Project ATLANTIS
(ArTificial Language uNdersTanding In robotS)
Stephen McGregor – Laboratoire Lattice (ENS/CNRS)

The objective of our project is to understand the first stages of grounded language learning, by which an agent interacting with an unknown environment comes to develop representations that can be applied to productive linguistic communication, and to apply this understanding to the development of artificial linguistic agents.

- Laborattoire Lattice (ENS/CNRS)
- Artificial Intelligence Lab Vrije Universiteit Brussel
- Austrian Research Institute for Artificial Intelligence (OFAI)
- Evolutionary Biology Lab Universitat Pompeu Fabra
- Sony Computer Science Laboratory

Linguistic Flexibility

Language as it is used by humans is fundamentally *flexible*.

Linguistic Flexibility

Language as it is used by humans is fundamentally *flexible*.

- Words are ambiguous: single words can be mapped to multiple extensions, and multiple words can have a single extension.

Bank

Bank

Sofa

Couch

Linguistic Flexibility

Language as it is used by humans is fundamentally *flexible*.

- Words are ambiguous: single words can be mapped to multiple extensions, and multiple words can have a single extension.
- But, more to the point, lexical semantics are specified in the context of a particular situation in the world.

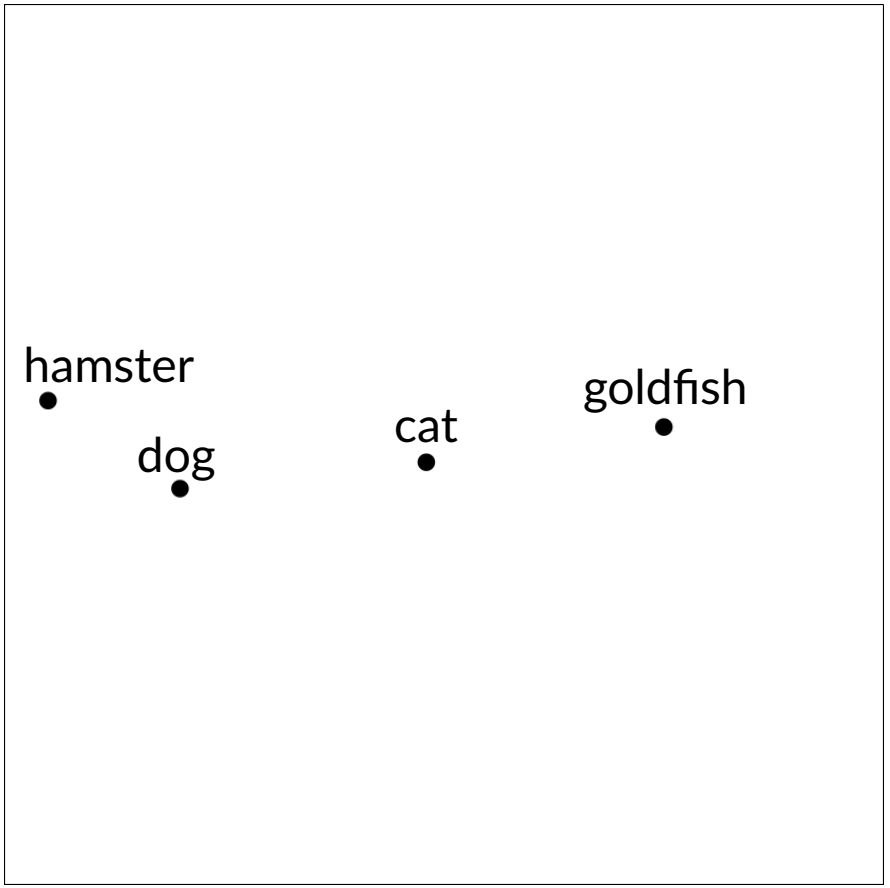
That surgeon was a butcher.

My boss is a bulldozer.

Linguistic Flexibility

We propose that flexibility should be built into lexical semantic representations from the ground up.

The distributional semantic paradigm uses statistics about the way that words co-occur in large scale corpora to represent words as vectors in spaces in which geometry corresponds to semantic properties.

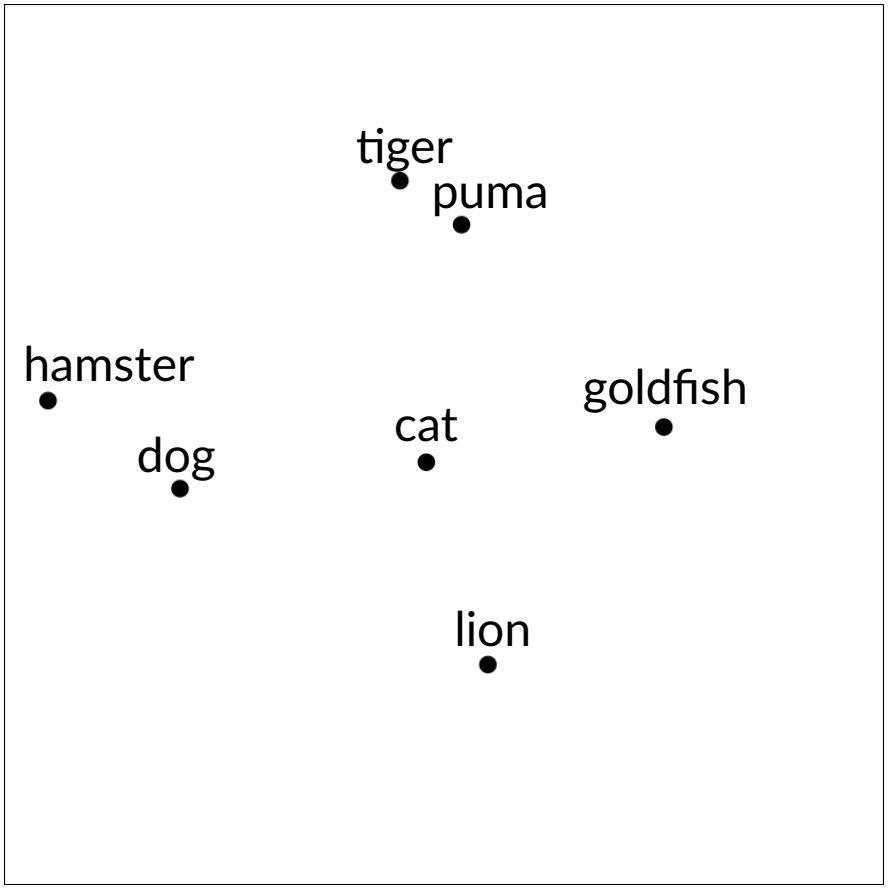


Linguistic Flexibility

We propose that flexibility should be built into lexical semantic representations from the ground up.

The distributional semantic paradigm uses statistics about the way that words co-occur in large scale corpora to represent words as vectors in spaces in which geometry corresponds to semantic properties.

But all the messiness inherent in language in use is also incorporated into these spaces.

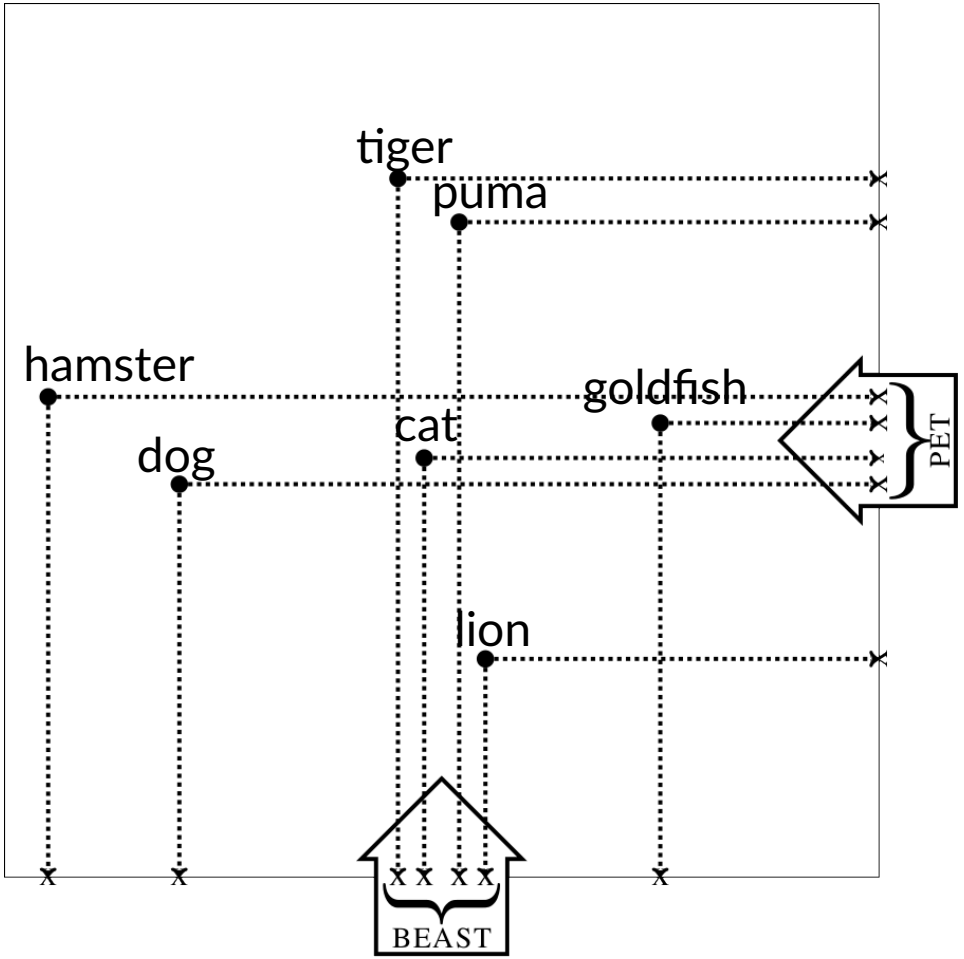


Linguistic Flexibility

We propose that flexibility should be built into lexical semantic representations from the ground up.

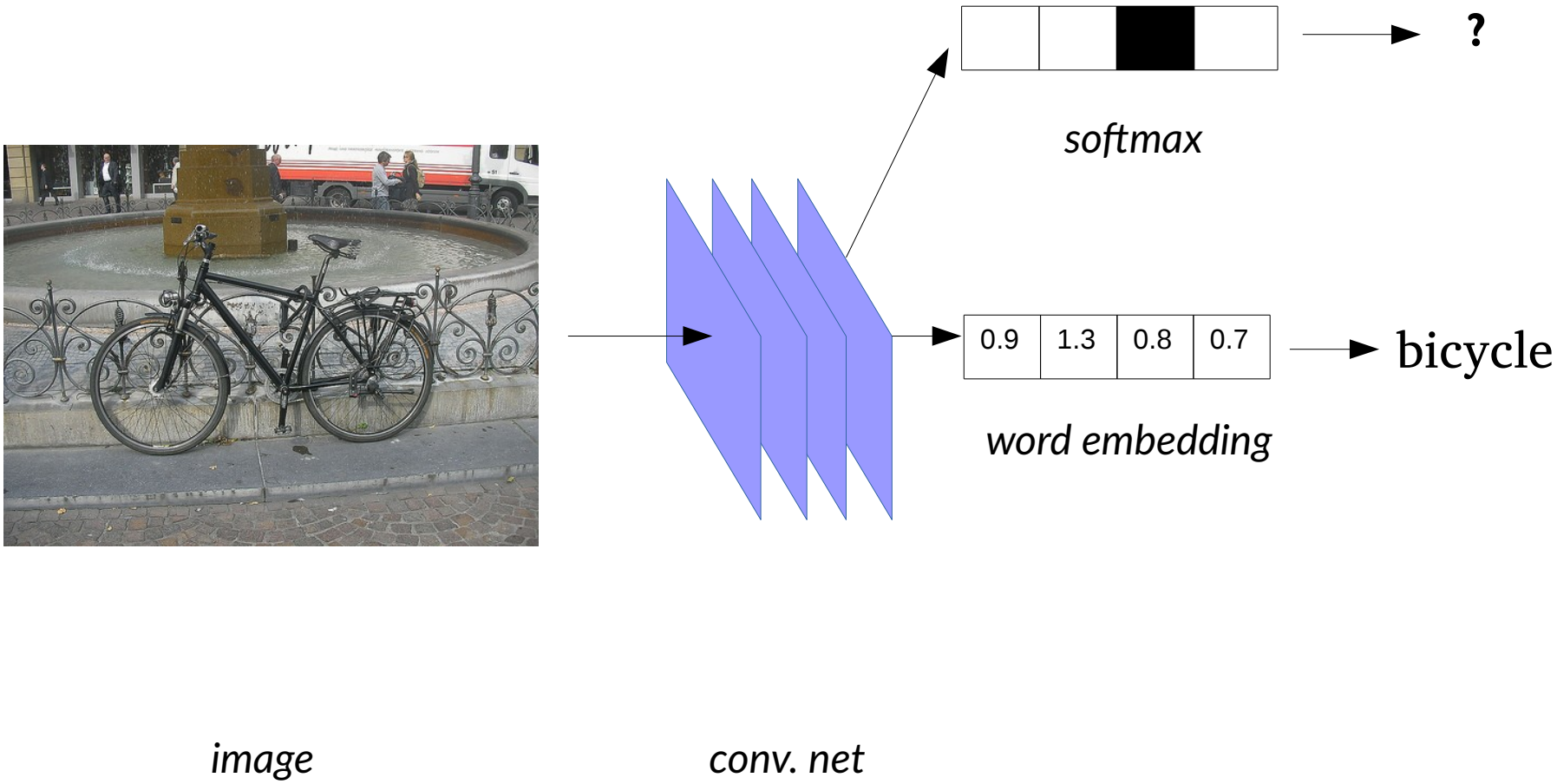
We’ve developed a technique for taking lower dimensional *perspectives* on these noisy semantic spaces in order to reflect the way that meaning arises in a particular context.

In this way, we can project ad hoc semantic geometries in response to on-line input.



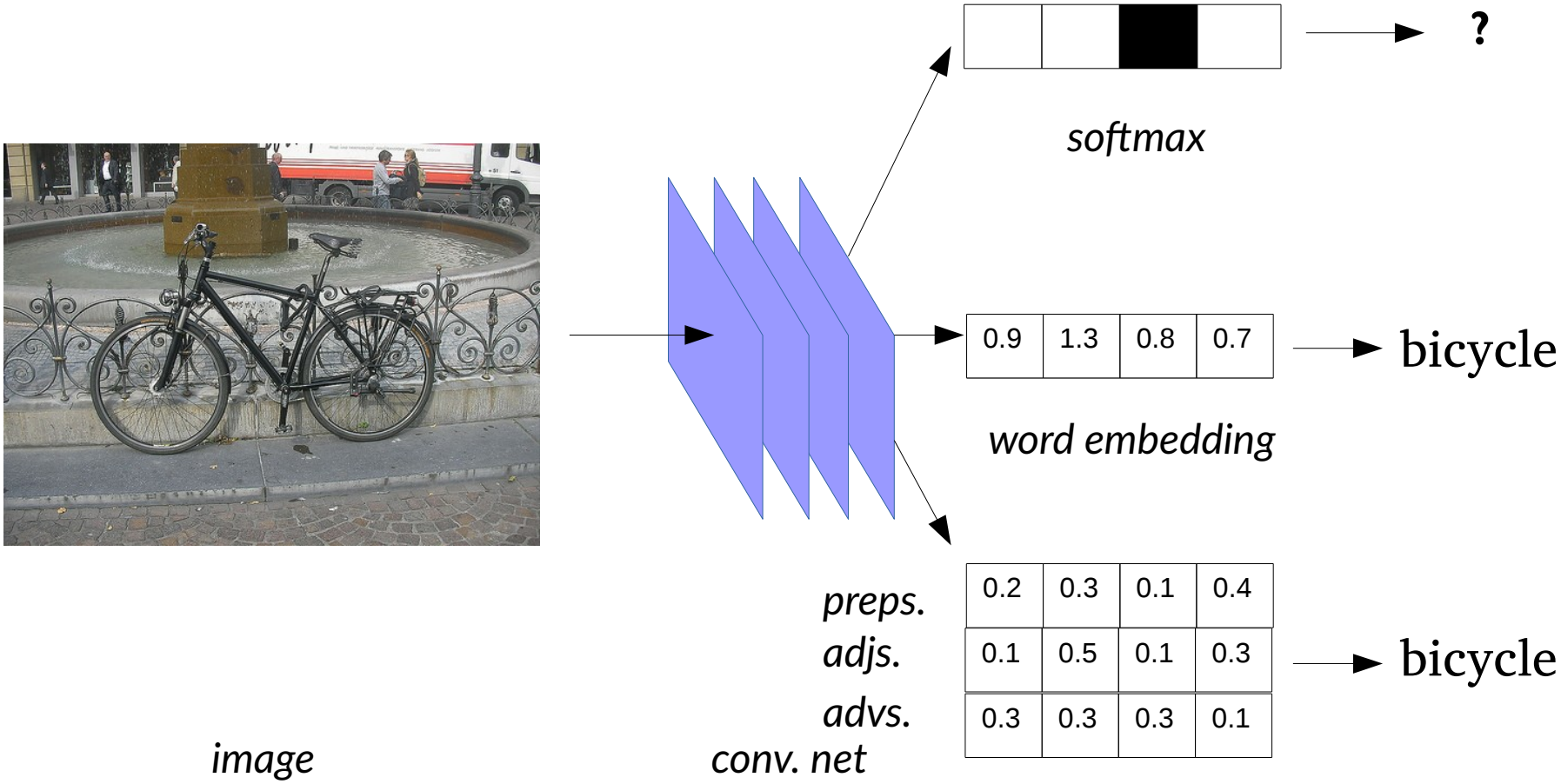
Affordances in Grounded Language Learning

It has been demonstrated that distriutional semantic representations can facilitate the grounded language learning task of learning to classify images.



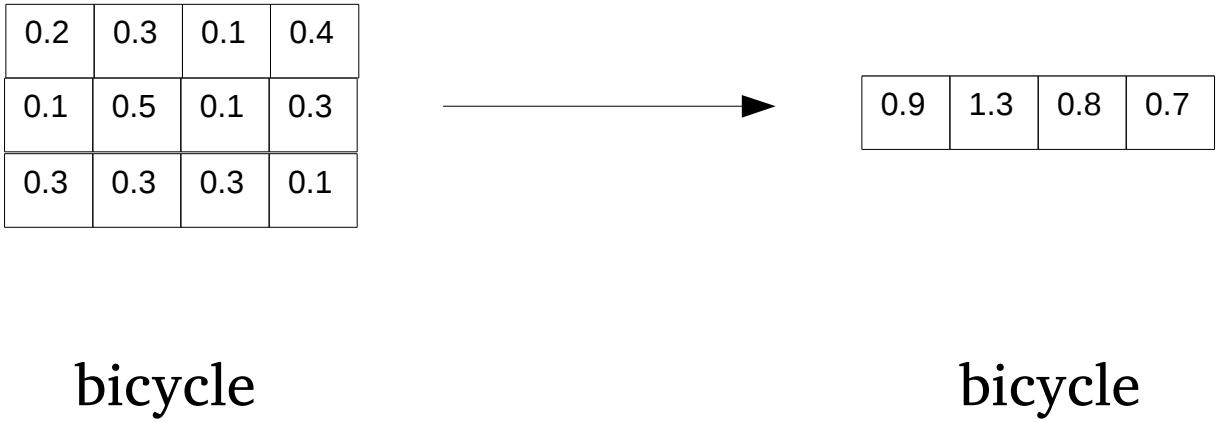
Affordances in Grounded Language Learning

We show that comparable results can be found by mapping network output to matrices of probabilities that represent co-occurrences in particular dependency relationships, which we argue weakly simulate *affordances* (McGregor and Lim, 2018).



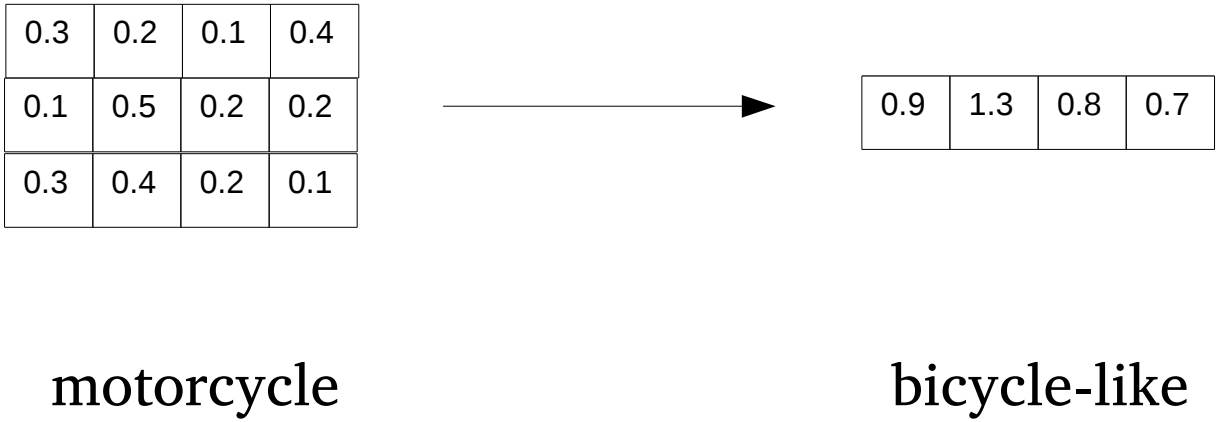
Affordances in Grounded Language Learning

Furthermore, we show that we can learn to systematically map from concrete matrices of co-occurrence distributions to word embedding type representations, meaning that if we can learn to identify some of the basic affordances of unseen objects, we can move towards a true model of zero shot learning in which an agent can make a pretty good guess at the general category to which a new object encountered in its environment belongs.



Affordances in Grounded Language Learning

Furthermore, we show that we can learn to systematically map from concrete matrices of co-occurrence distributions to word embedding type representations, meaning that if we can learn to identify some of the basic affordances of unseen objects, we can move towards a true model of zero shot learning in which an agent can make a pretty good guess at the general category to which a new object encountered in its environment belongs.



Interpreting Semantic Type Coercion

We've used this technique to model semantic type coercion, in which a verb that typically selects for one type takes an object of another type as an argument (McGregor and Jezek, 2019).

The student finished the book.

Interpreting Semantic Type Coercion

We build context-specific probability distributions of particular grammatical classes (adjectives and prepositions in particular) associated with the component of a sentence, and then use these distributions to construct tensor that project the object into an interpretive semantic space.

smart
tired
poor

$$\begin{array}{cc} & \text{distribution of objects over } A \\ \otimes & \begin{array}{ccc} n_A^1 & n_A^2 & n_A^k \end{array} \\ \text{distribution of subjects over } A \begin{array}{c} s_A^1 \\ s_A^2 \\ \\ s_A^k \end{array} & \begin{bmatrix} s_A^1 \times n_A^1 & s_A^1 \times n_A^2 \cdots s_A^k \times n_A^1 \\ s_A^2 \times n_A^1 & s_A^2 \times n_A^2 \cdots s_A^k \times n_A^2 \\ \vdots & \vdots \quad \ddots \quad \vdots \\ s_A^k \times n_A^1 & s_A^k \times n_A^2 \cdots s_A^k \times n_A^k \end{bmatrix} \end{array} \times \begin{bmatrix} w_{i1} \\ w_{i2} \\ \vdots \\ w_{ik} \end{bmatrix} = \begin{bmatrix} z_{i1}, z_{i2} \ldots z_{ik} \end{bmatrix}$$

tensor T *contextual word-vector* *vector in Z*

The student finished the book.

Interpreting Semantic Type Coercion

We build context-specific probability distributions of particular grammatical classes (adjectives and prepositions in particular) associated with the component of a sentence, and then use these distributions to construct tensor that project the object into an interpretive semantic space.

smart
tired
poor

$$\begin{array}{cc} & \text{distribution of objects over } A \\ \otimes & \begin{array}{ccc} n_A^1 & n_A^2 & n_A^k \end{array} \\ \text{distribution of subjects over } A \begin{array}{c} s_A^1 \\ s_A^2 \\ \\ s_A^k \end{array} & \begin{bmatrix} s_A^1 \times n_A^1 & s_A^1 \times n_A^2 \cdots s_A^k \times n_A^1 \\ s_A^2 \times n_A^1 & s_A^2 \times n_A^2 \cdots s_A^k \times n_A^2 \\ \vdots & \vdots \quad \ddots \quad \vdots \\ s_A^k \times n_A^1 & s_A^k \times n_A^2 \cdots s_A^k \times n_A^k \end{bmatrix} \end{array} \times \begin{bmatrix} w_{i1} \\ w_{i2} \\ \vdots \\ w_{ik} \end{bmatrix} = \begin{bmatrix} z_{i1}, z_{i2} \ldots z_{ik} \end{bmatrix}$$

tensor T *contextual word-vector* *vector in Z*

The student finished [*reading*] the book.

Interpreting Semantic Type Coercion

We build context-specific probability distributions of particular grammatical classes (adjectives and prepositions in particular) associated with the component of a sentence, and then use these distributions to construct tensor that project the object into an interpretive semantic space.

famous
eccentric
unknown

$$
\begin{array}{cc}
 & \begin{array}{cccc} \otimes & \multicolumn{3}{c}{\textit{distribution of objects over } A} \\ & n_A^1 & n_A^2 & n_A^k \end{array} \\
\textit{distribution of subjects over } A \begin{array}{c} s_A^1 \\ s_A^2 \\ \\ s_A^k \end{array} & \begin{bmatrix} s_A^1 \times n_A^1 & s_A^1 \times n_A^2 \cdots & s_A^k \times n_A^1 \\ s_A^2 \times n_A^1 & s_A^2 \times n_A^2 \cdots & s_A^k \times n_A^2 \\ \vdots & \vdots \quad \ddots & \vdots \\ s_A^k \times n_A^1 & s_A^k \times n_A^2 \cdots & s_A^k \times n_A^k \end{bmatrix}
\end{array}
\times
\begin{bmatrix} w_{i1} \\ w_{i2} \\ \vdots \\ w_{ik} \end{bmatrix}
= \begin{bmatrix} z_{i1}, z_{i2} \ldots z_{ik} \end{bmatrix}
$$

tensor T *contextual word-vector* *vector in Z*

The student finished [*reading*] the book.

The author finished the book.

Interpreting Semantic Type Coercion

We build context-specific probability distributions of particular grammatical classes (adjectives and prepositions in particular) associated with the component of a sentence, and then use these distributions to construct tensor that project the object into an interpretive semantic space.

famous
eccentric
unknown

$$\begin{array}{cc} & \text{distribution of objects over } A \\ \otimes & \begin{array}{ccc} n_A^1 & n_A^2 & n_A^k \end{array} \\ \text{distribution of subjects over } A \begin{array}{c} s_A^1 \\ s_A^2 \\ \\ s_A^k \end{array} & \begin{bmatrix} s_A^1 \times n_A^1 & s_A^1 \times n_A^2 \cdots s_A^k \times n_A^1 \\ s_A^2 \times n_A^1 & s_A^2 \times n_A^2 \cdots s_A^k \times n_A^2 \\ \vdots & \vdots \quad \ddots \quad \vdots \\ s_A^k \times n_A^1 & s_A^k \times n_A^2 \cdots s_A^k \times n_A^k \end{bmatrix} \end{array} \times \begin{bmatrix} w_{i1} \\ w_{i2} \\ \vdots \\ w_{ik} \end{bmatrix} = \begin{bmatrix} z_{i1}, z_{i2} \ldots z_{ik} \end{bmatrix}$$

tensor T *contextual word-vector* *vector in Z*

The student finished [*reading*] the book.

The author finished [*writing*] the book.

Interpreting Semantic Type Coercion

We build context-specific probability distributions of particular grammatical classes (adjectives and prepositions in particular) associated with the component of a sentence, and then use these distributions to construct tensor that project the object into an interpretive semantic space.

stubborn
old
billy

$$\begin{array}{cc} & \text{distribution of objects over } A \\ \otimes & \begin{array}{ccc} n_A^1 & n_A^2 & n_A^k \end{array} \\ \text{distribution of subjects over } A \begin{array}{c} s_A^1 \\ s_A^2 \\ \\ s_A^k \end{array} & \begin{bmatrix} s_A^1 \times n_A^1 & s_A^1 \times n_A^2 \cdots s_A^k \times n_A^1 \\ s_A^2 \times n_A^1 & s_A^2 \times n_A^2 \cdots s_A^k \times n_A^2 \\ \vdots & \vdots \quad \ddots \quad \vdots \\ s_A^k \times n_A^1 & s_A^k \times n_A^2 \cdots s_A^k \times n_A^k \end{bmatrix} \end{array} \times \begin{bmatrix} w_{i1} \\ w_{i2} \\ \vdots \\ w_{ik} \end{bmatrix} = \begin{bmatrix} z_{i1}, z_{i2} \ldots z_{ik} \end{bmatrix}$$

tensor T *contextual word-vector* *vector in Z*

The student finished [*reading*] the book.

The author finished [*writing*] the book.

The goat finished the book.

Interpreting Semantic Type Coercion

We build context-specific probability distributions of particular grammatical classes (adjectives and prepositions in particular) associated with the component of a sentence, and then use these distributions to construct tensor that project the object into an interpretive semantic space.

stubborn
old
billy

$$\begin{array}{cc} & \text{distribution of objects over } A \\ \otimes & \begin{array}{ccc} n_A^1 & n_A^2 & n_A^k \end{array} \\ \text{distribution of subjects over } A \begin{array}{c} s_A^1 \\ s_A^2 \\ \\ s_A^k \end{array} & \begin{bmatrix} s_A^1 \times n_A^1 & s_A^1 \times n_A^2 \cdots s_A^k \times n_A^1 \\ s_A^2 \times n_A^1 & s_A^2 \times n_A^2 \cdots s_A^k \times n_A^2 \\ \vdots & \vdots \quad \ddots \quad \vdots \\ s_A^k \times n_A^1 & s_A^k \times n_A^2 \cdots s_A^k \times n_A^k \end{bmatrix} \end{array} \times \begin{bmatrix} w_{i1} \\ w_{i2} \\ \vdots \\ w_{ik} \end{bmatrix} = \begin{bmatrix} z_{i1}, z_{i2} \ldots z_{ik} \end{bmatrix}$$

tensor T *contextual word-vector* *vector in Z*

The student finished [*reading*] the book.

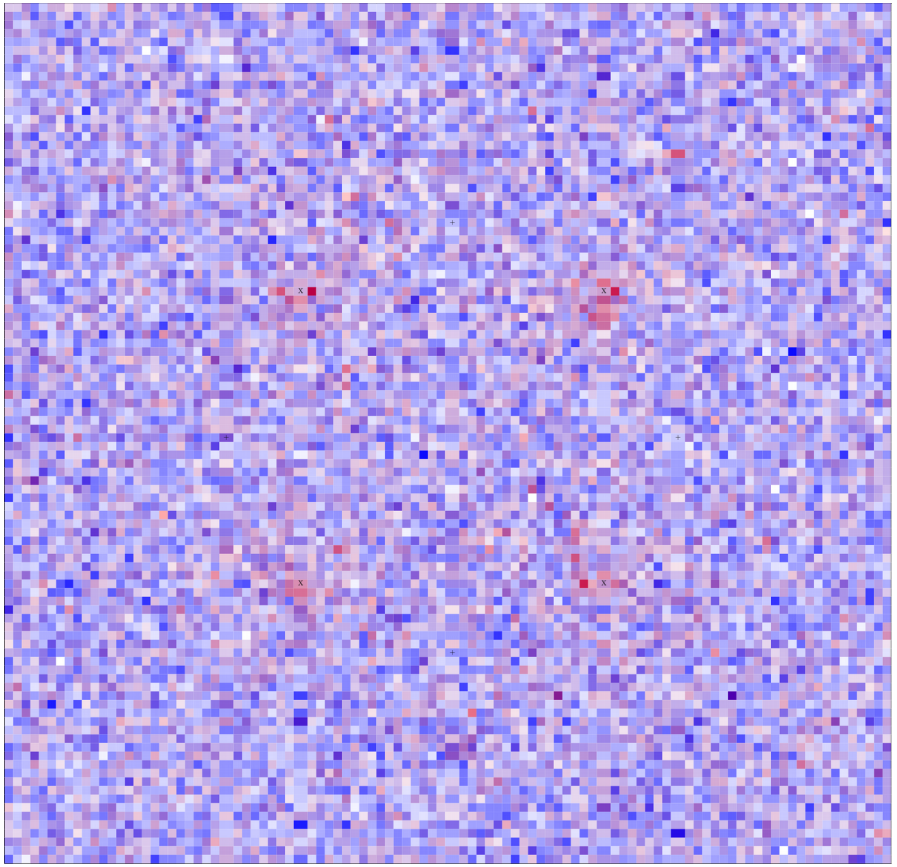
The author finished [*writing*] the book.

The goat finished [*eating*] the book.

Embodied Linguistic Agents

We propose that, for semantic representations to acquire their flexibility from the ground up, they should be semiotically entwined with the physiognomy of the agents that use them (McGregor and Poibeau 2019).

To explore this idea, we've designed a series of simulations in which very simple agents existing in a grid-like world learn to communicate with one other using their bodies.

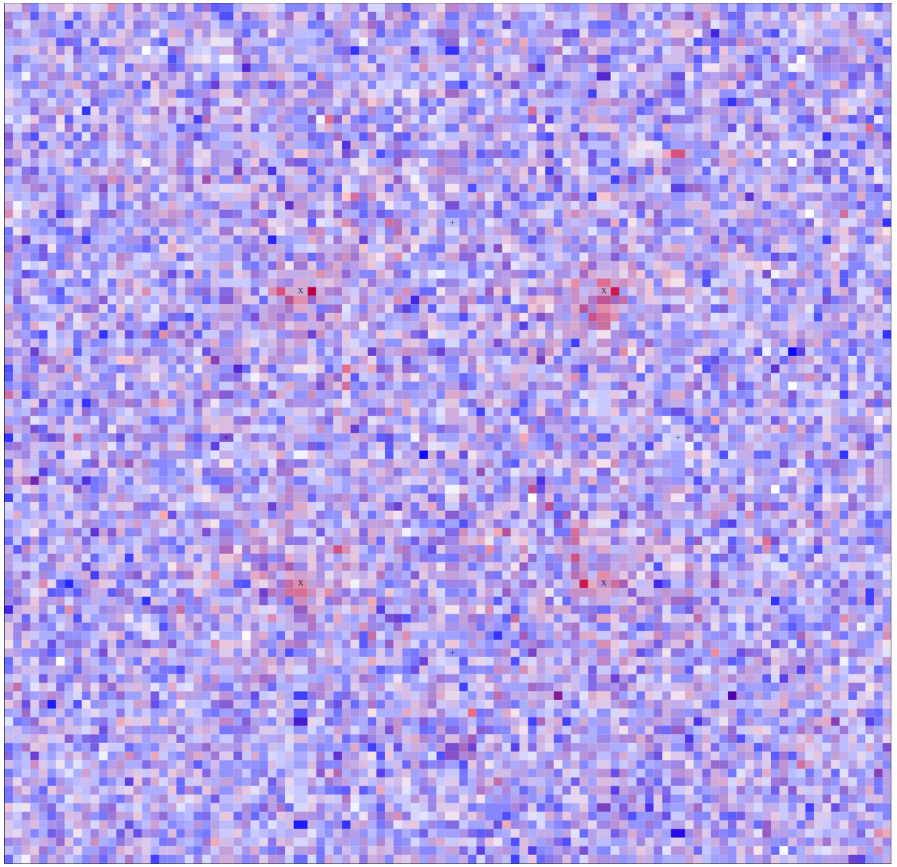


Embodied Linguistic Agents

We propose that, for semantic representations to acquire their flexibility from the ground up, they should be semiotically entwined with the physiognomy of the agents that use them (McGregor and Poibeau 2019).

The rules of this world are straightforward:

- Agents have no internal representations.
- Agents have no pre-defined goals; they simply survive or parish based on their actions in the world.
- Agents have no access to a global sense of the situation in the world.



Human Robot Interactions

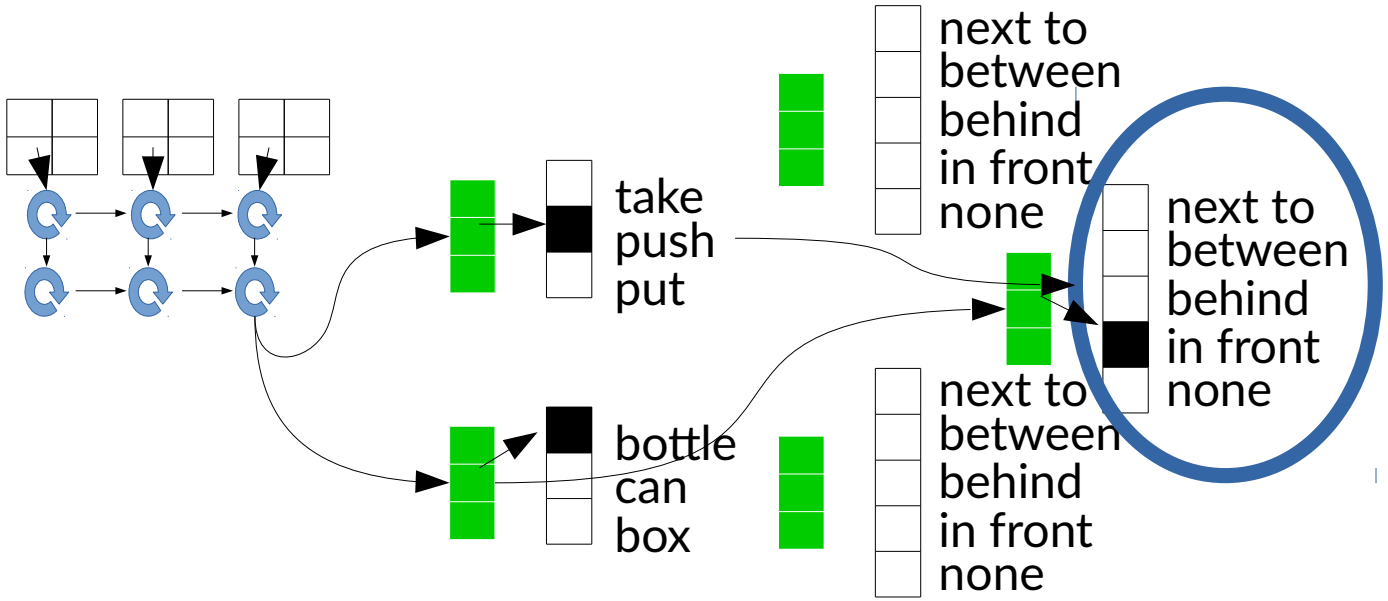
We have modelled a dataset of videos involving humans teaching robots some basic actions in the world, developed by our colleagues at OFAI (Gross et al., 2018).

The data involves humans using a VR interface to explain some straightforward actions on a small set of objects in a simple environment.

The idea is to explore the role of a *tutor* in the process of an artificial agent acquiring semantic representations.

Human Robot Interactions

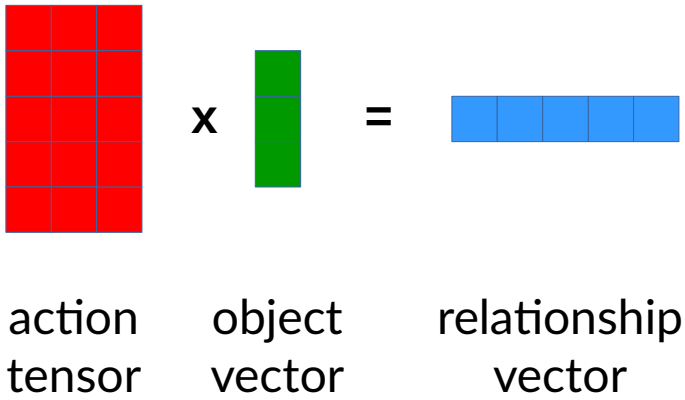
We've built a neural network model that learns to predict actions, objects on which these actions are performed, and resulting spatial relationships based on the videos.



Human Robot Interactions

By incorporating the network with our idea of context-specific distributional representations, we are able to explore the way that an agent might begin to acquire a system of *compositional* symbols.

The key component of this methodology is that the model learns to represent actions as contextually activated *transformations* that project object representations onto relationship representations.



A Symposium on Language Learning for Artificial Agents

Finally, we will be holding a seminar on Language Learning for Artificial Agents (L2A2) at the AISB conference in Falmouth later this month (16-18 April).

- A keynote lecture from Prof. Tony Cohn (Leeds).
- Five presentations based on peer-reviewed submissions.
- An invited talk from Dr. Yasemin Erden (St. Mary's), followed by a panel session on "Language Games and Language Technology".

Please join us!

References

Stephanie Gross, Matthias Hirschmanner, Brigitte Krenn, Friedrich Neubarth and Michael Zillich. 2018. Action Verb Corpus. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC).*

Stephen McGregor and Elisabetta Jezek. 2019. A Distributional Model of Affordances in Semantic Type Coercion. *Proceedings of the International Workshop on Computational Semantics (IWCS).*

Stephen McGregor and KyungTae Lim. 2018. Affordances in Grounded Language Learning. *COGACLL.*

Stephen McGregor and Thierry Poibeau. 2019. Semantic Flexibility and Grounded Language Learning. *AISB Symposium on Language Learning for Artificial Agents (L2A2).*