Cognitive Development 78 (2026) 101694

Contents lists available at ScienceDirect

Cognitive Development

journal homepage: www.elsevier.com/locate/cogdev

Gesture as a mechanism of change in the interface between spatial language and cognitive development

Dilay Z. Karadöller[a,b,*], Ercenur Ünal[b], Beyza Sümer[c], Demet Özer[d], Aslı Özyürek[b]

[a] Middle East Technical University, Ankara, Turkiye
[b] Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[c] University of Amsterdam, Amsterdam, the Netherlands
[d] Bilkent University, Ankara, Turkiye

A R T I C L E I N F O

Keywords:
Spatial language
Spatial cognition
Gesture

A B S T R A C T

Children often use gestures to express concepts before expressing them in speech, particularly in domains rich in visual-spatial information. For example, spatial relations such as left-right are cognitively and linguistically challenging for children. Consequently, 8-year-olds struggle to convey these complex relations verbally, but they frequently rely on gestures to describe these spatial concepts informatively. This study builds on this prior work on the descriptions of left-right relations to investigate further (1) the differing functions of gestures (complementary or redundant) in relation to speech in both children and adults; (2) the change in semantic information conveyed in both of these types of gestures in childhood to adulthood, aiming to shed light on the interaction between spatial language and spatial cognitive development. Eight-year-old and adult monolingual Turkish speakers described pictures of objects in left-right spatial relations shown among a quartet of other types of spatial relations between these objects. Results demonstrated that when describing left-right relations between objects, children, compared to adults, provided more under-informative descriptions (i.e., using "side" instead of "left-right") in speech but used complementary gestures to convey missing information multimodally. Adults already used informative spatial terms in speech and used gestures mostly redundantly. Moreover, children showed a preference for iconic gestures depicting the relative locations of objects over directional pointing gestures indicating single locations, especially when gestures complemented speech. In contrast, adults showed no reliable preference for either gesture type. These results indicate the significance of gestures as mechanisms of change, alongside speech, in spatial language and cognition, particularly in the context of describing cognitively complex left-right spatial relations between objects.

Children, from an early age, learn how to map linguistic labels of their own language to objects, relations, actions, and events surrounding them. This can often be challenging for children, especially when they need to express cognitively complex and challenging relations, actions, and events. In such instances, children often express these concepts through gestures before expressing them in speech (see Avcılar et al., 2021; Goldin-Meadow, 2015; Karadöller el., 2025a, b for reviews). This is seen especially in domains that are rich in visual-spatial information (e.g., Austin & Sweller, 2018; Calero et al., 2019; Furman et al., 2014; Göksun et al., 2010; Hostetter & Alibali, 2008; Karadöller et al., 2021, 2024; Kita, 2000; Miller et al., 2020; Ünal et al., 2025, 2026; Sauter et al., 2012;

* Correspondence to: Middle East Technical University, Dumlupınar Bulvarı 1, Ankara 06800, Turkiye.
E-mail address: dilayk@metu.edu.tr (D.Z. Karadöller).

https://doi.org/10.1016/j.cogdev.2026.101694
Received 2 July 2025; Received in revised form 25 December 2025; Accepted 4 March 2026

Sekine, 2009; Sekine, 2020). However, how the relationship between speech and gesture develops and changes into adulthood is less understood. It is also less explored how the semantic information conveyed in such gestures varies in relation to speech and across children and adults. In this study, we seek to investigate how the relationship between speech and gesture differs in childhood and adulthood in encoding cognitively challenging spatial relations. To do so, we focused on descriptions of left-right relations by monolingual Turkish-speaking 8-year-old children and adults and investigated (1) the differing functions of gestures (complementary or redundant) in relation to speech among children and adults; (2) the change in semantic information conveyed in gestures (i.e., iconic gestures depicting relative locations of objects or directional pointing gestures indicating single locations for left/right space) in relation to speech and in childhood and adulthood. This will inform our understanding of the possible role of gestures at the interface between spatial language and cognitive development.

We will first synthesize what has been suggested on the role of co-speech gestures in spatial language and cognitive development. Next, we will focus on the domain of conceptually challenging left-right spatial relations between objects, where this role has recently been shown to be more noticeable. Finally, we will introduce our hypotheses and report our experimental design investigating whether and how the role of speech and gestures in spatial descriptions differs between childhood and adulthood to better understand the role of gestures in spatial language development and cognition, drawing evidence from Turkish-speaking children and adults.

1. Role of co-speech gestures in development of spatial language and spatial cognition

Human communication frequently relies on visual-spatial articulators such as gestures (Kendon, 2004; Kita & Özyürek, 2003; McNeill, 1992, 2005; Özyürek, 2014, 2018). Co-speech gestures have different affordances than speech for representing and packaging visual-spatial information (Kita et al., 2017; Kita, 2000). While speech conveys information in arbitrary and linear ways (Levelt, 1980, 1981), gestures convey information in a holistic manner by integrating meaning elements into a single expression (McNeill, 1992) and typically accompanying speech simultaneously (Slonimska et al., 2020, 2021, 2022).

In the context of spatial relations, iconic gestures visually represent the size and shape properties of objects, spatial arrangements between objects, and spatial events through hand shapes, movements, or the relative positioning of the hands (McNeill, 1992). These gestures exhibit visually-motivated similarities between the handshapes and the spatial properties of the objects, event components, locations, or directions depicted (e.g., Furman et al., 2014; Göksun et al., 2010; Karadöller et al., 2024).

Prior work has shown that children express information about some concepts in gestures before expressing them in speech (e.g., Avcılar et al., 2021; Alibali & Goldin-Meadow, 1993; Alibali & Kita, 2010; Capone & McGregor, 2004; Church & Goldin-Meadow, 1986; Goldin-Meadow, 2015; Kita et al., 2017). Such patterns are most evident in domains where visual-spatial information plays a central role, such as geometry (Calero, et al., 2019), caused motion events (Furman et al., 2014), instruments of causal actions (Göksun et al., 2010), spatial reasoning (Ehrlich et al., 2006; Miller et al., 2020), and route descriptions (Austin & Sweller, 2018; Sauter et al., 2012; Sekine, 2009). In such domains, gestures are considered to play a critical role in spatial cognition and language, serving as a bridge between internal representations and communication (e.g., Alibali, 2005; Goldin-Meadow & Alibali, 2013; Hostetter & Alibali, 2008). It is plausible that gestures help children express conceptually challenging concepts by physically grounding abstract mental representations. For example, when representing routes, children map their internal spatial concepts (i.e., directional sequences) onto external hand movements (e.g., gestures tracing the path; Austin & Sweller, 2018; Sauter et al., 2012; Sekine, 2009). This process leverages embodied experiences by utilizing sensorimotor knowledge via gestures before speech (Chu & Kita, 2008; Emmorey, 2014; Goldin-Meadow, 2016; Hostetter & Alibali, 2008; Imai & Kita, 2014; Motamedi et al., 2021, 2024). Thus, gestures serve as bridging tools that transform bodily experiences into communicable representations, scaffolding linguistic expression of complex ideas (Goldin-Meadow, 2015).

Previous research concentrates mainly on instances where gestures precede and function as complements to speech and how such instances play a role in spatial language and cognitive development. However, both adults (e.g., Cooperrider et al., 2016; Karadöller et al., 2021, 2024; Kita et al., 2017; McNeill, 2005; Peeters & Özyürek, 2016) and children (e.g., Capone & McGregor, 2004; Karadöller et al., 2021, 2025a, b; Ping & Goldin-Meadow, 2008) use gestures also redundantly, that is, when both speech and gestures convey similar information. Redundant gestures may reinforce spatial representations, particularly during cognitively challenging tasks (Alibali, 2005; Hostetter & Alibali, 2008). In development, such redundancy may help children acquire concepts through embodied reinforcement (Iverson & Goldin-Meadow, 2005). Although prior accounts of gesture in cognitive change have emphasized the importance of gestures that complement speech during transitional knowledge states—developmental moments characterized by emerging but not yet consolidated understanding (Goldin-Meadow, 2015), the accumulated evidence on gesture-speech redundancy suggests that such redundancy remains meaningful beyond these transitional periods.

Overall, both complementary and redundant uses of gestures underscore claims arguing for a multimodal nature of language (Hagoort & Özyürek, 2024; Karadöller et al., 2025b; Kendon, 2004) and demonstrate how speech and gestures are integrated systems of communication (see McNeil, 1992), with both complementary and redundant uses optimizing spatial encoding through sensorimotor grounding (e.g., Emmorey, 2014; Hostetter & Alibali, 2008; Imai & Kita, 2014; Motamedi et al., 2021, 2024; Perniss et al., 2010; Perniss & Vigliocco, 2014). Recent work has demonstrated that such uses of gestures extend to descriptions of object locations using locative terms that involve cognitively challenging left-right spatial relations between objects (Karadöller et al., 2021; Ünal et al., 2025, 2026).

2. Developmental trajectory of different types of locative terms

Linguistic encoding of spatial relations between objects requires an explicit reference to the figure (the smaller, foregrounded object) and ground (the larger, backgrounded object), as well as their spatial relation (Talmy, 1985). While some descriptions may adopt viewpoint-independent perspectives, left-right and front-behind depend on observer perspective (Landau, 2017; Levinson, 2003; Martin & Sera, 2006). Furthermore, front-behind can benefit from asymmetrical cues (visibility/occlusion) as well as mapping to the asymmetry of the body, whereas left-right lacks distinguishing cues, remaining symmetrical and, thus, being the most conceptually complex relation to describe.

The order in which children acquire locative terms may reflect this conceptual complexity (Johnston & Slobin, 1979), aligning with the view that linguistic input is mapped onto pre-existing conceptual representations (Arunachalam & Waxman, 2010; Hespos & Spelke, 2004). In this view, linguistic emergence patterns reflect underlying conceptual development (Clark, 1973; Dromi, 1987; Huttenlocher et al., 1983; Smiley & Huttenlocher, 1995). Prior empirical findings demonstrate that spatial terms like in, on, and under emerge around age two, followed by front and back, which emerge around age five; however, learning to encode left and right shows a protracted acquisition (Grigoroglou et al., 2019; Johnston & Slobin, 1979; Landau, 2017; Sümer, 2015). Mastering the expression of left-right relations in speech typically occurs at later stages of spatial cognition and language development (Benton, 1959; Corballis & Beale, 1976; Clark, 2004; Harris, 1972; Ünal et al., 2025, 2026; Piaget, 1972; Rigal, 1994, 1996).

Children acquiring many spoken languages often do not use left-right spatial terms to talk about spatial relations between objects until about 8 years of age (Abarbanell & Li, 2021; Karadöller et al., 2021, 2024; Ünal et al., 2025, 2026; Sümer et al., 2014; Sümer, 2015). Instead, they rely on under-informative spatial terms that fail to convey exact spatial information (Karadöller et al., 2021, 2024; Ünal et al., 2025, 2026; Sümer et al., 2014; Sümer, 2015). Recently, it has been discovered that 8-year-old monolingual Turkish-speaking children primarily used terms like "next to" or "side," which often failed to convey precise information on left-right spatial relations between objects (Karadöller et al., 2024; Ünal et al., 2025, 2026; Sümer, 2015). However, an examination of the children's gestures alongside their speech revealed that they frequently enriched these under-informative spoken descriptions with gestures that represented the locations of objects (Karadöller et al., 2024; Ünal et al., 2025, 2026).

In such cases, where spoken expressions are under-informative, children are found to provide information in gesture, either pointing to abstract space, indicating left or right directionality with single points (see Fig. 3b), or positioning their hands in front of their gesture space to depict the relative locations of objects (see Fig. 3a). These findings align well with evidence on the acquisition of spatial expressions in sign languages (Karadöller, 2022; Karadöller et al., 2021, 2024; Sümer, 2015; Sümer et al., 2014). This body of work on the development of spatial descriptions by deaf children shows the expression of left-right relations in adult-like ways earlier (around 5 years of age and beyond) than their speaking peers (around age 8–10). This early acquisition of linguistic forms to express left-right spatial relations, but not necessarily others, such as in-on (see Sümer & Özyürek, 2020), by signing children is attributed to the affordances of visual-spatial and body-anchored nature of sign language expressions (see Karadöller et al., 2025a, b for a review; Sümer, 2015; Sümer & Özyürek, 2020). Thus, signing children have an advantage in expressing left-right relations compared to speaking children, due to the visual-spatial affordances of sign language, which allows them to map space to space.

Summarizing, while foundational research has established how gesture-speech integration supports cognitive and linguistic development across many domains (e.g., Goldin-Meadow, 2015; Iverson & Goldin-Meadow, 2005), less is known about the changes in spatial information conveyed through gesture in childhood and adulthood, particularly when semantic informativeness in spatial speech increases and the opportunity for complementary gestures diminishes. This gap is especially salient for cognitively challenging spatial concepts, such as left-right relations, where gesture may play a unique role. Our study addresses this by examining how the specificity of spatial information in gesture (e.g., conveying information through iconic representations depicting relative locations of objects or directional pointing gestures indicating single locations) varies in childhood and adulthood, alongside the development of informativeness of spatial spoken expressions, illuminating the role of gestures as mechanisms of change at the spatial language and cognition interface.

3. The present study

In order to explore the abovementioned gaps in the literature and build on the previous work on gestures complementing speech, here, we investigate how monolingual Turkish-speaking 8-year-old children and adults describe left-right relations by specifically exploring (1) the differing roles of gestures (whether complementary or redundant) in relation to speech among children and adults

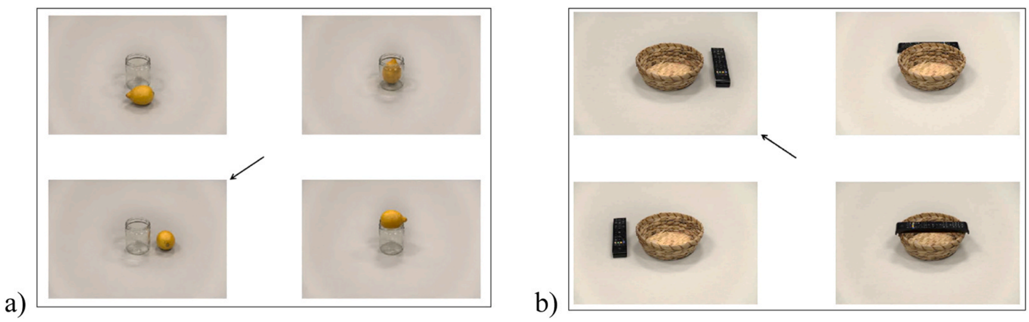


Fig. 1. Examples of non-contrastive (a) and contrastive (b) experimental displays.

and (2) the change of semantic information conveyed through gestures (i.e., iconic gestures depicting relative locations of objects or directional pointing gestures indicating single locations) in relation to speech in childhood and adulthood. This investigation will help us understand the potential role of gestures at the intersection of spatial language and cognitive development, specifically in encoding the cognitively challenging left-right relations between objects.

To do so, we conducted a study involving a picture description task where both children and adults were shown 4-picture displays of two objects in differing spatial relationships to each other (i.e., *lemon in front of the jar*; *lemon in the jar*; *lemon right to the jar*; *lemon on the jar*) and asked to describe the picture pointed with an arrow to an addressee (see Fig. 1 for display examples). Later, we coded the spatial information conveyed in participants' descriptions, both in speech and co-speech gestures.

4. Hypotheses

Based on the previous work on the development of spatial terms (Abarbanell & Li, 2021; Karadöller et al., 2021, 2024; Ünal et al., 2025, 2026; Sümer et al., 2014) and co-speech gestures conveying information missing in speech (Austin & Sweller, 2018; Calero, et al., 2019; Furman et al., 2014; Göksun et al., 2010; Karadöller et al., 2021, 2024; Ünal et al., 2025, 2026; Sauter et al., 2012; Sekine et al., 2009, 2020), we propose three hypotheses:

(1) Children will provide less spatial information in speech but more spatial information in their gestures compared to adults
This will align well with previous accounts demonstrating children's difficulty in expressing left-right terms early in development (e.g., Abarbanell & Li, 2021; Clark, 1973; Karadöller et al., 2021, 2024; Ünal et al., 2025, 2026; Rigal, 1994, 1996; Sümer, 2015; Sümer et al., 2014) and their tendency to ground challenging concepts into here and now through gestures (e.g., Motamedi et al., 2024; Perniss et al., 2010; Perniss & Vigliocco, 2014).

(2) Gestures will transition from being complementary to redundant in childhood to adulthood.
As speech becomes more specific in conveying spatial information in adults, gestures will have less potential to express complementary information (Emmorey, 2014; Imai & Kita, 2014; Karadöller et al., 2021, 2024; Motamedi et al., 2021, 2024; Perniss et al., 2010; Perniss & Vigliocco, 2014; Ünal et al., 2025, 2026). In such instances, gestures can still be used redundantly (e.g., Alibali, 2005).

(3) Semantic information conveyed in gestures will change in relation to changes in informativeness in speech from childhood to adulthood.
Gestures are expected to shift from childhood to adulthood from iconic representations of object relations, such as conveying relative spatial locations by directly mapping perceived relations onto gesture space and thus grounding them in real space (Fig. 3a), to more abstract directional gestures that refer to single locations, for example, pointing gestures to the left or right space (Fig. 3b). This developmental shift is expected to parallel increasing informativeness in speech from childhood to adulthood, with children relying more heavily on iconic gestures grounded in real space.

5. Method

The method reported in this study was approved by the Ethics Review Board of the Radboud University Nijmegen. Additionally, consent from the Survey and Research Commission, Department of the National Education, Ministry of the Republic of Turkey, was obtained to collect data from children in primary schools in İstanbul, Turkey.

6. Participants

Participants consisted of 26 child (*14 females*; Mean Age = 8;6; Age Range = 6;7 – 9;5) and 25 adult (*14 females*; Mean Age = 35;9; Age Range = 19;8 – 50) monolingual Turkish speakers. This data was collected as part of a larger project that also involved deaf signers residing in Turkey. The ages of the adult and child participants were determined to match the ages of the deaf signers recruited for that larger project. We only included monolingual speakers of Turkish to avoid confounds due to knowing another language, which could create differences in the way spatial relations are organized in a different language. One additional child and two adults were tested but excluded from the study because they were bilingual in Turkish and another language. Moreover, children with diagnosed developmental delays and/or disorders (e.g., ADHD) were excluded prior to data collection upon consultation with teachers and school administration. Participation was voluntary. Informed consent was obtained from adults, and verbal assent was obtained from children, prior to the experiment. At the end of the study, adults received monetary compensation of 50 Turkish Liras (approximately $10 at the time of testing), and children received crayons as gifts, also valued at 50 Turkish Liras, for their participation.

7. Materials

Two sets of displays, each containing a total of 84 displays (experimental displays, n = 28; filler displays, n = 56), were created for the current study. Each display had 4 pictures of the same objects in various spatial configurations. The ground objects (bigger objects, e.g., a cup) were always in the center of the pictures, and the figure objects (smaller objects, e.g., a pencil) were located in relation to the ground object. In each display, an arrow pointed to one of the four pictures. The picture that was pointed to with an arrow was the target picture to be described. Experimental displays (n = 28 in each set) consisted of left-right spatial configurations between items (e.g., the pencil is to the left of the cup). In half of the experimental displays, only the target picture contained a left or right spatial configuration between objects, and all other pictures contained spatial configurations other than left-right, such as front, behind, in, or on. In the remaining half of the experimental displays, one additional picture also contained the contrastive spatial configuration (i.e., if the target picture contains the left spatial configuration, the contrast picture contains the right spatial configuration or vice versa), and the remaining pictures contain the front, behind, in, or on spatial configurations. This variation was included to control for different demands for informativeness when describing spatial relations between objects in the target pictures (see Karadöller et al., 2024 and Manhardt et al., 2020, 2021 for a similar procedure). See Fig. 1 for display examples. In addition to the experimental displays, each set included 56 filler displays to prevent excessive emphasis on the left-right spatial configurations. All filler displays consisted of target pictures in front (n = 14), behind (n = 14), in (n = 14), and on (n = 14) spatial configurations between objects.

All visual displays were piloted to make sure both children and adults could identify and name the objects in each display. Within all displays, figure objects were presented only once. Ground objects (e.g., jar) were presented 4 times, but always with a different figure object (e.g., jar-watch, jar-mandarin, jar-lemon, jar-screwdriver). The same ground objects were never presented twice in a row. Moreover, the same spatial configuration as a target picture was not presented more than twice in a row to avoid biases in one type of spatial relationship. There were two sets of displays with the same ground objects but different figure objects. All other configurations were similar across the two sets. The order of the displays and the locations of the pictures in each display were randomized for each participant.

8. Procedure

Participants were seated in front of a computer in a quiet room. Sitting across was an addressee who was a researcher, other than the experimenter delivering the instructions. All tasks were administered in the same order using a Dell Precision M4800 laptop with software Presentation NBS 16.4 (Neurobehavioral Systems, Albany, CA). Instructions for all tasks were given orally to avoid potential misunderstandings that could arise from written instructions. We applied the same procedure to adults to ensure identical experimental strategies. The researcher provided identical instructions to children and adults without modeling how the participants should engage with the task. Before each task (i.e., Familiarization and Description), participants were provided with three practice trials. These trials were repeated if necessary.

Each participant first performed the Familiarization task, followed by the Description task. The familiarization task aimed to introduce participants to the general complexity of the displays, which featured a 2 × 2 grid with two objects in various spatial configurations relative to each other. To ensure that participants were paying attention to the displays, they were engaged in a mock task where they were randomly presented with pictures of one of the objects within the displays and asked to report whether or not the presented item was part of the previous display. These questions randomly appeared after 30 of the 84 displays. Participants answered these questions by pressing a green button to indicate “Yes” or a red button to indicate “No” on the keyboard. Before starting the familiarization task, participants were given three practice trials that had different displays from the original display set.

Next, participants performed the Description task. In this task, participants were instructed to describe the target picture to the addressee during the visual white noise screen after each trial. Participants were informed that the addressee had the display with the same pictures in a randomized order without the arrow indicating the target picture on their tablets. Following the participant’s description, the addressee pretended to choose the picture described on her tablet. The addressee did not give feedback on whether the description was correct or not. In cases where spatial information is missing in the participants' descriptions, the addressee only asks where the figure item is located (e.g., "Where is the pencil?"). After completing each trial, participants moved to the next one by pressing the ENTER key on the keyboard. Participants were randomly assigned to either set of displays and received the other set in the Description task. See Fig. 2 for the illustration of the experimental flow of tasks.

Participants were video recorded from two cameras. One camera was placed to capture the front view from a horizontal angle, and the other was placed to capture the side view from a high angle for later coding of the speech and gestures during the description task. Participants were not instructed to gesture or not to gesture to obtain gestures spontaneously. Participants’ eye gaze was also recorded. Yet, this paper only focuses on the description data.

9. Coding and reliability

All descriptions were transcribed and coded for Target Pictures by native speakers of Turkish. Speech and gesture coding was done using ELAN (Version 4.9.3), which is a free annotation tool (http://tla.mpi.nl/tools/tla-tools/elan/) for multimedia resources

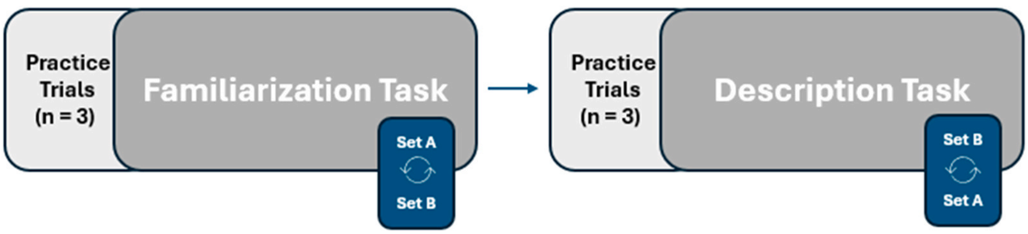


Fig. 2. Illustration of the flow of the Familiarization and Description Tasks and utilization of display sets.

developed by the Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands (Wittenburg et al., 2006).

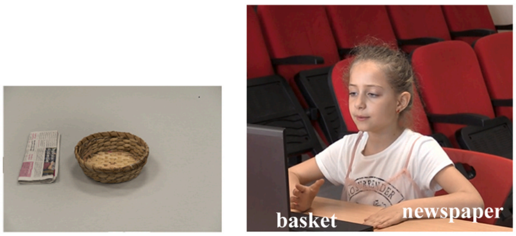
9.1. Speech

Speech coding was done per description to identify the presence and use of spatial terms produced by participants that convey information on the spatial relation between the items. Participants described the pictures using informative spatial nouns (i.e., left *sol* or right *sağ*) to refer to the target spatial relationship between the objects in their speech or using under-informative general nouns (i.e., side y*an*). In some instances, participants just labeled the figure and ground objects but not the relation between them, or provided descriptions that referred to the target relation using other specific spatial terms (e.g., front *ön*). We did not have a reliability coding for speech, as speech coding involved the presence of spatial terms (e.g., left) that were unambiguously heard and identified by a native speaker of Turkish. However, the accuracy of the coding was checked by another native speaker of Turkish to ensure no errors were made during the initial coding.

9.2. Gesture

Gesture coding was done per description, regardless of the speech used in the description. We coded for spontaneous co-speech spatial gestures as identified by strokes (see Kita, van der Hulst, & van Gijn, 1998; see also Karadöller et al., 2024 for a similar approach) that convey information concerning the location of the two items or the spatial relation between the items. We did not consider other types of gestures, such as beat gestures or gestures representing the size or shape of the items. We also did not code the number of gestures used in a single trial.

Participants used two types of gestures that convey spatial information. The first type included iconic gestures that convey relative locations of both objects by grounding them in the gesture space on the table in front of them. See Fig. 3a for an example of a description from a child using iconic gestures in her description that convey the relative location of two objects (Basket and Newspaper). The second type includes either a directional gesture that indicates the location of a single object by pointing to left-right areas in gesture space or by locating a single object on the left or right of their gesture space on the table. For this category, participants only indicate the location of one of the objects via gestures. See Fig. 3b for an example from a child using a directional gesture indicating the location of only one item (Toothbrush) being on the left side of the gesture space. Twenty percent of the data (5 children and 5 adults) was also coded by another coder for the reliability checks. There was substantial agreement for the type of spatial gestures used to localize figure (88% Agreement, kappa = 0.77) and ground (92% Agreement, kappa = 0.79) items. All disagreements were discussed to reach 100% agreement for 20% of the data. Next, the primary coder audited the remaining 80% of data to align coding with resolved disagreement patterns.

a)


Turkish: Örgülü sepet var, onun yanında da bir tane gazete var.

English: There is a basket, next to it, there is a newspaper.

b)


Turkish: Bardağın solunda diş fırçası var.

English: The toothbrush is to the left of the cup.

Fig. 3. Example of a) iconic gestures that convey the relative location of two objects (Basket and Newspaper); b) directional gestures that convey a single location of one of the objects (Toothbrush) to the left side of the gesture space. Underlined words denote the gesture onset.

As a final step, we categorized the descriptions based on whether they unimodally (i.e., speech alone) or multimodally (i.e., speech and gestures combined) provide a uniquely referring informative expression that distinguishes the spatial relation between the objects in the target picture from other pictures in the display (see Karadöller et al., 2024 for a similar approach). There were also a few instances of missing or incorrect spatial information in participants' responses, which were categorized accordingly. See also Table 1 for the frequency of each description.

1. Unimodal Descriptions: Descriptions that fall under this category are unimodal and do not include any spatial gestures. The first type of descriptions in this category consists of descriptions that are informative in speech only (i.e., left or right). The other type of descriptions in this category consists of descriptions that are under-informative in speech only (i.e., side).
2. Multimodal Descriptions: Descriptions that fall under this category include speech together with co-speech gestures. The first type of description in this category consists of descriptions that are informative in speech (i.e., left or right) coupled with spatial gestures of either form. Thus, for this category, spatial gesture use was redundant, as speech already provides information concerning the spatial relations between items (See Fig. 3b.) The other type of descriptions in this category consists of descriptions that are under-informative in speech (i.e., side) coupled with spatial gestures of either from, which disambiguate the missing information in speech. Thus, for this category, spatial gesture use was complementary to speech (See Fig. 3a)
3. Missing and Incorrect Descriptions: Descriptions that fall under this category are either unimodal or multimodal and consist of descriptions that include specific spatial terms like front, as well as descriptions where the spatial relationship was not specified, and participants only identified the objects without explaining their spatial relationship between them. Even when gestures were considered alongside speech, these descriptions remained under-informative. In other words, gestures did not provide additional information beyond what was already conveyed through speech.

10. Results

Data were analyzed using generalized binomial linear mixed-effects modeling regression (*glmer*) with random intercepts for Subjects and Items. This mixed-effects approach enabled us to account for the random variability resulting from the inclusion of different participants and items. All models were fitted using the *lme4* package (version 1.1.17; Bates et al., 2014) in R (R Core Team, 2018) with the optimizer *bobyqa* (Powell, 2009). The analysis script and the data can be found on OSF: https://osf.io/fpxm7/overview?view_only=dc410df42bee4403b97226ac3d37ac92

First, we investigated whether children provide less spatial information in speech compared to adults. Children produced informative speech in 32% of their descriptions and under-informative speech in 66%, whereas adults produced informative speech in 86% of their descriptions and under-informative speech in 12% (see Table 1). To statistically test this difference, we focused on all descriptions irrespective of gesture use, and examined the fixed effect of age group dummy-coded with children as the reference group (0 = children, 1 = adults) on binary values for the presence of informative spatial terms (1 = informative spatial term use; 0 = under-informative spatial term use) at the item level as the dependent measure. Because of dummy coding, the Intercept indicates if there is a reliable difference in the distribution of informative and under-informative speech for children. The fixed effect of age group then indicates whether and how this distribution differs for adults compared to children. The model revealed a significant negative intercept ($\beta = -3.31$, $SE = 1.16$, $p = .004$), indicating that children were less likely to produce informative speech than under-informative speech. The model also revealed a significant positive fixed effect for age group ($\beta = 7.66$, $SE = 1.72$, $p < .001$), indicating that this pattern was reversed for adults, and they were more likely to produce informative speech compared to children.

Next, we tested whether children provide more spatial information in their gestures compared to adults. Children used gestures in 64% of their descriptions, whereas adults used gestures in 32% of their descriptions (see Table 1). To statistically test this difference,

Table 1
Mean proportions (Standard deviations) and frequencies for all descriptions as a function of age group and type of description.

	Adults		Children	
	Mean (SD)	*N*	*Mean (SD)*	*N*
Unimodal Descriptions				
Informative Speech-Only	0.62 (0.49)	400	0.15 (0.36)	96
Under-Informative Speech-Only	0.04 (0.19)	25	0.19 (0.39)	121
Multimodal Descriptions				
Informative Speech with Redundant Gesture	0.24 (0.43)	153	0.17 (0.37)	107
└ Directional Single Location	0.11 (0.31)	69	0.05 (0.22)	34
└ Iconic Relative Location	0.13 (0.34)	84	0.11 (0.32)	73
Under-Informative Speech with Complementary Gesture	0.08 (0.28)	53	0.47 (0.50)	302
└ Directional Single Location	0.03 (0.18)	21	0.14 (0.35)	90
└ Iconic Relative Location	0.05 (0.22)	32	0.33 (0.47)	212
Missing or Incorrect Descriptions	0.02 (0.03)	13	0.02 (0.04)	20

Notes. Proportions were calculated separately for adults and children. *N* indicates the total number of descriptions observed in each category.

we focused on all descriptions regardless of the type of speech, and examined the fixed effect of age group dummy-coded with children as the reference group (0 = children, 1 = adults) on the binary values for the presence of gesture use (1 = present; 0 = absent) at the item level as the dependent measure. Because of dummy coding, the Intercept indicates if there is a reliable difference in the distribution of gesture presence and absence for children. The fixed effect of age group then indicates whether and how this distribution differs for adults compared to children. The model revealed a significant positive intercept ($\beta = 1.12$, $SE = 0.50$, $p = .025$), indicating that children were more likely to produce gestures than not. The model revealed a significant negative fixed effect for age group ($\beta = -2.60$, $SE = 0.72$, $p < .001$), indicating that this pattern was reversed for adults, and they were less likely than children to produce gestures.

Then, we investigated the differing functions of gestures (complementary or redundant) across adults and children. Children used gestures complementarily in 47% of their descriptions and redundantly in 17%, whereas adults used gestures complementarily in 8% of their descriptions and redundantly in 24% (see Table 1). To statistically test this difference, we focused only on the multimodal descriptions and examined the fixed effect of age group dummy-coded with children as the reference group (0 = children, 1 = adults) on the binary dependent measure (1 = informative speech with redundant gesture, 0 = under-informative speech with complementary gesture) at the item level. Because of dummy coding, the Intercept indicates if there is a reliable difference in the distribution of redundant and complementary gestures for children. The fixed effect of age group then indicates whether and how this distribution differs for adults compared to children. The model revealed a significant negative intercept ($\beta = -4.02$, $SE = 1.78$, $p = .024$), indicating that children were less likely to produce gestures redundantly than complementarily. The model revealed a significant positive fixed effect for age group ($\beta = 8.14$, $SE = 2.88$, $p < .01$), indicating that this pattern was reversed for adults, and they were more likely to produce gestures redundantly compared to children.

Finally, we investigated whether and how the semantic information conveyed in gestures varies in relation to speech across adults and children. Children produced iconic relative location gestures complementarily in 33% of their descriptions and redundantly in 11%, and they produced directional single location gestures complementarily in 14% of their descriptions and redundantly in 5%. On the other hand, adults produced iconic relative location gestures complementarily in 5% of their descriptions and redundantly in 13%, and they produced directional single location gestures complementarily in 3% and redundantly in 13% of their descriptions.

Because these descriptive patterns show substantial age-related differences in the frequency distribution of gesture types, we conducted separate statistical analyses for children and adults. Specifically, we focused only on the multimodal descriptions and conducted separate binomial mixed-effects models for children and adults on the binary dependent variable (1 = Iconic Relative Location Gesture, 0 = Directional Single Location Gesture) at the item level. The model also included the fixed effect of gesture function, dummy coded with complementary gestures as the reference level (0 = complementary, 1 = redundant). Because of dummy coding, the Intercept, indicates if there is a reliable difference in the distribution of iconic relative location and directional single location gestures when these gestures were complementary to speech. The fixed effect of gesture function then indicates how this distribution differs for redundant gestures compared to complementary gestures.

The model for children revealed a significant positive intercept ($\beta = 1.05$, $SE = 0.53$, $p = .047$), indicating that children were more likely to produce iconic relative location gestures than directional single location gestures when these gestures were complementary to speech. The model also had a negative parameter estimate for the fixed effect of gesture function ($\beta = -0.37$, $SE = 0.48$, $p = 0.440$), suggesting that iconic relative location gestures tended to be used less frequently when gestures were redundant to speech compared to when they were complementary; nevertheless, this difference did not reach significance.

The model for adults revealed that neither the intercept ($\beta = 0.33$, $SE = 0.63$, $p = .601$) nor the fixed effect of gesture function ($\beta = -0.25$, $SE = 0.59$, $p = .667$) was statistically significant. This indicates that there was no statistically significant evidence that adults preferentially produce iconic relative location compared to directional single location gestures when these gestures were complementary to speech. Moreover, the distribution of iconic relative location and directional single location gestures did not differ statistically between when they were redundant or complementary.

11. Discussion

In this study, we investigated the potential role of co-speech gestures at the interface between the development of spatial language and cognition. To do so, we focused on how monolingual Turkish-speaking 8-year-old children and adults describe cognitively complex left-right relations and examined (1) the differing functions of gestures (complementary or redundant) among children and adults; (2) the change in the semantic information conveyed through gestures in relation to speech in childhood and adulthood. Our results demonstrated that when describing left-right relations between objects, children provide more under-informative descriptions in speech but use complementary gestures to convey missing information multimodally. This pattern was reversed for adults, who already used informative spatial terms frequently in their descriptions, and when they produced gestures, they mostly did so redundantly. Moreover, our findings reveal an age-related difference in the use of gesture types. Children produced iconic relative location gestures more frequently when these gestures complemented their speech. Although the frequency of iconic relative location gestures appeared to decrease in redundant contexts compared to complementary contexts, this difference was not statistically significant. Adults, however, showed no consistent preference for one gesture type over another, and their patterns remained stable regardless of whether these gestures were complementary or redundant. This suggests that children's gesture choices are more sensitive to the semantic contribution that gestures make to under-informative speech, unlike adults' choices, as these types of gestures tend to diminish in adulthood when speech becomes more informative. This shift in gestural representations is accompanied by a corresponding shift in speech. Moreover, iconic gestures seem to ground information about the spatial relation observed when speech is less informative and specific, possibly scaffolding conceptual mastery and playing a role in driving the alignment of cognitive and linguistic systems.

11.1. The challenge of left-right in spatial language development

First of all, our findings substantiate the broader empirical consensus that children experience fundamental challenges in acquiring left-right distinctions conceptually and linguistically (Abarbanell & Li, 2021; Clark, 1973; Karadöller et al., 2021, 2024; Ünal et al., 2025, 2026; Rigal, 1994, 1996; Sümer, 2015; Sümer et al., 2014). Notably, even at age 8, children used left-right terms in significantly lower frequencies. Previous work showed that this difficulty persists even among deaf children acquiring sign languages (Karadöller et al., 2024), despite the body-anchored and visual-spatial nature of these descriptions in sign languages and their relatively earlier acquisition of linguistic forms to describe left-right spatial relations between objects (Sümer, 2015; Sümer et al., 2014). Collectively, these results highlight that the cognitive challenges of left-right distinctions surpass the modality of communication, pointing to domain-general complexity rather than modality-specific constraints.

11.2. Differing functions of gestures (complementary and redundant)

Our result about children's tendency to disambiguate under-informative spatial terms through the use of gestures confirms the previous evidence demonstrating that gestures complement children's speech for conceptually challenging domains, especially those that rely on visual-spatial reasoning (e.g., Alibali & Goldin-Meadow, 1993; Alibali & Kita, 2010; Austin & Sweller, 2018; Capone & McGregor, 2004; Church & Goldin-Meadow, 1986; Calero et al., 2019; Furman et al., 2014; Göksun et al., 2010; Karadöller et al., 2021, 2024; Miller et al., 2020; Ünal et al., 2025, 2026; Sauter et al., 2012; Sekine et al., 2009).

This prior work on the role and function of gestures during childhood primarily focuses on cases where gestures are used to complement missing spatial information in speech (e.g., Karadöller et al., 2024; Ünal et al., 2025, 2026). In fact, gestures can convey the same information already expressed in speech, and thus, can be used redundantly by adults as well as by children (see Karadöller et al., 2025a, b and Özer & Göksun, 2020 for reviews). Here, we contribute to the existing literature by demonstrating that the use of redundant gestures for left-right relations is present in both children and adults, even after they have mastered the use of spatial terms in speech, confirming that gestures are inherent components of the language system (Kendon, 2004; McNeill, 1992). Rather than disappearing with linguistic mastery, gestures provide continuous embodied grounding of spatial concepts.

Building on theories of embodiment (e.g., Chu & Kita, 2008; Goldin-Meadow, 2016), it is possible that variability in gesture use may reflect individual differences in the recruitment of sensorimotor experiences during spatial encoding (see also Özer & Göksun, 2020). For instance, individuals who rely more strongly on embodied representations might produce more gestures to ground abstract spatial concepts (e.g., left-right) in physical experience (e.g., using body-anchored pointing to leverage proprioceptive awareness). Additionally, the cognitive demands of the task might predict gesture production, with higher demands prompting increased gesture use to offload processing (also see Hostetter and Alibali, 2008). Thus, embodiment frameworks help explain why gestures persist even when speech is adequate by providing a multimodal foundation for conceptualizing and communicating spatial relations.

Overall, our findings add to previous accounts (e.g., Goldin-Meadow, 2016 and Hostetter & Alibali, 2008) by demonstrating that gestures persist in relation to their deep integration with spatial language and cognition. They constitute the multimodal foundation of language (see Hagoort & Özyürek, 2024; Karadöller et al., 2025a, b), operating from early scaffolding to adult-like usage, with redundancy also serving as a resource throughout the lifespan.

11.3. Semantic information conveyed through gestures in relation to speech differs in childhood and adulthood

Our findings also reveal age-related differences in the types of gestures used. Children tended to produce iconic gestures depicting the relative locations of objects primarily when these gestures complemented their speech. Although the use of such gestures appeared to decline in redundant contexts, this decrease was not statistically significant. By contrast, adults showed no reliable preference for one gesture type over another, and their gesture use remained stable across complementary and redundant contexts—although they produced fewer gestures overall than children. Building on the prior work on the types of changes in speech and gesture during development (e.g., Alibali & Goldin-Meadow, 1993; Alibali & Kita, 2010; Austin & Sweller, 2018; Calero et al., 2019; Capone & McGregor, 2004; Church & Goldin-Meadow, 1986; Furman et al., 2014; Goldin-Meadow, 2015; Göksun et al., 2010; Karadöller et al., 2021, 2024; Miller et al., 2020; Ünal et al., 2025, 2026; Sauter et al., 2012; Sekine et al., 2009), our findings also suggest a developmental shift in the semantic information encoded in gestures when expressing left-right relations. Specifically, children appeared to rely more on two-handed placement gestures that convey relative locations in an analog and grounded manner. This reliance seems to diminish in adulthood. One possibility is that such gestures reflect children's spatial conceptualizations, which may initially be more grounded in real space analog mappings and gradually become more abstract and directional as speech becomes more specific and informative with age. Because two-handed iconic gestures directly represent spatial relations in physical space, they are well-suited to compensating for missing spatial information in speech (e.g., Austin & Sweller, 2018; Calero et al., 2019; Furman et al., 2014; Göksun et al., 2010; Karadöller et al., 2021, 2024; Miller et al., 2020; Sauter et al., 2012; Sekine et al., 2009). By contrast, directional gestures that merely point to left-right space may be less informative in under-informative speech contexts, as they do not specify which object occupies the left or right position. Directional pointing gestures indicating single locations may therefore be more effective when speech itself explicitly encodes left-right information.

Children's reliance on grounded and iconic gestures to express spatial concepts (e.g., object locations) suggests that gestures serve as a cognitive and communicative bridge, helping them to externalize sensorimotor experiences when they are confronted with abstract linguistic categories (e.g., Clark, 2006; Glenberg & Kaschak, 2002; Perniss & Vigliocco, 2014; Volterra et al., 2017), such as left-right. This reliance aligns well with theories of embodied cognition, where physical actions scaffold the internalization of complex spatial language (Clark, 2006; Glenberg & Kashcak, 2002; Hostetter & Alibali, 2008; Perniss & Vigliocco, 2014). Thus, the prevalence of grounded iconic gestures in children's descriptions supports the idea that language development progresses from concrete and embodied representations to abstract symbolic systems. In this abstraction process, gestures may act as a "stepping stone" (Goldin-Meadow, 2015), helping children to anchor challenging spatial concepts in physical experience before fully integrating them into speech (Perniss & Vigliocco, 2014).

From a communicative perspective, children's reliance on more grounded and iconic gestures when describing object locations resonates well with the work on child-directed communication (Campisi & Özyürek, 2013; Kandemir et al., 2024; Motamedi et al., 2024; Perniss et al., 2010). It seems that children not only themselves produce these grounded and iconic representations when communicating conceptually challenging concepts, but they are also exposed more to iconic descriptions of events or objects when their parents (Perniss et al., 2010) or other adults (Campisi & Özyürek, 2013; Kandemir et al., 2024) communicate with them (see also Motamedi et al., 2024 for a review). The fact that children are both exposed to (via caregivers' and adults' child-directed communication) and produce iconic representations may imply a feedback loop in early communication. That is, while adults use iconic gestures/descriptions (Campisi & Özyürek, 2013; Kandemir et al., 2024; Perniss et al., 2010; Motamedi et al., 2024) when communicating with children, children in turn, may use ways to simplify abstract concepts into perceptually accessible forms. This way, adults may be modeling how to map bodily experiences onto linguistic labels and/or reinforcing children's own multimodal strategies for conceptual mastery. In sum, these findings position gesture as a dynamic interface between embodied experience and linguistic abstraction, offering insights into how children bootstrap complex spatial descriptions through multimodal interactions.

11.4. Limitations and future directions

Our findings reflect participants' multimodal encoding patterns when describing spatial relations to an adult interlocutor. However, prior research indicates that adults adjust their gesture production based on perceived listener knowledge, particularly by simplifying and increasing the use of iconic gestures with children (e.g., Campisi & Özyürek, 2013; Kandemir et al., 2024; Perniss et al., 2010). Critically, in our study, child participants also had an adult listener, raising important questions: Would children exhibit different spatial encoding strategies (e.g., reduced gesture complexity, simplified speech) when communicating with peer addressees? This bidirectional flexibility, where both adults and children can recalibrate multimodal strategies based on partner expectations, may shed light on whether and how communicative context influences spatial encoding. Future studies should manipulate interlocutor characteristics to examine whether adults amplify the informativeness of gestures for left-right relations with child listeners, and whether children adjust their speech and gestures when addressing peers versus adults, potentially revealing developmental milestones in adapting communication patterns according to partner characteristics.

This study provides insights into multimodal encoding of conceptually challenging spatial relations, but is constrained by a couple of limitations. First, our focus on 8-year-old children and adults restricts understanding of developmental trajectories, as the absence of younger children (e.g., 4–6 years) or adolescents leaves gaps in how spatial strategies emerge and consolidate across critical stages. Second, we did not directly assess children's comprehension of left-right, which prevented us from determining whether the observed shift from gesture reliance to speech is related to conceptual understanding and/or broader changes in integrating gestures into language structures. Future research should: (1) systematically investigate age-related changes across expanded cohorts to map the development of left-right use; (2) incorporate standardized measures of left-right comprehension (e.g., Rigal, 1994; Gilligan-Lee et al., 2021) to test if mastery of this specific spatial concept predicts the timing of the transition; and (3) employ longitudinal designs tracking gesture-speech mismatches (e.g., Church & Goldin-Meadow, 1986), particularly for left/right concepts, alongside linguistic and conceptual proficiency. Such approaches could clarify whether mismatches signal learning readiness while disentangling conceptual maturation in spatial understanding.

Another limitation of the present work was that measures of children's spatial skills or linguistic proficiency were not included. Consequently, we were unable to investigate how variability in these abilities influences gesture strategies in encoding cognitively challenging left-right spatial relations between objects. Future research should systematically examine whether and how individual differences in spatial reasoning (e.g., mental rotation, perspective-taking) and language skills (e.g., vocabulary, syntactic complexity) predict gesture production patterns in describing spatial relations (see Özer & Göksun, 2020). Understanding these links would offer critical insights into the cognitive prerequisites for multimodal communication and potentially expand our understanding of language acquisition from a multimodal lens (Karadöller et al., 2025a, b).

12. Conclusions

This study explores how co-speech gestures mediate the interplay between spatial language and cognitive development by analyzing how monolingual Turkish-speaking 8-year-olds and adults verbally and gesturally encode cognitively demanding left-right spatial configurations between objects. We specifically found (1) the differing functions of gestures (complementary vs. redundant) in relation to speech among children and adults and (2) the change in the semantic information conveyed in gestures in childhood to adulthood. Findings revealed that children, unlike adults, produced less informative descriptions of left-right relations in their speech; however, they compensated for this missing information by using complementary gestures. Children showed a preference for iconic gestures indicating the relative location of objects over directional pointing gestures, especially when these gestures complemented speech. In contrast, adults showed no reliable preference for either gesture type. Our results add to the literature on gestures preceding language development in children (Alibali & Goldin-Meadow, 1993; Austin & Sweller, 2018; Calero et al., 2019; Ehrlich et al., 2006; Furman et al., 2014; Göksun et al., 2010; Hurst et al., 2022; Karadöller et al., 2021, 2024; Miller et al., 2020; Sauter et al., 2012; Sekine, 2009). We extend this body of research by demonstrating how gestures complement speech and what functions they may serve in scaffolding conceptual mastery of spatial concepts and spatial language development. Furthermore, our study highlights the significance of gestures as mechanisms of change (Goldin-Meadow, 2015), illustrating how they progress from being more to less grounded in childhood to adulthood (Emmorey, 2014; Imai & Kita, 2014; Motamedi et al., 2021, 2024; Perniss et al., 2010; Perniss & Vigliocco, 2014), accompanied by changes in speech, particularly in the context of describing cognitively complex left-right spatial relations between objects (Karadöller et al., 2024).

Author note

This research is supported by the NWO VICI Grant awarded to the last author. Portions of this work were presented at several international conferences between 2019 and 2023. We thank our research assistants, Hükümran Sümer, Ekin Tünçok, İrem Yeşilçavdar, and Yeşim Özüer for their help in collecting, transcribing, and coding data. The authors are very grateful to Jeroen Geerts for helping us with processing the video data.

CRediT authorship contribution statement

Aslı Özyürek: Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Demet Özer:** Writing – review & editing, Data curation. **Dilay Z. Karadöller:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ercenur Ünal:** Writing – review & editing, Methodology, Data curation, Conceptualization, Formal analysis. **Beyza Sümer:** Writing – review & editing, Methodology, Data curation, Conceptualization.

Data Availability

OSF link provied withing the manuscript. Link directs to data and the analysis script.

References

Abarbanell, L., & Li, P. (2021). Unraveling the contribution of left-right language on spatial perspective taking. *Spatial Cognition & Computation*, 1–38.

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial cognition and Computation, 5*(4), 307–331.

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child′ s state of mind. *Cognitive Psychology, 25*(4), 468–523.

Alibali, M. W., & Kita, S. (2010). Gesture highlights perceptually present information for speakers. *Gesture, 10*(1), 3–28.

Arunachalam, S., & Waxman, S. R. (2010). Language and conceptual development. *WIREs Cognitive Science, 1*(4), 548–558.

Austin, E. E., & Sweller, N. (2018). Gesturing along the way: Adults' and preschoolers' communication of route direction information. *Journal of Nonverbal Behavior, 42*(2), 199–220.

Avcılar, G., Bilir, B., Cihankaya, Ş., Coşkun, Z., Piri, İ., & Ünal, E. (2021). Jestlerin çocukların dil-iletişim ve bilişsel becerilerinin gelişimindeki destekleyici rolü. *Psikoloji ÇalışMalarıaı, 41*(3), 789–816.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Benton, A. (1959). *Right-left discrimination and finger localization*. New York: Hoeber-Harper.

Calero, C. I., Shalom, D. E., Spelke, E. S., & Sigman, M. (2019). Language, gesture, and judgment: Children's paths to abstract geometry. *Journal of Experimental Child Psychology, 177*, 70–85.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics, 47*(1), 14–27.

Capone, N. C., & McGregor, K. K. (2004). Gesture development. Gesture.

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General, 137*(4), 706.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23*(1), 43–71.

Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E.Moore. (Ed.), *Cognitive development and acquisition of language* (pp. 65–110). Academic Press.

Clark, E. V. (2004). How language acquisition builds on cognitive development. *Trends in Cognitive Sciences, 8*(10), 472–478.

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences, 10*(8), 370–374.

Cooperrider, K., Gentner, D., & Goldin-Meadow, S. (2016). Spatial analogies pervade complex relational reasoning: Evidence from spontaneous gestures. *Cognitive research: principles and implications, 1*, 1–17.

Corballis, M. C., & Beale, I. L. (1976). *The psychology of left-right*. Hillsdale, NJ: Erlbaum.

Dromi, E. (1987). *Early lexical development*. Cambridge University Press.

Ehrlich, S. B., Levine, S. C., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology, 42*(6), 1259.

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), 20130301. https://doi.org/10.1098/rstb.2013.0301

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.

Gilligan-Lee, K. A., Hodgkiss, A., Thomas, M. S., Patel, P. K., & Farran, E. K. (2021). Aged-based differences in spatial language skills from 6 to 10 years: Relations with spatial and mathematics skills. *Learning and Instruction, 73*, Article 101417.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558–565.

Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2010). How do preschoolers express cause in gesture and speech? *Cognitive Development, 25*(1), 56–68.

Goldin-Meadow, S. (2015). From action to abstraction: Gesture as a mechanism of change. *Developmental Review, 38*, 167–184.

Goldin-Meadow, S. (2016). L'enfant parle d'abord avec les mains (Children speak first with their hands). *Enfance, 4*(4), 435–443.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology, 64*(1), 257–283.
Grigoroglou, M., Johanson, M., & Papafragou, A. (2019). Pragmatics and spatial language: The acquisition of front and back. *Developmental Psychology, 55*(4), 729.
Hagoort, P., & Özyürek, A. (2024). Extending the architecture of language from a multimodal perspective. *Topics in Cognitive Sciene, 17*(2025), 877–887.
Harris, L. (1972). Discrimination of left and right, and development of the logic relations. *Merrill-Palmer Quarterly, 18*, 307–320.
Hespos, S. J., & Spelke, E. S. (2004). Conceptual precursors to language. *Nature, 430*(6998). Article 6998.
Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15*, 495–514.
Hurst, M. A., Wong, A., Gordon, R., Alam, A., & Cordes, S. (2022). Children's gesture use provides insight into proportional reasoning strategies. *Journal of Experimental Child Psychology, 214*, Article 105277.
Huttenlocher, J., Smiley, P., & Charney, R. (1983). Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review, 90*(1), 72–93.
Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), Article 20130298.
Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*(5), 367–371.
Johnston, J. R., & Slobin, D. I. (1979). The development of locative expressions in English, Italian, Serbo-Croatian and Turkish. *Journal of Child Language, 6*(3), 529–545.
Kandemir, S., Özer, D., & Aktan-Erciyes, A. (2024). Multimodal language in child-directed versus adult-directed speech. *Quarterly Journal of Experimental Psychology, 77*(4), 716–728.
Karadöller, D. Z., Sümer, B., Ünal, E., & Özyürek, A. (2024). Sign advantage: Both children and adults' spatial expressions in sign are more informative than those in speech and gestures combined. *Journal of Child Language, 51*(4), 876–902.
Karadöller, D. Z., Sümer, B., & Özyürek, A. (2025a). First-language acquisition in a multimodal language framework: Insights from speech, gesture, and sign. *First Language, 45*(6), 673–710.
Karadöller, D. Z., Sümer, B., & Özyürek, A. (2025b). Advancing the multimodal language acquisition framework through collaborative dialogue. *First Language, 45*(6), 797–809.
Karadöller, D. Z., Sümer, B., Ünal, E., & Özyürek, A. (2021). Spatial Language Use Predicts Spatial Memory of Children: Evidence from Sign, Speech, and Speech-plus-gesture. In T. Fitch, C. Lamm, H. Leder, & K. Tessmar (Eds.), *Proceedings of the 43th Annual Conference of the Cognitive Science Society (CogSci 2021)* (pp. 672–678). Vienna: Cognitive Science Society.
Karadöller, D. Z. (2022). Development of Spatial Language and Memory: Effects of Language Modality and Late Sign Language Exposure. Published Doctoral Dissertation, Radboud University Nijmegen. MPI Series in Psycholinguistics, Nijmegen, The Netherlands. ISBN: 978-94-92910-33-2.
Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
Kita, S. (2000). How representational gestures help speaking. *Language and gesture, 1*, 162–185.
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.
Kita, S., & Özyurek, 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.
Kita, S., van der Hulst, H., & van Gijn, I. (1998). Movement phases in signs and co- speech gestures, and their transcription by human coders. In I. Wachsmuth, & M. Fröhlich (Eds.), *Gesture and sign language in human-computer interaction* (pp. 23–35). Berlin: Springer.
Landau, B. (2017). Update on "what" and "where" in spatial language: A new division of labor for spatial terms. *Cognitive Science, 41*, 321–350.
Levelt, W. J. M. (1980). On-line processing constraints on the properties of signed and spoken language. In U. Bellugi, & M. Studdert-Kennedy (Eds.), *Signed and spoken language: Biological constraints on linguistic form* (pp. 141–160). Verlag Chemie.
Levelt, W. J. M. (1981). The speaker's linearization problem. *Philosophical Transactions of the Royal Society of London, 295*(1077), 305–315.
Levinson, S. C. (2003). *Space in language and cognition: explorations in cognitive diversity*. New York, NY: Cambridge University Press.
Manhardt, F., Brouwer, S., & Özyurek, A. (2021). A tale of two modalities: Sign and speech influence in each other in bimodal bilinguals. *Psychological Science, 32*(3), 424–436.
Manhardt, F., Özyürek, A., Sümer, B., Mulder, K., Karadöller, D. Z., & Brouwer, S. (2020). Iconicity in spatial language guides visual attention: A comparison between signers' and speakers' eye gaze during message preparation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(9), 1735–1753.
Martin, A. J., & Sera, M. D. (2006). The acquisition of spatial constructions in American Sign Language and English. *Journal of Deaf Studies and Deaf Education, 11*(4), 391–402.
McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
Miller, H. E., Andrews, C. A., & Simmering, V. R. (2020). Speech and gesture production provide unique insights into young children's spatial reasoning. *Child Development, 91*(6), 1934–1952.
Motamedi, Y., Murgiano, M., Grzyb, B., Gu, Y., Kewenig, V., Brieke, R., … Vigliocco, G. (2024). Language development beyond the here-and-now: Iconicity and displacement in child-directed communication. *Child Development, 95*(5), 1539–1557.
Motamedi, Y., Murgiano, M., Perniss, P., Wonnacott, E., Marshall, C., Goldin-Meadow, S., & Vigliocco, G. (2021). Linking language to sensory experience: Onomatopoeia in early language development. *Developmental Science, 24*(3), Article e13066.
Özer, D., & Göksun, T. (2020). Gesture use and processing: A review on individual differences in cognitive resources. *Frontiers in Psychology, 11*, Article 573555.
Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), Article 20130296.
Özyürek, A. (2018). Cross-linguistic variation in children's multimodal utterances. In In. M. Hickmann, E. Veneziano, & H. Jisa (Eds.), *Sources of variation in first language acquisition: Languages, contexts, and learners* (pp. 123–138). Amsterdam: Benjamins.
Peeters, D., & Özyürek, A. (2016). This and that revisited: A social and multimodal approach to spatial demonstratives. *Frontiers in Psychology, 7*, 222.
Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology, 1*, 227.
Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B, 369*.
Piaget, J. (1972). Judgment and reasoning in the child. Littlefield, Adams. (Originally work published 1928).
Ping, R. M., & Goldin-Meadow, S. (2008). Hands in the air: using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology, 44*(5), 1277.
Powell, M.J. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, 26-46.
R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL ⟨https://www.R-project.org/⟩.
Rigal, R. (1994). Right-left orientation: Development of correct use of right and left terms. *Perceptual and Motor Skills, 79*(3), 1259.
Rigal, R. (1996). Right-left orientation, mental rotation, and perspective-taking: When can children imagine what people see from their own viewpoint? *Perceptual and Motor Skills, 83*(3), 831–843.
Sauter, M., Uttal, D. H., Alman, A. S., Goldin-Meadow, S., & Levine, S. C. (2012). Learning what children know about space from looking at their hands: The added value of gesture in spatial communication. *Journal of Experimental Child Psychology, 111*(4), 587–606.
Sekine, K. (2009). Changes in frame of reference use across the preschool years: A longitudinal study of the gestures and speech produced during route descriptions. *Language and Cognitive Processes, 24*, 218–238.

Sekine, K. (2020). Changes in frame of reference use across the preschool years: A longitudinal study of the gestures and speech produced during route descriptions. *Speech Accompanying-Gesture* (pp. 218–238). Psychology Press.

Slonimska, A., Özyürek, A., & Capirci, O. (2020). The role of iconicity and simultaneity for efficient communication: The case of Italian Sign Language (LIS). *Cognition, 200*, Article 104246.

Slonimska, A., Özyürek, A., & Capirci, O. (2021). Using depiction for efficient communication in LIS (Italian Sign Language). *Language and Cognition, 13*(3), 367–396.

Slonimska, A., Özyürek, A., & Capirci, O. (2022). Simultaneity as an emergent property of efficient communication in language: A comparison of silent gesture and sign language. *Cognitive Science, 46*(5), Article e13133.

Smiley, P., & Huttenlocher, J. (1995). Conceptual development and the child's early words for events, objects, and persons. In M. Tomasello, & W. Merriman (Eds.), *Beyond names for thing: Young children's acquisition of verbs* (pp. 21–62). Erlbaum.

Sümer, B. (2015). Acquisition of spatial language by signing and speaking children: A comparison of Turkish Sign Language (TİD) and Turkish. [Doctoral dissertation]. Radboud University Nijmegen, Nijmegen.

Sümer, B., & Özyürek, A. (2020). No effects of modality in development of locative expressions of space in signing and speaking children. *Journal of Child Language, 47* (6), 1101, 31.

Sümer, B., Perniss, P. M., Zwitserlood, I. E. P., & Özyürek, A. (2014). Learning to express "left-right" & "front-behind" in a sign versus spoken language. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 1550–1555). Cognitive Science Society.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. *Language Typology and Syntactic Description, 3*, 57–149.

Ünal, E., Karadöller, D. Z., & Özyürek, A. (2026). Children sustain their attention on spatial scenes when planning to describe spatial relations multimodally in speech and gesture. *Developmental Science, 29*(2), Article e70128.

Ünal, E., Kırbaşoğlu, K., Karadöller, D. Z., Sümer, B., & Özyürek, A. (2025). Gesture reduces mapping difficulties in the development of spatial language depending on the complexity of spatial relations. *Cognitive Science, 49*(2), Article e70046.

Volterra, V., Capirci, O., Caselli, M. C., Rinaldi, P., & Sparaci, L. (2017). Developmental evidence for continuity from action to gesture to sign/word. *Language, Interaction and Acquisition, 8*(1), 13–41.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*.