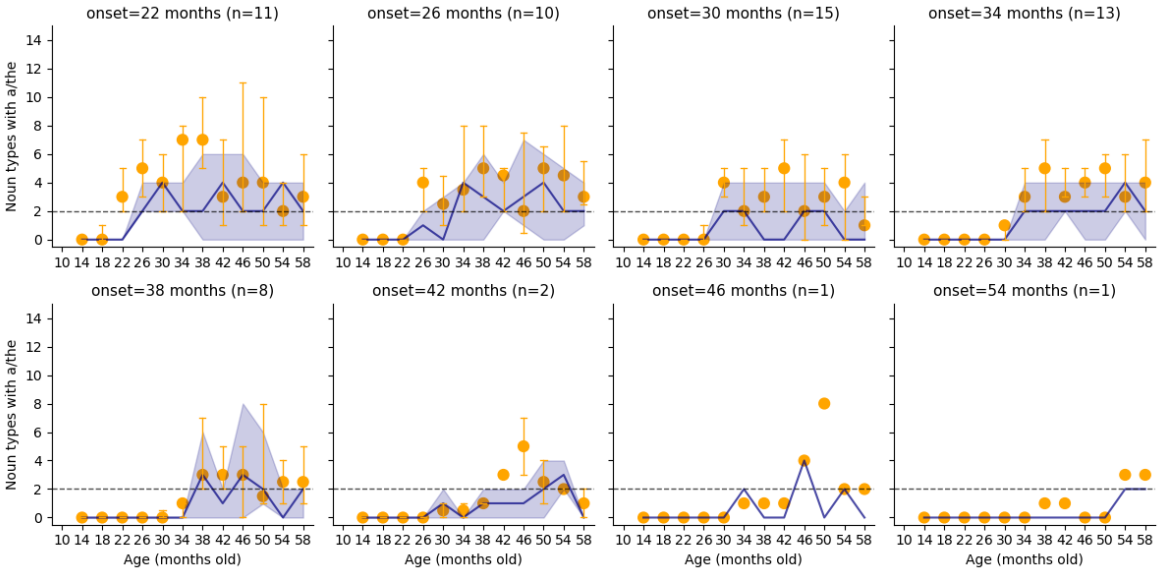
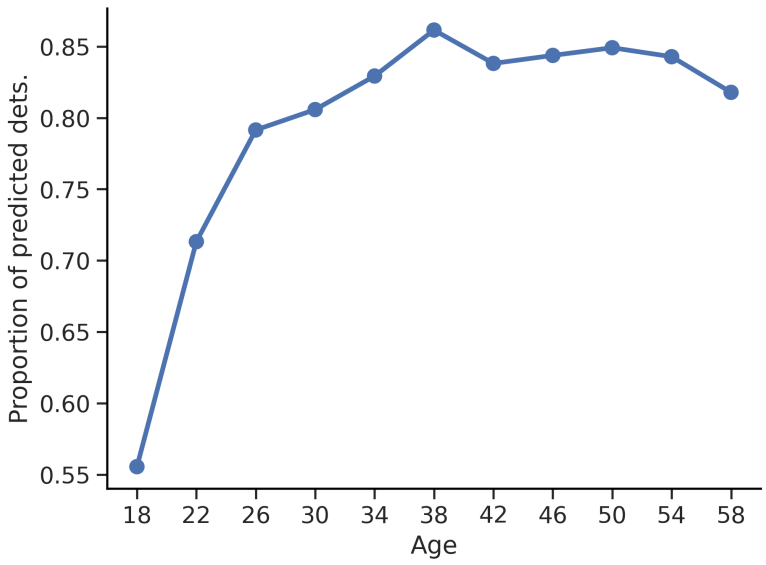
Appendix SI: Applying an n-gram model to our behavioral data

Figure S.1. The figure displays the median number of noun types produced with both determiners (*a, the*), grouped by the age at which the child produced (orange dots) or an n-gram model predicted (blue line) determiner-noun combinations that meet our criterion for productivity. The dashed horizontal line denotes two different nouns, each combined with both *a* and *the*; the figure should be compared with Figure 1 in the text. The n-gram model was a trigram model with backoff. The 'masked' (or predicted) word in each trigram was the middle one, and the context was formed by the first and third word. In the event of backing off to a bigram model (when a trigram from the children's productions was not part of the input data), the predicted word was the first one, and the context was formed by the second word in the bigram. This design accounts for the fact that nouns appear after determiners in English (hence conditioning on the words appearing after the determiner is necessary). Compared to the Transformer model used in our analyses (and to the children), the n-gram model underestimates determiner-noun productivity (measured in the same way as reported for Figure 1).



SI Appendix: Learning to predict a determiner

Figure S.2. The figure displays the accuracy of the model in predicting determiners (*a* or *the*) where children produced them, at every child observation session. We omit data for the first observation session (age 14 months old) because few children produced determiners.



SI Appendix: Descriptive statistics of the LDP child language corpus

Table S.1. The table displays the means and ranges for the total number of utterances, the total number of noun types, the total number of noun tokens combined with *a* or *the,* and the total number of noun types combined with *a* or *the* that the children produced at each of the 12 observation sessions.

Age (mos)	*Utterances*		*Noun types*		*a/the-Noun tokens*		*a/the-Noun types*	
	Range	**M**	**Range**	**M**	**Range**	**M**	**Range**	**M**
14	1-218	52	1-3	1	0-21	1	0-3	0
18	6-631	178	1-36	2	0-290	7	0-36	1
22	3-1198	357	1-63	8	0-188	15	0-73	8
26	4-1174	534	1-55	18	0-172	35	0-67	18
30	68-1266	583	1-75	27	0-267	53	0-92	29
34	223-1365	651	1-61	31	0-189	62	0-82	35
38	247-1276	698	5-91	37	0-176	76	0-107	43
42	126-1559	668	3-74	36	0-161	70	0-86	40
46	79-1574	688	4-84	41	0-228	84	0-113	43
50	203-1045	639	10-96	40	0-224	81	0-114	45
54	126-1254	594	5-85	38	0-218	70	0-103	40
58	126-1161	607	5-77	39	0-176	72	0-93	41