



Children do not overuse “the” in natural production

Yuanfan Ying, Valentine Hacquard, Alexander Williams, Jeffrey Lidz

University of Maryland, College Park



MAIN FINDINGS

Children have long been assumed to overuse definites (“the NP”) based on elicited production studies ^[1-7].

We show: **“the”-overuse does not hold in natural production.**

- Children’s pattern of use of definites follows their mothers’.
- Their use of “the” (vs. “a”) is consistent with adult expectations.

BACKGROUND

Elicited production studies report “the”-overuse in 3- to 5-year-olds: they sometimes use it when the referent was either not in the common ground ^[2-4] or not unique ^[1,6,10], where indefinite “a” would be better.

Table 1. Percentage of *the*-misuse by children and adults in the literature

Study	Children	Adult controls
Maratsos (1976)	Exp: 17% (3yrs)	
Schafer & de Villiers (2000)	Cond8: 49% (3yrs)	Cond8: 30%
Matthewson et al. (2001)	72% (108/150)	16% (24/148)
Schaeffer & Matthewson (2005)	25% (19/76)	2% (2/113)
van Hout et al. (2010)	Exp1: 50% Exp2: 26%	Exp1: 17% Exp2: 24%
de Cat (2013)	17% (age 2;6-3;3)	

However, adult controls are prone to similar errors when tested. Moreover, comprehension studies tend to show children’s adult-like responses to “the” ^[8,9].

HYPOTHESES & PREDICTIONS

Does overuse of “the” in elicitation tasks reflect a non-adultlike knowledge of “the”, which comprehension tasks overestimate?

Predictions: If yes, this overuse should be present in natural production as well.

Specifically, we would expect that:

- Children use “the” more often than their mothers
- Naïve adults should have difficulties guessing child uses of “the” (vs. “a”), relative to mother/adult uses.

STUDY 1: CORPUS

Corpus: CHILDES (912,530 words; age: 1;0-3;11) ^[11-14]

★ **Children do not use more definites than their mothers.**
If anything, they use them less.

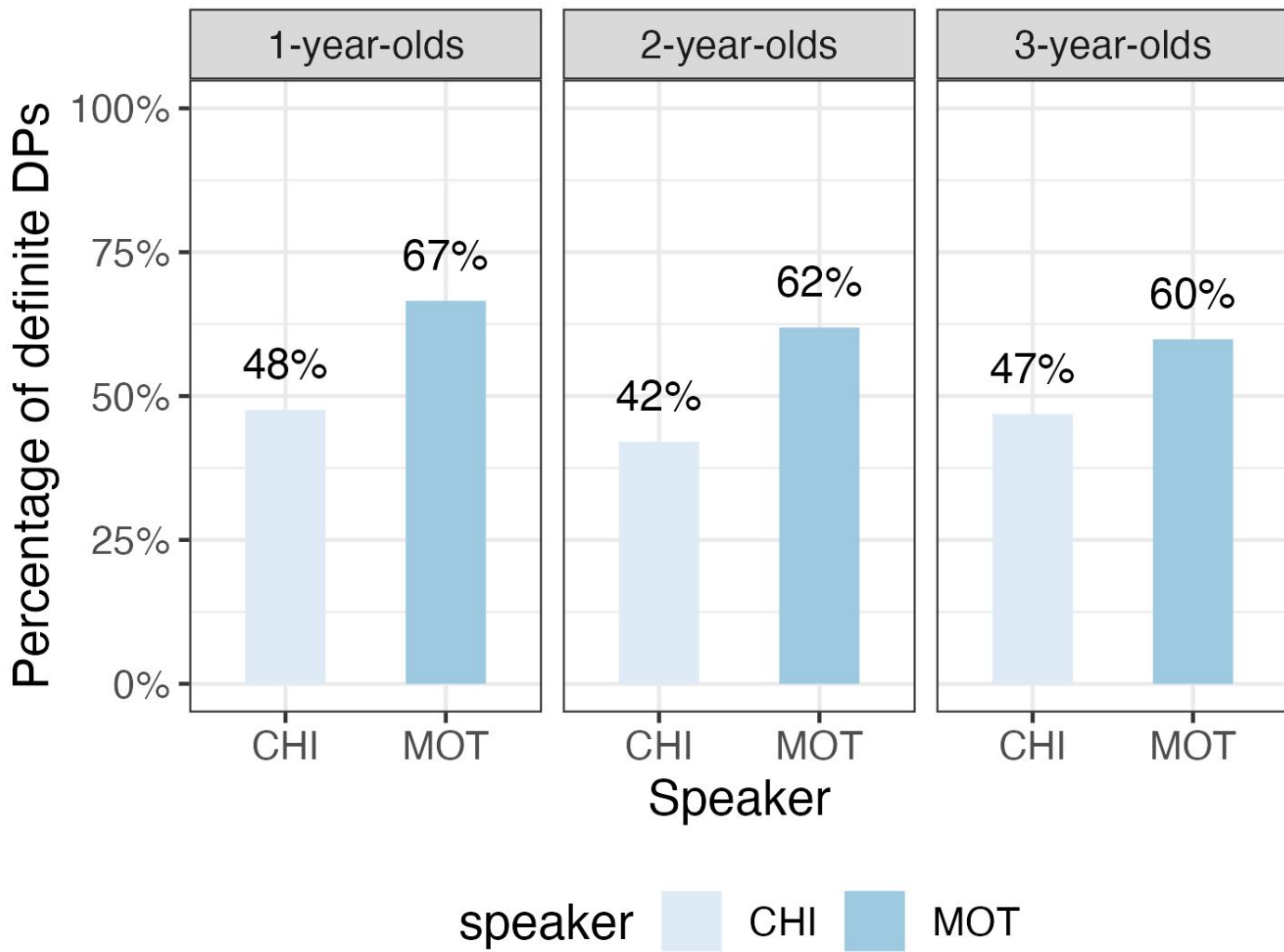


Figure 1. Percentage of definites by children and mothers
Note: Percentage of definites = $\#[\text{the } N_{SG}] / \#[\text{'a'+the } N_{SG}]$

Children match their mothers even in pragmatically distinct environments: (a) declaratives vs. interrogatives, and (b) subjects vs. objects vs. fragments.
⇒ Like their mothers, they tend to use more definites in subjects than in objects. ^[15]

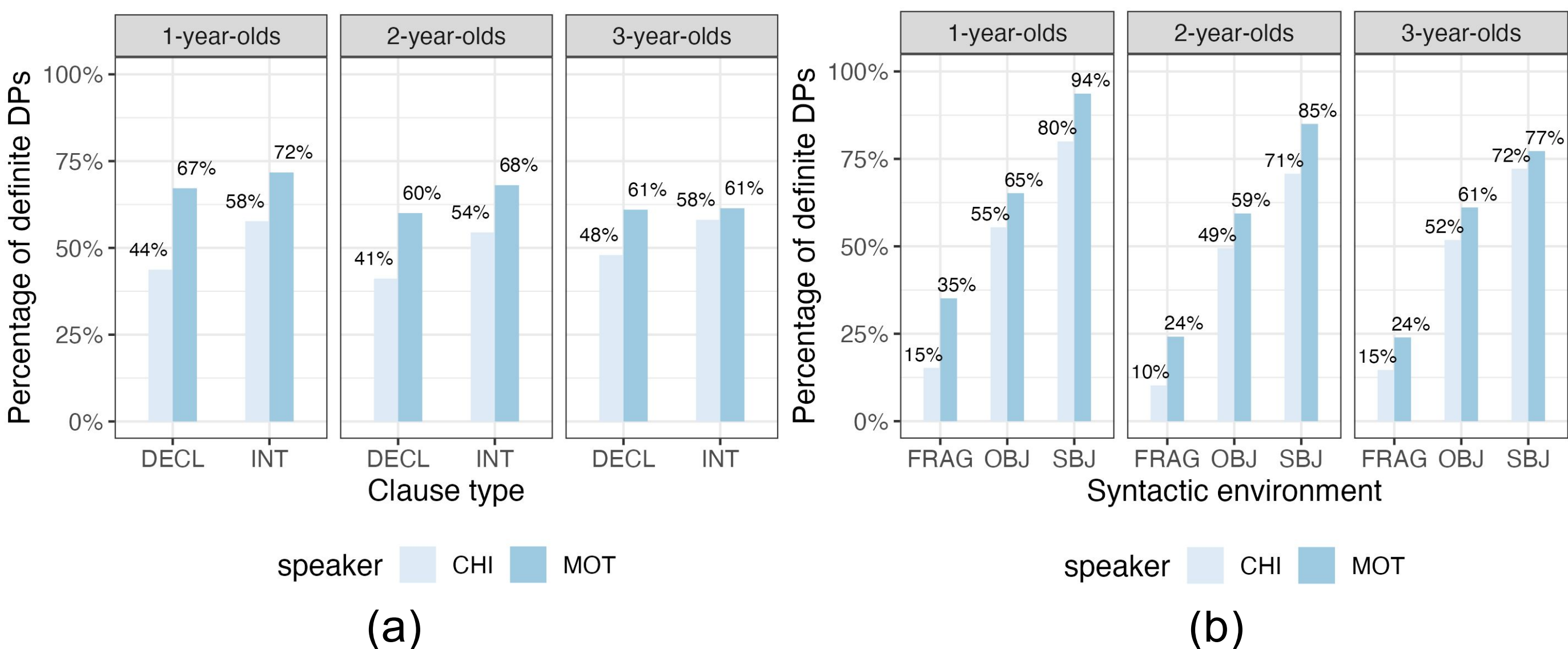


Figure 2. Percentage of definites across (a) clause types and (b) syntactic environments

★ **Rate of reference-related miscommunication is low**
Mothers are rarely confused by children’s use of ‘the N_{SG} ’ due to unclear or non-unique reference.

Table 2. Rate of reference-related miscommunication by children and adults

Corpus	#miscom	#[the N]	#word-by-speaker	rate_mis (%)
1-year-olds	1	205	23,850	0.49%
2-year-olds	10	2575	158,880	0.39%
Adult ^[16]	4	3121	301,805	0.13%

STUDY 2: EXPERIMENT

Corpus: CHILDES (80,347 words; age: 2;1-4;11) ^[17]

Participants (N=240) were presented with 10-line dialogues and asked to choose determiners (used by a child or a mother).

Materials: Excerpts of mother-child conversations from 23 mother-child pairs. Each participant received 40 items: 10 definite and 10 indefinite targets (‘the’ vs. ‘a’) plus 20 fillers (‘this’ vs. ‘that’ or ‘this/that’ vs. ‘the/a’).

★ **Children’s use of “the” is predictable**
Adults were above chance at predicting “the” vs. “a” used by children and mothers.

★ **Children’s “the” is just as predictable as their mothers’**
Children’s “the” is predictable given linguistic contexts only.

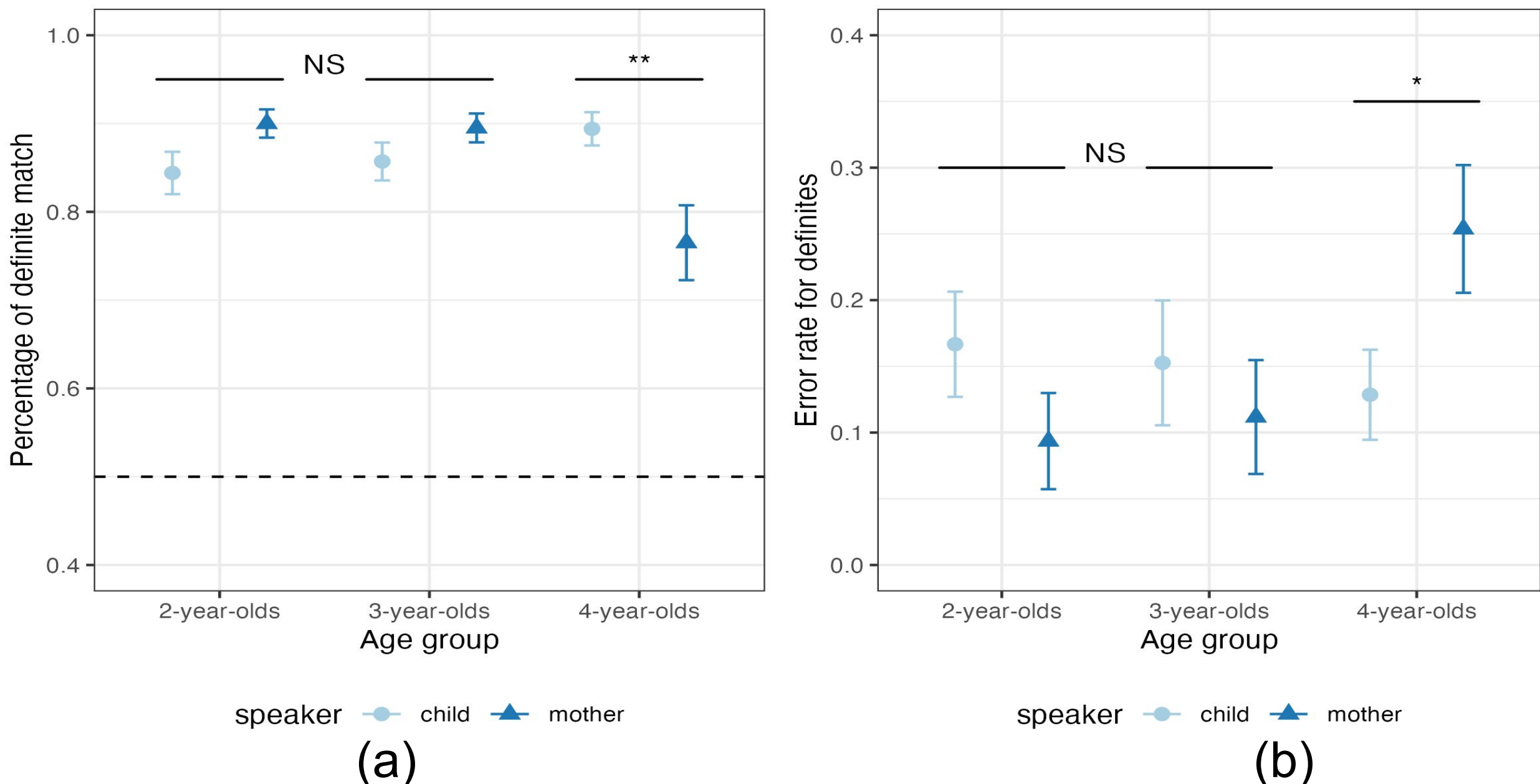


Figure 3. (a) Percentage of determiner-matched responses for definite items, and (b) Average error rate for definite items

TAKE-HOMES

Children do not overuse “the” in natural production.

Elicited production errors may be due to *experimental artifacts* and explained by a *performance account*.

REFERENCES

[1] Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference*. Cambridge University Press. [2] Maratsos, M. (1976). *The use of definite and indefinite reference in young children*. Cambridge University Press. [3] Schaeffer, J., & Matthewson, L. (2005). Grammar and pragmatics in the acquisition of article systems. *Natural Language and Linguistic Theory*, 23(1), 53-101. [4] Schafer, R., & de Villiers, J. (2000). Imagining articles: What a and the can tell us about the emergence of DP. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Boston University Language Development Conference* (pp. 609-620). Cascadia Press. [5] Matthewson, L., Bryant, T., & Roeper, T. (2001). A Salish stage in the acquisition of English determiners: Unfamiliar ‘definites’. *The Proceedings of SULA: The Semantics of Under-Represented Languages in the Americas*. University of Massachusetts Occasional Papers in Linguistics, 25, 63-71. [6] van Hout, A., Harrigan, K., & de Villiers, J. (2010). Asymmetries in the acquisition of definite and indefinite NPs. *Lingua*, 120, 1973-1990. [7] de Cat, C. (2013). Egocentric definiteness errors and perspective evaluation in preschool children. *Journal of Pragmatics*, 56, 58-69. [8] Choi, Y. J., Song, H. J., & Luo, Y. (2018). Infants’ understanding of the definite/indefinite article in a third-party communicative situation. *Cognition*, 175, 69-76. [9] Syrett, K., Kennedy, C., & Lidz, J. (2010). Meaning and context in children’s understanding of gradable adjectives. *Journal of Semantics*, 27(1), 1-35. [10] Wexler, K. (2011). Cues don’t explain learning: Maximal trouble in the determiner system. In E. Gibson, & N. J. Pearlmuter. (eds.), *The processing and acquisition of reference* (pp. 15-42). MIT Press. [11] Brown, R. (1973). *A first language*. Harvard University Press, Cambridge, MA. [12] Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, 35, 869-902. [13] Suppes, P. (1974). The semantics of children’s language. *American Psychologist*, 29, 103-114. [14] Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, 40, 21-81. [15] Prince, E. F. (1992). The ZPG letter: Subjects, definiteness, and information-status. In W. C. Mann and S. A. Thompson (Eds.), *Discourse descriptions: Diverse linguistic analyses of a fund-raising text* (pp. 295-325). Philadelphia: John Benjamins. [16] Kingsbury, P., Strassel, S., McLemore, C., & McIntyre, R. (1997). CALLHOME: American English transcripts. LDC97T14. Philadelphia: Linguistic Data Consortium. [17] Menn, L. & Gleason, J. (1986). Babytalk as a Stereotype and Register: Adult Reports of Children’s Speech Patterns. In J. Fishman (Ed.), *The Fergusonian Impact: In Honor of Charles A. Ferguson on the Occasion of his 65th Birthday. Volume 1: From Phonology to Society. Volume 2: Sociolinguistics and the Sociology of Language* (pp. 111-126). Berlin, Boston: De Gruyter Mouton.