Polysemy: a new bias in child language acquisition?

Bernardino Casas Neus Català Antoni Hernández-Fernández Ramon Ferrer-i-Cancho Jaume Baixeries

QUALICO 2018 July 5-8, Wroclaw, Poland

◆□ ▶ < @ ▶ < @ ▶ < @ ▶ < @ ▶ < @ ■ < の < 0 </p>

2/43

CQL Lab (UPC - BarcelonaTech)



UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Complexity and Quantitative Linguistics Lab

www.cs.upc.edu/~cqllab



Bernardino Casas UPC-BarcelonaTech Polysemy: a new bias in child language acquisition?

□ ▶ ▲ ヨ ▶ ▲ ヨ ▶ ヨ ヨ つくぐ

3/43

Contents









Contents









- ▲母 ▶ ▲目 ▶ ▲目 ■ のへで

5/43

Learning vocabulary

- Children are exposed to thousands and thousands of word tokens through an accumulation of small interactions grounded in context.
- During the word acquisition process of a child, some words are learned first instead of others.
- What is the reason for some word is learned before than another one?

Biases in the learning process

- Many biases have been hypothesized in order to explain why some words are learned earlier:
 - frequency (Goodman, Dale, & Li, 2008)
 - basic taxonomic level, i.e. less generic words (Mervis, 1987; Tomasello, 2000)
 - neighborhood density, i.e. number of words that sound similar to a given word (Storkel, 2004; Storkel, Armbruster, & Hogan, 2006)
 - nouns, e.g., children learn first nouns before verbs (Gentner, 1982, 2006)
 - ...
- We hypothesize that polysemy is a factor that influences the vocabulary learning process of children.

→ @ ▶ ★ E ▶ ★ E ▶ 로 E ● 9 Q @

7/43

What is polysemy?

Definition 1: Polysemy

Polysemy is the capacity of a word to have multiple meanings.

- Our measure of polysemy of a word is the number of meanings that have this word according the syntactic category.
- For example, the word cat
 - as noun has 8 meanings \rightarrow polysemy 8
 - as verb has 2 meanings \rightarrow polysemy 2

Main goals

First goal

Investigate the relationship between the polysemy of a word and child age for English.

Second goal

Validate the nouns bias in language acquisition for English.

▲圖▶ ▲ 필▶ ▲ 필▶ - 필|필 - 의 ۹ (○)

9/43

Motherese/Infant-Directed Speech/Baby talk

- When exploring biases in child speech, it is crucial to distinguish between biases which are genuine and biases which may just be the result of mimicking some form of adult speech.
- Motherese is the special language that adults use to speak with children.
- The adult speech is the target stage in the evolution of the verbal production of a child.
- If we take as usual adult speech as control then it is crucial to know what kind of input is receiving the child and if the adult is performing adaptive efforts.

Results

Conclusions and future work

◎ ▶ ▲ 토 ▶ ▲ 토 ▶ 三目目 のへで

10/43

Fighting versus Motherese

IMPORTANT!!

We study each adult role that appears in a conversation with a child separately. Thus, we improve the perspective of the use of language in the adults for every role: MOTHER, FATHER and INVESTIGATOR.

Bernardino Casas UPC-BarcelonaTech Polysemy: a new bias in child language acquisition?

Contents





3 Results



Research Materials I

- Princeton WordNet is a large lexical database that we use to obtain a measure of potential polysemy (number of meanings of a word taking into account the syntactic category of that word). WordNet has 155,287 lemmas and 117,659 synsets (meanings).
- SemCor corpus is a syntactically tagged corpus that we use to obtain a more real measure of polysemy (used meanings).
- Treetagger is a tool that annotates text with part-of-speech (POS) and lemma information (canonical form).

< ロ > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 < つ < ○

▲■▶▲■▶▲■▶ 三国国 のく(?)

13/43

Research Materials II

 CHILDES database contains transcriptions of conversations between children and adults (usually mothers and sons). The data source of this study are CHILDES conversations in English. Children age between 10 and 70 months aprox.

Role	Тур	es	Tokens		
	WN	SC	WN	SC	
Children	9,187	4,702	1,358,219	994,269	
Mother	13,225	6,996	2,269,801	1,871,262	
Father	6,808	4,221	313,593	256,979	
Investig	3,977	2,793	182,402	151,133	

Research Methods

- We analyze syntactically every conversation (using Treetagger) and we extract the words that someone "knows" in every point of time (age of children).
- We take into account only content words (nouns, verbs, adjectives and adverbs).
- The distribution of the data over time is not homogeneous (some time points have more data than other).
- Two ways to calculate the polysemy of a word: by using WordNet or SemCor.

< ロ > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 < つ < ○

Contents







4 Conclusions and future work

< ロ > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 < つ < ○

31= 9QQ

16/43

Evolution of mean polysemy vs child age



Bernardino Casas UPC-BarcelonaTech Polysemy: a new bias in child language acquisition?

31= 9QQ

17/43

< E

Correlations between mean polysemy and child age



WordNet

SemCor



S+ : Significant positive correlations

S-: Significant negative correlations

S? : Non Significant correlations

Validating results using ANOVA I

- In order to validate the results, we have divided the timeline into ten segments of the same time size.
- In each segment, we compare the mean polysemy (both WordNet and SemCor) of children with that of adults.
- The analysis is based on a one-way independent samples ANOVA with role (two levels: children and adults) as fixed factor and the identifier of the individual as random factor.

< ロ > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 < つ < ○

・ロ · ・ (日) · ・ モ · モ · 王 · つ へ · 19/43

ANOVA test results

			WordNet		SemCor	
segment	from / to	Ν	F	p-val	F	p-val
1	11.9 / 17.4	14	13.331	0.002	18.141	$< 10^{-3}$
2	17.4 / 22.9	99	158.401	$< 10^{-21}$	241.664	$< 10^{-27}$
3	22.9 / 28.4	139	90.728	$< 10^{-16}$	173.963	$< 10^{-25}$
4	28.4 / 33.9	139	35.397	$< 10^{-7}$	82.031	$< 10^{-15}$
5	33.9 / 39.4	127	11.723	$< 10^{-3}$	31.37	< 10 ⁻⁷
6	39.4 / 44.9	96	7.659	0.003	15.817	$< 10^{-4}$
7	44.9 / 50.3	25	3.827	0.031	12.408	$< 10^{-3}$
8	50.3 / 55.8	18	1.751	0.102	8.731	0.005
9	55.8 / 61.3	48	0.859	0.179	3.313	0.038
10	61.3 / 66.8	9	0.005	0.473	0.602	0.232

 H_0 = "" (null hypothesis) and H_1 = "" (alternative hypothesis)

Validating results using ANOVA II

- The results of this test show that the average WordNet polysemy in children is significantly lower than that of adults until the **7th segment** for WordNet polysemy, with an F-value that decreases gradually at each segment.
- The results of this test show the same pattern for SemCor polysemy: the mean SemCor polysemy in children is significantly lower than adults until the **9th segment**.

< ◎ > < ■ > < ■ > < ■ > ■ ■ の < ♡ 21/43

Preliminar Conclusions I

- We have shown that polysemy tends to increase markedly over time in children compared to adults.
- Children exhibit a two-phase (fast- slow) growth of the mean polysemy, delimited by a breakpoint (31th month).
- In adults there is neither clear positive nor negative tendency.
- WordNet and SemCor polysemy show a similar trend.

Percentage of word tokens over time



breakpoint:

nouns: 30.0 ± 0.1 months verbs: 33.0 ± 0.1 months

▲□▶ ▲□▶ ▲□▶ ▲□▶ ▲□■ のQ@

▲□ → ▲ 三 → ▲ 三 → 三 三 → ○ ○ 23/43

Preliminar Conclusions II

- The percentage of nouns decreases over time in children: it starts at 80%, drops to 40% and then, stabilizes.
- Verbs exhibit an opposite tendency in children: they start at 10% and increase their contribution to 40%, and finally they stabilize.
- These results are consistent with a well-known phenomenon: children tend to learn nouns earlier than verbs (Saxton, 2010).

Results

Conclusions and future work

Controlling by syntactic category

Controlling by morpho-syntactic category

We have analyzed the conversations considering only those tokens that belong to a specific morpho-syntactic category: noun, adjective, verb or adverb.

Results

Conclusions and future work

三日 のへの

25/43

▶ ▲ 臣

Evolution of nouns vs child age



三日 のへの

26/43

- 4 王

Correlations between nouns vs child age



WordNet

SemCor



S+ : Significant positive correlations

- S-: Significant negative correlations
- S? : Non Significant correlations

Results

Conclusions and future work

三日 のへの

27/43

Evolution of verbs vs child age



三日 のへの

28/43

- 4 更

Correlations between verbs and child age



WordNet

SemCor



S+ : Significant positive correlations

- S-: Significant negative correlations
- S? : Non Significant correlations

□ ▶ ▲ ■ ▶ ▲ ■ ▶ ▲ ■ ■ ● ● ● ●

29/43

Preliminar conclusions III

- The tendency for mean polysemy to increase in children seems to blur if we control by syntactic category (nouns, verbs, adjectives and adverbs).
- Verbs polysemy is higher than nouns polysemy.

Contents









Preference for non-polysemous words

The potential preference for non-polysemous words may be:

- a standalone bias. This hypothesis would be consistent with the lower uncertainty for those words with respect to their meaning, a factor that may reduce the cognitive cost of learning them as the cost for the listener would be smaller (Zipf, 1949).
- a side-effect of another bias, for instance an initial preference for nouns, the so-called noun-bias (Gentner, 1982, 2006; McDonough, Song, Hirsh-Pasek, Golinkoff, & Lannon, 2011) combined with the fact that nouns have lower polysemy (Fausey, Yoshida, Asmuth, & Gentner, 2006).

・ロト・(日)・(日)・(日)・(日)・

▲ @ ▶ ▲ E ▶ ▲ E ▶ 三目目 のへで

32/43

Something to think

Preference for nouns or for low polysemy?

Children learn firstly nouns because they prefer nouns or maybe are they choosing nouns by the low polysemy of this category?

Conclusions I

- There is a non-trivial pattern in the evolution of polysemy over time.
- Children increase their mean polysemy in two phases: an initial phase with a fast growth of polysemy and a second phase with a slower growth of polysemy.
- In contrast, adults interacting with them do not show this tendency.
- This non-trivial pattern weakens when the analysis is segmented by syntactic category.

< ロ > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 < つ < ○

▲■▶▲■▶▲■▶ 三国 のへで 34/43

Conclusions II

- Children show a tendency to learn nouns first and then verbs, which is consistent with previous research (Gentner, 1982, 2006; Goodman et al., 2008).
- Verbs have a significantly higher mean polysemy than nouns in all roles: children and adults.
- The last two facts could explain the pattern of the evolution of polysemy over time to some extent.

◎ ▶ ▲ 臣 ▶ ▲ 臣 ▶ 三国 = つくぐ

35/43

Last remark

L1 vs L2

That role of a standalone bias for low polysemy cannot be neglected. Our findings and (Crossley, Salsbury, & McNamara, 2010) suggest that L1 and L2 learners resemble each other in their dominant biases on polysemy: a preference for low polysemy words prevails in both kinds of learners.

●><</p>
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●</p

36/43

Future work

- A deeper understanding of why the bias for low polysemy weakens within specific syntactic categories and how it interacts with frequency is a challenge for future research.
- The relationship between the senses that a speaker really knows about a word and its potential number of synsets should be investigated in detail.

□ > < E > < E > E = のへの

37/43

Thank you!

Thank you for your attention!

Bernardino Casas UPC-BarcelonaTech Polysemy: a new bias in child language acquisition?

Conclusions and future work

Questions?



Bibliography I



Crossley, S., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning*, *60*(3), 573–605.

Fausey, C. M., Yoshida, H., Asmuth, J., & Gentner, D. (2006). The verb mutability effect: noun and verb semantics in English and Japanese. In *Proceedings of the Twenty-eighth Annual Meeting of the Cognitive Science Society* (pp. 214–219).

Bibliography II

 Gentner, D. (1982). Why nouns are learned before verbs:
 Linguistic relativity versus natural partitioning. In S. A.
 Kuczaj II (Ed.), Language development: Vol. 2. Language, thought and culture (Chap. 11, pp. 301–334). Hillsdale:
 Lawrence Erlbaum Associates.

Gentner, D. (2006). Why verbs are hard to learn.

(pp. 544–564). Action meets word: How children learn verbs. Oxford University Press.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, 35(3), 515–531. doi:10.1017/S0305000907008641

Bibliography III

McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). An image is worth a thousand words: why nouns tend to dominate verbs in early word learning. *Developmental Science*, *14*, 181–189. doi:10.1111/j.1467-7687.2010.00968.x
 Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.), *Concepts and conceptual development: echological and intellectual factors in categorization* (pp. 201–233). New York, NY.

US: Cambridge University Press.

Saxton, M. (2010). Child language: acquisition and development (1st ed.). London, UK: SAGE publications.

Bibliography IV

Storkel, H. L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. Applied Psycholinguistics, 25, 201–221. Storkel, H. L., Armbruster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. Journal of Speech, Language, and Hearing Research, 49, 1175–1192. Tomasello, M. (2000). Do young children have adult syntactic competence? Cognition, 74, 209-253. doi:10.1016/S0010-0277(99)00069-4 Zipf, G. K. (1949). Human behaviour and the principle of least

□ > < E > < E > E = のへで

License



This work is licensed under a Creative Commons "Attribution-NonCommercial-ShareAlike 4.0 International" license.

・ロト <
一 ト <
三 ト <
三 ト <
三 ト <
三 ト <
、 2 | = の へ の 43/43
</p>