CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Learning to express causal events in Mandarin Chinese: A multimodal perspective

Chenxi NIU[1,*] , Alan CIENKI[1] , Gerardo ORTEGA[2] and Martine COENE[1]

[1]Department of Language, Literature and Communication, Vrije Universiteit Amsterdam, Netherlands
[2]Department of English Language and Linguistics, University of Birmingham, UK
*Corresponding author: Chenxi Niu, Faculty of Humanities (c/o Cienki, postvak 4.14), Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands. Email: cxniu0322@gmail.com.

Abstract

Previous research has shown language-specific features play a guiding role in how children develop expression of events with speech and gestures. This study adopts a multimodal approach and examines Mandarin Chinese, a language that features context use and verb serializations. Forty children (four-to-seven years old) and ten adults were asked to describe fourteen video stimuli depicting different types of causal events involving location/state changes. Participants' speech was segmented into clauses and co-occurring gestures were analyzed in relation to causation. The results show that the older the children, the greater the use of contextual clauses which contribute meaning to event descriptions. It is not until the age of six that children used adult-like structures – namely, using single gestures representing causing actions and aligning them with verb serializations in single clauses. We discuss the implications of these findings for the guiding role of language specificity in multimodal language development.

Keywords: language development; gesture; causal event; Mandarin Chinese; typological features

Introduction

Causal events are an important component of children's daily talk. A causal event can generally be considered to consist of two parts – namely, a causing subevent (cause) and a caused subevent (result). It takes time for children to acquire causative structures to encode the two subevents, the process of which has been shown to be guided by language-specific features (Berman & Slobin, 2013; Slobin, 2004; Slobin, Bowerman, Brown, Eisenbeiss & Narasimhan, 2011). Previous studies have mainly focused on languages with specific morphemes for causal meaning (e.g., "-dir-" in Turkish) (Furman, 2012; Göksun, Küntay & Naigles, 2008) or with a strict grammatical sequence like the subject-verb-object (SVO) in English (e.g., Ammon & Slobin, 1979). These kinds of devices enable children to express "who does what to whom" succinctly in a causative clause. However, less attention has been paid to the development of causal event expressions in children learning a language that does not always use a fixed grammatical sequence like SVO.

In such languages, flexible grammatical sequences are allowed. Learners of such types of languages may have to rely on other devices for event expressions, such as an implicit topic-comment structure which places topic information at the beginning of sentences. Such topic information may be represented by the subject in a clause, or may be syntactically separated from the rest of the sentence (Shi, 2000). For example, *Tā bǎ qiú rēng chūqù le* 'He threw a ball out' can equally be expressed as *Qiú tā rēng chūqù le* 'The ball [topic], he threw out', or *Tā wán qiú ne, rēng chūqù le* 'He was playing ball [Topic]. Threw out'. This possibility suggests the potential value of investigating what could be called contextual clauses in children's speech, i.e., clauses that precede causative clauses and form part of the event meaning. These clauses introduce information about event participants such as locations and/or activities before a causal event occurs. They do not convey the meaning of causal relations themselves but form an essential part of causal events.

Another research gap concerns the types of languages examined in the encoding of causal events, and the potential impacts of language typological features on how children develop causative expressions. A few studies have compared so-called "Satellite-framed" and "Verb-framed" languages (Talmy, 2000) which show contrasting typological differences in encoding causal events, such as English, French, Turkish, etc. (Furman, Özyürek & Allen, 2006; Hickmann, Hendriks, Harr & Bonnet, 2018; Slobin et al., 2011). These studies show that the contrasting typological features of different languages lead children on different developmental paths to using adult-like expressions. However, less is known about a third type of languages, called "Equipollently-framed" languages, where the causing and caused subevents in a causal event can be represented with either two separate verbal clauses, or with a single clause that is headed by a serial verb construction. When encoding a causal event in this third type of languages, both two-clause and one-clause structures are acceptable. Studies exploring the impact these structures have on language learning are relatively rare (but see Chen, 2008 for caused state change events and Ji, Hendriks & Hickmann, 2011b for caused motion events).

A different channel where we can also see children expressing causal events in different ways is that of co-speech gesture. Children produce spontaneous hand gestures throughout the language development process (McNeill, 1992; Morgenstern, 2014; Özçalışkan & Goldin-Meadow, 2005). Children's gestures have been shown to integrate language-specific features at various levels, including verb semantics (Gullberg & Narasimhan, 2010), lexicalization patterns (Kita & Özyürek, 2003), and clause structures (Özyürek, Kita, Allen, Brown, Furman & Ishizuka, 2008; Özyürek & Özçalışkan, 2000). For instance, one-clause structures tend to co-occur with a single gesture, whereas two-clause expressions are more likely to co-occur with two separate gestures (Gullberg & Narasimhan, 2010; Kita & Özyürek, 2003; Özyürek et al., 2008; Özyürek & Özçalışkan, 2000). The close link between verbal elements in a clause and gesture use can be captured by the Interface Hypothesis, which argues that gestures represent aspects of events in ways that are compatible with formulations in speech (Kita & Özyürek, 2003; Özyürek, Kita, Allen, Furman & Brown, 2005). However, the findings from studies based on this hypothesis have concentrated on languages where a clause unit is normally defined by a single main verb.

In other languages, especially in the so-called "Equipollently-framed" languages, the predicate that heads a clause usually consists of more than two consecutive verbs. These serial verbs (verb compounds) are widely accepted as referring to single conceptual events (Comrie, 1995; Haspelmath, 2016), but few studies have investigated how gestures co-occur with such verb compounds. A notable exception is Defina's (2016) study on

Avatime (a Kwa language of eastern Ghana), which found that verb compounds tended to be synchronized with single rather than multiple gestures by Avatime speakers. However, that study was based on data from adults, so little is known yet about how children's gestures occur with verb compounds during the development process.

The present study takes contextual clauses into consideration and explores how children (four-to-seven years) learning Mandarin Chinese express causal events using speech and gestures. Chinese has been argued to be a topic-prominent language (Li & Thompson, 1976). In addition, it has typical features of "Equipollently-framed" languages (Ji, Hendriks & Hickmann, 2011a; Slobin, 2004). As such, Chinese offers an important and understudied test case that could further our understanding of how language typology shapes the multimodal development of expressions for causal events (i.e., both speech and co-occurring gestures are considered).

Language diversity and language development

As noted above, languages show typological differences in terms of how they combine and represent semantic components in causal events (Bohnemeyer, Enfield, Essegbey & Kita, 2011; Talmy, 2000). Based on characteristic ways of encoding causal events, languages generally fall into three types (Talmy, 2000). Satellite-framed languages (S-languages), like English, typically encode a causing subevent in a main verb and a caused subevent in a particle or prepositional phrase outside the main verb (e.g.,'hit' and 'off a bench' in example 1a). In contrast, Verb-framed languages (V-languages), like Turkish, typically encode the two subevents with two verbal clauses (e.g., *tekmele-* 'kick' and *in-* 'descend' in 1b). Consequently, S-languages tend to use one-clause structures to express causal events whereas V-languages tend to use two-clause structures. Different from these two contrasting types, Equipollently-framed languages (E-languages), like Chinese, show the features of both types due to their prevalent use of verb serializations. In an E-language, causing and caused subevents can be encoded with two separate verbal clauses, as in V-languages, or with one verbal construction that serializes the individual verbs together, as in S-languages. Taking (1c) and (1d) as examples, the two subevents can be represented with two verbal clauses headed by *dǎ* 'to hit' and *diào xià* 'to descend' respectively. Meanwhile, they can also be serialized in one verbal construction *dǎ xià* 'to hit something off somewhere'. As such, an E-language provides its speakers with two equally available ways to express a causal event.

(1)
(a)   English: He **hit** a ball **off a bench**.
(b)   Turkish: Fıçı-yı   **tekmele**-yerek aşağı-ya   **in**-dir-di.
                  Barrel-ACC[1] kick-CONN   downness-DAT   descend-CAU-PST
                  '(He/She/It) caused the barrel to descend by kicking it.' (Furman et al., 2006)
(c)   Chinese: Tā   **dǎ**   qiú.   Qiú   **diào xià**   le   bǎn dèng.
                  He   hit   ball. Ball   descend   ASP   bench.
                  "He hit a ball. The ball fell off a bench."

---

[1]The following abbreviations are used in this paper: ACC: accusative; CONN: connective; DAT: dative; CAU: causative; PST: past tense; BA: disposal marker *ba*; ASP: aspectual marker; DUR: durative marker *zhe*; PREP: preposition; SD: standard deviation.

(d)	Chinese:	Tā	bǎ[2]	qiú	**dǎ**	**xià**	le	bǎn dèng.
		He	BA	ball	hit	descend	ASP	bench.

‘He hit the ball off a bench.’

What implications do language typological features have for language development? Prior studies which compared children learning S- and V-languages have revealed that their causative expressions show both early shared tendencies and language-specific features. For instance, Furman et al. (2006) compared causal event narratives by children and adults speaking English (an S-language) and Turkish (a V-language) and found that, congruent with the typological features of the two languages, English adults tended to use one clause to encode causal events whereas Turkish adults preferred two clauses. Such typological variance became evident in children's causative expressions after a certain age: children of both groups preferred one-clause structures at age 3, but after age 5, Turkish children increased the use of two-clause structures whereas their English-speaking counterparts kept the predominant use of one-clause structures. Hickmann et al. (2018) compared groups of children speaking English and German (S-languages) with French (a V-language) and found that English and German children prefer one-clause structures, as their adult counterparts do, from as early as 3 years old. In contrast, French children were found to progress from one-clause to two-clause structures, tuning to adult-like patterns after age 6. These studies are great examples of research comparing children's use of S-languages and V-languages, which supported the view that typological features of languages guide how children express events along different tracks over a span of years.

Ji et al. (2011b), one of the few studies in which E-languages were examined, compared the development of expressions for caused-motion events by speakers of Chinese (an E-language) and English. They found that Chinese adults tended to use two-clause structures to package the multiple event elements shown in the study's stimuli (i.e., a main clause and a subordinate clause that were linked with a ZHE construction). Chinese children preferred to use such two-clause structures from around age 4, producing causative expressions as informative as those spoken by adults. As for English speakers, child participants from age 3 preferred one-clause structures as adult participants did but did not achieve adult-level of informativeness in their expressions until around age 10. The authors argued that the easy availability of verb serializations in Chinese could possibly facilitate the relevant language development for Chinese children. This study has provided us many insights into how children develop an E-language. Nevertheless, E-languages remain under-explored compared with the other two types of languages, especially in domains beyond caused motion events.

There are additional theoretical reasons to study this type of language. Chinese provides a topic-comment structure for organizing and expressing event meaning. Instead of following a strict grammatical sequence like SVO, topic-elements can be prioritized to anchor a theme and set up a shared context. These topic-elements can take the form of words, phrases, and even clauses. In addition, once topicalized, the elements are allowed to be omitted in subsequent clauses. For example, if the ball in (1d) is highlighted as a topic element, then (1d) can be re-organized and expressed as (1e), where the ball is introduced first and then omitted in the following clause. In this way, contextual clauses can be regarded as a cohesive part of the causative descriptions. Thus,

---

[2]Bǎ is regarded as a grammatical marker without referential meaning in this study. It marks that the noun right after it is the direct object of the action denoted by the main verb.

focusing on targeted event expressions alone inevitably overlooks the information encoded in neighboring clauses, a potentially important aspect of children's acquisition of event expressions. Although there has been some indication that children speaking other languages, particularly V-languages like French, use such contextual clauses when expressing events (Hickmann et al., 2018; Slobin, 2004), this topic has not yet received major attention in the field.

(e) Chinese: Nà   yǒu   yí gè qiú.   Tā dǎ xià qù   le.
           There have   a   ball. He hit descend ASP.
           'There was a ball. He hit (it) off.'

Chinese as a context-based E-framed language

As examples (1c-d) pointed out, Chinese is an E-framed language where one-clause and two-clause structures are equally acceptable to encode a causal event. This is mainly a consequence of the common use of Resultative Verb Compounds (RVCs), a specific type of verb serialization. An RVC consists of two verbal elements: one is a transitive verb that indicates an agent's action towards a patient, and the other is an intransitive verb which indicates a resulting change in the patient. Due to the lack of morphological changes in Chinese, the two verbal predicates in an RVC remain in the same forms when used separately or together.

RVCs can be used to refer to both caused-motion events (e.g., *dǎ pǎo* 'to hit away' in 2a and 2a') and caused-state change events (e.g., *dǎsuì* 'to smash' in 2b and 2b'). In addition, as shown in (2a') and (2b'), the form 'bǎ' is frequently used with RVCs to syntactically place a direct object of an RVC immediately before it (Jing-Schmidt & Tao, 2009).

(2)
(a)   Tā   **dǎ pǎo**   le   qiú.
      He   hit run   ASP   ball.
(a') Tā   bǎ   qiú   **dǎ pǎo**   le.
      He   BA   ball   hit run   ASP.
      'He hit the ball away.'
(b)   Tā   **dǎ   suì**   le   wǎn.
      He   hit   get smashed ASP   bowl.
(b') Tā   bǎ   wǎn   **dǎ   suì**   le.
      He   BA   bowl hit   get smashed ASP.
      'He smashed the bowl.'

The bǎ construction is an important linguistic structure for the description of causal events, but one that is highly debated in the literature. It marks that the noun right after it is the direct object of the action denoted by the main verb. Relevant to our study, research shows that Chinese children acquire the collocation of bǎ with simple causative verbs as early as around age 2 (Erbaugh, 1982), but they do not fully acquire bǎ with RVCs until age 5;6 (Tsung & Gong, 2021) or age 6 (Fahn, 1993). As this study focuses on verb compounding features of Chinese (i.e., considering all the four types of RVCs in (2)), the use of bǎ will not be focused on in our analysis.

Besides showing the properties of an E-framed language, Chinese also uses context to convey meaning (Li & Thompson, 1989). Contextual clauses resolve the potential

Figure 1. Screenshots of a video stimulus used in this study. It shows a man hitting a ball off a bench with a tennis racket.

ambiguity in Chinese sentences caused by the flexible word order and sometimes even the omission of arguments. Example (3) gives an example of how a contextual clause was used with bare causative verbs to describe the causal event shown in Figure 1.

> (3)
> Yígè   rén   bǎ   qiú   fàng zài   zhè   bǎndèng   shàng miàn.
> A   person BA   ball   place to   this   bench   surface.
> 'A person placed a ball on this bench.'
> Yì   pāi,
> Once   hit,
> 'Gave a hit,'
> pāi   xià qù   le.
> hit   descend   ASP.
> 'hit off.'
> 'A person placed a ball on this bench. The person gave it a hit. The ball was hit off.'

The last two verbal clauses *yì pāi* and *pāi xià qù le*, which literally mean 'hit' and 'hit off', expressed the targeted causing and caused subevents separately. However, this two-clause response only expressed verbs, leaving all the participants unspecified. Nevertheless, the preceding clause informs us of a person placing a ball on a bench, which sets up the causal event and suggests the participants involved. It is in this way that contextual information constitutes a significant part of event meaning and its expression in Chinese.

Given this special role of contextual expressions, an interesting but unexplored question concerns how Chinese children start using contextual clauses to describe causal events. Further, when children have at their disposal two accessible ways of expressing a causal event (i.e., one-clause and two-clause structures), how do the children develop in terms of their use of these causative structures?

Gestures show distinctive features in language during language development

Children produce various types of gestures when speaking, such as representational gestures, pointing gestures, beats, etc. (McNeill, 1992). This study focuses on representational gestures, which are defined as those gestures whose forms resemble in some way what is being referred to in the speech segments that they co-occur with (McNeill, 1992).

Representational gestures can be analyzed in terms of how they depict meaning through modes of representation, such as enacting an action, embodying an object with the hand (as when the middle and index fingers represent the blades of a pair of scissors), tracing a path with the hand or fingers, and appearing to hold or touch an imaginary object with one's hand(s) (Müller, 2014). With these modes of representation, representational gestures not only refer to specific actions or attributes of objects (McNeill, 1986; Streeck, 2008), but can also involve less detailed iconic mappings, such as in schematically showing a generalized path of motion (e.g., moving an index finger horizontally to indicate in a generic way the motion of a rolling ball) (Kita, Alibali & Chu, 2017).

As children's speech develops, their use of representational gestures changes accordingly (Gullberg, de Bot & Volterra, 2008; Özçalışkan & Goldin-Meadow, 2011; Özyürek & Özçalışkan, 2000). Previous studies have argued that such representational gestures are produced from the interface between mental representation of imagery information and language-specific structural possibilities (Kita & Özyürek, 2003; Mittelberg & Evola, 2014). In a study on Turkish children, for instance, Furman, Küntay, and Özyürek (2014) found that Turkish children started to use representational gestures at around 22-23 months, which was earlier than the '26 month' finding reported for English-speaking children (Özçalışkan & Goldin-Meadow, 2011). Furman et al. (2014) argued that the early use of representational gestures by Turkish children has to do with the dominant use of verbs in Turkish in event representations (Allen, Özyürek, Kita, Brown, Furman, Ishizuka & Fuji, 2007), and that the action meaning encoded in verbs invited Turkish children to use representational gestures from a young age. One could predict that the different linguistic possibilities available in Chinese (RVCs) will have an impact on children's gestural production: that is, the ways Chinese children conceptualize an event when using an RVC may be reflected in their use of representational gestures.

Other studies have found a close relation between the number of verbal clauses the speaker used and their gesture use. Özyürek et al. (2008) found that speakers of English and Turkish begin to talk in language-specific ways from age 3 (i.e., English children tended to conflate event components in one-clause structures while Turkish children preferred to map them onto two-clause structures). Turkish children, furthermore, started using separate gestures right after age 3 and English children conflated motion path and manner in gestures until age 9. The study argued that children's early use of separate gestures is induced by children's language experiences through which children understand that events are segmentable, and that each segment can be represented with a separate linguistic item.

Despite the rich insights gained from studying gestures, earlier research has mostly focused on the speech units that are headed by a single main verb, since the languages under investigation normally only allowed one verb as a main predicate. Defina (2016) is one of the few studies which examined serial verb constructions in relation to gestures. She found that verb serializations tended to occur with single gestures and supported that verb serializations refer to single conceptual events. Relatedly, Duncan (2001) examined the gestural representation of motion events by Mandarin Chinese speakers and also found that single gestures were preferred to accompany Chinese RVCs. However, she treated Chinese as an S-framed language and did not fully explore the autonomy of the verbs in Chinese RVCs and corresponding gestures. Moreover, these studies were based on adult data. An important gap in the literature, and one that we aim to address, is a description of how children develop event expressions multimodally in a verb-serializing language.

To sum up, in speech, most earlier studies have overlooked some linguistic devices that might be important in encoding causal events, such as contextual clauses. Furthermore, how children learning an E-framed language develop the use of verbal clauses with representational gestures remains understudied. The current study aims to explore these phenomena in Chinese children.

The current study

When representing causal events, Chinese shows features that distinguish it from the languages that have been studied extensively before, including using contextual clauses and verb serialization. However, there is little documentation of how Chinese-speaking children develop the expressions for causal events multimodally under the potential guiding impact of Chinese-specific features. In this study, we adopt an explorative approach and ask the following three questions:

(1) Do Chinese children use contextual clauses to express causal events, and if so, how? We predict that children do use contextual clauses, but it is unclear how they use them and whether there are differences across age groups.

(2) Which clause structures are preferred by children when expressing a causal event (i.e., one-clause structures headed by RVCs or two-clause structures headed by separate verbs)? Considering that Chinese adults predominantly use RVCs in one-clause structures to express simple causal events and that young children, regardless of their language, have an early tendency to use single clauses for expressing events (Allen et al., 2007), we predict that Chinese children will start with one-clause structures, and thus be tuned to an adult-like pattern, in that respect, from an early age.

(3) How do children represent causal events with gestures when they use RVCs versus two-clause structures in speech?

Method

Participants

Forty children (ages: four to seven years old) and ten adults speaking Mandarin Chinese as their first and only language were recruited in Hebei Province, China. The numbers of participants were based on previous studies using convenience sampling without a power analysis. The children were evenly divided (with 10 participants each) among four age groups: four-year-olds (5 females, $M = 52.9$ months, $SD = 4.18$), five-year-olds (4 females, $M = 67.7$ months, $SD = 2.83$), six-year-olds (5 females, $M = 77.3$ months, $SD = 3.06$), and seven-year-olds (5 females, $M = 89.5$ months, $SD = 3.10$). None of the child participants had sensory deficits or learning disorders and they were from families of middle-level socio-economic status.

Materials

Fourteen video clips (each 5-10 seconds long) designed by the project Causality Across Languages (CAL[3]) were used in this study to investigate causal expressions at the clause

---

[3]We thank Prof. Jürgen Bohnemeyer for allowing the use of the videos. All the video stimuli used in this study are available on the CAL website (https://causalityacrosslanguages.wordpress.com/).

level. Two additional video clips were used as warm-up material. Each of these video clips depicts a person's action directly causing a change in an object's location or form: of the fourteen test videos, six of them displayed caused motion events and eight of them displayed caused change-of-state events. Each of the stimuli could typically be encoded in terms of Chinese RVCs. The depictions of these sixteen stimuli and the corresponding RVCs are listed in Appendix 1. These video clips were presented on a laptop in a PowerPoint file, as described below, with a white slide inserted after each video stimulus. The white slide was used to indicate to the participants that the video had ended and that they should start to describe what they saw.

Procedure

Guardians of all the participants signed consent forms in advance. On the consent forms, they were shown an image exemplifying the set-up of the task and how the child participant's data might be presented in publications (i.e., in an anonymized image) if they agreed. Every participant was tested with an experimenter in a quiet room in a local kindergarten (Nanpu No.1 Kindergarten). The participants sat on a sofa before a laptop and the experimenter sat a distance away from them such that she could not tell what was happening in the videos. For four-year-old children, the experimenter waved a doll at participants to engage them and promote their narration to the doll; whereas for the remaining groups, the experimenter pretended to be a curious listener. The instructions for all participants were similar: "We are going to watch several short video clips. The doll/I want(s) to know what happens in the videos. Can you please describe the videos for the doll/me?" After participants agreed, the experiment started. Children were first shown the two warm-up video clips. Some child participants started to describe the trial videos with very short utterances containing only a single verb (e.g., *Suì le* 'Smashed!' was used to describe a woman smashing a cup into pieces. Or *Wán zhǐqiú ne* "Playing the paper ball." was used to describe that someone threw a ball of paper into a bowl). It is known that children tend to omit arguments (Bloom, 1990), but since this study aimed to examine children's skills in formulating event descriptions, when a participant used bare verbs in the warm-up phase, the experimenter raised questions in order to stimulate them to talk more about the displayed events, such as "What else did you see?", "Why?", "Can you talk more about it?", etc. However, the warm-up phase was the only time when the experimenter intervened and prompted children's descriptions. Some children tended to refrain from talking loudly and using gestures because they were educated to conform to classroom etiquette, such as keeping quiet and sitting straight with hands placed on their knees or behind their backs. In such cases, the experimenter asked things like 'what did you say just now?' to encourage them to speak louder and clearer. The experimenter used this second-time answer to help figure out the first-time replies.

For those who had difficulty recalling what had just happened, the video clips were replayed until the children understood the event. Participants' responses were video recorded[4]. After the task, each of the participants was provided with a gift for having taken part.

---

[4]This study has received the Ethical Approval from Vrije Universiteit Amsterdam and the University of Birmingham with EC 19.13 as its reference number. The storage of data follows the General Data Protection Regulation (GDPR) both in the UK and in the Netherlands. The data collected is now stored on a secure SurfDrive at the VU Amsterdam.

Figure 2. The stimulus shows that a man knocks over a cup tower.

Coding

Speech coding and data analysis

All the elicited responses were transcribed by the first author, a native speaker of Chinese, using the annotation tool ELAN (Sloetjes, 2017). For each response, the words were first counted in order to subsequently obtain the relative gesture rates per word, as will be shown in the following section.

Next, every response was segmented into clauses (i.e., units that contain one verb and its arguments) and coded as either causative clauses or contextual clauses. CAUSATIVE CLAUSES refer to the one or two clauses with explicit causative verbs that described causal events as stipulated in the CAL project (listed in Appendix 1). CONTEXTUAL CLAUSES refer to the clauses that preceded the causative clauses. As explained earlier, they introduce participants' locations or activities and provide information that is essentially part of causal events. Figure 3 and Figure 4 illustrate how a six-year-old participant and an adult participant responded to the stimuli shown in Figure 2, and how their verbal responses were coded.

> (4a)   Preceding contextual clauses
> Yǒu   yígè   xiǎotǎ,   tā   bǎichéng le   yígè   sān jiǎoxíng.
> Have   a   little tower,   it   place-become-ASP   a   triangle
> 'There was a little tower which formed a pyramid.'
> (4b)   Causative clause
> Ránhòu   yǒu   yígè   nánhái   bǎ   tā   nòng dǎo le.
> Then   have   a   boy   BA   it   make-fall-ASP.
> 'Then a boy made it fall.'
>
> (5)   Causative clause
> Yígè   rén   bǎ   dāhǎode   bēizi   pèng luàn   le.
> A   person   BA   well-placed   cups   knock-get   messed-ASP.
> 'A person knocked over the cups which had been carefully placed.'

Lastly, the main causative clauses were focused on and coded as either representing a causing and/or caused subevent, with either an RVC in one clause or with two separate verbs in two clauses. The complete scheme for speech coding is illustrated in Figure 5.
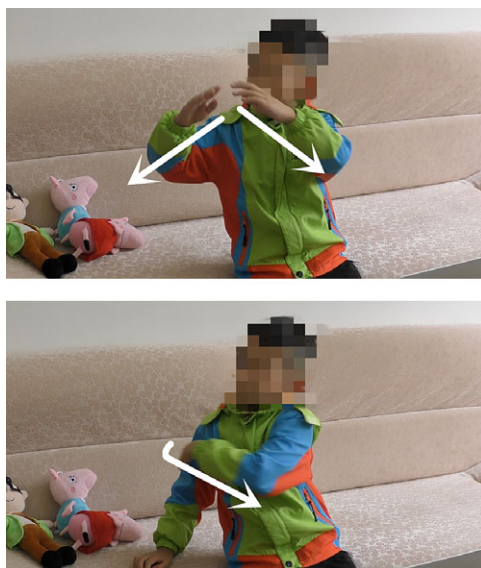
Figure 3. The six-year-old participant produced (a) preceding contextual clauses and (b) the main causative clause to describe the causal event shown in Figure 2.

Figure 4. The adult participant produced a causative clause with the resultative verb compound.
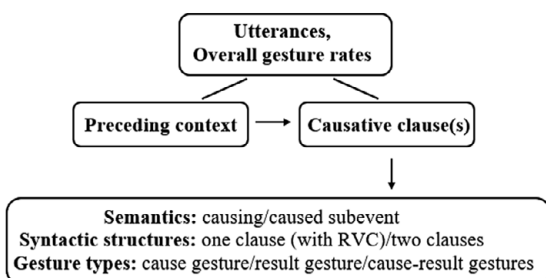

Figure 5. The complete coding schema for speech and gestures in participants' utterances.

Gesture coding

We first counted the total numbers of representational gestures produced by every participant in their descriptions of all the video clips. Combined with the total numbers of words, this enabled us to get the overall gesture frequencies in each age group. We also included body movements since there were a few cases where children used their whole bodies to express caused changes in objects (e.g., a child dropped his upper body onto the sofa which he was sitting on to express that the cup tower fell down).

Next, in order to explore whether children's use of representational gestures changed in relation to the developments in syntactic packaging, we focused on the gestures that accompanied the causative clauses where both causing and caused subevents were mentioned. The representational gestures under examination were coded as one of the following categories: (1) cause gestures, i.e., gestures that represented the force dynamic aspects in causing actions; (2) result gestures, the gestures that represented the motoric-spatial aspects in the resulting motion or resulting state changes; (3) cause-result gestures, a sequence of two separate gestures representing a causative action and a caused change respectively; and (4) other gestures, those that both coders found did not fit one of the categories above. Table 1 gives the examples of gesture types (1)-(3).

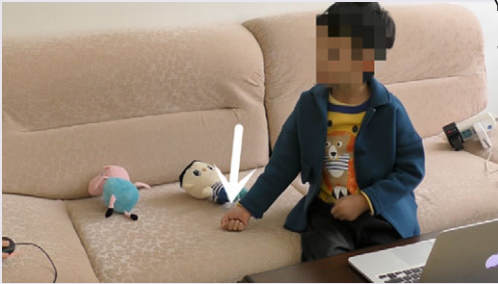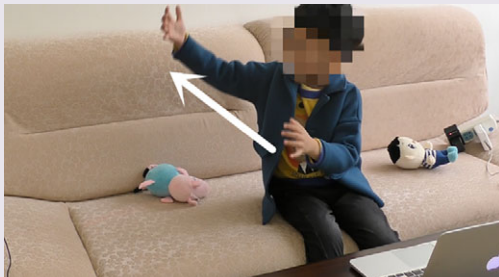**Table 1.** Examples of Cause Gestures, Result Gestures, and Cause-result Gestures

Gesture descriptions and categorizations	Utterances
The participant's closed fist moved down rapidly and hit the sofa. **Cause gesture**	Yígè nǚrén [názhe tiáozhou **dǎpò le** yígè wǎn]. *'A girl [__hit-broke__ a plate with a broom.]'*
The participant's right hand was curved, with fingers slightly bent towards the palm. It moved down vertically. **Result gesture**	Nán rén zài sǎo dì, bù xiǎo xīn [bǎ zhǐ bēi **pèng dǎo le**]. *'The man was sweeping the floor. (He) carelessly [__knocked-fell__ the paper cups.]'*

Table 1. *(Continued)*

Gesture descriptions and categorizations	Utterances

…yòng zhè tiáozhou [bǎ zhègè bēizi gěi **nòngdǎo le**].
"…he used a broom [and **hit**]-[**fell** the **cups**.]"

The first gesture (left hand) moved quickly and clapped on the other hand.

The second gesture (right hand with palm open) moved upward diagonally.
Cause-result gestures

Note. The videos of these three examples can be found at https://osf.io/etxzp/, together with more examples of Cause and Result gestures. The underlined words are resultative verb compounds (RVCs). The square brackets [ ] indicate the starting and ending points of individual gestures, and the **words in bold** indicate where the stroke of a gesture occurred (i.e., the most effortful phase in a gesture).

Reliability

Participants' utterances (700: 50 participants x 14 video stimuli) and gestures were initially coded by the first author. A second coder randomly chose and coded 30% of the speech data. Agreement between the coders for the number of contextual clauses was 82% (n = 210). Agreement for syntactic structures in speech was 92% (Cohen's kappa (κ) = .87, n = 210). As for gesture coding, the two coders coded the first 50 gestures for the types of meaning, listed above. Agreement in coding representation of cause and/or result in the gestures was 84% (Cohen's kappa (κ) = .75, n = 50). All instances of disagreement in coding were discussed by the two coders until consensus was reached. Then the first author coded the rest of the gestures using the revised coding scheme.

Results

Speech

Context

First we calculated the number of clauses that provided context for each causal event for all descriptions across age groups. Table 2 summarizes the means and standard deviations of contextual clauses across the five age groups. It indicates that older age groups (at ages five to seven years) were more likely to use contextual clauses. Next, we focused on the four children's groups and ran a generalized linear mixed effects model (with a log link function) to examine the potential association between age and the number of contextual clauses used. The model contains age (in months) as the continuous fixed factor, and stimuli items and participants as random effects factors. The results show that with each additional month, the increase in the log count equals 0.025, $Z = 3.893$, 95% CI [0.012, 0.038], $p < .001$. Thus, the number of contextual clauses is significantly associated with age (in months). The older the children were, the more contextual clauses were used.

Next, we analyzed the context clauses qualitatively to understand why the older the children were, the more contextual clauses they used. The analysis revealed that contextual clauses were mainly used to provide new information to set up a causal event, such as specifying participants and their properties or activities (e.g., 6a-b) and older children provided more information in these clauses.

(6)
(a)   A six-year-old participant:
Yǒu   yí gè   qiú   gē   zài   le   yí gè   yǐ zi   shàng. [preceding context]
Have   a   ball   put   PREP   ASP   a   chair   on.
'There was a ball sitting on a chair.'
Rán hòu   yǒu   yí gè   rén   ná zhe   yí gè   yǔ máo qiú
Then   have   a   person   holding   a   badminton
pāi. [preceding context]
racket.
'Then there was a person holding a badminton racket.'
Tā   bǎ   tā   pāi pǎo   le. [causative description]
He   BA   it   hit run   ASP.
'He hit it away.'

Table 2. Estimated Marginal Means (log value) and the Probabilities of Using Contextual Clauses across the Five Age Groups.

AGE	emmean	asymp.LCL.x	asymp.UCL.x	prob	asymp.LCL.y	asymp.UCL.y
4	−1.8563	−2.4453	−1.2672	0.1563	0.0867	0.2816
5	−0.9332	−1.4547	−0.4117	0.3933	0.2335	0.6625
6	−0.7639	−1.2790	−0.2488	0.4658	0.2783	0.7797
7	−0.8449	−1.3625	−0.3274	0.4296	0.2560	0.7208
Adults	−1.3105	−1.8542	−0.7669	0.2697	0.1566	0.4645

Note. The asymptotic degrees-of-freedom method was used (asymp). Emmean = estimated marginal means. LCL/UCL.x = lower/upper limit of the confidence interval for estimated marginal means. Prob = probabilities. LCL/UCL.y = lower/upper limit of the confidence interval for probabilities.

'There was a ball sitting on a chair.[context] Then there was a person holding a badminton racket. [context] He hit it away. [causative description]'

(b)   A five-year-old participant:

Yí gè	rén	bǎ	tā	fàng zài	bǎndèng	shàngmiàn. [preceding context]
One	person	BA	it	place to	bench	surface.

'A person placed it on this bench.'

Yì	pāi,
Once	hit,

'Gave a hit,'

pāi	xiàqù	le. [causative description]
hit	descended	ASP.

'hit off.'

'A person placed a ball on this bench. The person gave it a hit. The ball was hit off.'

In contrast, four-year-olds were less likely to give contextual information. Instead, they used pronouns directly which led to ambiguous depictions. Take (6c) as an example; when describing 'a man pushed a swing and the swing moved back and forth', the four-year-old participant used *Tā tuī nà gè le* 'He pushed that', where both the *tā* 'He' and *nàgè* 'that' were unspecified as to their referents.

(c)   A four-year-old participant

Tā	tuī	nà gè	le.	[causative description]
He/She/It*	push	that	ASP.	

'*He/She/It pushed that.* [causative description]'

Adults also gave information to identify participants but did so in a compact way with verb strings and various modifiers instead of additional clauses. Example (6d) was produced by an adult participant. Comparing (6a-b) and (6d), we can see that what was encoded in contextual clauses by older children was encoded in a co-verb *ná zhe qiú pāi* 'while holding a racket' and a modifier *dèng zi shàng de* 'on the chair' by an adult.

(d)   An adult participant

Yí gè	rén	ná zhe	qiú pāi	bǎ	dèng zi shàng de	qiú
A	person	hold DUR	racket	BA	on-chair	ball
dǎ	dào	le	dì shàng. [causative description]			
hit	arrive	ASP	on-ground.			

'A person hit off the ball which was on the chair while holding a racket. [causative description]'

Syntactic packaging of causal events in terms of the number of clauses

Focusing on causative descriptions, we first compared the difference in using one-clause and two-clause structures across the five age groups. Table 3 summarizes the estimated marginal means and probabilities of (a) one-clause responses and (b) two-clause responses for each age group. In order to test whether there were differences between children's groups and the adult group in using the two types of syntactic structures, pair-

Table 3. Estimated Marginal Means (logit value) and the Probabilities of Using (a) One-clause Responses and (b) Two-clause Structures across the Five Age Groups.

(a) AGE	emmean	asymp.LCL.x	asymp.UCL.x	prob	asymp.LCL.y	asymp.UCL.y
4	−0.7879	−1.6046	0.0287	0.3126	0.1673	0.5072
5	−0.8806	−2.0604	0.2992	0.2931	0.1130	0.5743
6	−0.3927	−1.3823	0.5969	0.4031	0.2006	0.6449
7	0.0163	−0.9201	0.9527	0.5041	0.2849	0.7217
Adults	1.1701	−0.0912	2.4314	0.7632	0.4772	0.9192

(b) AGE	emmean	asymp.LCL.x	asymp.UCL.x	prob	asymp.LCL.y	asymp.UCL.y
4	−0.9293	−1.7466	−0.1121	0.2831	0.1485	0.4720
5	−0.6244	−1.7183	0.4696	0.3488	0.1521	0.6153
6	−1.3284	−2.3971	−0.2597	0.2094	0.0834	0.4354
7	−1.5897	−2.3863	−0.7932	0.1694	0.0842	0.3115
Adults	−3.1315	−4.2735	−1.9894	0.0418	0.0137	0.1203

Note. The asymptotic degrees-of-freedom method was used (asymp). Emmean = estimated marginal means. LCL/UCL.x = lower/upper limit of the confidence interval for estimated marginal means. Prob = probabilities. LCL/UCL.y = lower/upper limit of the confidence interval for probabilities.

wise comparisons with Tukey adjustment were used. The results showed that there were significant differences in using one-clause structures between the group of four-year-olds and the adults ($z = −3.386$, $p = 0.01$), as well as the five-year-olds and the adults ($z = −3.200$, $p = 0.01$). There were also significant differences in using two-clause structures between four-year-olds and adults ($z = 3.751$, $p = 0.00$), five-year-olds and adults ($z = 4.602$, $p < 0.00$), and six-year-olds and adults ($z = 3.302$, $p = 0.01$).

Next, we focused on just the child participants and tested the effects of age on the probabilities of children using one-clause responses. Given that the outcome is a binary variable (i.e., children may either use or not use one-clauses structures), a mixed effects model with the logit link function was used. The model contained the children's age (in months) as the continuous fixed factor, and participants and stimuli items as the random factors. The results showed that the fixed effects estimate for the age variable equals 0.03, 95% CI [0.01, 0.05] ($z = 2.62$, $p = .01$). Thus, there is a statistically significant relation between age and the probability of using a one-clause response. The older the children, the more likely they were observed to use one-clause structures.

Likewise, the same model was run on the probability of using two-clause structures. The results showed that the fixed effects estimate for the age variable was -.022, 95% CI [-0.014, 0.00] ($Z = −2.144$, $p = 0.03$), indicating a statistically significant relation between age (in months) and the probability of a two-clause response. The older the children, the less likely they were observed to use two-clause structures. The predicted probabilities in relation to age (in months) are illustrated in Figure 6.
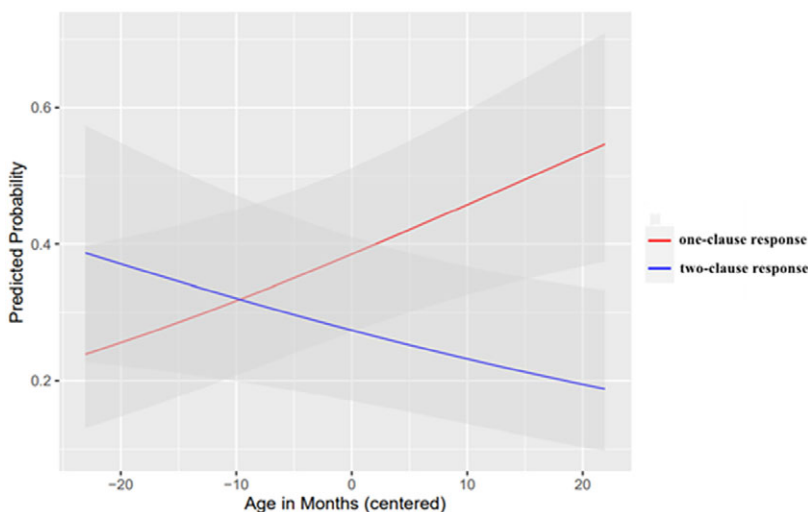
Figure 6. Predicted probabilities of using one-clause structures and two-clause structures as a function of age (in months).

Gesture

Overall gesture rates

Table 4 shows the average gesture rates across the five age groups in the preceding contextual clauses and causative clauses. Overall, children gestured almost three times as much as adults did.

Gestural representations of causing and caused subevents

Table 5 shows the mean proportions of the three types of gestures that accompanied two-clause responses. Overall, the four groups of children used cause-result gestures (ranging from 39% to 48%) most frequently. They also used moderate proportions of cause gestures (ranging from 29% to 36%). In comparison, adults used cause gestures most frequently and they did not produce representational gestures as much as children did.

   Likewise, Table 6 presents the mean proportions of the three types of representational gestures that accompanied RVCs (i.e., one-clause responses). The result shows that

Table 4. Means of Representational Gesture Rates across the Five Groups in Preceding Contextual Clauses and Causative Clauses.

Age groups	4 yr (n = 10)	5 yr (n = 10)	6 yr (n = 10)	7 yr (n = 10)	Adults (n = 10)
Gestures (per 100 words)	9.88	11.62	9.56	9.37	3.52

Table 5. Means Proportions (SD) of the Three Types of Representational Gestures that Occurred in Two-clause Responses.

Age groups	Cause gesture	Result Gesture	Cause-result gestures	Two-clause responses
4 (n = 10)	0.29 (0.18)	0.08 (0.17)	0.39 (0.33)	47
5 (n = 10)	0.35 (0.19)	0.09 (0.14)	0.41 (0.21)	58
6 (n = 10)	0.28 (0.24)	0.15 (0.33)	0.48 (0.29)	42
7 (n = 10)	0.36 (0.33)	0.05 (0.11)	0.46 (0.32)	33
Adults (n = 10)	0.17 (0.26)	0.08 (0.20)	0.02 (0.05)	14

Note. In the table, n refers to the number of participants. Cause gestures refer to the gestures representing causing actions. Result gestures represent caused changes (e.g., motion or state change). Finally, cause-result gestures are two sequential gestures, in which the first gesture represents a causing action, and the other gesture presents a resultant change. The total number of two-clause responses are shown in the last column.

Table 6. Means Proportions (SD) of the Three Types of the Gestures that Accompanied RVCs in One-clause Responses.

Age groups	Cause gesture	Result gesture	Cause-result gestures	One-clause responses
4 (n = 10)	0.43 (0.27)	0.12 (0.17)	0.06 (0.10)	50
5 (n = 10)	0.68 (0.30)	0.08 (0.11)	0.15 (0.18)	52
6 (n = 10)	0.63 (0.27)	0.03 (0.05)	0.08 (0.09)	61
7 (n = 10)	0.47 (0.25)	0.03 (0.07)	0.17 (0.26)	67
Adults (n = 10)	0.28 (0.30)	0.02 (0.04)	0.01 (0.02)	91

Note. N = number of participants. The total number of the one-clause responses are shown in the last column.

children showed a clear preference for cause-only gestures, with 43% to 68% of RVCs accompanied by cause-only gestures. Adults also preferred to use cause-only gestures, which co-occurred with about 1/3 of RVCs in their speech, but adults gestured less frequently than children did.

To summarize, four-to-seven-year-old children's gestural representations of causal events were consistent with the number of clauses they produced, which was slightly different from the pattern found in adults. When expressing causing and caused subevents with two verbal clauses, children tended to use two separate gestures for cause and result respectively, whereas adults tended to use cause-only gestures. When combining causing and caused subevents in one-clause structures – namely, with RVCs – both children and adults tended to use cause-only gestures. Adults also produced far fewer representational gestures than children (four-to-seven-years old) did.

Discussion

Previous research on causal events has shown that, regardless of their languages, children start describing events with single clauses (Allen et al., 2007; Furman et al., 2006) and over time they start mapping cause and result across different clauses depending on the

patterns of their specific language (Hickmann et al., 2018; Ji et al., 2011a). Here we explored a gap in the literature and asked how children give context to support their descriptions of causal events. We studied Chinese, an Equipollently-framed language, to shed light on the developmental path of causative structures in this language where children have at their disposal separate verbal clauses or single clauses. More important, we adopted a multimodal approach by studying children's use of representational gestures with separate or single verbal clauses, which has not been studied before in detail. In order to address these gaps, this study examined children's use of representational gestures co-occurring with separate clauses and single clauses that were headed by resultative verb compounds.

We set out to answer the following questions:

- whether, and if so, how children develop the use of context in causal event expressions;
- what syntactic structures (one- or two-clause structures) Chinese children prefer to use in describing a causal event, given that both of them are available due to verb-serializing features of Chinese;
- what representational gestures tend to occur with one- and two-clause structures.

Speech: Effects of age and language-specificity on syntactic structures

The results for contextual clauses show that older children used more contextual clauses. These contextual clauses did not encode cause and result directly but specified the relevant features, locations, or activities of participants before a causal event took place. These contextual clauses seem to form part of a coherent causative description together with the causative clauses whose arguments were sometimes omitted. We consider that such contextual clauses function as a prelude to the causal event rather than having other pragmatic functions. They are like a stage, displaying the settings of the scene before a causal event is put under the spotlight. In this regard, young children's increasing use of contextual clauses suggests their developing skills to anchor thematic objects and elaborate relevant details in a given event. Chinese children might have exhibited this developmental change because of the topic-comment structure that is prevalent in Chinese.

Alternatively, contextual clauses may be used to reduce the cognitive load of speech production for young children. These additional verbal clauses may have relieved the pressure on causative descriptions by pre-loading some relevant information. We might say that children learned to carve out the participant-related information and represent it in context, so that the remaining information to be expressed can be packaged in a way that fits their cognitive capacities. Comparing children and adults, this possibility seems plausible. Specifically, contextual information was provided in adults' descriptions, although it went from being represented in contextual clauses to being conflated in causative descriptions with devices like modifiers. Thus, contextual clauses may have served like a device to segment complex information into smaller units by children. If this possibility is true, then using additional contextual clauses may not be peculiar to Chinese learners, but universal to children learning different languages. Future research should explore if children learning other languages use descriptions of context in ways similar to what was observed here in Chinese.

As for the clausal packaging of causal events, our finding that Chinese children's use of RVCs (i.e., one-clause structures) did not reach an adult-like level until age 6 may suggest

the impact of Chinese-specific features on language development. This is evident when we compare Chinese children with children learning S-framed languages like English and German. Although in both Chinese and S-languages, adult speakers tend to use single clauses to encode a causal event, children learning S-languages become tuned to one-clause structures as early as age 3 (Hickmann et al., 2018), whereas Chinese children do not become tuned to these patterns until age 6. The later age found in this study may be attributed to the particular properties of Chinese RVCs. RVCs can be used as two separate predicates without inflectional changes. This may have invited children to use the separate verbal clauses to represent causing and caused subevents for a longer period of time before eventually using RVCs in one-clause structures. Additionally, bǎ constructions may also have caused some difficulties. Earlier studies found that Chinese children do not fully acquire the semantic constraints related to bǎ-RVC constructions until age 6 (Fahn, 1993). Young children may have not fully acquired related constructions that could be used in our task, thereby not conflating cause and result at an adult-like level until a later age. This possibility is merely speculative, but it opens the door to future study to explore the link between bǎ and the conceptualization of causal events.

In our study, Chinese children's use of RVCs approached an adult-like level later than that found in similar studies (namely, Ji et al., 2011b). This is mainly attributed to the different causal relations that our stimuli displayed. The key difference has to do with whether the causing agent keeps contacting the affected object after exerting the causative force. Unlike the events in the stimuli used by Ji et al. (e.g., someone pushes a chair and keeps pushing it until the chair moves into a cave), causing and caused subevents are separated temporally and spatially in our design (e.g., a man gave a ball a hit, and then the ball rolled away without the man continuing to touch it). This might be the reason that young children in our study tended to map causing and caused subevents onto two separate clauses, rather than conflating them in RVCs from a very early age.

Concerning the early shared tendency in compact syntactic packaging (Allen et al., 2007), our study shows contrasting results. Allen et al. (2007) argued that there might be an early shared bias to conflate event information into one unit because young children tend to "pursue cohesion between cognitive and linguistic representations of an event" (p. 43). The contradicting results in this study could relate to the different event types we examined (in addition to language differences). Allen et al.'s (2007) proposal is based on motion events, in which event elements like path of motion and manner of motion are closely related and take place almost simultaneously. A causal event, however, involves the opposition of an agent and a patient. It involves agency and is more complex than a motion event or state change event (Talmy, 2000, p. 221). We illustrate the distinction in
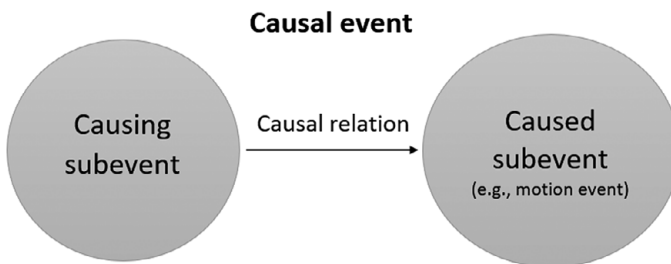

Figure 7. The semantic structure of a causal event. A causal event may contain a motion event but not vice versa (simplified, based on Talmy, 2000).

Figure 7: a causal event may contain a motion or state change event as a component, but not vice versa. Therefore, to understand whether, cross-linguistically, children initially represent events in one single unit, future studies may first need to distinguish event types. For example, one can compare young children's depictions for motion/state change events (e.g., a box moved down/a plate broke into pieces) versus those for caused motion/caused state change events (e.g., someone kicked a box and the box moved down/someome hit a plate and the plate broke into pieces).

Gestures: relationship between syntactic structures and representational gestures

We found that when two-clause structures were used, children tended to use separate gestures, whereas adults preferred single gestures representing causing subevents. Our results about adults' use of gestures differ from those in past studies which have shown that two-clause responses were most likely to elicit separate gestures in adult speakers across languages (Allen et al., 2007; Kita & Özyürek, 2003). However, earlier studies mainly investigated the gestures representing path of motion and manner of motion which are concurrent with a motion event (e.g., a ball descended while rolling). In comparison, the causing and caused subevents we investigated are temporally separate from each other (e.g., someone hit a ball and caused it to descend while rolling). As a result, adults in our study tended to express only one part of a causal event. Adult speakers preferred to produce cause gestures without result gestures, possibly because causing actions are regarded as more prominent in starting and realizing a causal event. In comparison, children gestured both cause and result when separate clauses were used, suggesting that they may treat the two subevents as equal in terms of information prominence. Thus, the difference in gesture patterns may reveal children's conceptualization of causal events and their journey to reach adult-like structures for expressing them.

Chinese RVCs tended to be accompanied by single gestures across all the age groups, which is compatible with the claim that Chinese RVCs have the same grammatical properties as a single verb (Li & Thompson, 1989). Chinese speakers seem to conceptualize causal events as single event units when using RVCs. Interestingly, such single gestures tended to depict causing actions rather than resulting changes. The reason might relate to the high gesturability of cause. The causative actions like 'hit' or 'rip' are motoric events that can be re-enacted, in contrast to changes in objects such as 'the ball rolls away' or 'the paper becomes into pieces' (glosses of some Chinese examples), which are not ones that can be simulated as easily from a first-person perspective. Therefore, this might lead to the preferred use of cause gestures.

It is important to note that the 14 video stimuli involved action-based causes, so it may be possible that this could have skewed the results towards higher gesture rates in causes than in results. Research has shown that the properties of the referent influence the types of gesture produced (Masson-Carro, Goudbeek & Krahmer, 2016), so we cannot rule out that something similar may have happened in our analysis. That said, children in our study did sometimes gesture events that were not bodily actions (e.g., spreading their hands and fingers forward to depict a plate was smashed into pieces, or falling to one side with their bodies to represent a tower of cups falling down). Future studies should use less embodied stimuli (e.g., the wind blew the umbrella away) to explore whether this changes the proportion of cause-related gestures.

Another possible reason might relate to children's close attention to animate agents, which leads them to gesture about agent's causing actions more frequently when producing RVCs in speech. For instance, during the age span investigated in this study, animacy proves to be salient in children's conception of events (Piaget, 1954). McNeill (2013) reported that while adults produce gestures reflecting an observer's viewpoint of the event (e.g., wriggling fingers to represent 'someone is running forward'), children preferred to adopt a character viewpoint and produced arm movements as if they themselves were running like the character. Therefore, it might be children's shared perspective with animate participants that led them to gesture more about causing actions.

Conclusions

There is growing consensus that language is a part of a multimodal phenomenon consisting of speech along with co-occurring gestures (Cienki, 2017; Kendon, 1980, 2004; McNeill, 1985, 1992; Morgenstern, 2014). These multimodal utterances provide an exciting window through which the cognitive and linguistic development of children can be viewed. Our study makes an important contribution to the field in that it describes the developmental changes in how children, learning a language that is understudied in this respect, describe causal events in both speech and gesture. Our findings show that contextual clauses provide a special device that can be exploited for event expressions by children. We have also found language-specific impacts on children's multimodal expressions of causal events. In terms of speech, Chinese adult speakers used Resultative Verb Compounds (RVCs) prevalently to encode causal events in single clauses. However, Chinese children used two-clause structures for a long period of time before tuning in to one-clause structures at around age 6. We argue that this relates to the separability and autonomy of the component verbs in RVCs. In terms of gestures, children's use of representational gestures corresponded to the syntactic structures used. In particular, one-clause structures containing RVCs are possibly conceptualized as encoding single events, as they co-occurred most frequently with single representational gestures. The results go beyond previous findings by suggesting that gestures are predominantly used to represent human agents' causing actions as opposed to caused changes in objects, possibly due to children's preference for the character viewpoint when expressing causal events.

References

Allen, S., **Özyürek, A.**, **Kita, S.**, **Brown, A.**, **Furman, R.**, **Ishizuka, T.**, & **Fuji, M.** (2007). Language-specific and universal influences in children's syntactic packaging of manner and path: A comparison of English, Japanese, and Turkish. *Cognition*, **102**(1), 16–48.

Ammon, M. S., & **Slobin, D. I.** (1979). A cross-linguistic study of the processing of causative sentences. *Cognition*, **7**(1), 3–17. https://doi.org/10.1016/0010-0277(79)90007-6

Berman, R. A., & **Slobin, D. I.** (2013). *Relating events in narrative: A crosslinguistic developmental study*. Psychology Press.

Bloom, P. (1990). Subjectless sentences in child language. *Linguistic Inquiry*, **21**(4), 491–504.

Bohnemeyer, J., **Enfield, N. J.**, **Essegbey, J.**, & **Kita, S.** (2011). The macro-event property: The segmentation of causal chains. In J. Bohnemeyer, & E. Pederson (Eds.), *Event representation in language and cognition* (pp. 43–67). Cambridge University Press.

Chen, J. (2008). *The acquisition of verb compounding in Mandarin Chinese* (Doctoral dissertation, Vrije Universiteit Amsterdam, Amsterdam, Netherlands). Retrieved from https://pure.mpg.de/rest/items/item_57993_7/component/file_2293749/content

Cienki, A. (2017). *Ten lectures on spoken language and gesture from the perspective of cognitive linguistics: Issues of dynamicity and multimodality*. Brill.

Comrie, B. (1995). Serial verbs in Haruai (Papua New Guinea) and their theoretical implications. In J. Bouscaren, J.-J. Franckel, & S. Robert (Eds.), *Langues et langage: Problèmes et raisonnement en linguistique: Mélanges offerts à Antoine Culioli* (pp. 25–37). Presses Universitaires de France.

Defina, R. (2016). Do serial verb constructions describe single events?: A study of co-speech gestures in Avatime. *Language*, **92**, 890–910.

Duncan, S. (2001). Co-expressivity of speech and gesture: Manner of motion in Spanish, English, and Chinese. In C. Chang, M, J. Houser, Y. Kim, D. Mortensen, M. Park-Dobb, & M. Toosarvandani (Eds.), *Proceedings of the 27th Berkeley Linguistics Society Annual Meeting* (pp. 353–370). Berkeley Linguistics Society.

Erbaugh, M. (1982). *Coming to order: natural selection and the origin of syntax in the Mandarin speaking child.* https://escholarship.org/uc/item/5gq4q2kj

Fahn, R. S. (1993). *The acquisition of Mandarin Chinese ba construction.* (Unpublished doctoral dissertation), University of Hawaii, Honolulu, US.

Furman, R. (2012). *Caused motion events in Turkish: Verbal and gestural representation in adults and children* (Doctoral dissertation, Radboud University, Nijmegen, Netherlands). Retrieved from https://pure.mpg.de/rest/items/item_1478431_2/component/file_1478430/content

Furman, R., **Küntay, A. C.**, & **Özyürek, A.** (2014). Early language-specificity of children's event encoding in speech and gesture: evidence from caused motion in Turkish. *Language, Cognition and Neuroscience*, **29**(5), 620–634. https://doi.org/10.1080/01690965.2013.824993

Furman, R., **Özyürek, A.**, & **Allen, S.** (2006). Learning to express causal events across languages: What do speech and gesture patterns reveal? In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 190–201). Cascadilla Press.

Göksun, T., **Küntay, A. C.**, & **Naigles, L. R.** (2008). Turkish children use morphosyntactic bootstrapping in interpreting verb meaning. *Journal of Child Language*, **35**(2), 291–323. https://doi.org/10.1017/S0305000907008471

Gullberg, M., **de Bot, K.**, & **Volterra, V.** (2008). Gestures and some key issues in the study of language development. *Gesture*, **8**(2), 149–179. https://doi.org/10.1075/gest.8.2.03gul

Gullberg, M., & **Narasimhan, B.** (2010). What gestures reveal about how semantic distinctions develop in Dutch children's placement verbs. *Cognitive Linguistics*, **21**(2), 239–262. https://doi.org/10.1515/COGL.2010.009

Haspelmath, M. (2016). The serial verb construction: Comparative concept and cross-linguistic generalizations. *Language and Linguistics*, **17**(3), 291–319. https://doi.org/10.1177/2397002215626895

Hickmann, M., **Hendriks, H.**, **Harr, A.-K.**, & **Bonnet, P.** (2018). Caused motion across child languages: A comparison of English, German, and French. *Journal of Child Language*, **45**(6), 1247–1274. https://doi.org/10.1017/S0305000918000168

Ji, Y., **Hendriks, H.**, & **Hickmann, M.** (2011a). The expression of caused motion events in Chinese and in English: Some typological issues. *Linguistics*, **49**(5), 1041–1077. https://doi.org/10.1515/ling.2011.029

Ji, Y., **Hendriks, H.**, & **Hickmann, M.** (2011b). How children express caused motion events in Chinese and English: Universal and language-specific influences. *Lingua*, **121**(12), 1796–1819. https://doi.org/10.1016/j.lingua.2011.07.001

Jing-Schmidt, Z., & **Tao, H.** (2009). The Mandarin disposal constructions: Usage and development. *Language and Linguistics*, **10**, 29–58.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. Ritchie Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207–227). Mouton and Co.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, **124**(3), 245–266. https://doi.org/10.1037/rev0000059

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, **48**(1), 16–32. https://doi.org/10.1016/S0749-596X(02)00505-3

Li, C. N., & Thompson, S. A. (1976). Subject and topic: A new typology of language. In C. N. Li (Ed.), *Subject and topic* (pp. 457–489). Academic Press.

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. University of California Press.

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2016). Can you handle this? The impact of object affordances on how co-speech gestures are produced. *Language, Cognition and Neuroscience*, **31**(3), 430–440. https://doi.org/10.1080/23273798.2015.1108448

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, **92**(3), 350–371.

McNeill, D. (1986). Iconic gestures of children and adults. *Semiotica*, **62**(1-2), 107–128. https://doi-org.vu-nl.idm.oclc.org/10.1515/semi.1986.62.1-2.107

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

McNeill, D. (2013). Gesture as a window onto mind and brain, and the relationship to linguistic relativity and ontogenesis. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill & S. Tessendorf (Eds.), *Body – language – communication: An international handbook on multimodality in human interaction*, Volume **1** (pp. 28–54). De Gruyter Mouton. https://doi-org.vu-nl.idm.oclc.org/10.1515/9783110261318.28

Mittelberg, I., & Evola, V. (2014). Iconic and representational gestures. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body – language – communication: An international handbook on multimodality in human interaction*, Volume **2** (pp. 1732–1746). De Gruyter Mouton. https://doi.org/10.1515/9783110302028.1732

Morgenstern, A. (2014). Children's multimodal language development. In C. Fäcke (Ed.), *Manual of language acquisition* (pp. 123–142). De Gruyter. https://doi.org/10.1515/9783110302257.123

Müller, C. (2014). Gestural modes of representation as techniques of depiction. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body – language – communication: An international handbook on multimodality in human interaction*, Volume **2** (pp. 1687–1702). De Gruyter Mouton. https://doi/10.1515/9783110302028.1687/html

Özçalışkan, S., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, **96**, B101–B113.

Özçalışkan, Ş., & Goldin-Meadow, S. (2011). Is there an iconic gesture spurt at 26 months? In G. Stam & M. Ishino (Eds.), *Integrating gestures: The interdisciplinary nature of gesture* (pp. 163–174). John Benjamins Publishing Company. https://doi.org/10.1075/gs.4.14ozc

Özyürek, A., Kita, S., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2008). Development of cross-linguistic variation in speech and gesture: Motion events in English and Turkish. *Developmental Psychology*, **44**(4), 1040–1054. https://doi.org/10.1037/0012-1649.44.4.1040

Özyürek, A., Kita, S., Allen, S., Furman, R., & Brown, A. (2005). How does linguistic framing influence co-speech gestures? Insights from crosslinguistic differences and similarities. *Gesture*, **5**(1–2), 219–240.

Özyürek, A., & Özçalışkan, S. (2000). Early language-specificity of children's event encoding in speech and gesture: Evidence from caused motion in Turkish. In *The proceedings of the Thirtieth Child Language Research Forum* (pp. 77–85). CSLI Publications. https://www.tandfonline.com/doi/abs/10.1080/01690965.2013.824993

Piaget, J. (1954). *The construction of reality in the child*. (M. Cook, Trans.). Basic Books. https://doi.org/10.1037/11168-000

Shi, D. (2000). Topic and topic-comment constructions in Mandarin Chinese. *Language*, **76**(2), 383–408. doi:10.1353/lan.2000.0070.

Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömqvist & L. Verhoeven (Eds.) *Relating events in narrative, Volume 2: Typological and contextual perspectives* (pp. 219–257). Lawrence Erlbaum Associates Publishers.

Slobin, D. I., Bowerman, M., Brown, P., Eisenbeiss, S., & Narasimhan, B. (2011). Putting things in places: Developmental consequences of linguistic typology. In J. Bohnemeyer & E. Pedersen (Eds.), *Event representation in language and cognition* (pp. 134–165). Cambridge University Press.

Sloetjes, H. (2017). *ELAN* (5.0.0) [Computer software]. https://archive.mpi.nl/tla/elan/cite

Streeck, J. (2008). Depicting by gesture. *Gesture*, **8**(3), 285–301. https://doi.org/10.1075/gest.8.3.02str

Talmy, L. (2000). *Toward a cognitive semantics, Volume* **2**: *Typology and process in concept structure.* MIT Press.

Tsung, L., & **Gong, Y. F.** (2021). A corpus-based study on the pragmatic use of the ba construction in early childhood Mandarin Chinese. *Frontiers in Psychology*, **11**, 607818. https://doi.org/10.3389/fpsyg.2020.607818

Appendix 1

Causal events shown in the video stimuli

	Description of the video stimuli	Target RVCs in Chinese (with literal translations in English)
Trial 1	A woman smashes a cup into pieces on the floor.	摔碎 'smash-break"
Trial 2	A woman throws a paper ball into a bowl.	扔进去 'throw-enter"
1	A woman tears a piece of paper in half.	撕坏 'tear-break'
2	A man pushes a swing with a tennis racquet, and it moves back and forth.	摇晃 'sway-move'
3	A woman cracks an egg at the edge of a bowl.	磕碎 'knock-break'
4	A man hits a ball off a bench with a tennis racket.	打到地上 'hit-arrive at the ground'
5	A man knocks over a cup tower.	打掉 'hit-fall'
6	A woman trips while she is carrying several eggs, and accidentally smashes them into a bowl.	打碎 'hit-smash'
7	A woman is flipping through a book and accidentally tears a page.	撕坏 'tear-break'
8	A woman breaks a plate with a broom.	打碎 'hit-smash'
9	A man throws a ball out of the door.	扔出门外 'throw-exit the door'
10	A woman cuts a piece of paper into pieces with scissors.	剪碎 'cut-break'
11	A man pushes a swing and it moves back and forth.	摇晃 'sway-move'
12	A man bumps into a swing, causing it to move back and forth.	摇晃 'sway-move'
13	A man is sweeping next to his stack of cups, he turns and accidentally knocks the cups over with the broom handle.	打到地上 'hit-fall to the ground'
14	A woman is playing with a ball, and accidentally kicks it over her head and out of the room.	踢出门外 'kick-exit the door'