$), indicating a more pronounced N400-P600 tradeoff in the DLD group when compared to the TD group. This suggests that, particularly for the omission-correct condition, as the magnitude of the N400 increased, the amplitude of the P600 decreased more steeply in children with DLD than in TD children (Fig. 4a). No significant group differences in slopes were found for the other conditions.

Within-group contrasts were performed to examine differences in the tradeoff slopes between conditions for each group (see Supplementary Materials Table S18). In the DLD group, significant differences were found between the default-correct and incorrect-correct conditions ($p < 0.001$) as well as between the default-correct and omission-correct conditions ($p = 0.02$). Marginally significant difference was found between the incorrect-correct and omission-correct conditions ($p = 0.09$). In other words, for children with DLD, numerically steeper tradeoff slope was observed in the incorrect (−1.00) condition compared to the omission (−0.87) condition; both conditions had significantly steeper slopes than the default condition (−0.69). In contrast, for the TD group, significant contrasts emerged between all condition pairs ($ps < 0.05$), reflecting a clearer distinction in processing between these conditions. Specifically speaking, the incorrect condition elicited the steepest slope (−0.86), followed by the default condition (−0.62), and finally the omission condition (−0.31).

Collectively, these results indicate that children with DLD showed a different pattern for the N400-P600 tradeoff across conditions compared

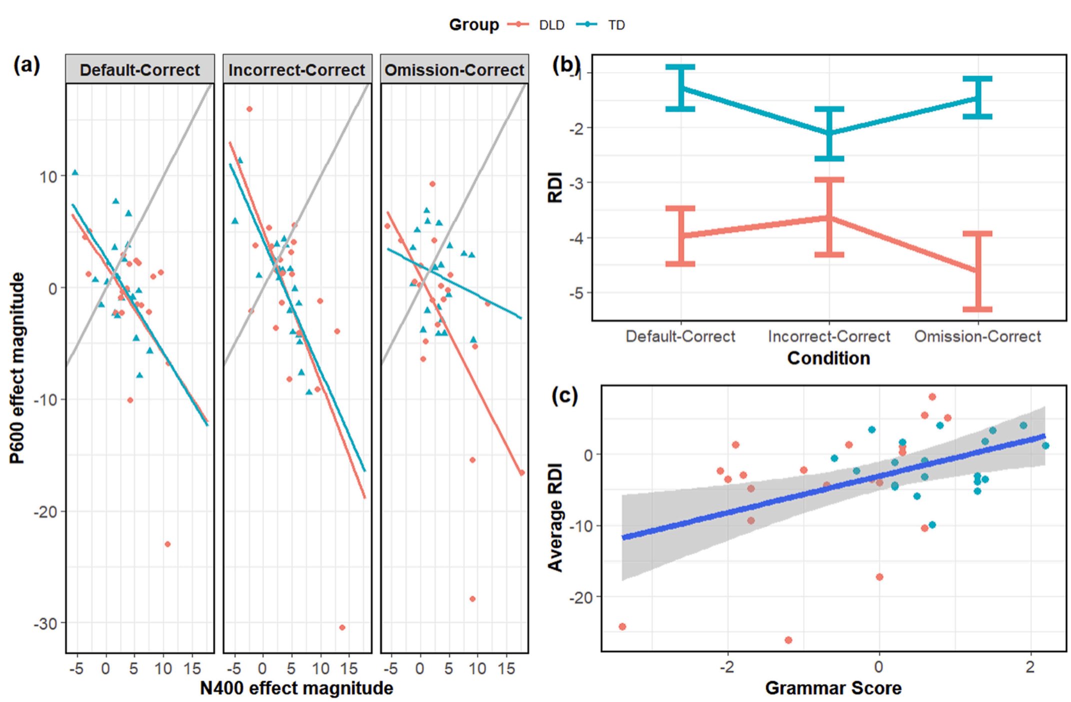

Fig. 4. The N400-P600 interplay. (a) Tradeoff between the N400 and P600 effects under the three conditions. Each point represents one participant. For clear visualization, a full range of lines is applied. The positions above or below the gray line (y = x) indicate P600- or N400- dominance, respectively. (b) RDI values for the TD and DLD groups in the three conditions (default-correct, incorrect-correct, omission-correct). The error bars represent one SEM. (c) The relationship between RDI (averaged across selected electrodes for each participant; one spot represents one participant for clear visualization) and grammar scores under the omission condition, collapsing the two groups (marked in different colors though).

to the TD children, particularly for the omission condition.

3.2.4. RDI and its relationship with grammatical performance

As mentioned earlier, the RDI measures the relative response dominance between the P600 and N400 within an individual. As shown in Fig. 4a, individual datapoints located above or below the equal effect size line indicate P600- and N400-dominant responses respectively; relative distance from the equal effect size line captures the degree of such dominance. A larger (more positive) RDI indicates greater P600-dominance.

The visualization of the RDI data was also shown in Fig. 4b. The RDI analysis revealed a significant interaction between group and condition ($F(2,1028) = 3.78$, $p < 0.05$), while no other effects reached significance. Post-hoc between-group comparisons revealed no significant differences between groups under any condition. Albeit non-significant, a numerical trend of larger RDI (i.e., greater P600 dominance) was observed in the TD group compared to the DLD group, especially in the omission condition ($p = 0.11$). Post-hoc within-group comparisons revealed marginally significant differences between the omission and incorrect condition in the DLD group ($M_{\text{incorret-omission}} = 0.99$, $t(1028) = 2.27$, $p_{\text{corrected}} = 0.06$ while $p_{\text{uncorrected}} = 0.02$), whereas no other contrasts approached statistical significance. Altogether, these results implied particularly degraded P600 dominance in the omission condition in the DLD group, opposite to the expectation of heightened P600 dominance given the outright syntactic violation in this condition.

To better understand individual differences in the P600 dominance pattern and their relationship with language performance profiles, a subsequent analysis further explored the relations between the RDI and children's grammatical scores from the standardized HKCOLAS test. Note that some children with DLD had normal grammatical scores (i.e., above − 1.25 SD), so we collapsed the two groups; the aim was to explore whether better grammatical abilities were associated with stronger P600 dominance, irrespective of the group label (DLD or TD). Since the omission condition carried major syntactic anomaly and was expected to elicit the clearest P600 dominance, the RDI under this condition was selected to explore its relationship with grammatical performance, with vocabulary abilities controlled for. A linear-mixed effect model was constructed with RDI under the omission condition as the dependent variable, where the fixed effect was the age-normed grammatical scores from the HKCOLAS and the age-normed vocabulary scores from the HKCOLAS was entered as a controlled covariate, with the by-participant and by-electrode random intercepts included. The result revealed that the grammatical score was a significant positive predictor ($t(37) = 2.419$, $p < 0.05$) for the RDI under the omission condition (see Fig. 4c). Children with better grammatical performance in the HKCOLAS tended to have stronger P600 dominance (i.e., larger RDI) in response to classifier omission. Note that the response dominance of N400 was the opposite of the response dominance of P600 according to the RDI formula. Therefore, the current results indicated two related findings: better grammatical performance was associated with stronger P600 dominance and concurrently with weaker N400 dominance.

4. Discussion

In the current study, we utilized the classifier-noun agreement in Chinese and employed ERPs to examine real-time semantic processing, syntactic processing, and their dynamic interplay (tradeoff and response dominance) during a grammaticality judgment task among Cantonese TD and DLD children. The behavioral results showed that the DLD group had an overall lower accuracy compared to the TD group in judging the grammaticality of the classifier-noun agreement. This finding is consistent with earlier studies showing that Chinese children with DLD have difficulties in the production of the numeral-classifier-noun structure (Cheung, 2008; Stokes & So, 1997). Apart from this general accuracy difference, no other significant between-group differences were found in accuracy or RT. A notable observation is the lower accuracy in the correct condition for both groups. This finding may suggest that the children in this study have not fully acquired nuanced semantic knowledge about the classifier-noun agreement and appeared cautious when judging the classifier-noun agreement. As children are still mastering these linguistic subtleties, they may tend to be overly cautious, leading to more false negatives in the correct condition, but effectively rejecting inappropriate classifier-noun associations in the other anomaly conditions. This pattern echoes previous findings (e.g., Stokes & So, 1997), which also pointed out a cautious strategy in terms of avoiding incorrect use of specific classifiers.

Importantly, based on the PDH, we hypothesized that children with DLD would exhibit syntactic deficits, whereas their semantic processing might be largely intact and might even compensate for their syntactic processing deficits during real-time language processing. Overall, the ERP results confirmed our predictions, as discussed in the following subsections.

4.1. Semantic processing: N400 effects

In TD children, we found significant N400 effects elicited by all the anomaly conditions compared to the correct baseline condition, which confirmed that TD children were adept at detecting semantic incongruities in classifier-noun agreement. Particularly, the N400 effect triggered by the default condition was significantly milder than the incorrect and omission condition. This observation was in line with prior research and our predictions that the default classifier *go3*, due to its lower semantic specificity, induced a milder semantic violation than the incorrect classifiers or classifier omissions (Frankowsky et al., 2022; Qian & Garnsey, 2016). Another intriguing finding is that the omission condition elicited comparable N400 amplitude to the incorrect condition. The N400 effects elicited in the incorrect condition confirmed our prediction, but the N400 effects elicited in the omission condition were not expected, as we mainly predicted a P600 effect. This may suggest that the omission condition also induced effortful semantic integration among the participants. While the absence of a classifier clearly violates the syntactic rule, it also creates semantic incongruency, which may increase the semantic integration difficulty, especially in children who are still acquiring the full complexity of semantic knowledge in classifier-noun agreement.

Overall, the findings in the TD group revealed their nuanced neural responses to different severities of semantic anomaly and captured the gradations in semantic processing that might be overlooked in less sophisticated designs (e.g., only correct vs incorrect). This highlighted the advantage of the current design in discriminating between different levels of semantic anomaly in classifier-noun agreement in both children with typical development and language disorders.

Previous ERP studies on DLD have often reported intact semantic processing abilities, as evidenced by N400 effects comparable to those observed in TD children. These findings have been interpreted as indicating that DLD children possess a relatively preserved capacity for detecting semantic incongruities (Courteau et al., 2023; Fonteneau & van der Lely, 2008; Haebig et al., 2017; Weber-Fox et al., 2010). Our results largely align with these observations. Our DLD children showed N400 effects across conditions that were not significantly different from those in the TD group. However, a notable exception was observed in the default condition, where the DLD group exhibited a numerically stronger N400 effect, although it did not reach statistical significance after correction ($p = 0.10$).

Despite these similarities, a critical difference was that the DLD group demonstrated similar magnitudes of N400 effects across the three anomaly conditions, whereas the TD group clearly showed a milder N400 effect for the default condition compared to the others. The lack of such modulation across the conditions in the DLD group, aligning with previous findings by Sabisch et al. (2006), Neville et al. (1993), Popescu et al. (2009), and Helenius et al. (2009), indicates potentially less differentiated semantic processing when compared to their TD counterparts. Specifically, the N400 response in the DLD group tended to be maximally engaged even for a milder semantic anomaly, thus obscuring finer distinctions between different severities of semantic anomalies (i.e., default vs incorrect and default vs omission). This suggests a potentially over-active semantic processing mechanism in the children with DLD.

Thus, the current study demonstrated two kinds of findings observed in previous N400 research on DLD: preserved N400 effects comparable to TD children in response to semantic anomalies and a lack of fine-tuned modulation across conditions. Our design allowed these seemingly inconsistent findings to emerge within a single framework, potentially providing a more integrated perspective on the N400 dynamics in children with DLD.

Evidences from previous behavioral studies can also corroborate this interpretation of semantic overactivation in DLD. Pizzioli and Schelstraete (2011) observed that children with DLD exhibited larger semantic priming effects and comparable accuracy than the language-matched TD children in an auditory pair-primed paradigm, reflecting a heightened engagement with semantic cues, analogous to the overactive N400 responses we observed. Additionally, Benham and Goffman (2020) found that introducing lexical-semantic information helped stabilize phonological patterns of the children with DLD in learning and producing new words, underscoring the significant influence of semantic processing on broader linguistic abilities in these children.

In sum, these findings suggest that children with DLD likely engage in semantic processing in a potentially excessive manner. While these children with DLD could process semantic information as robustly as their TD counterparts, there was a critical difference in quality: their processing style lacked the gradation and specificity observed in TD children.

4.2. Syntactic processing: P600 effects

In the TD group, the significant P600 modulation in response to classifier omissions compared to the baseline correct condition confirmed the sensitivity of this component to the presence of an outright syntactic violation. This finding aligns with existing literature that associates the P600 component with complex syntactic reanalysis and repair processes (e.g., Hahne & Friederici, 1999; Osterhout & Holcomb, 1992).

It is also worth noting that the incorrect and default conditions elicited numerically larger P600 amplitude than the correct baseline among the TD group, although they did not reach significance. This is in line with previous studies on the N400-P600 interplay (Grey et al., 2017; Kim et al., 2018; Tanner et al., 2013), where similar conditions can elicit mixed N400 and P600 responses, and whether the individual response was relatively N400-dominant or P600 dominant depended on which cue (i.e., semantic vs syntactic) won out. More specifically, the current result could reflect a P600 modulation based on the severities of syntactic violation. Note that the incorrect or default conditions still provided some syntactic structure, albeit atypical, for the children to process. In contrast, omissions left a syntactic gap that required most substantial reanalysis and repair, thus inducing a significant P600. An alternative explanation for the lack of significant P600 effects in these two conditions is component overlap (Luck, 2014). The ERP waveform at any time point is a summation of all the latent components at the time (Luck, 2014), and previous studies indicated temporal overlap between the N400 and P600 (Brouwer & Crocker, 2017). In other words, the significant N400 effects under the incorrect and default conditions might attenuate the later P600, depending on their relative sizes. However, in the current design, this possibility may be largely ruled out. First, temporally non-adjacent time windows were chosen for the N400 ($250 \sim 500$ ms) and P600 ($750 \sim 1000$ ms) analysis deliberately (Kim et al., 2018). Second, the omission condition triggered comparable N400 responses in size to the incorrect condition but still showed a significant P600 effect.

Compared to the TD group, children with DLD exhibited a particular vulnerability in the omission condition, as evidenced by the significantly smaller P600 effect than the TD group in this condition. This attenuated P600 effect suggests a disruption in the syntactic integration process, which is critical for constructing coherent sentence structures. The significant difference between groups for the omission-correct condition points to a salient syntactic challenge in managing syntactic gaps in DLD. This finding is remarkable, given that we did not specifically confine children in the DLD group to only those with a grammatical deficit. Indeed, children failing any two subtests of HKCOLAS were included in the DLD group, following the standard practice. Thus, it is intriguing to observe a syntactic deficit in children with diverse profiles of language difficulties.

Similar syntactic challenges in DLD were also observed with the omission of the third person singular −s tense marker in previous studies (Haebig et al., 2017; Weber-Fox et al., 2010), providing support for the hypothesized syntactic deficits in DLD across different language backgrounds. In addition, it is worth mentioning that previous studies have also explored syntactic processing through commission errors (i.e., errors involving incorrect additions, e.g., "I say**s** that") in grammaticality judgement tasks (Haebig et al., 2017; Purdy et al., 2014; Weber-Fox et al., 2010). However, a commission error design was not directly applicable in our current study due to the specific nature of classifier-noun agreement in Chinese. Future studies might explore other designs to probe aspects of commission errors in Chinese grammatical processing other than classifier-noun agreement.

4.3. Dynamic interplay of semantic and syntactic processing: The N400-P600 tradeoff and P600 response dominance

Regarding the interplay between semantic and syntactic processing, our analysis revealed that in general, the magnitude of the N400 effect was negatively correlated with the P600 effect across individuals, confirming the N400-P600 tradeoff (Kim et al., 2018). Further, the N400-P600 tradeoff varied across conditions and between groups. The TD group's response patterns served as a baseline, showcasing distinct tradeoff slopes between the incorrect and omission conditions. The steepest tradeoff was observed in the incorrect condition, suggesting a stronger N400-P600 dependence, and a flatter slope in the omission condition, suggesting a weakened N400-P600 dependence. In the incorrect condition, where there was an overt semantic incongruency between the classifier and noun, the N400 effect magnitude strongly and negatively modulated that of the following P600: individuals who exhibited a large N400 effect in response to the semantic incongruency showed a small P600 (and vice versa). Intriguingly, this dependence was weakened in the omission condition, suggesting that TD children were capable of prioritizing syntactic processing over semantic processing, ensuring a relatively robust P600 effect, which is less dependent on the preceding N400 effect. These results suggested that the TD children can moderate the tradeoff patterns according to different violations to prioritize one processing over the other. This is consistent with the intended manipulation and results of previous studies (Chan, 2019; Qian & Garnsey, 2016; Zhang et al., 2012).

Contrastingly, the DLD group lacked such moderated patterns – they exhibited similar degrees of steepness of tradeoff slopes between incorrect and omission conditions, reflecting a marked reliance on semantic processing, possibly at the expense of syntactic processing, particularly when confronted with outright syntactic violations. This pattern suggests that the DLD group may not prioritize syntactic processing as effectively as TD children in the omission condition.

The RDI results further corroborated this interpretation. The DLD group demonstrated a marginally significantly lower RDI for the omission condition, indicative of a more N400-dominant response particularly when faced with a syntactic gap. This reduced P600 dominance pattern not only echoes the N400-P600 tradeoff findings above but also highlights the differential allocation of processing resources in the DLD group, who may over-rely on semantic processing as a compensation for syntactic deficits.

Further analysis between individual RDI scores and grammar scores from the standardized language test confirmed that individual differences in the degree of P600 dominance in the omission condition were linked with the individual grammatical abilities. Children with better grammatical abilities tended to have a more P600-dominant and less N400-dominant response, a pattern that holds true across groups, even when their vocabulary abilities were controlled for. This finding implies that the RDI might serve as a neural marker for children's grammatical abilities.

As noted by an anonymous reviewer, there is substantial overlap between the two groups in terms of the P600 and N400 effect magnitudes, suggesting that the P600 is not exclusive to the TD group, nor is N400 dominance exclusive to the DLD group (see Fig. 4). Directly linking these magnitudes to individual language abilities is intrinsically challenging, as N400 and P600 effect magnitudes may be influenced by other factors such as participant variability and task-related processing context. The RDI offers an advantage by considering the relative dominance of the P600 and N400 within an individual, thereby partially controlling for these influences. This advantage may further strengthen the plausibility of the RDI serving as a neural marker for children's grammatical abilities.

From the developmental standpoint, the capacity to proficiently navigate the interplay between semantic and syntactic processing and to engage in more intensive syntactic integration when warranted, likely emerges as a hallmark of linguistic maturation in TD children. On the contrary, the absence of similar modulations of the N400-P600 tradeoff across conditions in children with DLD underscores their difficulties in prioritizing syntactic processing over semantic processing and might suggest potential developmental divergences in the neurocognitive foundations of syntactic processing. Rather than an absence of syntactic processing as null P600 responses, this difficulty of prioritizing syntactic processing, when necessary, may be a key neural signature that subserves DLD children's lower accuracy in judging classifier-noun agreement and grammatical performance in general.

4.4. General discussion and theoretical implications

Our findings align with previous research indicating that semantic processing is relatively intact in children with DLD, while syntactic processing shows deficits (Fonteneau & van der Lely, 2008; Haebig et al., 2017; Weber-Fox et al., 2010). Importantly, this study extends prior work by examining the N400-P600 tradeoff and response dominance within the specific context of classifier-noun agreement. The dynamic N400-P600 tradeoff, especially under the omission condition, suggests that children with DLD rely heavily on semantic processing, potentially as a compensatory mechanism to support syntactic challenges deficits (Evans et al., 2022; Haebig et al., 2017; Pijnacker et al., 2017).

We suggest this compensation is, at least, suboptimal from two aspects: task performance and processing mechanism (Cabeza et al., 2018). First, such a compensatory mechanism does not appear to fully mitigate the challenges posed by syntactic processing demands, as evidenced by the DLD children's overall lower accuracy in grammaticality judgment of classifier-noun agreement. Second, the children with DLD demonstrated over-active semantic processing and N400 dominance for outright syntactic violations. This indicates an ill match in processing mechanism where semantic processing was employed to handle syntactic violations. Previous studies have proposed a distinction between adaptive and maladaptive compensation (Cabeza et al., 2018; Liu et al., 2021; Zhao et al., 2023). For example, (Liu et al., 2021) reported evidence for maladaptive compensation in developmental dyslexia. Essentially, the authors found a negative correlation between the right fusiform gyrus network connectivity and pseudoword reading accuracy in dyslexics children, namely, higher right fusiform gyrus network connectivity being associated with lower reading accuracy. In line with this idea, we found a negative relation between the N400 response dominance under the omission condition and the general grammatical performance. Thus, there is some evidence for maladaptive compensation of over-active semantic processing from the current study. Future studies should further probe different compensatory mechanisms to reach a better understanding of adaptive and maladaptive compensations in DLD.

The findings of the current study are consistent with the DP model (Ullman, 2016) and the PDH in DLD (Ullman et al., 2020; Ullman & Pierpont, 2005). According to the PDH, the grammatical and other deficits in DLD can be largely explained by abnormalities of the circuit underlying procedural memory, in particular of the basal ganglia, whereas declarative memory remains relatively spared and might play a compensatory role (Ullman et al., 2024; Ullman et al., 2020; Ullman & Pullman, 2015). As argued by previous ERP studies, the N400 components reflected declarative memory process, and the elicited N400 effects in response to morphosyntactic violations underscore an early reliance on declarative memory (Courteau et al., 2023; Tanner et al., 2013; Ullman, 2004, 2016). This reliance aligns with the DP model's predictions for novice language learners, such as children, who typically depend on declarative memory for language tasks. As proficiency increases, a developmental shift towards procedural memory for grammatical processing is expected—a shift that appears to be hindered in children with DLD due to the procedural memory deficits posited by the PDH. The current results showing degraded P600 effects and relative N400 dominance in the DLD group, particularly in response to the outright syntactic omission, echo the PDH's predictions of procedural memory deficits (Ullman et al., 2020) and a compensatory role of declarative memory (Ullman & Pullman, 2015). This pattern of attenuated P600 responses is analogous to those found in individuals with basal ganglia lesion (Friederici et al., 2003), a neural structure critical for procedural memory (Ullman et al., 2020), and substantially impacted in DLD (Ullman et al., 2024), substantiating the PDH's claims. Although our study did not directly assess the LAN component—often linked to syntactic processing and procedural memory (Tanner et al., 2013; Ullman, 2004)—the attenuated P600 effects observed offered some support for procedural memory dysfunction in DLD. Future research that employs more direct measures of procedural memory such as LAN or fMRI methods, should provide a more nuanced understanding of the interaction between memory systems and language processing in DLD.

While the findings of the current study provide insights into the neurocognitive profiles of DLD, it is essential to acknowledge the absence of data from a group of younger and language-matched children with TD in the current study. Future studies that include younger children with TD can further verify whether the observed patterns are unique to the DLD population or also characteristic of younger children with TD at earlier developmental stages. Secondly, we did not fully follow the persistence and functional impact criteria of CATALISE in the diagnosis of DLD children. Future studies may consider assessing whether the children had a prior diagnosis of language disorder or DLD or received therapy prior to the experiment (i.e., persistence) and whether they demonstrated any functional deficits in life (i.e., functional impact) for better alignment with the CATALISE diagnostic criteria. Thirdly, future studies should assess verbal working memory in children with DLD. Individual differences in verbal working memory capacities are found to underlie the N400-P600 tradeoff patterns (Kim et al., 2018), and verbal working memory deficits are often observed in children with DLD (Leonard, 2014). Apart from verbal working memory, including measures of procedural memory (e.g., the serial reaction time task) may offer a means to directly assess the relationship between a procedural memory deficit and the N400-P600 interplay in children with DLD. Fourthly, as pointed out by an anonymous reviewer, we did not exclude trials with incorrect responses from data analysis, in order to retain more trials for better signal-to-noise ratio. Future studies with a larger number of trials for each condition may exclude incorrect responses, to further verify the findings of the current study. Fifthly, we deliberately presented the spoken sentences without any pause before the target noun to ensure naturalness of the stimuli. For this reason, the baseline period of the ERP waveforms might contain some activities spread over from the preceding words. Future studies may consider inserting a brief pause before the target noun or using deconvolution methods to mitigate the carry-over influence of preceding words. Sixthly, we followed previous EEGs studies on the N400-P600 tradeoff and chose standard N400 (250–500 ms) and P600 (750–1000 ms) windows for data analysis. Future studies may consider adopting a data-driven window selection approach (e.g., cluster-based analysis) to further verify the current findings. Lastly, future studies may pre-register their hypotheses and data analysis plan to avoid *p*-hacking and implement a blinding procedure during data analysis to reduce potential observer bias.

Finally, these findings also have implications for intervention strategies for children with DLD. Targeted intervention like the sequential application of conversational recast treatment followed by auditory bombardment can be effective (Plante et al., 2018). These methods synergistically enhance the syntactic processing capabilities of children with DLD by reinforcing the correct grammatical structures through both immediate contextual feedback and repeated auditory exposure. Such integrated interventions are crucial as they facilitate not only the learning and internalization of syntactic rules but also their automatic application across new linguistic contexts, significantly aiding in the generalization of these skills. The N400-P600 interplay may be included as a key intervention outcome, to assess whether the intervention enhances automatic prioritization of syntactic processing. A grammaticality judgement task with syntactic anomaly conditions can be included before and after the intervention, to assess the N400-P600 interplay. It is expected that the intervention, if efficient, would lead to positive changes in the N400-P600 interplay (e.g., less dependency between the N400 and P600 and increased RDI) together with improved behavioral performance.

5. Conclusion

The current study confirmed the intertwined semantic and syntactic processing that underlies Chinese classifier-noun agreement judgment during real-time language comprehension. Chinese children with DLD have syntactic deficits in processing Chinese classifier-noun agreement, particularly with syntactic gaps (i.e., the omission condition), as indicated by attenuated P600 effects. Although semantic processing was relatively comparable between children with DLD and TD, children with DLD lacked the gradation of N400 effects according to different severities of semantic violations, indicating potentially less differentiated and over-active semantic processing in this group. The N400-P600 tradeoff illustrated a deficit in the automatic prioritization of syntactic processing in children with DLD, together with their lower P600-dominance during the omission condition. This suggested an over-reliance of semantic processing as a suboptimal and possibly maladaptive compensation. These findings not only provide a deeper understanding of the distinct neurocognitive profiles of children with DLD during on-line language processing, but also have important implications for intervention strategies to mitigate linguistic difficulties in this population.

CRediT authorship contribution statement

Jueyao Lin: Writing – original draft, Visualization, Formal analysis. **Xiaocong Chen:** Writing – review & editing, Formal analysis. **Xunan Huang:** Formal analysis. **Patrick Chun Man Wong:** Writing – review & editing, Methodology, Conceptualization. **Angel Wing Shan Chan:** Writing – review & editing, Resources. **Michael T. Ullman:** Writing – review & editing. **Caicai Zhang:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the Project of Strategic Importance funded by the Hong Kong Polytechnic University, Hong Kong SAR. [grant number P0032004] and the "Sci-Tech Innovation 2030 – Brain Science and Brain-like Research Major Project" funded by the Ministry of Science and Technology of the People's Republic of China [grant number 2022ZD0208400] to CZ. We thank Stephanie Wing Laam Siu, Viann Wing Yan Leung, Janeane Chun Ling Mak, Jiayi Lu, Cherry Chung, Yixin Cui, Xinye Huang and Jianfei Ma for their assistance with subject recruitment and EEG data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2025.105571.

Data availability

I have shared the link to my data in the manuscript file

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benham, S., & Goffman, L. (2020). Lexical-Semantic Cues Induce Sound Pattern Stability in Children With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research, 63*(12), 4109–4126. https://doi.org/10.1044/2020_jslhr-20-00244

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, 58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Brouwer, H., & Crocker, M. W. (2017). On the Proper Treatment of the N400 and P600 in Language Comprehension [Opinion]. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01327

Cabeza, R., Albert, M., Belleville, S., Craik, F. I., Duarte, A., Grady, C. L., Lindenberger, U., Nyberg, L., Park, D. C., & Reuter-Lorenz, P. A. (2018). Maintenance, reserve and compensation: The cognitive neuroscience of healthy ageing. *Nature Reviews Neuroscience, 19*(11), 701–710.

Chan, S.-H. (2019). An elephant needs a head but a horse does not: An ERP study of classifier-noun agreement in Mandarin. *Journal of Neurolinguistics, 52*, Article 100852. https://doi.org/10.1016/j.jneuroling.2019.100852

Cheung, H. (2008). Chapter 3. Grammatical Characteristics of Mandarin-speaking Children with Specific Language Impairment. In L. Sam-Po, W. Brendan, & M. Y. W. Anita (Eds.), *Language Disorders in Speakers of Chinese* (pp. 33-52). Multilingual Matters. Doi: doi:10.21832/9781847691170-005.

Coderre, E. L., & Cohn, N. (2023). Individual differences in the neural dynamics of visual narrative comprehension: The effects of proficiency and age of acquisition. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02334-x

Conti-Ramsden, G., & Durkin, K. (2012). Language Development and Assessment in the Preschool Period. *Neuropsychology Review, 22*(4), 384–401. https://doi.org/10.1007/s11065-012-9208-z

Courteau, É., Royle, P., & Steinhauer, K. (2023). Number agreement processing in adolescents with and without developmental language disorder (DLD): Evidence from event-related brain potentials. *Scientific Reports, 13*(1), 22836. https://doi.org/10.1038/s41598-023-49121-1

Delogu, F., Brouwer, H., & Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and cognition, 135*, Article 103569.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Evans, J. L., Maguire, M. J., & Sizemore, M. L. (2022). Neural patterns elicited by lexical processing in adolescents with specific language impairment: Support for the procedural deficit hypothesis? *Journal of Neurodevelopmental Disorders, 14*(1), 20. https://doi.org/10.1186/s11689-022-09419-z

Fonteneau, E., & van der Lely, H. K. (2008). Electrical brain responses in language-impaired children reveal grammar-specific deficits. *PLoS One1, 3*(3), e1832.
Frankowsky, M., Ke, D., Zwitserlood, P., Michel, R., & Bölte, J. (2022). The interplay between classifier choice and animacy in Mandarin-Chinese noun phrases: An ERP study. *Language, cognition and neuroscience*, 1–17.
Friederici, A. D., Kotz, S. A., Werheid, K., Hein, G., & von Cramon, D. Y. (2003). Syntactic comprehension in Parkinson's disease: Investigating early automatic and late integrational processes using event-related brain potentials. *Neuropsychology, 17*(1), 133.
Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7* (4), 493–498. https://doi.org/10.1111/2041-210X.12504
Grey, S., Tanner, D., & van Hell, J. G. (2017). How right is left? Handedness modulates neural responses during morphosyntactic processing. *Brain Research, 1669*, 27–43. https://doi.org/10.1016/j.brainres.2017.05.024
Haebig, E., Weber, C., Leonard, L. B., Deevy, P., & Tomblin, J. B. (2017). Neural patterns elicited by sentence processing uniquely characterize typical development, SLI recovery, and SLI persistence. *Journal of Neurodevelopmental Disorders, 9*(1), 22. https://doi.org/10.1186/s11689-017-9201-1
Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition, 4*(2), 123–141.
Helenius, P., Parviainen, T., Paetau, R., & Salmelin, R. (2009). Neural processing of spoken words in specific language impairment and dyslexia. *Brain, 132*(7), 1918–1927. https://doi.org/10.1093/brain/awp134
Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*(2), 159–201.
Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of memory and language, 52*(2), 205–225. https://doi.org/10.1016/j.jml.2004.10.002
Kim, A. E., Oines, L., & Miyake, A. (2018). Individual Differences in Verbal Working Memory Underlie a Tradeoff Between Semantic and Structural Processing Difficulty During Language Comprehension: An ERP Investigation. *Journal of experimental psychology. Learning, memory, and cognition, 44*(3), 406–420. https://doi.org/10.1037/xlm0000457
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13
Lenth, R. V. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means. In (Version R package version 1.8.5)*.
Leonard, L. B. (2014). *Children with specific language impairment* (Second edition. ed.). The MIT Press.
Liu, T., de Schotten, M. T., Altarelli, I., Ramus, F., & Zhao, J. (2021). Maladaptive compensation of right fusiform gyrus in developmental dyslexia: A hub-based white matter network analysis. *Cortex, 145*, 57–66.
Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials [Technology Report]. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00213
Luck, S. J. (2014). *An introduction to the event-related potential technique* (Second edition. ed.). The MIT Press.
Mak, D. L.-W. (1991). *The acquisition of classifiers in Cantonese* [Doctoral Dissertation, University of Reading].
Matthews, S., & Yip, V. (2011). Cantonese: A Comprehensive Grammar. Florence: Routledge. Doi: 10.4324/9780203835012.
McGregor, K. K. (2020). How We Fail Children With Developmental Language Disorder. *Language, Speech, and Hearing Services in Schools*, *51*(4), 981-992. doi:10.1044/2020_LSHSS-20-00003.
Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and Span: Working Memory Capacity Impacts the Use of Animacy but Not of World Knowledge during Spoken Sentence Comprehension. *Journal of cognitive neuroscience, 22*(12), 2886–2898. https://doi.org/10.1162/jocn.2009.21400
Neville, H. J., Coffey, S. A., Holcomb, P. J., & Tallal, P. (1993). The Neurobiology of Sensory and Language Processing in Language-Impaired Children. *Journal of Cognitive Neuroscience, 5*(2), 235–253. https://doi.org/10.1162/jocn.1993.5.2.235
Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z
Palmer, J. A., Kreutz-Delgado, K., & Makeig, S. (2012). AMICA: An adaptive mixture of independent component analyzers with shared components. *Swartz Center for Computatonal Neursoscience, University of California San Diego, Tech. Rep.*
Pijnacker, J., Davids, N., van Weerdenburg, M., Verhoeven, L., Knoors, H., & van Alphen, P. (2017). Semantic Processing of Sentences in Preschoolers With Specific Language Impairment: Evidence From the N400 Effect. *Journal of Speech, Language, and Hearing Research, 60*(3), 627–639. https://doi.org/10.1044/2016_jslhr-l-15-0299
Pizzioli, F., & Schelstraete, M.-A. (2011). Lexico-semantic processing in children with specific language impairment: The overactivation hypothesis. *Journal of Communication Disorders, 44*(1), 75–90. https://doi.org/10.1016/j.jcomdis.2010.07.004
Plante, E., Tucci, A., Nicholas, K., Arizmendi, G. D., & Vance, R. (2018). Effective Use of Auditory Bombardment as a Therapy Adjunct for Children With Developmental Language Disorders. *Lang Speech Hear Serv Sch, 49*(2), 320–333. https://doi.org/10.1044/2017_lshss-17-0077
Popescu, M., Fey, M. E., Lewine, J. D., Finestack, L. H., & Popescu, E. A. (2009). N400 responses of children with primary language disorder: Intervention effects. *Neuroreport, 20*(12), 1104–1108. https://doi.org/10.1097/WNR.0b013e32832e9c97
Purdy, J. D., Leonard, L. B., Weber-Fox, C., & Kaganovich, N. (2014). Decreased sensitivity to long-distance dependencies in children with a history of specific language impairment: Electrophysiological evidence. *Journal of Speech, Language, and Hearing Research, 57*(3), 1040–1059. https://doi.org/10.1044/2014_jslhr-l-13-0176
Qian, Z., & Garnsey, S. (2016). An ERP study of the processing of Mandarin classifiers. In H. Tao (Ed.), *Integrating Chinese linguistic research and language teaching and learning*. John Benjamins Publishing Company.
R Core Team. (2023). R: A Language and Environment for Statistical Computing. In R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Sabisch, B., Hahne, A., Glass, E., von Suchodoletz, W., & Friederici, A. D. (2006). Lexical-semantic processes in children with specific language impairment. *Neuroreport, 17* (14), 1511–1514. https://doi.org/10.1097/01.wnr.0000236850.61306.91
Sheng, L., Yu, J., Su, P. L., Wang, D., Lu, T.-H., Shen, L., Hao, Y., & Lam, B. P. W. (2023). Developmental language disorder in Chinese children: A systematic review of research from 1997 to 2022. *Brain and Language, 241*, Article 105268. https://doi.org/10.1016/j.bandl.2023.105268
Stokes, S. F., & So, L. K. H. (1997). Classifier use by language-disordered and age-matched Cantonese-speaking children. *Asia Pacific Journal of Speech, Language and Hearing, 2*(2), 83–101. https://doi.org/10.1179/136132897805577413
T'sou, B., Lee, T., Tung, P., Man, Y., Chan, A., To, C., & Chan, Y. (2006). Hong Kong Cantonese oral language assessment scale. *Hong Kong: City University of Hong Kong*.
Tanner, D. (2019). Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach. *Cortex, 111*, 210–237. https://doi.org/10.1016/j.cortex.2018.10.011
Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. E. E. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism (Cambridge, England), 16*(2), 367–382. https://doi.org/10.1017/S1366728912000302
Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*(1), 231–270. https://doi.org/10.1016/j.cognition.2003.10.008
Ullman, M. T. (2016). The Declarative/Procedural Model: A Neurobiological Model of Language Learning, Knowledge, and Use. In G. Hickok, & S. L. Small (Eds.), *Neurobiology of Language* (Vol. 76, pp. 953–968). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00076-6.
Ullman, M. T., Clark, G. M., Pullman, M. Y., Lovelett, J. T., Pierpont, E. I., Jiang, X., & Turkeltaub, P. E. (2024). The neuroanatomy of developmental language disorder: A systematic review and meta-analysis. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-024-01843-6
Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2020). The neurocognition of developmental disorders of language. *Annual Review of Psychology, 71*, 389–417.
Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex, 41*(3), 399–433.
Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. *Neuroscience & Biobehavioral Reviews, 51*, 205–222.
Weber-Fox, C., Leonard, L. B., Wray, A. H., & Bruce Tomblin, J. (2010). Electrophysiological correlates of rapid auditory and linguistic processing in adolescents with specific language impairment. *Brain and Language, 115*(3), 162–181. https://doi.org/10.1016/j.bandl.2010.09.001
Winkler, I., Debener, S., Müller, K. R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*.
Wu, S., Zhao, J., de Villiers, J., Liu, X. L., Rolfhus, E., Sun, X., Li, X., Pan, H., Wang, H., Zhu, Q., Dong, Y., Zhang, Y., & Jiang, F. (2023). Prevalence, co-occurring difficulties, and risk factors of developmental language disorder: First evidence for Mandarin-speaking children in a population-based study. *LancetReg Health West Pac, 34*, Article 100713. https://doi.org/10.1016/j.lanwpc.2023.100713
Zhang, Y., Zhang, J., & Min, B. (2012). Neural dynamics of animacy processing in language comprehension: ERP evidence from the interpretation of classifier–noun combinations. *Brain and Language, 120*(3), 321–331.
Zhao, J., Zhao, Y., Song, Z., Thiebaut de Schotten, M., Altarelli, I., & Ramus, F. (2023). Adaptive compensation of arcuate fasciculus lateralization in developmental dyslexia. *Cortex, 167*, 1–11. https://doi.org/10.1016/j.cortex.2023.05.017