Children’s Pronoun Case Error Revisit

A Computational Model on Distributional Cues

Xiaomeng Ma

April 25, 2019

Graudate Center, CUNY

Roadmap

Introduction

Children's Pronoun Case Error

English speaking children make pronoun case errors:

- Pronoun case errors usually occur in children at the age of 2 to 4.
- Most two common types:
  - Accusative case as Subject:

    (1) * Me bite. (Abe, 1;9)

  - Nominative case as Object:

    (2) * When me see he again? (Eve, 2;0)

- Pronoun case errors have been considered as **frequent, systematic** and **characteristic** mistakes.

Why Children make Pronoun Case Errors

ATOM (Agreement/Tense Omission Model)

Wexler (1994); Wexler et al. (1998)

Paradigm building Model

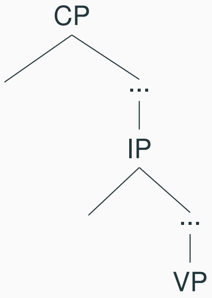
Rispoli (1994, 1998, 1999, 2005)

Input Driven Model

Theakston et al. (2001, 2002, 2004)

Basic Syntactic Structure of the Case Error

`[CP...[IP(tense and agreement)...[VP...]]]`



ATOM Agreement/Tense Omission Model

Explanation

Agreement is manifested on use of Nominative case. When children fail to check agreement and/or tense, they produce errors:

- *Mummy do it.
- *Me do it.

ATOM Agreement/Tense Omission Model

Explanation

Agreement is manifested on use of Nominative case. When children fail to check agreement and/or tense, they produce errors:

- *Mummy do it.
- *Me do it.

Prediction

Accusative cases are more likely to occur with infinite verbs.

- *Her want cookie » *Her wants cookie.

Paradigm Building Model

Explanation

`case`, `person` and `number` form a 3x3 paradigm for all nouns. Children have difficulties access to all the pronouns in the paradigm.

Paradigm Building Model

Explanation

`case`, `person` and `number` form a 3x3 paradigm for all nouns. Children have difficulties access to all the pronouns in the paradigm.

Prediction

Children will make more errors on the pronouns the use more frequently. Error rate is also related in their use of finiteness structures.

Input Driven Analysis

Explanation

Children learn language largely from input. Not only frequency of the pronouns but also sequences of word could be play a role in the errors.

Input Driven Analysis

Explanation

Children learn language largely from input. Not only frequency of the pronouns but also sequences of word could be play a role in the errors.

Prediction

If the children hear "Acc + V" ("Let `me do` that.") and "Nom + V" ("`I do` that everyday.") more often, they are more likely to make errors, probably on the same verb too.

Some Questions

- **Which explanation is the most appropriate?**
- **What is the average error rate for this frequent and characteristic error?**
- **What kinds of invididual differences are there?**

First Study: Corpus Analysis

Data and Corpora

Data[1] Selection Criteria:

- North American English
- Files are in `xml` form that has been tagged on the *%mor* tier and *%gra* tier
- Pronoun rate (Cased Pronouns/Total words) > 5% in each file

[1] All data are from CHILDES (MacWhinney, 2014)

Data and Corpora

This study included **22** children with longitudinal data and **173** children with cross-sectional data.

Corpora	Child	Age	Corpora	Child	Age
Bloom et al. (1974)	Peter	1;9-3;2	Suppes (1974)	Nina	2;0-3;4
Braunwald (1971)	Laura	1;5-4;0	Kuczaj (1978)	Abe	2;5-4;0
Brown (1973)	Adam	2;3-4;0	Demetras (1986)	Trevor	2;1-4;0
	Eve	1;6-2;3	Weist et al. (2009)	Ben	2;4-3;4
	Sarah	2;3-4;0		Emily	2;6-3;4
Demetras (1989)	Jimmy	2;2-2;10		Emma	2;7-3;9
Clark (1978)	Shem	2;3-3;2		Jilian	2;1-2;10
Sachs (1983)	Naomi	1;3-4;9		Matt	2;5-5;0
MacWhinney (2014)	Ross	1;4-5;0		Roman	2;3-4;0
Post (1993)	She	1;8-2;5	Snow (1990)	Nathaniel	3;1-3;3
	Tow	1;9-2;5	Hayes and Ahrens (1988)	Geraldine	1;6-2;2

	No.	Mean Age		No.	Mean Age
Bates et al. (1991)	11	2;4	Bohannon III (1977)	2	3;6
Gleason (1980)	19	4;8	Snow et al. (1995)	79	3;11
Snow (1989)	25	2;8	Valian (1991)	17	2;5
Van Kleeck (1980)	19	3;9			

Data Collection

For each child:

General Data

Age, MLU, case errors

Data to ATOM

total number of verbs, past tense, perfective, and infinite forms

Data for Input Driven Hypothesis

Parent's: total number of words, pronouns, cased pronouns, verbs, past tense verbs, perfective, infinite forms

Data for Paradigm Building Model

Child's: total number of words, pronouns, cased pronouns, each pronoun

Data Collection

For each child:

General Data

Age, MLU, case errors

Data to ATOM

total number of verbs, past tense, perfective, and infinite forms

Data for Input Driven Hypothesis

Parent's: total number of words, pronouns, cased pronouns, verbs, past tense verbs, perfective, infinite forms

Data for Paradigm Building Model

Child's: total number of words, pronouns, cased pronouns, each pronoun

All the data are generated automatically using `nltk` python package

Automatic Data Generalization

nltk package basic

```
import nltk
from nltk.corpus.reader import CHILDESCorpusReader
valian = CHILDESCorpusReader('corpora/childes/data-xml/Eng-USA-M
valian.age('Valian/01a.xml',month=True)
[25]
valian.MLU('Valian/01a.xml')
[2.35746606334..]
len(valian.words('Valian/01a.xml',speaker='CHI'))
[571]
```

Automatic Data Generalization

Case Errors

By searching `%gra` tier[1], when a nominative pronoun is tagged as `OBJ` or an accusative pronoun is tagged as `SUBJ`, then it counted as an error.

Example

```
4297 *CHI: when me <see Fra> [//] see he again ?
4298 %mor:  conj|when pro:obj|me v|see pro:sub|he
adv|again ? 4299 %gra:  1|3|LINK 2|3|SUBJ 3|0|ROOT 4|3|OBJ
5|4|JCT 6|3|PUNCT
```

(Brown/Eve/020100.xml)

[1] The reliability of this automated morphosyntactic annotation system has been reported to have high-level accuracy, that the precision for SUBJ is 95.8% and for OBJ is 94.1% (Sagae et al., 2010)

Data Summary

Longitudinal	pronouns	cased pronouns	errors	error rate	mean error rate per child	mean max error rate
	62538	41581	224	0.54%	1.32%	9.79%

Cross-sectional	pronouns	cased pronouns	errors	error rate	mean error rate per child	mean max error rate
	13286	8932	75	0.84%	0.75%	11.21%

The error rate for children is really small.

Pronoun case errors are rare in children's utterances.

Data Summary

Name	MLU	words	pronouns	cased pronouns	errors	error-rate	age for Max error
Abe	7.6	158538	21754	13404	47	2.2%(13.6%)	3;1
Adam	5.23	145563	19883	11924	82	1%(11.4%)	2;11
Ben	5.92	13730	1459	1107	19	3.14%(13.79%)	2;8
Emily	6.52	32325	4267	2894	14	0.82%(2.27%	2;11
Emma	5.28	26829	3104	2012	3	0.43%(3.85%)	2;8
Eve	4.34	21123	2967	1759	11	1.59%(2.52%)	2;1
Geraldine	3.74	3294	407	288	0	0.00%	
Nathaniel	5.39	5220	531	375	0	0%	
Nina	4.38	87938	9027	6505	100	4.91%(32.22%)	2;6
Naomi	3.70	32325	3871	2440	22	2.07%(21.69%)	2;1
Roman	6.72	47943	5264	3462	10	0.39%(2.9%)	2;6
Sarah	3.96	95969	12418	8142	37	1.64%(10.3%)	2;9
She	3.04	6768	665	434	9	2.94%(15.8%)	1;9
Tow	7.56	158538	21754	13404	47	2.23%(13.6%)	3;1
Jillian	4.87	20492	2138	1578	2	0.11%(0.39%)	2;5
Jimmy	4.45	17859	1852	1309	11	0.82%(1.56%)	2;7
laura	3.82	64469	7689	5259	28	0.52%(2.75%)	3;0
Matt	6.47	43989	5634	3672	23	0.76% (7.95%)	2;5

Mean Age	MLU	words	pronouns	cased pronouns	errors	error-rate	Mean Age for Max error
3;3	3.66	97907	13286	8932	75	0.8% (24.4%)	3;10

Results

Correlation between Error count/rate vs Age and MLU:

- There is no unified correlational relationship among all children.
- Age and Errors are moderately to slightly negatively correlated: Abe(-.52**), Nina(-.63**), Eve(-.29*)
- Children whose age and errors are not correlated: Roman(.07), Sarah(-.02)
- Children whose age and errors are positively correlated: Jimmy (0.46*)

Results

Correlation between Error count/rate vs Age and MLU:

- There is no unified correlational relationship among all children.
- Age and Errors are moderately to slightly negatively correlated: Abe(-.52**), Nina(-.63**), Eve(-.29*)
- Children whose age and errors are not correlated: Roman(.07), Sarah(-.02)
- Children whose age and errors are positively correlated: Jimmy (0.46*)
- Errors and MLU positively correlated: Eve(.41*), Emily(.34)
- Errors and MLU not correlated: Sarah (.09), Jillian (.07)
- Errors and MLU negatively correlated: Nina (-.55**), Matt (-.40**)

Similar unified pattern also found in other variables (including verbs, children's use of pronouns, parent's input, e.t.c)

Conclusion and Discussion

Conclusion

- Children's pronoun case error is surprisingly low and not frequent.
- Some children don't make pronoun case errors at all.
- For those who make pronoun case errors, there's such high individual differences that one single theory might fail to account for all

Conclusion and Discussion

Conclusion

- Children's pronoun case error is surprisingly low and not frequent.
- Some children don't make pronoun case errors at all.
- For those who make pronoun case errors, there's such high individual differences that one single theory might fail to account for all

Discussion

- Children's learning on abstract pronoun case is quite effective.
- We've been focusing on why they made errors, instead we never attempt to explain how they learn the cases.
- What is the learning mechanism?

Second Study: Computational Model on Distributional Cues

Background

Distributional cues in pronoun case errors

Tomasello (2000, 2009) proposed that the frame structure the children are being exposed to might mislead their interpretation of the correct use of certain case. For example, they may derive *Me do it.* from *Let me do it.*

Background

Distributional cues in pronoun case errors

Tomasello (2000, 2009) proposed that the frame structure the children are being exposed to might mislead their interpretation of the correct use of certain case. For example, they may derive *Me do it.* from *Let me do it.*

Distributional Cue in word category learning

Children might use bigram (aX,Xb) or trigram(aXb) frame to infer the grammatical category of word X. (Mintz, 2003; Clair et al., 2010) have built computational models on the distributional patterns of parents' input and showed that frequent frames are effective and efficient source of information for categorizing words.

Data and Corpora

Followed procedures in Clair et al. (2010)

Corpora Analyzed:

Peter, Nina, Eve, Naomi

Data Collection

- Get bigrams and trigrams containing pronouns
- Devide the bigrams and trigrams into aX, Xb and aXb frames
  e.g. ('I', "don't") ("want","me") ('want', 'me', 'to')
- label the pronoun as NOM or ACC

Analysis

Uniqueness

Basic frequency accounts to see how many of the frame is unique in NOM or ACC case.

Analysis

Uniqueness

Basic frequency accounts to see how many of the frame is unique in NOM or ACC case.

Feedforward Neural Network

With one 100-units hidden layer feedforward neural network

Decision Tree

Term Frequency-Inverse Document Frequency (TD-IDF) Vectorized Trained on Naive Bayesian and Random Forrest

Evaluation

Intrinsic

Train on Parents' utterances (80%)

Test on Parents' utterances (20%)

Extrinsic

Train on Parents' utterances

Test on Children's utterances

Results

Uniqueness

Child	Frame	No.Token	No.Type	NOM Token	NOM Type	ACC Token	ACC Type
Peter	aX	649	254	0.66	0.57	0.95	0.67
	Xb	649	312	0.84	0.61	0.64	0.60
Naomi	aX	926	205	0.57	0.53	0.95	0.63
	Xb	926	288	0.83	0.53	0.72	0.53
Eve	aX	1221	206	0.52	0.5	0.88	0.61
	Xb	1217	325	0.77	0.55	0.57	0.6
Nina	aX	7778	500	0.34	0.44	0.70	0.48
	Xb	7766	1094	0.49	0.40	0.38	0.42

Results

Evaluation

Child	Frame	Feedforward NN		Naive Bayesian		Random Forest	
		Intrinsic	Extrinsic	Intrinsic	Extrinsic	Intrinsic	Extrinsic
Peter	aX	0.7025	0.7794	0.86503	0.8965	0.9018	0.90184
	Xb	0.6854	0.7549	0.8733	0.8947	0.9354	0.9287
	aXb	0.7633	0.7820	0.8159	0.7826	0.82822	0.8349
Naomi	aX	0.7635	0.87764	0.88235	0.88234	0.92213	0.920703
	Xb	0.7248	0.89558	0.8916	0.879	0.91388	0.913168
	aXb	0.8023	0.86437	0.87616	0.87925	0,874	0,88589
Eve	aX	0.89714	0.909217	0.96338	0.96198	0.935193	0.935193
	Xb	0.80857	0.877094	0.90160	0.897025	0.88363	0.88789
	aXb	0.89428	0.76821	0.86956	0.867276	0.90160	0.90255
Nina	aX	0.973	0.8919	0.97079	0.97188	0.9022	0.8903
	Xb	0.88321	0.86552	0.8868	0.88499	0.88363	0.887890
	aXb	0.96629	0.838	0.91639	0.9142	0.84024	0.836419

Conclusion and Discussion

Conclusion

- Distributional Cues are effective and informative enough for to decide case
- Even with small sample size, the aX, Xb and aXb frames are powerful enough to generate results better than chance and better than frequency count

Conclusion and Discussion

Conclusion

- Distributional Cues are effective and informative enough for to decide case
- Even with small sample size, the aX, Xb and aXb frames are powerful enough to generate results better than chance and better than frequency count

Discussion

- How to interpret the extrinsic evaluation? What does it mean when it is higher than the intrinsic evaluation?
- How to relate this result to real life learning?
- How to model developmental pattern and errors?

References

Bates, E., Bretherton, I., and Snyder, L. S. (1991). *From first words to grammar: Individual differences and dissociable mechanisms*, volume 20. Cambridge University Press.

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

Bohannon III, John Neil, M. A. L. (1977). Children's control of adult speech. *Child Development*, pages 1002–1008.

Braunwald, S. R. (1971). Mother-child communication: the function of maternal-language input. *Word*, 27(1-3):28–50.

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.

Clark, E. V. (1978). Awareness of language: Some evidence from what children say and do. In *The childâĂŹs conception of language*, pages 17–43. Springer.

Demetras, M. (1989). Changes in parentsâĂŹ conversational responses: A function of grammatical development. *ASHA, St. Louis, MO*.

Demetras, M. J.-A. (1986). Working parents'conversational responses to their two-year-old sons (linguistic input, language acquisition).

Gleason, J. B. (1980). The acquisition of social speech routines and politeness formulas. In *Language*, pages 21–27. Elsevier.

Hayes, D. P. and Ahrens, M. G. (1988). Vocabulary simplification for children: A special case of âĂŸmothereseâĂŹ? *Journal of child language*, 15(2):395–410.

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.

Post, K. N. (1993). The language learning environment of laterborns in a rural florida community.

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

Rispoli, M. (1998). Me or my: Two different patterns of pronoun case errors. *Journal of Speech, Language, and Hearing Research*, 41(2):385–393.

Rispoli, M. (1999). Case and agreement in english language development. *Journal of Child Language*, 26(2):357–372.

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.

Sagae, K., Davis, E., Lavie, A., MacWhinney, B., and Wintner, S. (2010). Morphosyntactic annotation of childes transcripts. *Journal of child language*, 37(3):705–729.

Snow, C. (1990). The child language data exchange system: An update. *Journal of child language*, 17(2):457–472.

Snow, C. E. (1989). Imitativeness: A trait or a skill? In *The many faces of imitation in language learning*, pages 73–90. Springer.

Snow, C. E., Tabors, P. O., Nicholson, P. A., and Kurland, B. F. (1995). Shell: Oral language and early literacy skills in kindergarten and first-grade children. *Journal of Research in Childhood education*, 10(1):37–48.

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.

Theakston, A. L., Lieven, E. V., Pine, J. M., and Rowland, C. F. (2002). Going, going, gone: The acquisition of the verb âĂŸgoâĂŹ. *Journal of Child Language*, 29(4):783–811.

Theakston, A. L., Lieven, E. V., Pine, J. M., and Rowland, C. F. (2004). Semantic generality, input frequency and the acquisition of syntax. *Journal of child language*, 31(1):61–99.

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

Tomasello, M. (2009). *Constructing a language*. Harvard university press.

Valian, V. (1991). Syntactic subjects in the early speech of american and italian children. *Cognition*, 40(1-2):21–81.

Van Kleeck, A. (1980). The effects of children's language comprehension level on adults' child-directed talk. *Journal of Speech, Language, and Hearing Research*, 23(3):546–569.

Weist, R. M., Pawlak, A., and Hoffman, K. (2009). Finiteness systems and lexical aspect in child polish and english. *Linguistics*, 47(6).

Wexler, K. (1994). 14 optional infinitives, head movement and the economy of derivations1. *Verb movement*, page 305.

Wexler, K., Schütze, C. T., and Rice, M. (1998). Subject case in children with sli and unaffected controls: Evidence for the agr/tns omission model. *Language Acquisition*, 7(2-4):317–344.

Thanks

Big Thanks to:

The Lund Symposium on Cognition, Communication and Learning

Virginia Valian

Martin Chodorow