Development of Supplementary Gestures with Speech in Mandarin-Speaking Children

Kawai Chui (kawai@nccu.edu.tw)
Department of English & Research Center for Mind, Brain, and Learning, National Chengchi University
No. 64, Sec. 2, Zhinan Rd., Wenshan Dist., Taipei, Taiwan (R.O.C.)
Chan-Tat Ng (tatng@g.nccu.edu.tw)
Department of Psychology, National Chengchi University
No. 64, Sec. 2, Zhinan Rd., Wenshan Dist., Taipei, Taiwan (R.O.C.)
***Huei-Mei Liu (liumei@ntnu.edu.tw)**
Department of Special Education, National Taiwan Normal University
No. 129, Shida Rd., Da'an Dist., Taipei, Taiwan (R.O.C.)
*corresponding author
Feng-Ming Tsao (tsaosph@ntu.edu.tw)
Department of Psychology, National Taiwan University
No. 1, Sec. 4, Roosevelt Rd., Da'an Dist., Taipei, Taiwan (R.O.C.)

Abstract

This study examined the developmental trajectory of supplementary gestures, which conveyed information distinct from speech, and their semantic relationships with accompanying speech in 16 Mandarin-speaking children. Children were observed during mother–child interactions between 12 and 21 months of age at three-month intervals, with minimal age variation within each time point. Findings showed a marked increase in supplementary gestures at 18 months, alongside the predominance of declarative pointing within gesture-speech acts and one-word utterances. This specific cross-modal pattern remained stable through 21 months. Most cross-modal combinations conveyed entity ideas, with predicative ideas occurring infrequently and propositional ideas even less often. The frequently expressed 'entity-related', 'what is done to what', and 'who initiates what and who/what is affected' knowledge was linguistically encode only about three to six months later than in gesture-speech acts. The overall findings support the enduring role of supplementary gestures in enriching children's communicative expressiveness within a single act.

Keywords: supplementary gesture, cross-modal semantic patterns, gesture-speech combination, speech equivalent, communicative development

Introduction

Gestural communication develops along with language acquisition (Capirci et al., 2005; Colonnesi et al., 2010; Butcher & Goldin-Meadow, 2000; Goldin-Meadow, 2015; Iverson & Goldin-Meadow, 2005; LeBarton et al., 2015; McNeill, 1992, 2015; Özçalişkan & Goldin-Meadow, 2014; Rowe & Goldin-Meadow, 2009; Rowe et al., 2008). Children typically begin using gestures between 8 and 12 months of age. The earliest to emerge are deictic gestures, which are associated with context-dependent objects, people, locations, or events. These are most often expressed through pointing, showing, or giving objects, each serving distinct communicative goals. Pointing gestures are produced to request, share interests, or provide information (Liszkowski et al., 2004, 2006; Liszkowski & Tomasello, 2011; Tomasello et al., 2007). Children later use conventional gestures, such as waving to signify 'goodbye', and iconic gestures that convey semantic information, for example, raising a hand to indicate something flying high. Around the age of two, beat gestures emerge; these are brief, rhythmic flicks of one or both hands, moving up-and-down or back-and-forth, which accompany the rhythm or stressed syllables of speech (Capone & McGregor, 2004; McNeill, 1992; Özçalişkan & Goldin-Meadow, 2005; Özçalişkan et al., 2016). The combination of gesture and speech marks a pivotal milestone in communicative development, signifying the integration of two modalities for the expression of meaning. When children combine gestures with words, these cross-modal occurrences can foreshadow the onset of two-word speech. For example, a child might say 'nap' while pointing to the bird, conveying two distinct meanings simultaneously; later, this evolved into a two-word utterance such as 'bird nap' (Iverson & Goldin-Meadow, 2005). Studies of Taiwanese children have relied on parental reports to examine verbal and nonverbal development, including some questions about gesture use (Chi, 2002; Chiou & Liu, 2018; Liu & Chen, 2015; Liu & Tsao, 2010). These questionnaires overlooked gesture-speech combinations, leaving unexplored how children use gestures to enhance expressiveness by conveying non-redundant information beyond speech content.

Özçalişkan and Goldin-Meadow (2005) pioneered the examination of how supplementary gestures conveyed information not expressed in speech from 14 to 22 months. They categorized gesture-speech combinations into "two or three arguments without a predicate," "a predicate with one or two arguments," and "two predicates with or without arguments." Their findings showed that gesture-speech combinations provided additional arguments and/or predicates within utterances beyond what multi-word speech encoded across ages. In a subsequent study of children aged 14 to 34 months (Özçalişkan & Goldin-Meadow, 2009), supplementary gestures expressed argument–predicate relationships before their appearance in speech alone; however, once encoded linguistically, supplementary gesture-speech use declined. Fasolo and D'Odorico (2012) reported similar results in Italian children at 18 months. Compared to complementary gestures, which conveyed meanings overlapping with speech, supplementary combinations predicted greater utterance complexity at 24 months, such as expressions with a verb and two or more arguments. What remains unclear from these studies is the nature of cross-modal argument-predicate relations.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

To address this gap, we investigated the developmental trajectory of supplementary gestures, and their semantic relationships with accompanying speech in 16 Mandarin-speaking children. We examined children's gesture and spoken language production longitudinally at 12, 15, 18, and 21 months during naturalistic mother-child interactions. An activity-based setting facilitated dyadic communication, ensuring equal opportunities for verbal and gestural expression. We observed communicative acts conveying entity, predicative, and propositional information across these age points, which were optimal for capturing the emergence, timeline, and cross-modal semantic relations of supplementary gesture-speech combinations, and their verbal equivalents. The study would provide a comprehensive understanding of the role of supplementary gestures in children's progression in coordinating and mapping distinct information across modalities to express integrated meanings within single communicative acts.

Methods

Participants

Sixteen Taiwanese children (7 male; 9 female) and their mothers participated in this longitudinal study, which followed the children from 12 to 21 months of age, with minimal age variation within each group. Each dyad completed four sessions at three-month intervals. Mean ages for each session are presented in Table 1. Participants were recruited via social media parenting groups and local infant laboratories. All children were being raised in monolingual Mandarin-speaking families, in which caregivers primarily communicated with them in Mandarin Chinese. According to parental reports, none of the children exhibited notable intellectual disabilities, sensory impairments or significant emotional difficulties. No cases of language disorders or family history of diagnosed developmental disorders were reported. The study protocol was approved by the local research ethics committee, and informed consent was obtained from all participants during the initial session. Their participation was remunerated.

Table 1. Mean child age at each session (days)

	Female (*N*=9)	Male (*N*=7)
12-month session	384.1 (11.13)	375.6 (7.58)
15-month session	477.3 (3.87)	465.4 (4.92)
18-month session	553.0 (5.48)	558.8 (7.48)
21-month session	659.6 (3.93)	647.4 (5.53)

Note. Standard deviations are in parentheses.

Data Collection

Gesture and speech data were collected in a playroom designed to study interactions between children and their mothers. The room was equipped with two wall-mounted cameras capturing multiple angles of the participants and objects, along with two camcorders positioned near the activity area for close-up audio-video recording. During every session, the mother and child sat facing each other on a mat, and the child was free to move around. Each pair took part in a standardized 48-minute session, comprising 15 minutes of free play, 10 minutes viewing 19 pictures selected based on the common early words among Taiwanese children (Chi, 2002), and five semi-structured tasks adapted from the Communication Play Protocol (Adamson et al., 2009). These tasks were employed to encourage the use of speech and gestures in response to unexpected events, such as when a new toy suddenly appeared amid ongoing activity. All activities were administered by an assistant. Dyads participated at three-month intervals, yielding a total of 64 sessions across 16 children.

Transcription and Categorization

Children's speech production was transcribed. The analysis focused on utterances that included discernible words. Following Özçalışkan and Goldin-Meadow (2005), 'words' were sounds used consistently to refer to specific objects, actions, or qualities such as *əm* for 'to eat', onomatopoeia like *būbū* for 'car', conventional evaluative expressions such as *wa* to indicate surprise, and actual words, for example, *huāhuā* for 'flower' and *dǎkāi* for 'open'. We also classified each utterance by word count (e.g., one-word, two-word, or three-word).

Children's gestures were categorized into "declarative pointing" (using index finger or whole hand to share interest or information), "imperative pointing" (using index finger or whole hand to request), "showing" (holding up objects for the addressee to see), "giving" (placing objects in or near the addressee's hand), "iconic" (depicting semantic information, such as hands moving up and down to indicate drumming), or "conventional" (expressing culturally established meanings, such as palms upward to signal disappearance). Each gesture's referent or meaning was identified. In order to assess the role of supplementary gestures (e.g., pointing to a giraffe while saying *hóng*sè 'red') in relation to other gesture types, the analysis included both gesture-only productions and complementary gestures expressing information similar to speech (e.g., saying *yúyú* 'fish' while showing a toy fish) for comparison. Supplementary gestures with speech were further distinguished by idea type: "entity" (e.g., *zān* 'three'+point to apples to mean 'three apples'), "predicative" (e.g., *yáoyáo* 'shake'+point to a rattle for 'shake the rattle'), and "propositional" (e.g., *zǒulù* 'walk'+point to a bird for 'bird walks'). For each idea type, corresponding semantic patterns were identified, such as "number-entity" ('three apples'), "action-affected entity" ('shake the rattle'), and "agent-action" ('The bird walks'). The same sets of semantic patterns were employed to examine whether they were also realized in children's word combinations, both within and across age groups.

To assess inter-rater reliability, a second independent coder, who was blind to the study's objectives, analyzed a randomly selected 20% subset of the dataset. Agreement was high across categories: 98.12% for speech content, 98.67% for speech type, 97.46% for gesture type, 97.02% for gestural meaning, 100% in distinguishing gesture-only acts from the two types of gesture-speech combinations, 89.69% for the relation between gesture and speech,

92.20% for semantic relations in cross-modal cases, and 86% for the type of idea conveyed by gesture-speech combinations. Disagreements were resolved through discussion.

Data Analysis

To model developmental changes, speech and gesture outcomes were analyzed using Bayesian generalized mixed-effects models, with random intercepts for child; age point was modeled as an ordinal predictor. Weakly informative priors were used to regularize estimates under sparse cell counts. Speech was analyzed using a binary contrast capturing the emergence of multi-word utterances (1 vs. 2+ words), as well as an ordinal analysis of utterance length across four categories (1, 2, 3, and 4+ words). Gesture analyses included contrasts between gesture-only and gesture-speech productions, categorical analyses of gesture types within gesture-speech combinations (declarative pointing, imperative pointing, showing, giving, iconic, conventional), a gesture modality analysis (gesture-only, complementary, supplementary), and a categorical analysis of idea types within supplementary combinations (entity, predication, proposition). Models yielded estimated probabilities with associated 95% credible intervals, reported in the Results and summarized in Figures 1–5. For analyses involving finer-grained semantic patterns with sparse observations (Figures 6–8), results are shown descriptively to illustrate distributional trends rather than formal statistical inference.

Results

Speech Production

Between 12 and 21 months of age, we identified 4,502 word-containing utterances, with speech remaining overwhelmingly one-word through 18 months and then diversifying by 21 months (Figure 1).

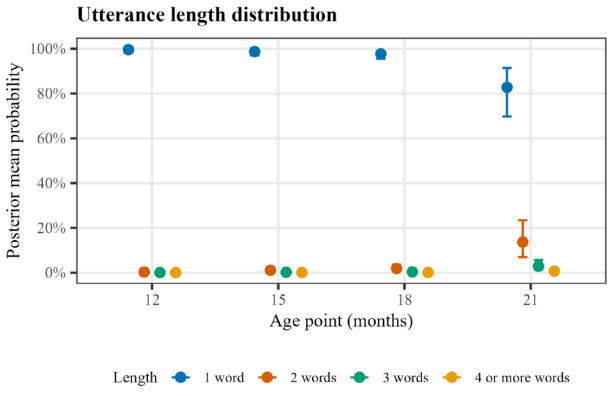


Figure 1. Utterance lengths across age points.

One-word utterances dominated early production, with posterior probabilities of 1.00 [.98, 1.00] at 12 months, .99 [.97, 1.00] at 15 months, and .98 [.96, .99] at 18 months. By 21 months, the probability of one-word utterances declined to .83 [.70, .91], accompanied by a clear rise in multi-word production. Specifically, two-word utterances increased to .14 [.07, .23] at 21 months, and longer utterances, though still infrequent, were present (three-word: .03 [.01, .06]; four-or-more word: .01 [.00, .01], respectively). Overall, speech across the second year was marked by sustained reliance on single-word utterances followed by a late shift toward greater utterance-length diversity by 21 months.

Production of Gesture-Speech Combinations

A total of 2,555 gestures were identified between 12 and 21 months—12.33% observed at 12 months, 19.53% at 15 months, 31.82% at 18 months, and 36.32% at 21 months. These distributions reflect a consistent upward trajectory in children's overall gesture production.

Beyond changes in overall gesture frequency, children increasingly integrated gestures with speech. Specifically, the estimated probability of producing speech-gesture combinations, rather than gestures alone, rose across age points (Figure 2): .02 [.01, .05] at 12 months, .15 [.06, .29] at 15 months, .50 [.30, .71] at 18 months, and .61 [.40, .80] at 21 months. This pattern indicates a developmental shift from unimodal gesture toward cross-modal communication through the second year of life.

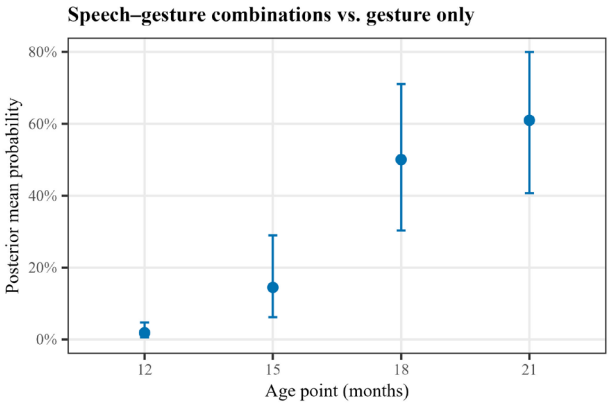


Figure 2. Gesture-speech occurrences across age points.

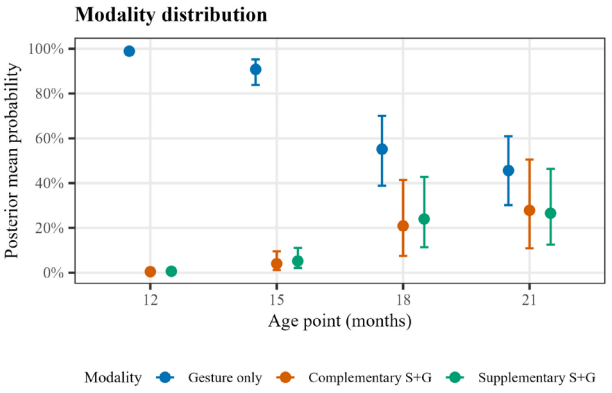


Figure 3. Production of gesture-only, complementary, and supplementary gesture-with-speech.

The three-level modality model clarified the nature of this shift (Figure 3). Gesture-only acts were predominant at 12 months (.99 [.97, 1.00]) and remained common at 15 months (.91 [.84, .95]), but declined substantially by 18 months (.55 [.39, .70]) and further by 21 months (.46 [.30, .61]). In contrast, complementary and supplementary combinations increased with age. Complementary combinations rose from .00 [.00, .01] at 12 months to .28 [.11, .51] at 21 months; supplementary combinations increased from .01 [.00, .02] at 12 months to .27 [.13, .46]

at 21 months. These combinations were comparable in frequency across age points, indicating that gestures were increasingly used to convey information that extended beyond spoken content.

Within the 1,177 cross-modal combinations, gesture type distributions also changed with age (Figure 4). At 12 months, conventional gestures were most prevalent (.56 [.18, .87]) while declarative pointing was lower (.22 [.02, .57]). From 15 months onward, declarative pointing became the dominant type (.69 [.58, .80] at 15 months; .64 [.56, .71] at 21 months), whereas imperative pointing stayed secondary across age points (.08–.13). Showing gestures were near-zero at 12 months and remained low from 15 to 21 months (.07–.08), and giving gestures were generally low but showed a brief increase at 18 months (.14 [.07, .23]) before returning to .06 [.02, .10] at 21 months; iconic gestures were rare throughout (≤ .04). Together, these patterns reflect children's developmental progress in joint attention and referential communication, driven primarily by the rise of declarative pointing.

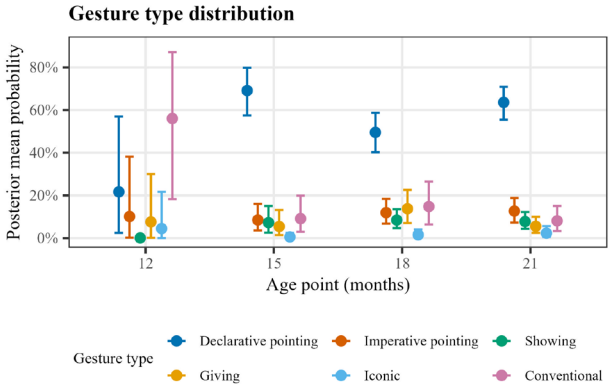


Figure 4. Types of gestures in gesture-speech combinations across ages.

Idea Types within Supplementary Combinations

A total of 441 supplementary gesture-speech combinations were identified, with the idea types conveyed varied across ages (Figure 5).

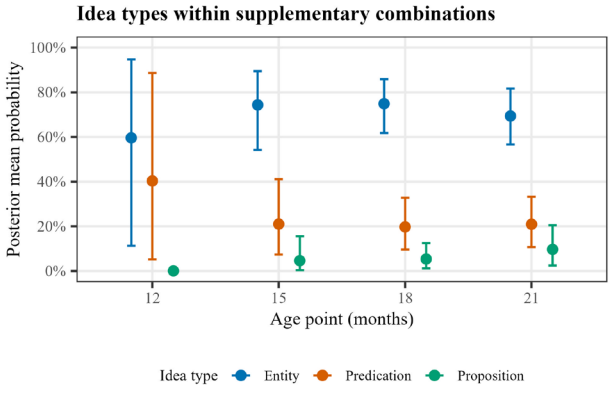
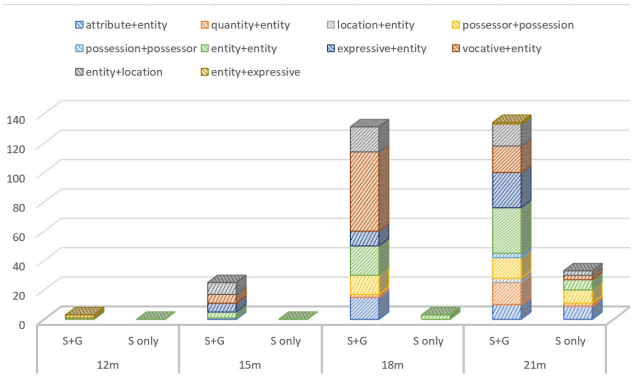


Figure 5. Types of ideas in supplementary-speech combinations across ages.

Entity-related ideas were the most frequent at all ages, with estimated probabilities of .60 [.11, .95] at 12 months, .74 [.54, .90] at 15 months, .75 [.62, .86] at 18 months, and .69 [.57, .82] at 21 months. Predicative ideas were less frequent and relatively stable (.20–.21 between 15 and 21 months). Propositional ideas were rare early in development (.00 at 12 months) but increased gradually, reaching .10 [.02, .21] by 21 months. These results indicate that supplementary gesture-speech combinations mainly conveyed entity information throughout development.

Cross-Modal Semantic Relations in Supplementary Combinations

Semantic relations in supplementary combinations derived from distinct information types integrated across modalities. Because data were sparse for certain patterns and gesture categories, descriptive distributional analyses were conducted. Among 293 entity instances, ten semantic patterns characterized the meaning relationships in these combinations. Eight patterns involved gestures specifying entities or places, with words expressing attributes (*lánsè* 'blue'), quantity (*liù* 'six'), location (*nǎlǐ* 'there'), possessor (*bàba* 'father') or possession (*yǎnjīng* 'eye'), another entity, expressive content (*wā* 'surprise'), or forms of address (*māma* 'mother'). The remaining two patterns paired entity words in speech with gestures conveying locative or expressive information (e.g., pointing to a place and hand-shaking for disagreement). These cross-modal patterns were rare at 12 months ('S+G' in Figure 6). By 15 months, five patterns appeared as children frequently uttered a vocative or expressive word alongside a referential gestural act, or an object word with a gesture indicating location, accounting for 80% of combinations. Their occurrences decreased to 62% at 18 months and 43% at 21 months, as new patterns emerged and others increased in frequency, including attribute+entity, quantity+entity, location+entity, possessor+possession, and possession+possessor. These diverse cross-modal patterns were rarely realized solely through speech, with only 36 instances of word combination from 12 to 21 months ('S only'). Utterances containing two nominal words, such as *māmā* 'mother' *bàbà* 'father', appeared by 18 months but remained infrequent. By 21 months, six semantic patterns appeared in the speech data, still fewer in number than those conveyed through gesture-speech combinations for entity ideas.



*In each combination, the left component represents linguistic information, and the right component is gestural information; both separated by the plus sign (+).

Figure 6. Entity patterns in supplementary gestures with speech (S+G) versus speech alone (S only).

In communicating predicative ideas, nine patterns were identified among 133 supplementary combinations. Seven patterns paired a verb in speech (*ná* 'take') with gestures providing additional information—the affected entity, action manner, event location, instrument, another involved entity, consecutive action, or expressive content. The other two patterns were predicates depicted by gestures alongside a spoken affected entity or vocative term. Cross-modal predication was rare at 12 and 15 months, but the pattern of spoken predicates with gestures indicating affected entities emerged at 15 months (Figure 7). This supplementary combination became predominant thereafter, comprising 75% of instances at 18 months and 78% at 21 months. In the last age point, diverse patterns appeared, with gestures conveying locative, instrumental, conjoined action or entity, or expressive information. Speech-only predication totaled 79 instances, nearly all observed at 21 months. Of these, 99% conveyed a predicate–affected entity relationship (e.g., *yào* 'want' *tiēzhǐ* 'sticker'). One case involved the verb *zuò* 'sit' followed by the noun *báiyún* 'white cloud', expressing an action-location relation. Gesture-speech integration produced richer predicative meanings than speech alone.

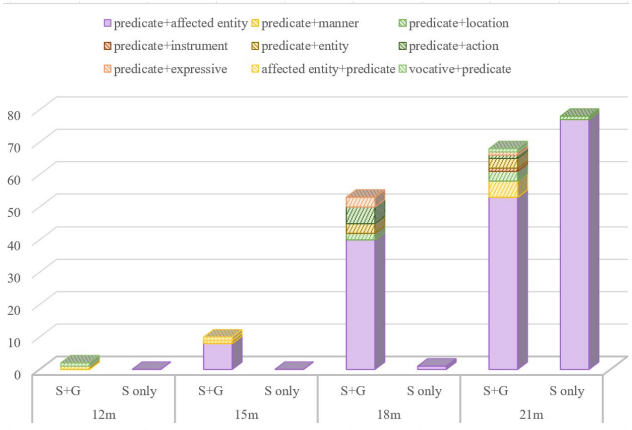
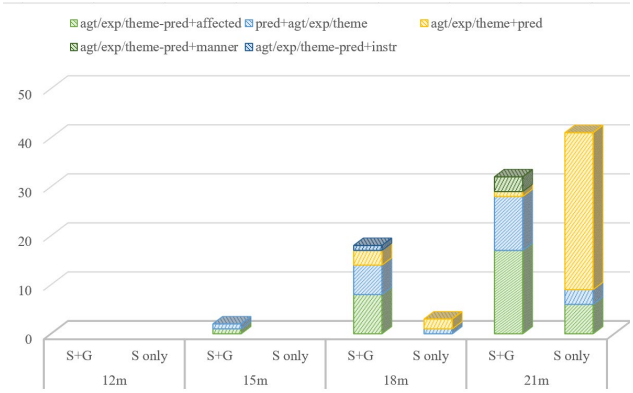


Figure 7. Predicative patterns in supplementary gestures with speech (S+G) versus speech alone (S only).

Finally, a total of 52 instances of propositional cross-modal patterns were recognized, rarely observed before 15 months but showing a marked increase at 18 months and consistently present by 21 months (Figure 8). These patterns combined agents, experiencers, or themes with predicative components, and gestures enriched spoken utterances in diverse ways. In one cross-modal pattern, speech encoded the agent/experiencer/theme and the predicate, while a supplementary gesture depicted the affected referent (e.g., *māma ná* 'mother takes' accompanied by pointing to the target container with a toy inside). In another pattern, speech conveyed the predicate while gesture indicated the agent/experiencer/theme (e.g., *hǎo zhòng* 'very heavy' with pointing to a large roly-poly). A third pattern involved speech expressing only the agent/experiencer/theme, with predicative information conveyed in gesture (e.g., *húdié* 'butterfly' combined with a flapping gesture). Other patterns had speech encode these components while gestures provided further manner information (e.g., *níngning ná* 'Ningning (child's name) takes (giraffe)' with a hold-object gesture) or instrumental information (e.g., *māma cā* 'mother wipes (my nose)' with a give-tissue gesture). In contrast, speech-only realizations of these supplementary relationships totaled 44 instances, primarily featuring agent/experiencer/theme combined with predicate in spoken words (e.g., *āyí kāichē* 'Aunt drives'), which accounted for 85% of cases. Verbal representations demonstrate a narrower range of semantic relations than those expressed through gesture-speech integration.



*Abbreviations: 'agt/exp' as 'agent/experiencer'; 'pred' as predicate, 'instr' as instrument; 'affected' as affected entity. Linguistic components precede the plus sign (+) and the hyphen (-) indicates the linguistic sequence; gestural components follow the plus sign.

Figure 8. Propositional patterns in supplementary gestures with speech (S+G) versus speech alone (S only).

Discussion

When children use supplementary gestures with speech to express meanings, their integrative capability to coordinate and map distinct pieces of information across modalities to enhance expressiveness within a single communicative act is revealed. This study examined the development and nature of this ability in Mandarin-speaking children across four sessions of participation from 12 to 21 months, exhibiting minimal age variation within each time point. This consistency sharpened inference about age-related change in supplementary gesture-speech use, conceptual knowledge underlying their semantic relationships with accompanying speech, and speech equivalents of the cross-modal combined meanings.

One-word utterances dominated across all age points in children's language development; two-word combinations remained rare through 18 months before increasing markedly by 21 months. Across the same period, gesture production increased steadily and gesture-speech integration rose sharply by 18 months, suggesting that multimodal communication growth preceded the surge in multi-word speech. Among gesture types, declarative pointing was consistently the most common for establishing joint attention to referents and events in real-time interactional contexts. Children also continued using other gesture types to achieve different communicative goals despite variability in usage rates, demonstrating the crucial role of gestures in developing the multimodal communicative system alongside language acquisition.

Gestural communication underwent notable changes. Across the second year, children shifted from predominantly gesture-only production to increasingly frequent gesture-speech integration. This transition became especially clear around 18 months, when one-word utterances still dominated and children more often used gesture and speech together to convey meanings. This integrative developmental trend aligns with observations of English-speaking children, who similarly transitioned from using gestures alone to combining gestures with speech between 14 and 22 months of age (Özçalışkan & Goldin-Meadow, 2005). Our study specified the shift for Chinese children at around 18 months, a phase dominated by one-word productions. Of the combinations, English and Italian children were found to have significantly more redundant combinations early in development (Butcher & Goldin-Meadow, 2000; Fasolo & D'Odorico, 2012; Iverson & Goldin-Meadow, 2005). In contrast, Chinese children produced gestures that conveyed information either to enrich or to overlap with speech content, with complementary and supplementary gesture-speech combinations increasing with age at comparable rates.

Regardless of how frequent they were produced, supplementary gestures were highly effective in conveying diverse complex messages. Their integration with speech made children's underlying knowledge more explicit, particularly during the stage when children could not yet fully express through language. During the observed period, the large majority of cross-modal combinations conveyed entity ideas, highlighting the significance of entity-based cognition prior to two years of age. Predicative and propositional ideas were infrequently expressed, likely due to the greater demands of cognitive maturity and effort to represent actions involving entities, diverse entity roles within events, causal relations, and events situated in time and space, among others. Although children might not fully master the associations among these components, the emergence of certain predicative and propositional relations was nonetheless evident in the production of supplementary gestures with speech.

We identified three sets of cross-modal semantic patterns to characterize children's knowledge of entity, predicative, and propositional ideas. In most entity-related patterns, gestures specified entities and words encoded attributes, quantity, location, possession, evaluation, and vocative terms. These supplementary relationships were pervasive and revealed the diversity of children's entity knowledge. Verbal equivalents of these conjoined information appeared three to six months later, indicating a developmental lag relative to their cross-modal counterparts. For predicative ideas, most semantic patterns involved spoken verbs combined with supplementary gestures indicating entities, manner, location, instrument, evaluation, and associated actions. The most frequent pattern at 18 months was the uttering of a predicative word with a gesture representing the entity affected by the event. This 'what is done to what' knowledge was not yet completely expressed in spoken form at the same age, but became dominant by 21 months. The cross-modal representation of propositional ideas typically comprised a lexical predicate accompanied by a gesture that indicated the initiating entity of the event, or the mention of these two components in speech with a gesture representing the affected entity. These patterns revealed children's understanding of 'who initiates what and who/what is affected', appearing from 18 months of age, whereas expression through speech alone did not become frequent until three months later. Many supplementary semantic patterns still lacked speech equivalents by 21 months, indicating the need for a longer period of linguistic development. These findings corroborate previous studies and affirm gestures as precursors to language acquisition (Iverson & Goldin-Meadow, 2005; Özçalışkan & Goldin-Meadow, 2005; Özçalışkan et al., 2016; Rowe & Goldin-Meadow, 2009).

Cognitive accounts have been proposed to explain the precedence of supplementary gesture-speech combinations over articulation entirely through speech, suggesting that gestures demand less cognitive effort and help alleviate cognitive load. We also consider that combining words to form a syntactic unit is cognitively challenging as it requires understanding linguistic constructs such as constituency and grammatical structure—concepts not inherent in the spatio-motoric manifestation of information through gestures. Meanwhile, children may as well have produced supplementary combinations for communicative purposes. Experimental studies by Liszkowski and colleagues (2004, 2006) showed that 12-month-olds pointed declaratively and significantly more often in the joint-attention condition as the experimenter alternated gaze between the event and the infant's face while talking and commenting on the stimulus, or when the adult searched for a toy only visible to the infant. These gestural behaviors marked the emergence of social cognition at 12 months. Mandarin-speaking children exhibited this cognitive ability in naturalistic mother-child interactions, where declarative pointing gestures were also observed when children were one year old. From 15 months onward, children increasingly and actively engaged their addressees in establishing a joint attention frame and sharing interests or information, using declarative pointing in combination with speech to convey more complex meanings. From the linguistic perspective, children may have lacked the lexical items required to express cross-modal information or the necessary linguistic knowledge to use them effectively, resulting in limited representations of cross-modal relations in both type and frequency.

Together, cross-modal semantic patterns demonstrate children's substantial reliance on supplementary gestures to enrich their speech, enabling them to convey a wider range of messages concerning entities, predications, and propositions. Verbal equivalents for these gesture-speech patterns typically emerge at least three months later, with some still absent at 21 months, attesting that cross-modal communication effectively supports early language development. The gesture-speech semantic patterns and their later corresponding linguistic forms reveal children's emerging grammatical competence.

References

Adamson, L. B., Bakeman, R., Deckner, D. F., & Romski, M. (2009). Joint engagement and the emergence of language in children with autism and Down syndrome.

Journal of Autism and Developmental Disorders, 39(1), 84–96.

Bang, J. Y., Bohn, M., Ramírez, J., Marchman, V. A., & Fernald, A. (2023). Spanish-speaking caregivers' use of referential labels with toddlers is a better predictor of later vocabulary than their use of referential gestures. *Developmental Science, 26*(4), e13354.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms.* Cambridge University Press.

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 96–151). Blackwell Publishing.

Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In D. McNeill (Ed.), *Language and gesture* (pp. 235–258). Cambridge University Press.

Camaioni, L., Perucchini, P., Bellagamba, F., & Colonnesi, C. (2004). The role of declarative pointing in developing a theory of mind. *Infancy, 5*(3), 291–308.

Capirci, O., Contaldo, A., Caselli, M. C., & Volterra, V. (2005). From action to language through gesture: A longitudinal perspective. *Gesture, 5*(1–2), 155–177.

Capone, N. C., & McGregor, K. K. (2004). Gesture development: A review for clinical and research practices. *Journal of Speech, Language, and Hearing Research, 47*(1), 173–186.

Cartmill, E. A., Hunsicker, D., & Goldin-Meadow, S. (2014). Pointing and naming are not redundant: Children use gesture to modify nouns before they modify nouns in speech. *Developmental Psychology, 50*(6), 1660–1666.

Caselli, M. C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language, 26*(3), 69–91.

Chi, P.-H. (2002). The development of language and communication abilities of Mandarin Chinese-speaking infants and toddlers: Parental report. *Journal of Special Education, 16*, 23–64.

Chiou, Y.-T., & Liu, H.-M. (2018, June). Early gesture development and its correlation with language comprehension and production in Mandarin-speaking infants [Conference presentation]. XXI International Conference on Infant Studies, Philadelphia, United States.

Colonnesi, C., Stams, G. J. J. M., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review, 30*(4), 352–366.

Fasolo, M., & D'Odorico, L. (2012). Gesture-plus-word combinations, transitional forms, and language development. *Gesture, 12*(1), 1–15.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), 1–185.

Goldin-Meadow, S. (2015). Gesture as a window onto communicative abilities: Implication for diagnosis and intervention. *Perspectives on Language Learning and Education, 22*(2), 50–60.

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development, 39*, 42–52.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language, 37*(2), 229–261.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*(5), 367–371.

LeBarton, E. S., Goldin-Meadow, S., & Raudenbush, S. (2015). Experimentally induced increases in early gesture lead to increases in spoken vocabulary. *Journal of Cognition and Development, 16*(2), 199–220.

Liszkowski, U., & Tomasello, M. (2011). Individual differences in social, cognitive, and morphological aspects of infant pointing. *Cognitive Development, 26*(1), 16–29.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Reference and attitude in infant pointing. *Journal of Child Language, 34*(1), 1–20.

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science, 7*(3), 297–307.

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). Twelve- and eighteen-month-olds point to provide information for others. *Journal of Cognition and Development, 7*(2), 173–187.

Liu, H.-M., & Chen, Y.-C. (2015). Developmental changes in the content and composition of early expressive vocabulary in Mandarin-speaking infants and toddlers. *Bulletin of Educational Psychology, 47*(2), 217–242.

Liu, H.-M., & Tsao, F.-M. (2010). The standardization and application of Mandarin-Chinese communicative developmental inventory for infants and toddlers. *Formosa Journal of Mental Health, 23*, 503–534.

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

McNeill, D. (2015). *Why we gesture: The surprising role of hand movements in communication.* Cambridge University Press.

Özçalışkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition, 96*(3), B101–B113.

Özçalışkan, Ş., & Goldin-Meadow, S. (2009). When gesture-speech combinations do and do not index linguistic change. *Language and Cognitive Processes*, *24*(2), 190–217.

Özçalışkan, Ş., Adamson, L. B., & Dimitrova, N. (2016). Early deictic but not other gestures predict later vocabulary in both typical development and autism. *Autism, 20*(6), 754–763.

Özçalışkan, Ş., Gentner, D., & Goldin-Meadow, S. (2014). Do iconic gestures pave the way for children's early verbs? *Applied Psycholinguistics, 35*(6), 1143–1162.

Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science, 12*(1), 182–187.

Rowe, M. L., Özçalışkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting

vocabulary development. *First Language, 28*(2), 182–199.
Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development, 78*(3), 705–722.