Get out but don't fall down: computational investigations of verbs and verb-particle constructions in child language

Aline Villavicencio and Marco Idiart
Federal University of Rio Grande do Sul (Brazil)

Language Acquisition

Children acquire language:

- in a short time
- based on data that can be noisy, incomplete and ungrammatical
- on the presence of other languages/dialects/ regional variations

Language Acquisition

- For language acquisition studies
  - large-scale naturalistic acquisition data important for assessing theories and empirical predictions helping to compare alternative theories
  - CHILDES - Child Language Data Exchange System
    - corpora for over 25 languages in raw format
    - some contain audio or video recordings

Language Acquisition

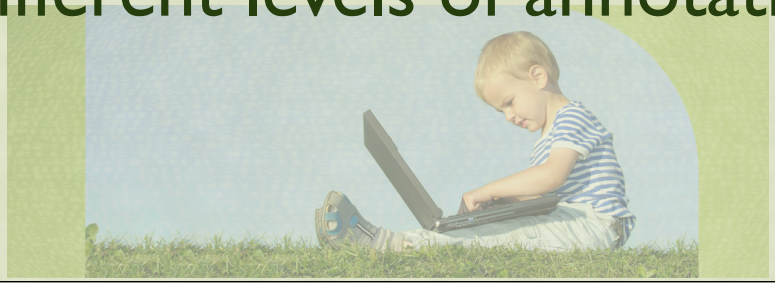
- With increasing availability of robust NLP systems and electronic resources
  - possibility of adding further linguistic, psycholinguistic and distributional annotation
  - application: profiling of MWEs in CHILDES

Outline

- English CHILDES Verb Construction Database
- Child Usage of Verb-Particle Constructions
- Conclusions and Future Works

English CHILDES Verb Construction Database - ECVCD

- Initiative for extending CHILDES annotation with:
  - grammatical information
  - semantic classes
  - psycholinguistic and
  - distributional information
- Integrated resource that allows complex searches involving different levels of annotation



English CHILDES Verb Construction Database

- Syntactic annotation with
  - MEGRASP (Sagae et al. 2010)
  - RASP (Briscoe et al., 2006, Buttery and Korhonen, 2005)
    - syntactic trees (ST) and
    - grammatical relations (GR)
- Combine annotations provided by multiple parsers with complementary strengths
  - inter-parsing agreement
  - coverage

English CHILDES Verb Construction Database

- Levin (1993) English Verb Classes - syntactic and semantic properties
  - 3,100 verb types and 90 classes
    - 4,167 verb tokens

Verbs of Contact by Impact
{bang, bash, hit, pound, ...}

Verbs of Motion
{advance, arrive, bounce, ...}

Change of State Verbs
{break, chip, divide, frost, ...}

Measure Verbs
{measure, weigh, appraise, ...}

English CHILDES Verb Construction Database

- MRC Psycholinguistic Database (Coltheart, 1981)
  - 150,837 entries with information about 26 properties
    - **Familiarity**: apple vs hard-drive
    - **Concreteness**: cat vs love
    - **Imageability**
    - **Age of Acquisition**
    - **Number of syllables**

English CHILDES Verb Construction Database

- Frequency information
  - from CHILDES
    - adults' sentences
    - children's sentences
  - from the Web
  - from the MRC Database

English CHILDES Verb Construction Database

RASP

Syntactic Tree

Grammatical Relations

MEGRASP

Dependency Parse

Sentence

w1

w2

w3

...

wn

Levin EVCA

Class1

Class2

...

Classn

Wordnet

Synset1

Synset2

...

Synsetn

MRC

Conc

Fam

...

Freq

Freq

Web

Childes

English CHILDES Verb Construction Database

Information	Sentences
Total Raw	4.84 million
MEGRASP Parsed	109,629
RASP Parsed	2.21 million
MEGRASP & RASP	98,456

English CHILDES Verb Construction Database

- Development of user-friendly interface for complex patterns
- Evaluation with users

Outline

- English CHILDES Verb Construction Database
- Child Usage of Verb-Particle Constructions
- Conclusions and Future Works

Child Usage of Verb-Particle Constructions

- Verb-Particle Constructions (VPCs):
  - carry in, put off and move on
- With variable degrees of syntactic flexibility
  - *eat up the desert* and *eat the desert up*
  - *hang around the park* and *?hang the park around*
- *Semantic idiosyncrasies*
  - *carry the suitcases up*
  - *to play down X: to (try to) make X seem less important than it really is*

Why VPCs?

- Semi-productivity
  - some verbs combine with almost every particle (*get, fall, go*)
  - some combine with only a few (*book up* and *sober up*)
  - others do not combine well (e.g., *know, want, resemble*,...) (Fraser, 1976)

Why VPCs?

- Indication of impact in learning for children
  - With increase in number of obligatory arguments children with Specific Language Impairments (SLI) use more general and fewer specific verbs

    Boynton-Hauerwas, 1998

Why VPCs?

- Given possibly higher complexity in relation to simplex verbs
  - are there less VPCs in child-produced than in child-directed speech?
  - what kind of VPCs do they use?
  - are children's VPCs more conservative than adults'
    - with less variety of VPCs and verbs/particles?

Why VPCs?

- Profiling of VPCs in English and their usage in child-produced and child-directed sentences
- Ground work for computational models of VPC learning

VPCs in Child Language

- VPCs automatically identified from RASP annotation
  - using mwetoolkit (Ramisch et al. 2010)
    - V followed by Prt up to 5 words to the right
  - automatic removal of noise: a@l up, di, dat
  - Predominance of VPC sentences in younger ages

Sentences	Children Set	Adults Set
Parsed	482,137	988,101
with VPCs	44,305	83,098
with VPCs Cleaned	38,326	82,796

Age in Months	VPC sentences	Proportion
0-24	2799	6.4%
24-48	26152	59.9%
48-72	8038	18.4%
72-96	1337	3.1%
>96	514	1.2%
No age	4841	11.1%
Total	43681	100.0%

VPCs in Child Language

- *Do children use VPCs as frequently as adults do?*

VPCs in Child Language - Sentences

Total VPC	Children Set	Adults Set
Tokens	38,326	82,796
Types	1,579	2,468

- Absolute values:
  - adults produced more than double the number of VPC tokens than children
- Relative values:
  - similar proportion, 7.95% (children) vs 8.38% (adults)

VPCs in Child Language - Frequency

- Frequencies reflects Zipfian distribution found in natural languages:
  - many VPCs occur just once
- Similar frequencies per bin for two groups

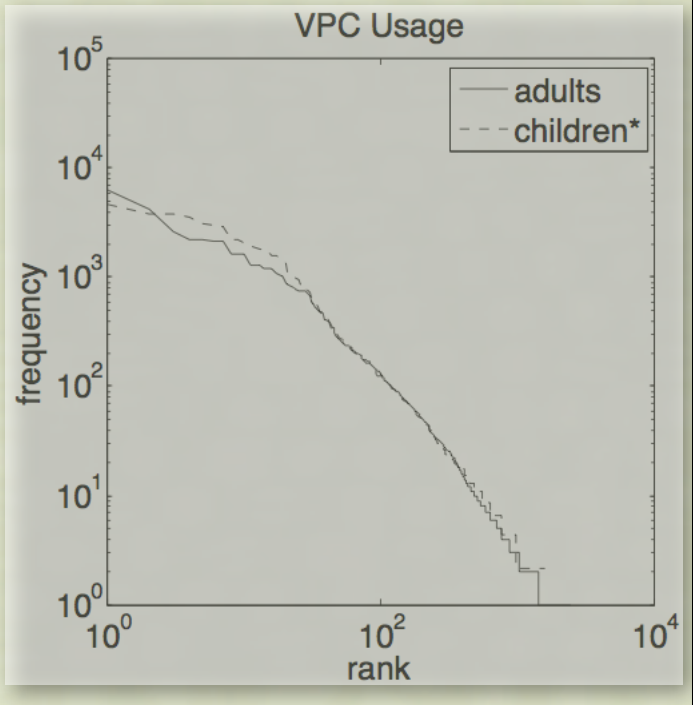
Frequency	Children Set	Adults Set
1	42.62%	43.03%
2	13.05%	15%
3	8.36%	6.48%
4	4.05%	4.5%
≥5	31.92%	31%

VPCs in Child Language

- *Quantitatively is the variety of VPCs and verbs used by children as rich as by adults?*

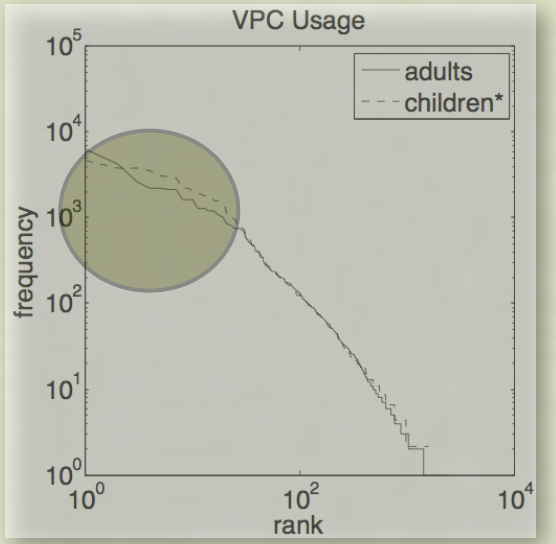
VPCs in Child Language - Rank of Types

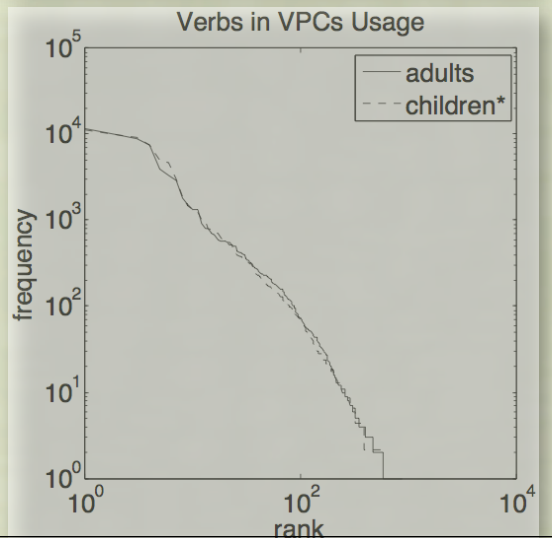
- Adults use larger VPC vocabulary
  - 1.56 more types than children
- But given lower number of children's sentences
  - distributions only differ in scale
  - adult rank = children rank * 2.16
    - 2.16 is the ratio between VPC tokens by adults and children



VPCs in Child Language - Rank of Types

- But discrepancy for higher frequency VPCs
  - Children have a more uniform distribution
  - Adults repeat more higher freq. VPCs
  - Found in VPCs but not in verbs used in VPCs
    - same scale of 2.16 gives very close match



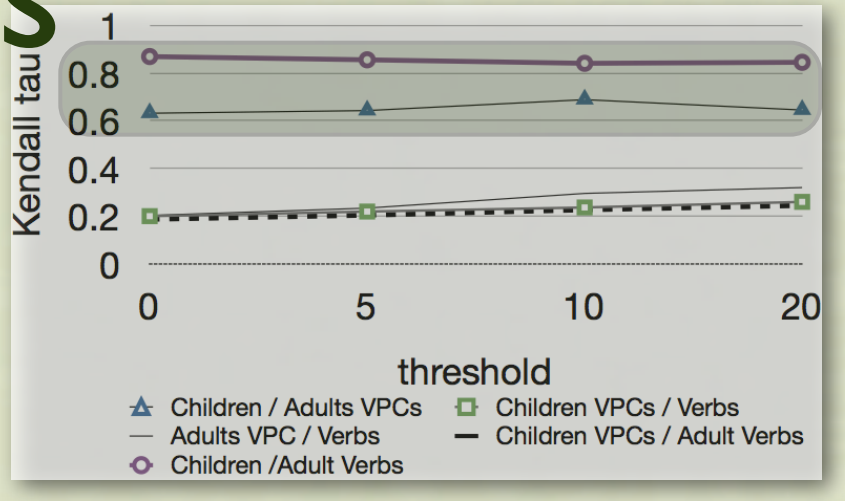


VPCs in Child Language

- *Do they use qualitatively similar VPC and verb types? Are these used with similar frequencies?*

VPCs in Child Language - Types

- Children vs Adult VPCs in vocabularies
  - Kendall τ score = 0.63
    - highly correlated (range is from -1 to 1)
- Children vs Adult Verbs in VPCs
  - Kendall τ score = 0.84
    - even stronger correlation
- Trends confirmed with frequency thresholds



VPCs in Child Language - Types

- Top 10 VPC types
  - 9/10 are the same with only different order

Rank	Chidren VPC	Children Freq	Adult VPC	Adult Freq	Child Rank
1	put on	2005	come on	6244	7
2	go in	1608	put on	4217	1
3	get out	1542	go on	2660	9
4	take off	1525	get out	2251	3
5	fall down	1329	take off	2249	4
6	put in	1284	put in	2177	6
7	come on	1001	sit down	2133	8
8	sit down	981	go in	1661	2
9	go on	933	come out	1654	10
10	come out	872	pick up	1650	18

VPCs in Child Language - Types

- Shared VPCs
  - 72.32% of children's
  - 89.48% >= 5
- Children only VPCs:
  - erase off and crash down
- Adults only VPCs:
  - 93.44% have low frequency (< 5)
  - crawl in and creep up
  - Lower frequency cases may not yet be incorporated in children's active vocabulary

	Children VPCs	Adult VPCs	Children ∩Adult VPCs	Children only VPCs	Adult only VP
VPCs	1579	2468	1142	437	1243
Verb in VPCs	561	884	401	160	483
Particle in VPCs	28	35	24	4	9
VPCs ≥ 5	504	766	451	53	278
Verb in VPCs ≥ 5	207	282	183	24	99
Particle in VPCs ≥ 5	18	20	17	1	3

VPCs in Child Language - Verbs

	Children VPCs	Adult VPCs
VPCs	1579	2468
Verb in VPCs	561	884
Particle in VPCs	28	35
VPCs ≥ 5	504	766
Verb in VPCs ≥ 5	207	282
Particle in VPCs ≥ 5	18	20

- Absolute terms:
  - adults use more verbs in VPCs
- Relative terms:
  - groups have similar ratios of verb in VPCs
    - 2.81 VPCs for children
    - 2.79 for adults

VPCs in Child Language - Verbs

Children		Adults	
Types	Tokens	Types	Tokens
7.02%	47.81%	5.83%	43.76%

- Verbs in more VPC types are also in frequent VPC tokens
  - Very general and frequent verbs
    - *go, get, come, take, put, make* and *move*
  - Among the first to be learned (Goldberg, 1999)
- Facilitated acquisition and use in VPCs

VPCs in Child Language - Distances

Distance	Children Set	Adults Set
0	65.13%	64.14%
1	23.48%	22.15%
2	9.33%	10.90%
3	1.65%	2.15%
4	0.29%	0.47%
5	0.09%	0.16%

- Distance from verb to particle
  - strong preference for joint VPCs
    - over 97% of VPCs have at most 2 words between them

Outline

- English CHILDES Verb Construction Database
- Child Usage of Verb-Particle Constructions
- Conclusions and Future Works

Conclusions and future work

- Profile of VPCs in acquisition data in English
- VPCs are used as much in children's data as in adults'
- Children and adult usage shows agreement on:
  - types and frequencies with similar distributions
  - particle placement

Conclusions and future work

- But some discrepancies
  - children with more uniform distribution for higher frequency VPCs than adults
  - children have modest but significant dispreference for longer distances between verb and particle
  - Do they reflect different strategies or efficiency considerations?
- Latitudinal vs longitudinal results
- Computational models for VPC acquisition

Acknowledgements

- This research was partly supported by CNPq Projects 551964/2011-1, 202007/2010-3, 305256/2008-4 and 309569/2009-5
- UFRGS: Carlos Ramisch, Rodrigo Wilkens, Vitor Araujo
- MIT: Robert Berwick, Beracah Yankama, Igor Malioutov

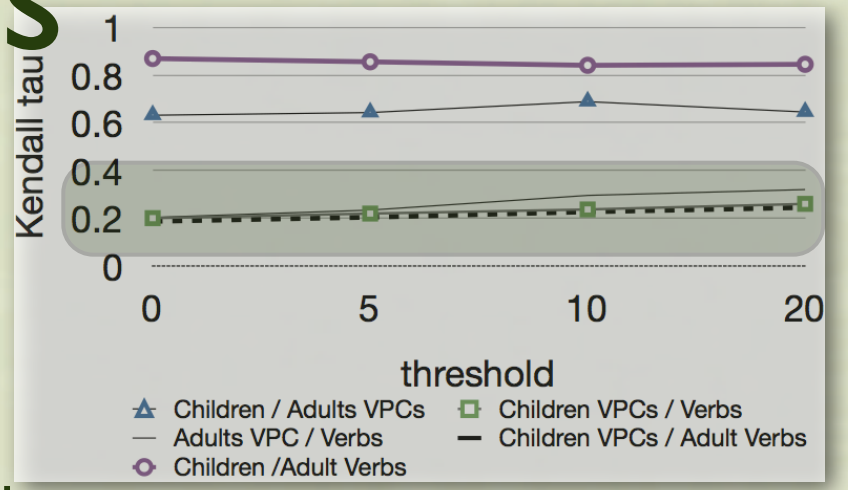
Get out but don't fall down: computational investigations of verbs and verb-particle constructions in child language

Aline Villavicencio and Marco Idiart
Federal University of Rio Grande do Sul (Brazil)
alinev@gmail.com, marco.idiart@gmail.com

Acquisition of VPCs

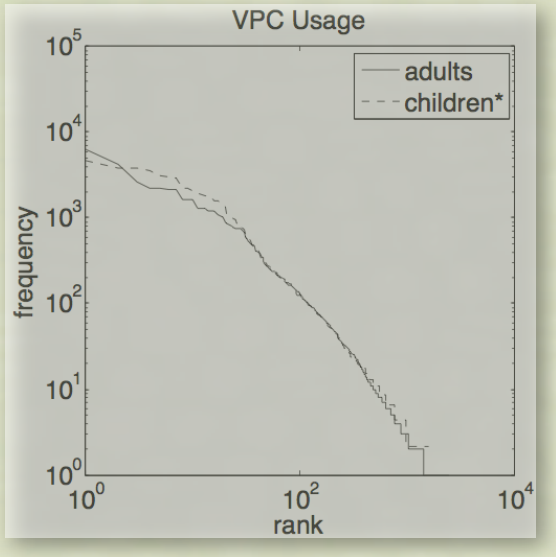
- Children seem to treat aspectual and compositional VPCs differently
  - aspectuals are more frequent and varied
  - sources of error:
    - compositional: tend to be lexical,
    - aspectuals: predominance of syntactic errors
      - 92% of object dropping errors in split configuration for children under 5 (Sawyer, 1999).
      - SLI children tended to produce more object dropping errors than TD children (Juhasz and Grela, 2008)

VPCs in Child Language - Types



- Children vs Adults, VPCs vs verbs
  - children VPC ranks vs verb ranks
  - adults VPC ranks vs verb ranks
  - children VPC ranks vs adult verb ranks
- Kendall scores of 0.2 for all
  - order of the verbs in the data is not predictive of the relative frequencies of VPCs

VPCs in Child Language - Rank of Types



- Lower number of VPC types by children due to lower number of sentences produced
  - The hypothesis that difficulties in VPCs would lead to their avoidance is not confirmed by the data

Why VPCs?

- What about particle placement (split vs joint VPCs)?
  - 84% success for (adult) placement in 403 VPCs from the BNC using multifactorial analysis with 25 variables (Gries, 2002)
    - type of NP (pronoun or lexical) and size of the direct object (in syllables and words), type of determiner (indefinite or definite)
  - For Lohse et al. (2004) this is due to processing efficiency
    - size of the object NP and dependencies among verb, particle, and object
  - the type of NP (pronoun or lexical) and semantics of the particle (spatial vs non-spatial) were good predictors on child language data (Diessel and Tomasello, 2005)

Why VPCs?

- Type of verb and number of arguments seem to have an impact in learning for children
  - more consistent use of obligatory arguments and inflectional morphology with general verbs than with more specific ones
  - But as the number of obligatory arguments increases children with Specific Language Impairments (SLI) seem to produce more general and fewer specific verbs (Boynton-Hauerwas, 1998)
- How can learner decide in which frame VPC should be realized?

VPCs in Child Language - Dictionaries

Threshold	Children	Adults
1	75.87%	72%
5	79.82%	87.72%

- Most VPCs found in dictionaries
  - VPC dataset - 3,078 VPCs (Baldwin,2008)
  - Comlex - 10,478 PVs (Macleod and Grishman, 1998)
  - ANLT - 6,351 PVs (Carroll and Grover, 1989)
- It is in the lower frequencies that novel and non-standard usages can be found
  - Adults: *crawl in* and *creep up*
  - Children: *erase off* and *crash down*

Acknowledgements

- To Valia Kordoni for kindly presenting this work
- This research was partly supported by CNPq Projects 551964/2011-1, 202007/2010-3, 305256/2008-4 and 309569/2009-5, 478222/2011-4