Affordance Extraction and Inference based on Semantic Role Labeling

Daniel Loureiro, Alípio Jorge
University of Porto

Fact Extraction and Verification (FEVER) Workshop
EMNLP 2018

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO
[dcc]
LIAAD
INESCTEC
LABORATÓRIO ASSOCIADO
U.PORTO

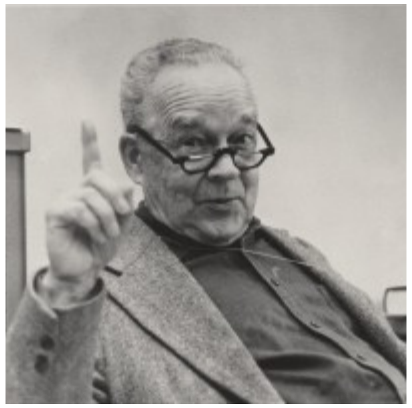
Overview

1. Affordances → What are they and why are they relevant?
2. FEVER → How may this relate to FEVER? (Suggestions)
3. Affordance Extraction
4. Affordance Inference

→ Method

5. Evaluation
6. Conclusions

What is affordance?

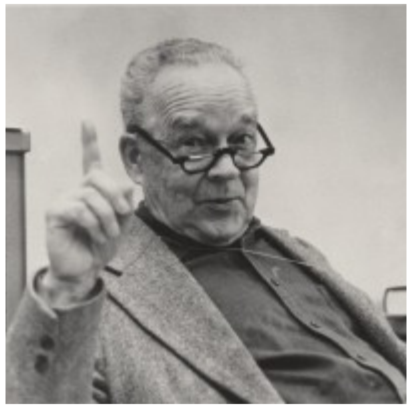

Gibson 1979


Norman 1988


Glenberg 2000

Depends on who you ask.

What is affordance?



Gibson 1979

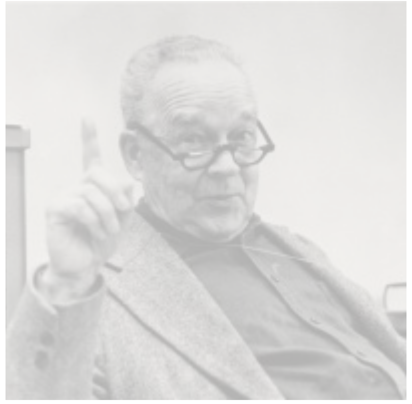
Norman 1988

Glenberg 2000

Psychology

Affordance: What the environment provides the animal.

What is affordance?



Gibson 1979

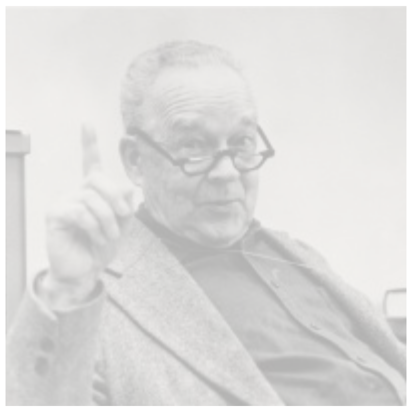
Norman 1988

Glenberg 2000

Design

Affordance: Perceived action possibilities (suggestive).

What is affordance?



Gibson 1979

Norman 1988

Glenberg 2000

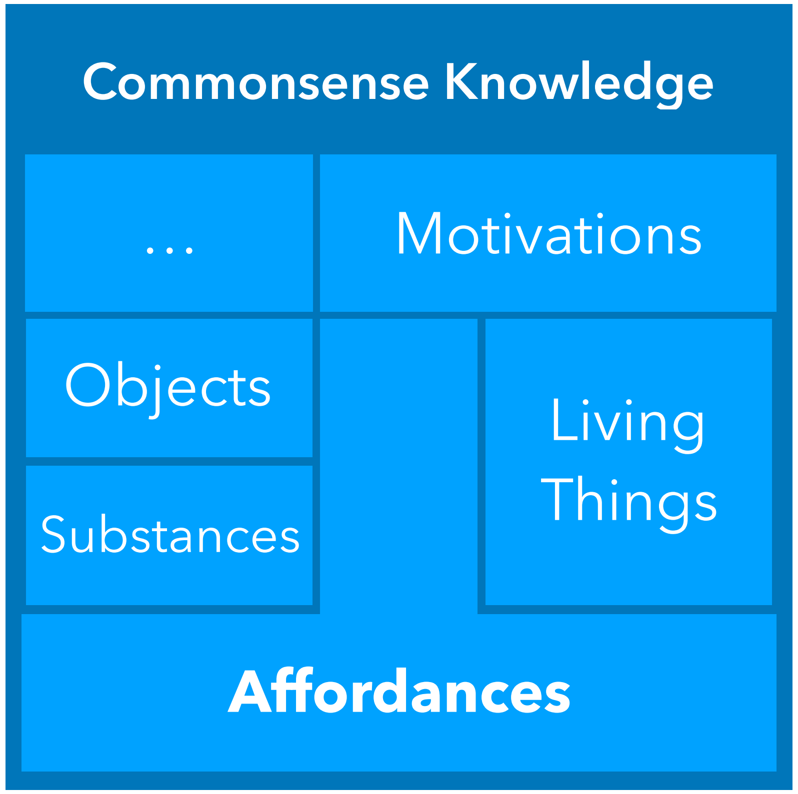
Language

Affordance: Basis for grounding meaning under the Indexical Hypothesis.

=

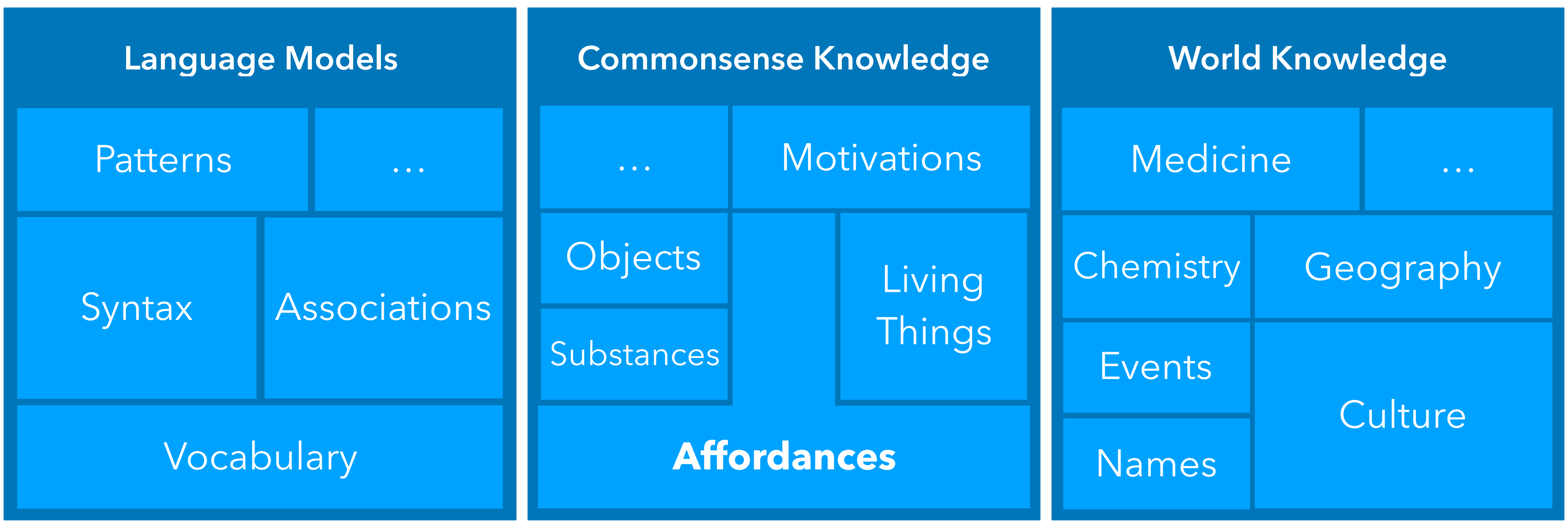
Why affordance?

- Commonsense acquisition and representation in Distributional Semantic Models is still an open question [Camacho-Collados, Pilhevar 2018].
- Affordances are a **relational** component of Commonsense Knowledge.



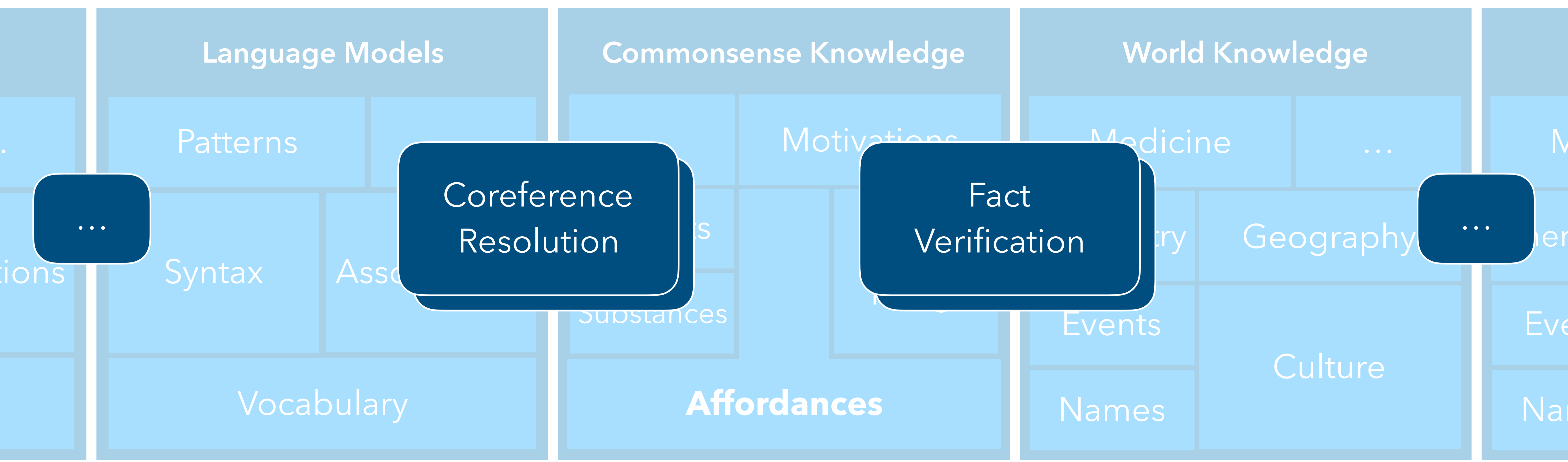
Why affordance?

- Commonsense acquisition and representation in Distributional Semantic Models is still an open question [Camacho-Collados, Pilhevar 2018].
- Affordances are a **relational** component of Commonsense Knowledge.



Why affordance?

- Commonsense acquisition and representation in Distributional Semantic Models is still an open question [Camacho-Collados, Pilhevar 2018].
- Affordances are a **relational** component of Commonsense Knowledge.



Fact Extraction

ugh Whitemore about Turing. The play ran in London's West End beginning in November 1986 and on Broadway from 15
ese performances Turing was played by Derek Jacobi. The Broadway production was nominated for three Tony Awards
atured Actor in a Play, and Best Direction of a Play, and for two Drama Desk Awards, for Best Actor and Best Featured
acobi in the 1996 television film adaptation of *Breaking the Code*.[231]

nnial, American Lyric Theater commissioned an operatic exploration of the life and death of Turing from composer Justine
232] Titled *The Life and Death(s) of Alan Turing*, the opera is a historical fantasia on the life of Turing. In November 2014,
rks inspired by Turing's life were featured on Studio 360.[233] The opera received its first public performance in January

Turing police have jurisdiction over AIs. (1984)[235]

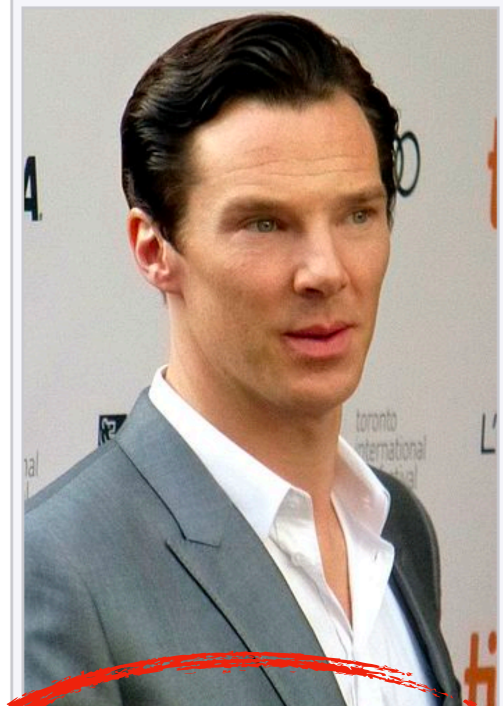
son novel *Cryptonomicon* (1999).[236]

g Test features Turing and the writer Graham Greene.[237]

Turing Machines contrasts fictionalised accounts of the lives and ideas of Turing and Kurt Gödel.[238]

sa Hall, includes a series of fictional letters written from Turing to his best friend's mother throughout his life, detailing his
9][240]

ch a fictionalised version of WWII plays out involving superhuman soldiers called "Tank-Men", Turing is one of the researchers as well as a Tank-Man



Benedict Cumberbatch portrayed Turing in the 2014 film *The Imitation Game*

https://en.wikipedia.org/wiki/Alan_Turing

Fact Extraction

Benedict Cumberbatch <u>portrayed</u> Turing in The Imitation Game.

With good statistics on affordances, you can infer additional extractions:

- Those who portray usually personify.
  - Benedict Cumberbatch personified Turing.
- Things portrayed are usually film characters.
  - Turing is a film character. (not exclusive)
- Places where portrayal occurs are usually films.
  - The Imitation Game is a film.

Fact Extraction

Benedict Cumberbatch portrayed Turing in The Imitation Game.

With good statistics on affordances, you can infer additional extractions:

- Those who portray usually personify.
  - Benedict Cumberbatch personified Turing.
- Things portrayed are usually film characters.
  - Turing is a film character. (not exclusive)
- Places where portrayal occurs are usually films.
  - The Imitation Game is a film.

cf. Selectional Preferences, Argument Typicality, Frame Semantics.

Fact Verification

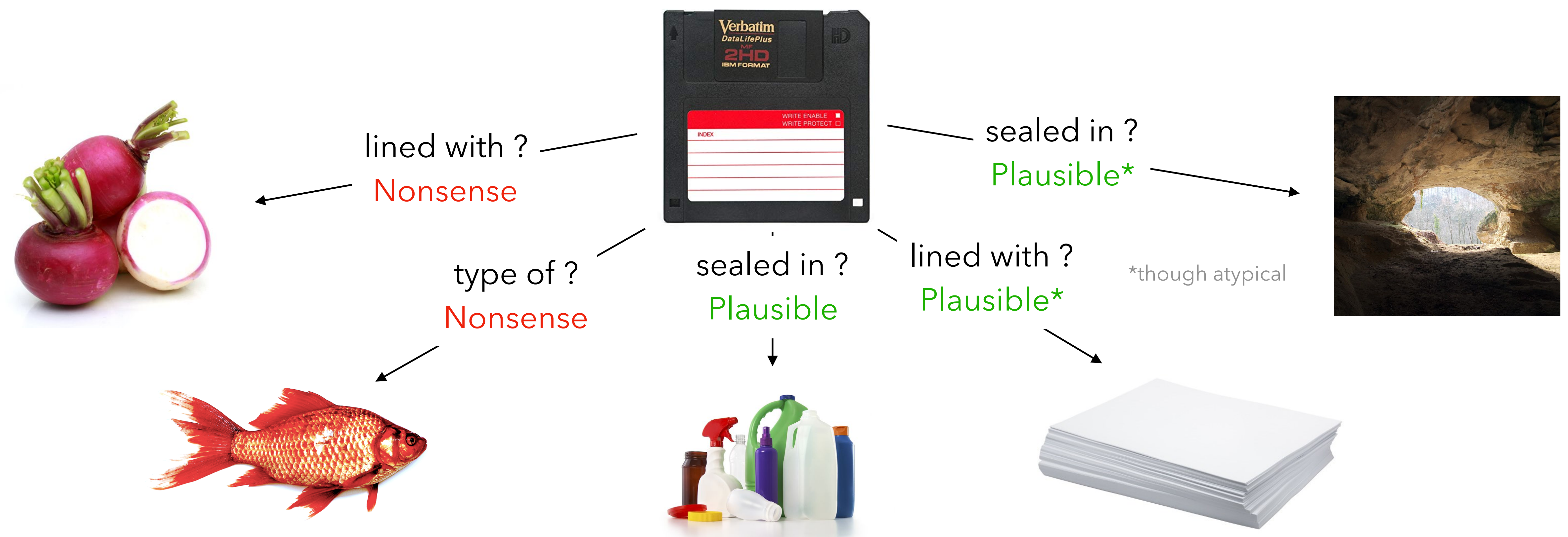
Example claims from the FEVER dataset:

- A Floppy disk is lined with turnips.
- A Floppy disk is a type of fish.
- A Floppy disk is sealed in a cave.
- A Floppy disk is lined with paper.
- A Floppy disk is sealed in plastic.

Fact Verification

Example claims from the FEVER dataset:

- A Floppy disk is lined with turnips.
- A Floppy disk is a type of fish.
- A Floppy disk is sealed in a cave.
- A Floppy disk is lined with paper.
- A Floppy disk is sealed in plastic.



Fact Verification

Semantic Plausibility as a **prior bias** for Fact Verification.

- If **implausible** (i.e. nonsense):
  - Probably refutable and no explicit evidence.

    E.g. “A Floppy disk is a type of fish.”
- If **plausible** and **typical** (i.e. obvious):
  - Probably supported with implicit evidence.

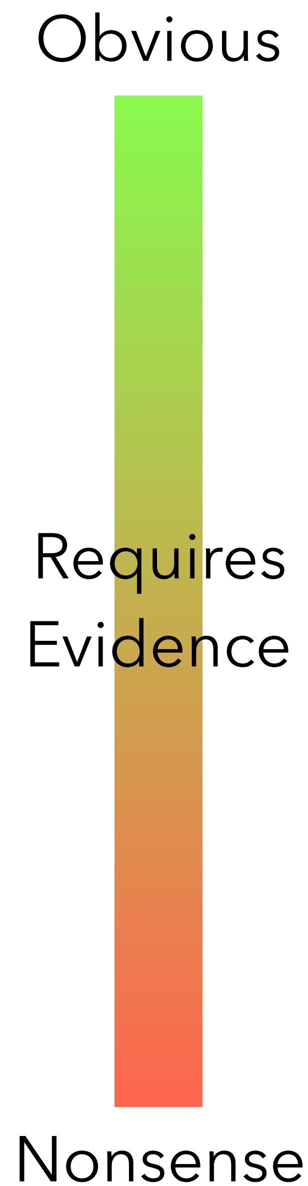
    E.g. “Dan Trachtenberg is a person.”
- If **plausible** and **atypical** (i.e. others):
  - Unknown refutability, explicit evidence should exist.

    E.g. “Sarah Hyland is a New Yorker.”

Fact Verification

Semantic Plausibility as a **prior bias** for Fact Verification.

- If **implausible** (i.e. nonsense):
  - Probably refutable and no explicit evidence.
    E.g. "A Floppy disk is a type of fish."
- If **plausible** and **typical** (i.e. obvious):
  - Probably supported with implicit evidence.
    E.g. "Dan Trachtenberg is a person."
- If **plausible** and **atypical** (i.e. others):
  - Unknown refutability, explicit evidence should exist.
    E.g. "Sarah Hyland is a New Yorker."



Intuition: Plausibility should be easier to assess than Truth.

Affordance Extraction

Affordance Representation:
Every symbol (i.e. token) is represented by a vector whose dimensions signal affordances.

	Can eat ?	Can jump ?	Used for riding ?	Place for getting lost?
dog	Yes	Yes	No	No
cat	Yes	Yes	No	No
horse	Yes	Yes	Yes	No
brussels	No	No	No	Yes
thought	No	No	No	No

Assignment

Affordance Extraction

Affordance Representation:
Every symbol (i.e. token) is represented by a vector whose dimensions signal affordances.

	Can eat ?	Can jump ?	Used for riding ?	Place for getting lost?
dog	1.0	1.0	0.2	0.0
cat	1.0	1.0	0.0	0.0
horse	1.0	0.8	1.0	0.0
brussels	0.2	0.0	0.0	1.0
thought	0.0	0.2	0.0	0.2

Assignment > Grading

Affordance Extraction

Affordance Representation:
Every symbol (i.e. token) is represented by a vector whose dimensions signal affordances.

	eat | AGENT	jump | AGENT	ride | PATIENT	lose | LOCATION
dog	1.0	1.0	0.2	0.0
cat	1.0	1.0	0.0	0.0
horse	1.0	0.8	1.0	0.0
brussels	0.2	0.0	0.0	1.0
thought	0.0	0.2	0.0	0.2

Assignment > Grading > Formalizing

Affordance Extraction

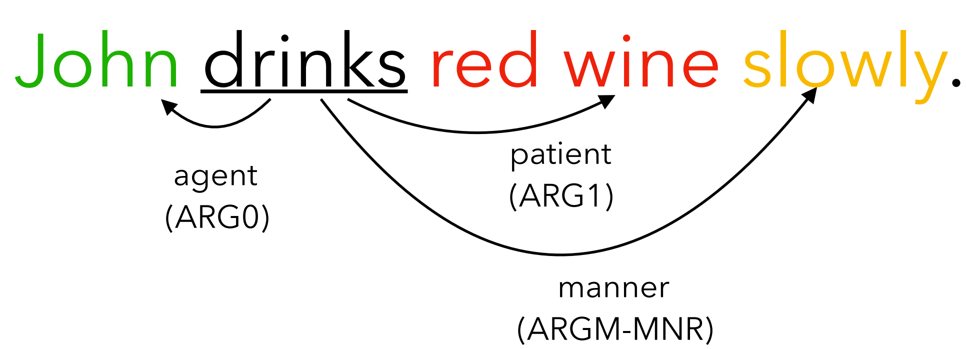
- Affordances are based on **Predicate-Argument Structures** (PASs) extracted from Natural Language using **Semantic Role Labeling** (SRL). We use [He et. al 2017]'s end-to-end neural SRL to process Wikipedia.

- After extraction, PASs are organised into a **co-occurrence matrix** and **weighted using PPMI**, similarly to [Levy and Goldberg 2014].

Affordance Extraction

- Affordances are based on **Predicate-Argument Structures** (PASs) extracted from Natural Language using **Semantic Role Labeling** (SRL). We use [He et. al 2017]'s end-to-end neural SRL to process Wikipedia.

- After extraction, PASs are organised into a **co-occurrence matrix** and **weighted using PPMI**, similarly to [Levy and Goldberg 2014].



PropBank annotations [Palmer 2012]

Affordance Extraction

- Affordances are based on **Predicate-Argument Structures** (PASs) extracted from Natural Language using **Semantic Role Labeling** (SRL). We use [He et. al 2017]'s end-to-end neural SRL to process Wikipedia.

- After extraction, PASs are organised into a **co-occurrence matrix** and **weighted using PPMI**, similarly to [Levy and Goldberg 2014].

	drink | ARG0	drink | ARG1	drink | ARGM-MNR	...
John	0.8	0.0	0.0	0.0
red	0.0	0.6	0.0	0.0
wine	0.0	0.9	0.0	0.0
slowly	0.0	0.0	0.7	0.0
...	0.0	0.0	0.0	0.0

Affordance Extraction

- Affordances are based on **Predicate-Argument Structures** (PASs) extracted from Natural Language using **Semantic Role Labeling** (SRL). We use [He et. al 2017]'s end-to-end neural SRL to process Wikipedia.

- After extraction, PASs are organised into a **co-occurrence matrix** and **weighted using PPMI**, similarly to [Levy and Goldberg 2014].

	drink | ARG0	**drink | ARG1**	**drink | ARGM-MNR**	**...**
John	0.8			
red		0.6		
wine		0.9		
slowly			0.7	
...				

Affordance Extraction

- Affordances are based on **Predicate-Argument Structures** (PASs) extracted from Natural Language using **Semantic Role Labeling** (SRL). We use [He et. al 2017]'s end-to-end neural SRL to process Wikipedia.

- After extraction, PASs are organised into a **co-occurrence matrix** and **weighted using PPMI**, similarly to [Levy and Goldberg 2014].

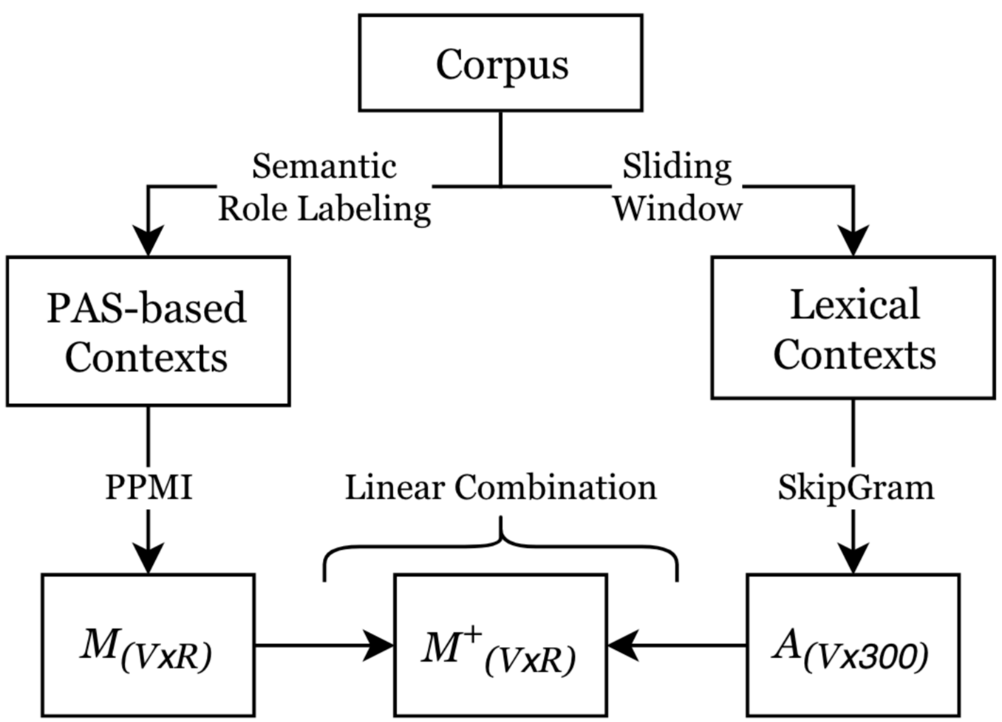
	drink | ARG0	drink | ARG1	drink | ARGM-MNR	...
John	0.8			
red		0.8		
wine		0.9		
slowly			0.7	
...				

Too Sparse ...

(# role contexts << # adj contexts)

Affordance Extraction

- To address sparsity we perform linear combination with adjacency-based representations obtained from the same corpus. Inspired by work in translation [Zhao et al. 2015].

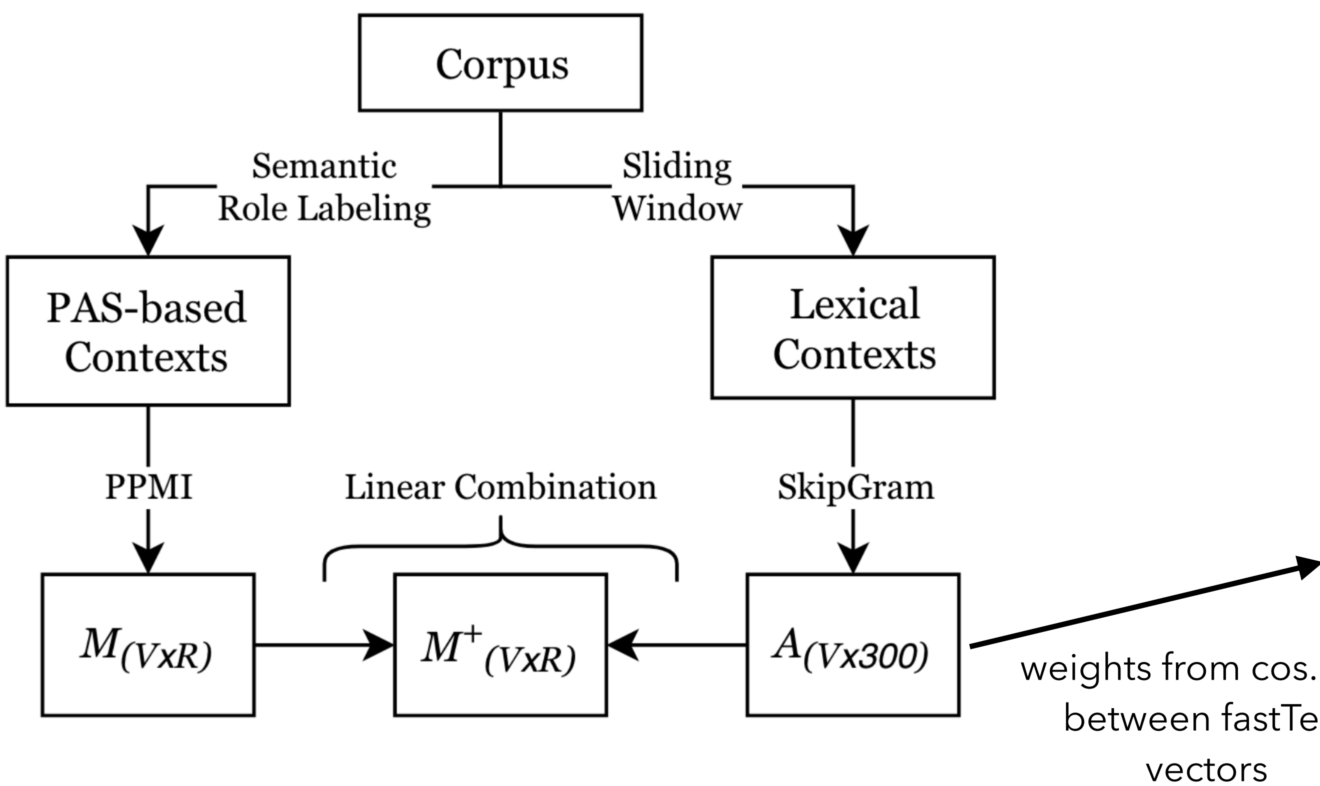


$$\vec{v}_w = \frac{\vec{v}_1 * \alpha_1 + ... + \vec{v}_n * \alpha_n}{\sum_{i=1}^{n} \alpha_i}$$

$$\alpha_i = \begin{cases} \frac{A_{\vec{v}_w} \cdot A_{\vec{v}_i}}{\|A_{\vec{v}_w}\| \|A_{\vec{v}_i}\|} = cos_A(\vec{v}_w, \vec{v}_i), \\ \qquad \text{if } cos_A(\vec{v}_w, \vec{v}_i) > 0.5 \\ 0, \quad \text{otherwise} \end{cases}$$

Affordance Extraction

- To address sparsity we perform linear interpolation with adjacency-based representations obtained from the same corpus. Inspired by work in translation [Zhao et al. 2015].

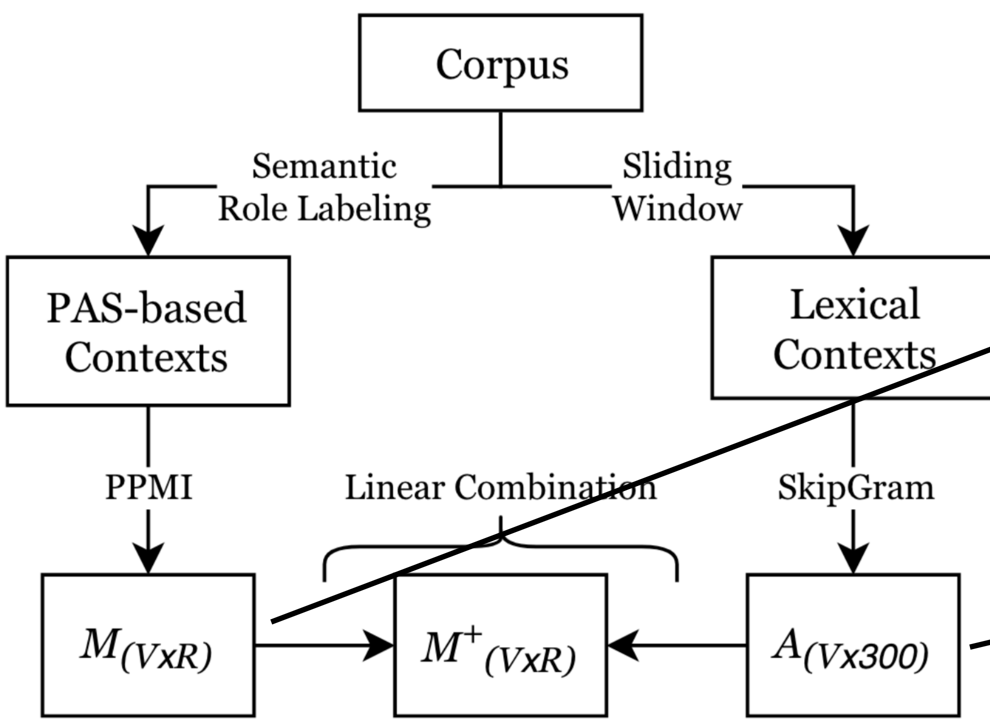


$$\vec{v}_w = \frac{\vec{v}_1 * \alpha_1 + \ldots + \vec{v}_n * \alpha_n}{\sum_{i=1}^{n} \alpha_i}$$

$$\alpha_i = \begin{cases} \frac{A_{\vec{v}_w} \cdot A_{\vec{v}_i}}{\|A_{\vec{v}_w}\| \|A_{\vec{v}_i}\|} = cos_A(\vec{v}_w, \vec{v}_i), & \text{if } cos_A(\vec{v}_w, \vec{v}_i) > 0.5 \\ 0, & \text{otherwise} \end{cases}$$

Affordance Extraction

- To address sparsity we perform linear interpolation with adjacency-based representations obtained from the same corpus. Inspired by work in translation [Zhao et al. 2015].



each PAS-based vector becomes a weighted combination of other vectors

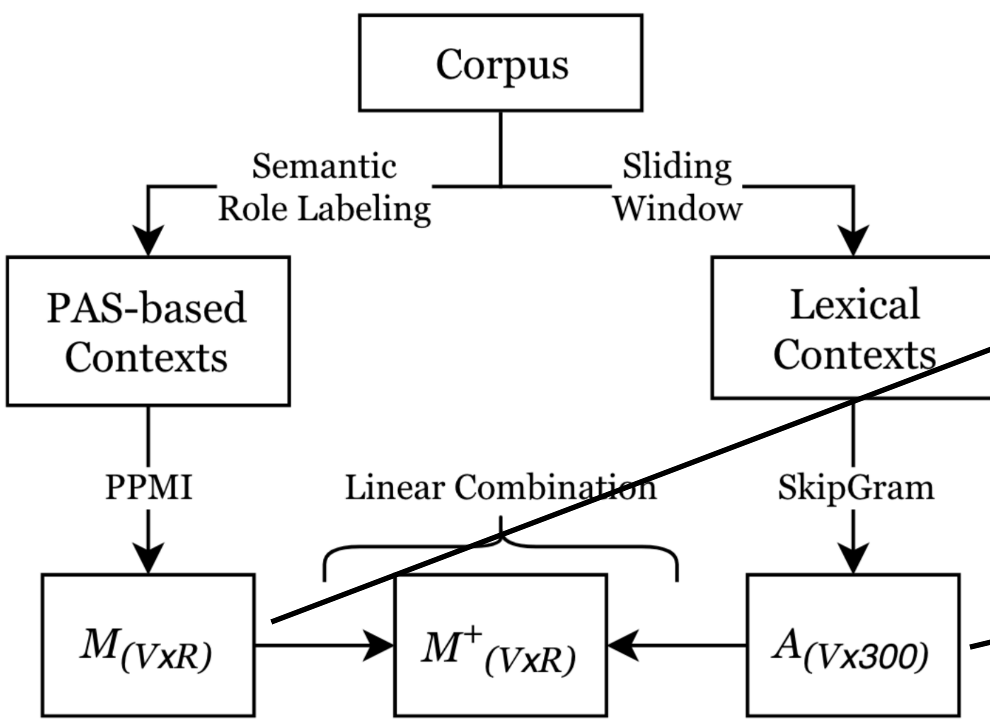
$$\vec{v}_w = \frac{\vec{v}_1 * \alpha_1 + ... + \vec{v}_n * \alpha_n}{\sum_{i=1}^{n} \alpha_i}$$

$$\alpha_i = \begin{cases} \frac{A_{\vec{v}_w} \cdot A_{\vec{v}_i}}{\|A_{\vec{v}_w}\| \|A_{\vec{v}_i}\|} = cos_A(\vec{v}_w, \vec{v}_i), \\ \qquad \text{if } cos_A(\vec{v}_w, \vec{v}_i) > 0.5 \\ 0, \quad \text{otherwise} \end{cases}$$

weights from cos. sim. between fastText vectors

Affordance Extraction

- To address sparsity we perform linear interpolation with adjacency-based representations obtained from the same corpus. Inspired by work in translation [Zhao et al. 2015].



each PAS-based vector becomes a weighted combination of other vectors

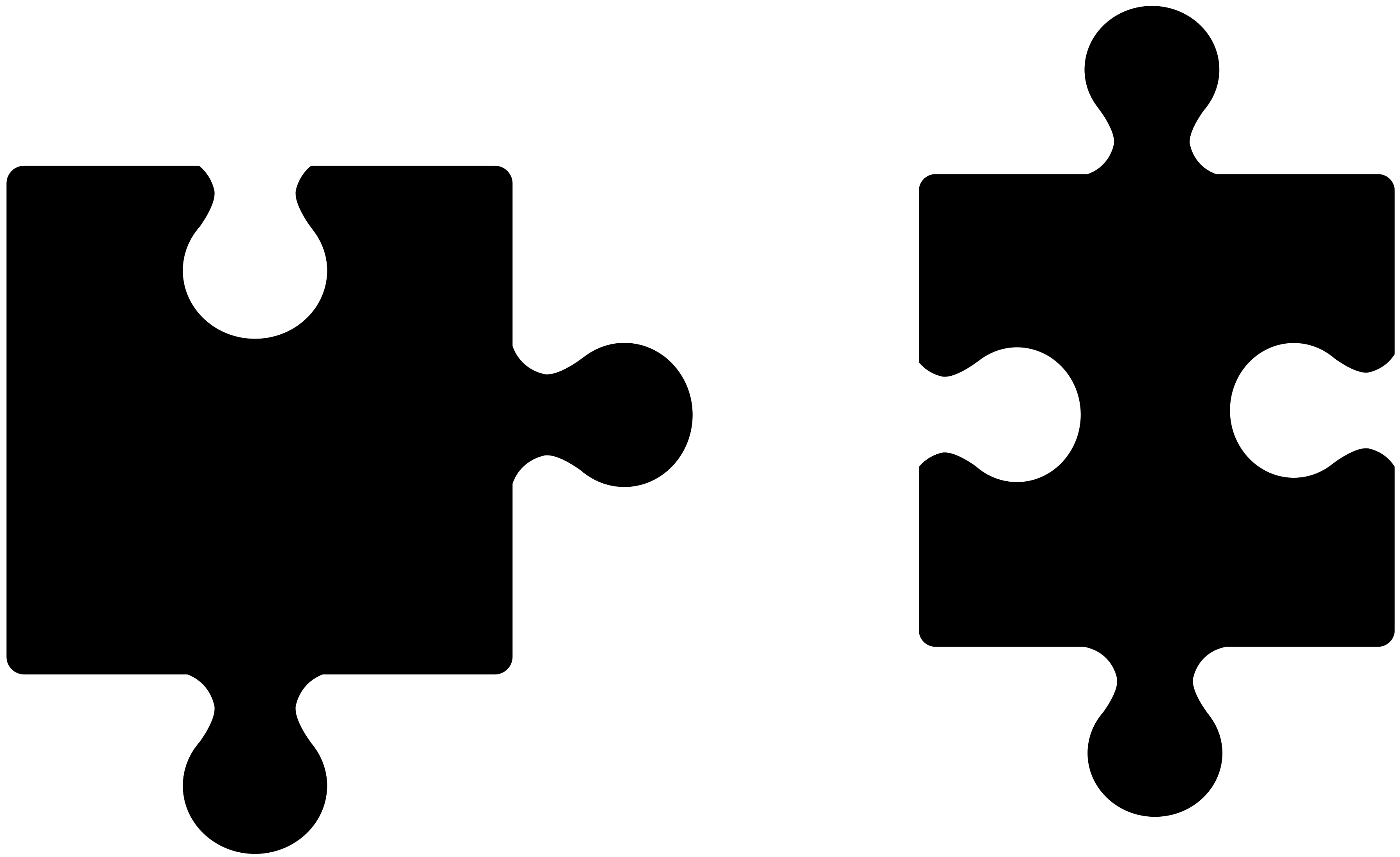
$$\vec{v}_w = \frac{\vec{v}_1 * \alpha_1 + ... + \vec{v}_n * \alpha_n}{\sum_{i=1}^{n} \alpha_i}$$

$$\alpha_i = \begin{cases} \frac{A_{\vec{v}_w} \cdot A_{\vec{v}_i}}{\|A_{\vec{v}_w}\| \|A_{\vec{v}_i}\|} = cos_A(\vec{v}_w, \vec{v}_i), \\ \qquad \text{if } cos_A(\vec{v}_w, \vec{v}_i) > 0.5 \\ 0, \quad \text{otherwise} \end{cases}$$

weights from cos. sim. between fastText vectors

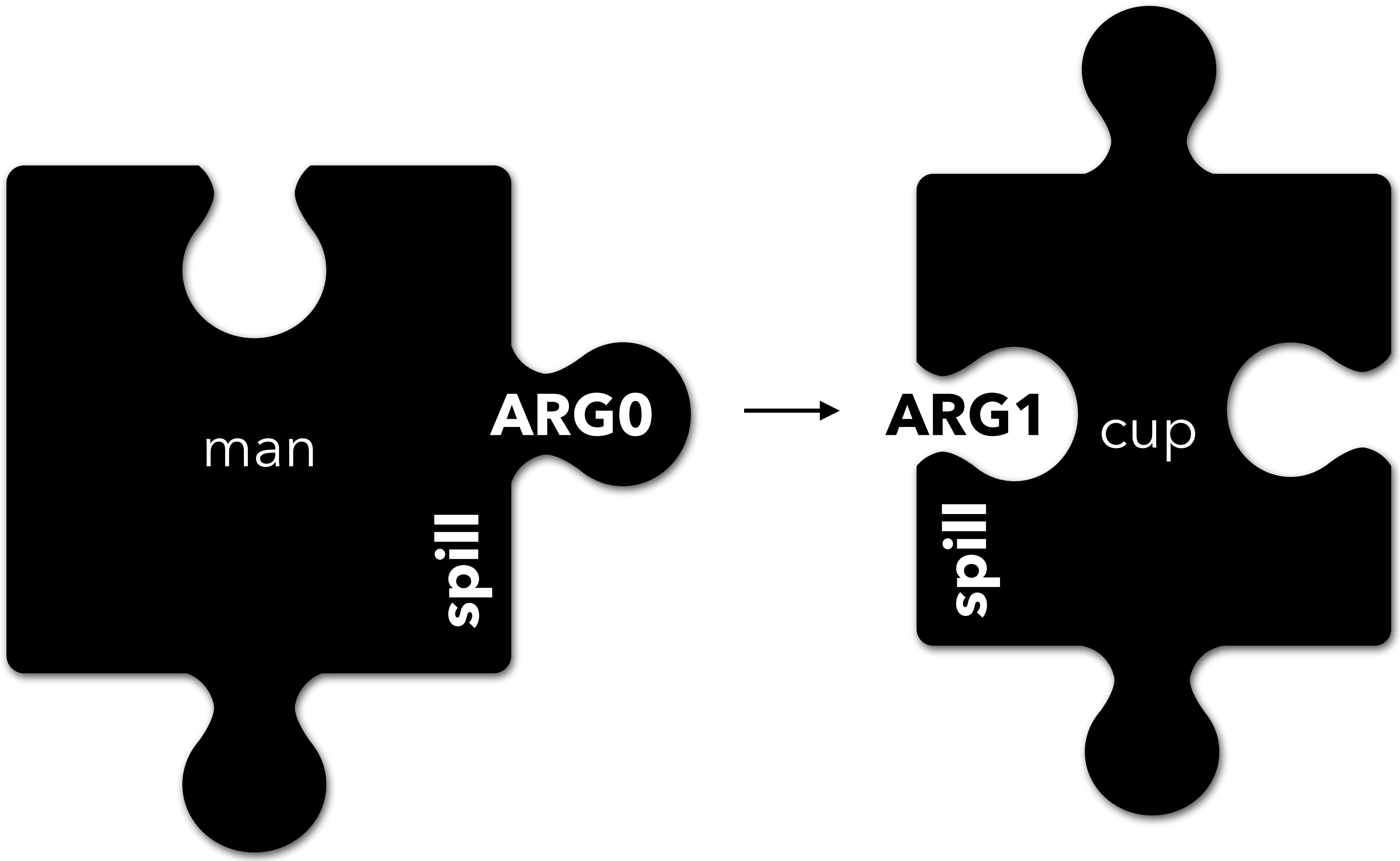
- This redefines existing vectors as well as creates new ones.

Affordance Inference



Indexical Hypothesis' Meshing

Affordance Inference



i.e. Role Complementarity

Affordance Inference

Algorithm 1 Affordance Meshing Algorithm

procedure INFERENCE(M^+, w_1, w_2, a_1, a_2)
  $relations \leftarrow []$
  $\vec{v}_1, \vec{v}_2 \leftarrow get\_vec(w_1, M^+), get\_vec(w_2, M^+)$
  for $f_1 \in features(\vec{v}_1) \wedge arg(f_1) = a_1$ **do**
    for $f_2 \in features(\vec{v}_2) \wedge arg(f_2) = a_2$ **do**
      if $pred(f_1) = pred(f_2)$ **then**
        $relations.add((f_1 * f_2, pred(f_1)))$
  return $sorted(relations)$

Word Pairs (w_1, w_2)	**Affordances** (w_1 as ARG0, w_2 as ARG1)
shop, tea	sell, import, cure
doctor, patient	diagnose, prescribe, treat
newspaper, face	cover, expose, poke
man, cup	drink, pour, spill

Simple algorithm using interpolated PAS-based vectors.

Word Representations that are relational and interpretable

Affordance Inference

Algorithm 1 Affordance Meshing Algorithm

procedure INFERENCE(M^+, w_1, w_2, a_1, a_2)
  $relations \leftarrow []$
  $\vec{v}_1, \vec{v}_2 \leftarrow get\_vec(w_1, M^+), get\_vec(w_2, M^+)$
  for $f_1 \in features(\vec{v}_1) \wedge arg(f_1) = a_1$ **do**
    for $f_2 \in features(\vec{v}_2) \wedge arg(f_2) = a_2$ **do**
      if $pred(f_1) = pred(f_2)$ **then**
        $relations.add((f_1 * f_2, pred(f_1)))$
  return $sorted(relations)$

Word Pairs (w_1, w_2)	**Affordances** (w_1 as ARG0, w_2 as ARG1)
shop, tea	sell, import, cure
doctor, patient	diagnose, prescribe, treat
newspaper, face	cover, expose, poke
man, cup	drink, pour, spill

Simple algorithm using interpolated PAS-based vectors.

Word Representations that are relational and interpretable

But are these ***accurate*** word representations?

Evaluation

- Word Similarity Tasks are the standard for evaluating word representations.

Context	Model	SL-666	SL-999	WS-SIM	WS-ALL	MEN	RG-65
Lexical	word2vec	.426	.414	.762	.672	.721	.793
	GloVe	.333	.325	.637	.535	.636	.601
	fastText (A)	.426	.419	**.779**	**.702**	**.751**	.799
Syntactic	★ Deps	**.475**	**.446**	.758	.629	.606	.765
	Open IE	.397	.390	.746	.696	.281	.801
	A2Avecs (M^+)	.461	.412	.734	.577	.687	**.802**
	A2Avecs (SVD(M^+))	.436	.386	.672	.509	.599	.789

All trained on Wikipedia

- Ours (A2Avecs) performs competitively with adjacency-based lexical contexts, but the dependency-based embeddings of Levy and Goldberg 2014 still perform better.

- Curiously, applying SVD to reduce our explicit 18k dimensions into the standard 300 latent dimensions hurts performance significantly.

Evaluation

- However, what if we try concatenating our PAS-based vectors with latent embeddings trained on larger corpora (fastText 600B)?

Context	Model	SL-666	SL-999	WS-SIM	WS-ALL	MEN	RG-65
Lexical SOTA	fastText 600B (A)	.523	.504	.839	**.791**	**.836**	**.859**
Intp. w/SOTA	A2Avecs (M^+)	.513	.468	.780	.619	.744	.814
Intp. & Conc.	★A2Avecs ($M^+ \parallel A$)	**.540** ↑	**.521** ↑	**.846** ↑	.771 ↓	.829 ↓	.857 ↔
Deps Conc.	Deps $\parallel$ A	.524 ↔	.503 ↔	.818 ↓	.752 ↓	.770 ↓	.835 ↓

- Interestingly, this solution is markedly better, significantly outperforming the SOTA on challenging tasks such as SimLex-999 (specially nouns).

- To be rigorous, we concatenated the same latent embeddings to the dependency-based embeddings, and found that this combination wasn't beneficial.

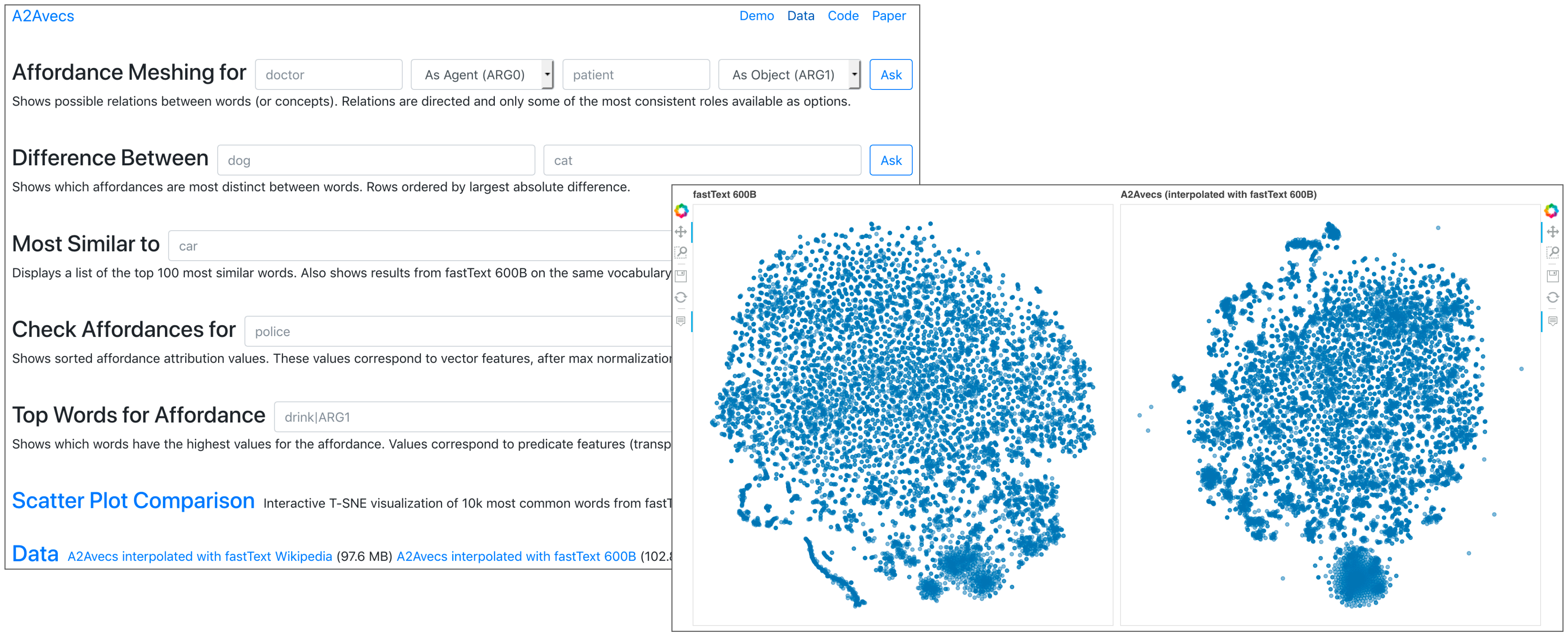
Conclusions

- SRL can be useful for deriving word representations with information that is **complementary to adjacency-based contexts** (and dependency-based).
- Within the same vector space, you can perform **relational inferences** while still using cosine similarity for semantics.
- This representation of affordances may be a useful way to integrate Commonsense knowledge into applications such as Fact Verification, particularly by enabling **semantic plausibility assessments**.

Conclusions

- SRL can be useful for deriving word representations with information that is **complementary to adjacency-based contexts** (and dependency-based).
- Within the same vector space, you can perform **relational inferences** while still using cosine similarity for semantics.
- This representation of affordances may be a useful way to integrate Commonsense knowledge into applications such as Fact Verification, particularly by enabling **semantic plausibility assessments**.
- In future work, we'll evaluate on more tasks and propose better ways to exploit PAS-based relational knowledge. (on-going)

Questions ?



Demo and more at: a2avecs.github.io

Thank You!

dloureiro@fc.up.pt

danielbloureiro