Concepts, Composition, and Conversational Coordination

Semantic Competence for Situated Interaction

David Schlangen
University of Potsdam, Germany

http://clp.ling.uni-potsdam.de

https://github.com/davidschlangen/cosine-paris

plan

today:

- where this is coming from
- sketch of the proposal, approach, data

whole seminar series:

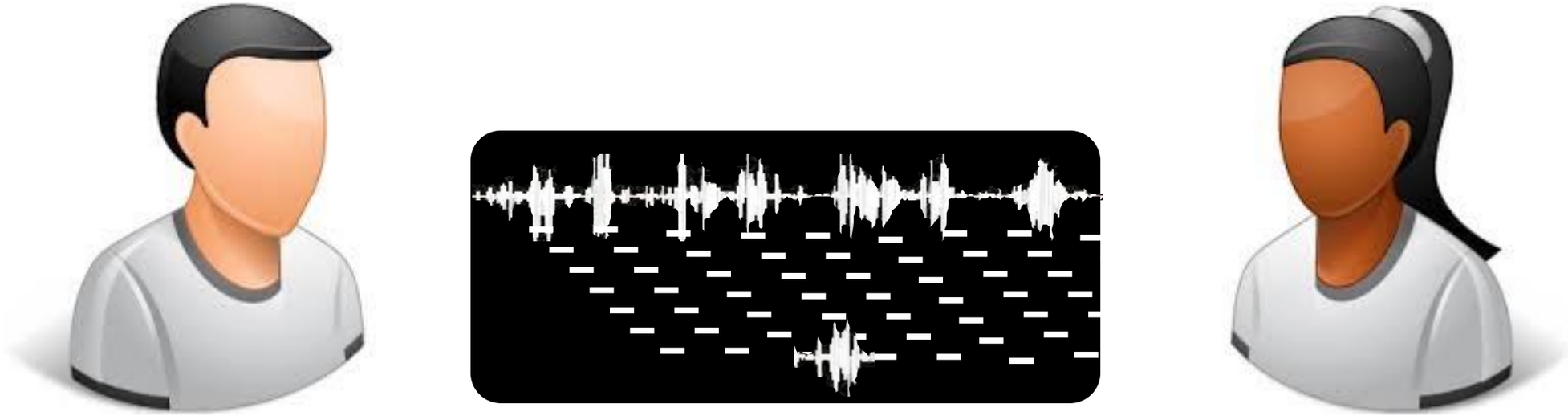
- intro: the problem & the approach
- concepts [Mon, Sep 23]
- composition [Mon, Sep 30]
- conversational coordination / dialogue [Mon, Oct 7]

Concepts, Composition, and Conversational Coordination

Semantic Competence for Situated Interaction

situated interaction

Interaction between participants that perceive themselves to currently be in the same situation.



The „incremental units“ (IU) model.
(Schlangen & Skantze 2009, 2011)

situated interaction

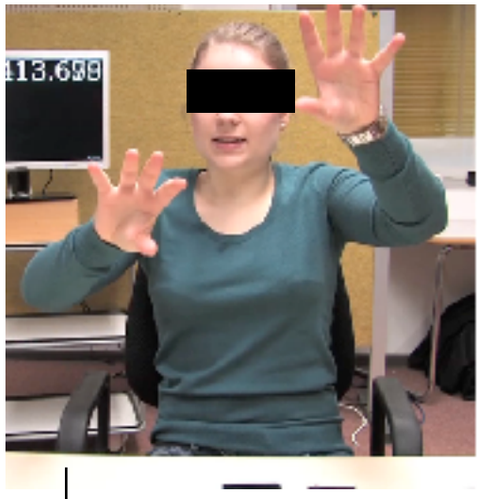
Interaction between participants that perceive themselves to currently be in the same situation.

situated interaction

Interaction between participants that perceive themselves to currently be in the same situation.

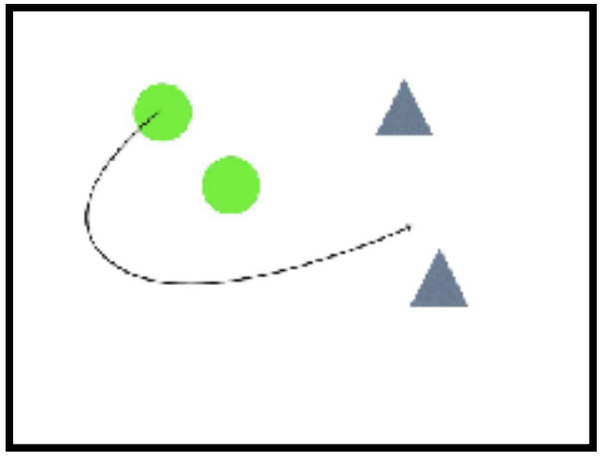
hier ist ein graues Dreieck
here is a gray triangle

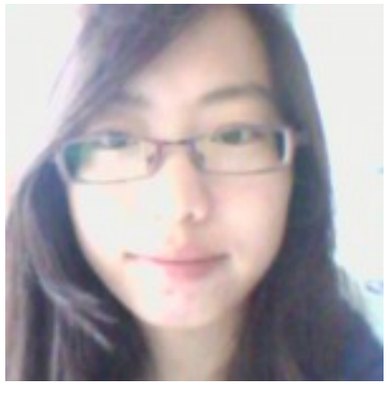
und hier ist ein grüner Kreis
and here is a green circle



hier ist noch ein grüner Kreis
here is another green circle

und hier ist noch ein graues Dreieck
and here is another gray triangle

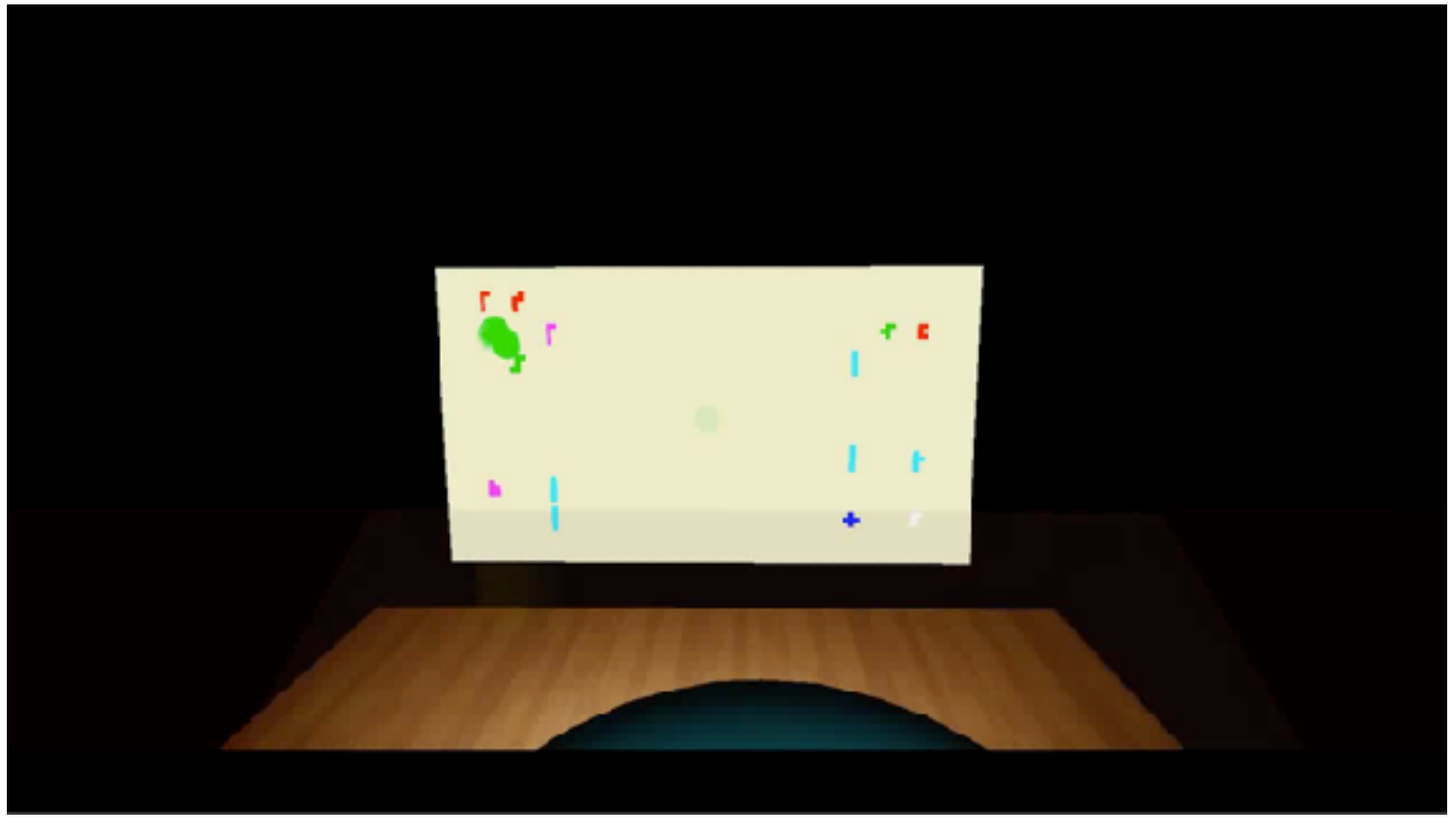




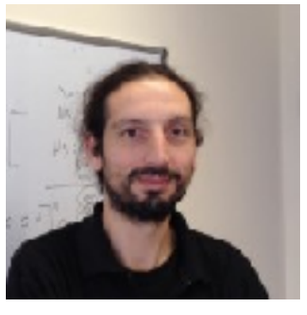
Ting Han, PhD Bielefeld 2018

situated interaction

Interaction between participants that perceive themselves to currently be in the same situation.



(Kennington *et al.*, SIGdial 2013, 2014)
(Kousidis *et al.*, SIGdial 2013, 2014)



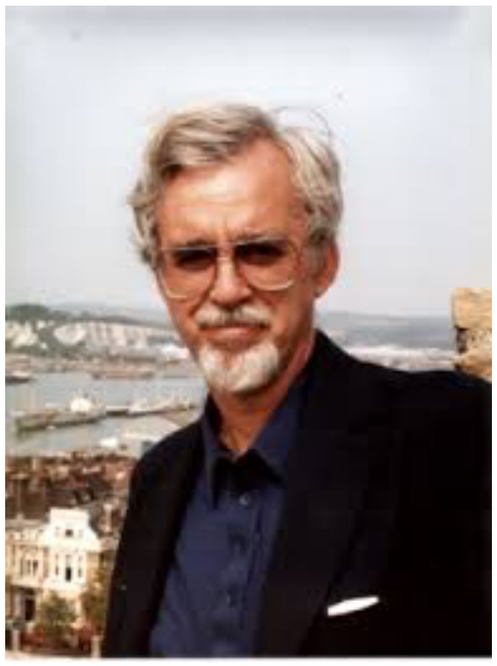
situated interaction

Interaction between participants that perceive themselves to currently be in the same situation.

Sina Zarrieß, Casey Kennington
(Schlangen *et al.* ACL 2016)
& see website

“Simple statements about people, things, places and times are the bedrock of language. They are the basic units of conversation, of literature, of the language of practical affairs.”

–John Perry (2012)

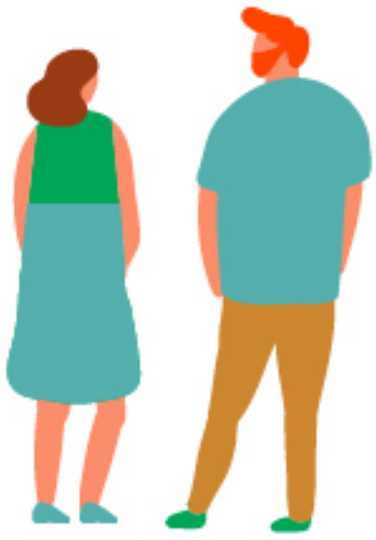


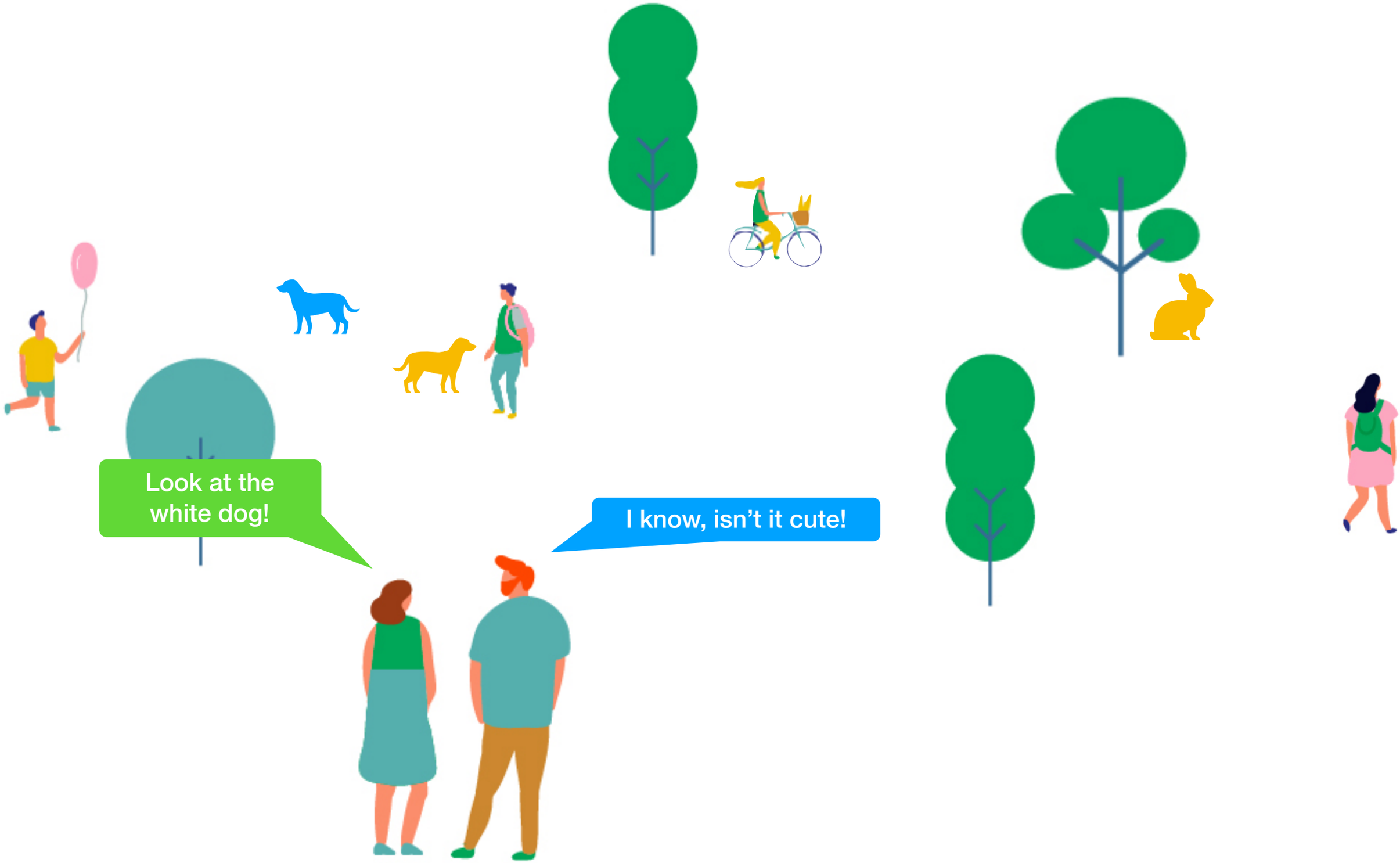
today

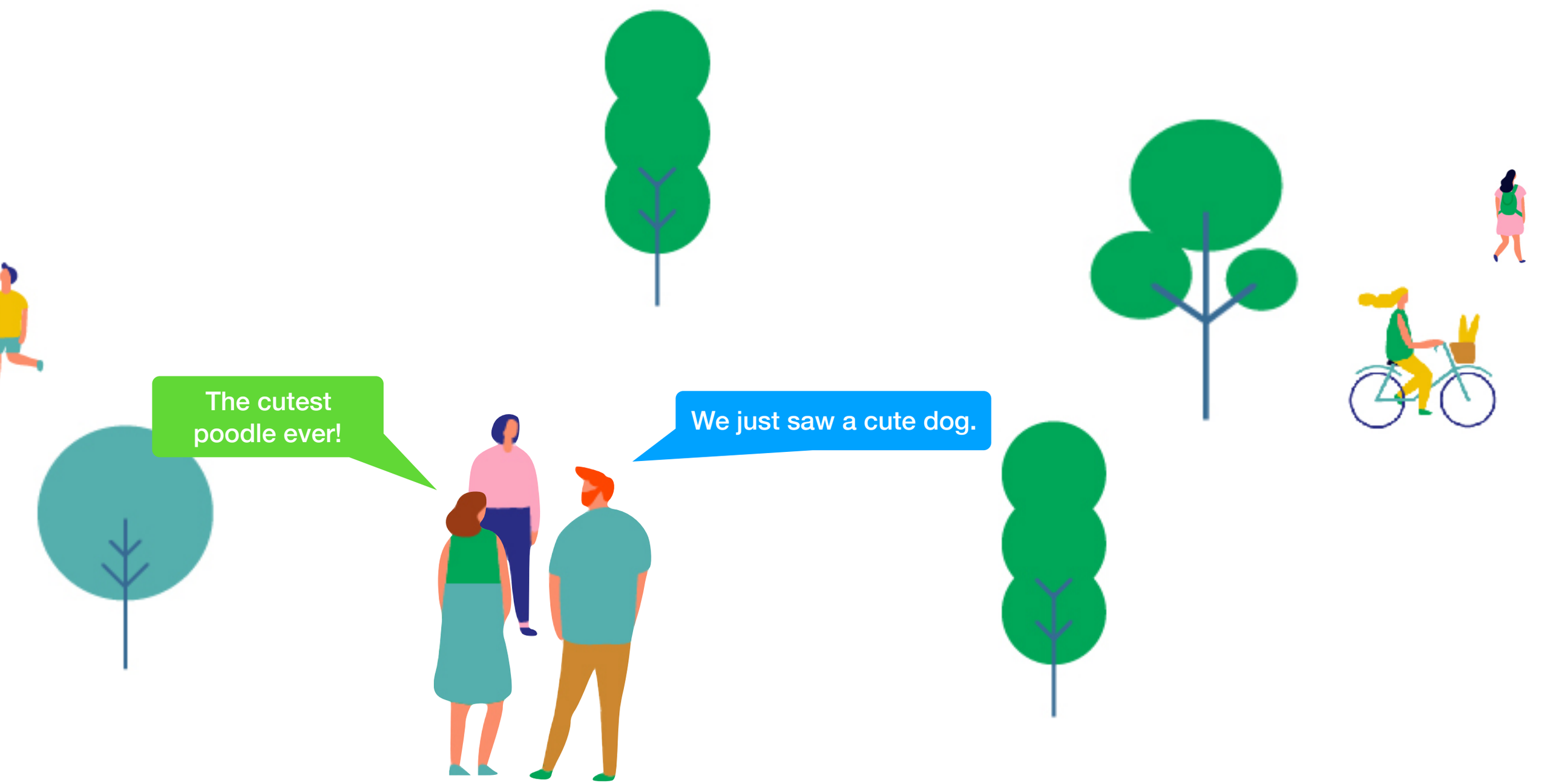
- wider context
- the explanandum, in the form of a story
- constraints on the explanation
- sketch of the proposal
- empirical evidence / data
- why not deep end-to-end learning?

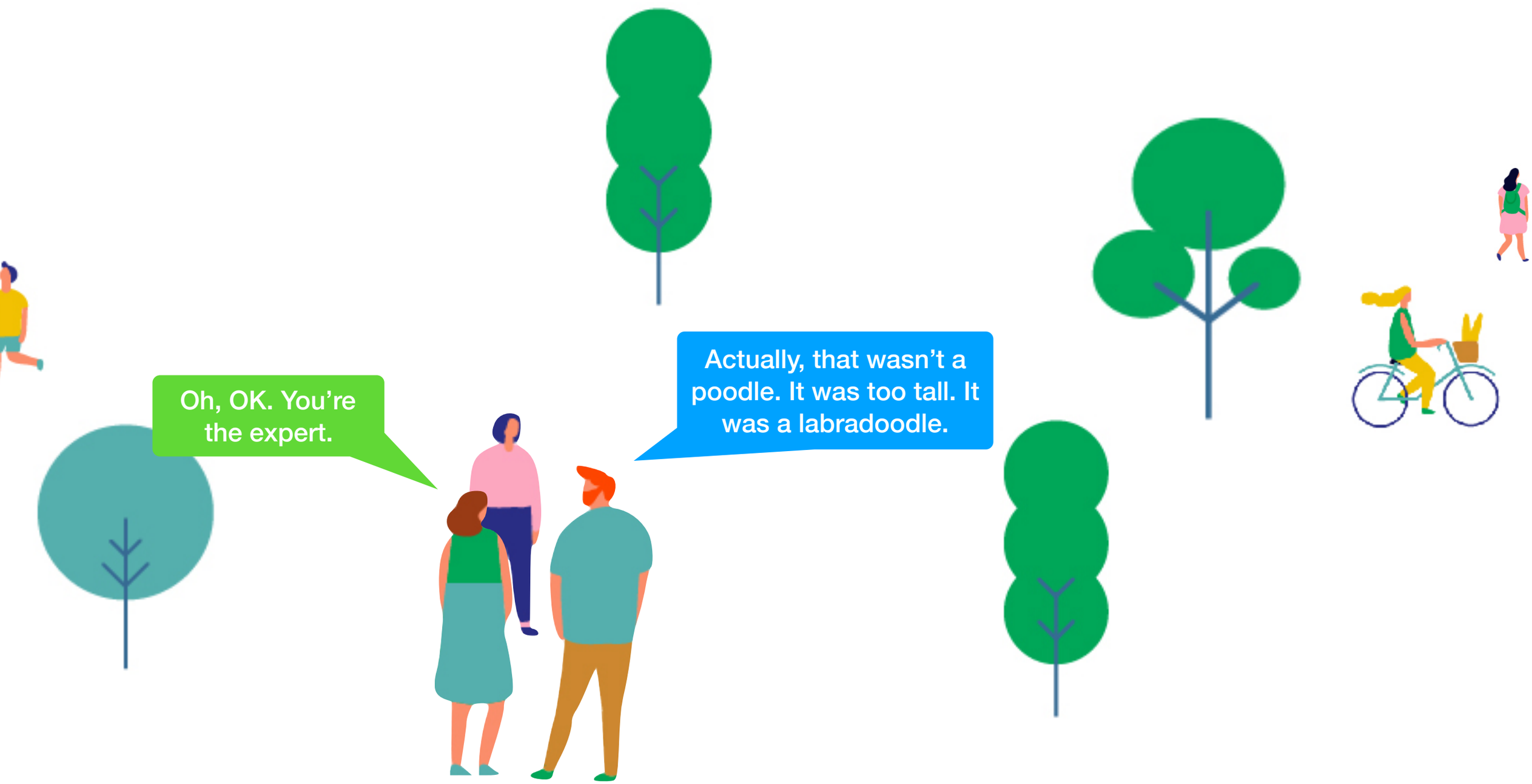
today

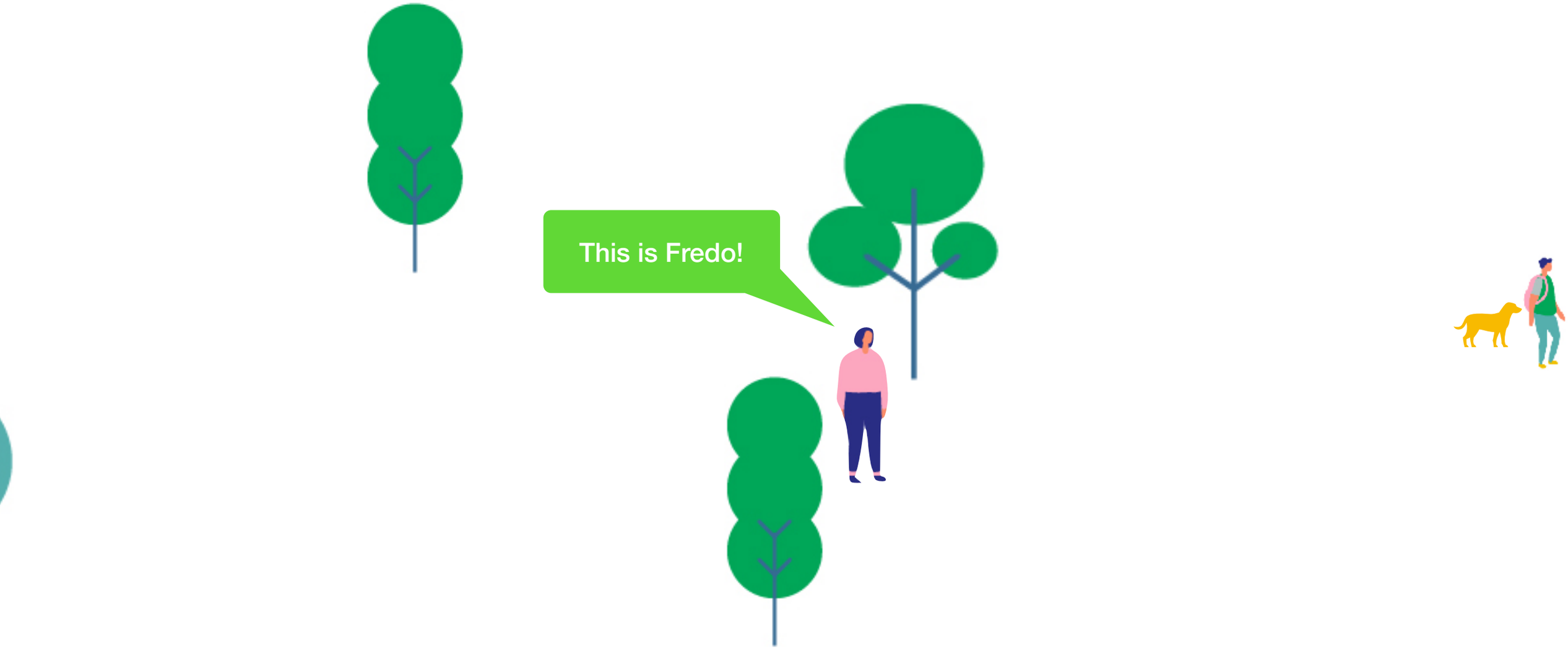
- wider context
- the explanandum, in the form of a story
- constraints on the explanation
- sketch of the proposal
- empirical evidence / data
- why not deep end-to-end learning?




Look at the white dog!
I know, isn’t it cute!

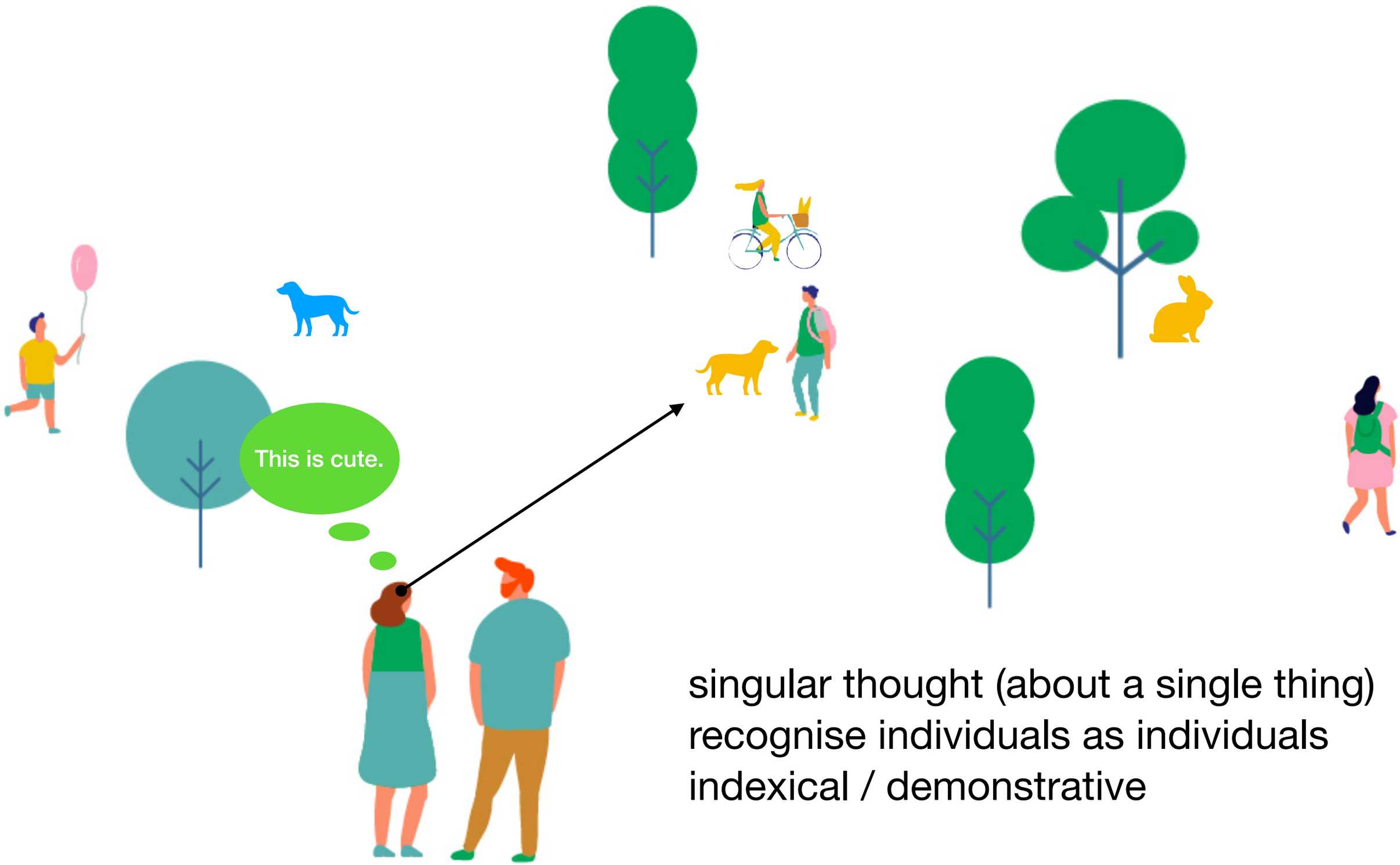

The cutest poodle ever!
We just saw a cute dog.


Oh, OK. You're the expert.
Actually, that wasn't a poodle. It was too tall. It was a labradoodle.

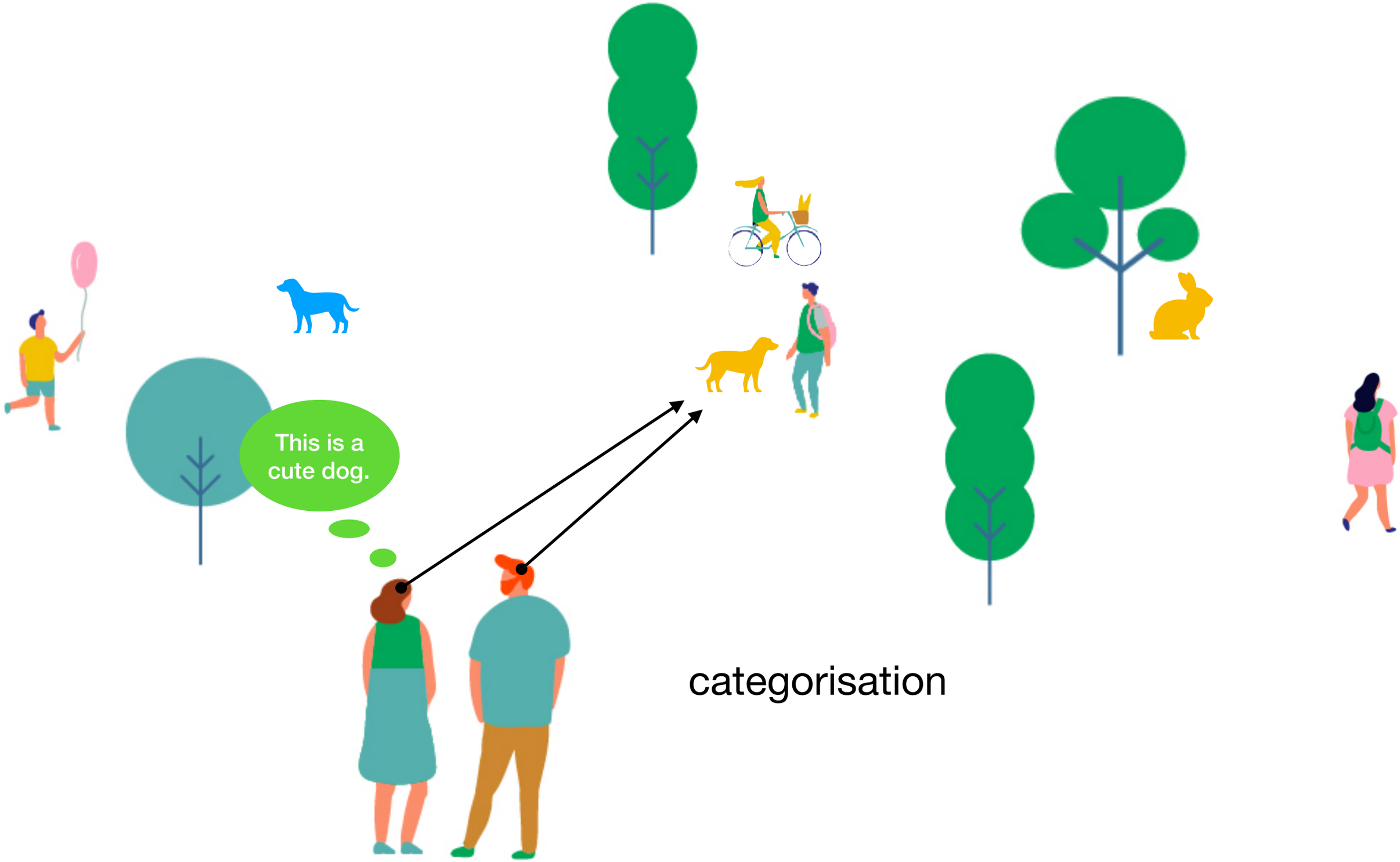

This is Fredo!

rewind!


Look at the white dog!



singular thought (about a single thing)
recognise individuals as individuals
indexical / demonstrative

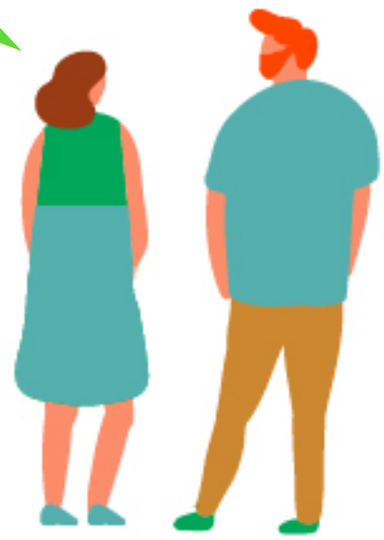


categorisation

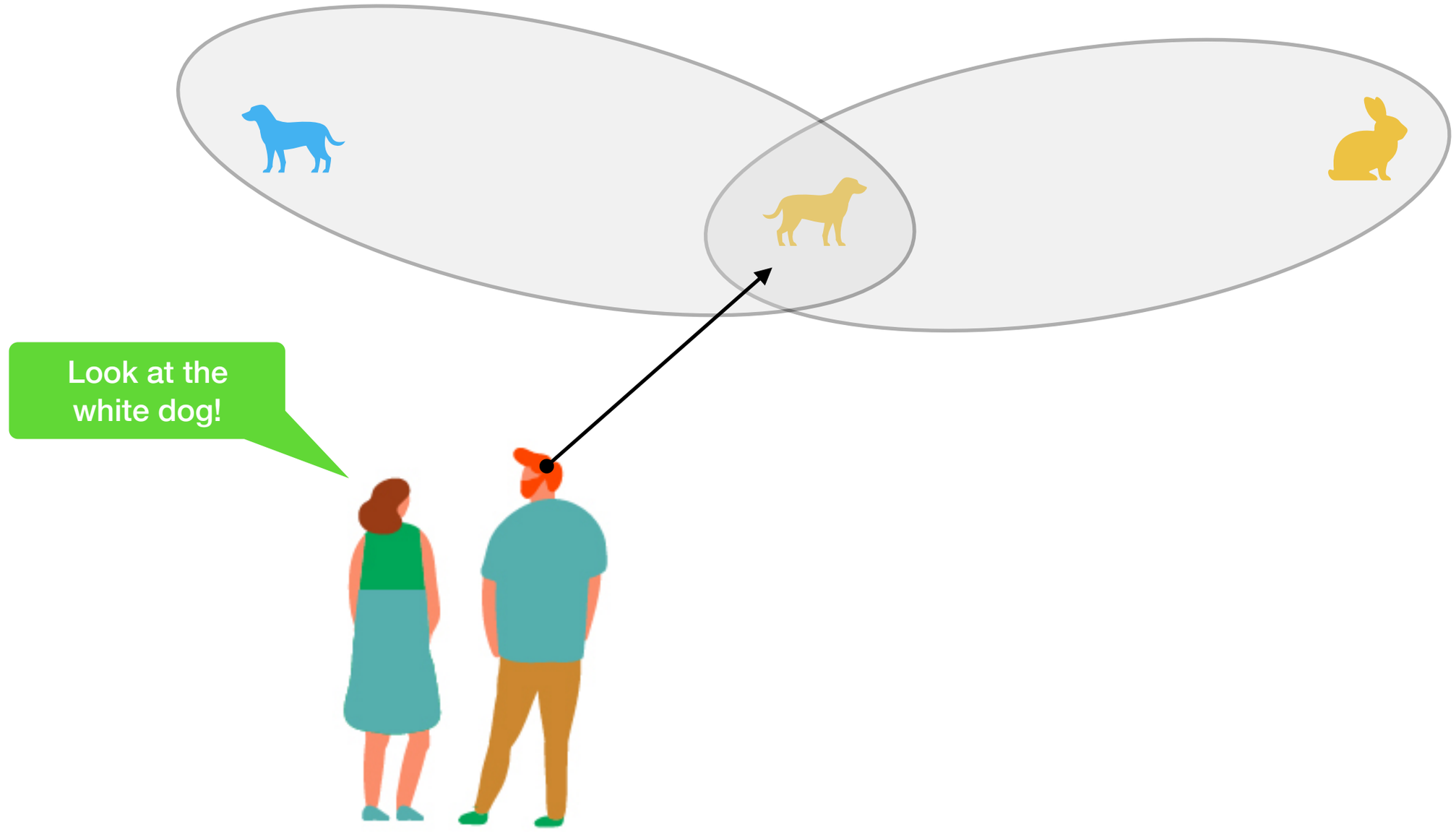
animal!

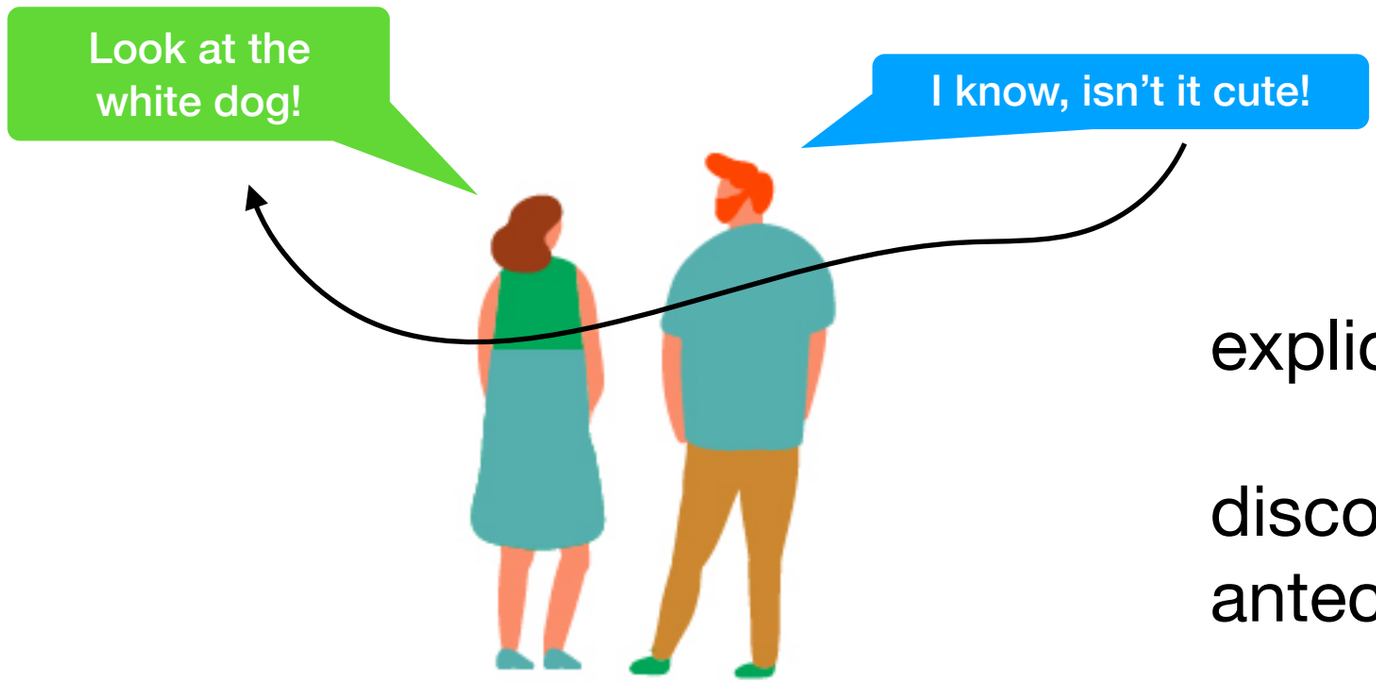
dog!

white!



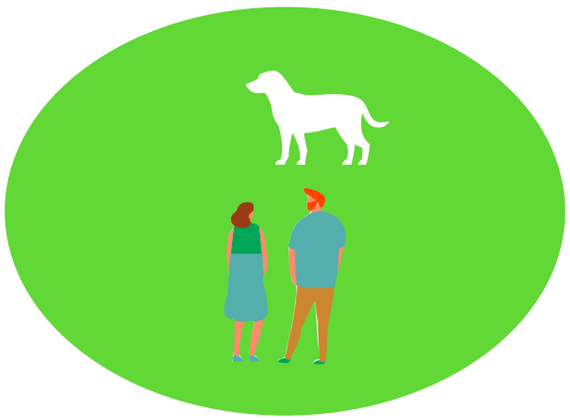
Look at the white dog!


Look at the white dog!



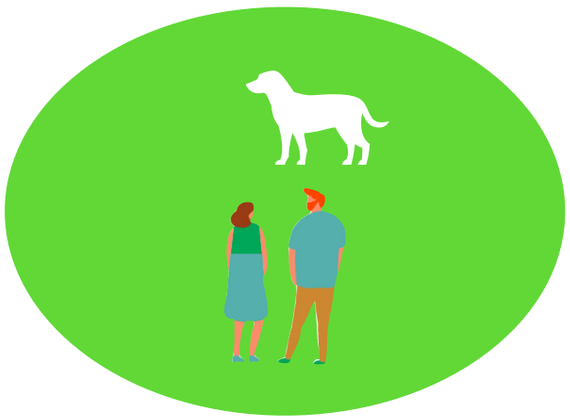
explicit grounding / uptake

discourse reference (“it” needs antecedent in discourse)

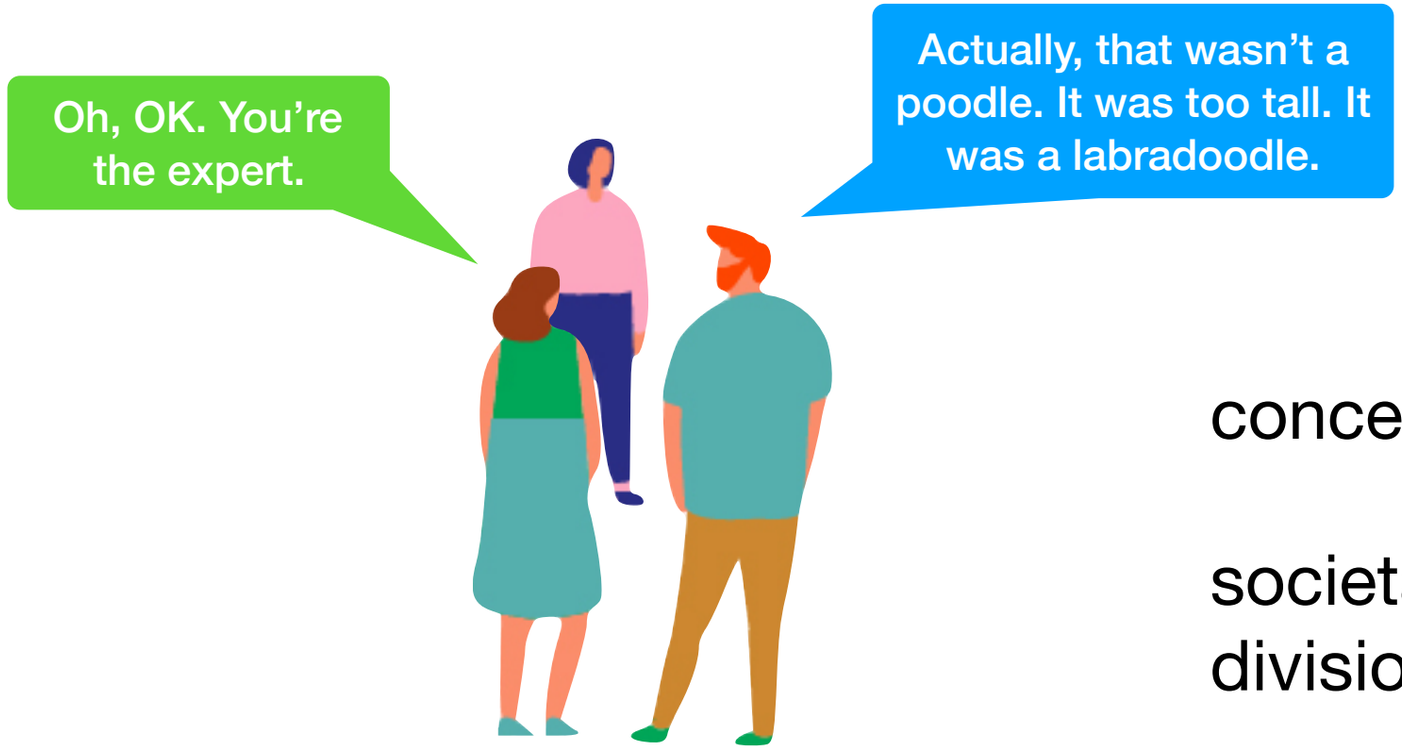


displacement

discourse reference (“a cute dog”)



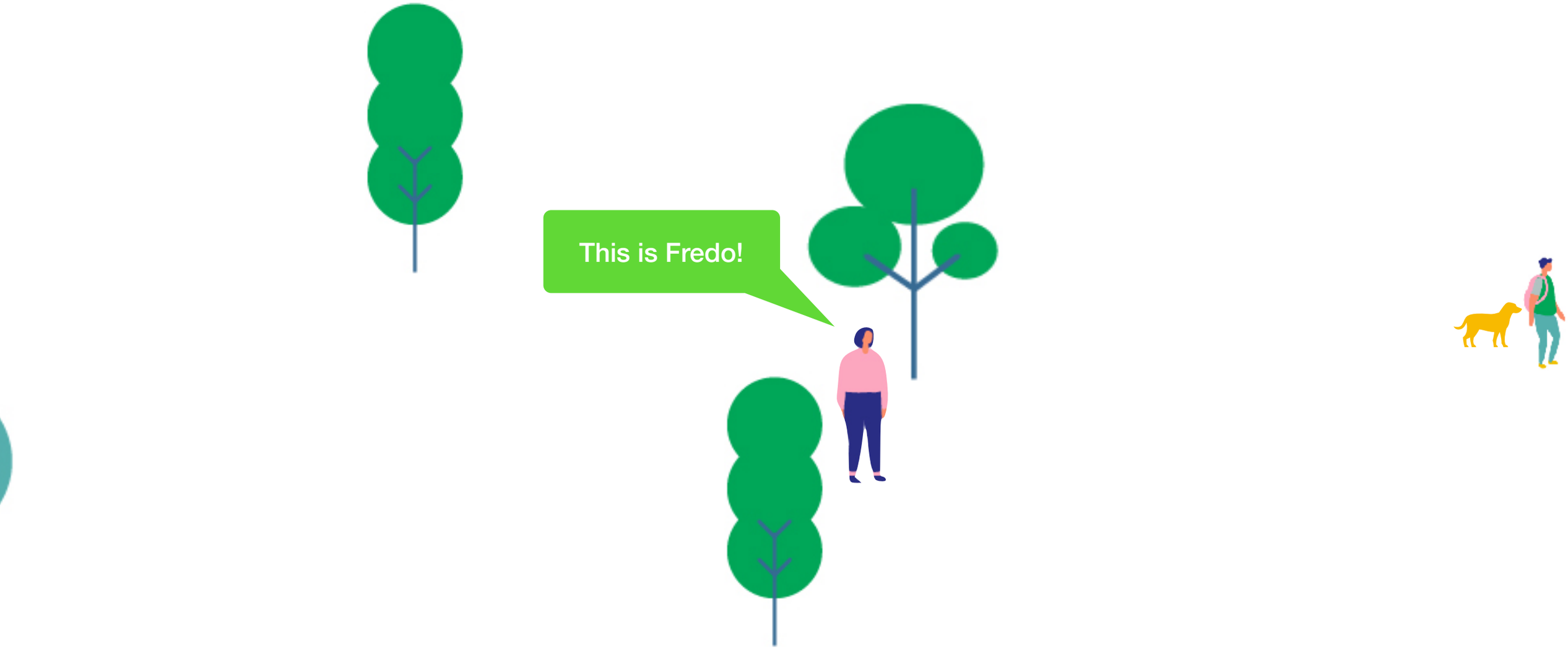
discourse co-reference, with inference required



concept adaptation

societal grounding, linguistic division of labour

zero-shot learning (for C)



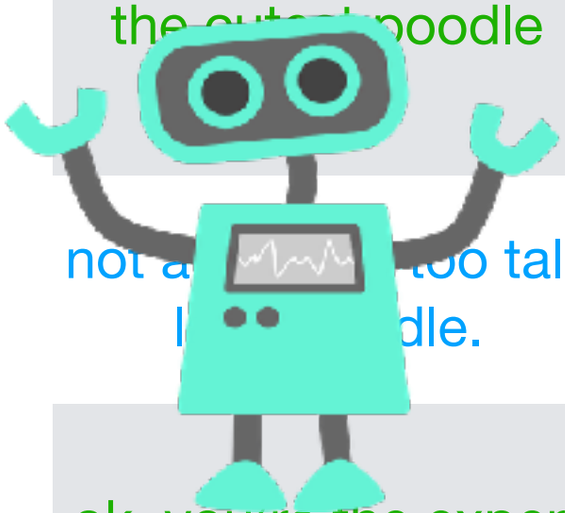
unification of mental files

utterance	Ann	Bert	Chris
look at the white dog!	• have thought about object • categorise obj. • … helpful to B	• categorise objects to resolve reference	
isn't it cute	• resolve reference to disc. referent	• generate discourse reference (discourse old)	
we saw a cute dog	• resolve to memory, & previous disc.	• generate discourse reference (discourse new)	• create new discourse referent
the cutest poodle ever	• discourse old • categorisation		• resolve discourse reference, bridging inference
not a poodle. too tall. labradoodle.		• re-categorisation • justification	
ok, you're the expert	• update concept • rely on authority		• new concept "labradoodle"

utterance	Ann	Bert	Chris
look at the white dog!	• have thought about object • c • …	• categorise objects to resolve	
isn't it cute	• re to		
we saw a cute dog			nt
the cutest poodle ever	• d • c		e ng
not a poodle. too tall. labradoodle.			
ok, you're the expert	• update concept • rely on authority		• new concept "labradoodle"

- Concepts (dog, poodle, labradoodle, cute, white, tall)
- Composition ("white dog")
- Coordination (uptake; "not a poodle." "ok")

utterance	Ann	Bert	Chris
look at the white dog!	• have thought about object • c • ..	• categorise objects to resolve	
isn't it cute	• re to		
we saw a cute dog			nt
the cute poodle	• d • c		e ng
not a too tall. l dle.			
ok, you're the expert	• update concept • rely on authority		• new concept "labradoodle"



https://publicdomainvectors.org/en/free-clipart/Green-robot/38871.html

Explananda

- how do words link to concepts, and concepts to the world and to each other?
- how do expressions compose concepts?
- how do dialogue participants coordinate on this?

today

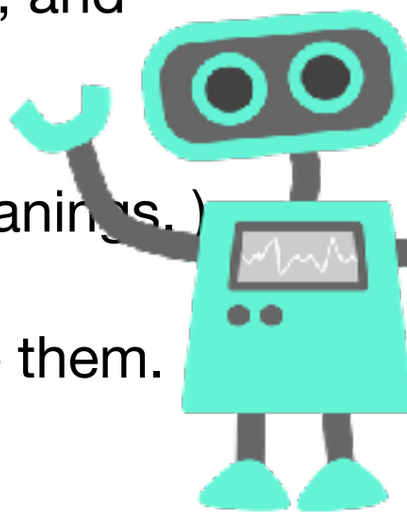
- wider context
- the explanandum, in the form of a story
- constraints on the explanation
- sketch of the proposal
- empirical evidence / data
- why not deep end-to-end learning?

today

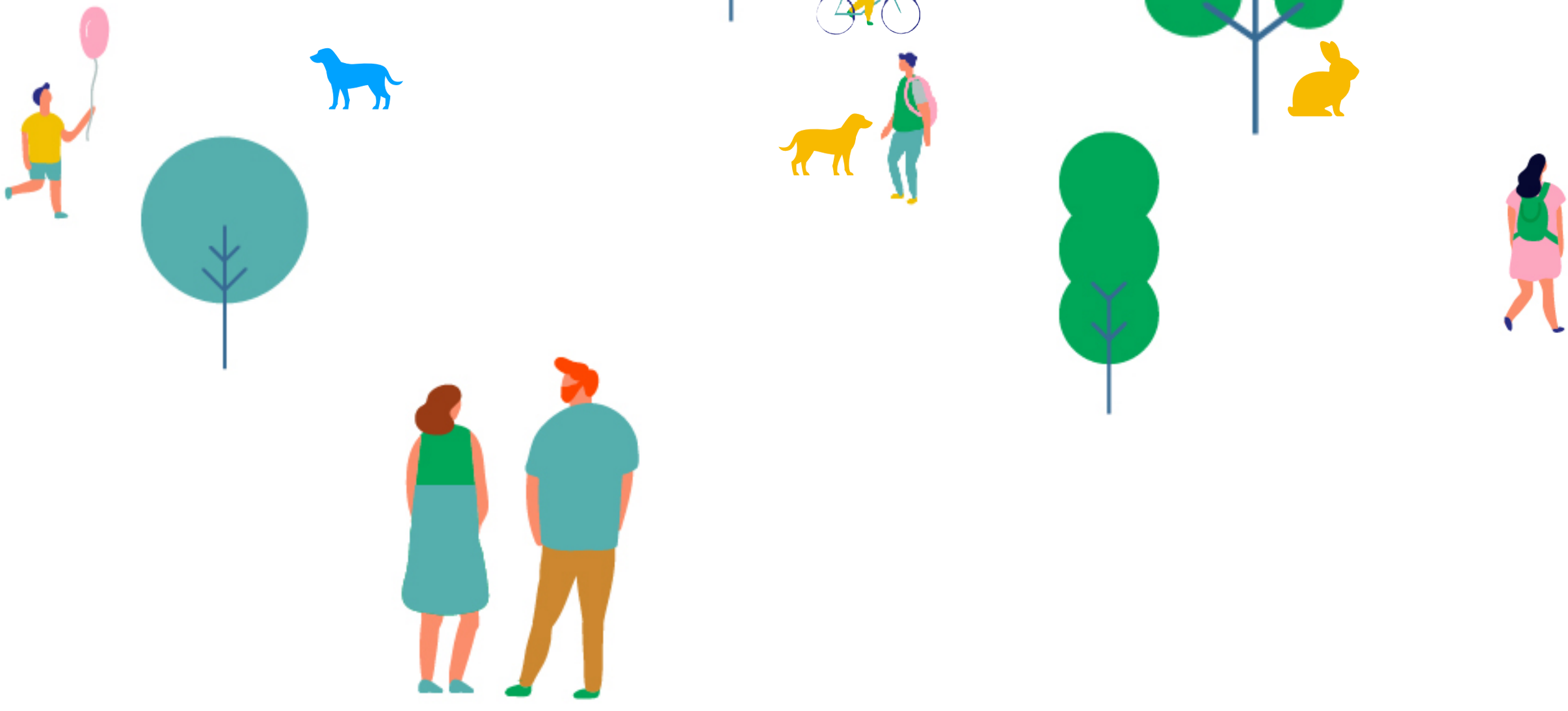
- wider context
- the explanandum, in the form of a story
- constraints on the explanation
- sketch of the proposal
- empirical evidence / data
- why not deep end-to-end learning?

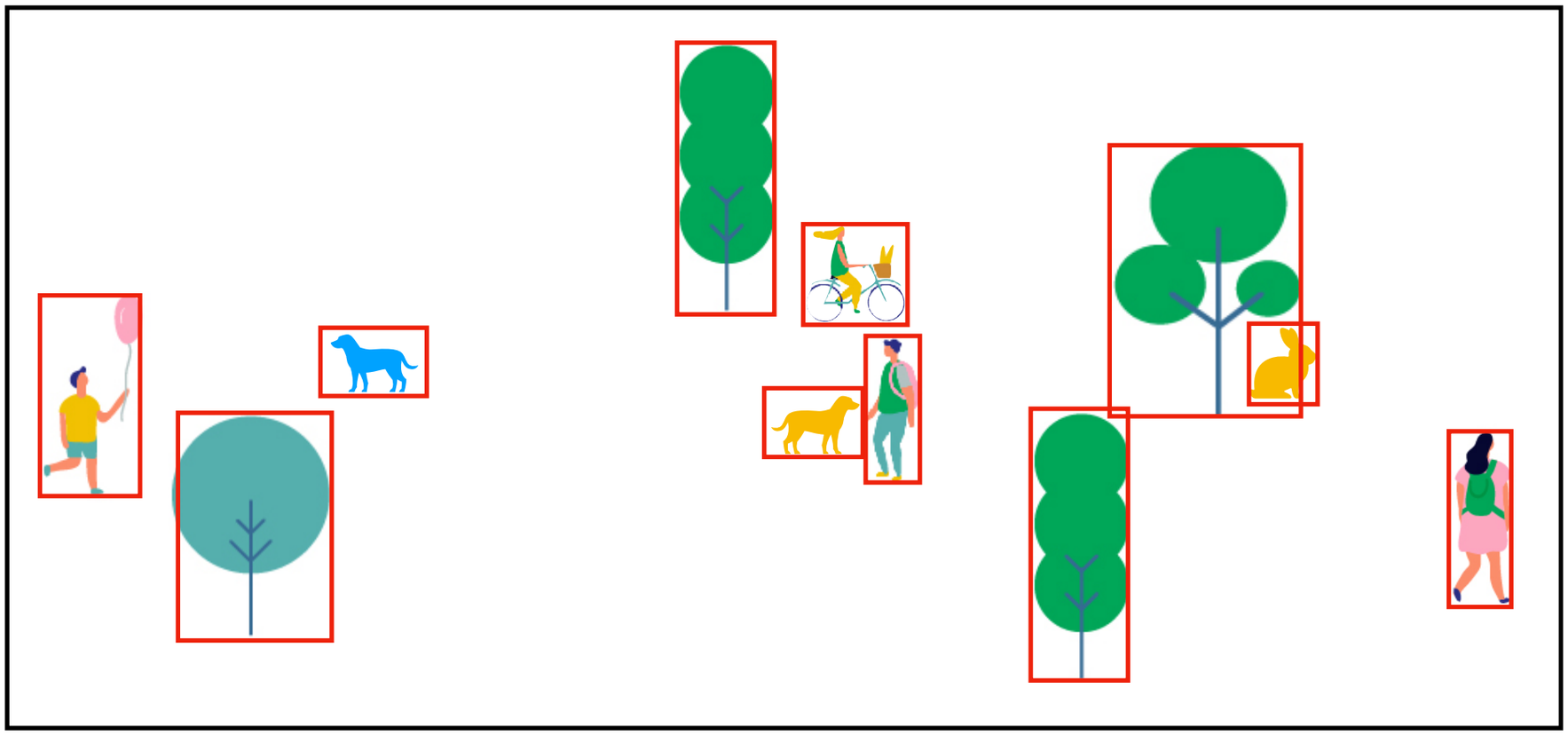
constraints on the explanation, starting points, methodology

- “In order to say what a meaning *is*, we may first ask what a meaning *does*, and then find something that does that.” (Lewis 1970: 22)
- Something has to be in the head. (Concepts, as mental reprs of word meanings.)
- There are objects out there, and the agent can (normally) visually separate them.



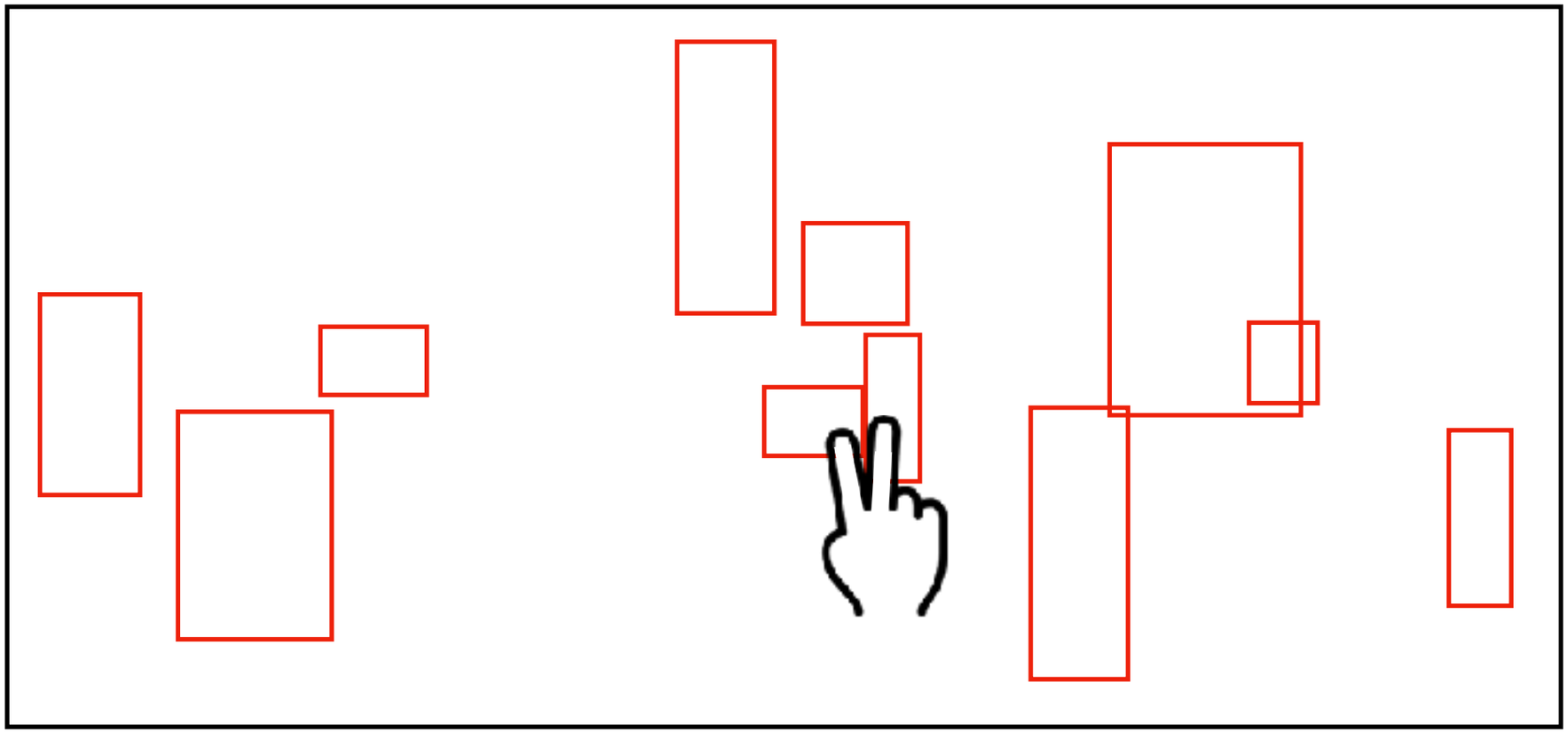
https://publicdomainvectors.org/en/free-clipart/Green-robot/38871.html





"Spelke objects" (Bloom 2000), (e.g. Spelke 1994):

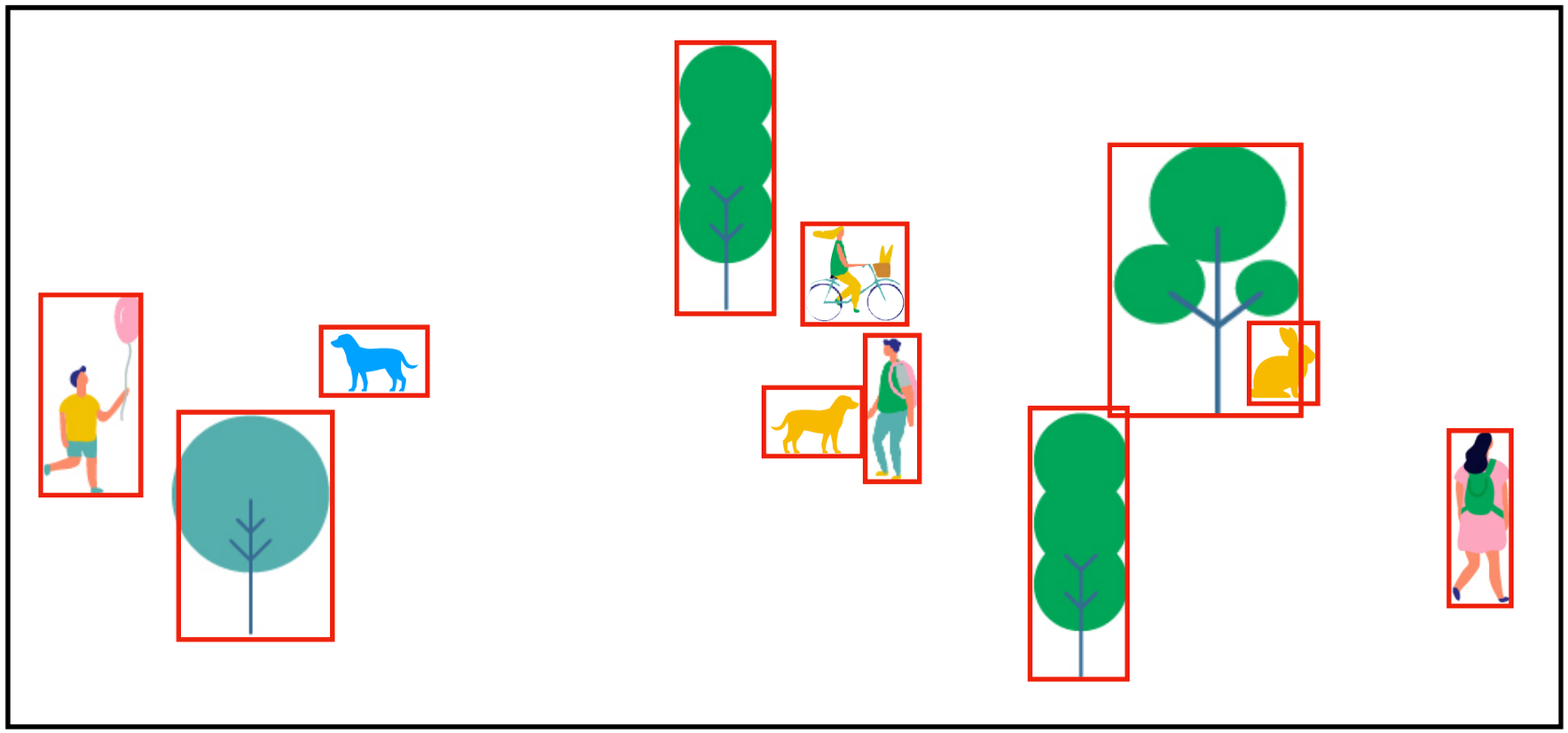
- *principle of cohesion*: if you pull one part, the rest follows
- *principle of continuity*: continuous pathway through space
- *principle of solidity*: don't move through each other
- *principle of contact*: inanimate obj. move only when pushed



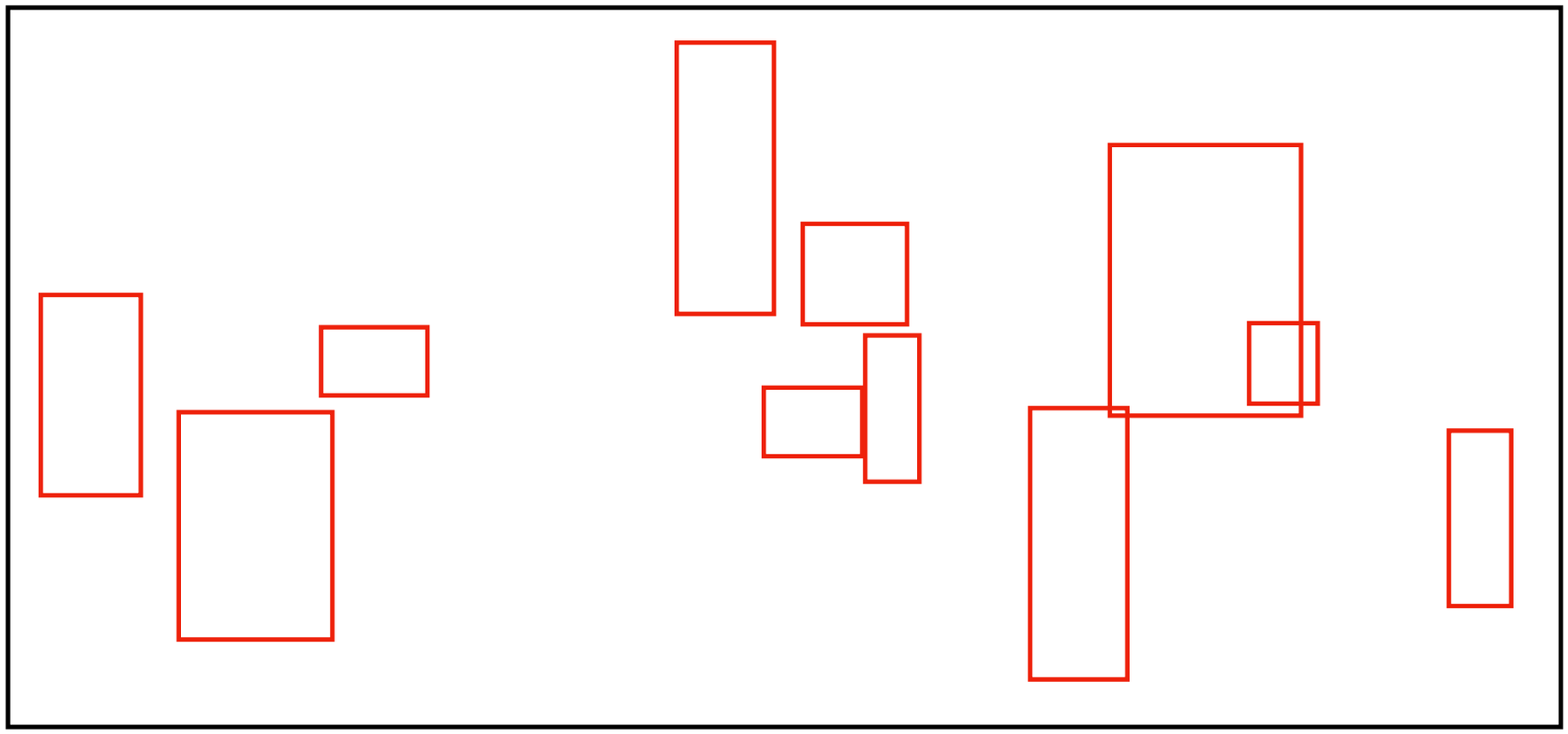
“Fingers of Instantiation” (FINST), Pylyshyn 1989, 2007

- pre-attential tracking and indexing mechanism
- pre-conceptual: location, not type

Fodor & Pylyshyn 2015



Situation representation, assumed to be shared

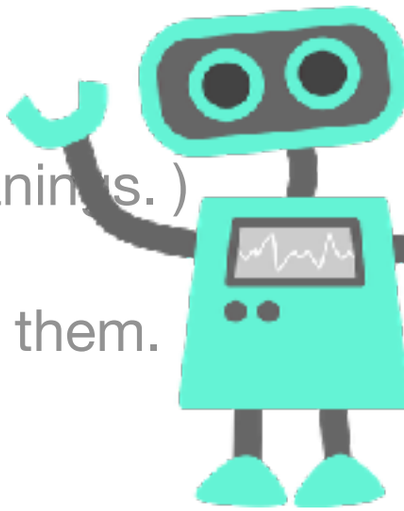


Discourse representation, coordinated on to make it shared

x
$dog(x)$ $white(x)$

constraints on the explanation, starting points, methodology

- “In order to say what a meaning is, we may first ask what a meaning does, and then find something that does that.” (Lewis 1970: 22)
- Something has to be in the head (Concepts, as mental reprs of word meanings.)
- There are objects out there, and the agent can (normally) visually separate them.



- There are discourses out there, with referents for things, places, times, …
- Concepts get in the head …
  - … through interaction with the world and with speaker,
  - … not all at once,
  - … in various ways,
  - … and generally fairly quickly

https://publicdomainvectors.org/en/free-clipart/Green-robot/38871.html

learning

- High school graduates know about 60 - 80k words. (Aitchinson 1994)
- Learning new words / concepts doesn’t stop then. New concepts are constructed, new words are coined, unfamiliar words are encountered.

Even if you don’t want to model the dynamics of this (slower start, later acceleration), can’t do batch learning. Must be *incremental learning*.

learning

There are multiple ways of learning about the meaning of a word:

- ostensive definition (labelled examples)
  Augustine (400): “"When they (my elders) named some object, and accordingly moved towards something, I saw this and I grasped that the thing was called by the sound they uttered when they meant to point it out. Their intention was shewn by their bodily movements, as it were the natural language of all peoples […]”

- explicit definition
  father to child: “when they have fingers in them they are called gloves and when the fingers are all put together they are called mittens” (E. Clark 2007; Cooper & Larsson 2009)
  “it’s not a poodle, it’s a labradoodle”

- implicit definition
  from linguistic context

learning

- implicit definition
  from linguistic context

L. Tolstoy, War and Peace:

The little princess went round the table with quick, short, swaying steps, her workbag on her arm, and gaily spreading out her dress sat down on a sofa near the silver ***samovar****, as if all she was doing was a pleasure to herself and to all around her.*

learning

- McDonald & Ramscar 2001

Context A: 'urn'

On his recent holiday in Ghazistan, Joe slipped easily into the customs of the locals. In the hotel restaurant there was a samovar dispensing tea at every table. Guests simply served themselves from the samovar whenever they liked. Joe's table had an elaborately crafted samovar. It was the first earthenware samovar that he had seen.

Context B: 'kettle'

On his recent holiday in Ghazistan, Joe slipped easily into the customs of the locals. His hotel room featured a samovar and a single hob. Each morning Joe boiled water in the samovar for tea. Like others he had seen on his holiday, Joe's samovar was blackened from years of use. He imagined that at some point it would be replaced with an electric samovar.

Figure 1. The *urn*-biased and *kettle*-biased paragraph contexts created for *samovar*.

learning

There are multiple ways of learning about the meaning of a word:

- ostensive definition (labelled examples)
  Augustine (400): “"When they (my elders) named some object, and accordingly moved towards something, I saw this and I grasped that the thing was called by the sound they uttered when they meant to point it out. Their intention was shewn by their bodily movements, as it were the natural language of all peoples […]”

- explicit definition
  father to child: “when they have fingers in them they are called gloves and when the fingers are all put together they are called mittens” (E. Clark 2007; Cooper & Larsson 2009)
  “it’s not a poodle, it’s a labradoodle”

- implicit definition
  from linguistic context

learning

- typically, not many exposures are required, to learn at least an initial underspecified meaning (Carey 1978, Bloom 2000)

- learning can be “cross-modal” (Gottfried *et al.* 1977); “zero-shot”

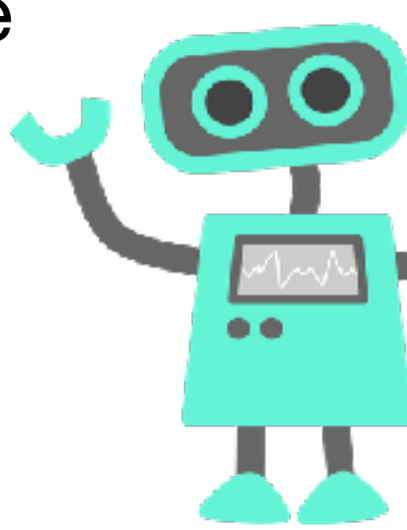
constraints on the explanation, starting points, methodology

- How does this relate to research on human language acquisition, on the mental lexicon, on language processing in general?

  - Doesn’t have to model all idiosyncrasies (learning curves, biases).

  - Can (and should) be inspired by this research.

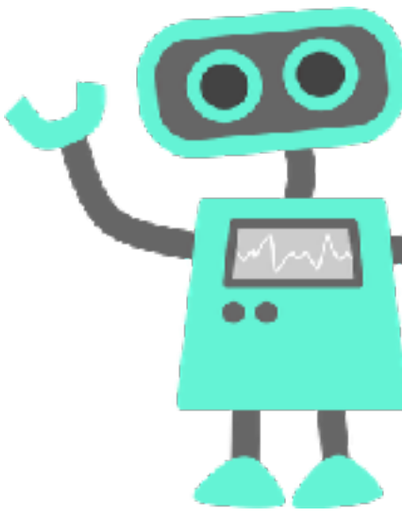
  - Tries to solve the same problem (i.e., is computational model in the sense of Marr 1982), but idealised specification that runs on different OS.



https://publicdomainvectors.org/en/free-clipart/Green-robot/38871.html

constraints on the explanation, starting points, methodology

- To sum up, we want solution (for interpretation / generation of references / inferences) that:
  - is implementable / implemented,
  - takes as input scenes parsed into objects,
  - learns incrementally, from different kinds of situations,
  - allows for explicit modification of word meanings,
  - connects to existing work.



https://publicdomainvectors.org/en/free-clipart/Green-robot/38871.html

today

- wider context
- the explanandum, in the form of a story
- constraints on the explanation
- sketch of the proposal
- empirical evidence / data
- why not deep end-to-end learning?

today

- wider context
- the explanandum, in the form of a story
- constraints on the explanation
- sketch of the proposal
- empirical evidence / data
- why not deep end-to-end learning?

functional

Conceptual Apparatus

Inference

- word to word
- discourse resolution

Reference

- word to world
- categori-sation
- naming / resolution

Sellars (1954): language-entry rules [observations to language], language-language rules [inferences], language-exit rules [orders, promises]

functional

Conceptual Apparatus

Inference

- word to word
- discourse resolution

Reference

- word to world
- categori-sation
- naming / resolution

Partee (1980): “No amount of intralinguistic connections can serve to tie down the extralinguistic content of intensions. For that there must be some language-to-world grounding.”

functional

Conceptual Apparatus

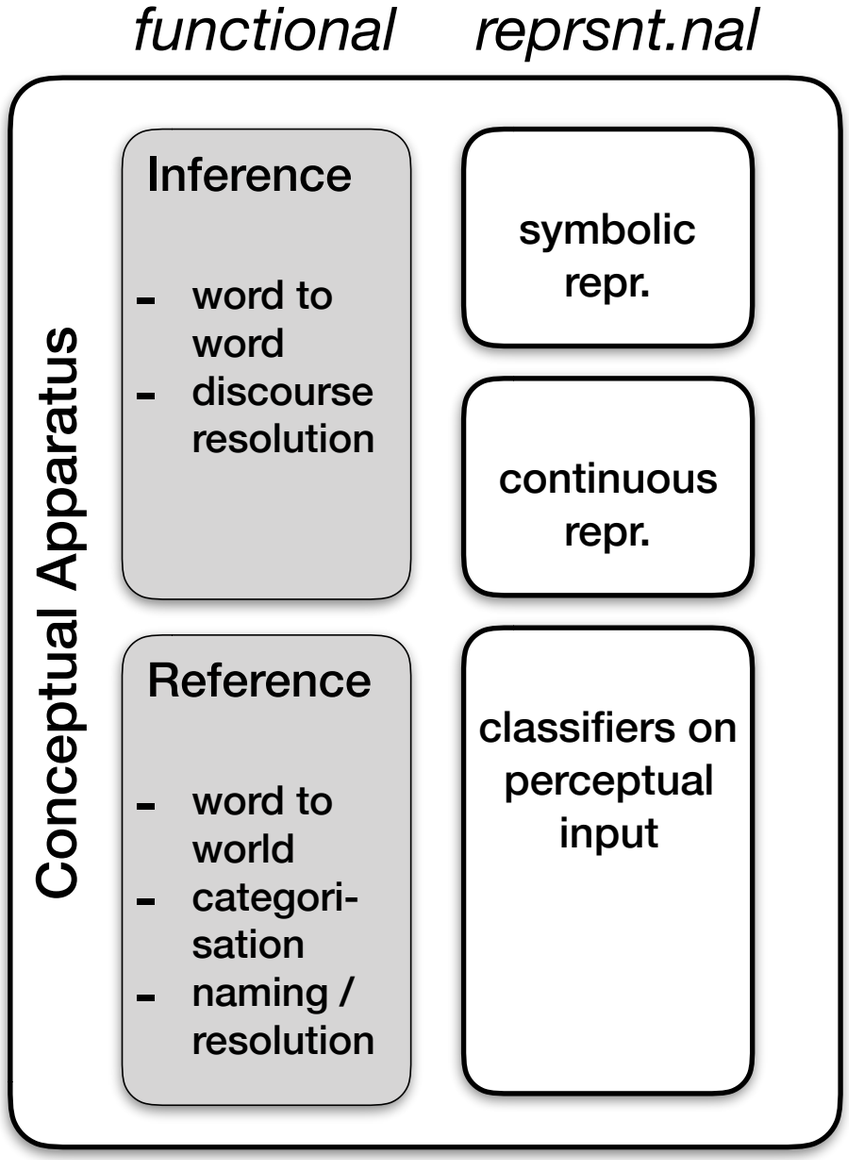
Inference

- word to word
- discourse resolution

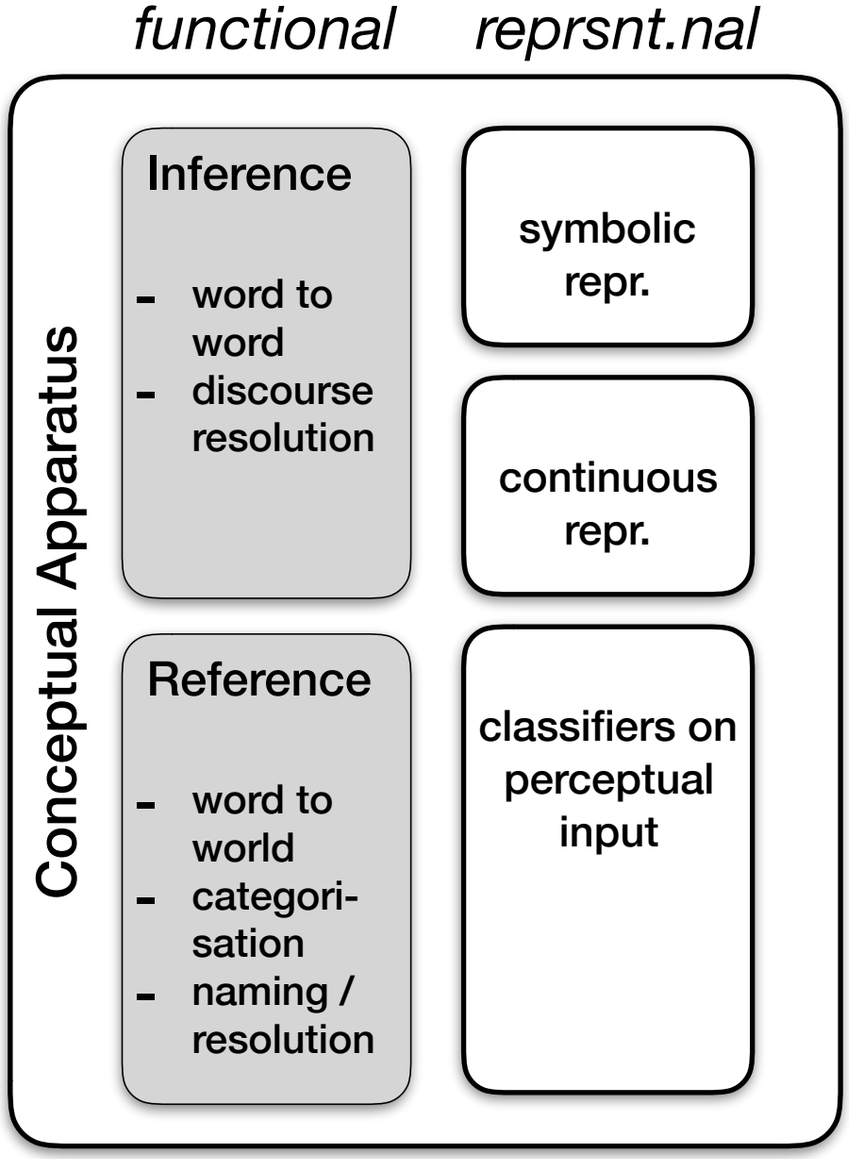
Reference

- word to world
- categori-sation
- naming / resolution

Marconi (1997): Referential and Inferential Lexical Competence



(Harnad 1990), The Symbol Grounding Problem:
"[H]ow can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads?
"[…] invariant features […] that will reliably distinguish a member of a category from any nonmembers […] Let us call the output of this ***category-specific feature detector*** *the* categorical representation.*"*



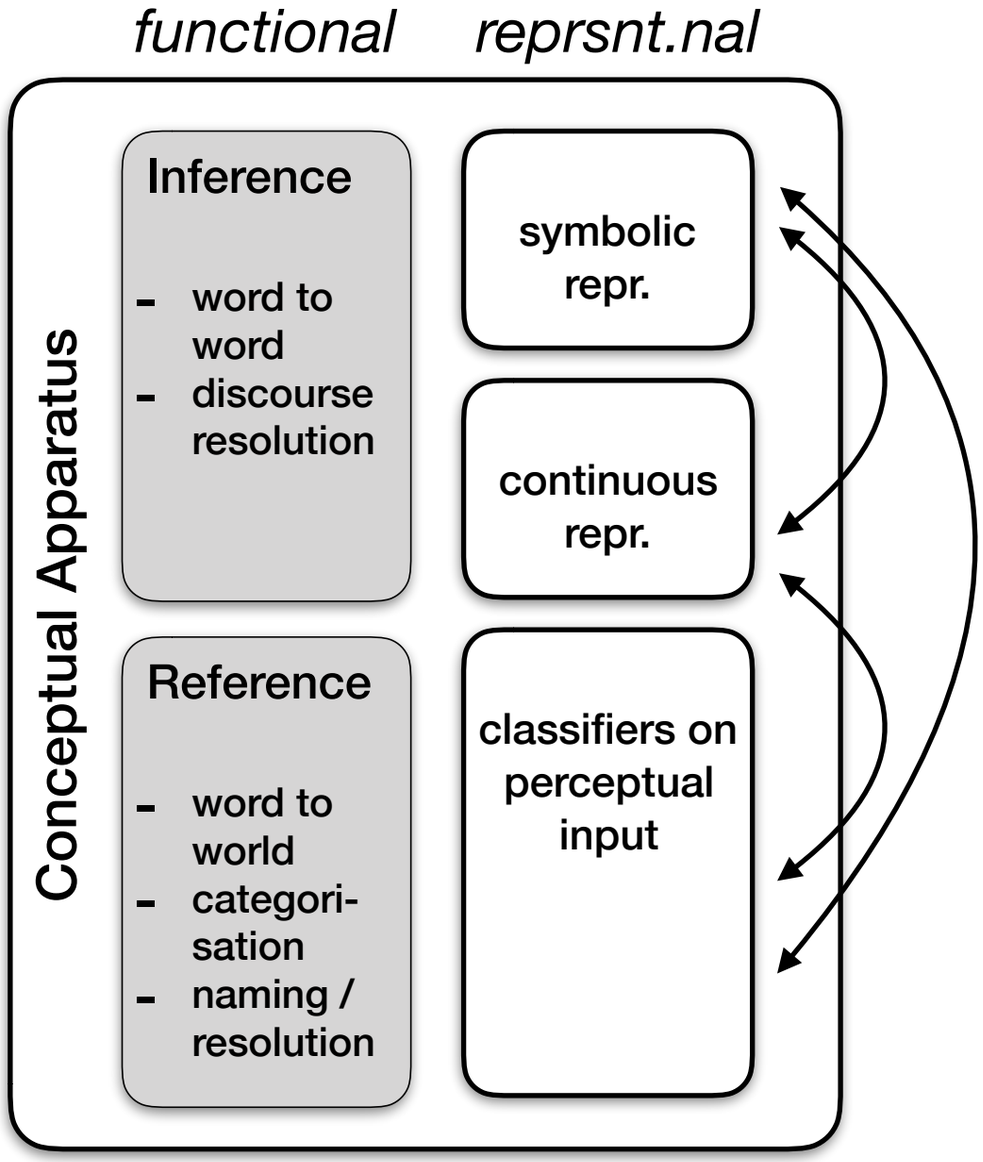
Centre for Linguistic Theory and Studies in Probability (CLaSP), Gothenburg

e.g. Larsson 2015, Formal Semantics for Perceptual Classfication

Staffan Larsson

Robin Cooper

Simon Dobnik


functional
reprsnt.nal
Conceptual Apparatus
Inference
- word to word
- discourse resolution
Reference
- word to world
- categori-sation
- naming / resolution
symbolic repr.
continuous repr.
classifiers on perceptual input

functional *reprsnt.nal* *composition* *coordination*

Conceptual Apparatus

Inference

- word to word
- discourse resolution

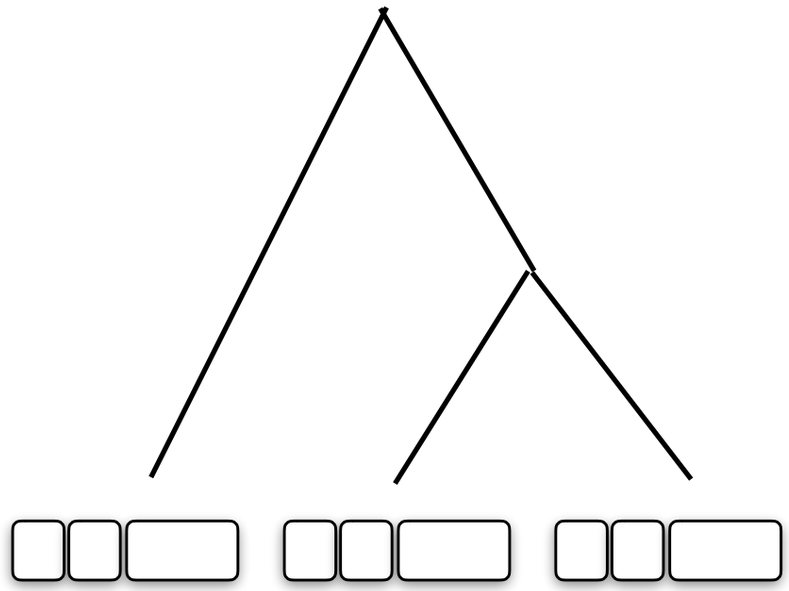
Reference

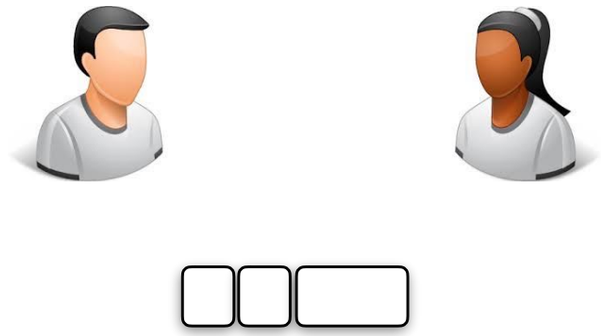
- word to world
- categori-sation
- naming / resolution

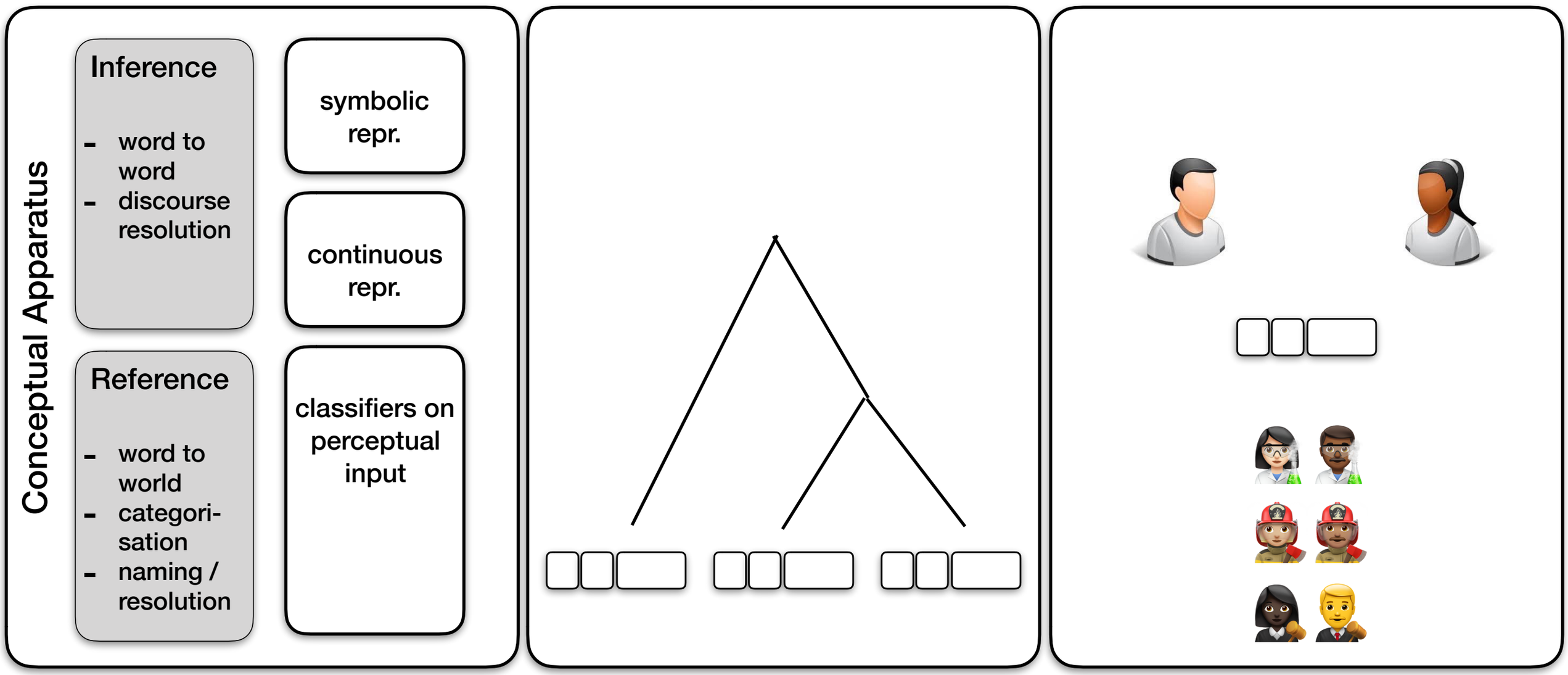
symbolic repr.

continuous repr.

classifiers on perceptual input







- Model word-to-world grounding & word-to-word grounding
- in the service of singular, exophoric reference & of discourse reference.
- Learning: Incremental, diff. sources
- Implemented
- Tested on actual data.

today

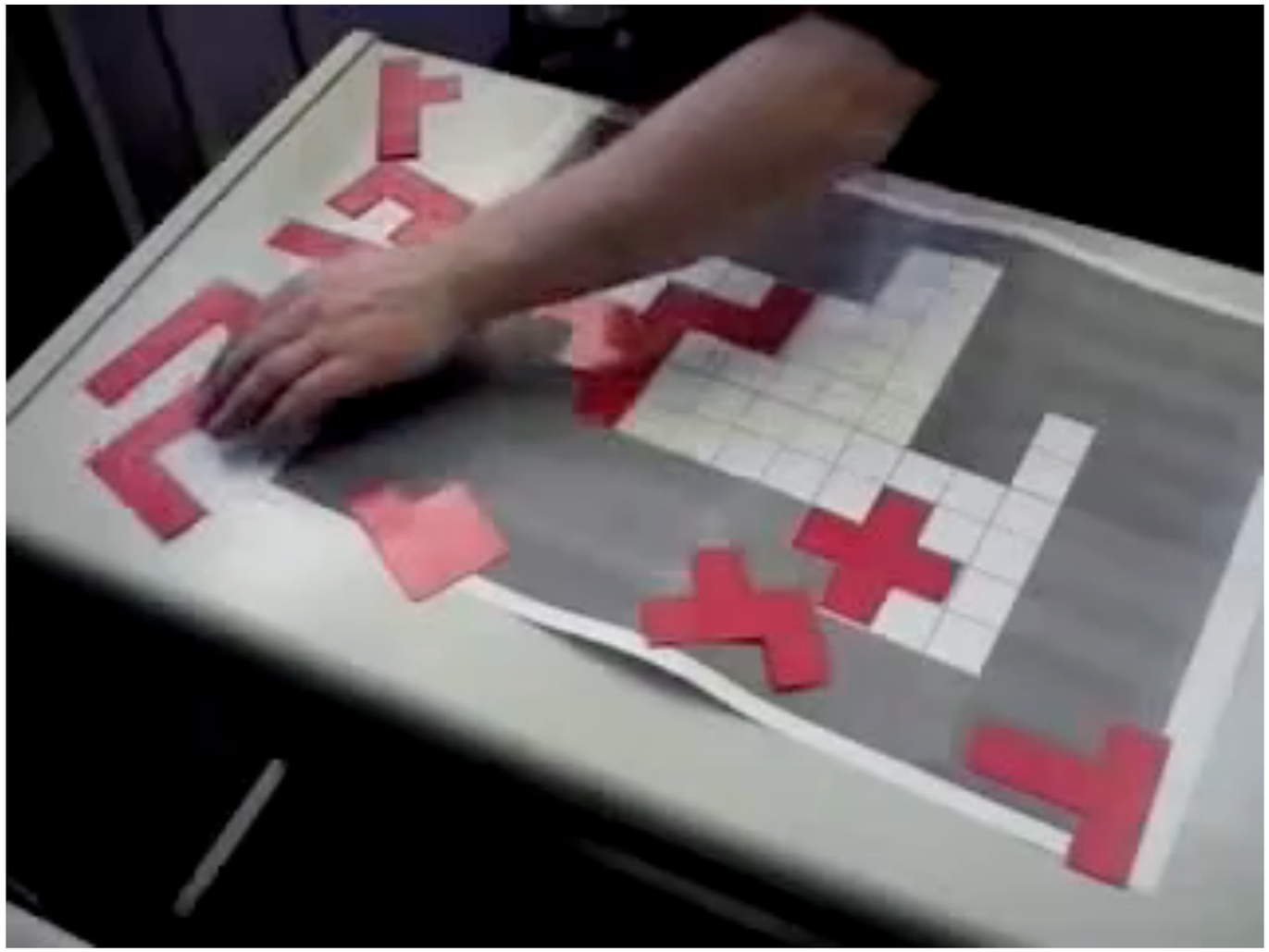
- wider context
- the explanandum, in the form of a story
- constraints on the explanation
- sketch of the proposal
- empirical evidence / data
- why not deep end-to-end learning?

data

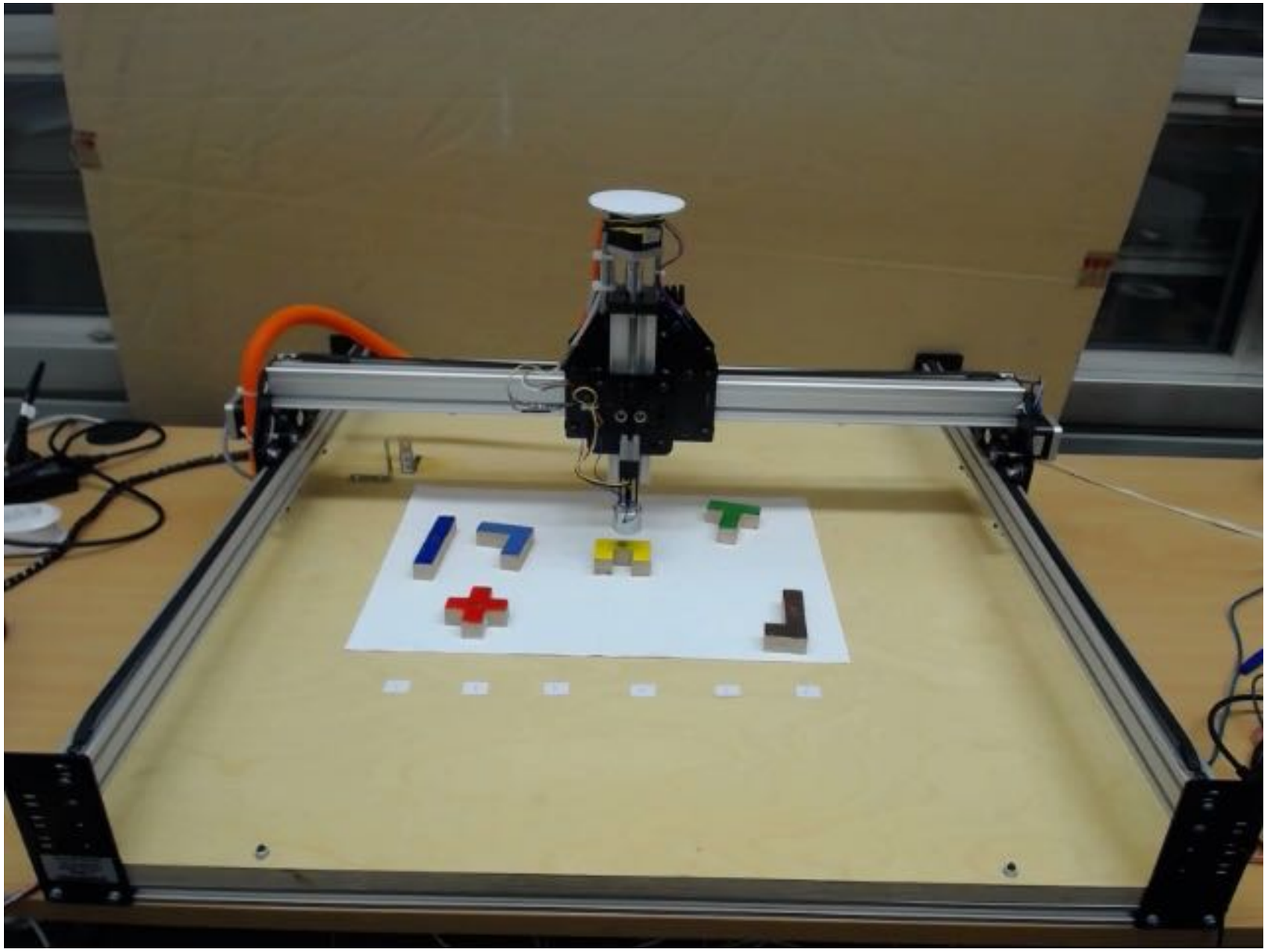
- we need learning situations: ostensive definitions + non-linguistic context, explicit linguistic definitions, implicit linguistic definitions

- we need testing situations: referring expressions, statements, & non-linguistic context.

- (Schlangen, IWCS 2019)



(Fernández & Schlangen, SIGdial 2007)



(Hough & Schlangen, semdial 2016)

Experiment	# tokens	# types	# utts	# games	# participants	Annotations
Human-wizard Interactions						
WOz Pento	9149	237	1686	284	12	scene-logical, target
Take	13863	383	1045	1214	8	scene-logical, target, dialogue act tags, disfluencies
Take-CV	15053	736	870	870	9	scene-perceptual, target, landmark, relation
Human-human Dialogues						
Noise/No-Noise	29057	1482	6073	11	22	scene-logical, target, disfluencies
Visual Pento	4610	907	1158	6	12	scene-logical, target, dialogue act tags
Pento-CV	89373	1828	6108	32	16	scene-perceptual, target, dialogue act tags, disfluencies
RDG-Pento (En)	55238	1371	8030	24	48	scene-perceptual, target, dialogue act tags, disfluencies
	216,343					

(Zarrieß, Hough, Kennington, Manuvinakurike, DeVault, Fernández, Schlangen, LREC 2016)

- SAIAPR (2006 ff.): 20k images (Grubinger *et al.* 2006; Escalante *et al.* 2010), 200,000 referring expressions (Kazemzadeh *et al.* 2014)

- Flickr30k (2014 ff.): 30k images, 160,000 captions (Young *et al.* 2014)

- MSCOCO (2014 ff.): 300k images, 400,000 captions (Lin *et al.* 2014), 280,000 referring expressions (Yu *et al.* 2016)

- VisualGenome (2016 ff.): 100k images, 2e6 region descriptions (Krishna *et al.* 2016)

Can we create plausible learning and application situations from that data?

the sempix / clp-vision package

SAIAPR

COCO Captions

Flickr30kEntities

VisualGenome
Captions, Region Descriptions,
VQA, Objects, Attributes, Relations

VQA

Visual Dialogue

GuessWhat?

CUB-Birds

	i_corpus	image_id	split	
0	1	378466	train	a man laying in bed next to his dog
1	1	378466	train	a shirtless man poses next to his dog
2	1	378466	train	a person that is laying next to a dog
3	1	378466	train	a man is lying down on the bed looking
4	1	378466	train	a man with a beard without a shirt and a

easy to use Pandas DataFrames

similarity relations

soon: treebanks!

https://github.com/davidschlangen/cosine-paris

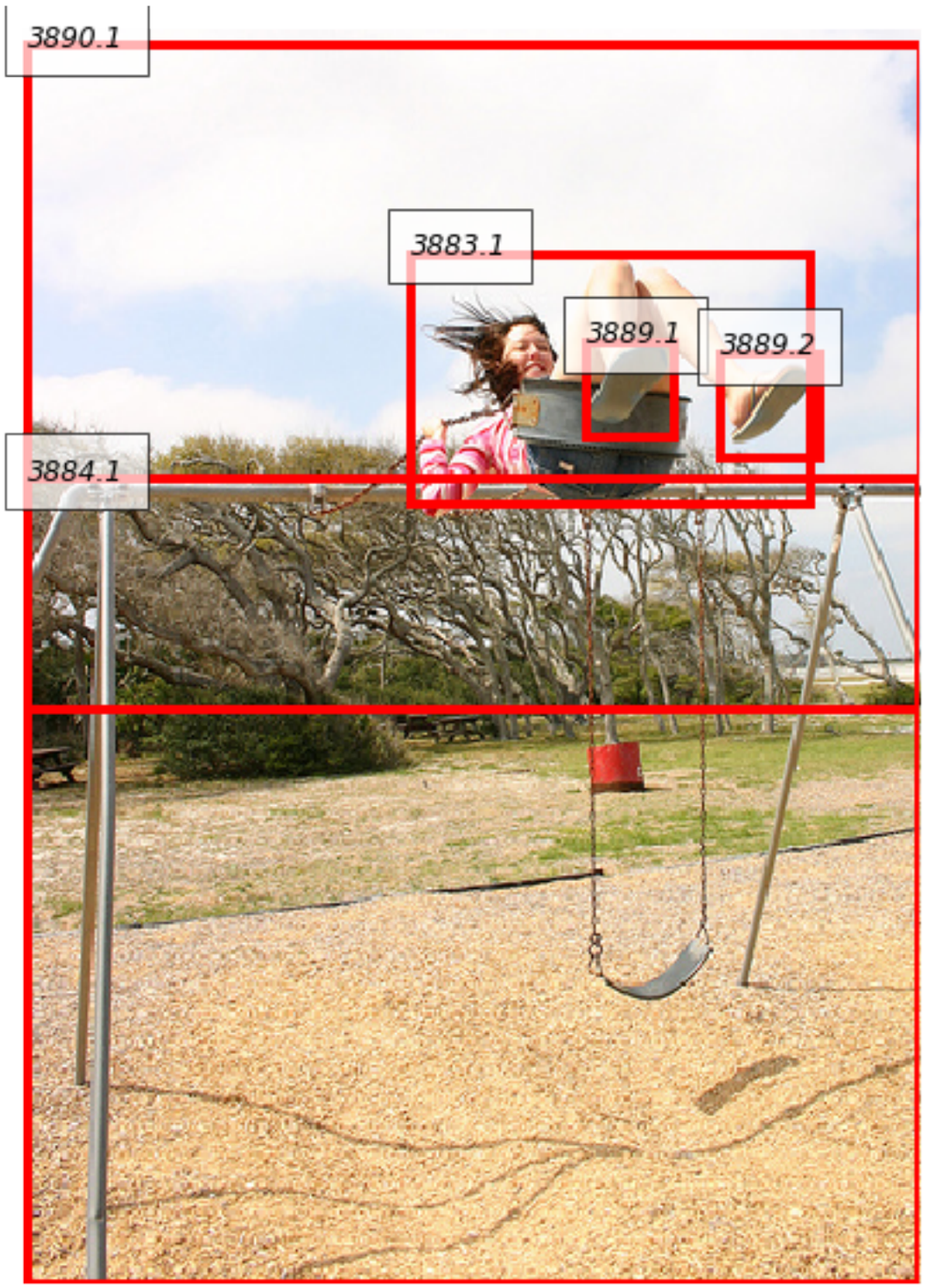
https://github.com/clp-research/sempix

SAIAPR

SAIAPR

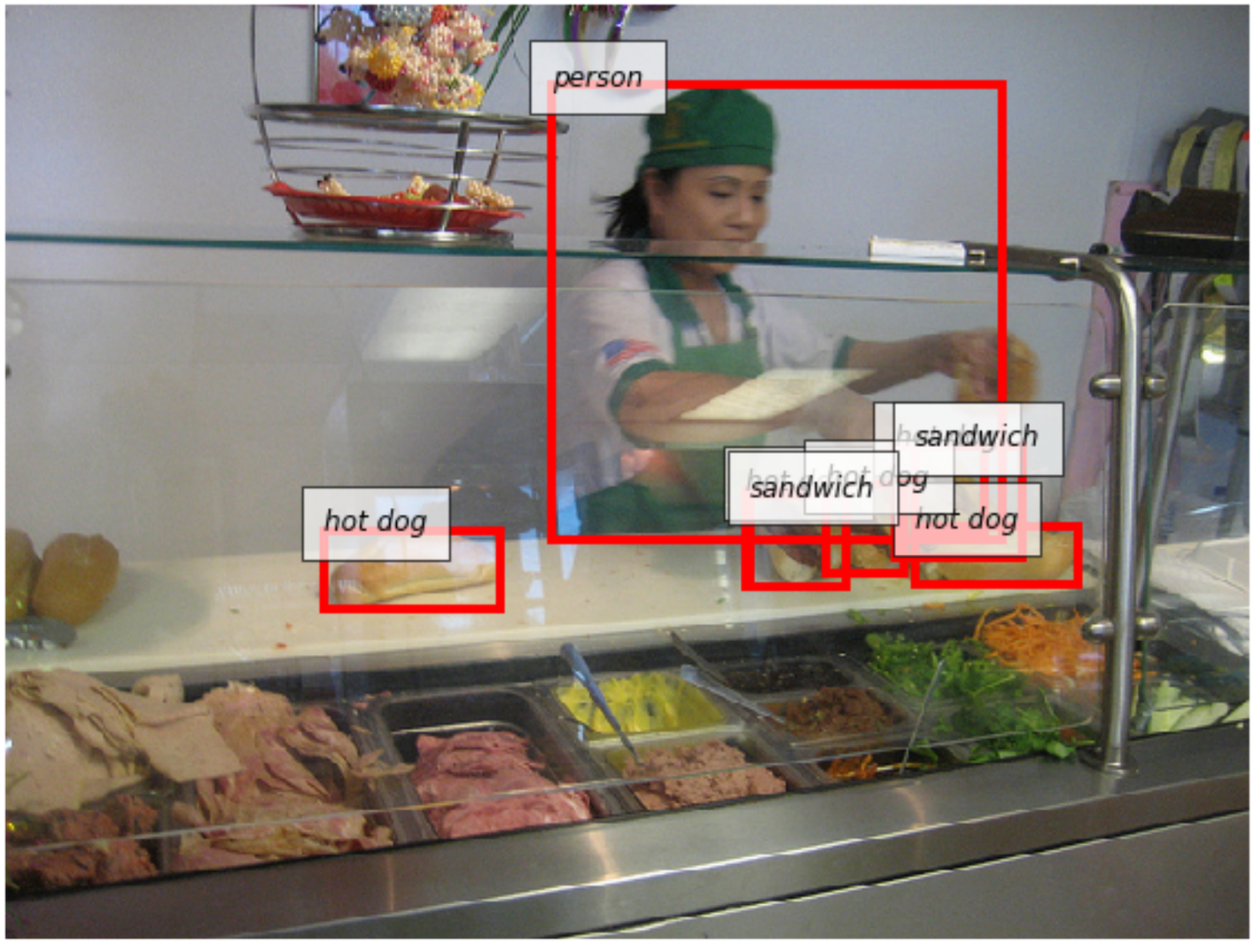
the girl in green

Flickr30k



a girl wearing flip-flop sandals swinging on a swing set in a park underneath fluffy white clouds

MS COCO



A woman prepares several sub sandwiches at a deli counter.

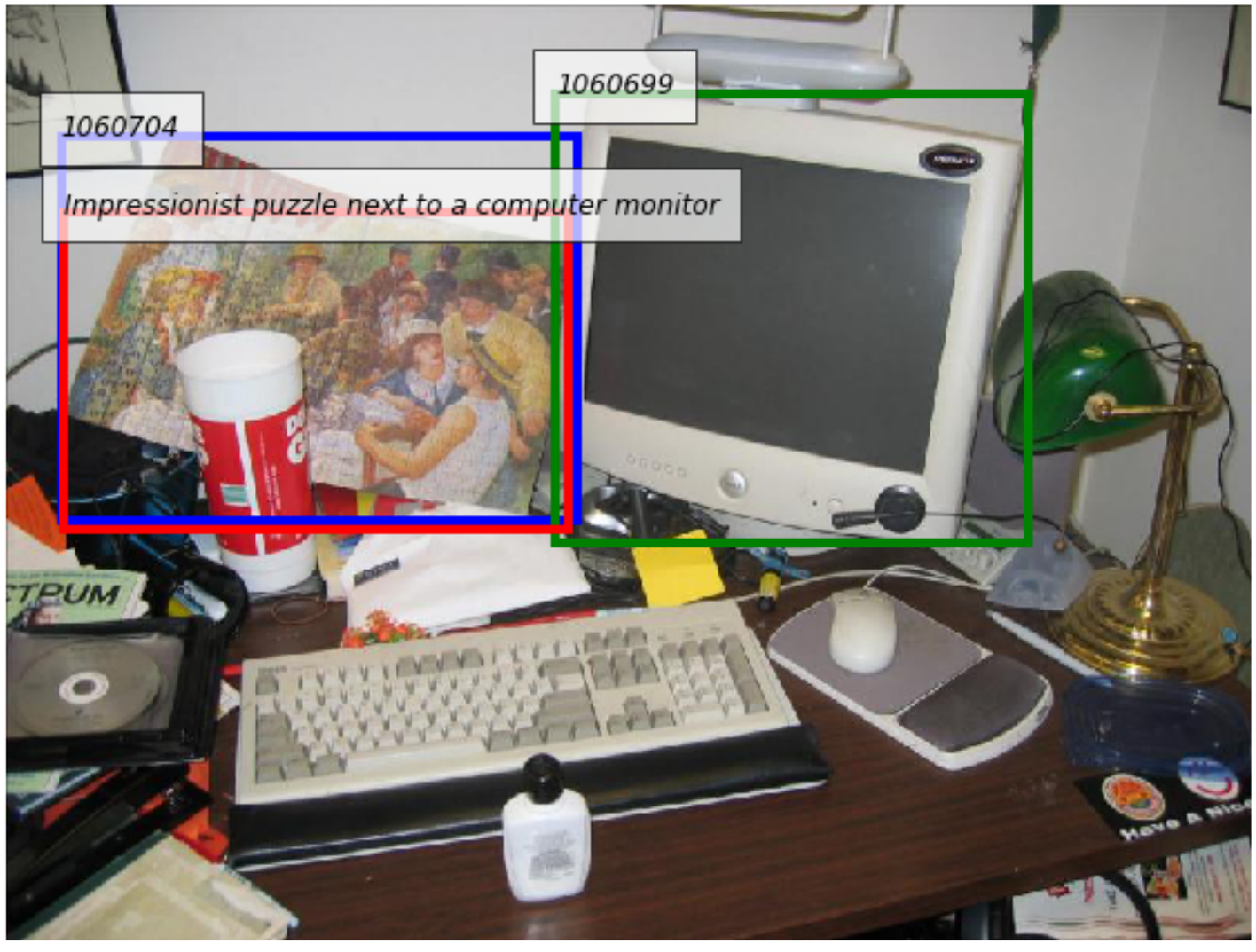
a person behind a display glass preparing food

A lady behind a sneeze guard making sub sandwiches.

THERE IS A WOMAN THAT IS MAKING SANDWICHES AT A DELI

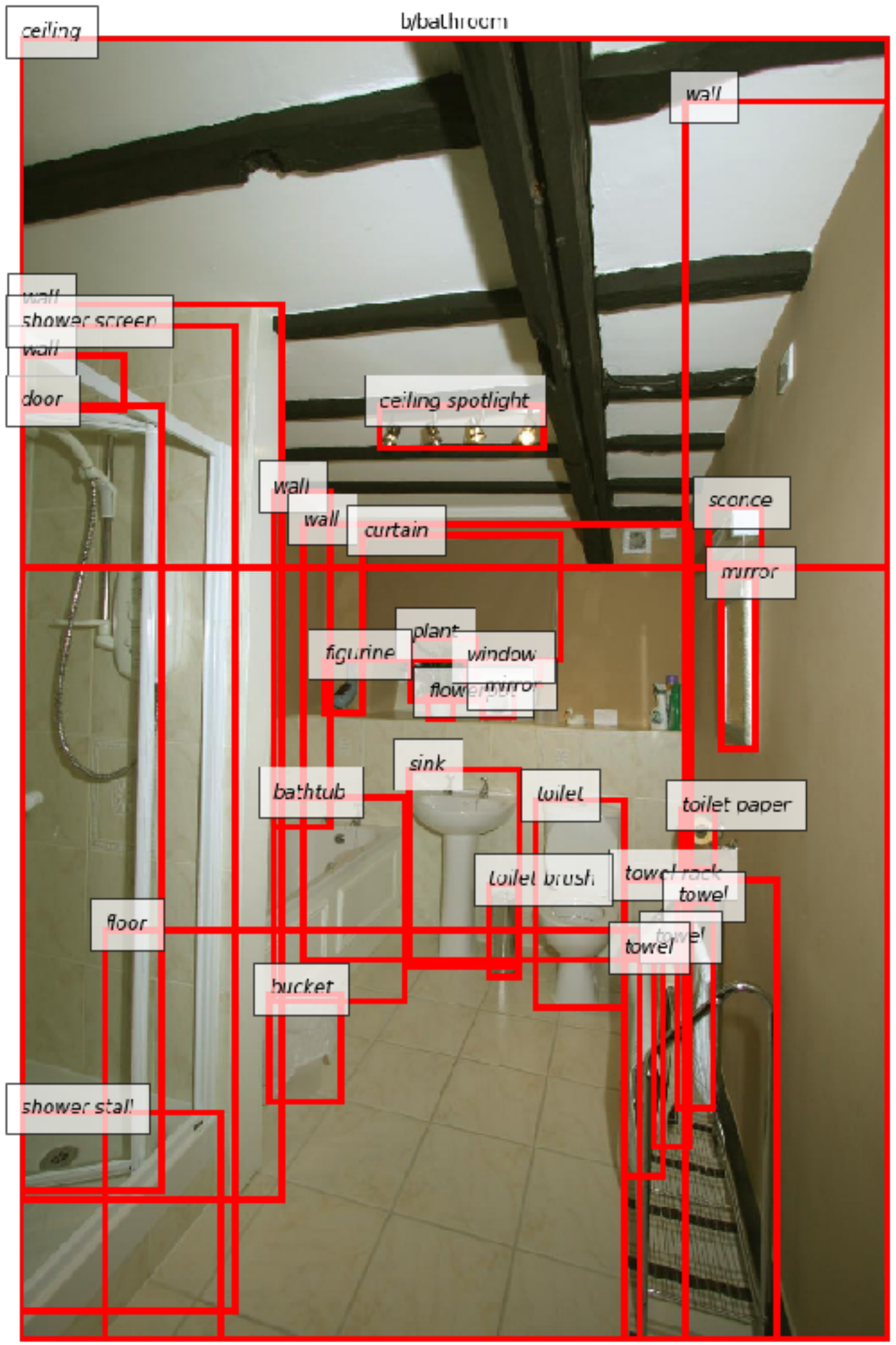
A woman behind a deli counter making sub sandwiches.

visual genome



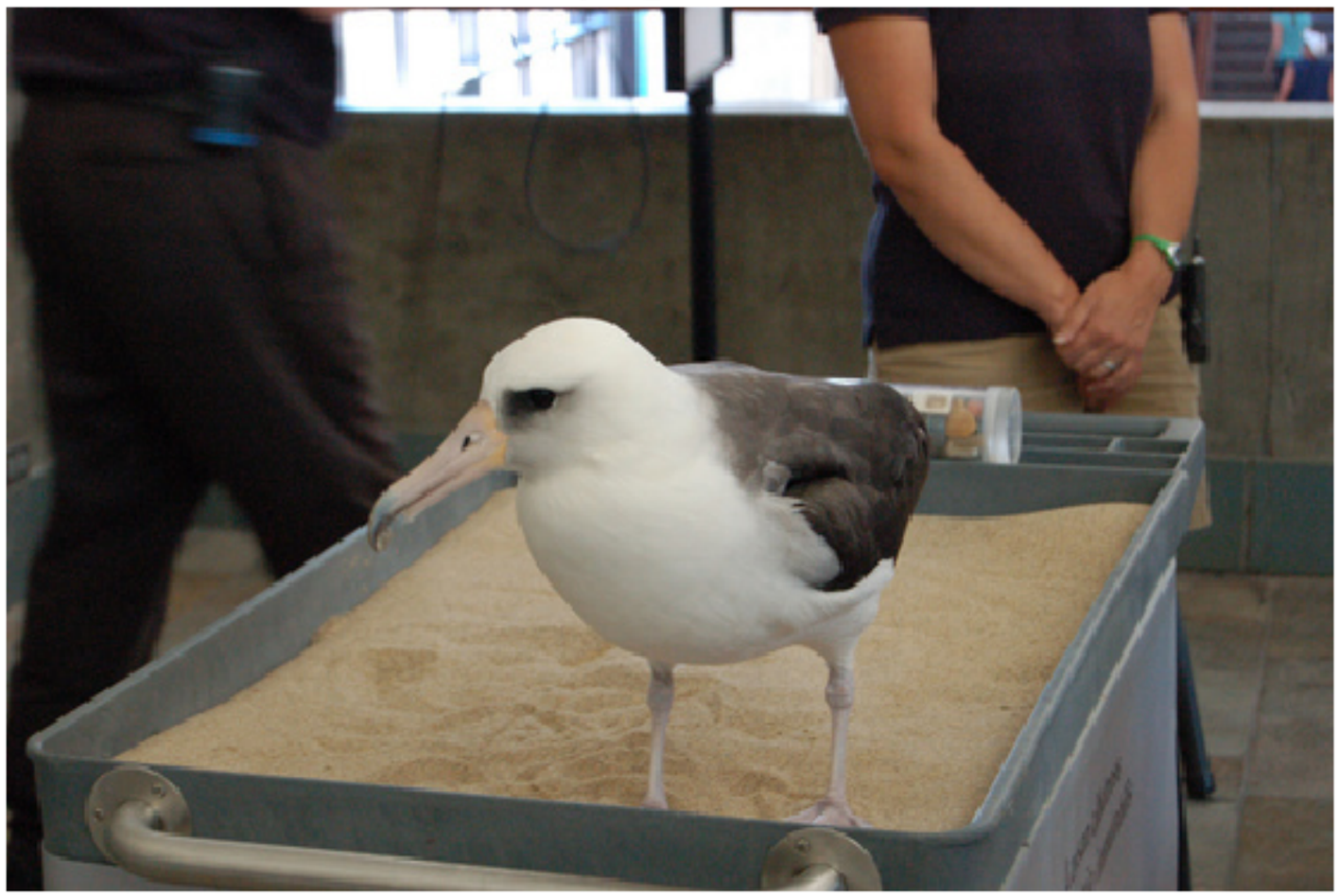
ADE20k

(Zhou *et al.* 2017)



CUB-200-2011 (Wah *et al.* 2011)

Laysan_Albatross



Large seabird with dark brown back, white head, neck and rump, dark eye patch. Bill is thick and yellow with gray hooked tip. Wings dark brown above and white below, with irregular brown-black borders, dark brown-black with white coverts, and pink legs and feet. Feeds on fish and invertebrates. Dynamic soaring, stays aloft for hours with little flapping of wings. Sexes similar.

this bird has a white body and head with black feathers and a long pointy beak.

a bird with a large, downward curved orange bill, white breast and black wings.

a white bird with a black secondaries with a pink a gray bill

this large bird has a white head and underside, brown feathering along the back and wings, with a long bill that curves at the tip.

Thank you.

I will be here (in Room 520) until Oct 13. Happy to meet up!

Thanks also to my Bielefeld PhD students, Postdocs, and collaborators: Julian Hough, Sina Zarrieß, Casey Kennington, Nikolai Ilinykh, Soledad Lopez, Ting Han, Nazia Attari, Spyros Kousidis.

Funding received from CITEC, DFG.

References

References to our own work can be resolved via http://clp.ling.uni-potsdam.de/publications/ (where also the PDFs are available).
(First authors: Han, Kennington, Kousidis, Lopez, Schlangen, Zarrieß.)

- Aitchinson, J. (1994). Words in the Mind: An Introduction to the Mental Lexicon 2nd ed. Oxford: Blackwell
- Bloom, Paul (2000). How Children Learn the Meaning of Words. MIT Press
- Carey, S. (1978). The child as word learner. In J. Bresnan & G. A. Miller (Eds.), Linguistic theory and psychological reality (pp. 264–293). Cambridge, MA, USA: MIT Press.
- Clark, E. V. (2007). Young children's uptake of new words in conversation. Language in Society, 36(02), 157–182.
- Erk, K. (2013). Towards a semantics for distributional representations. In IWCS 2013.
- Escalante, H. J., Hernández, C. a., Gonzalez, J. a., López-López, a., Montes, M., Morales, E. F., … Grubinger, M. (2010). The segmented and annotated IAPR TC-12 benchmark. Computer Vision and Image Understanding, 114(4), 419–428
- Gottfried, A. W., Rose, S. A., & Bridger, W. H. (1977). Cross-modal Transfer in Human Infants. Child Development, 48(1), 118–123.
- M. Grubinger, P. Clough, H. Müller et al., ``The IAPR TC-12 benchmark: a new evaluation resource for visual information systems'', Proceedings of the International Conference on Language Resources and Evaluation (LREC 2006), 2006
- Harnad, S. (1990). The Symbol Grounding Problem. Physica D, 42, 335–346.
- Hocket, C. F. (1960). The origin of speech. Scientific American, 203, 89–96.
- Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., … Fei-Fei, L. (2016). Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations. ArXiv
- Lewis, D. (1970). General Semantics. Synthese, 22(1), 18–67.
- T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Doll´ar, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in ECCV, 2014.
- Marr, D. (1982). Vision. Holt & Co., NY

References

- McDonald, S., & Ramscar, M. (2001). Testing the distributional hypothesis: The influence of context on judgements of semantic similarity. Proceedings of the 23rd Annual Conference of the Cognitive Science Society, 611–616.

- Partee, B. H. (1980). Montague Grammar, Mental Representation, And Reality. In S. Kanger & S. Öhman (Eds.), Philosophy and Grammar (pp. 59–78).

- John Perry (2012), Reference and Reflexivity (2nd ed), CSLI Press

- Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. Cognition, 32, 65–97.

- Pylyshy, Z. (2007). Things and Places: How the Mind Connects with the World. MIT Press

- Sellars, W. (1954). Some reflections on language games. Philosophy of Science 21 (3):204-228

- Spelke, E. (1994). Initial knowledge: six suggestions. Cognition, 50(1–3), 431–445.

- Wah, C., Branson, S., Welinder, P., Perona, P., & Belongie, S. (2011). The Caltech-UCSD Birds-200-2011 Dataset. Cns-Tr-2011-001.

- Young, P., Lai, A., Hodosh, M., & Hockenmaier, J. (2014). From Image Descriptions to Visual Denotations: New Similarity Metrics for Semantic Inference over Event Descriptions. Transactions of the Association for Computational Linguistics (TACL), 2(April), 67–78.

- Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). Scene parsing through ADE20K dataset. In Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, pp. 5122–5130