Distributional semantics for Child-Directed Speech

A multimodal approach

Cassani G [12], Baroni M [2]

[1] COMPUTATIONAL LINGUISTICS & PSYCHOLINGUISTICS RESEARCH CENTER / CLiPS

Universiteit Antwerpen

[2] CIMeC

UNIVERSITY OF TRENTO - Italy

Giovanni Cassani
CLiPS @ Uantwerpen

6 Feb 2015
CLIN25

Goal(s)

Test whether Child-Directed Speech (CDS) is a valuable source for distributional semantic models (DSMs) that encode information extracted from texts only and from texts and images together

Explore to which extent multimodal DSMs can explain or approximate certain aspects of early language acquisition

From linguistic to multimodal DSMs

Words occurring in similar contexts have similar meanings.
[Harris, 1954; Firth, 1957]

DSMs can shed light on cognitive processes
[Landauer & Dumais, 1997]

But…as they are, DSMs lack reference to the external world
[Vigliocco, 2009]

Why don't we enrich them with images then?
[Bruni & many others, from 2010 onward]

Computational models & language learning

Nativist vs emergentist theory: do children have a specialized and innate cognitive system for language or do they use general cognitive abilities?

Greater attention traditionally devoted to syntax learning and grammar induction

Few works about semantics and meaning (but more and more popular)

Fewer used distributional semantics

Nobody (to our knowledge) used multimodal DSMs

Data

- CDS extracted from 21 English (British and American) corpora from the CHILDES database
- Simple Wikipedia
- UkWac (a portion of)
- ImageNet database
- MEN dataset for semantic relatedness [Bruni et al, 2014]
- Concept Property Norms dataset from CSLB [Devereux et al, 2014]
- List with automatically generated concreteness scores by Turney et al [2011]

Experiment #1

Question: is CDS *more semantically informative* than general English when we consider concrete and imageable concepts?

Data: CDS from CHILDES, portion of UkWac, pairs of concepts from the MEN dataset

Setup: four DSMs (one for each factor on the dimensions CDS ~ general English and concrete words ~ all words) for computing semantic relatedness between concepts

Metric: cosine similarity between pairs of concepts

Experiment #1

Hypothesis: CDS-derived cosines correlates better with human judgments than general English for more concrete concepts

Results:

	CDS		General English	
	Pearson	Spearman	Pearson	Spearman
Concrete words	.542949	**.712464**	.536361	**.623990**
All words	.422870	.622733	.557074	.628807

Experiment #2

Question: is CDS *more perceptually grounded* than general English?

Data: CDS from CHILDES, portion of UkWac, CSLB concept property norms from a feature listing task

Setup: four DSMs (one for each factor on the dimensions CDS ~ general English and perceptual norms ~ all norms) for computing semantic relatedness between concepts

Metric: cosine similarity between pairs of concepts

Experiment #2

Hypothesis: CDS-derived cosines correlates better with human generated concept norms than general English when considering perceptual norms

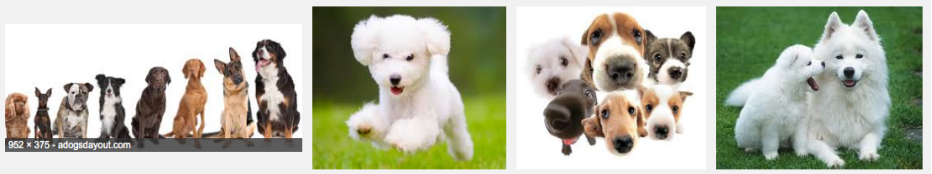
Results:

	CDS	General English
Perceptual norms	.253	.191
All norms	.398	.253

Zero-shot learning…

Train a mapping function from visual vectors for known images to distributional vectors for the words referring to the known images …

→ Dog

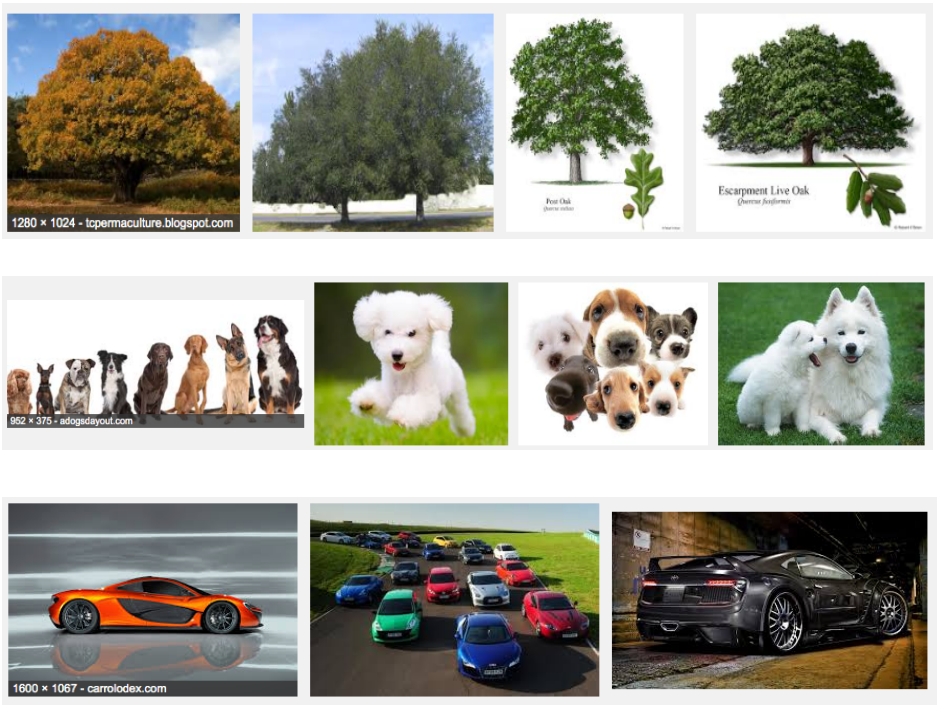


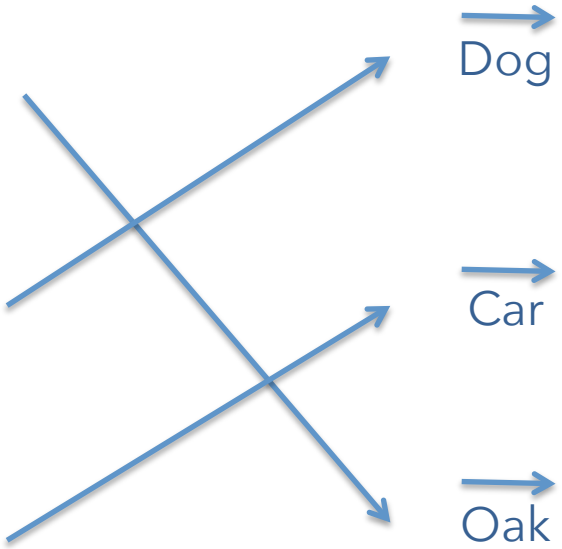
→ Car

→ Oak

Zero-shot learning…

Train a mapping function from visual vectors for known images to distributional vectors for the words referring to the known images …





Zero-shot learning…

Then feed a vector of a new image and a semantic space with new concepts, have the mapping function 'reconstruct' the linguistic vector for the given image and retrieve the nearest neighbors for this 'reconstructed' vector in the semantic space. Hopefully it will be the correct concept for the image.



→ Airplane

→ Truck

→ Duck

Zero-shot learning…

Then feed a vector of a new image and a semantic space with new concepts, have the mapping function 'reconstruct' the linguistic vector for the given image and retrieve the nearest neighbors for this 'reconstructed' vector in the semantic space. Hopefully it will be the correct concept for the image.

→ Reconstructed Vector

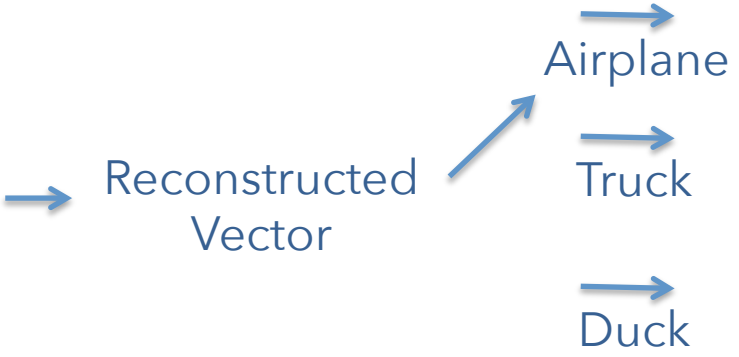
→ Airplane

→ Truck

→ Duck

Zero-shot learning…

Then feed a vector of a new image and a semantic space with new concepts, have the mapping function 'reconstruct' the linguistic vector for the given image and retrieve the nearest neighbors for this 'reconstructed' vector in the semantic space. Hopefully it will be the correct concept for the image.



… and language acquisition

Zero-shot learning can approximate the situation in which a child sees a new object while hearing someone naming it: she needs to map a new word to a new referent

However, zero-shot learning relies on huge corpora and hundreds of images, which a child can obviously not access

The more plausible setting of fast-mapping [Lazaridou et al, 2014] is not explored here

Experiment #3

Question: does zero-shot learning approximate what happens when children learn word-referent mappings?

Data: CDS from CHILDES, Simple Wikipedia, ImageNet, list with concreteness scores

Training: 2945 words vectors derived from CDS + Simple Wikipedia & correspondent images from ImageNet

Testing: 750 word vectors from CDS and correspondent images from ImageNet

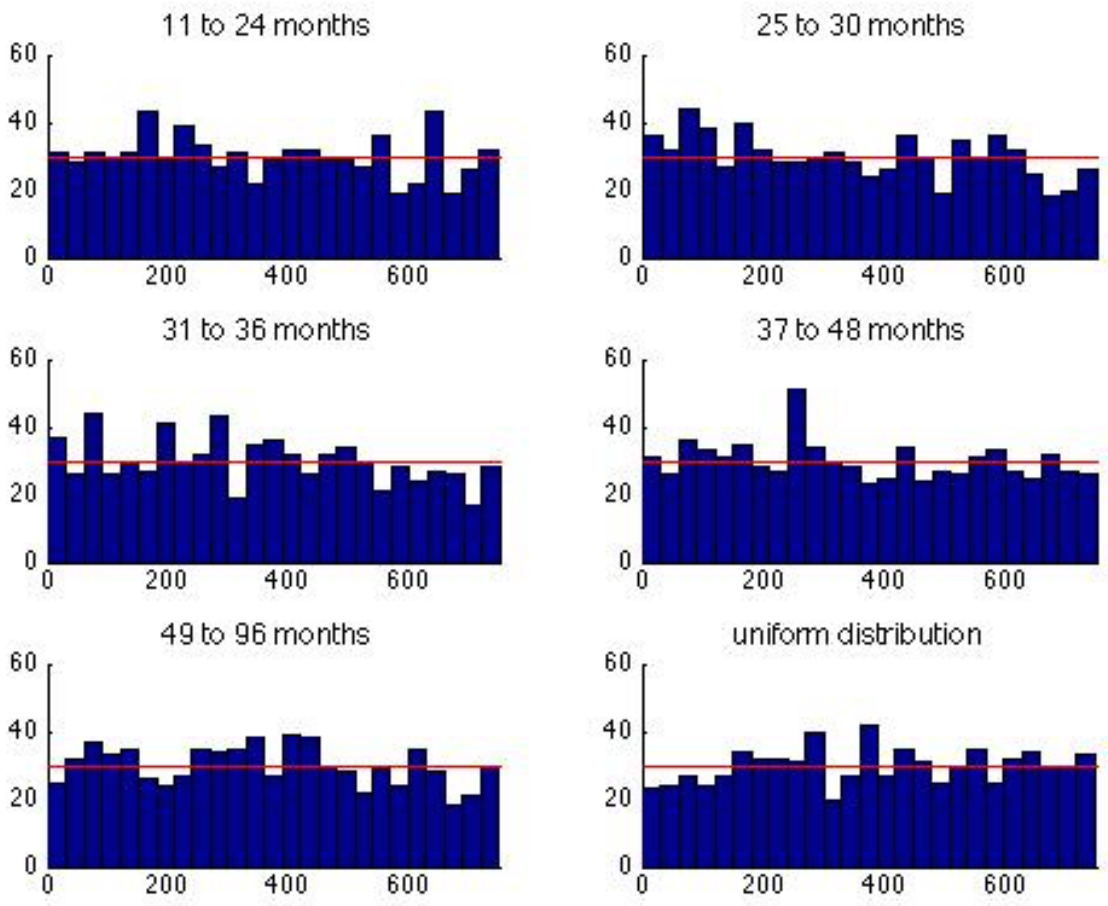
Mapping function: linear regression with regularization [Lazaridou et al, 2014]

Experiment #3

Results:

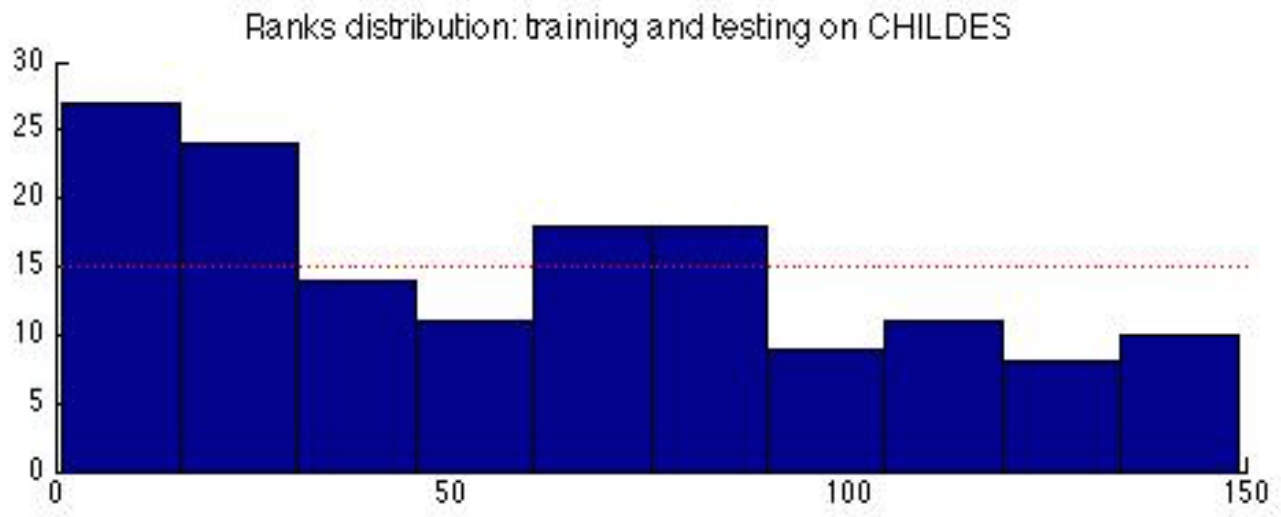
Ranks on the x

Frequencies on the y



Wait…

When training and testing using only CHILDES, results get *slightly* better:



Conclusions

CDS can be used to derive meaningful DSMs: it correlates better than general English with semantic relatedness judgments and property norms provided by human subjects

This provides an empirical basis for the exploration of semantics using child-caregiver interactions

The presented setting for zero-shot learning did not work

- Inconsistency between training corpora?
- Limited amount of data?
- Inadequate learning function?
- Too many target words?

Further challenges

- Explore this setting further
- Use more refined visual attributes
- Use different and hopefully more informative linguistic contexts
- Implement an incremental approach
- Implement fast mapping

References

- Bruni, E., Tran, N. K., & Baroni, M. (2014). Multimodal distributional semantics. Journal of Artificial Intelligence Research, 49, 1-47.
- Devereux, B.J., Tyler, L.K., Geertzen, J., Randall, B. (in press) The Centre for Speech, Language and the Brain (CSLB) Concept Property Norms. Behavior Research Methods.
- Firth, J.R. (1957). Papers in Linguistics, 1934-1951. Oxford University Press, Oxford, UK.
- Harris, Z. (1954). Distributional structure. Word, 10(2-3), 1456-1162.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological review, 104(2), 211.
- Lazaridou, A., Bruni, E., & Baroni, M. (2014). Is this a wampimuk? Cross-modal mapping between distributional semantics and the visual world.
- Turney, P. D., Neuman, Y., Assaf, D., & Cohen, Y. (2011). Literal and metaphorical sense identification through concrete and abstract context. In Proceedings of the 2011 Conference on the Empirical Methods in Natural Language Processing (pp. 680-690).
- Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. Language and Cognition, 1(2), 219-247

Resources

- VSEM: http://clic.cimec.unitn.it/vsem/
- DISSECT: http://clic.cimec.unitn.it/composes/toolkit/

- CHILDES: http://childes.psy.cmu.edu
- ImageNet: http://www.image-net.org
- UkWac: http://wacky.sslmit.unibo.it/doku.php?id=corpora
- MEN dataset: http://clic.cimec.unitn.it/~elia.bruni/MEN.html
- CSLB Norms: https://csl.psychol.cam.ac.uk/propertynorms/
- Simple Wikipedia: http://simple.wikipedia.org/wiki/Main_Page
- Concreteness Scores: contact author

Thank you!

And thanks to Elia Bruni and Angeliki Lazaridou for the precious help they gave me and the tools they developed.

- Questions?
- Suggestions?
- Criticisms?

Extra #1: DSM parameters

- We used content words as dimensions, making sure that they were present in all the corpora used for each experiment and applying different frequency thresholds; thus, each experiment has a different set of target and context words.
- We used a window of 20 words l/r, without considering sentence boundaries
- Raw frequencies were weighted using PLMI
- The dimensionality of the DSMs was reduced to 300 using SVD
- For the experiment with CSLB norms, we considered each concept~norm pair as a co-occurrence, and its production frequency as its frequency of occurrence.

Extra #2: Visual Features extraction

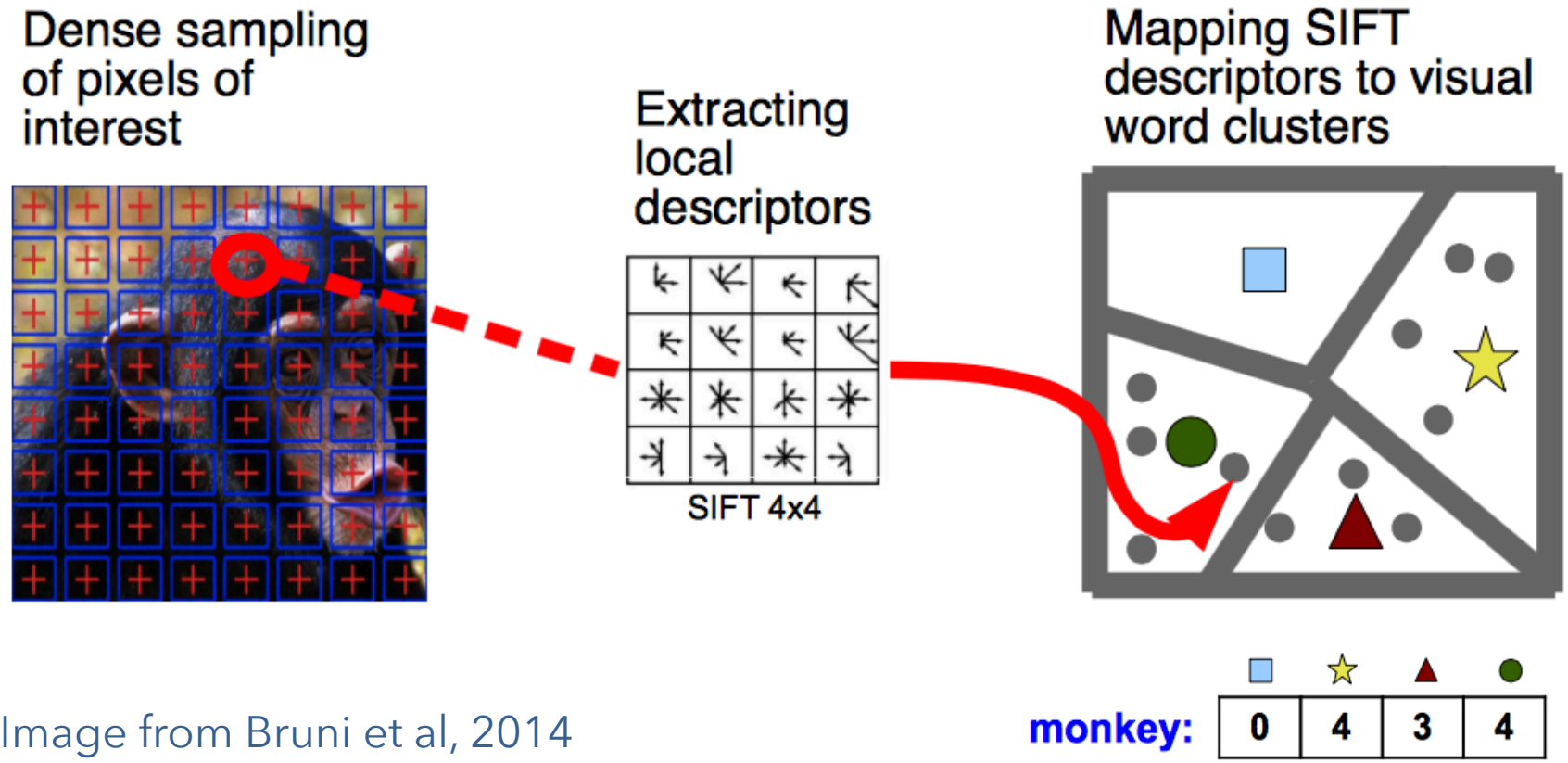


Image from Bruni et al, 2014