

Generic noun phrases in child speech: Supplemental materials

Samarth Mehrotra
Birla Institute of Technology and Science

Amy Perfors
University of Melbourne

Abstract

This document contains supplemental materials for the 2019 CogSci submission entitled *Generic noun phrases in child speech*. This includes details of our model architecture, description and citations for our 26 corpora, larger figures corresponding to the ones in the CogSci paper, and one additional analysis that didn't fit.

Model Architecture

As described in the paper, our model consisted of seven different classifiers, each constructed with a different architecture: a different arrangement of LSTMs and GRUs, with different numbers of units. We chose these seven through trial and error, aiming for a range of architectures so that each optimised different aspects of the decision problem. Decisions were made by majority vote among the seven. The architectures are shown in Table 1.

Architecture
input(64) → GRU(90) → GRU(60) → GRU(30) → Dense (2)
input(64) → GRU(45) → GRU(45) → GRU(45) → Dense (2)
input(64) → LSTM(30) → GRU(45) → GRU(45) → GRU(45) → GRU(30) → Dense (2)
input(64) → GRU(90) → LSTM(90) → Dense (2)
input(64) → GRU(30) → GRU(40) → GRU(50) → GRU(60) → GRU(70) → Dense (2)
input(64) → LSTM(196) → GRU(196) → GRU(196) → GRU(196) → LSTM(196) → Dense (2)
input(64) → GRU(396) → GRU(192) → GRU(98) → Dense (2)

Table 1

Architecture structure for the seven classifiers in our model. All classifiers began with the same input structure and ended with the same output, which followed the Dense(2) units at the end. Numbers in parentheses indicate number of hidden units.

CHILDES Corpora

#	Name	Reference	Children	Ages
1	Belfast	Henry (1995)	4 boys, 4 girls	2;4-4;5
2*	Bloom	Bloom, Hood, and Lightbown (1974)	Peter	1;9-3;2
3	Braunwald	Braunwald (1976)	Laura	1;0-6;0
4*	Brown	Brown (1973)	Adam, Eve, Sarah	1;6-5;1
5	Clark	Clark (1978)	Shem	2;2-3;2
6	Cornell	NA	8 files	1;6-4;0
7	Demetras 1	Demetras (1989b)	Trevor	2;0-3;11
8	Demetras 2	Demetras (1989a)	Michael, Jimmy, Tim	2;0-2;6
9	Feldman	Feldman and Menn (2003)	Steven	0;5-2;9
10	Forrester	Forrester (2002)	Ella	1;0-3;2
11	Hall	Hall, Nagy, and Linn (1984)	39 children	4;6-5;0
12	Higginson	Higginson (1985)	April, May, June	0;11-2;11
13*	Kuczaj	Kuczaj (1977)	Abe	2;4-5;0
14	MacWhinney	MacWhinney (1991)	Ross, Mark	0;7-8;0
15*	Morisset	Morisset, Barnard, and Booth (1995)	206 children	2;6
16	MPI-EVA-Manchester	Lieven, Salomo, and Tomasello (2009)	4 children	1;8-3;2
17	Peters	Peters (1987)	Seth	1;7-4;1
18	Providence	Demuth, Culbertson, and Alter (2006)	6 children	0;11-3;11
19*	Sachs	Sachs (1983)	Naomi	1;1-5;1
20	Sawyer	Sawyer (1997)	20 children	3;6-4;11
21*	Snow	MacWhinney and Snow (1990)	Nathaniel	2;3-3;9
22	Soderstrom	Soderstrom, Blossom, Foygel, and Morgan (2008)	Charles, Joseph	0;6-1;0
23	Suppes	Suppes (1974)	Nina	1;1-3;9
24	Thomas	Lieven et al. (2009)	Thomas	2;0-4;11
25	Warren	Warren-Leubecker and Bohannon (1984)	20 children	1;6-6;2
26	Wells	Wells (1981)	32 children	1;6-5;0

Table 2

Details of 26 CHILDES corpora used in our analysis. The number on the left corresponds to the number we used to denote it in the text of the paper. Starred items indicate corpora that were also used in Gelman, Sarnecka, and Flukes (2008).

Age	Child speech	Adult speech
Under 2	117551	220156
2	567749	787690
3	264232	352240
4	212423	224508
5	11279	7877
6	1127	1258
7+	997	1576

Table 3

Number of utterances in our corpora at each age in child and adult speech. We excluded all child speech from children less than two from our analyses.

WordNet Categories

A full description of the noun categories we took from WordNet is available here:
<https://wordnet.princeton.edu/documentation/lexnames5wn>.

Name	Description
animal	animals
artifact	man-made objects
attribute	attributes of people and objects
body	body parts
cognition	cognitive processes and contents
communication	communicative processes and contents
event	natural events
feeling	feelings and emotions
food	foods and drinks
group	groupings of people or objects
location	spatial position
motive	goals
object	natural objects (not man-made)
person	people
phenomenon	natural phenomena
plant	plants
possession	possession and transfer of possession
process	natural processes
quantity	quantities and units of measure
relation	relations between people or things or ideas
shape	two and three dimensional shapes
state	stable states of affairs
substance	substances
time	time and temporal relations

Table 4

WordNet category labels we used for our nouns. We combined animal and person to make animate. We kept the existing categories of food and artifact. Everything else was combined to form the other category.

Below, we reproduce the main figures from the paper much larger, so you can see them better.

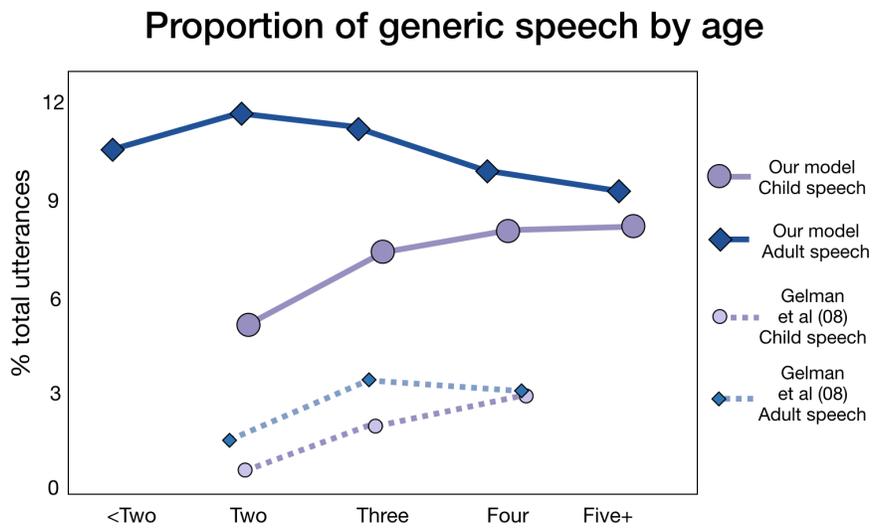
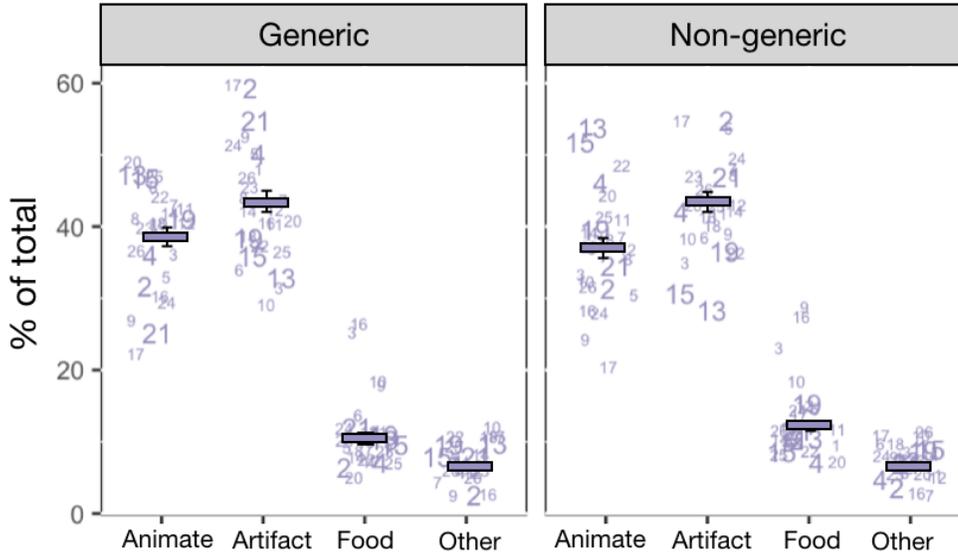


Figure 1. Proportion of generic speech by age. The overall percentage of all utterances coded as generic in our corpora (solid line), broken down by child and adult speech (purple and blue, respectively). For comparison, we plot analogous results from Gelman et al. (2008) with the dotted line. Although we estimated more total generics than they did, the qualitative patterns over development and between child and adult speech are extremely similar.

Division of generics and non-generics

Child speech



Adult speech

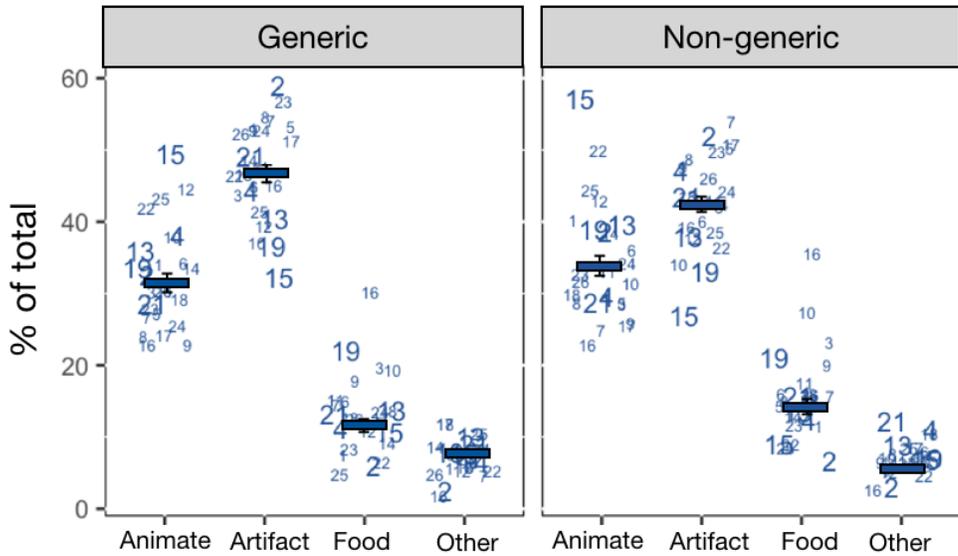


Figure 2. Proportion of generic and non-generic speech across categories. This figure shows the distribution of speech across the four noun categories, for both children (purple) and adults (blue). Lines show the mean when averaged by corpus; error bars indicate standard error. The numbers correspond to the relevant measure for each of the 26 corpora. The corpora from Gelman et al. (2008) are slightly larger and correspond to numbers 2, 4, 13, 15, 19, and 21. It is evident that there is high variability between corpora, but for the most part both children and adults speak about artifacts more often and that there is little difference between generics and non-generics in how they are distributed amongst the four noun categories.

Genericity by noun type

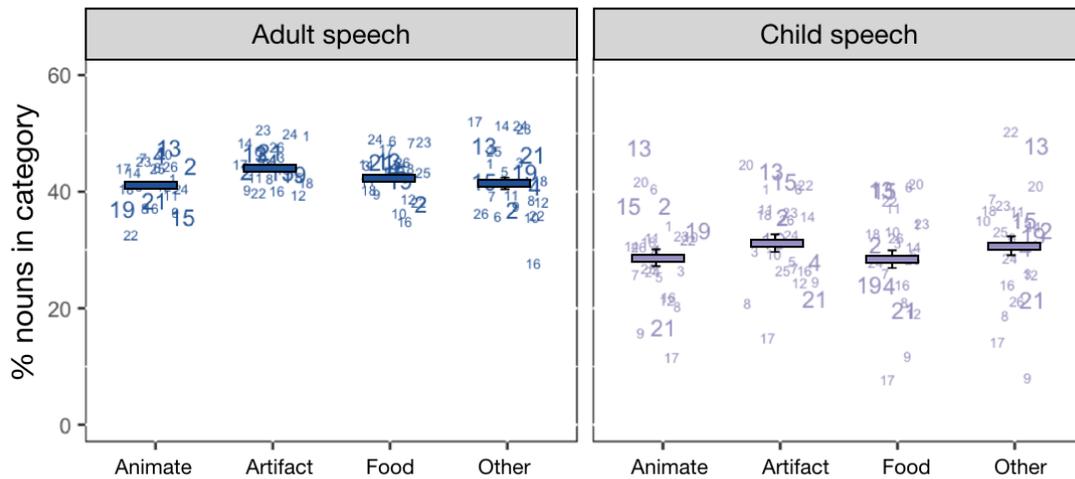


Figure 3. Proportion of generic speech within each noun category. For each of the four categories, this figure shows how often nouns in that category were generic. The large transparent bars indicate the aggregate proportion over all corpora, while the small boxes with error bars show the mean when averaged by corpus. The numbers correspond to the relevant measure for each of the 26 corpora. The corpora from Gelman et al. (2008) are slightly larger and correspond to numbers 2, 4, 13, 15, 19, and 21. There is high variability between corpora (especially for children). However, there is little difference in the pattern of generic usage across noun categories.

Genericity by noun type: Gelman et al (2008) coding

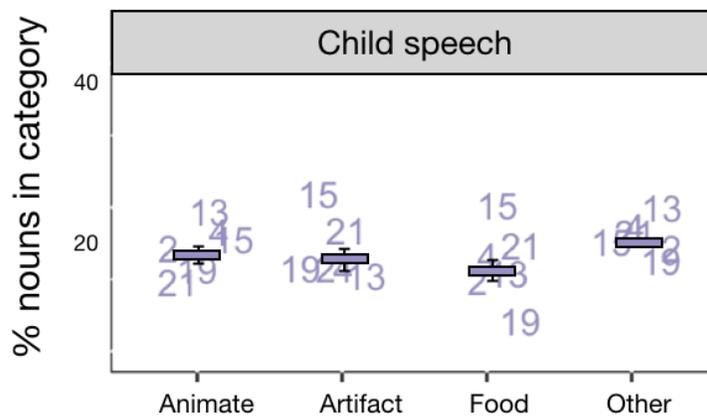
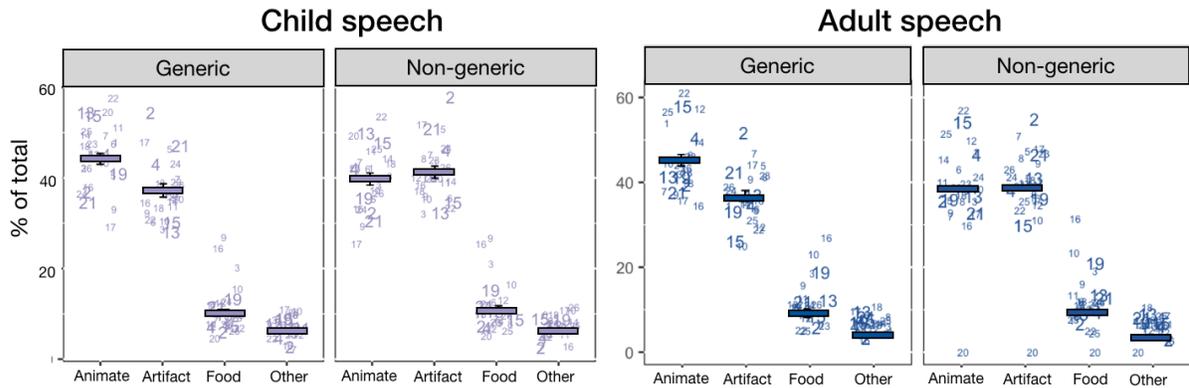


Figure 4. Proportion of generic speech within each noun category using the genericity identifications from Gelman et al. (2008). For each of the four categories, this figure shows how often nouns in that category were generic, using the six corpora and their classifications rather than the classifications from our model. Despite using their classifications, we replicate our previous result, suggesting that the difference between our findings and theirs did not arise due to poor classification performance by our model.

No singular pronouns: Division of generics and non-generics



No singular pronouns: Genericity by noun type

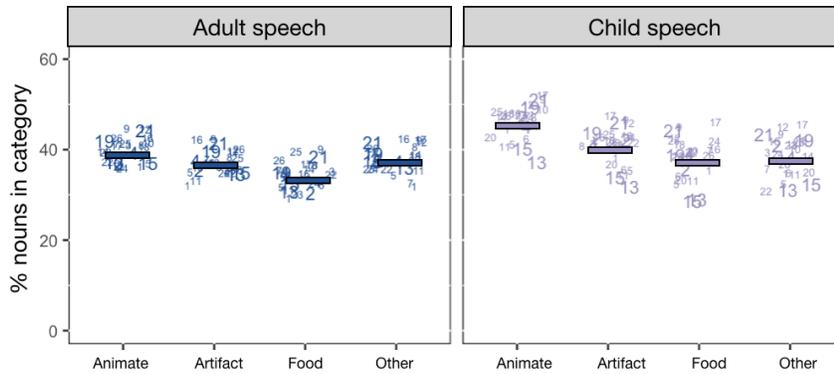


Figure 5. Proportion of generic speech across and within noun categories, on corpora without any singular pronouns. Since Gelman et al. (2008) excluded singular pronouns, we reran our analyses (using our classifications) on our corpora after excluding all singular pronouns. Results are now much more similar to their findings than ours. Generics but not non-generics are used more for animate than artifact categories (top); and for both adults and children, the proportion of generic utterances in animates is higher than in artifacts. This suggests that their exclusion of singular pronouns from the dataset may have driven their results.

Generic preference for animates

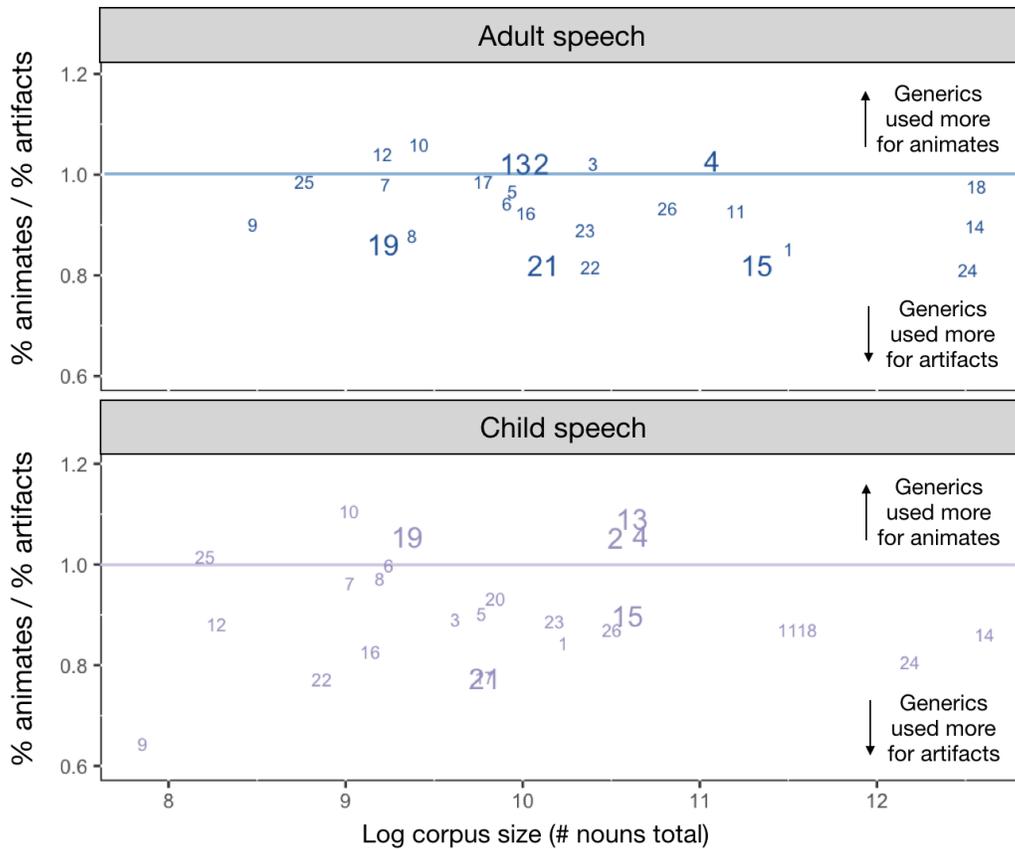


Figure 6. Animacy preference within each corpus as a function of corpus size. We investigate variability between corpora by plotting the degree to which each corpus reflects a preferential use of generics for animates over artifacts. The y axis reflects the percentage of animates that were coded as generic divided by the percentage of artifacts that were coded as generic in that corpus for each type of speaker. This is plotted against corpus size (log of nouns) in order to determine whether there are systematic differences based on quantity of data. For most corpora, generics were used slightly more often for artifacts (i.e., they are plotted below the line). However, the corpora from Gelman et al. (2008) are disproportionately found above the line, among those that use generics more in animates. Of those six corpora (2, 4, 13, 15, 19, 21), three are above the line for adult speech (2, 4, 13) and four are for child speech (2, 4, 13, 19).

References

- Bloom, L., Hood, L., & Lightbown, P. (1974). Imitation in language development: If, when and why. *Cognitive Psychology*, 6, 380–420.
- Braunwald, S. (1976). Mother-child communication: The function of maternal language input. In R. von Raffler-Engel (Ed.), *Child language* (pp. 28–50). International Linguistics Association.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Clark, E. (1978). Awareness of language: Some evidence from what children say and do. In R. Sinclair & W. Levelt (Eds.), *The child's conception of language*. Berlin: Springer Verlag.
- Demetras, M. (1989a). Changes in parents' conversational responses: A function of grammatical development. *ASHA*.
- Demetras, M. (1989b). Working parents conversational responses to their two-year-old sons. *Working paper*.
- Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis, and coda licensing in the acquisition of English. *Language Speech*, 49, 137–174.
- Feldman, A., & Menn, L. (2003). Up close and personal: A case study of the development of three English fillers. *Journal of Child Language*, 30, 735–768.
- Forrester, M. (2002). Appropriating cultural conceptions of childhood: Participation in conversation. *Childhood*, 9, 255–278.
- Gelman, S., Sarnecka, P. G., & Flukes, J. (2008). Generic language in parent-child conversations. *LL&D*, 4(1), 1–31.
- Hall, W., Nagy, W., & Linn, R. (1984). *Spoken words: Effects of situation and social group on oral word usage and frequency*. Hillsdale, NJ: Erlbaum.
- Henry, A. (1995). *Belfast English and standard English: Dialect variation and parameter setting*. New York: Oxford University Press.
- Higginson, R. (1985). Fixing-assimilation in language acquisition. *Unpublished doctoral dissertation*.
- Kuczaj, S. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, 16, 589–600.
- Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, 20, 481–508.
- MacWhinney, B. (1991). *The childes project: Tools for analyzing talk*. Hillsdale, NJ: Erlbaum.
- MacWhinney, B., & Snow, C. (1990). The child language data exchange system: An update. *Journal of Child Language*, 17, 457–472.
- Morisset, C., Barnard, K., & Booth, C. (1995). Toddlers' language development: Sex differences within social risk. *Developmental Psychology*, 31(5), 851–865.
- Peters, A. (1987). The role of imitation in the developing syntax of a blind child. *Text*, 7, 289–311.
- Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent-child discourse. In K. Nelson (Ed.), *Children's language*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sawyer, K. (1997). *Pretend play as improvisation*. Mahwah, NJ: Erlbaum.
- Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, 35, 869–902.
- Suppes, P. (1974). The semantics of children's language. *American Psychologist*, 29, 103–114.
- Warren-Leubecker, A., & Bohannon, J. (1984). Intonation patterns in child-directed speech: Mother-father speech. *Child Development*, 55, 1379–1385.
- Wells, C. (1981). *Learning through interaction: The study of language development*. Cambridge, UK: Cambridge University Press.