Do children overuse "the" phrases?

Jack Yuanfan Ying, Valentine Hacquard, Alexander Williams, Jeffrey Lidz
University of Maryland, College Park

English-learning young children have been reported to overuse "the" when the referent is either not established in the common ground or not unique. We present a corpus study and a determiner-guessing experiment on children's determiner use in naturalistic production – our results do not show strong evidence for *the*-overuse, potentially raise concerns about the claim.

Findings from our corpus study do not support *the*-overuse either globally or locally. Our global analysis targeted the percentage of definites in different clause types (interrogatives vs. declaratives) and across different syntactic environments (within-VP vs. above-VP vs. fragment) and demonstrated that children's percentage of definites does not exceed the baseline of determiner use by mothers. In addition, their pattern of use follows that of mothers – there are more definites in interrogatives than in declaratives, and for definites within VP (i.e., direct/indirect objects), the determiner used is not biased towards definite (i.e., it occurs for about 50% of the time). Moreover, the local analysis results in only 3.4% (7 out of 205) of suspicious cases where mothers seem to have trouble identifying the referent for 'the N' in 2-year-old children's questions.

Our determiner-guessing experiment does not find evidence for *the*-overuse, either. 240 adult participants (6 between-subject conditions with 40 per condition) were presented with 10-line mother-child conversations (drawn from Gleason's corpora in CHILDES, aged between 2;0 and 5;0) and asked to guess the missing determiner in the last line between two choices, given the context (40 trials: 10 definites, 10 indefinites, 20 fillers). Our results show that 1) adults were above chance at guessing definites used by either mothers or children for all age groups (see Table 1), and 2) they were never significantly better at guessing definites by mothers than by children (see Table 2). Regarding error rates for individual test items, no significant difference was found between those for mothers and those for children with 3- and 4-year-olds, but for 5-year-olds, adults' error rates were higher for definites by mothers than those by children (see Table 3).

Taken together, *the*-overuse seems ill-founded when artifacts of experiments are removed. Conceptual (e.g., Maratsos, 1974; Karmiloff-Smith, 1979; Schafer & de Villiers, 2000), semantic (Wexler, 2011), and pragmatic (Schaeffer & Matthewson, 2005; Munn et al., 2006; van Hout et al., 2010) accounts of children's errors seem incompatible with results from comprehension studies (e.g., Choi et al., 2018; Syrett et al., 2010) demonstrating their ability to identify infelicitous use of definites since their second year of life. We argue based on our findings that an alternative performance account is needed to explain children's errors reported in the literature – children's performance in using determiners is likely affected by experimental settings that involve perspective disruptors, trivialized common ground, and less motivated correction mechanisms for egocentric biases.

(Word count: 454 words)

Table 1. Adults' percentage of definite determiner match

Age	Speaker	N	Mean	SD	Wilcoxon test
3-year-olds	child	36	0.844	0.144	V = 595, p < .001
	mother	36	0.900	0.096	V = 666, p < .001
4-year-olds	child	35	0.857	0.127	V = 595, p < .001
	mother	39	0.895	0.102	V = 780, p < .001
5-year-olds	child	34	0.894	0.110	V = 595, p < .001
	mother	31	0.765	0.236	V = 389, p < .001

Table 2. Mann-Whitney U tests for determiner match with mothers and children

Age	3-year-olds	4-year-olds	5-year-olds
Tests	W = 516, p = .124	W = 571.5, p = .212	W = 726.5, p = .007*

Table 3. Kruskal-Wallis tests for error rates with mothers and children

Age	Speaker	N	Mean	SD	Kruskal-Wallis test
3-year-olds	child	20	16.667	17.763	p = .067
	mother	20	9.356	16.232	
4-year-olds	child	20	15.262	21.082	p = .320
	mother	20	11.173	19.225	
5-year-olds	child	20	12.854	15.206	p = .014
	mother	20	25.375	21.567	

References

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.

Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference*. Cambridge University Press.

Maratsos, M. (1974). Preschool children's use of definite and indefinite articles. *Child Development*, *45*(2), 446-455.

Munn, A., Miller, K., & Schmitt, C. (2006). "Maximality and plurality in children's interpretation of definiteness". In D. Bamman, T. Magnitskaia and C. Zaller (eds.), *BUCLD 30: Proceedings of the 30th Annual Boston University Conference on Language Development, Vol.2* (pp. 377-387). Boston, MA: Cascadilla Press.

Schaeffer, J., & Matthewson, L. (2005). Grammar and pragmatics in the acquisition of article systems. *Natural Language and Linguistic Theory*, *23*(1), 53-101.

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). Somerville, Cascadilla Press.

Syrett, K., Kennedy, C., & Lidz, J. (2010). Meaning and context in children's understanding of gradable adjectives. *Journal of Semantics*, *27*(1), 1-35.

Van Hout, A., Harrigan, K., & de Villiers, J. (2010). Asymmetries in the acquisition of definite and indefinite NPs. *Lingua*, *120*, 1973-1990.

Wexler, K. (2011). Cues don't explain learning: Maximal trouble in the determiner system. In E. Gibson, & N. J. Pearlmutter. (eds.), *The processing and acquisition of reference* (pp. 15-42). MIT Press.