Learning Pronoun Case from Distributional Cues

Flexible Frames for Case Acquisition

Xiaomeng Ma, Martin Chodorow, Virginia Valian

January 12, 2021

Graudate Center, CUNY

Virginia Valian

Martin Chodorow

contact me: xma3@gradcenter.cuny.edu

Introduction

Children's Pronoun Case Error

English Pronoun Case

English cases are expressed on pronouns, as nominative case (e.g. *I*, *he*), accusative case(e.g. *me*, *him*) and genitive case (e.g. *my*, *his*)

English speaking children make pronoun case errors

Pronoun case errors usually occur in children at the age of 2 to 4.

Example errors:

(1) ***Me*** bite. (Abe, 1;9 in Kuczaj (1978))

(2) All of ***they*** going go in here. (Nina, 2;11 in Suppes (1974))

(3) What ***my*** doing? (Eve, 2;1 in Brown (1973))

Why Children Make Pronoun Case Errors

Syntactic Explanation

- ◇ Non-finite verbs lead to non-nominative errors (e.g. Schütze and Wexler, 1996).

Usage-based Paradigm building model

- ◇ Case, person, gender and number forms a paradigm for each pronoun. The more varieties the children attempt to produce, the more errors they make (e.g Rispoli, 1994, 2005)

Input-based theory

- ◇ Children would say things like ‘*her* go home’ because they mistakenly repeated ‘Let her go home’ in parents’ input (e.g. Tomasello, 2000; Kirjavainen et al., 2009).

Children Rarely Make Pronoun Case Errors

Corpus Analysis

Corpora: 46 longitudinal children's data, 211 children with cross-sectional data in all the available English-speaking children's data in CHILDES (MacWhinney, 2014).

Cross-sectional data: 141 children didn't make any errors; average pronoun case error rate is 1.16%; 95% children's pronoun case error rate is lower than 5%.

Longitudinal data: average pronoun case error rate 1.56%, with median of 0.6%.

Error rate is not correlated with age or MLU.

Children Rarely Make Pronoun Case Errors

Table 1: Summary of Pronoun Case Error Data

Pronoun	Tokens	Error Type	Errors	Pronoun Correct Rate by Use[a]	N children made error	Maximum error/child
I	118607	I-for-me	9	99.99%	6	3
he	16966	he-for-him	27	99.84%	14	8
she	4955	she-for-her	4	99.92%	4	1
we	13525	we-for-us	4	99.97%	3	2
they	9703	they-for-them	4	99.96%	4	1
me	21280	me-for-I	1579	91.80%	41	858
		me-for-my	165		21	81
him	4732	him-for-he	148	95.79%	26	26
		him-for-his	51		11	30
her	4650	her-for-she	412	91.14%	30	169
us	727	us-for-we	13	98.21%	9	3
them	7181	them-for-they	194	95.95%	36	42
		them-for-their	97		23	17
my	35329	my-for-I	485	98.54%	25	124
		my-for-me	31		7	8
his	5109	his-for-he	9	99.82%	9	1
our	1265	our-for-we	1	99.92%	1	1
their	845	their-for-they	8	99.05%	6	2

Learning Pronoun Case

Questions

How did children learn pronoun case?

Is parents' input informative enough for them to distinguish different pronoun cases?

Are they able to learn in the face of ambiguity?

Hypothesis

Pronoun case can be distinguished by different distributional patterns. For example, 'help X cook', X is an accusative pronoun; 'can Y cook', Y is a nominative pronoun.

Related Work on Distributional Patterns

Distributional cues are effective in grammatical categorization

- Frequent trigram frames (`aXb` where `X` is the target word) are effective in grammatical categorization. For example, 'to `X` to', '`X`' is like to be a verb.
- Two bigram frames (`aX + Xb`) with a feedforward neural network can categorize more words with better categorization accuracy (Mintz, 2003).

Trigram (`aXb`) and flexible bigrams (`aX + Xb`) could be used to categorize pronoun cases

Methods

Corpora

Corpora

Following Mintz (2003) and Clair et al. (2010), we used the same six corpora of child-directed speech from CHILDES:
Anne and Aran (Theakston et al., 2001), Eve (Brown, 1973), Naomi (Sachs, 1983), Nina (Suppes, 1974), Peter (Bloom et al., 1974).

Included files where the child is younger than 2;6 years old

Data

Data

Each pronoun was extracted in aXb context, where X is the pronoun, e.g. ‘help me cook’.

Table 2: Token counts of three pronoun cases and type counts of three context frames

	Nominative	Accusative	Genitive	Pronoun Tokens	aX types	Xb types	aXb types
Aran	4518	1014	1454	6986	445	927	2489
Anne	4343	1080	1392	6815	428	707	2308
Eve	1292	479	1029	2800	278	500	1364
Naomi	599	249	503	1352	224	364	806
Nina	3490	1195	1571	6256	400	747	2376
Peter	339	135	207	681	187	250	475
Total	14581	4152	6156	24889	898	1672	7355

Model

Model Architecture

feedforward connectionist models with `aXb` and `aX + Xb` as different inputs

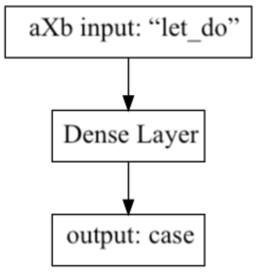


Figure 1: The architecture of `aXb` model

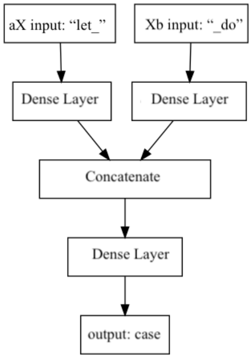


Figure 2: The architecture of `aX + Xb` model

Model

Evaluation

Classification accuracy

Asymmetric lambda value (following Clair et al. (2010)), which evaluates the association among the classes. Lambda is in the range of [0, 1], 0 as no association, 1 as perfect association.

Training and Testing

Each model was trained on all tokens (24889 tokens) and all types (7355 tokens of unique types).

Each model was trained using the same 10-fold cross-validation split.

Experiments

Experiment 1: Models aXb vs aX + Xb in Categorizing Grammatical Cases

Method

Input: aXb or aX , Xb, e.g. 'let X go' and 'let X, X go'

Output: The case of X: nominative, genitive or accusative

Results

	Training on 24889 total tokens				Training on 7355 tokens of unique types			
	aX + Xb		aXb		aX + Xb		aXb	
	Accuracy	λ	Accuracy	λ	Accuracy	λ	Accuracy	λ
Aran	0.984	0.956	0.962	0.894	0.968	0.94	0.849	0.631
Anne	0.984	0.957	0.962	0.897	0.963	0.936	0.841	0.639
Eve	0.979	0.961	0.96	0.928	0.968	0.931	0.872	0.648
Naomi	0.983	0.969	0.951	0.914	0.953	0.902	0.878	0.708
Nina	0.987	0.97	0.951	0.911	0.974	0.952	0.834	0.6
Peter	0.982	0.965	0.954	0.913	0.963	0.927	0.827	0.619
Total	**0.984**	**0.962**	**0.960**	**0.907**	**0.967**	**0.939**	**0.847**	**0.631**

Experiment 1: Heatmap results

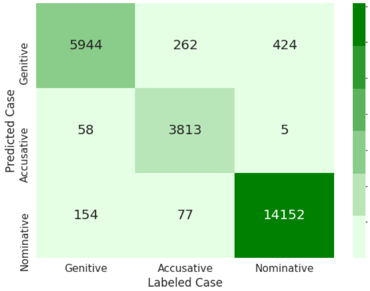


Figure 3: `aXb` model on 24889 tokens

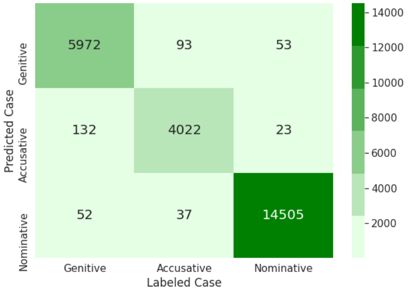


Figure 4: `aX + Xb` model on 24889 tokens

Experiment:1 Results on aXb model

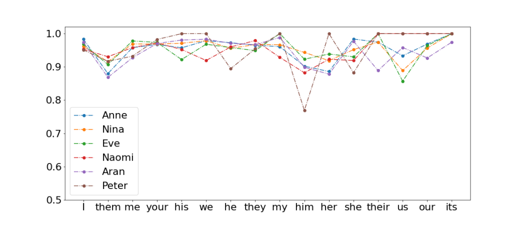


Figure 5: Training results of aXb model with 24889 tokens

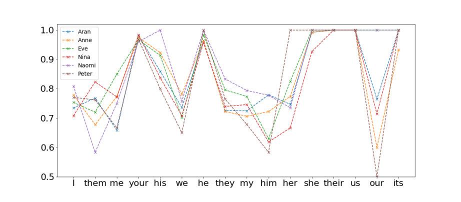


Figure 6: Training results of aXb model with 7355 tokens of unique types

Experiment 1: Results on aX + Xb model

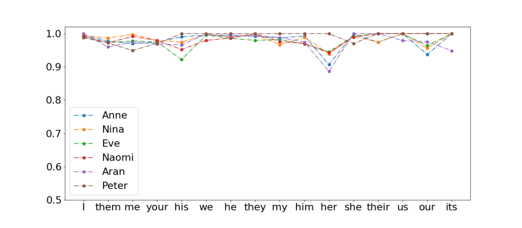


Figure 7: Training results of `aX + Xb` model with 24889 tokens

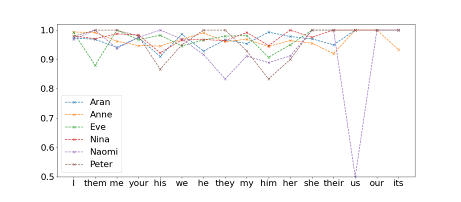


Figure 8: Training results of `aX + Xb` model with 7355 tokens of unique types

Experiment 2: Predicting the Pronoun Using `aX + Xb` Model with Person, Gender, Number Information

Method

Person, gender, number were used together to train the `aX + Xb` model to predict the pronoun.

For example, ‘help `X` cook’ plus *3rd person*, *masculine*, *singular* would be able to predict `X` as ‘*him*’.

Results

	24889 tokens		**7355 types**	
	Accuracy	λ	**Accuracy**	λ
Aran	0.994	0.992	0.980	0.971
Anne	0.994	0.992	0.980	0.976
Eve	0.993	0.990	0.983	0.972
Naomi	0.993	0.995	0.980	0.967
Nina	0.996	0.994	0.987	0.982
Peter	1.000	1.000	0.983	0.975
Total	**0.994**	**0.993**	**0.982**	**0.975**

Experiment 2. Heatmap Result

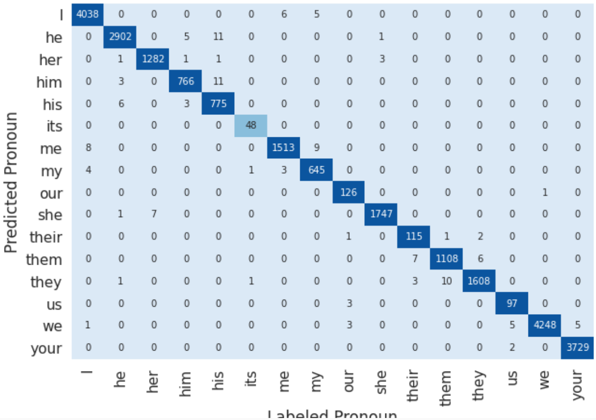


Figure 9: Heatmap of `aX + Xb` model in predicting the pronoun

Experiment 2. Results on predicting pronouns

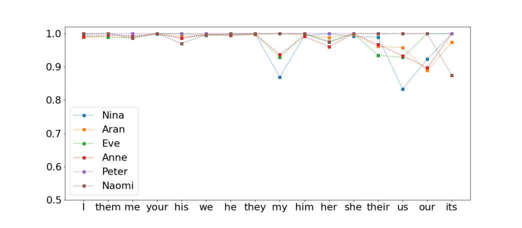


Figure 10: Training results of pronoun on `aX + Xb` model with 24889 tokens

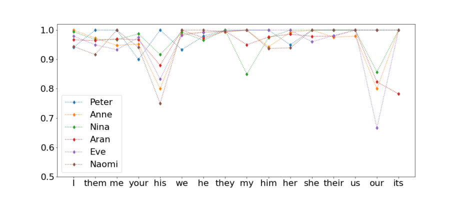


Figure 11: Training results of pronoun on `aX + Xb` model with 7355 tokens of unique types

Experiment 3: Corpus Analysis of Children's Pronoun Case Errors

Methods

Each child's pronoun case errors were searched in all available files.

	Errors	Total Pronouns	Accuracy
Anne	57	5009	0.989
Aran	25	8450	0.997
Peter	115	4077	0.971
Eve	49	2685	0.982
Naomi	64	3249	0.980
Nina	633	8609	0.926
Total	**943**	**32079**	**0.970**

Table 3: Results of each child's pronoun case errors and accuracy

Comparison between errors made by children and the model

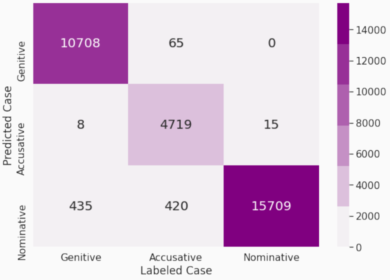


Figure 12: Children's pronoun case error heatmap

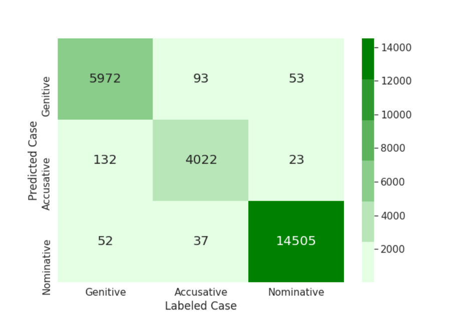
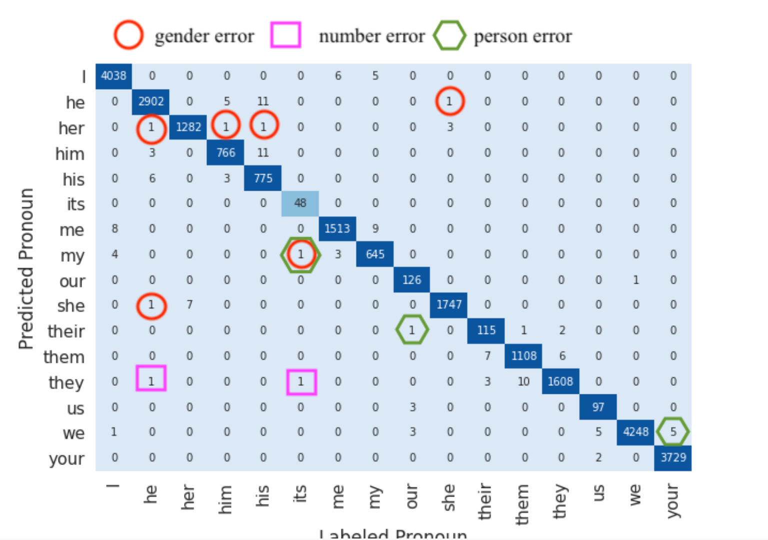


Figure 13: `aX + Xb` heatmap on 24889 tokens

Errors in predicting pronoun



Conclusion

- Distributional patterns in parents' input are very useful in categorizing grammatical cases.
- `aX + Xb` model showed similar accuracy rate as children in real life.
- Children and models made similar errors.
- However, these results are not evidence that children actually use distributional cues to acquire pronoun case.

Questions and Comments

References

Bloom, L., Hood, L., and Lightbown, P. (1974). Imitation in language development: If, when, and why. *Cognitive psychology*, 6(3):380–420.

Brown, R. (1973). *A first language: The early stages.* Harvard U. Press.

Clair, M. C. S., Monaghan, P., and Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition*, 116(3):341–360.

Kirjavainen, M., Theakston, A., and Lieven, E. (2009). Can input explain children's me-for-i errors? *Journal of Child Language*, 36(5):1091–1114.

Kuczaj, S. A. (1978). Why do children fail to overgeneralize the progressive inflection? *Journal of Child language*, 5(1):167–171.

MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1):91–117.

Rispoli, M. (1994). Pronoun case overextensions and paradigm building. *Journal of child language*, 21(1):157–172.

Rispoli, M. (2005). When children reach beyond their grasp: Why some children make pronoun case errors and others don't. *Journal of Child Language*, 32(1):93–116.

Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent-child discourse. *Children's language*, 4:1–28.

Schütze, C. and Wexler, K. (1996). Subject case licensing and english root infinitives. In *Proceedings of the 20th annual Boston University conference on language development*, volume 2, pages 670–681. Cascadilla Press Somerville, MA.

Suppes, P. (1974). The semantics of children's language. *American psychologist*, 29(2):103.

Theakston, A. L., Lieven, E. V., Pine, J. M., and Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of child language*, 28(1):127–152.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74(3):209–253.