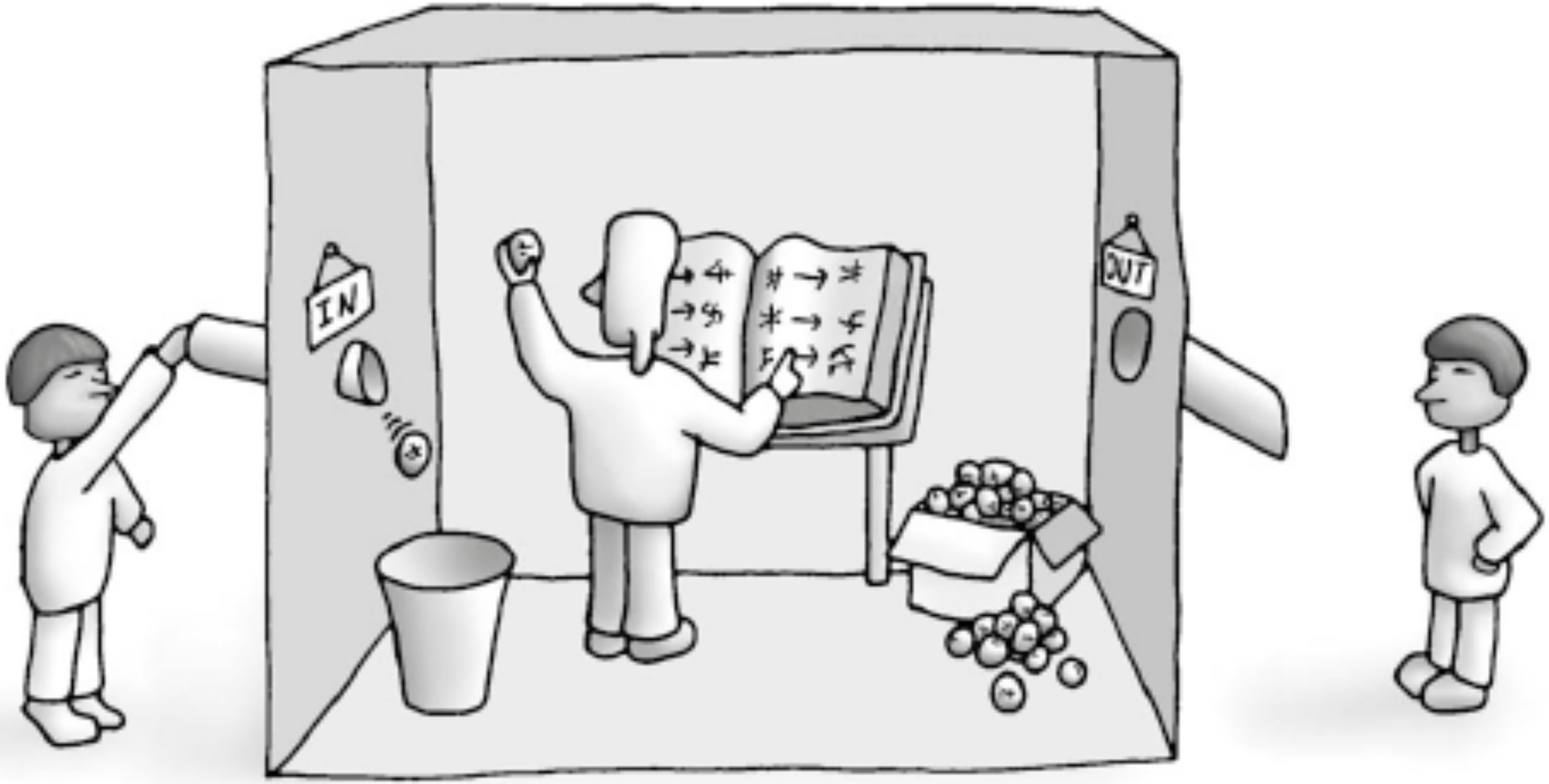
facebook
Artificial Intelligence Research

Reinterpreting Wittgenstein:
Grounded Multi-Agent Language Games

Douwe Kiela
Pronunciation guide: DOW-uh KEE-lah
Facebook AI Research (FAIR)
New York



How can you know the meaning of a symbol if it is defined only through other symbols?

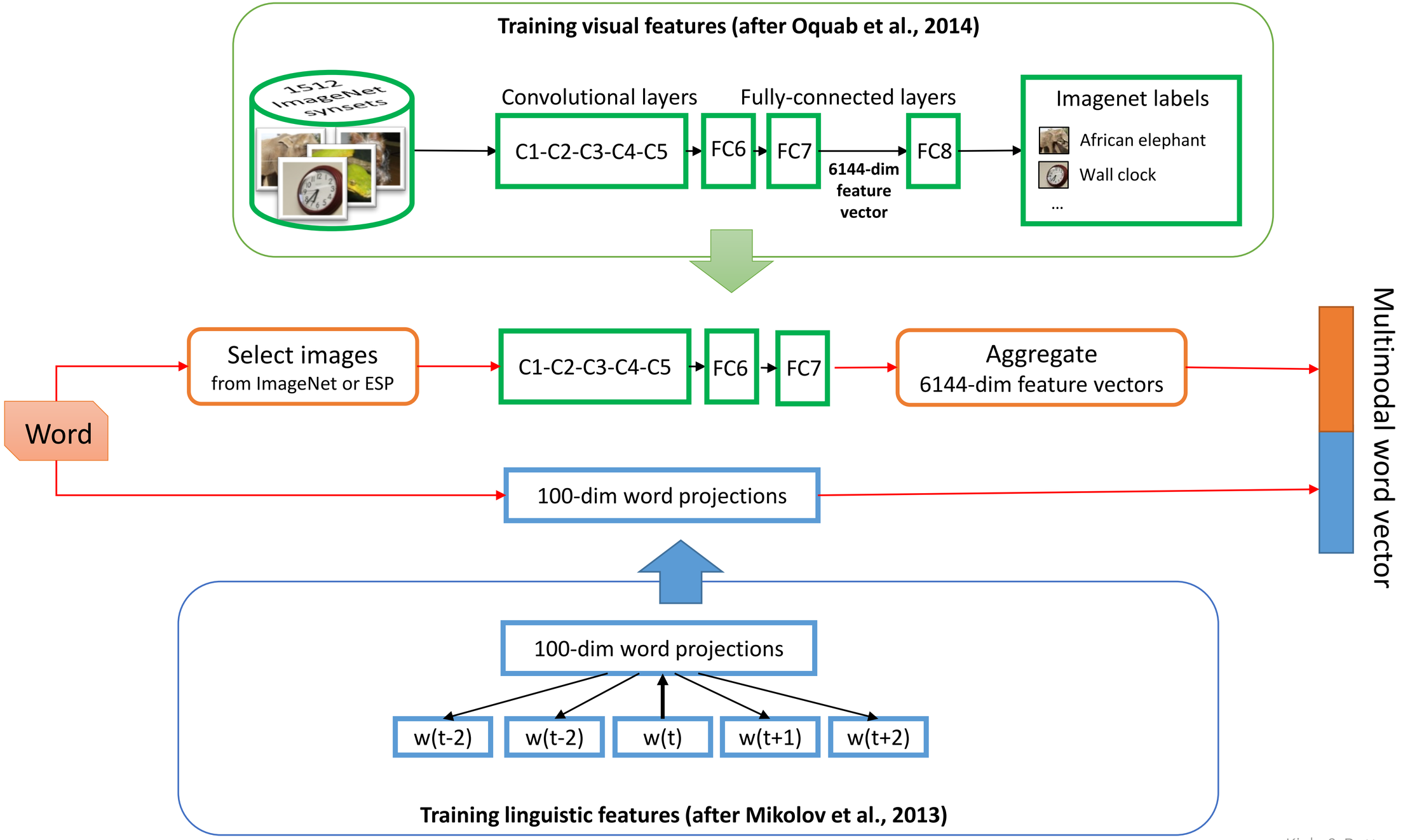
Source: WikiCommons

"Cat"

A small domesticated carnivorous mammal with soft fur, a short snout and retractile claws, widely kept as a pet or for catching mice.

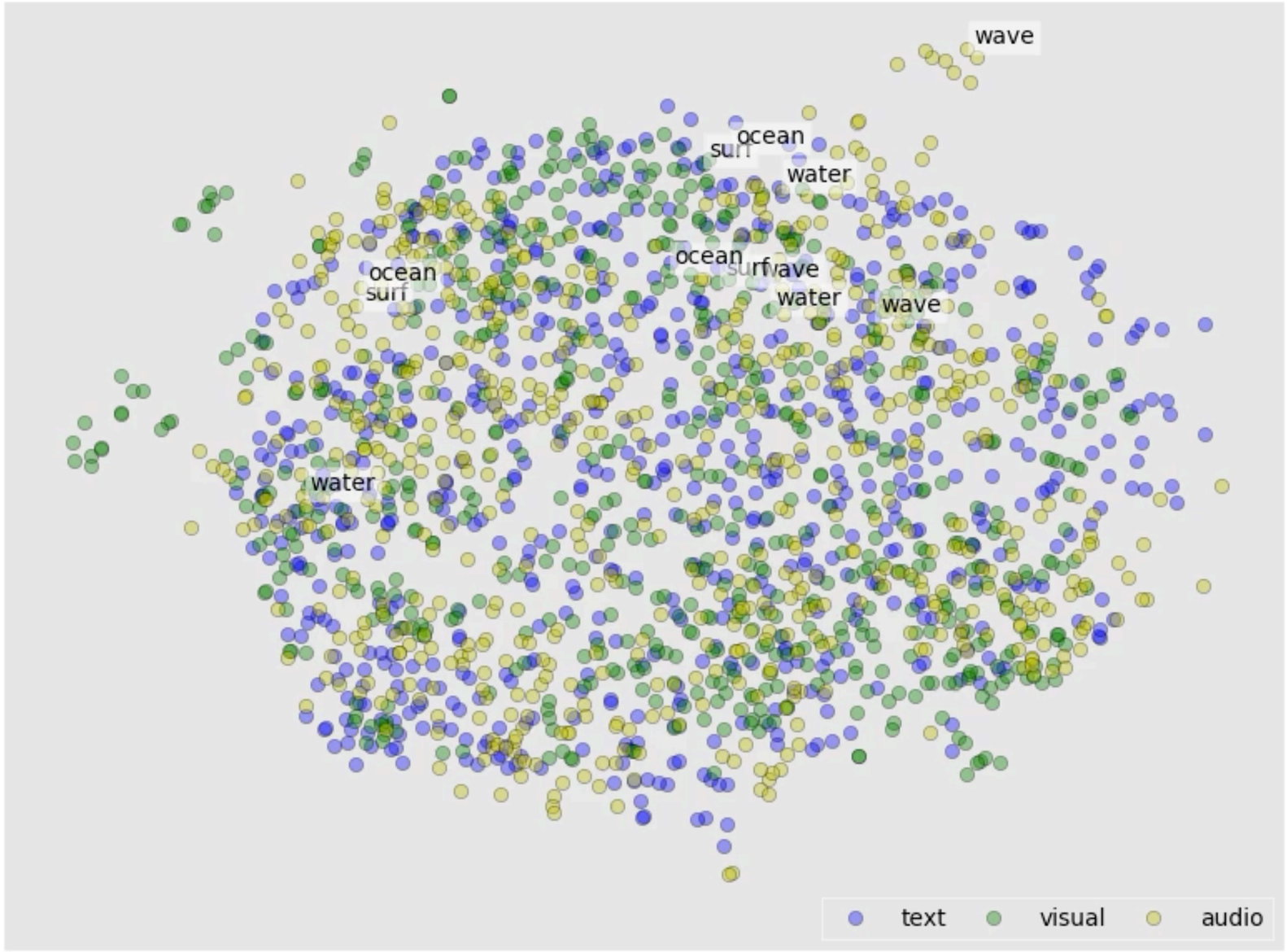
"Democracy"

A system of government by the whole population or all the eligible members of a state, typically through elected representatives.



Kiela & Bottou, EMNLP 2014

Grounding Semantics in Perceptual Modalities

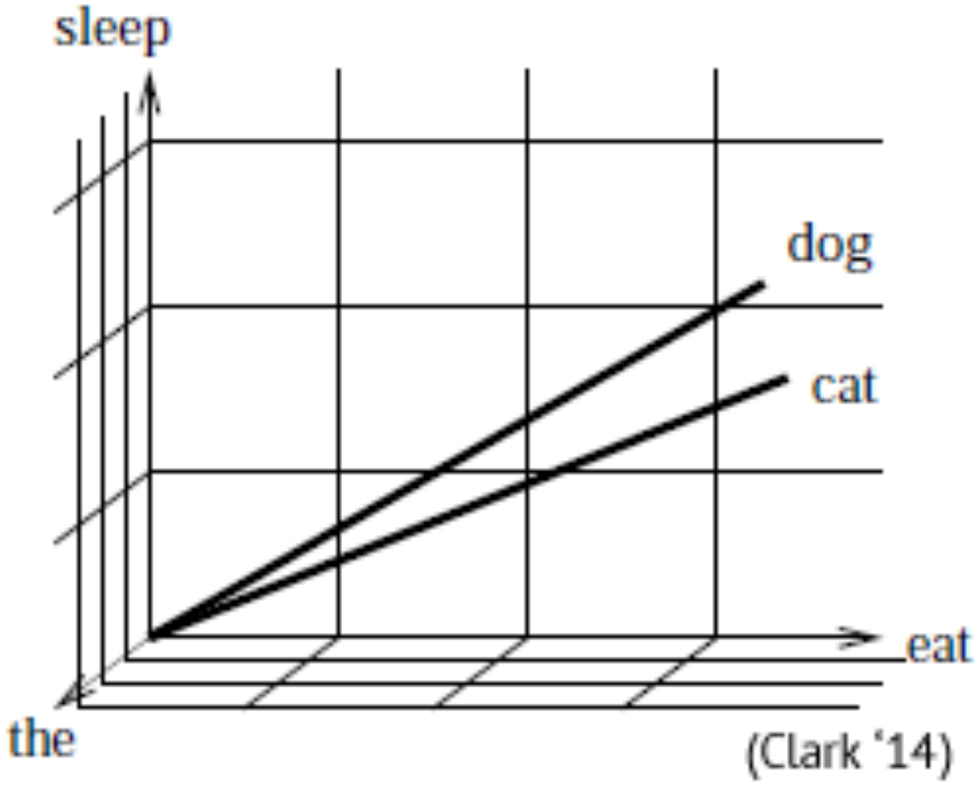

wave
ocean
surf
water
ocean
surf
ocean
surf
wave
water
wave
water
text
visual
audio

Distributional Hypothesis

"You shall know a word by the company it keeps"

- **Vector Space Models of Meaning**
- **Word Embeddings: Word2Vec/GloVe/FastText/etc**
- **Sentence Embeddings: SkipThought**
- **Unsupervised pretraining (ELMo, BERT, and friends)**

^ ALL SUFFER FROM THE GROUNDING PROBLEM.



Outline

What I will be talking about

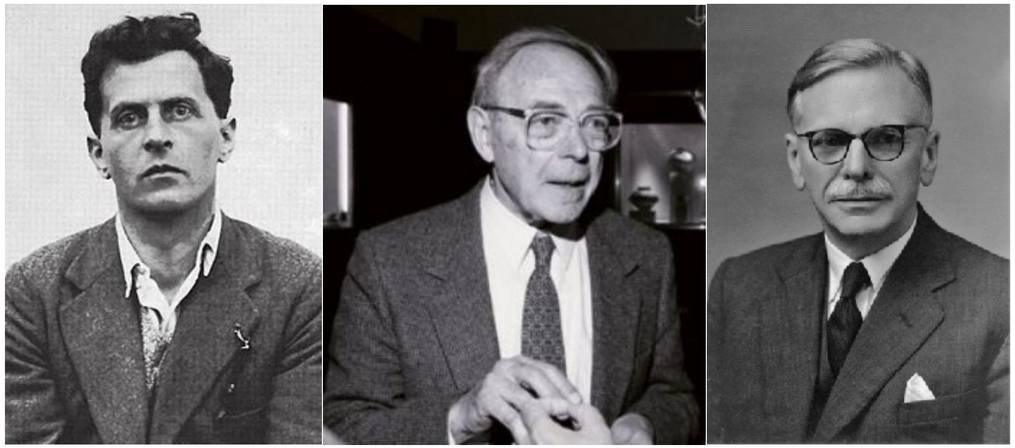
- **The Distributional Hypothesis, or, How We Misinterpreted Wittgenstein**
- **Recent Research:**
  - **Emergent Translation in Multi-Agent Communication**
  - **Countering Language Drift via Grounding**
  - **Talk The Walk: A New Dataset**
- **Future Directions**

Distributional Hypothesis

A Brief History

- **Zellig Harris** (1954): ″There is no way to define or describe the language and its occurrences except in such statements said in that same language or in another natural language.″ Thus, a science that aims to determine the nature of language is limited to investigation of the relationships of the elements of language to one another.

- **John Rupert Firth** (1957): ″The placing of a text as a constituent in a context of situation contributes to the statement of meaning since situations are set up to recognize use. **As Wittgenstein says, ‘the meaning of words lies in their use.’** (Phil. Investigations, 80, 109). The day-to-day practice of playing language games recognizes customs and rules. [..] **You shall know a word by the company it keeps!**



Philosophy, Yay!

Meaning = Use

Ludwig Wittgenstein

Philosophische Untersuchungen (1950)

"the meaning of words lies in their use"

But what is the meaning of the word "five"?—No such thing was in question here, only how the word "five" is used.

Philosophy of Language

Platonism and Pragmatism (via Christopher Mole)

Platonism (Plato, Wittgenstein I):
"Start with the fact that symbols have meaning. Explain the rational behavior of the symbol in light of that fact."
meaning is a relationship between symbols and what they refer to (i.e., objects)

vs.

Pragmatism (Wittgenstein II, Dewey, Brandom):
"Explain the fact that symbols have meaning by reference to the rational behavior in which they participate."
meaning is how a symbol is perceived by the listener or intended by the speaker

Language Games

Excerpts from the Philosophical Investigations

(2) That philosophical concept of meaning has its place in a primitive idea of the way language functions. But one can also say that it is the idea of a language more primitive than ours. Let us imagine a language for which the description given by Augustine [that words denote objects] is right. The language is meant to serve for communication between a **builder A and an assistant B**. A is building with building stones: there are blocks, pillars, slabs and beams. B has to pass the stones, and that in the order in which A needs them. For this purpose they use a language consisting of the words ″block″, ″pillar″, ″slab″, ″beam″. A calls them out;—B brings the stone which he has learnt to bring at such-and-such a call.—Conceive this as a complete primitive language.

(7) We can also think of the whole process of using words in (2) as one of those games by means of which children learn their native language. I will call these games ″language-games″ and will sometimes speak of a primitive language as a language-game. [...] **I shall also call the whole, consisting of language and the actions into which it is woven, the ″language-game″.**

The Meaning of "Meaning"

What the Philosophical Investigations *really* Say

(43) For a large class of cases—though not for all—in which we employ the word "meaning" it can be defined thus: the meaning of a word is its use in the language. And the meaning of a name is sometimes explained by pointing to its bearer.

The Meaning of ″Meaning″

What the Philosophical Investigations *really* Say (According to… Me)

(43) **For a large class of cases**—though not for all—in which we employ the word ″meaning″ it can be defined thus: **the meaning of a word is its use in the language**. And the meaning of a name is sometimes explained by pointing to its bearer.

The meaning of language **mostly depends on active usage** in a community of language speakers playing ″language games″.

The Meaning of "Meaning"

What the Philosophical Investigations *really* Say (According to… Me)

(43) For a large class of cases—**though not for all**—in which we employ the word "meaning" it can be defined thus: the meaning of a word is its use in the language. **And the meaning of a name is sometimes explained by pointing to its bearer.**

The meaning of language depends in part on **language-to-world grounding.**

The Meaning of "Meaning"

What the Philosophical Investigations *really* Say (According to... Me)

(43) **For a large class of cases**—**though not for all**—in which we employ the word "meaning" it can be defined thus: **the meaning of a word is its use in the language**. **And the meaning of a name is sometimes explained by pointing to its bearer.**

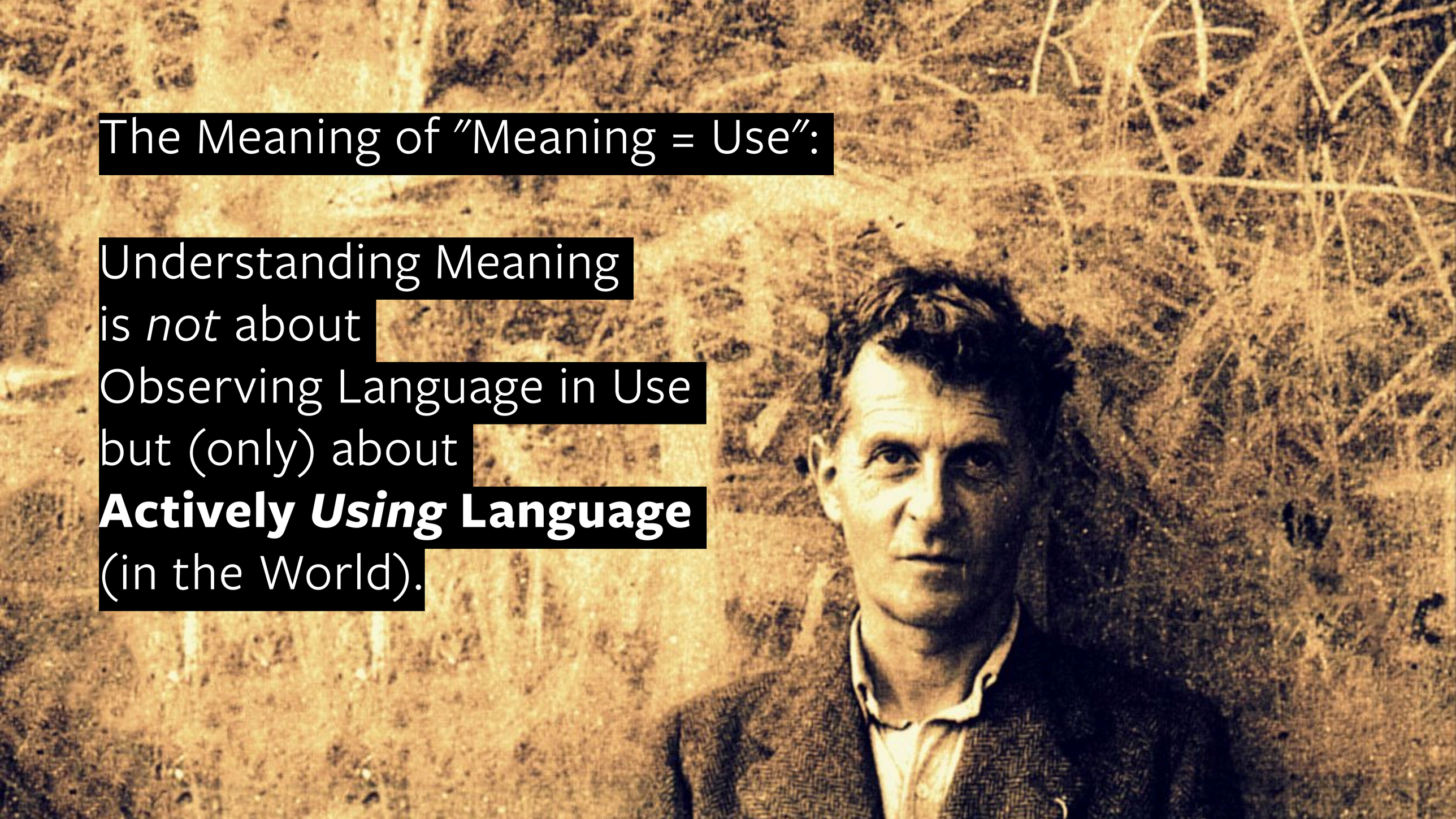
The meaning of language depends in part on **language-to-world grounding** and **mostly depends on active usage** in a community of language speakers playing "language games".

^ THIS IS NOT AT ALL WHAT NLP IS DOING AT THE MOMENT.

Has NLP become lazy?

"All we need is more data, more compute."

(which, to be fair, has been an extremely successful strategy)



The Meaning of ″Meaning = Use″:

Understanding Meaning
is *not* about
Observing Language in Use
but (only) about
Actively *Using* Language
(in the World).

Outline

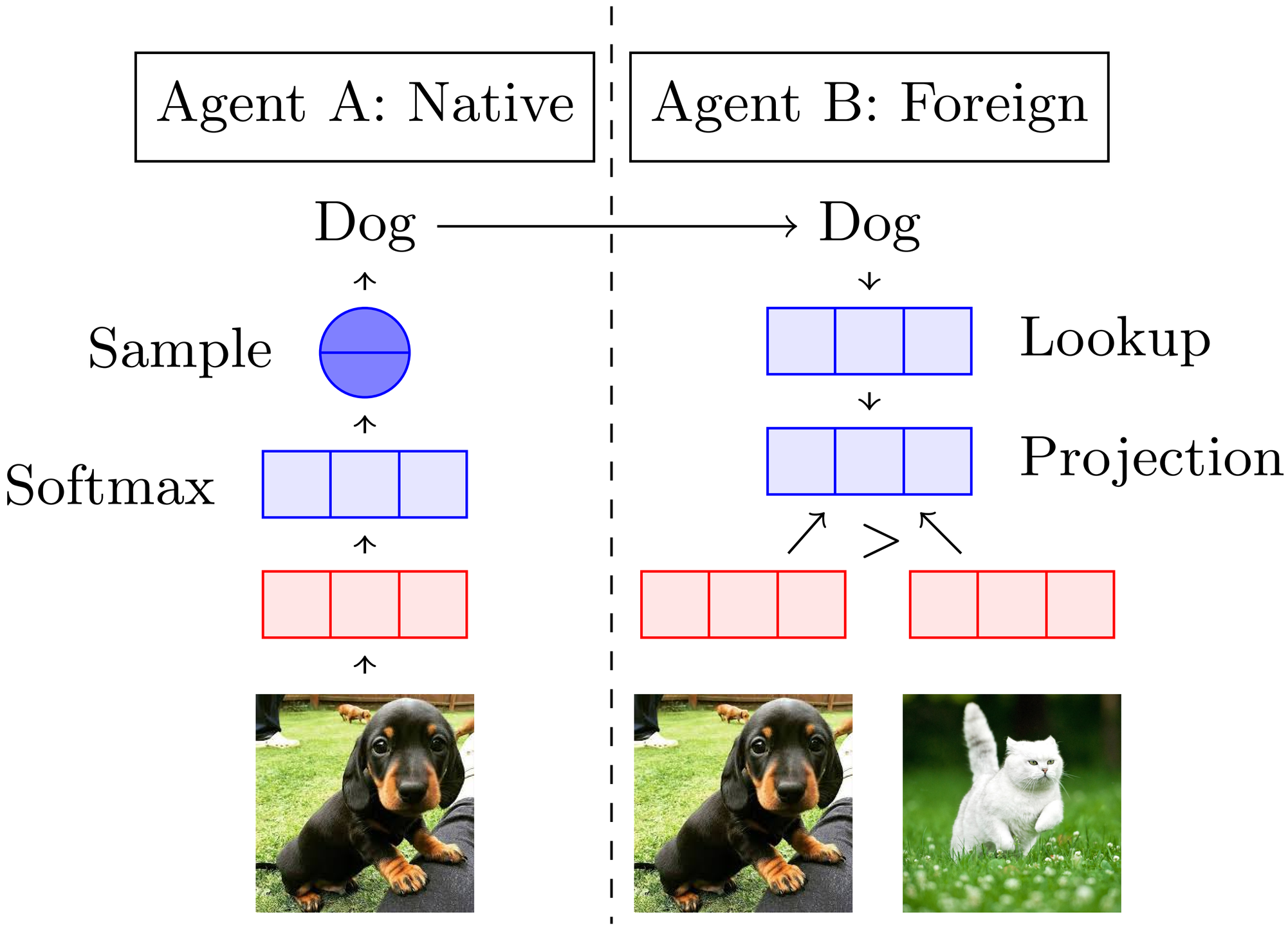
The rest of this talk

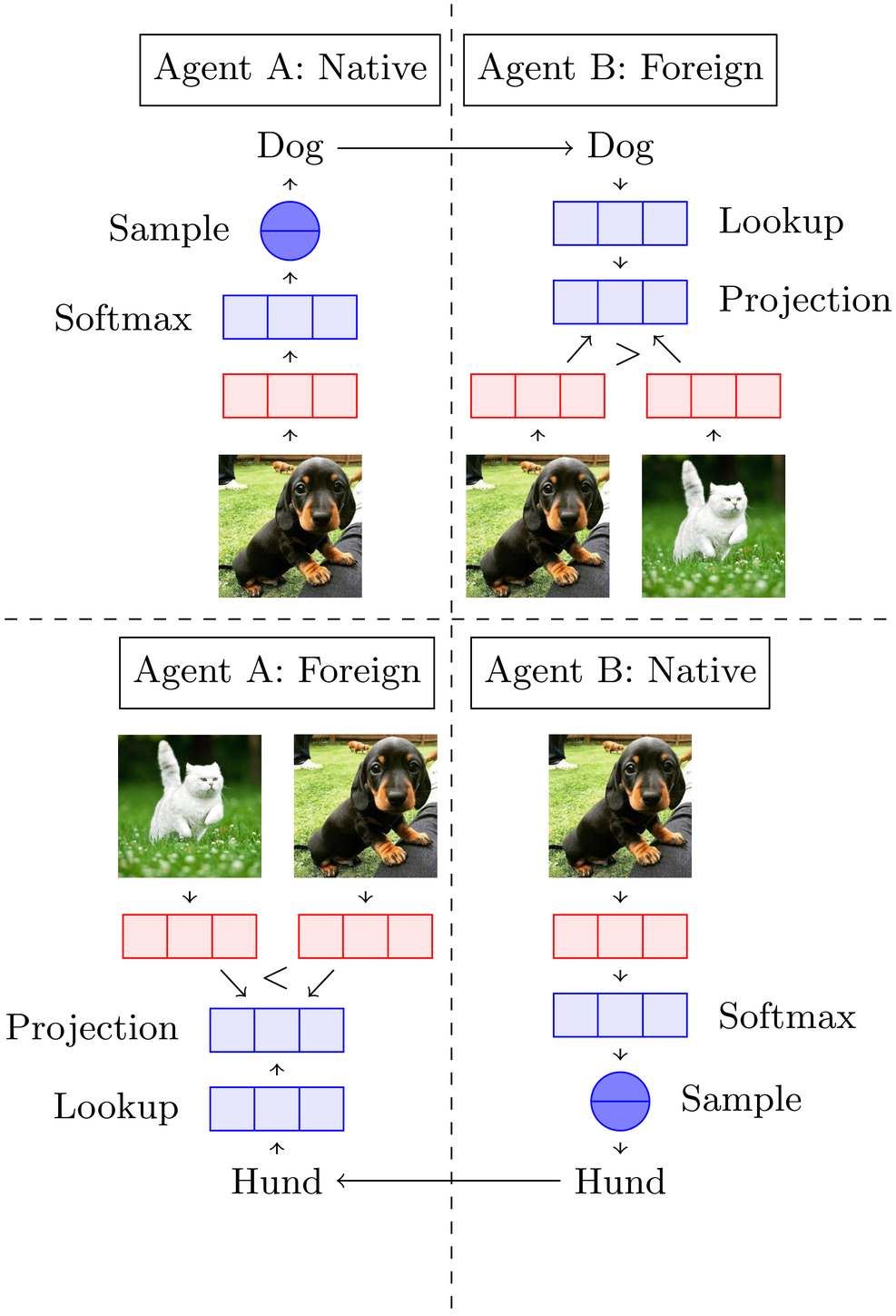
- **Take-aways:**
  - **The Distributional Hypothesis is great, but:**
  - The Grounding problem
  - The Passive Observation problem
  - **Take Wittgenstein seriously:**
    Grounded multi-agent language games

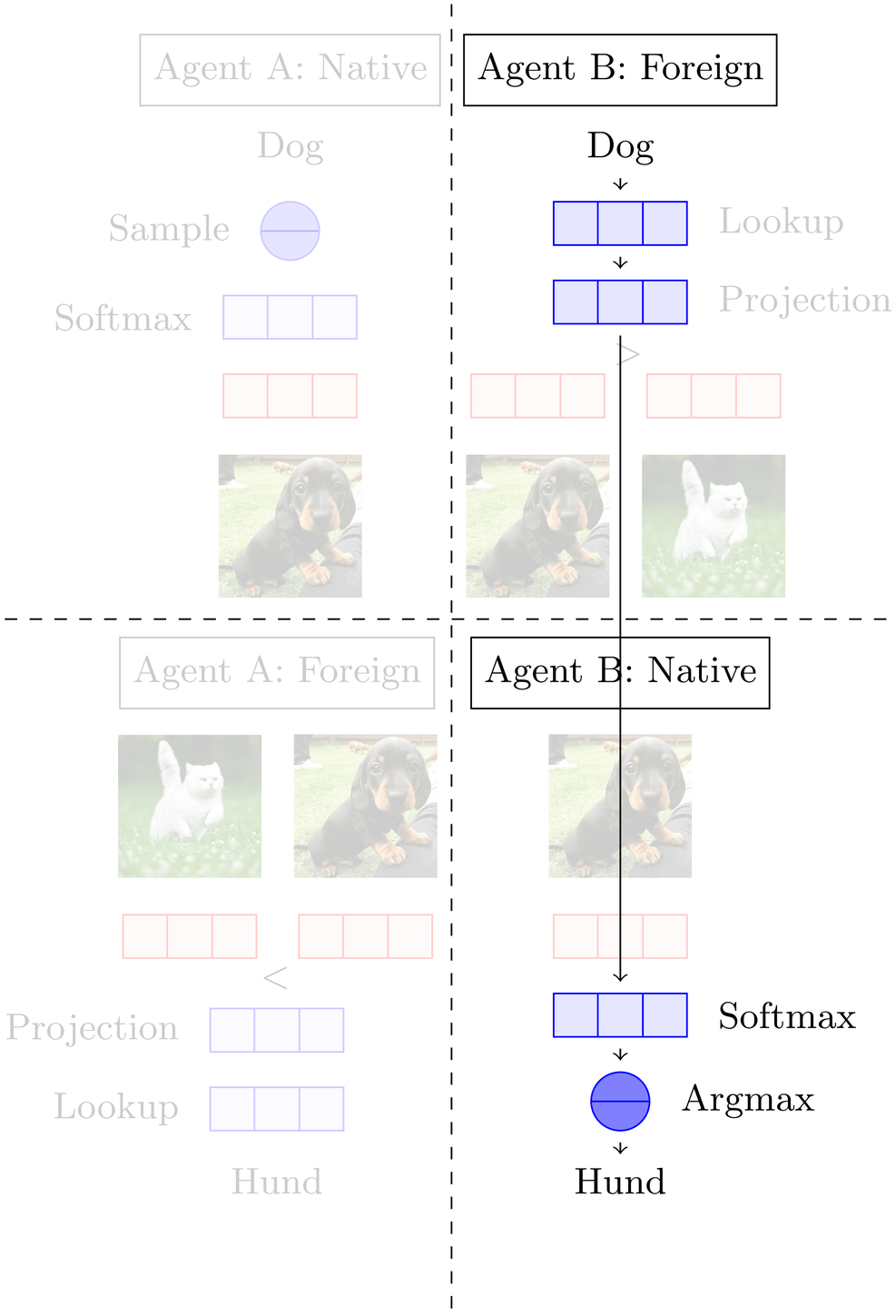
- **Next:**
  - Emergent Translation in Multi-Agent Communication
  - Countering Language Drift via Grounding
  - Talk The Walk: A New Dataset

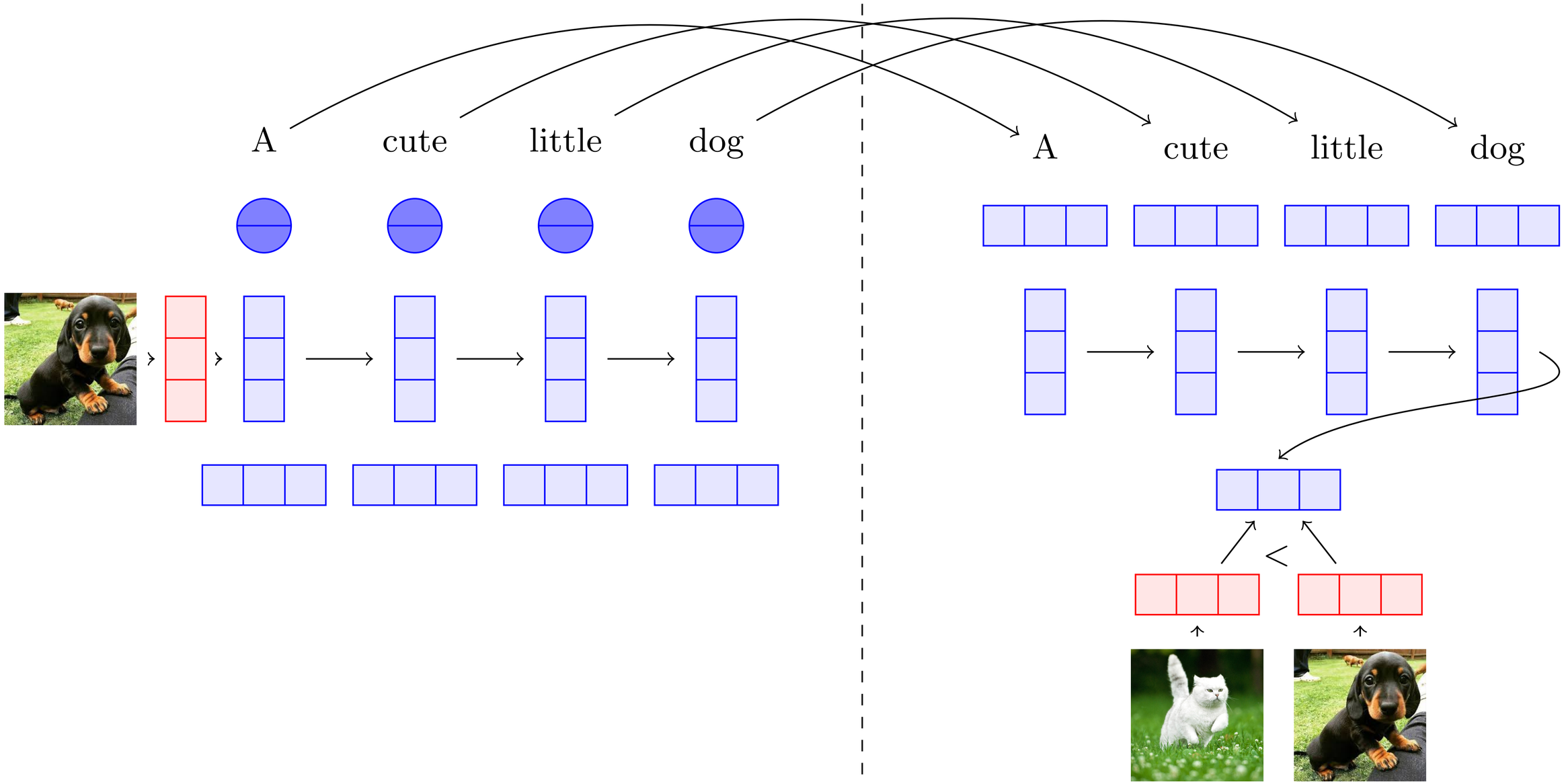
Emergent Translation in Multi-Agent Communication

Lee, Cho, Weston, Kiela. ICLR 2017.


Agent A: Native
Agent B: Foreign
Dog
Dog
Sample
Lookup
Softmax
Projection


Agent A: Native
Agent B: Foreign
Dog
Dog
Sample
Lookup
Softmax
Projection
Agent A: Foreign
Agent B: Native
Projection
Softmax
Lookup
Sample
Hund
Hund


Agent A: Native
Agent B: Foreign
Dog
Dog
Sample
Lookup
Softmax
Projection
Agent A: Foreign
Agent B: Native
Projection
Lookup
Softmax
Argmax
Hund
Hund


A
cute
little
dog
A
cute
little
dog

Emergent Translation in Multi-Agent Communication

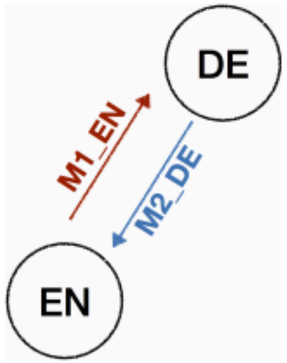
Unaligned Scores Approach Aligned (Parallel) Data

- **Datasets: Multi30k (EN-DE) and COCO-STAIR (EN-JP)**
- **Pre-training was important (more later).**
- **Without parallel data, just from playing a "reference game", we still get pretty close to aligned NMT performance.**
- **This is how humans would learn a completely new language.**

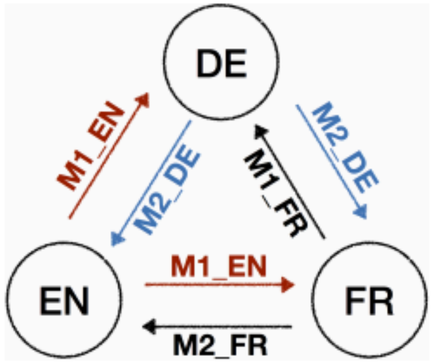
				Multi30k Task 1		Multi30k Task 2		COCO & STAIR	
				EN-DE	DE-EN	EN-DE	DE-EN	EN-JA	JA-EN
Unaligned Models			Nearest neighbor	1.41	1.77	3.75	5.87	15.88	10.94
			NMT with neighboring pairs	3.07	3.41	6.83	14.78	32.17	22.39
	Baselines	Their loss	N&N, 2-way, img	2.57	2.69	5.22	12.78	28.68	20.61
			N&N, 3-way, img	2.01	3.51	6.19	14.60	29.81	21.25
			N&N, 3-way, desc	3.34	3.87	9.66	15.96	27.53	17.51
			N&N, 3-way, both	1.50	3.62	9.89	15.50	31.01	20.59
		Our loss	N&N, 2-way, img	4.20	6.04	11.95	17.22	33.10	23.43
			N&N, 3-way, img	2.32	5.91	11.62	17.84	32.11	23.61
			N&N, 3-way, desc	5.13	6.02	11.07	17.01	26.65	17.82
			N&N, 3-way, both	4.89	6.59	13.53	18.48	32.84	23.28
	Our models		not pretrained	5.80	7.20	14.81	17.70	33.26	23.66
			pretrained, spk & enc fixed	5.81	7.36	13.87	18.68	**35.25**	**24.61**
			pretrained, spk fixed	**6.49**	**7.42**	**14.93**	**19.81**	33.01	23.59
			pretrained, not fixed	5.02	6.06	13.44	17.41	33.58	23.19
Aligned NMT				17.21	16.65	19.99	21.44	38.55	28.36

Emergent Translation in Multi-Agent Communication

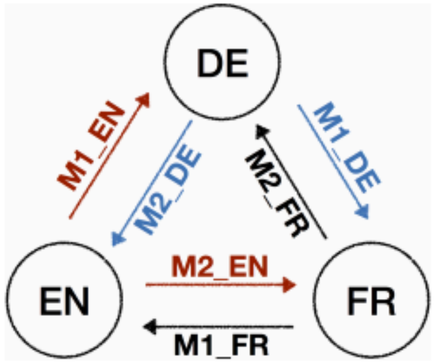
Multilingual Communities



(a) Single



(b) Fair



(c) Full

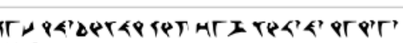
Model	EN-DE	DE-EN	EN-FR	FR-EN	DE-FR	FR-DE
Single	3.85	5.36	5.20	5.87	4.31	3.92
Fair	3.73	5.56	4.81	5.96	5.08	4.00
Full	**4.83**	**7.21**	**7.09**	**8.10**	**6.55**	**5.15**

Src	ein hund springt auf einer wiese vor einem weißen zaun in die luft .	ein mann hängt an einem seil mit rollen das über ein wasser gespannt ist .	ein skateboarder an einer böschung zu einem parkplatz .
Ref	a dog runs on the green grass near a wooden fence .	a man wearing bathing trunks is para-sailing in the water .	a skateboard is grinding on a curb with his skateboard .
NN	a brunette photographer is kneeling down to take a photo .	police watch some punk rock types at a protest .	a man is standing on the streets taking photographs .
NMT	a dog is jumping over a fence .	a man in a blue shirt is riding a bike .	a man in a blue shirt is riding a bike .
N&N	a brown dog is running on the grass .	a man in a wetsuit is surfing a large wave .	a man in a blue shirt is walking down the sidewalk .
Model	two dogs playing with a ball in the grass .	a man is parasailing in the ocean .	a man in a blue shirt and black pants is skateboarding .

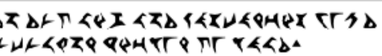
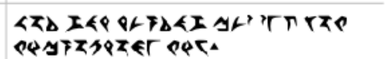
Emergent Translation in Multi-Agent Communication

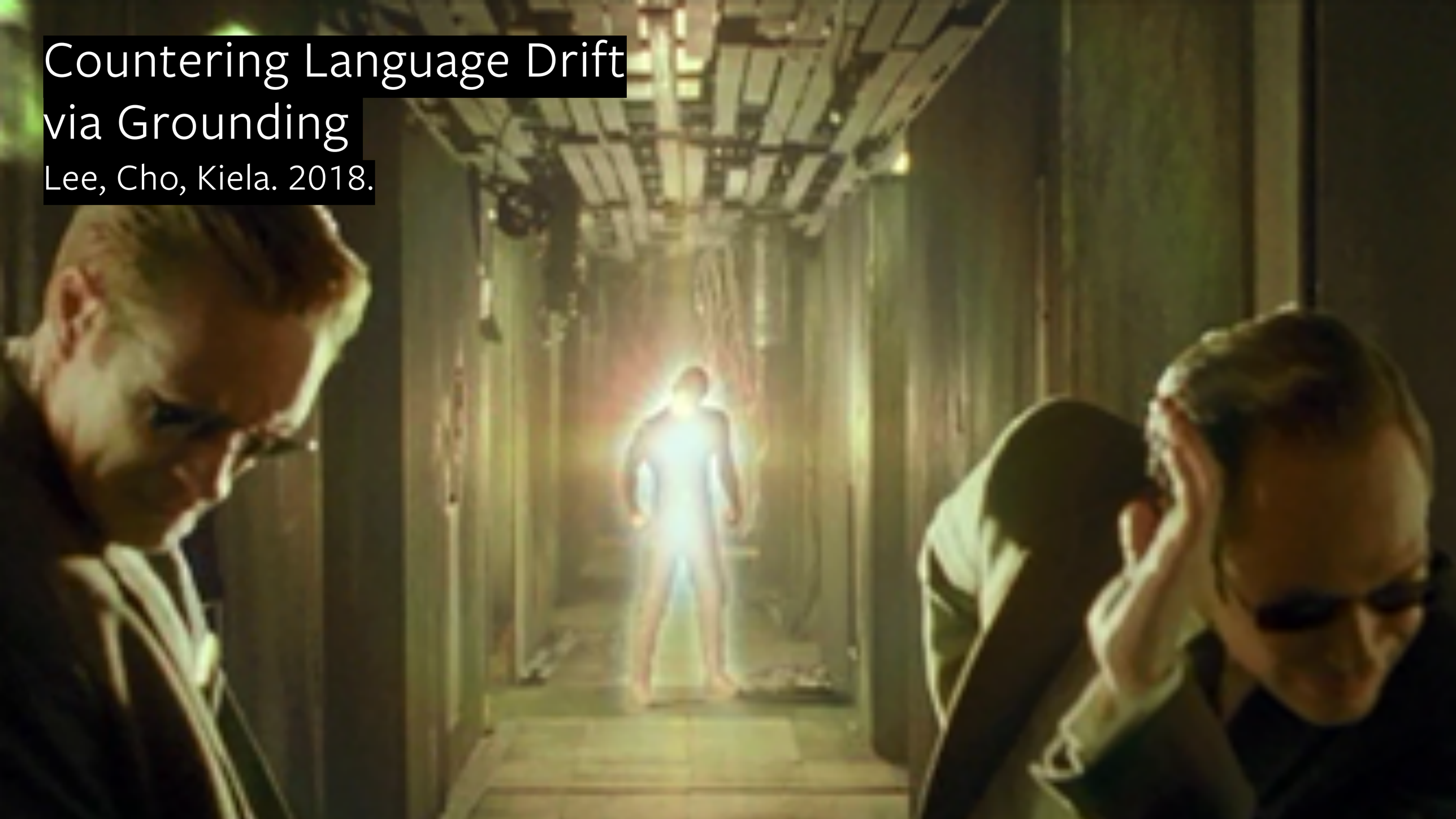
Klingon



Src	a black and white dog swimming in clear water .	a woman sitting on a very large rock smiling at the camera with trees in the background .	a young girl is swimming in a pool .
Hyp		 <UNK>	 <UNK>
Hyp (en)	the Black Dog ran, and go through the edge of the flood water.	We have caught her man some 200 top standing stone	only a young girl pink and suit bathing



Src	a brown dog picks up a <UNK> g from a stone surface .	a woman in a mostly black outfit and white helmet riding a bike with blurred trees in the background .	a man sings and plays the guitar into a microphone .
Hyp	 <UNK>		
Hyp (en)	Brown the meat in you must go through grass	bicycle ride on him, saying, this man shirt Red and purple, and bald.	It was, he was a man and song michrophone.



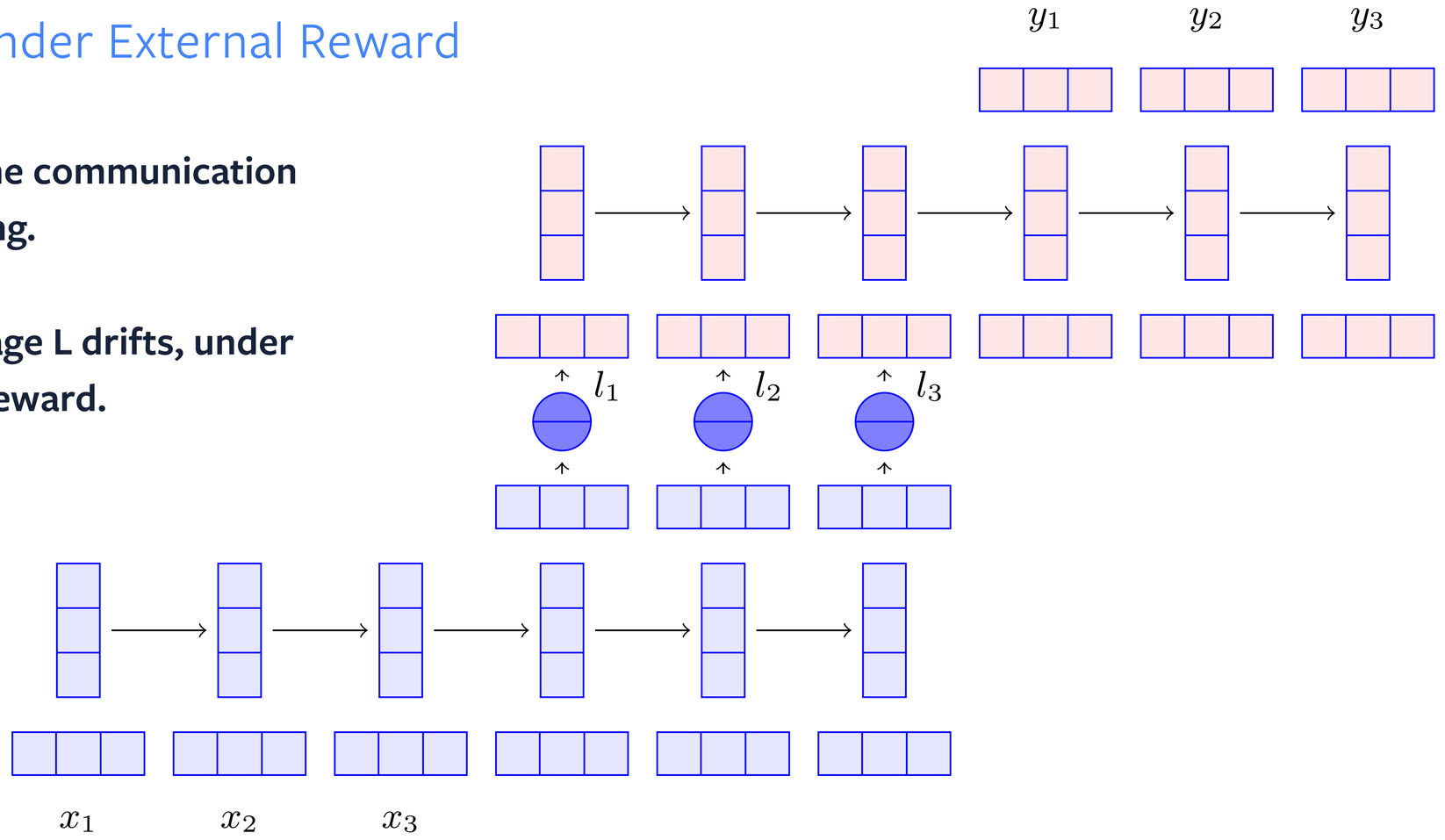
Countering Language Drift via Grounding

Lee, Cho, Kiela. 2018.

Language Drift

Preserving Meaning under External Reward

- **Massive search space in the communication channel. Hence pre-training.**
- **Even if pre-trained, language L drifts, under external (non-linguistic) reward.**



AI Gone Rogue

″FACEBOOK KILLED AN AI AFTER IT CAME UP WITH ITS OWN LANGUAGE″

- **Aside: The world needs to be AI-educated, quickly.**
- **Language drift is to be expected under external reward - no surprise.**
- **But, natural to ask what we can do to could avoid drift. Benefits:**
  - Fine-tuning one big ″language module″
  - Self-play for NLP
  - Playing ″language games″ without drifting: Optimizing for some reward while retaining natural language


THE Sun

THE SUN, A NEWS UK COMPANY

HOME | FOOTBALL | SPORT | TV & SHOWBIZ | NEWS | FABULOUS | MONEY

All Tech | Science | Phones & Gadgets | Gaming

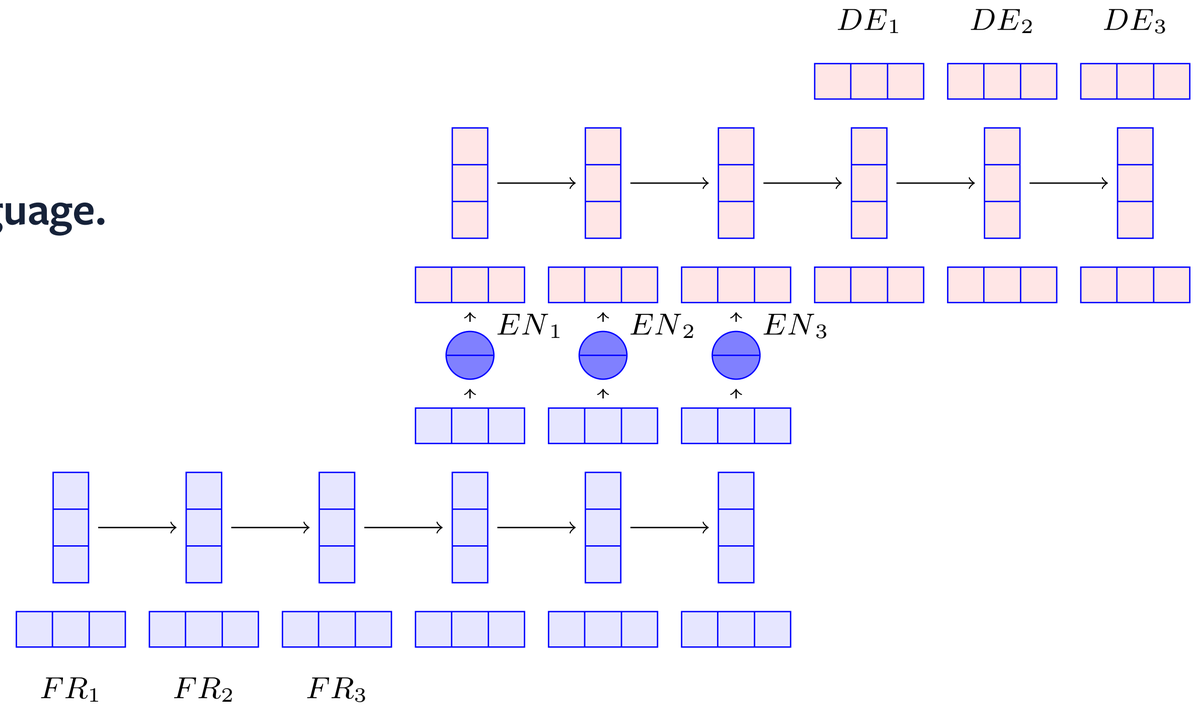
ROBOSTOP **Facebook shuts off AI experiment after two robots begin speaking in their OWN language only they can understand**

Experts have called the incident exciting but also incredibly scary

Countering Language Drift via Grounding

How do we avoid drift?

- **Three-way ”translation game” with English ”pivot” language. Two agents: FR->EN; EN->DE. Fully measurable setup.**
- **Semantic autoencoder of agents’ mental states**
- **Three experiments:**
  - Experiment 1: Policy gradient fine-tuning
  - Experiment 2: Add EN language model constraints
  - Experiment 3: Add Grounding EN by predicting images



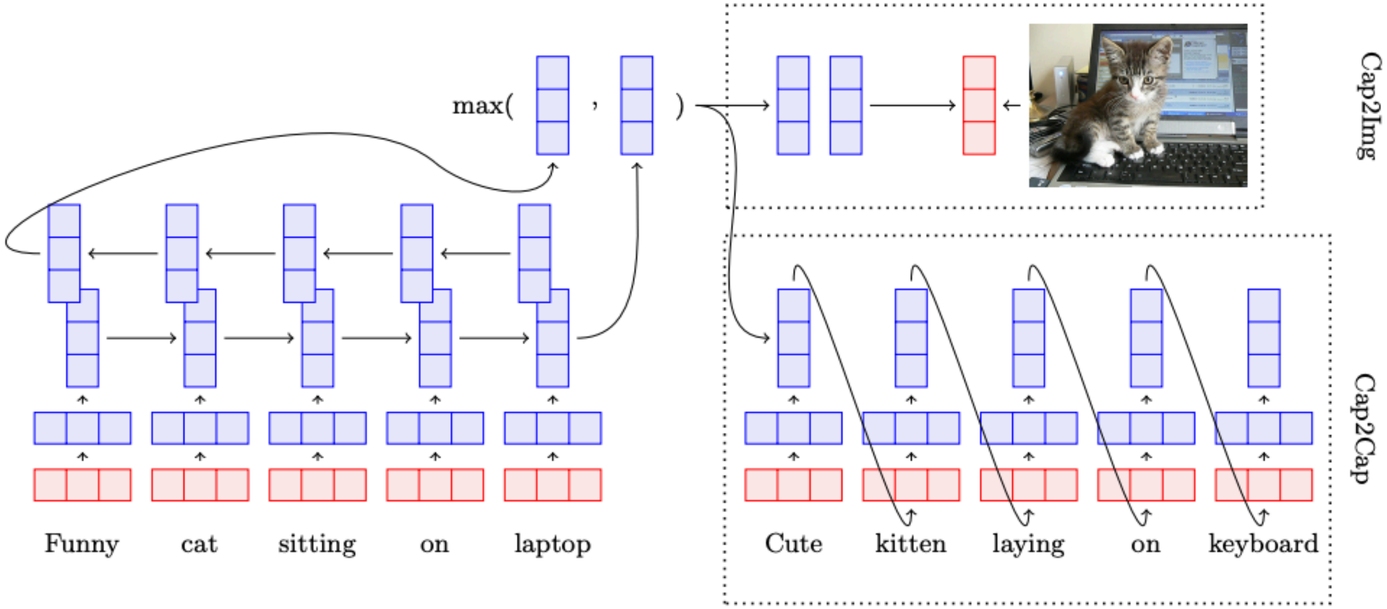
Countering Language Drift via Grounding

How do we ground sentences?

- **GroundSent (Kiela et al., NAACL 2017)**
  (cf. Chrupała et al. 2015; Elliott & Kádár 2017)

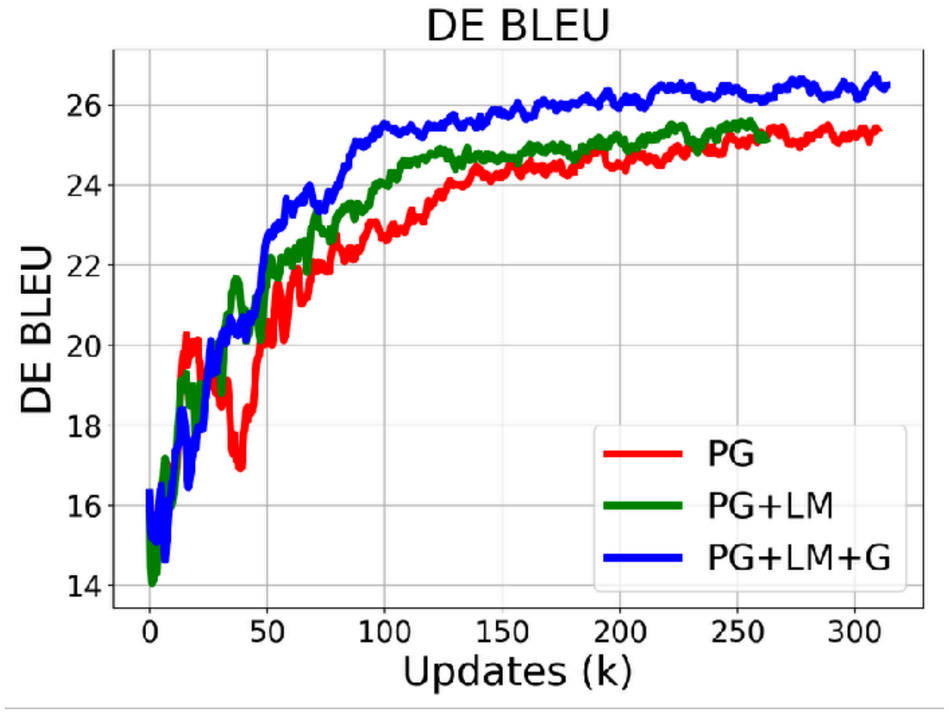
- **Grounding sentence meaning by predicting images (Cap2Img)**

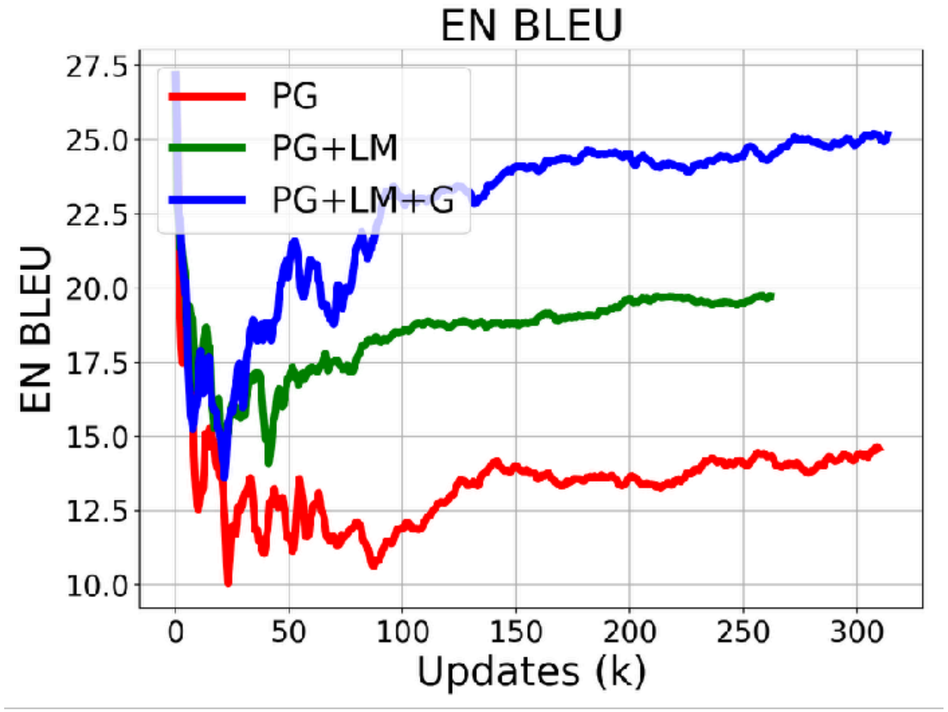
- **You understand a sentence when you can "imagine" its meaning: model theory provided by images.**



Countering Language Drift via Grounding

Language Model Constraints are *Not Enough*





Ref	Fr	un joueur de football américain en blanc et rouge parle à un entraîneur .
	De	ein rot-weiß gekleideter footballspieler spricht mit einem trainer .
	En	a football player in red and white is talking to a coach .
En	PG	a player football american football american and red talking talking a coach a coach
	PG+LM	a player of white and red talking to a coach . ” ” ”
	PG+LM+G	a football player in white and red talking to a coach .
De	PG	ein footballspieler spricht mit einem spieler in einem roten trikot .
	PG+LM	ein weiß gekleideter fußballspieler spricht zu einem trainer .
	PG+LM+G	ein fußballspieler in einem rot-weißen trikot spricht mit einem trainer .

Countering Language Drift via Grounding

Observations

- **Drift happens in all classes:**

	Function words			Content words			
	TO	.	DT	Noun	Verb	Adj	Adv
PG	0.22	0.36	0.57	0.38	0.17	0.32	0.26
PG+LM	0.55	0.84	0.72	0.39	0.18	0.21	0.25
PG+LM+G	0.62	0.88	0.74	0.43	0.26	0.33	0.29

- **Symbols become ungrounded:**

Fr src	un enfant assis sur un rocher.
En ref	a child sitting on a rock formation.
En hyp	a punk sitting sitting on on a broken
De ref	ein kind sitzt auf einem felsen .
De hyp	ein kind sitzt auf einem felsen .

Fr src	un petit enfant est assis à une table, en train de manger un goûter.
En ref	a toddler is sitting at a table eating a snack .
En hyp	a punk sits sitting sitting next next a airline
De ref	ein kleines kind sitzt an einem tisch und isst einen snack .
De hyp	ein kind sitzt an einem tisch und liest ein buch .

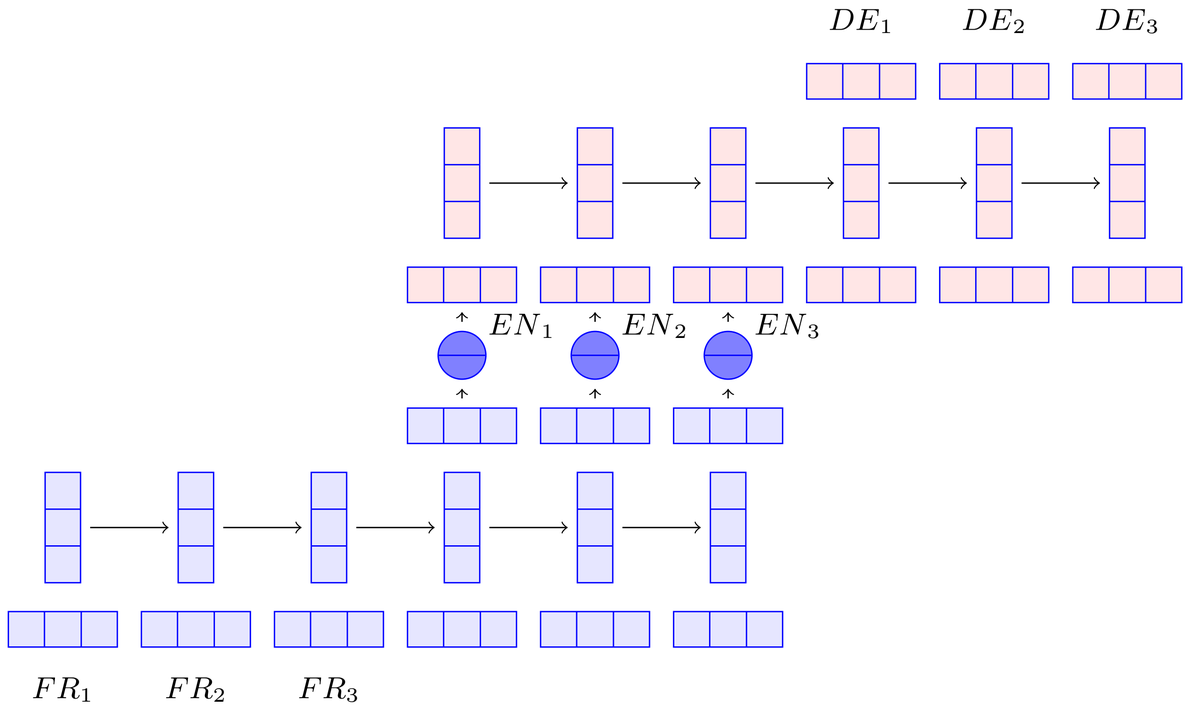
Countering Language Drift via Grounding

Observations

- **Input: ”A giraffe is standing next to a truck”**
  -> ”Democracy is a political system” (?)
  Output: ”A giraffe is standing next to a truck”

- **Constraining syntax is not enough**

- **You need semantic/model-theoretic constraints**
  = IN THIS CASE, PROVIDED VIA GROUNDING



Talk The Walk: Navigating New York City through Grounded Dialogue

De Vries, Batra, Parikh, Weston, Kiela. 2018.

Three Pillars of Meaning

Essential Components of Real Natural Language Semantics

Perception.

We have a shared sensorimotor experience of the world.

Action.

We take actions to manipulate and change the world around us.

Interaction.

We interact with other agents playing "language games".

Talk The Walk

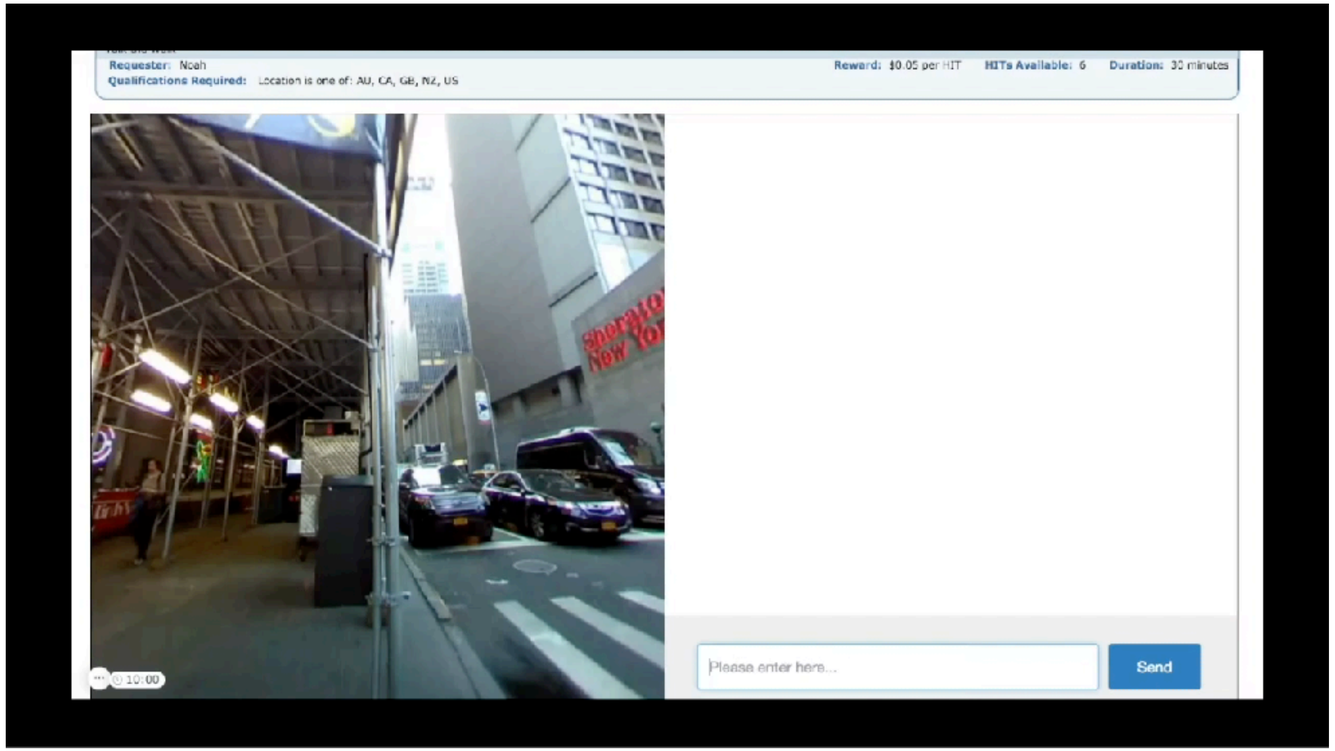
Dataset for Perception, Action & Interaction

- **Tourist and Guide work together to navigate NYC's grid system via dialogue.**

- **"Next Frontier" dataset:**

  - Combines 360 images (perception), navigation (action) and dialogue (interaction) into one single task.

- **Partially observable "navigation game": communication *required*.**



Talk The Walk

Dataset for Perception, Action & Interaction

- **Tourist:** Dynamic Observations (360 images/"perfect perception") + State (dialogue & action history) -> Actions/Speak
- **Guide:** Static Observations (map) + State (dialogue & action history) -> Stop/Speak
- **In the paper:**
  - Perfect perception
  - Emergent language upper-bounds
  - Localization with random walk navigation
  - Natural language baselines
- **A LOT OF WORK TO BE DONE.**
  https://github.com/facebookresearch/talkthewalk/

Conclusion

Teaching Machines to Understand by Using

- **We need:**

  - Active language usage, not passive language observation

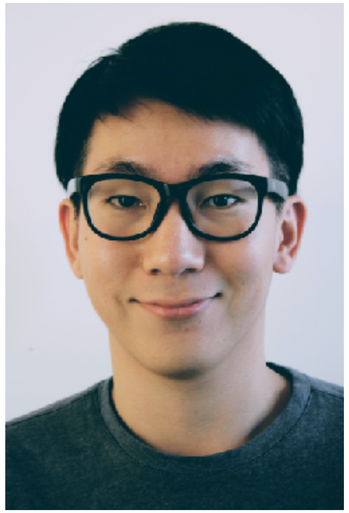
  - Embodied agents grounded in environments

  - => Grounded multi-agent language games that combine perception, action and interaction <= like TTW

  - Recent NLP improvements arguably already take small steps towards more active language usage?

    - ELMo, OpenAI GPT and BERT
    - Backtranslation
    - RL in NLP

Thanks






If this was interesting, come to the **Workshop on Emergent Communication** tomorrow!