Language Modelling Makes Sense

Propagating Representations through WordNet for Full-Coverage Word Sense Disambiguation

Daniel Loureiro, Alípio Jorge

ACL – Florence, 31 July 2019

U.PORTO
FC FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO
LIAAD
INESCTEC | LABORATÓRIO ASSOCIADO

Sense Embeddings

Exploiting the latest Neural Language Models (NLMs) for sense-level representation learning.

Sense Embeddings

Exploiting the latest Neural Language Models (NLMs) for sense-level representation learning.

- Beat SOTA for English Word Sense Disambiguation (WSD).
- Full WordNet in NLM-space (+100K common sense concepts).
- Concept-level analysis of NLMs.

Related Work

Related Work

[Iacobacci et al. (2016)]
[Zhong and Ng (2010)]

Bag-of-Features Classifiers

(SVM)

[Luo et al. (2018b)]
[Luo et al. (2018a)]
[Vial et al. (2018)]
[Raganato et al. (2017)]

Deep Sequence Classifiers

(BiLSTM)

[Peters et al. (2018)]
[Melamud et al. (2016)]
[Yuan et al. (2016)]

Sense-level Representations

(k-NN)
(over NLM reprs.)

Related Work

[Iacobacci et al. (2016)]
[Zhong and Ng (2010)]

Bag-of-Features Classifiers

(SVM)

[Luo et al. (2018b)]
[Luo et al. (2018a)]
[Vial et al. (2018)]
[Raganato et al. (2017)]

Deep Sequence Classifiers

(BiLSTM)

[Peters et al. (2018)]
[Melamud et al. (2016)]
[Yuan et al. (2016)]

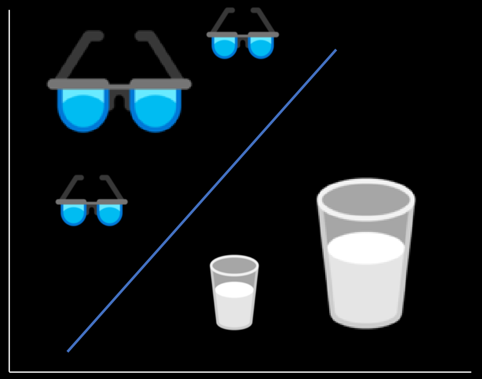
Sense-level Representations

(k-NN)
(over NLM reprs.)

Bag-of-Features Classifiers

It Makes Sense (IMS) **[Zhong and Ng (2010)] :**

- POS tags, surrounding words, local collocations.
- SVM for each word type in training.
- Fallback: Most Frequent Sense (MFS).
- Improved with word embedding features. [Iacobacci et al. (2016)]
- Still competitive (!)

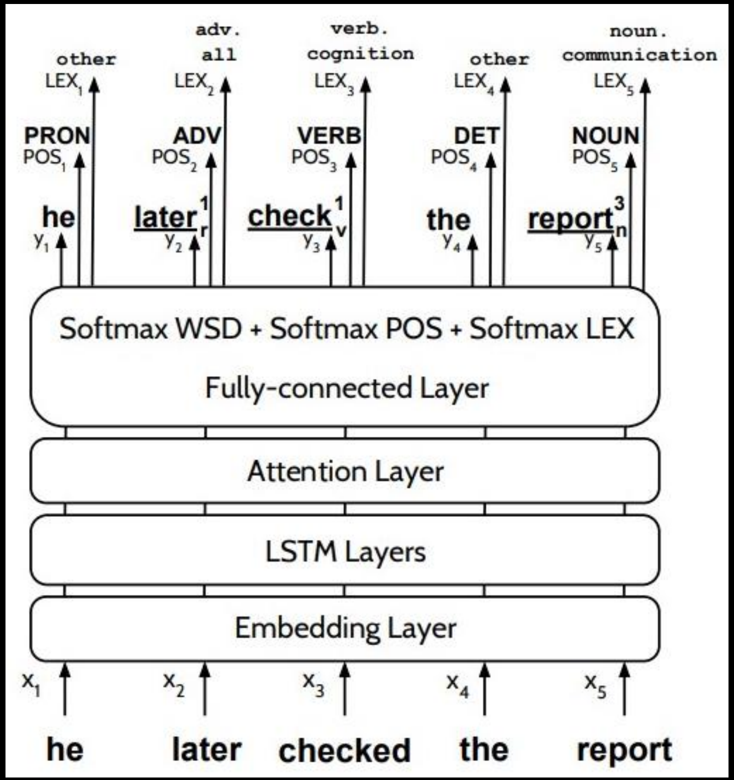


“glasses”

Deep Sequence Classifiers

Bi-directional LSTMs (BiLSTMs):

- Better with:
  - Attention (as everything else).
  - Auxiliary losses. (POS, lemmas, lexnames) [Raganato et al. (2017)]
  - Glosses, via co-attention mechanisms. [Luo et al. (2018)]
- Still must fallback on MFS.
- Not that much better than bag-of-features…

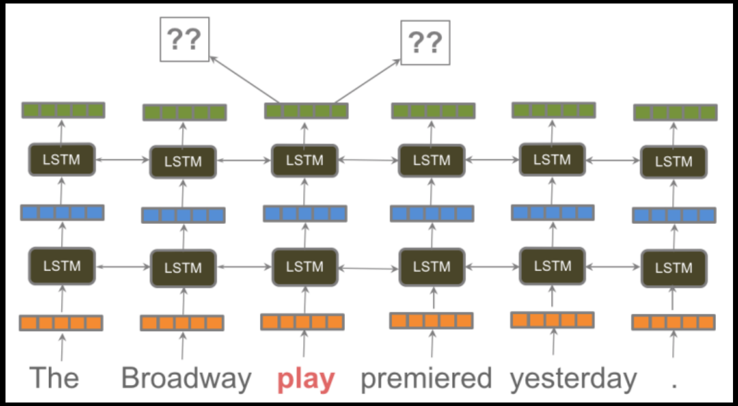


[Raganato et al. (2017)]

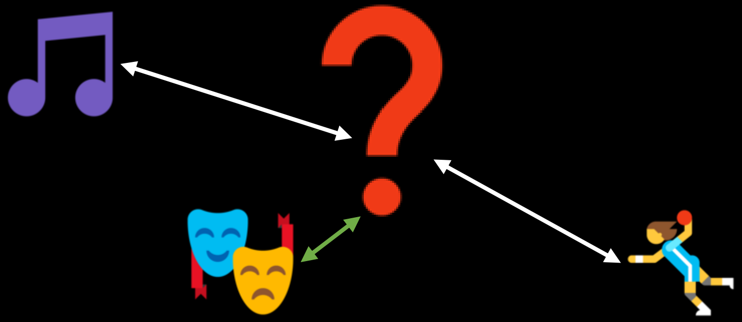
Contextual k-NN

Matching Contextual Word Embeddings:

- Produce Sense Embeddings from NLMs (averaging).
- Sense embs. can be compared with contextual embs.
- Disambiguation = Nearest Neighbour search (1-NN).
- Sense embs. limited to annotations. MFS required.
- Promising, but early attempts.



[Ruder (2018)]



Our Approach

Our Approach

- Expand the k-NN approach to full-coverage of WordNet.

Our Approach

- Expand the k-NN approach to full-coverage of WordNet.
- Matching senses becomes trivial, no MFS fallbacks needed.

Our Approach

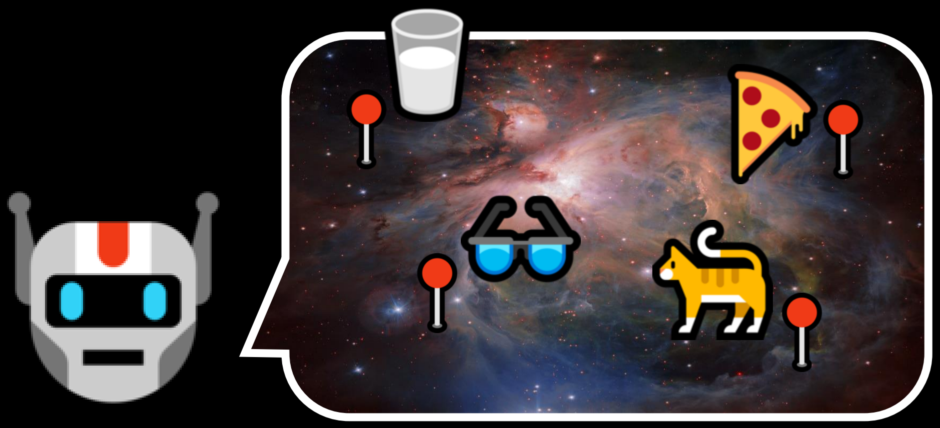
- Expand the k-NN approach to full-coverage of WordNet.
- Matching senses becomes trivial, no MFS fallbacks needed.
- Full-set of sense embeddings in NLM-space is useful beyond WSD.

Our Approach

- Expand the k-NN approach to full-coverage of WordNet.
- Matching senses becomes trivial, no MFS fallbacks needed.
- Full-set of sense embeddings in NLM-space is useful beyond WSD.

Our Approach

- Expand the k-NN approach to full-coverage of WordNet.
- Matching senses becomes trivial, no MFS fallbacks needed.
- Full-set of sense embeddings in NLM-space is useful beyond WSD.



Challenges

Challenges

- Overcome very limited sense annotations (covers 16% senses).

Challenges

- Overcome very limited sense annotations (covers 16% senses).
- Infer missing senses correctly so that task performance improves.

Challenges

- Overcome very limited sense annotations (covers 16% senses).
- Infer missing senses correctly so that task performance improves.
- Rely only on sense embeddings, no lemma or POS features.

Challenges

- Overcome very limited sense annotations (covers 16% senses).
- Infer missing senses correctly so that task performance improves.
- Rely only on sense embeddings, no lemma or POS features.

Bootstrap → Propagate → Enrich → Reinforce

Bootstrap	Propagate	Enrich	Reinforce
Annotated Dataset	WordNet Ontology	WordNet Glosses	Morphological Embeddings

Bootstrapping Sense Embeddings

Can your insurance company aid you in reducing administrative costs ?

Would it be feasible to limit the menu in order to reduce feeding costs ?

Bootstrapping Sense Embeddings

insurance_company%1:14:00:: reduce%2:30:00:: cost%1:21:00::

aid%2:41:00:: administrative%3:01:00::

Can your insurance company aid you in reducing administrative costs ?

Would it be feasible to limit the menu in order to reduce feeding costs ?

feasible%5:00:00:possible:00 menu%1:10:00:: feeding%1:04:01::

limit%2:30:00:: reduce%2:30:00:: cost%1:21:00::

Bootstrapping Sense Embeddings

insurance_company%1:14:00:: aid%2:41:00:: reduce%2:30:00:: administrative%3:01:00:: cost%1:21:00::

Can your insurance company aid you in reducing administrative costs ?

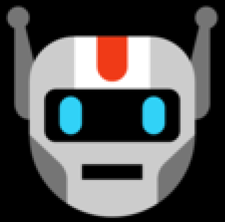
Would it be feasible to limit the menu in order to reduce feeding costs ?

feasible%5:00:00:possible:00 limit%2:30:00:: menu%1:10:00:: reduce%2:30:00:: feeding%1:04:01:: cost%1:21:00::

Bootstrapping Sense Embeddings

insurance_company%1:14:00::
aid%2:41:00::
reduce%2:30:00::
administrative%3:01:00::
cost%1:21:00::

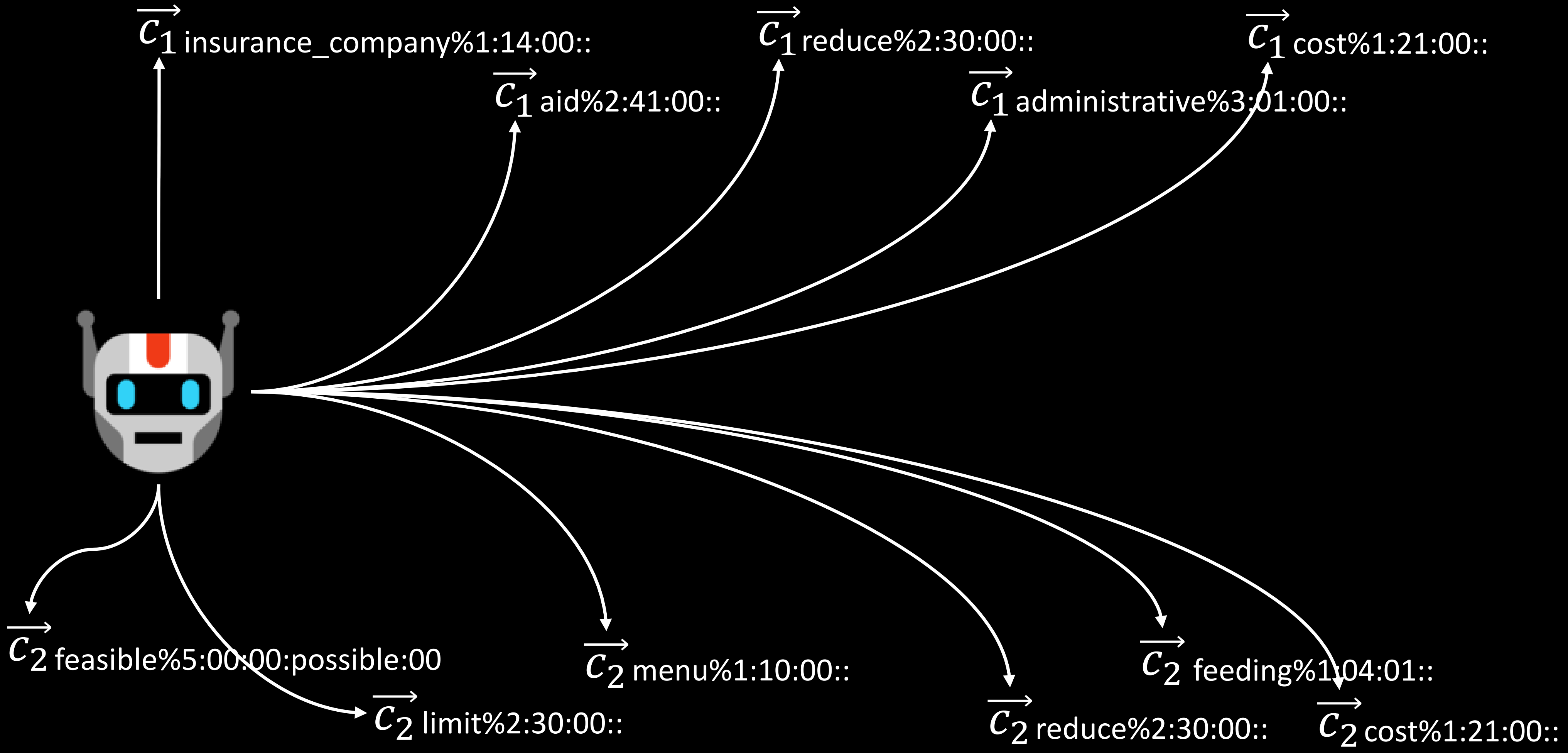
Can your insurance company aid you in reducing administrative costs ?



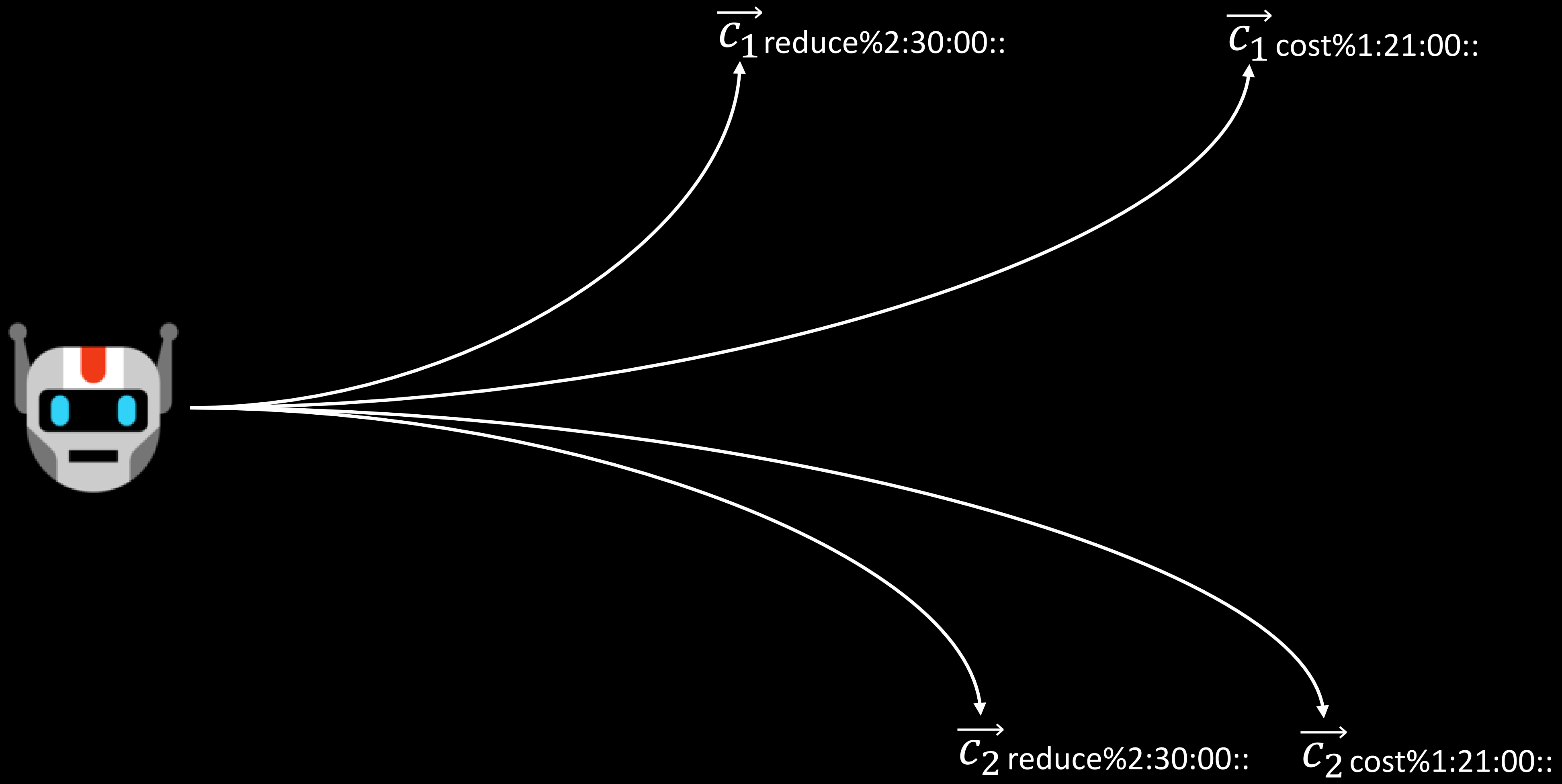
Would it be feasible to limit the menu in order to reduce feeding costs ?

feasible%5:00:00:possible:00
limit%2:30:00::
menu%1:10:00::
reduce%2:30:00::
feeding%1:04:01::
cost%1:21:00::

Bootstrapping Sense Embeddings



Bootstrapping Sense Embeddings



Bootstrapping Sense Embeddings

$$\vec{v}_{\text{reduce\%2:30:00::}} = \frac{\vec{c_1}_{\text{reduce\%2:30:00::}} + \vec{c_2}_{\text{reduce\%2:30:00::}} + \ldots + \vec{c_n}_{\text{reduce\%2:30:00::}}}{n}$$

$$\vec{v}_{\text{cost\%1:21:00::}} = \frac{\vec{c_1}_{\text{cost\%1:21:00::}} + \vec{c_2}_{\text{cost\%1:21:00::}} + \ldots + \vec{c_n}_{\text{cost\%1:21:00::}}}{n}$$

Bootstrapping Sense Embeddings

$$\vec{v}_{\text{reduce\%2:30:00::}} = \frac{\vec{c_1}_{\text{reduce\%2:30:00::}} + \vec{c_2}_{\text{reduce\%2:30:00::}} + \ldots + \vec{c_n}_{\text{reduce\%2:30:00::}}}{n}$$

$$\vec{v}_{\text{cost\%1:21:00::}} = \frac{\vec{c_1}_{\text{cost\%1:21:00::}} + \vec{c_2}_{\text{cost\%1:21:00::}} + \ldots + \vec{c_n}_{\text{cost\%1:21:00::}}}{n}$$

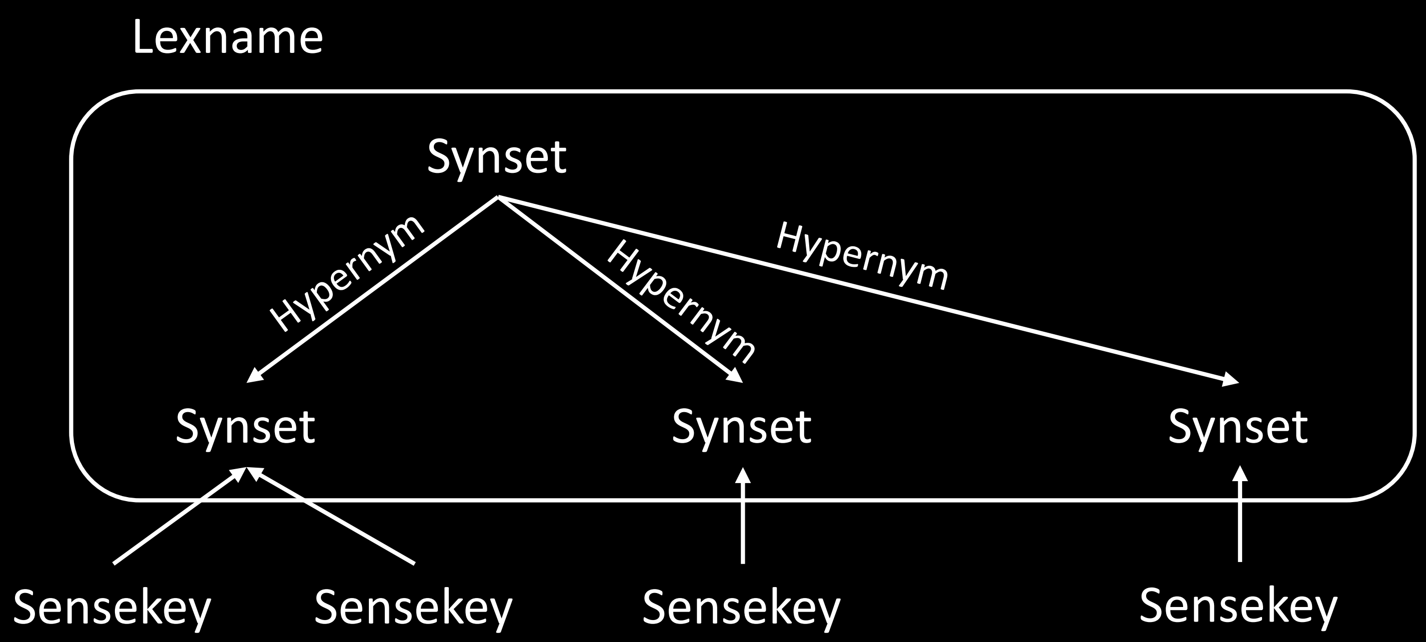
Outcome: 33,360 sense embeddings (16% coverage)

Propagating Sense Embeddings

WordNet's units, synsets, represent concepts at different levels.

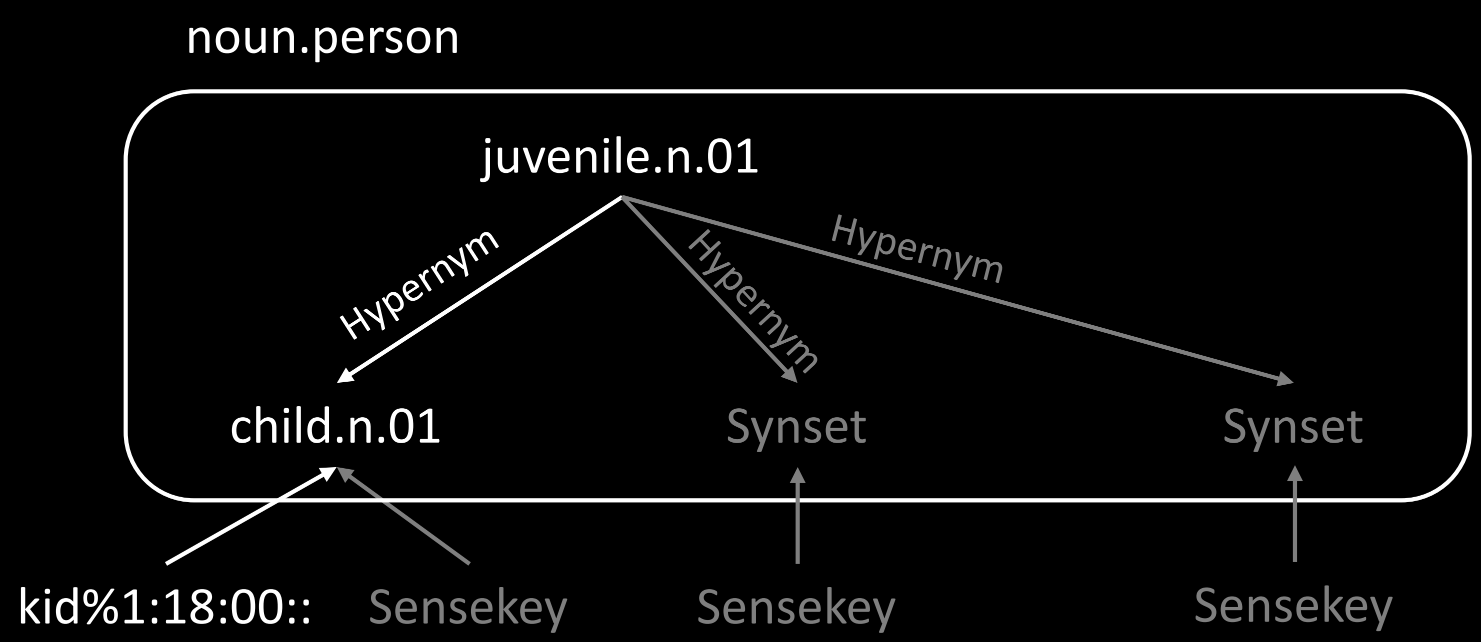
Propagating Sense Embeddings

WordNet's units, synsets, represent concepts at different levels.



Propagating Sense Embeddings

WordNet's units, synsets, represent concepts at different levels.



Propagating Sense Embeddings

burger%1:13:00::

hotdog%1:18:00::

hamburger%1:13:01::

sandwich%1:13:00::

wrap%1:13:00::

potato_chip%1:13:00::

Propagating Sense Embeddings

burger%1:13:00::

hotdog%1:18:00::

hamburger%1:13:01::

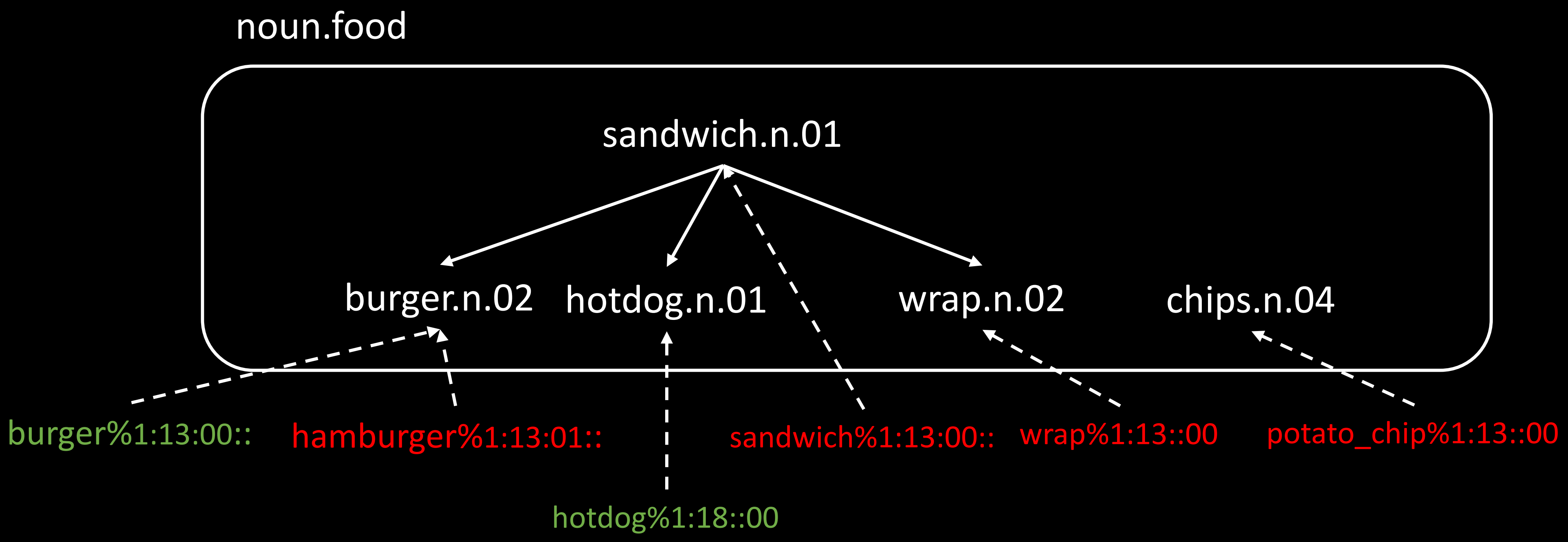
sandwich%1:13:00::

wrap%1:13:00::

potato_chip%1:13:00::

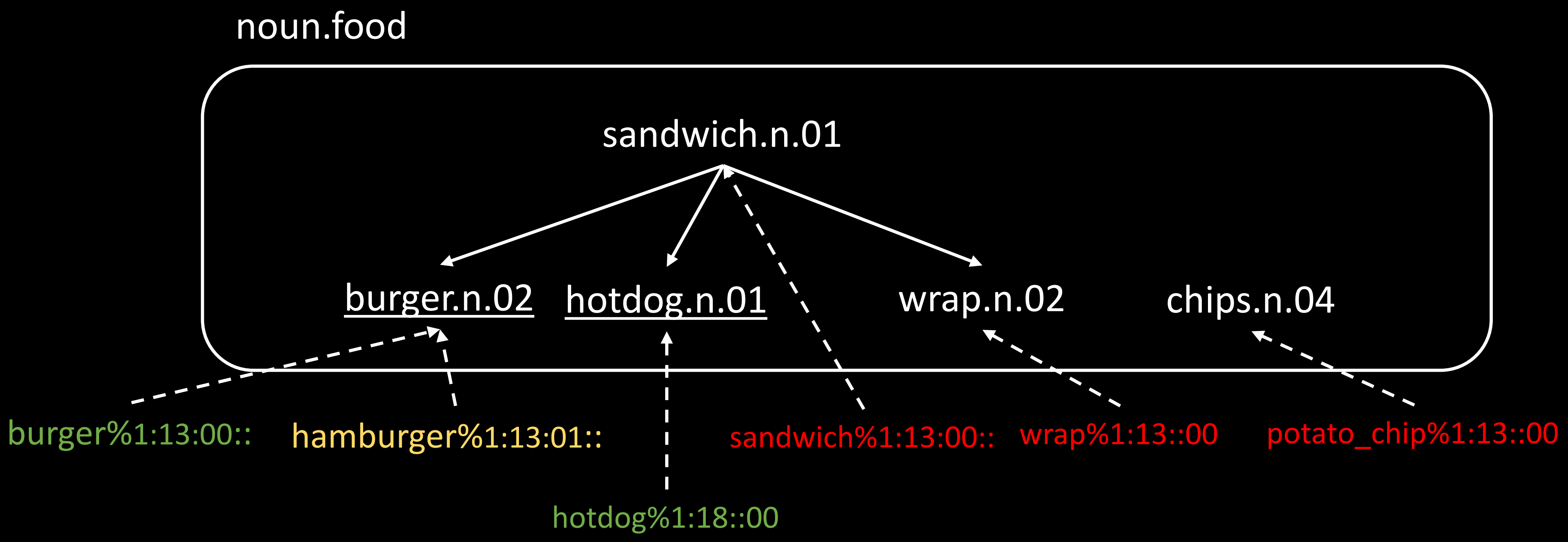
Propagating Sense Embeddings

Retrieve Synsets, Relations and Categories



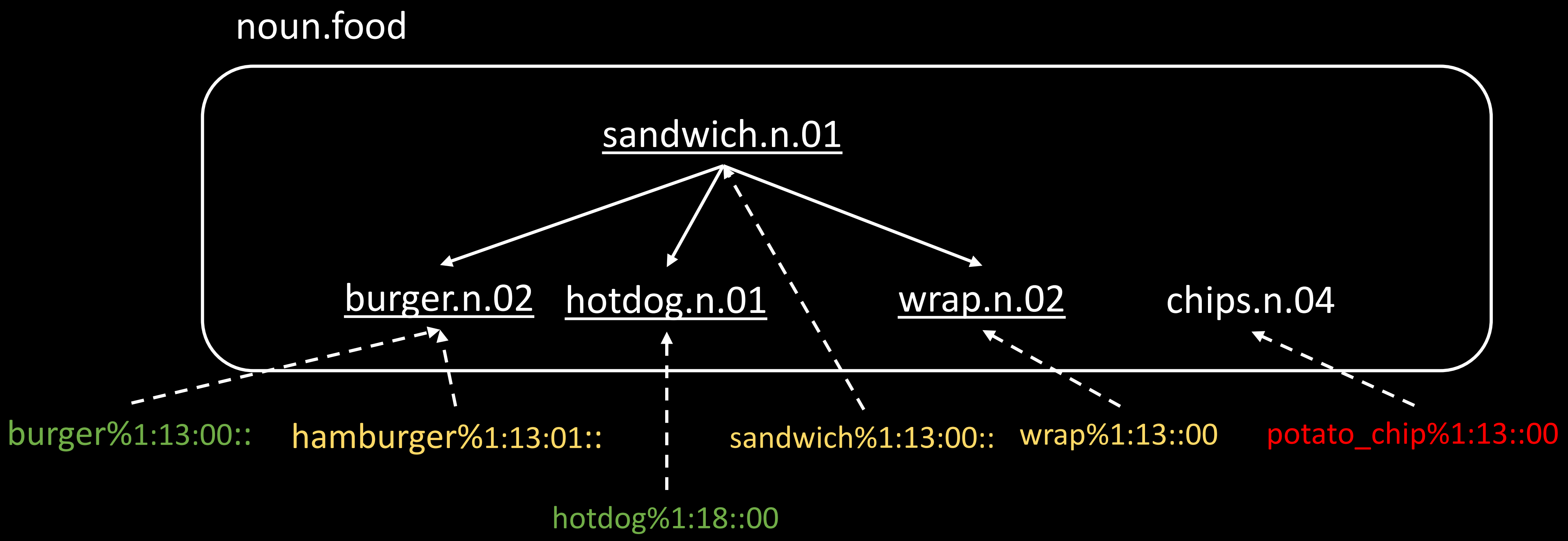
Propagating Sense Embeddings

1st stage: Synset Embeddings



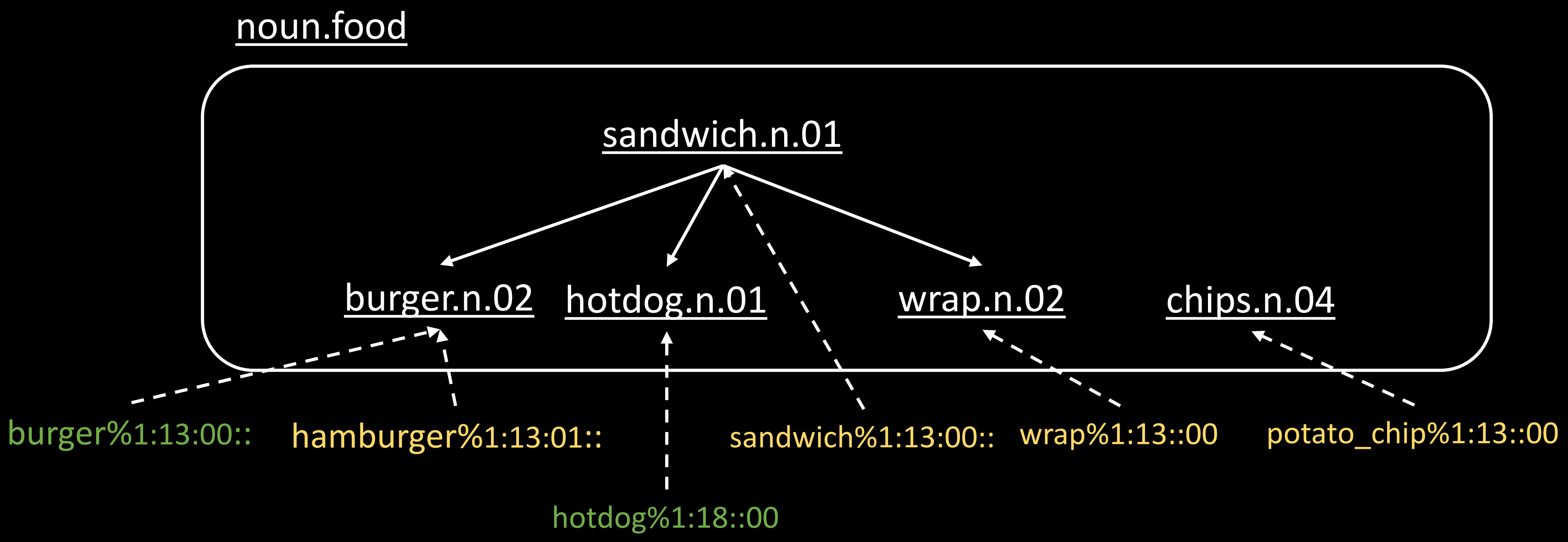
Propagating Sense Embeddings

2^{nd} Stage: Hypernym Embeddings (ind. Synsets)



Propagating Sense Embeddings

3rd Stage: Lexname Embeddings



Propagating Sense Embeddings

But  !=  …

noun.food

sandwich.n.01

burger.n.02 hotdog.n.01 wrap.n.02 chips.n.04

burger%1:13:00:: hamburger%1:13:01:: **sandwich%1:13:00::** **wrap%1:13::00** potato_chip%1:13::00

hotdog%1:18::00

Enriching Sense Embeddings

Leverage Synset Definitions and Lemmas for Differentiation

Enriching Sense Embeddings

Leverage Synset Definitions and Lemmas for Differentiation

sandwich:%1:13:00:: (sandwich.n.01)

Definition: two (or more) slices of bread with a filling between them

Lemmas: sandwich

wrap:%1:13:00:: (wrap.n.02)

Definition: a sandwich in which the filling is rolled up in a soft tortilla

Lemmas: wrap, tortilla

Enriching Sense Embeddings

Compose a new context

🥪

sandwich:%1:13:00:: (sandwich.n.01)

sandwich - two (or more) slices of bread with a filling between them

🌯

wrap:%1:13:00:: (wrap.n.02)

wrap, tortilla - a sandwich in which the filling is rolled up in a soft tortilla

Enriching Sense Embeddings

Make the context specific to sensekey (repeat lemma)

sandwich:%1:13:00::

sandwich - sandwich - two (or more) slices of bread with a filling between them

wrap%1:13:00::

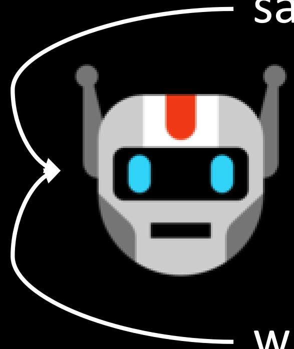
wrap - wrap, tortilla - a sandwich in which the filling is rolled up in a soft tortilla

Enriching Sense Embeddings

Make the context specific to sensekey (repeat lemma)

sandwich:%1:13:00::

sandwich - sandwich - two (or more) slices of bread with a filling between them



wrap%1:13:00::

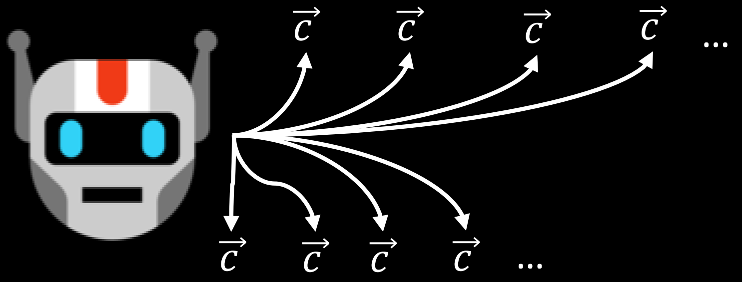
wrap - wrap, tortilla - a sandwich in which the filling is rolled up in a soft tortilla

Enriching Sense Embeddings

Obtain contextual embeddings for every token

sandwich:%1:13:00::

sandwich - sandwich - two (or more) slices of bread with a filling between them

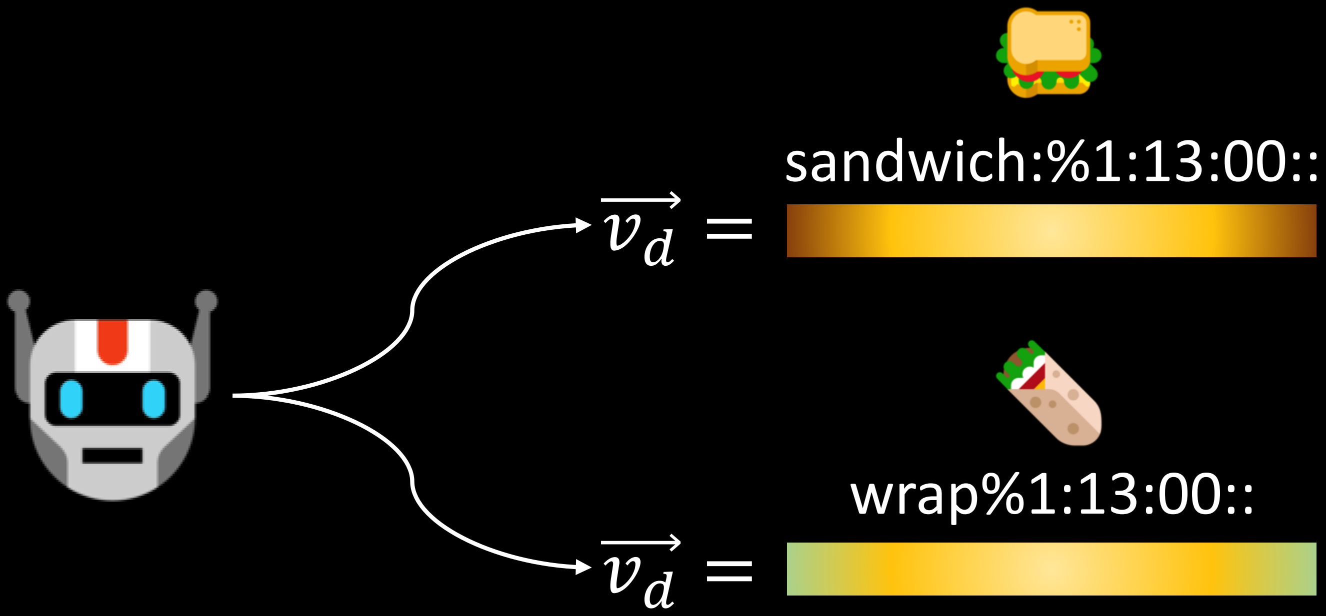


wrap%1:13:00::

wrap – wrap, tortilla - a sandwich in which the filling is rolled up in a soft tortilla

Enriching Sense Embeddings

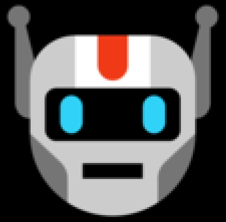
Sentence Embedding from avg. of Contextual Embeddings



$d = 1024$

Enriching Sense Embeddings

Merge Sentence Embedding with previous Sense Embedding



sandwich:%1:13:00::

$\overrightarrow{v_d} =$

sandwich:%1:13:00::

$\overrightarrow{v_s} =$

wrap%1:13:00::

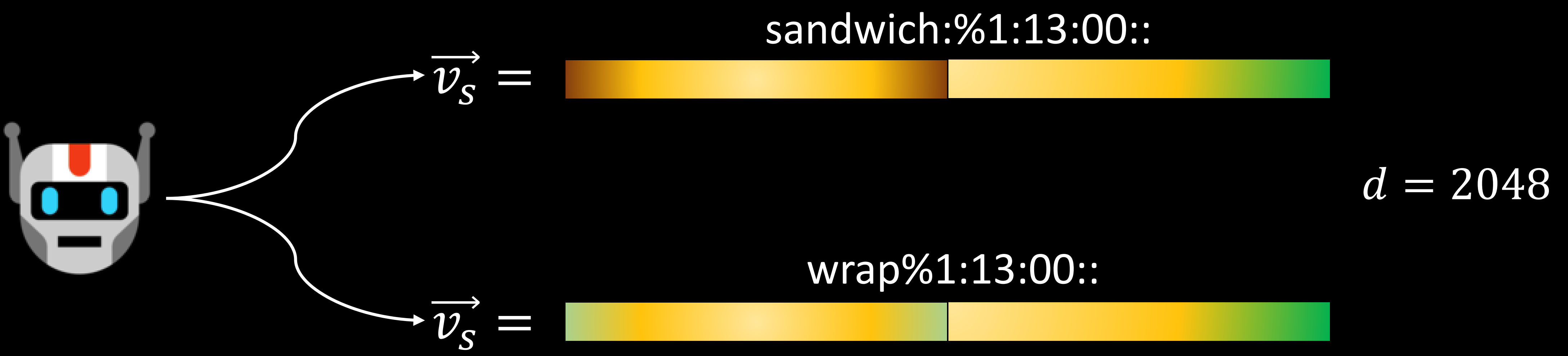
$\overrightarrow{v_d} =$

wrap:%1:13:00::

$\overrightarrow{v_s} =$

Enriching Sense Embeddings

Merge Sentence Embedding with previous Sense Embedding



Reinforcing Sense Embeddings

Contextual Embeddings aren’t good at preserving morphological relatedness

Reinforcing Sense Embeddings

Retrieve char-ngram embeddings (static) for lemmas

sandwich:%1:13:00::

$\vec{v_l} =$

wrap%1:13:00::

$\vec{v_l} =$

Reinforcing Sense Embeddings

Merge with previous sense embeddings

sandwich:%1:13:00::

$\overrightarrow{v_s} =$ $\overrightarrow{v_l} =$

wrap%1:13:00::

$\overrightarrow{v_s} =$ $\overrightarrow{v_l} =$

Reinforcing Sense Embeddings

Merge with previous sense embeddings

sandwich:%1:13:00::

$\overrightarrow{v_s} =$ 

$d = 2348$

wrap%1:13:00::

$\overrightarrow{v_s} =$ 

Matching Sense Embeddings

The glasses are in the cupboard.

Matching Sense Embeddings

The <u>glasses</u> are in the cupboard.

Matching Sense Embeddings

$$\overrightarrow{v_t} = \begin{array}{|c|c|c|}\hline \vec{c} & \vec{c} & \vec{v} \\ \hline \end{array}$$

$\longrightarrow \vec{v}$

$\longrightarrow \vec{c}$

The <u>glasses</u> are in the cupboard.

Matching Sense Embeddings

drinking_glass%1:06:00::

$\overrightarrow{v_s}$	$\overrightarrow{v_d}$	$\overrightarrow{v_l}$

glass%1:27:00::

$\overrightarrow{v_s}$	$\overrightarrow{v_d}$	$\overrightarrow{v_l}$

spectacles%1:06:00::

$\overrightarrow{v_s}$	$\overrightarrow{v_d}$	$\overrightarrow{v_l}$

$$\overrightarrow{v_t} = \boxed{\vec{c}}\,\boxed{\vec{c}}\,\boxed{\vec{v}}$$

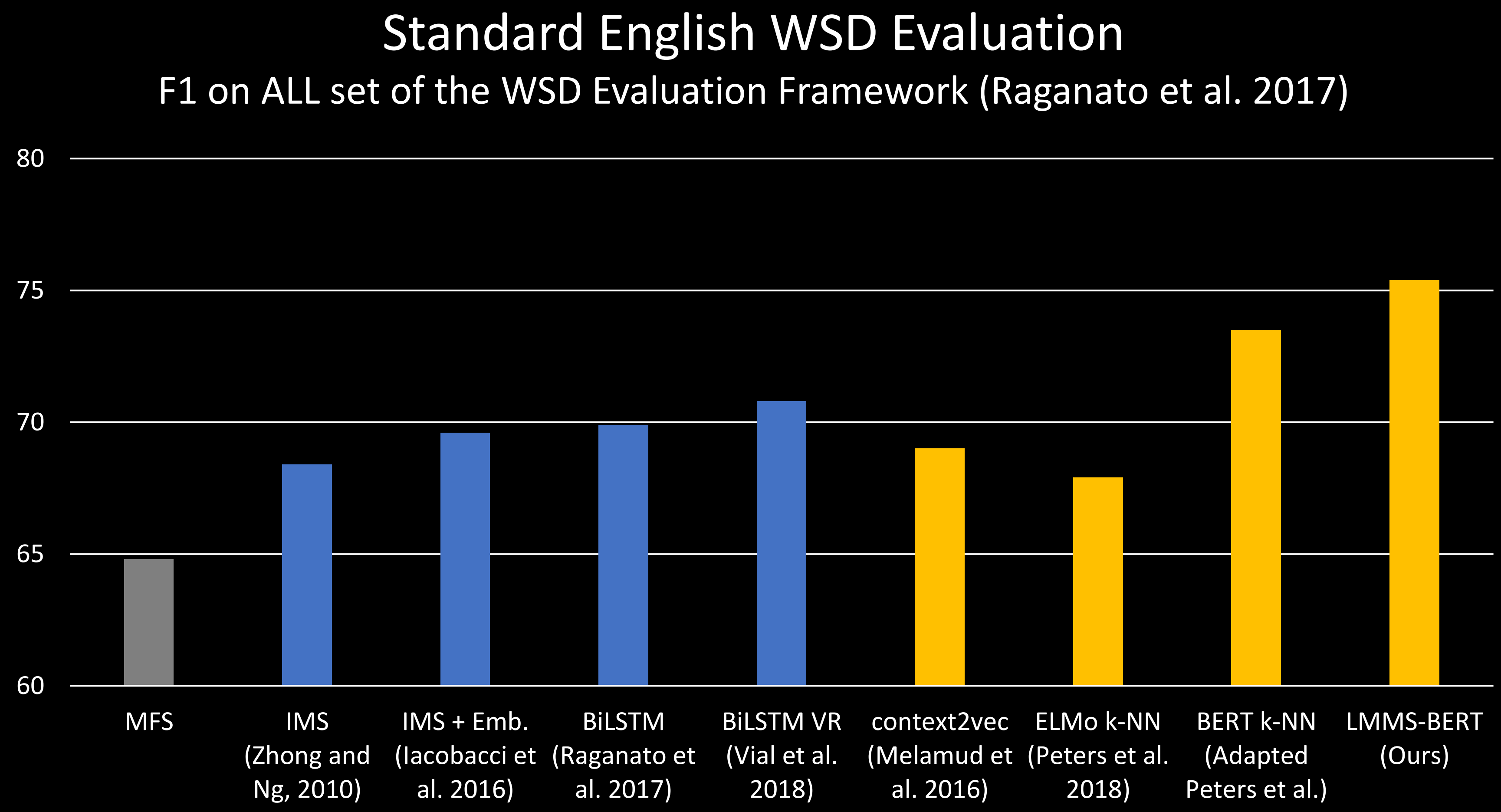
$\longrightarrow \vec{v}$

$\longrightarrow \vec{c}$

The glasses are in the cupboard.

WSD Results

WSD Results

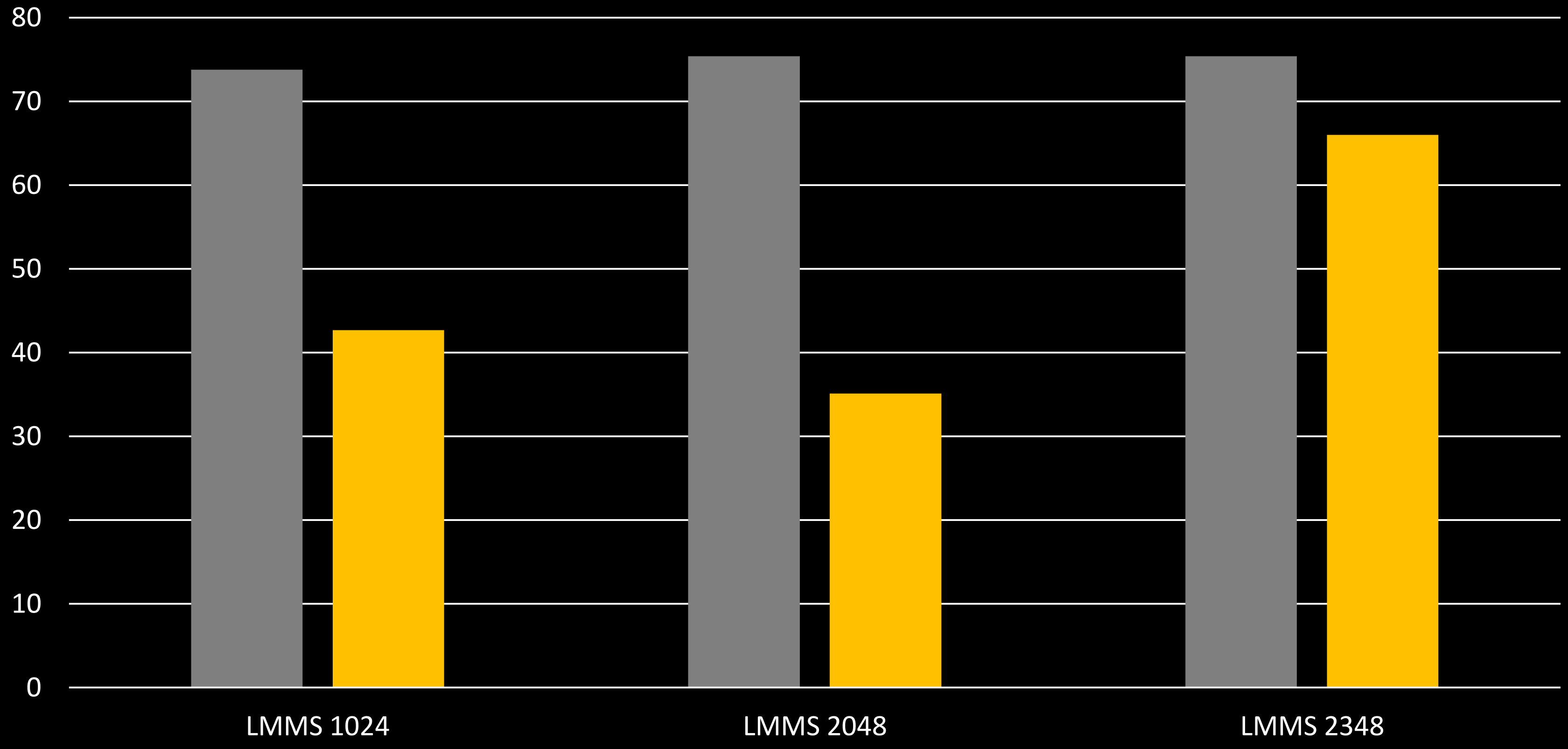


Standard English WSD Evaluation
F1 on ALL set of the WSD Evaluation Framework (Raganato et al. 2017)
80
75
70
65
60
MFS
IMS (Zhong and Ng, 2010)
IMS + Emb. (Iacobacci et al. 2016)
BiLSTM (Raganato et al. 2017)
BiLSTM VR (Vial et al. 2018)
context2vec (Melamud et al. 2016)
ELMo k-NN (Peters et al. 2018)
BERT k-NN (Adapted Peters et al.)
LMMS-BERT (Ours)

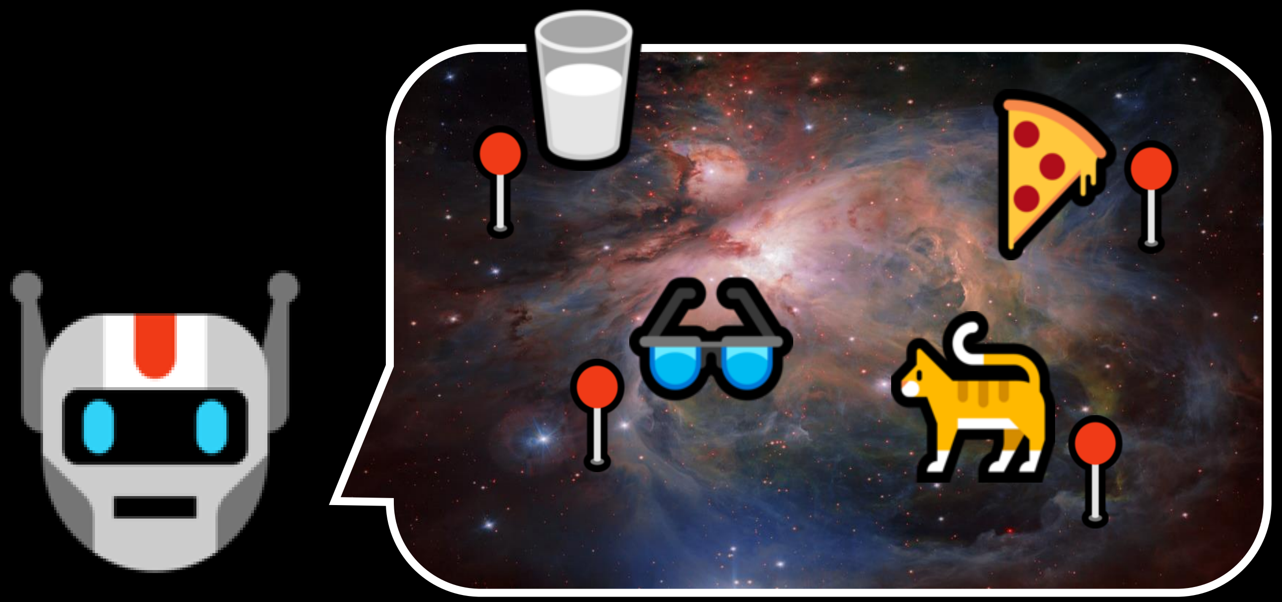
WSD Results

Uninformed Sense Matching (matching +200K)

Same standard but without filtering candidates by lemmas or POS



Applying Sense Embeddings



World Knowledge in NLMs

What's BERT thinking about when he reads?

World Knowledge in NLMs

[E1] played [E2] in [E3]

Marlon★	**Brando**★	**played**	**Corleone**★	**in**	**Godfather**★
$person_n^1$	$person_n^1$	act_v^3	$syndicate_n^1$	$movie_n^1$	$location_n^1$
$womanizer_n^1$	$group_n^1$	$make_v^{42}$	$mafia_n^1$	$telefilm_n^1$	$here_n^1$
$bustle_n^1$	$location_n^1$	$emote_v^1$	$person_n^1$	$final\_cut_n^1$	$there_n^1$

$\mathbf{act_v^3}$: play a role or part; $\mathbf{make_v^{42}}$: represent fictiously, as in a play, or pretend to be or act like; $\mathbf{emote_v^1}$: give expression or emotion to, in a stage or movie role.

Serena★	**Williams**	**played**	**Kerber**★	**in**	**Wimbledon**★
$person_n^1$	$professional\_tennis_n^1$	$play_v^1$	$person_n^1$	win_v^1	$tournament_n^1$
$therefore_r^1$	$tennis_n^1$	$line\_up_v^6$	$group_n^1$	$romp_v^3$	$world\_cup_n^1$
$reef_n^1$	$singles_n^1$	$curl_v^5$	$take\_orders_v^2$	$carry_v^{38}$	$elimination\_tournament_n^1$

$\mathbf{play_v^1}$: participate in games or sport; $\mathbf{line\_up_v^6}$: take one's position before a kick-off; $\mathbf{curl_v^5}$: play the Scottish game of curling.

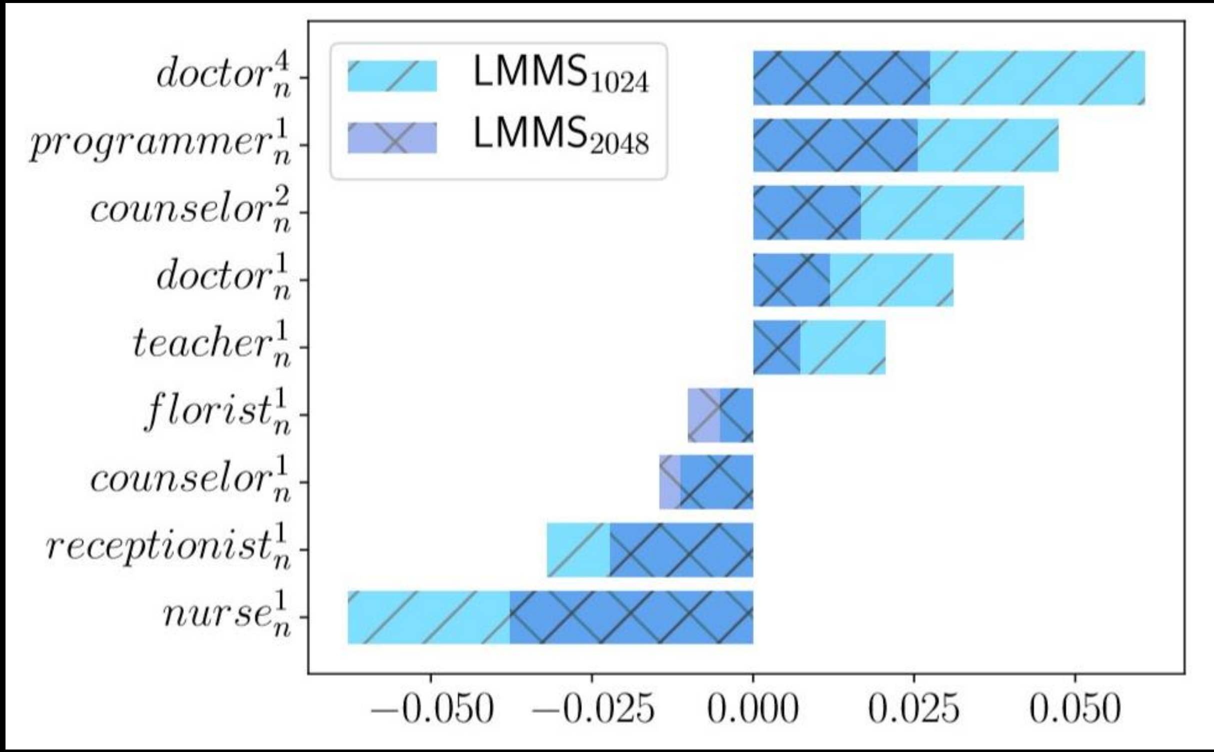
David	**Bowie**★	**played**	**Warszawa**★	**in**	**Tokyo**
$person_n^1$	$person_n^1$	$play_v^{14}$	$poland_n^1$	$originate\_in_v^1$	$tokyo_n^1$
$amati_n^2$	$folk\_song_n^1$	$play_v^6$	$location_n^1$	in_r^1	$japan_n^1$
$guarnerius_n^3$	$fado_n^1$	$riff_v^2$	$here_n^1$	$take\_the\_field_v^2$	$japanese_a^1$

$\mathbf{play_v^{14}}$: perform on a certain location; $\mathbf{play_v^6}$: replay (as a melody); $\mathbf{riff_v^2}$: play riffs.

Checking for Biases in NLMs

Putting BERT on the spot

Checking for Biases in NLMs



$$bias(s) = sim(\vec{v}_{man_n^1}, \vec{v}_s) - sim(\vec{v}_{woman_n^1}, \vec{v}_s)$$

Conclusion

- Powerful NLMs allow for a simple k-NN to perform really well for WSD.
- NLMs are improving very rapidly, progress in WSD should follow.
- Sense embeddings from NLMs are useful not only for WSD, but also for NLM inspection, and other probing or downstream tasks.

Future Work

- Pipeline Improvements: Better NLMs, sentence embeddings, char embeddings, use of WN, etc..

- Multilingual Sense Embeddings.

- Semi-supervised Refinement.

- Formalize inspection (probing task), other applications.

Thanks

Code and Sense Embeddings:
github.com/danlou/LMMS

dloureiro@fc.up.pt

@danielbloureiro