**The role of iconicity in children's production of adverbial clauses**

Shijie Zhang,[a] Silke Brandt,[b] Anna Theakston[a]

[a] Division of Psychology, Communication and Human Neuroscience, The University of Manchester, Manchester, UK. Email: shijie.zhang@manchester.ac.uk, anna.theakston@manchester.ac.uk

[b] Department of Linguistics and English Language, Lancaster University, Lancaster, UK. Email: s.brandt@lancaster.ac.uk

Corresponding author: Shijie Zhang, Division of Psychology, Communication and Human Neuroscience, The University of Manchester, Manchester, M13 9PL, UK. Email: shijie.zhang@manchester.ac.uk

Declarations of interest: None

**The role of iconicity in children's production of adverbial clauses**

**Abstract**

Young children's comprehension of adverbial clauses is significantly affected by iconicity, which refers to whether the order of information in the sentence reflects the order of events in the real world. In contrast, clause order (main-subordinate vs. subordinate-main) and input frequency of specific adverbial clauses do not seem to play independent roles (De Ruiter et al., 2018). The present study tests children's sentence production across four different connective types (*after, before*, *because*, *if*) to determine whether the factors that underpin the comprehension of adverbial clauses also apply to production, which involves utterance planning and articulation. 42 four-year-old, 42 five-year-old, and 22 eight-year-old monolingual English-speaking children, along with 20 adult controls, completed a sentence completion task. The results showed that both four- and five-year-olds produced all type of sentences in iconic order ("She builds a tower, before she breaks her train"; "After she builds a tower, she breaks her train") more accurately than in non-iconic order. This suggests that while comprehension and production likely impose different demands on children, iconicity as a general semantic strategy benefits children's early processing of adverbial clauses. Moreover, the effect of iconicity persisted in older children's production, but only for their *because*- and *if*-sentences, which could be related to their semantic complexity and the pragmatic properties they encode.

**Keyword:** Complex syntax, Language production, Language development, Iconicity

**1. Introduction**

At about three years of age, children start producing adverbial clauses that allow them to express temporal, causal and conditional relationships between events (e.g., temporality–*after*, *before*; causality and conditionality–*because*, *if*), such as "sit down now on the train (be)fore it goes" and "I'm just putting it on because I'm cold" (De Ruiter et al., 2021). However, the early correct use of adverbial clauses in spontaneous speech does not necessarily indicate that children acquire all types of adverbial clauses equally well at the same time, nor that they fully understand the relationship(s) that each connective expresses (Clark, 1971).

Several experiments have found that children have difficulty comprehending adverbial clauses even when they enter formal school education (e.g., Blything & Cain, 2016; Blything et al., 2015; De Ruiter et al., 2018, 2020). The difficulties in comprehension have been mainly observed in understanding the mapping between the events in the linguistic form and in the real world. Specifically, in English and related languages, adverbial clauses can occur in two orders: main-subordinate and subordinate-main. Compare Example (1a) with (1b), and (2a) with (2b). In (1a) and (2b), the order of events in the clauses matches the order of events in the real world – often referred to as iconic. In contrast, in (1b) and (2a), the event that happens later in the real world is mentioned first – it is non-iconic. In general, young children appear to find non-iconic sentences more difficult to comprehend compared to iconic ones (e.g., Blything & Cain, 2016; Blything et al., 2015; De Ruiter et al., 2018, 2020).

(1a) [She builds a tower]$_{\text{MAIN}}$ [before she breaks her train]$_{\text{SUBORDINATE}}$ Iconic

(1b) [Before she breaks her train]$_{\text{SUBORDINATE}}$ [she builds a tower]$_{\text{MAIN}}$ Non-iconic

(2a) [She finds the nests]$_{\text{MAIN}}$ [because she climbs the ladder]$_{\text{SUBORDINATE}}$ Non-iconic

(2b) [Because she climbs the ladder]$_{\text{SUBORDINATE}}$ [she finds the nests]$_{\text{MAIN}}$ Iconic

In addition, children's comprehension difficulties have also been linked to the semantic complexity of the connective (e.g., Clark, 1971; Blything & Cain, 2019), the syntactic structure (or processing load - whether the main clause precedes or follows the subordinate clause) (e.g., Diessel, 2005;

Zhang et al., 2023), and the frequency of the specific connective-clause order combination (e.g., Ambridge et al., 2015; De Ruiter & Theakston, 2017).

To date, it is unclear whether the factors that underpin children's comprehension of adverbial clauses also apply to production. Comprehension processes are inherently different from production processes: Comprehension requires the individual to construct a mental representation of the events based on their interpretation of what they hear, while in production the individual needs to find the appropriate language structure to articulate their mental representation of the events. Previous studies of production reveal some apparent differences from comprehension, which will be explored in more detail below. The impact of iconicity (see examples above) has been found in both comprehension and in production (e.g., Blything & Cain, 2016, 2019; Blything et al., 2015; De Ruiter et al., 2018, 2020). Yet, to our knowledge, no previous studies have investigated children's production of temporal, causal and conditional adverbial clauses within a single study (*after*, *before*, *because*, *if*) using the same methodology to systematically test the factors of iconicity, semantic complexity, syntactic structure, and frequency of the specific connective-clause order combination (see De Ruiter et al. (2021) Appendix E for an overview of experimental studies on children's processing of different types of adverbial clauses).

To provide a comprehensive picture of the differences and commonalities of complex sentence comprehension and production, the present study investigates children's production of four types of adverbial sentences (*after*, *before*, *because*, *if*), to test the predictions of three theoretical accounts of complex sentence processing in children. In the following sections, we discuss the details of the theoretical accounts and review the empirical evidence for each of the accounts by highlighting the differences between comprehension and production. In addition, we take individual differences in children's language abilities, memory and inhibitory control into account as control measures.

### *1.1. Theoretical accounts*

#### *1.1.1. Semantic account*

Clark (1971) proposed that children's early complex sentence processing relies on an "order-of-mention" strategy. This means that children assume a direct mapping between the order of events in the real world and in the linguistic form, and as a result, children should find iconic sentences easier to process than non-iconic sentences (see examples (1a, b) and (2a, b)).

The results from child comprehension studies are converging to support the "order-of-mention" strategy (e.g., Blything & Cain, 2016; Blything et al., 2015; Clark, 1971; De Ruiter et al., 2018, 2020; French & Brown, 1977; Stevenson & Pollitt, 1987; Wagner & Holt, 2023). For example, Blything et al. (2015) and Blything and Cain (2016) investigated three- to seven-year-old children's comprehension of *before*- and *after*-sentences using a forced-choice picture selection task. In the task, children were first asked to watch two animations that represented the actions of two clauses (e.g., put on the sandals; eat the burger). They then selected one out of two animations that represented the action that happened first (Blything et al. 2015), or that happened last (Blything & Cain, 2016), according to the test sentence that was played to them (e.g., "Before he ate the burger, he put on the sandals*"*). Blything et al. (2015) found that three- to four-year-olds comprehended both iconic *before*- and *after*-sentences more accurately than non-iconic ones, while Blything and Cain (2016) found the effect of iconicity for children up until the age of seven.

De Ruiter and colleagues' (2018) study further extended evidence of a comprehension advantage for iconic sentences to causal/conditional connectives (*because* and *if*). They also used the forced-choice picture selection task but with a slightly different design. Children were asked to select one out of two separate two-picture sequences (e.g., sequence 1: break her train + build a tower; sequence 2: build a tower + break her train) that matched a spoken sentence (e.g., "She builds a tower before she breaks her train"). They found that five-year-old children comprehended all types of sentences more accurately when they were iconic (i.e., *before*-sentences in main-subordinate order; *after*-, *because*- and *if*-sentences in subordinate-main order; see Table 1 for examples).

**Table 1**

*The predictions from three theoretical accounts concerning children's production of four types of adverbial clauses*

| Connective | Clause order | Sentence | Section 1.1.1: Semantic account | Section 1.1.2: Processing-based account | Section 1.1.3: Frequency-based account |
|---|---|---|---|---|---|
| *Before* | subordinate-main | ***Before*** *she breaks her train, she builds a tower.* | | | |
| | main-subordinate | *She builds a tower* ***before*** *she breaks her train.* | ✓ | ✓ | ✓ |
| *After* | subordinate-main | ***After*** *she builds a tower, she breaks her train.* | ✓ | | ✓ |
| | main-subordinate | *She breaks her train* ***after*** *she builds a tower.* | | ✓ | |
| *Because* | subordinate-main | ***Because*** *she climbs the ladder, she finds the nests.* | ✓ | | |
| | main-subordinate | *She finds the nests,* ***because*** *she climbs the ladder.* | | ✓ | ✓ |
| *If* | subordinate-main | ***If*** *she climbs the ladder, she finds the nests.* | ✓ | | ✓ |
| | main-subordinate | *She finds the nests,* ***if*** *she climbs the ladder.* | | ✓ | |

*Note*. Tick mark refers to the clause order that the account would predict to be easiest to process: The semantic account predicts that iconic sentences are easier to process than non-iconic sentences; the processing-based account predicts that main-subordinate sentences are easier to process than subordinate-main sentences; the frequency-based account predicts that high frequency connective-clause order combinations are easier to process.

The effect of iconicity has also been observed in child production studies, but the effect varies depending on the type of connective. Blything and Cain (2019) investigated three- to six-year-old children's production of *before*- and *after*-sentences using a sentence repetition task and a blocked elicited production task. In the sentence repetition task, children were first asked to listen to the target sentence (e.g., "Before he ate the burger, he put on the sandals") and watch the two animations that represented the actions of two clauses (e.g., put on the sandals; eat the burger). They were then asked to produce an exact copy of the target sentence to the experimenter. In the blocked production task, children were only asked to describe what happened using *before* or *after* after watching the two animations depicting the actions of two clauses. The results of both tasks showed a significant effect of iconicity in all age groups on children's production of *after*-sentences, but not *before*-sentences.

Blything and Cain's (2019) production results are not fully compatible with previous comprehension results targeting a similar age group, in which the effect of iconicity has been found

for both *before* and *after*-sentences (Blything & Cain, 2016; Blything et al., 2015; De Ruiter et al., 2018, 2020; but note that Wagner and Holt (2023), who investigated seven- to twelve-year-old children's comprehension of *before*- and *after*-sentences, only found the effect of iconicity for *after*-sentences as well). Blything and colleagues suggested that the asymmetry between production and comprehension was due to the greater involvement of children's language knowledge in production. *Before* is used more consistently as a temporal connective than *after*, and the more consistent form-meaning relationship makes *before* easier to produce in general. In contrast, *after* can also be used in other constructions such as "look after" and "after all" in addition to being a temporal connective (De Ruiter et al., 2021; Leech et al., 2014), which could require greater planning and processing effort in production, particularly when the event order is reversed.

Some earlier work has looked at children's production of causal/conditional connectives such as *because*, *so*, and *if*, but there is no consistent evidence for the effect of iconicity (e.g., Emerson & Gekoski, 1980; French, 1988; Homzie & Gravitt, 1976; Johnson & Chapman, 1980). Homzie and Gravitt (1976) used a sentence imitation task to test three- to five-year-old children's production of sentences linked by causal connectives such as *because* and *so* (e.g., "The man fell down because he slipped on a banana peel"). They found that although children overwhelmingly produced the sentences in the same order initially presented to them, they were slightly more likely to reverse non-iconic sentences into an iconic order in their production. However, Emerson and Gekoski (1980) and Johnson and Chapman (1980) used a sentence imitation task to test the same types of sentences but did not find any effect of iconicity in six-to-eleven-year-olds. French (1988) similarly found no evidence for an iconicity preference with *because*- and *so*-sentences in five-to-eight-year-olds. French (1988) used a sentence completion task, in which children heard a story describing causally related events (e.g., Jane was helping her mother carry the groceries home; Jane tripped; She dropped the groceries; The eggs broke), and they were then asked to complete one sentence based on the story (e.g., "Jane dropped the groceries because/so…"). Since these production studies used different methodologies to test children from different age groups, it is

difficult to fully interpret the discrepancies and conflicting findings between them. Moreover, it is unclear whether these studies were well-powered, given the small number of participants and the limited number of test sentences. For example, Homzie and Gravitt (1976) tested only 13 children with their production of causal sentences, while French (1988) tested 36 children aged five to eight years, with each child completing only nine test sentences.

To summarise, there is clear evidence that children rely on the "order-of-mention" strategy (i.e., iconicity) in their comprehension of adverbial clauses (*before*, *after*, *because*, *if*). In production, iconicity has also been found to impact children's processing of temporal *after*-sentences. However, the role of iconicity in children's production of causal/conditional adverbial clauses is far from clear. This is in part due to the use of varying methodologies and age groups.

*1.1.2. Processing-based account*

The syntactic form of complex sentences has been proposed to influence their ease of processing. Wasow (1997) proposed that speakers tend to arrange sentence constituents in orders that do not require an early commitment to a specific syntactic structure. This is to avoid having an extensive amount of utterance planning (see also Arnold et al., 2000; Wasow, 2002). In the case of adverbial sentences, the initial occurrence of a subordinate clause (see Table 1 for examples) requires an early commitment to produce a structure consisting of at least two clauses and therefore is thought to increase the amount of utterance planning. On the other hand, when the subordinate clause follows the main clause, the link between the main and subordinate clause is only established after the main clause has been processed.

Similarly, from the hearer's perspective, placing the subordinate clause later is expected to facilitate the efficiency of parsing. Based on Hawkins' "performance theory of order and constituency" (Hawkins, 1990, 1992, 1994), Diessel (2005) argued that when processing sentences with subordinate-main clause order, the listener needs to keep the subordinate clause in memory until the main clause is encountered, which makes it harder to parse. In contrast, sentences with main-subordinate clause order allow the main clause to be fully parsed before parsing the

subordinate clause. Based on this account, children should find adverbial sentences with main-subordinate clause order easier to comprehend and produce.

The results from child comprehension studies did not find an independent main effect of clause order for the connectives *before*, *after*, *because* and *if*, that is, neither clause order was comprehended better than the other (e.g., Blything & Cain, 2016; De Ruiter et al. 2018, 2020). Instead, these studies found an interaction between clause order and connective type, leading to the effect of iconicity: *before*-sentences were better comprehended in the iconic main-subordinate order, while *after*-, *because*- and *if*-sentences were better comprehended in the iconic subordinate-main order (see Table 1 for examples). Junge et al. (2015), who tested three- and five-year-old children's comprehension of *when*-clauses using an act-out task (e.g., "The cat is driving, when the sheep is swimming"), also found that young children were not sensitive to clause order, but rather to information structure of the clauses. That is, regardless of the clause order of *when*- and main clauses, children tended to first act out the clause that encoded previously mentioned information.

In production, however, Zhang et al. (2023) investigated three- to five-year-old children's recall of *when*-sentences (e.g., "When Pig is swimming, Little Panda is drinking very fast"), and found that three-year-olds tended to reverse *when*-main sentences to main-*when*, but this preference decreased with age. These findings were interpreted as offering some support to the processing-based account; Zhang et al. (2023) suggest that younger children, who have lower processing capacities than adults, prefer main-*when* sentences because they are easier to plan. However, the *when*-sentences in Zhang et al. (2023) encoded simultaneous events (i.e., there was no clear order of events), which is different from *before*-, *after*-, *because*-, and *if*-sentences that encode sequential events. When producing the latter, processing demands are unlikely to be the sole factor influencing the positioning of adverbial clauses, as semantic factors (e.g., iconicity) may also play a role. As described above, Blything and Cain (2019) investigated the production of *before*- and *after*-sentences and found an effect of iconicity, rather than a main effect of clause order.

*1.1.3. Frequency-based account*

Usage-based approaches see the input frequency of grammatical structures as an important factor in children's language acquisition (see Ambridge et al., 2015, for an overview). In the case of adverbial clauses, a frequency-based account would predict that high frequency connectives and connective-clause order combinations in the input would lead to earlier and more accurate comprehension and production, other things being equal.

Based on the data from two high-density developmental corpora of parent-child interactions from two British English-acquiring children (2;00–4;07), De Ruiter et al. (2021)[1] found that in both children's and their mothers' speech:

- *because*- and *if*-sentences occur more frequently than *after*- and *before*-sentences;
- *after*- and *if*-sentences occur more frequently in (iconic) subordinate-main order;
- *before*-sentences occur more frequently in (iconic) main-subordinate order, and *because*-sentences occur overwhelmingly in (non-iconic) main-subordinate order.

Based on these frequency distributions, children should find *because*- and *if*-sentences easier to process than *after*- and *before*-clauses, and should find high frequency connective-clause order combinations easier to process. Note that there is a large overlap in the predictions of the frequency-based account and the semantic account for connective-clause order combinations. As shown in Table 1, these two accounts provide the same predictions for *before*-, *after*- and *if*-sentences; that is, the iconic clause order occurs more frequently in child and child-directed speech. The only differing predictions emerge for *because*-sentences: the subordinate-main order is iconic, while the main-subordinate order is the most frequently used in child and child-directed speech.

In comprehension, the only two studies that have investigated these four connectives together found that young children comprehended *before*-sentences more accurately than the other three types of sentences and had quicker response times (RTs) to *before*- and *after*-sentences

[1] The distribution of different types of sentences may differ between child-directed and adult-directed speech, as well as between different genres (Roland et al., 2007; Zhang, 2022). To better investigate the effect of input frequency, we cited results from the corpora of parent-child interactions, rather than from the general language corpora (Diessel, 2001; Diessel, 2008).

compared to *because*- and *if*-sentences (De Ruiter et al., 2018, 2020). Moreover, sentences in iconic clause orders (e.g., *because*-sentences in subordinate-main order) were comprehended better than non-iconic ones overall. Taken together, the comprehension findings lend only limited support to the frequency-based account.

It has been suggested that the inconsistent findings could be partly due to the fact that the sentences used in experiments are different from those that children typically hear and use in everyday interaction with their caregivers (De Ruiter et al., 2018, 2021; Lemen et al., 2021, 2024). Frequency effects can occur at different levels of language, and children are sensitive not only to sentence structure but also to the pragmatic contexts associated with each type of sentence. De Ruiter et al. (2021) found that in everyday speech, children mostly hear and produce *before*- and *after*-sentences that express the pragmatic relationship which Sweetser (1990) calls "Content" (i.e., the clauses present real-world events independent of the speaker; see examples in Table 1). However, they mostly hear and produce *because*-sentences that express "Speech Act" relationships (i.e., the subordinate clause provides the speaker's reason for the speech act, e.g., "Don't step in puddles, because you are getting your shoes wet"). For *if*-sentences, the "Speech Act" relationship is used more frequently in main-subordinate order, while the "Content" relationship is used more frequently in subordinate-main order. Moreover, a high number of *if*-sentences contain "idiomatic phrases" such as "if you want" and "if you like" *(*Lemen et al., 2021). The pragmatic properties of the *because*- and *if*-sentences that children hear frequently in everyday interaction are somewhat different from the "Content" sentences used in experiments, which could explain a lack of clear frequency effects in experiments examining children's comprehension (Lemen et al., 2024).

To the best of our knowledge, there are no experimental studies investigating children's production of the four connectives (*before*, *after*, *because* and *if*) within a single study. We would expect that children are more likely to produce sentences if the sentences encode the structures and pragmatic contexts that they are familiar with in everyday interaction with their caregivers. Corpus findings seem to confirm this prediction (De Ruiter et al., 2021; Lemen et al., 2021). However,

spontaneous speech only provides part of the picture, that is, what children like to talk about. The other part of the picture, what they are able to say, also requires investigation.

*1.1.4. Individual differences*

There is a general trend indicating that children's language processing, such as the comprehension and production of adverbial clauses, improves with age (e.g., Blything & Cain, 2016, 2019; Blything et al., 2015; De Ruiter et al., 2018, 2020). However, children within the same age group may still show some variations in their processing. Some studies suggest that these variations are linked to children's individual differences in memory capacity, general language knowledge (e.g., vocabulary), or inhibitory control (see Kidd, 2013 and Kidd et al., 2018 for a detailed review). Therefore, we included individual differences in this study as control measures to examine whether they would account for additional variance in the children's performance in this specific experimental study. Previous research has observed that individual memory capacity makes an independent contribution to children's comprehension of adverbial clauses. Three- to seven-year-old children with higher memory capacities comprehended *before*- and *after*-sentences in both orders more accurately than those with lower memory capacities (Blything et al., 2015; Blything & Cain, 2016), though the findings for production are less clear (Blything & Cain, 2019). However, the exact measures used to capture memory capacity and distinct tasks used to assess children's knowledge may influence the results. For example, De Ruiter and colleagues (2018, 2020) did not find any independent contribution of memory to three to five-year-old children's comprehension of *before*-, *after*-, *because*- and *if*-sentences.

Previous comprehension studies did not observe an independent contribution of inhibitory control on children's comprehension of *before*-, *after*-, *because*- and *if*- sentences (De Ruiter et al., 2020; De Ruiter et al., 2018). To the best of our knowledge, there are no experiments investigating the relationship between children's individual differences in inhibitory control and production of adverbial clauses.

Please note that we do not expect a direct link between individual differences and the specific theoretical predictions regarding the production of adverbial clauses (see Table 1). Specifically, the semantic and frequency-based accounts do not make explicit predictions about the links between sentence processing efforts and individual differences, although there is no a priori reason why the effects of iconicity and frequency could not be modulated by individual differences (see Blything & Cain, 2016; Blything et al. 2015). The processing-based account makes an explicit prediction about the processing efforts related to memory capacity (i.e., when processing subordinate-main sentences, the subordinate clause needed to be kept in memory until the main clause is encountered). However, as discussed above, clause order (main-subordinate vs. subordinate-main) is unlikely to be the sole factor influencing the processing efforts of adverbial clauses.

### *1.2. The present study*

The present study tested three- to five-year-old children's production of four types of adverbial clauses (*after*, *before*, *because*, *if*) using a sentence completion task, comparing them to seven- to nine-year-old children[2] and adult controls. By examining all four types of adverbial clauses within a single study, we can tease apart the predictions of three theoretical accounts of complex sentence processing in children (see Table 1 for a comparison of the theoretical accounts). We also looked at the links between children's individual differences in language ability, memory, and inhibitory control in relation to their sentence production. Based on the existing literature, we formulated the following main hypotheses in relation to both accuracy and speed of responses:

- Based on the semantic account, iconic sentences (i.e., *before*-sentences in main-subordinate order; *after*-, *because*- and *if*-sentences in subordinate-main order) are easier to produce than non-iconic sentences.

[2] The initial study plan only included three- to five-year-olds, as well as adult controls (see pre-registration). However, upon finding that five-year-olds' performance was far from that of adults, a smaller sample of eight-year-olds was tested to obtain a more comprehensive understanding of the developmental trajectory.

- Based on the processing-based account, main-subordinate sentences are easier to produce than subordinate-main sentences.
- Based on the frequency-based account, *before*- and *because*-sentences are easier to produce in main-subordinate order, while *after*- and *if*-sentences are easier to produce in subordinate-main order.
- In comparing each type of sentence, *before*-sentences are easier to produce than *after*-sentences (semantic account); *because*- and *if*-sentences are easier to produce than *after*- and *before*-sentences (frequency-based account).
- Performance improves with age.
- Language ability, memory, and inhibitory control correlate positively with children's production of adverbial clauses.
- Language ability, memory, and inhibitory control make independent contributions to children's production of adverbial clauses.

This study was pre-registered in the Open Science Framework repository, which specified research hypotheses, study design, sampling and analysis plan (https://osf.io/zkvje/?view_only=4e39759c33a84250a2845450805417e8 (anonymous link)).

## 2. Methods and materials

### *2.1. Participants*

106 children and 20 adults were included in the final sample. The children were recruited through nurseries and primary schools in the North West of England, and through the Universities' (anonymised) child participant databases. They were monolingual, native speakers of English without any known history of speech or language problems or developmental delays. Of the 106 children, 42 were between 3;06 and 4;06 years old (M = 48.06 months, SD = 3.03 months, 26 girls), 42 were between 4;07 and 5;07 years old (M = 60.95 months, SD = 3.16 months, 19 girls)[3], 22 were

[3] The sample size of four- and five-year-olds was decided based on a priori power analysis based on the data from De Ruiter et al. (2018), which tested four- and five-year-olds' comprehension of four adverbials (*after*,

between 7;04 and 9;11 years old (M = 104.27 months, SD = 8.26 months, 11 girls). We will refer to the first group as the four-year-olds, the second group as the five-year-olds, and the third group as the eight-year-olds. Two of the 42 four-year-olds only completed the first session, as they refused, or were not able to do the second session. One of the 42 four-year-olds took part in both sessions but refused to do the Flanker task. In addition, four additional four-year-olds and one additional five-year-old were tested but excluded from the final sample because they did not cooperate in the first session and refused to do the second session, or because they did not understand the instructions and therefore needed extra help to prompt or name the events on the screen. The adult participants (M = 18.65 years, SD = 1.01 years, 14 women) were undergraduate students from the University (anonymised), and native speakers of English.

### *2.2. General procedure*

Children were tested in a quiet area in their nurseries, primary schools, and in the Universities' child labs. In addition to the main sentence completion task, children completed five tasks to assess language ability, short-term and working memory and inhibitory control. The tasks were spread over two sessions on two separate days within a 2-week window, but one four-year old child completed two sessions in a single day. Each session lasted around 40 minutes. Children completed half of all items of the sentence completion task in Session one, and the other half in Session two. The two language ability tasks were administered in Session one. The short-term and working memory tasks and the inhibitory control task were administered in Session two. In both sessions, children always completed a nonverbal task before completing the sentence production task, followed by other verbal tasks. Adults were tested in the University lab, and they only did the sentence completion task. They completed all items in one session, with a short break.

*before*, *because*, *if*) in two clause orders (main-subordinate, subordinate-main). In order to obtain at least 80% power by adopting the effect sizes from published data, 40 children were needed in each age group.

### *2.3. Sentence completion task*

Children's and adults' production of adverbial clauses was tested using a sentence completion task (French, 1988). In the task, participants were presented with three pictures that depicted everyday events happening in a sequence, and then were asked to complete a sentence after the experimenter's prompt. Table 2 provides an example for *before-*/*after*-sentences, and Table 3 provides an example for *because-*/*if*-sentences. The experimenter's prompt always described the middle event along with a connective, and participants were required to use either the first or the last event to complete the sentence. Both response accuracy and RTs were measured.

**Table 2**

*An example of before-*/*after*-sentences

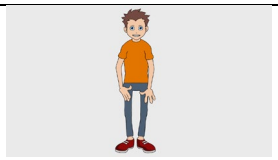
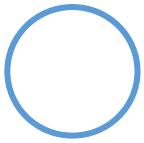
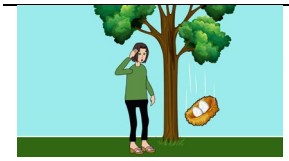
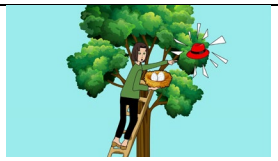
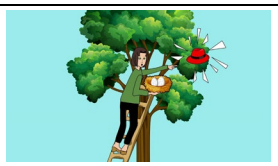
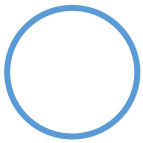
| Visual presentation | Auditory presentation |
|---|---|
|  | |
| (Blank screen) | "Look and listen carefully!<br>Finish the story, but wait for the beep!" |
|  | "First" |
|  | "Next" |
|  | "Last" |
|    | "Before he plays the big drum, ... (he drinks water)"/<br>"After he plays the big drum, ... (he eats a pear)"<br><br>*beep* |
|  | |

**Table 3**

*An example of because-/if*-sentences

| Visual presentation | Auditory presentation |
|---|---|
|  | |
| (Blank screen) | "Look and listen carefully!<br>Finish the story, but wait for the beep!" |
|  | "First" |
|  | "Next" |
|  | "Last" |
|    | "She climbs the ladder, because/if ...<br>(she finds the nest)"<br><br>*beep* |
|  | |

*Note*. The picture story illustrates that the nest falling from the tree causes the girl to climb the ladder to put it back (i.e., "She climbs the ladder, because she finds the nest"), or that the nest falling from the tree creates a condition for the girl to climb the ladder to put it back (i.e., "She climbs the ladder, if she finds the nest").

### *2.3.1. Design*

The main child experiment had three factors: one between-subjects factor (AgeGroup), and two within-subjects factors (ConnectiveType, ClauseOrder), each with the following levels:

- AgeGroup: 4 years, 5 years (8 years, adults)
- ConnectiveType: *after*, *before*, *because*, *if*

- ClauseOrder: main-subordinate, subordinate-main

Please note that the initial study plan only included four- and five-year-olds in the AgeGroup (see pre-registration), with the intention to analyse the adult data separately; in addition, eight-year-olds were added later and not included in the pre-registration. However, as suggested by anonymous reviewers, to obtain a more direct and comprehensive comparison of developmental differences across age groups, we ran exploratory analyses that included eight-year-olds and adults as additional levels of AgeGroup in a single analysis.

*2.3.2. Materials*

*2.3.2.1. Test sentences.* 24 adverbial sentences were constructed, each containing a main and subordinate clause representing two events performed by a single actor (a boy in half of the sentences, and a girl in the other half). All sentences occurred in both clause orders: main-subordinate (e.g., "He plays the big drum, before... (he eats a pear)") and subordinate-main (e.g., "Before he plays the big drum … (he drinks water)"), resulting in 48 sentences. In total, there were eight conditions (i.e., four connective types x two clause orders) with six items per condition.

The two clauses in the *after*- and *before*-sentences were arbitrarily related (the events could happen in either order (e.g., play the big drum, eat a pear), only the connective indicated the correct order), while the *because*- and *if*-sentences were bi-directionally related (one event was the cause or condition for the other, but both events could serve as the cause/condition (e.g., climb the ladder, find the nest), only the connective indicated the correct order). All sentences expressed a "Content" relationship (i.e., the clauses presented real-world events independent of the speaker). All sentences had a pronominal subject (i.e., "he" or "she"), and the objects of the transitive verbs were always inanimate. Each event was described with a different verb in the present tense[4]. All event clauses were between 4 and 6 syllables long, and the experimenter's prompt (the clause that described the

[4] We did not use the past tense, as it is not suitable for *if*-sentences. When we use the past tense for *if*-sentences, this refers to hypothetical situations that could have happened in the past and are now likely to be impossible (e.g., "If I won the cup, I could sing a song in the stadium"). However, our study described events that are happening in the present (or future) (e.g., "If she wins a big cup, she sings a song"). Therefore, the present tense is the most suitable tense for all four types of sentences with the current design.

middle event) was always 5 syllables long. A list of experimental sentences and corresponding events can be found in Table S1 in the Supplemental Material.

*2.3.2.2. Visual and audio stimuli.* As shown in Table 2 and Table 3, for each test trial, a visual stimulus of the character that the short story was about (i.e., a picture of Tom or Sue) was first presented in the centre of the screen. Then, three visual stimuli depicting one event from a sequence of three events involving the same character were presented from left to right of the screen to indicate the order of the events. Each of the three visual event stimuli was presented for 3 seconds and accompanied by the audio-recorded sequence words "first", "next", and "last", which also functioned to indicate to participants the order of the events. At the end, the three visual stimuli depicting the three events were all presented in left-to-right orientation and accompanied by an audio prompt describing the middle event along with a connective. After each prompt, a 500-millisecond "beep" sound was added with a 250-millisecond pause in between. RT was measured from the offset of the "beep" sound. Once the participant completed the sentence, the screen showed a blue circle indicating successful completion of the trial. Part of the visual stimuli were borrowed from De Ruiter et al. (2018) and part of the stimuli were newly created in the same style. The audio prompts were recorded by a young female native speaker of British English in a quiet room. The peak amplitude of all audio prompts was normalized to 0dB using the software Audacity, version 3.1.13.

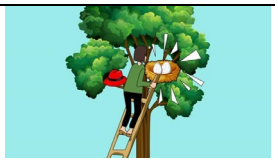
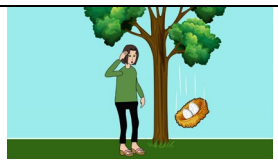
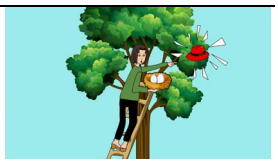
*2.3.2.3. Experimental lists*. Eight different experimental lists were constructed. Each list consisted of two sessions. Each sentence (N = 24) occurred once in each session (recall that each sentence occurred in two clause orders to create 48 sentences in total), with half of the sentences in each session being in main-subordinate clause order and the other half in subordinate-main clause order. The order of the trials within each session was pseudo-randomised, as was the position of the correct picture (first or last in a sequence of three events) on the screen. Specifically, there was a maximum of two consecutive trials containing the same connective type, a maximum of two

consecutive trials containing the same clause order, and a maximum of two consecutive trials that were iconic.

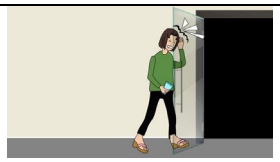
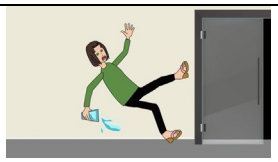
List 2 was created by swapping session 1 and session 2 of List 1. Lists 3 and 4 were the same as Lists 1 and 2, with the difference that all *after*-sentences were turned into *before*-sentences and vice versa, and all *if*-sentences were changed into *because*-sentences and vice versa. Lists 5-8 were the same as Lists 1-4, but all pictures sets in Lists 5-8 were the reversal of the pictures in Lists 1-4[5]. For example, for the sequences matching the sentence "Before he plays the big drum…", in List 1 the actor first drinks water, then plays the big drum, and last eats a pear, while in List 5, the actor first eats a pear, and last drinks water. In this case, the event used to complete the same test sentence in List 1 was different from that of List 5. For the *because*- and *if*-sentences, the pictures that represented the same event were slightly different in the two sequences, as the relation between the events changed slightly when the events were reversed (for examples, see Table 4). Participants were randomly assigned to one of the eight experimental lists.

**Table 4**

*Examples of the picture set occurring in two sequences*

| | | | |
|---|---|---|---|
| **List 1**: *She climbs the ladder, if …* *(she spots a red hat)* |  |  |  |
| | she spots a red hat | she climbs the ladder | she finds the nest |
| **List 5**: *She climbs the ladder, if …* *(she finds the nest)* |  |  |  |
| | she finds the nest | she climbs the ladder | she spots a red hat |

[5] The initial study plan only included four experimental lists (see pre-registration). However, as events in the current study could happen in either order (only the connective indicated the correct order), we added four more lists with the reversed order of the pictures in the first four lists to control for any order effects that might be present in the visual stimuli.

| | | | |
|---|---|---|---|
| **List 1:** *Because she slips to the ground, …* *(she bangs her head hard)* |  |  |  |
| | she spills the water | she slips to the ground | she bangs her head hard |
| **List 5:** *Because she slips to the ground, …* *(she spills the water)* |  |  |  |
| | she bangs her head hard | she slips to the ground | she spills the water |

*2.3.3. Procedure*

Children were asked to sit in front of a 13-inch MacBook Air. A puppet was positioned behind the laptop, but faced towards the children and the experimenter. The children were told that they were going to play a game, in which they would be presented with stories about two characters, Sue and Tom and that their task was to let the puppet know what was happening by verbally completing the stories after they heard a beep. The stimuli were presented using Microsoft PowerPoint on the laptop, and the sound was presented via the laptop speaker. The children's responses were recorded by an Olympus DS-3500 digital voice recorder.

Before the start of the experiment, there was a warm-up phase to familiarise the children with the task and the left-to-right reading of the picture sequences. In the warm-up, the presentation of the visual stimuli was not automatic, but manually controlled by the experimenter, which allowed the experimenter to explain the layout of the screen ("Look, this is Sue. Here we see the first thing she does", while pointing to the appropriate picture). The four warm-up trials contained four connective types and two clause orders (see examples (3)-(6)). While the first two warm-up trials only had two pictures that depicted the two events in the sentences, the other warm-up trials were like the test trials that provided three pictures.

(3) *After she bounces away*, … (*she hoovers the house*)

(4) *Because she feels really warm*, … (*she dives in the pool*)

(5) *She presses the button*, *if* … (*she hears the doorbell*)

(6) *He waves happily*, *before*… (*he watches TV*)

As shown in Table 2 and Table 3, at the beginning of each trial, the experimenter showed the visual stimulus of the character that the short story was about (i.e., a picture of Tom or Sue) and said "Ah, here's another story about Tom. Let's see what he/she does!" to focus the child's attention. Then, before each of the first three test trials and then after every three subsequent test trails, an audio instruction "Look and listen carefully! Finish the story, but wait for the beep!" was played to remind the child of what to do. If the child did not provide a verbal response, provided an ambiguous or unintelligible response, or repeated the prompt, the experimenter repeated the trial one more time. In 21.76% of these repetitions, the experimenter provided an extra instruction "this one (pointing to the first picture on the screen) or that one (pointing to the last picture on the screen)" to motivate the children to use either the first or the last picture on the screen to complete the sentence. In the Results section, we report the number of repetitions, and analysed children's responses both with and without trial replay to see if extra repetitions and instructions affected their production patterns (see Footnote 6). After every three test trials, a smiley face was shown on the screen to give the children a small break and encourage them to continue. It took 30-40 minutes for the children to complete both sessions of the sentence completion task.

The adult participants were tested using the same setup as the children, but no puppet was used. During the warm-up phase, they were instructed to look at the pictures and listen to the prompt carefully, and then complete a sentence after the "beep" sound. After the adults successfully completed the warm-up, they went through half of the test trials, followed by a short break, and then they completed the other half of the test trials. It took them 20-30 minutes to complete both sessions of the sentence completion task.

### *2.4. Individual measures*

#### *2.4.1. Language ability*

All children's receptive vocabulary was tested using the British Picture Vocabulary Scale III (BPVS III; Dunn et al., 2009) through a forced-choice picture selection from four illustrations. The BPVS test took 10-15 minutes.

Four- and five-year-olds' ability to recall and repeat sentences was tested using the "Recalling sentences in context" sub-test of the Clinical Evaluation of Language Fundamentals-Preschool 2 UK (CELF-Preschool 2 UK; Wiig et al., 2006). The test requires the child to recall and repeat lines from a picture story. As the story progresses, the number of morphemes, syntactic complexity, and number of prepositions in each item increases. The "Recalling sentences" sub-test of the Clinical Evaluation of Language Fundamentals-Fifth Edition (CELF-5; Wiig et al., 2013) was used to test eight-year-olds' ability to recall and repeat isolated sentences of increasing length and complexity. Both sentence recall tests took about 5 minutes.

*2.4.2. Short-term and working memory*

All children's short-term and working memory were tested using the forward digit span task from the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV, Wechsler, 2003), and the missing scan task from Roman et al. (2014). The forward digit span task requires the child to listen to strings of digits that increase in length and then repeat them back to the experimenter in the same order. The missing scan task requires the child to recall the missing animal that has just been removed from a set of animals. Each task took about 5 minutes.

*2.4.3. Inhibitory control*

All children's inhibitory control was tested using the computer-based Flanker task from Massonnié et al. (2019), which was programmed in Gorilla.sc. The Flanker task requires the child to focus on the direction of a central (left- or right-facing) fish on the screen and press the matching key on the keyboard (i.e., left or right), while ignoring the flanking distractor fish. The task took 5-10 minutes.

## 3. Results

### *3.1. Verbal data coding*

*3.1.1. Verbal response accuracy*

The verbal responses were transcribed and coded into the following categories based on accuracy:

(a) Correct response: the verbal response matched the correct picture on the screen (e.g., to complete the sentence “Before he plays the big drum…”, the participant answered “he drinks water”, or a similar answer that enabled the experimenter to identify the first picture in Table 2 such as “has a little drink”);

(b) Reversal response: the verbal response matched the reversed picture on the screen (e.g., “he eats a pear”, or a similar answer that enabled the experimenter to identify the last picture in Table 2 such as “eating apple”);

(c) Other response: the verbal response did not enable the experimenter to identify either the first or the last picture on the screen.

The Other responses were further broken down into the following sub-categories:

(c1) Prompt repetition: the verbal response matched the middle picture on the screen (e.g., repetition of the prompt “he plays the big drum”, or a similar answer that enabled the experimenter to identify the middle picture in Table 2 such as “drum”);

(c2) Ambiguous response: the verbal response described an event but did not match any pictures on the screen (e.g., “he is tired”), or the verbal response did not enable the experimenter to identify which picture the participant was referring to (e.g., “he is standing up”);

(c3) Unintelligible response: the experimenter was not able to understand the participant’s verbal response;

(c4) No response: the participant did not provide a verbal response, or their verbal response did not describe any event (e.g., “I don’t know”);

(c5) Experimenter error: the participant’s verbal response was missing because the experimenter either forgot to turn on the recorder when the participant began talking or did not save the recording.

The first coder transcribed and coded all verbal responses based on the above eight coding categories. A second coder then transcribed and coded ten percent of the verbal responses

(randomly selected) to check the interrater reliability. Agreement between the two coders on coding of the verbal responses was high (91.90%, Cohen's kappa = 0.91).

*3.1.2. Verbal response times*

Only Correct verbal responses were coded for the RT. Each RT was measured in milliseconds from the offset of the "beep" sound to the onset of the Correct verbal response using the software Audacity, version 3.1.13. However, if the participant was distracted during the first play of the trial or required trial repetition, the RT was coded as Unknown and was excluded from the RT analysis.

***3.2. Verbal data analysis***

*3.2.1. Verbal response accuracy*

In total, we recorded 1968 verbal responses from the four-year-olds, 2016 verbal responses from the five-year-olds, 1056 verbal responses from the eight-year-olds, and 960 verbal responses from the adult participants. Table 5 provides the number and percentage of verbal responses from each category and sub-category across all age groups. The four-year-olds responded correctly in 48.68% of all trials, and the five-year-olds responded correctly in 57.39% of all trials. As age increased, the eight-year-olds and adults reached much higher accuracy, at 82.67% and 96.98%, respectively. Moreover, with increasing age, the percentage of Other responses decreased considerably. The four-year-olds and five-year-olds produced 10.77% and 4.22% Other responses, mainly because they either did not provide a verbal response or gave an ambiguous one. Eight-year-olds and adults produced very few Other responses, at 1.23% and 0.21%, respectively. These Other responses were not interpretable and were therefore excluded from the following analyses. Additionally, it is worth pointing out that out of a total of 27 Reversal responses and 2 Other responses produced by adults, one adult contributed 22 reversal responses and 2 Other responses. This suggests that, in general, adults found completing our task very straightforward.

**Table 5**

*The number and percentage of verbal responses from each category and sub-category across all age groups*

| | Four-year-olds | Five-year-olds | Eight-year-olds | Adults |
|---|---|---|---|---|
| **Correct responses** | 958 (48.68%) | 1157 (57.39%) | 873 (82.67%) | 931 (96.98%) |
| **Reversal responses** | 798 (40.55%) | 774 (38.39%) | 170 (16.10%) | 27 (2.81%) |
| **Other responses** | 212 (10.77%) | 85 (4.22%) | 13 (1.23%) | 2 (0.21%) |
| Prompt repetition | 28 | 14 | 0 | 0 |
| Ambiguous response | 62 | 43 | 10 | 2 |
| Unintelligible response | 20 | 5 | 2 | 0 |
| No response | 97 | 23 | 0 | 0 |
| Experimenter error | 5 | 0 | 1 | 0 |

Figure 1 shows the proportion of Correct and Reversal responses for all age groups across four types of adverbial clauses in two clause orders. The figure indicates a clear interaction between ClauseOrder and Type across AgeGroup, which we will analyse statistically below to examine the significance of this interaction.

**Figure 1**

*Proportion of correct responses for children's and adults' production of adverbial clauses*

The statistical analyses were carried out using Generalized Linear Mixed-effects Models (Baayen et al., 2008) with the lme4 package (Bates et al., 2015) in R, version 4.2.0. The null model included random intercepts for participants and items. ClauseOrder (sub-main, main-sub), Type (*before*, *after*, *because*, *if*), and AgeGroup (four-year-olds, five-year-olds, eight-year-olds, adults) were fixed effects. We first compared each fixed effect to the null model one at a time with the ANOVA function. All fixed effects were significant and were therefore retained to the next stage. We then added each two-way interaction (ClauseOrder x Type, ClauseOrder x AgeGroup, Type x AgeGroup) to the model that included all significant fixed effects one at a time, and all two-way interactions were significant. Lastly, we compared the model with all two-way interactions (along with the included fixed effects) with the model with the three-way interaction (ClauseOrder x Type x AgeGroup) to determine the significance of the interaction between ClauseOrder and Type across AgeGroup. However, the three-way interaction failed to converge, which may be due to the large performance gap between age groups, particularly between young children and adults.

We therefore created another model using the same strategy, but including only the three child age groups (four-year-olds, five-year-olds, eight-year-olds). This model indicated a significant three-way interaction (along with the included fixed effects and two-way interactions), as shown in Table 6. To directly compare different levels of fixed effects and their interactions, post-hoc comparisons were conducted using the emmeans package (Lenth et al., 2023) with Bonferroni correction.

**Table 6**

*Significant fixed effects and interactions in the final model for children's production of adverbial clauses (accuracy)*

| | *β* | *SE(β)* | *z* | *p* |
|---|---|---|---|---|
| (Intercept) | 0.57 | 0.19 | 2.98 | 0.003 |
| **ClauseOrder_sub-main** | **-0.97** | **0.23** | **-4.20** | **< .001** |
| **AgeGroup_five-year-olds** | **0.57** | **0.27** | **2.11** | **0.035** |
| **AgeGroup_eight-year-olds** | **1.41** | **0.36** | **3.89** | **< .001** |
| **Type_*after*** | **-0.82** | **0.20** | **-4.04** | **< .001** |

| | | | | |
|---|---|---|---|---|
| **Type_*because*** | **-0.86** | **0.23** | **-3.76** | **< .001** |
| **Type_*if*** | **-0.79** | **0.24** | **-3.25** | **0.001** |
| **ClauseOrder_sub-main: Type_*after*** | **2.17** | **0.29** | **7.46** | **< .001** |
| **ClauseOrder_sub-main: Type_*because*** | **1.71** | **0.29** | **5.98** | **< .001** |
| **ClauseOrder_sub-main: Type_*if*** | **1.90** | **0.29** | **6.56** | **< .001** |
| ClauseOrder_sub-main: AgeGroup_five-year-olds | -0.53 | 0.33 | -1.60 | 0.110 |
| **ClauseOrder_sub-main: AgeGroup_eight-year-olds** | **1.44** | **0.48** | **3.00** | **0.003** |
| **AgeGroup_five-year-olds: Type_*after*** | **-0.63** | **0.29** | **-2.17** | **0.030** |
| **AgeGroup_eight-year-olds: Type_*after*** | **1.09** | **0.43** | **2.54** | **0.011** |
| AgeGroup_five-year-olds: Type_*because* | -0.36 | 0.31 | -1.18 | 0.240 |
| AgeGroup_eight-year-olds: Type_*because* | -0.19 | 0.41 | -0.49 | 0.647 |
| **AgeGroup_five-year-olds: Type_*if*** | **-1.13** | **0.33** | **-3.43** | **< .001** |
| AgeGroup_eight-year-olds: Type_*if* | -0.16 | 0.43 | -0.38 | 0.706 |
| **ClauseOrder_sub-main: AgeGroup_five-year-olds: Type_*after*** | **1.20** | **0.42** | **2.85** | **0.004** |
| **ClauseOrder_sub-main: AgeGroup_eight-year-olds: Type_*after*** | **-2.24** | **0.63** | **-3.58** | **< .001** |
| **ClauseOrder_sub-main: AgeGroup_five-year-olds: Type_*because*** | **1.13** | **0.41** | **2.76** | **0.006** |
| **ClauseOrder_sub-main: AgeGroup_eight-year-olds: Type_*because*** | **-1.44** | **0.57** | **-2.52** | **0.012** |
| **ClauseOrder_sub-main: AgeGroup_five-year-olds: Type_*if*** | **2.11** | **0.42** | **4.96** | **< .001** |
| ClauseOrder_sub-main: AgeGroup_eight-year-olds: Type_*if* | -1.08 | 0.61 | -1.78 | 0.075 |

*Note.* ClauseOrder = sub-main vs. main-sub (reference level). AgeGroup = eight-year-olds vs. five-year-olds vs. four-year-olds (reference level). Type = *after* vs. *because* vs. *if* vs. *before* (reference level). Number of observations = 4730. Significant effects are highlighted in bold.

As shown in Figure 1 and supported by post-hoc analysis (see Table S2 in the Supplemental Material), both four- and five-year-olds were significantly more accurate on *before*-sentences in the iconic main-subordinate order (e.g., “She builds a tower before she breaks her train”) than in the non-iconic subordinate-main order. They were significantly more accurate on *after*-, *because*-, and *if*- sentences in the iconic subordinate-main order (e.g., “After she builds a tower, she breaks her train”; “Because/If she climbs the ladder, she finds the nests”) than in the non-iconic main-subordinate order. When comparing *after*-, *because*-, and *if*-sentences, post-hoc analyses (see Table S2) show that four- and five-year-olds produced them equally well in both clause orders (with one exception: five-year-olds were significantly more accurate on *because*-sentences than on *if*-sentences in main-subordinate order). However, four- and five-year olds produced these three types of sentences significantly less accurately than *before*-sentences in main-subordinate order, but more accurately in subordinate-main order. Taken together, our results suggest that young children were

significantly more accurate on all types of sentences in iconic clause order. Their performance with sentences in iconic clause order was clearly above chance, while their performance with sentences in non-iconic clause order hovered around chance.

When comparing four- and five-year-olds, as shown in Figure 1, there was a general trend indicating that five-year-olds were more accurate than four-year-olds in their production of iconic sentences, with the differences between the two groups for iconic *because*- and *if*-sentences reaching significance (see Table S2). On the other hand, five-year-olds were significantly less accurate than four-year-olds in their production of *if*-sentences in the non-iconic main-subordinate order. This finding indicates that five-year-olds showed a stronger tendency to produce iconic sentences accurately than four-year-olds in general, and the effect of iconicity was most pronounced in five-year-olds' *if*-sentences[6].

Compared to the two younger age groups, eight-year-olds' production of adverbial sentences was generally better (above chance performance for all types of sentences in both clause orders), with only five-year-olds achieving comparable performance in the production of iconic sentences. However, eight-year-olds showed a different pattern; the effect of iconicity was found only for their *because*- and *if*-sentences. Post-hoc analyses (see Table S2) show that they were significantly more accurate with *because*- and *if*-sentences in the iconic subordinate-main order than in the non-iconic main-subordinate order, while they produced *before*- and *after*-sentences equally well in both clause orders. When comparing these four types of sentences, *because*- and *if*-sentences were produced less accurately than *before*- and *after*-sentences in main-subordinate order, with only the *before-if* comparison marginally failing to reach significance. In subordinate-

[6] The accuracy results for four- and five-year-olds (i.e., from the model that included the three groups of child participants in a single analysis) were consistent with the results based on our pre-registered analysis plan, where the data for four- and five-year-olds were analysed separately from the other age groups. Moreover, as we mentioned previously, we provided trial repetitions when four- and five-year-olds did not provide a verbal response (N=654), provided an ambiguous response (N=66) or unintelligible response (N=25), or repeated the prompt (N=33). To examine whether extra repetitions would affect the response patterns of four- and five-year-olds, we conducted additional analyses of their responses with trial repetition (see Table S3) and without trial repetition (see Table S4). The results suggest no difference, meaning that extra repetitions did not affect young children's production patterns.

main order, *because*- and *if*-sentences were produced equally well as *before*- and *after*-sentences. There were no significant differences between *because*- and *if*-sentences, nor between *before*- and *after*-sentences in either clause order. Additionally, eight-year-olds' performance resembled that of adults, although the adults generally performed near ceiling, except for one participant who made the vast majority of errors.

To conclude, the verbal response accuracy analyses show that both four- and five-year-olds produced iconic sentences more accurately than non-iconic ones, while eight-year-olds were more adult-like in showing the effect of iconicity only in *because*- and *if*-sentences. When comparing the groups, there was a general trend indicating that eight-year-olds produced all types of sentences more accurately than the younger groups (but five-year-olds achieved a similar performance to eight-year-olds in their production of iconic sentences), and that five-year-olds showed a stronger effect of iconicity than four-year-olds, especially in their production of *if*-sentences.

*3.2.2. Verbal response times*

For the analyses of verbal RTs, only the Correct verbal responses were included. From the Correct verbal responses, we removed the responses with RTs that were coded as Unknown (i.e., participants were distracted during the first play of the trial or required trial repetition). After inspection of the data, we also removed the responses with RTs that exceeded 2.5 standard deviations above or below the mean RT (i.e., outliers, as outlined in the pre-registration). In total, 549 out of 958 Correct responses from the four-year-olds were included (excluded 395 Unknown and 14 outliers), 890 out of 1157 Correct responses from the five-year-olds (excluded 243 Unknown, 24 outliers), 806 out of 873 Correct responses from the eight-year-olds (excluded 46 Unknown RTs and 21 outliers) and 900 out of 931 Correct responses from adults (excluded 10 Unknown RTs and 21 outliers). The mean RT for the four-year-olds was 1578.55ms, for the five-year-olds 1672.13ms, for the eight-year-olds 1270.57ms, and for the adults 560.99ms. Figure 2 shows the RTs for all groups of participants in the production of four types of adverbial clauses.

**Figure 2**

*RTs for children's and adults' production of four types of adverbial clauses*

We conducted an exploratory analysis of the verbal RTs for all age groups (four-year-olds, five-year-olds, eight-year-olds, adults) using the same strategy as for verbal accuracy. The final model indicates that the fixed effect of Type significantly added to the model (see Table 7). As shown in Figure 2, there was a general trend for all groups of participants to produce *because*- and *if*-sentences more slowly than *before*- and *after*-sentences, with post-hoc analyses (see Table S5) indicating that *because*-sentences were produced significantly more slowly than *before*-sentences, and that the comparison between *if*- and *before*-sentences, as well as between *because*- and *after*-sentences marginally reached significance. There were no significant differences between *because*- and *if*-sentences, nor between *before*- and *after*-sentences[7].

The fixed effect of AgeGroup also significantly added to the model (see Table 7). Post-hoc analyses (see Table S5) show that all groups of children produced sentences significantly more slowly than adults. There were no significant differences among the child groups, but the comparison

[7] The RT results for four- and five-year-olds, as well as for adults (i.e., from the model that included all participant groups in a single analysis), were consistent with the results based on our pre-registered analysis plan, where the data for four- and five-year-olds and for adults were analysed separately.

between five-year-olds and eight-year-olds marginally reached significance. However, the RT results from young children should be interpreted with caution, as a substantial proportion of the accuracy data (40.55% from four-year-olds and 38.39% from five-year-olds) were excluded from the RT statistical analyses, as only Correct responses were included. Having said that, our inspection of the dataset suggests that there was no clear pattern in terms of which specific children or specific adverbial type responses were included in the RT data, so the dataset is unlikely to represent only a subset of the participants or adverbial types investigated.

**Table 7**

*Significant fixed effects in the final models for children's and adults' production of four types of adverbial clauses (RTs)*

| Fixed effects | *β* | *SE(β)* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| (Intercept) | 1412.24 | 127.15 | 136.62 | 11.11 | < .001 |
| Type_*after* | 38.66 | 66.56 | 98.22 | 0.58 | 0.563 |
| **Type_*because*** | **315.13** | **117.89** | **42.99** | **42.99** | **0.011** |
| **Type_*if*** | **281.05** | **109.99** | **34.12** | **34.12** | **0.015** |
| AgeGroup_five-year-olds | 164.31 | 135.17 | 122.25 | 1.22 | 0.227 |
| AgeGroup_eight-year-olds | -217.80 | 153.63 | 107.58 | -1.42 | 0.159 |
| **AgeGroup_adults** | **-909.48** | **156.33** | **101.97** | **-5.82** | **< .001** |

*Note.* Type = *after* vs. *because* vs. *if* vs. *before* (reference level). AgeGroup = adults vs. eight-year-olds vs. five-year-olds vs. four-year-olds (reference level). Number of observations = 3166. Significant effects are highlighted in bold.

### *3.3. Individual difference measures*

For the individual difference analyses, the data from two four-year-olds who did not complete the forward digit span task, the missing scan task and the Flanker task (in the second session) and the data from one four-year-old who did not complete the Flanker task were excluded. The descriptive statistics for the five individual difference measures are presented in Table 8.

We first tested whether any of the individual difference measures were correlated with the children's mean accuracy and RTs in the sentence completion task. Tables 9 and 10 show the standard inter-correlations between the five measures, and their correlations with mean accuracy and RTs for four- and five-year-olds, and the eight-year-olds. For the four- and five-year-olds, all the

individual measures were significantly positively correlated with one another and with mean accuracy. This means that children who scored higher on one individual measure also tended to score higher on another, and performed better on the sentence completion task. However, none of the individual difference measures significantly correlated with mean RT.

For the eight-year-olds, only the BPVSIII vocabulary score was significantly positively correlated with mean accuracy. The BPVSIII vocabulary score was also significantly positively correlated with another language measure (i.e., recalling sentences) and one memory measure (i.e., forward digit span). Like for the younger children, none of the individual difference measures significantly correlated with mean RT.

**Table 8**

*Means and standard deviations of the raw scores for the five individual difference measures*

| | Maximum score | **Four-year-olds** | | **Five-year-olds** | | **Eight-year-olds** | |
|---|---|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD | Mean | SD |
| Recalling sentences in context | 39 | 13.31 | 8.75 | 23.10 | 8.86 | / | / |
| Recalling sentences | 63 | / | / | / | / | 31.32 | 11.29 |
| Vocabulary | 168 | 57.95 | 12.62 | 74.10 | 14.85 | 117.95 | 18.71 |
| Forward digit span | 16 | 5.54 | 1.60 | 7.36 | 2.07 | 10.27 | 2.05 |
| Missing scan | 10 | 3.56 | 1.12 | 4.00 | 1.33 | / | / |
| | 12 | / | / | / | / | 5.95 | 2.19 |
| Flanker | 96 | 60.31 | 14.84 | 82.36 | 13.67 | 93.91 | 1.77 |

**Table 9**

*Standard inter-correlations between the five individual difference measures, and mean accuracy and RT in the sentence completion task for four- and five-year-olds*

| Task | Mean accuracy | Mean RTs | Recalling sentences in context | Vocabulary | Forward digit span | Missing scan task |
|---|---|---|---|---|---|---|
| Recalling sentences in context | 0.34** | -0.09 | - | | | |
| Vocabulary | 0.41*** | -0.02 | 0.58*** | - | | |
| Forward digit span | 0.32** | -0.04 | 0.57*** | 0.53*** | - | |
| Missing scan | 0.28* | -0.10 | 0.31** | 0.32** | 0.38*** | - |
| Flanker | 0.30** | -0.08 | 0.56*** | 0.45*** | 0.30** | 0.30** |

*Note. *p < 0.05. **p < 0.01. ***p < 0.001.*

**Table 10**

*Standard inter-correlations between the five individual difference measures, and mean accuracy and RTs in the sentence completion task for eight-year-olds*

| Task | Mean verbal accuracy | Mean verbal RTs | Recalling sentences | Vocabulary | Forward digit span | Missing scan task |
|---|---|---|---|---|---|---|
| Recalling sentences | 0.39 | -0.33 | - | | | |
| Vocabulary | 0.71*** | -0.32 | 0.66*** | - | | |
| Forward digit span | 0.24 | -0.19 | 0.49* | 0.37 | - | |
| Missing scan | 0.26 | -0.33 | 0.12 | 0.41 | -0.12 | - |
| Flanker | -0.30 | 0.07 | -0.27 | -0.35 | -0.26 | -0.07 |

*Note. *p < 0.05. ***p < 0.001.*

In line with our pre-registered analysis plan, we entered each individual difference measure that was significantly correlated with mean accuracy one by one (by decreasing strength of correlation) as a predictor into the final statistical model for accuracy of four- and five-year-olds, as well as eight-year-olds[8], to determine their independent contributions. Since none of the individual measures correlated with mean RTs, no further statistical analyses were conducted. The results show that for four- and five-year-olds, only the BPVSIII vocabulary test score, $\beta$ = 0.01, *SE(β)* = 0.004, $z$ = 3.10, $p$ = 0.002, remained a significant predictor, making an independent contribution to younger children's production of adverbial clauses over and above the manipulated sentence-level variables. Similarly, vocabulary score was the sole significant predictor for eight-year-olds, $\beta$ = 0.03, *SE(β)* = 0.009, $z$ = 3.64, $p$ < .001.

**4. Discussion**

To our knowledge, the current study is the first to test the production of four types of adverbial clauses (*after*, *before*, *because*, *if*) within a single study using the same type of sentence completion task. We aimed to test three theoretical accounts to determine which can best explain our findings. As listed in Table 1 in the Introduction, the semantic account predicts that iconic

[8] Please note that all groups of children were analysed in a single model for both the accuracy and RT analyses in the sentence completion task. However, for individual difference analyses, we had to separate younger and older children. This was because we only included the individual difference measures that correlated with the children's mean accuracy, and these measures differed for each age group.

sentences (i.e., *after*-, *because*- and *if*-sentences in subordinate-main order; *before*-sentences in main-subordinate order) should be easier to produce than non-iconic sentences. The processing-based account predicts that main-subordinate sentences should be easier to produce than subordinate-main sentences across the board. The frequency-based account predicts that high frequency connective-clause order combinations should be easier to produce (i.e., *after*- and *if*-sentences in subordinate-main order; *before*- and *because*-sentences in main-subordinate order). We also explored the link between children's adverbial sentence production and their individual differences in language skills, short-term and working memory, and inhibitory control.

***4.1. The competition between semantic, processing- and frequency-based factors***

Using the sentence completion task, our results reveal that four- and five-year-olds completed all types of adverbial sentences more accurately when the sentences were in iconic clause order than in non-iconic clause order. Their performance with sentences in non-iconic clause order hovered around chance. These results suggest that young children have not yet developed a clear understanding of the meaning of each connective, and heavily rely on the "order-of-mention" strategy (i.e., the semantic account) to construct adverbial sentences, assuming a direct mapping between the order of events in the real world and in the linguistic form.

When comparing four- and five-year-olds, five-year-olds showed a stronger effect for iconicity and a stronger tendency to follow the "order-of-mention" strategy in general. We suggest that this could be because four-year-olds have not yet developed flexible temporal–causal reasoning or have not fully understood the importance of chronological organization (e.g., Povinelli et al., 1999; McCormack & Hoerl, 2005; McColgan & McCormack, 2008), which results in a relatively more random selection between pictures. In addition, we found that the effect of iconicity was most pronounced in five-year-olds' *if*-sentences, possibly because *if*-sentences pose greater processing difficulties compared to other types, leading to a stronger reliance on the "order of mention" strategy. The processing difficulties associated with *if*-sentences will be explained in detail later in this section. However, it is worth noting that there is only limited evidence suggesting that both

four- and five-year-olds understand the differences between the four types of adverbial clauses. They did take longer to produce *because*- and *if*-sentences than *before*- and *after*-sentences (this finding should be interpreted with caution, as a substantial proportion of the data was excluded from the RT analyses); however, the strategy they used for producing *because*- and *if*-sentences was similar to that for the other two types (i.e., the effect of iconicity on accuracy).

Compared to younger children, eight-year-olds generally completed all types of sentences more accurately. They still showed the effect of iconicity, but only for *because*- and *if*-sentences. The non-iconic *because*- and *if*-sentences were produced less accurately than iconic *because*- and *if*-sentences as well as *before*- and *after*-sentences in both orders. Consistent with this accuracy pattern, eight-year-olds also produced *because*- and *if*-sentences more slowly than *before*- and *after*-sentences. These results suggest that older children have developed a more robust knowledge of *before*- and *after*-sentences. However, they still experience some difficulty with *because*- and *if*-sentences, and therefore continue to show some reliance on the "order-of-mention" strategy to produce these sentences. We suggest that the processing difficulty of *because*- and *if*-sentences could be explained from the following perspectives.

First of all, the pragmatic properties *because*- and *if*-sentences encoded in the experiment differ from the properties they typically encode in everyday speech. As we discussed in the Introduction, in everyday speech, *because*-sentences occur overwhelmingly in the non-iconic main-subordinate order and express the "Speech Act" relationship (e.g., "Don't step in puddles, because you are getting your shoes wet"), which is different from the "Content" sentences used in experiment. *If*-sentences occur more frequently in the iconic subordinate-main order and express the "Content" relationship, while in the non-iconic main-subordinate order, they tend to express the "Speech Act" relationship (De Ruiter et al., 2018, 2021; Lemen et al., 2021, 2024). The unmatched pragmatic property encoded could lead to greater processing difficulties of *because*- and *if*-sentences, especially in the non-iconic order.

Moreover, *because*- and *if*-sentences are semantically more complex. They require an understanding of both temporal and causal/conditional relations between the events, while *before*- and *after*-sentences only involve temporality. However, we acknowledge that the left-to-right design of the experiment could potentially lead to a more pronounced temporal interpretation between events. This might explain why younger children, who have not yet developed robust knowledge of each connective, show similar accuracy patterns for all four types of sentences (i.e., the effect of iconicity, using a more temporal interpretation of all types of sentences). However, for older children, who have developed a more robust knowledge of connectives, the more temporal design of the *because*- and *if*-sentences could result in greater difficulty relative to *before*- and *after*-sentences, as it is inconsistent with the *because*- and *if*-sentences they typically encounter in everyday speech.

### *4.2. Asymmetries and symmetries between production and comprehension*

The results from the current production study are consistent with De Ruiter and colleagues' (2018) comprehension study in observing the effect of iconicity on accuracy, and in observing the slower responses to *because*- and *if*-sentences compared to *before*- and *after*-sentences in terms of RTs. However, it is worth noting that De Ruiter and colleagues (2018), who tested the comprehension of the same types of adverbial sentences in children of the same age, only observed the effect of iconicity in five-year-olds, but not in four-year-olds. The four-year-olds' comprehension of all types of sentences in both clause orders was at chance level, while five-year-olds' comprehension of *after*-, *because*- and *if*-sentences in non-iconic order was at chance level (consistent with our production results).

De Ruiter and colleagues suggested that the difference between their five- and four-year-olds could also be because four-year-olds have not yet developed flexible temporal–causal reasoning. In the comprehension studies, children had to construct a mental representation of the events based on the sentence they heard. When they had not understood the importance of chronological organization, they chose randomly between the two picture sequences. However, in

our production study, each visual stimulus was presented first and accompanied by the audio-recorded sequence words “first”, “next”, and “last”. This could have helped children to develop a better idea of the order of the events. Therefore, four-year-olds were able to use the “order-of-mention” strategy to interpret the sentences in production, though the effect of iconicity in four-year-olds was not as strong as in five-year-olds.

Another difference between the results from the current production study and De Ruiter and colleagues’ (2018, 2020) comprehension studies is in the processing of *before*- and *after*-sentences. In De Ruiter and colleagues’ studies (2018, 2020), *before*-sentences were comprehended more accurately than *after*-sentences. Compared to *after*, *before* has been suggested to be easier to process, as it is used more consistently as a temporal connective (Clark, 1971). Blything and Cain’s (2019) production results also support this argument. They found that iconicity only affected children’s production of *after*-sentences, but not *before*-sentences, arguing that the more consistent form-meaning relationship makes *before* easier to produce in general. However, in our production study, we did not observe a difference between *before*- and *after*-sentences. This could be related to the design of our study, which emphasizes the order of the events, but in our sentence completion task, participants did not need to produce the connectives in their verbal responses.

Our study is not the first one that did not observe a difference between *before*- and *after*-sentences. In Johnson (1975), two different types of command tasks were used, and the results varied across the tasks. In the picture-command task, children were asked to make a picture based on the command sentence that included two different verbs (e.g., “Before you **paste** the car on, **draw** the road”), while in the other command task children were asked to move different coloured cars based on the command sentence that included only one verb (e.g., “After you **move** the pink car, **move** the green car”). The results show that children’s performance in the picture-command task was significantly worse than in the other command task due to the greater difficulty of the task by involving two different verbs. Moreover, the advantage of *before*-sentences over *after*-sentences was only observed in the easier command task, but not in the picture-command task. In contrast,

Gorrell et al. (1989) used the easier version of the command task but did not find a difference between *before*- and *after*-sentences. They provided an opposite explanation that reducing the task demand could facilitate performance and therefore the difference between *before*-sentences and *after*-sentences diminished. Though the results of these two studies are not consistent, they suggest the same idea that the design of the task and the different test stimuli used could mask differences between *before*- and *after*-sentences.

### *4.3. Vocabulary predicts children's production*

The individual difference results show that young children's production of adverbial sentences was correlated with their vocabulary, short-term and working memory and inhibitory control. However, only vocabulary (assessed by the BPVSIII) made an independent contribution over and above the manipulated sentence-level variables such that children aged four, five, and eight years with better vocabulary produced adverbial clauses more accurately in the sentence completion task. We suggest that the individual difference results we observed clearly reflect how children process adverbial sentences in the production task. In the task, children were asked to complete a sentence after the experimenter's prompt, placing greater demands on their ability to access and retrieve language knowledge to complete the sentence. Therefore, children who accessed and retrieved language knowledge more efficiently had less difficulty in the task. Our results are consistent with Blything and Cain's (2019) production study, in which vocabulary also predicted young children's production accuracy of adverbial clauses in the blocked elicited production task. However, our results are not consistent with De Ruiter et al.'s (2018, 2020) comprehension studies, which did not find any independent contribution of vocabulary. This suggests that the skills needed to succeed in production may differ from those needed for comprehension. Production tasks that require greater demands on the retrieval of language knowledge are more influenced by vocabulary knowledge (Blything et al., 2015; Blything & Cain, 2016, 2019).

## 5. Conclusions

In the current study, we evaluated three theoretical models of the factors underpinning children's production of four types of adverbial clauses (*after*, *before*, *because*, *if*). Consistent with the previous comprehension findings (De Ruiter et al., 2018, 2020), the semantic account provided the best fit for young children's production results. Young children have limited understanding of these adverbial clauses, and rely heavily on the semantic factor of iconicity (i.e., the interaction between clause order and type) to produce all types of adverbial clauses. Our findings thus extend the literature, suggesting that while comprehension and production likely impose different demands on children, iconicity as a general semantic strategy benefits children's early processing of adverbial clauses. Moreover, we observed a developmental trajectory in children's production of adverbial clauses. With increasing age, eight-year-olds develop a robust understanding of *before*- and *after*-sentences. However, they still experience some difficulty with *because*- and *if*-sentences, which continue to show the effect of iconicity. The semantic and frequency-based explanations could account for the processing difficulties associated with *because*- and *if*-sentences. These sentences are semantically more complex, involving both temporal and causal/conditional relationships between events, and their pragmatic properties vary in everyday speech and in the experiment. However, the processing-based account, which focuses solely on clause order, cannot explain our results. Lastly, some variance in the children's performance was observed. Children with better vocabulary knowledge produced adverbial clauses more accurately.

**Data availability**

The datasets generated and analysed during the current study are available in the Open Science Framework repository (https://osf.io/skpm2/?view_only=cb6f16caa3744bcb91c82cce72d679ad (anonymous link))

**Ethics**

This study was approved by The University of Manchester's Research Ethics Committee [Reference number: 2022-13877-24174; Approval date: 21/06/2022] and Lancaster University's Research Ethics

Committee [Reference number: FASSLUMS-2023-3836-ExRev-2; Approval date: 14/07/2023], and follows all ethical and legal guidelines. Prior written consent was obtained from all adult participants and caregivers, and verbal assent was obtained from all child participants.

**Funding**

This study was supported by the Economic and Social Research Council (Grant number: ES/S007113/1). For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising

**CRediT authorship contribution statement**

Shijie Zhang: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing; Silke Brandt: Conceptualization, Methodology, Writing – review & editing, Supervision; Anna Theakston: Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition

**Acknowledgements**

The authors would like to thank Ciara Dutton for her support in data collection and reliability check. Warm thanks also to all the children and parents who participated in our study, and to the nurseries and primary schools for their support, and to Dr Katharina Kaduk and Lancaster Babylab for helping us with recruitment and letting us use their lab space for testing.

**References**


Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(2), 239-273. https://doi.org/10.1017/S030500091400049X.

Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, *76*(1), 28-55. https://doi.org/10.1353/lan.2000.0045.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005.

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.

Blything, L. P., & Cain, K. (2016). Children's processing and comprehension of complex sentences containing temporal connectives: The influence of memory on the time course of accurate responses. *Developmental Psychology*, *52*(10), 1517. https://doi.org/10.1037/dev0000201.

Blything, L. P., & Cain, K. (2019). The role of memory and language ability in children's production of two-clause sentences containing before and after. *Journal of Experimental Child Psychology*, *182*, 61-85. https://doi.org/10.1016/j.jecp.2019.01.011.

Blything, L. P., Davies, R., & Cain, K. (2015). Young children's comprehension of temporal relations in complex sentences: The influence of memory on performance. *Child Development*, *86*(6), 1922-1934. https://doi.org/10.1111/cdev.12412.

Clark, E. V. (1971). On the acquisition of the meaning of before and after. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 266-275. https://doi.org/10.1016/S0022-5371(71)80054-3.

Diessel, H. (2001). The ordering distribution of main and adverbial clauses: a typological study. *Language*, *77*(3), 433-455. http://dx.doi.org/10.1353/lan.2001.0152.

Diessel, H. (2005). Competing motivations for the ordering of main and adverbial clauses. *Linguistics*, *43*(3), 449-470. https://doi.org/10.1515/ling.2005.43.3.449.

Diessel, H. (2008). Iconicity of sequence: A corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics*, *19*(3), 465-490. http://dx. doi.org/10.1515/COGL.2008.018.

De Ruiter, L. E., Lieven, E. V., Brandt, S., & Theakston, A. L. (2020). Interactions between givenness and clause order in children's processing of complex sentences. *Cognition*, *198*, 104130. https://doi.org/10.1016/j.cognition.2019.104130.

De Ruiter, L. E., Lemen, H. C., Lieven, E. V., Brandt, S., & Theakston, A. L. (2021). Structural and interactional aspects of adverbial sentences in English mother-child interactions: an analysis of two dense corpora. *Journal of Child Language*, *48(*6), 1150-1184. https://doi.org/10.1017/S0305000920000641.

De Ruiter, L., & Theakston, A. (2017). First language acquisition. In B. Dancygier (Ed.), *The Cambridge Handbook of Cognitive Linguistics* (pp. 59-72). Cambridge University Press. https://doi.org/10.1017/9781316339732.005.

De Ruiter, L. E., Theakston, A. L., Brandt, S., & Lieven, E. V. (2018). Iconicity affects children's comprehension of complex sentences: The role of semantics, clause order, input and individual differences. *Cognition*, *171*, 202-224. https://doi.org/10.1016/j.cognition.2017.10.015.

Dunn, D. M., Dunn, L. M., Styles, B., & Sewell, J. (2009). *The British picture vocabulary scale - Third Edition (BPVS3).* GL Assessment.

Emerson, H. F., & Gekoski, W. L. (1980). Development of comprehension of sentences with "because" or "if". *Journal of Experimental Child Psychology*, *29*(2), 202-224. https://doi.org/10.1016/0022-0965(80)90016-8.

French, L. A. (1988). The development of children's understanding of “because” and “so”. *Journal of Experimental Child Psychology*, *45*(2), 262-279. https://doi.org/10.1016/0022-0965(88)90032-X.

French, L. A., & Brown, A. L. (1977). Comprehension of before and after in logical and arbitrary sequences. *Journal of Child Language*, *4*(2), 247-256. https://doi.org/10.1017/S0305000900001641.

Gorrell, P., Crain, S., & Fodor, J. D. (1989). Contextual information and temporal terms. Journal of Child Language, 16(3), 623–632. https://doi.org/10.1017/ S0305000900010758

Homzie, M. J., & Gravitt, C. B. (1977). Children's reproductions: effects of event order and implied vs. directly stated causation. *Journal of Child Language*, *4*(2), 237-246. https://doi.org/10.1017/S030500090000163X.

Hawkins, J. A. (1990). A parsing theory of word order universals. *Linguistic Inquiry*, *21*(2), 223-261.

Hawkins, J.A. (1992). Syntactic Weight Versus Information Structure in Word Order Variation. In: Jacobs, J. (Ed), *Informationsstruktur und Grammatik* (pp. 196-219). VS Verlag für Sozialwissenschaften Wiesbaden. https://doi.org/10.1007/978-3-663-12176-3_7.

Hawkins, J. A. (1994). *A performance theory of order and constituency* (No. 73). Cambridge University Press.

Johnson, H. L. (1975). The meaning of before and after for preschool children. *Journal of Experimental Child Psychology*, *19*(1), 88-99. https://doi.org/10.1016/0022-0965(75)90151-4.

Johnson, H. L., & Chapman, R. S. (1980). Children's judgment and recall of causal connectives: A developmental study of “because,”“so,” and “and”. *Journal of Psycholinguistic Research*, *9*, 243-260. http://dx.doi.org/10.1007/BF01067240.

Junge, B., Theakston, A. L., & Lieven, E. V. (2015). Given–new/new–given? Children's sensitivity to the ordering of information in complex sentences. *Applied Psycholinguistics*, *36*(3), 589-612. https://doi.org/10.1017/S0142716413000350.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, *99*(1), 122. https://doi.org/10.1037/0033-295X.99.1.122.

Kidd, E. (2013). The role of verbal working memory in children's sentence comprehension: A critical review. *Topics in Language Disorders*, *33*(3), 208-223. https://doi.org/10.1097/TLD.0b013e31829d623e.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154-169. https://doi.org/10.1016/j.tics.2017.11.006.

Leech, G., & Rayson, P. (2014). *Word frequencies in written and spoken English: Based on the British National Corpus*. Routledge.

Lemen, H. C. P., Lieven, E. V. M., & Theakston, A. L. (2021). A comparison of the pragmatic patterns in the spontaneous because-and if-sentences produced by children and their caregivers. *Journal of Pragmatics*, *185*, 15-34. https://doi.org/10.1016/j.pragma.2021.07.016.

Lemen, H. C. P., Lieven, E. V. M., & Theakston, A. L. (2024). The influence of pragmatic function on children's comprehension of complex because- and if-sentences. *Frontiers in Language Sciences*, *3*. https://doi.org/10.3389/flang.2024.1420237.

Lenth, R. V., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2023). Emmeans: Estimated Marginal Means, aka Least-Squares Means. R Package Version 1.8.8.

MacDonald, M. C. (2016). Speak, act, remember: The language-production basis of serial order and maintenance in verbal memory. *Current Directions in Psychological Science*, *25*(1), 47-53. https://doi.org/10.1177/0963721415620776.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*(1), 35–54. https://doi.org/10.1037/0033-295X.109.1.35.

Massonnié, J., Rogers, C. J., Mareschal, D., & Kirkham, N. Z. (2019). Is classroom noise always bad for children? The contribution of age and selective attention to creative performance in noise. *Frontiers in Psychology*, *10*, 381. https://doi.org/10.3389/fpsyg.2019.00381.

McCormack, T., & Hoerl, C. (2005). Children's reasoning about the causal significance of the temporal order of events. *Developmental Psychology*, *41*(1), 54. https://doi.org/10.1037/0012-1649.41.1.54.

McColgan, K. L., & McCormack, T. (2008). Searching and planning: Young children's reasoning about past and future event sequences. *Child Development*, *79*(5), 1477-1497. https://doi.org/10.1111/j.1467-8624.2008.01200.x.

Povinelli, D. J., Landry, A. M., Theall, L. A., Clark, B. R., & Castille, C. M. (1999). Development of young children's understanding that the recent past is causally bound to the present. *Developmental Psychology*, *35*(6), 1426. https://doi.org/10.1037/0012-1649.35.6.1426.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348-379. https://doi.org/10.1016/j.jml.2007.03.002.

Roman, A. S., Pisoni, D. B., & Kronenberger, W. G. (2014). Assessment of working memory capacity in preschool children using the missing scan task. *Infant and Child Development*, *23*(6), 575-587. https://doi.org/10.1002/icd.1849.

Stevenson, R. J., & Pollitt, C. (1987). The acquisition of temporal terms. *Journal of Child Language*, *14*(3), 533-545. https://doi.org/10.1017/S0305000900010278.

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure* (Vol. 54). Cambridge University Press. https://doi.org/10.1017/CBO9780511620904.

Wasow, T. (1997). End-weight from the speaker's perspective. *Journal of Psycholinguistic Research*, *26*(3), 347-361. https://doi.org/10.1023/A:1025080709112.

Wasow, T. (2002). *Postverbal behavior*. CSLI Publications.

Wagner, L., & Holt, R. F. (2023). Time after time: Factors influencing children's comprehension of Before and After. *Journal of Child Language*, 1-9. https://doi.org/10.1017/S0305000923000612.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children - Fourth Edition (WISC-IV)*. Pearson Clinical Assessment UK.

Wiig, E. H., Secord, W.A., & Semel, E. (2006). *CELF-Preschool 2 UK*. Pearson Clinical Assessment UK.

Wiig, E. H., & Semel, E., & Secord, W.A. (2006). *Clinical Evaluation of Language Fundamental - Fifth Edition (CELF-5)*. Pearson Clinical Assessment UK.

Zhang, S. (2022). *The acquisition and use of Mandarin relative clauses by monolingual and bilingual children and adults* [Doctoral dissertation, Lancaster University]. ProQuest Dissertations and Theses Global.

Zhang, S., Junge, B., Lieven, E., Brandt, S., & Theakston, A. (2023). The Competition Between Processing and Discourse-Pragmatic Factors in Children's and Adults' Production of Adverbial When-Clauses. *Journal of Speech, Language, and Hearing Research*, *66*(12), 5048-5060. https://doi.org/10.1044/2023_JSLHR-23-00238.