A Deeper Look into Dependency-Based Word Embeddings

Sean MacAvaney and Amir Zeldes

To be presented at NAACL SRW 2018
Preprint: arxiv.org/abs/1804.05972

Word Embeddings

- Goal: Represent the meanings of words in a dense vector

$$V_{cat} = [-0.182, -0.037, -0.166, -0.165, 0.013, \dots, 0.028, -0.088, 0.113]$$

- Train a neural model that predicts a word given the words around it (or vise versa)
- Words with *similar* meanings should be located at a similar place in the embedding space

- Problem: How do you define word similarity?

Relatedness vs. Similarity

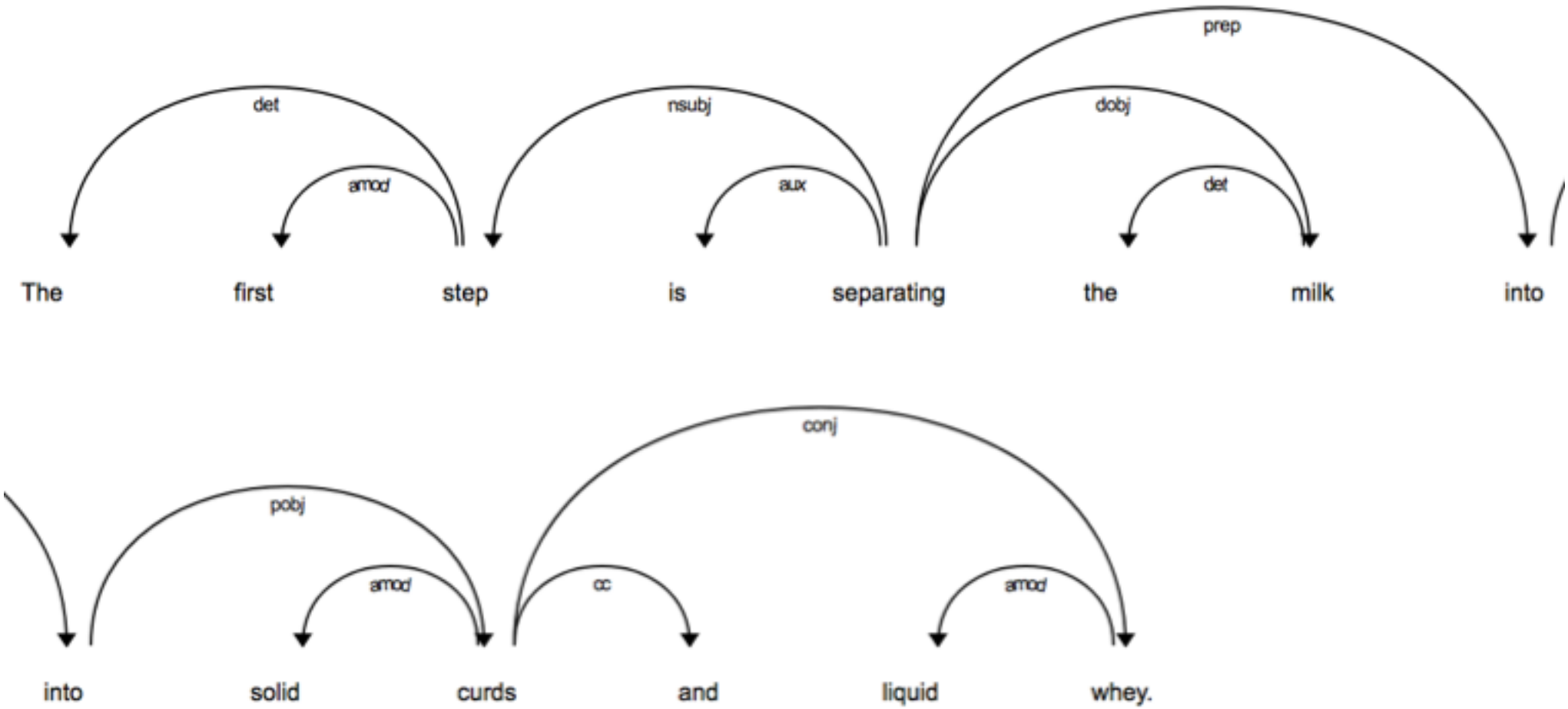
Related terms: (topically similar)	Similar terms: (functionally similar)
Hogwarts + Dumbledore	Hogwarts + Sunnydale
Camel + Hump	Camel + Bear
Physics + Proton	Physics + Chemistry
Ball + Bat	Ball + Bat

Functionally similar words can often fill the same slots in a sentence:
My major was _____ in college.

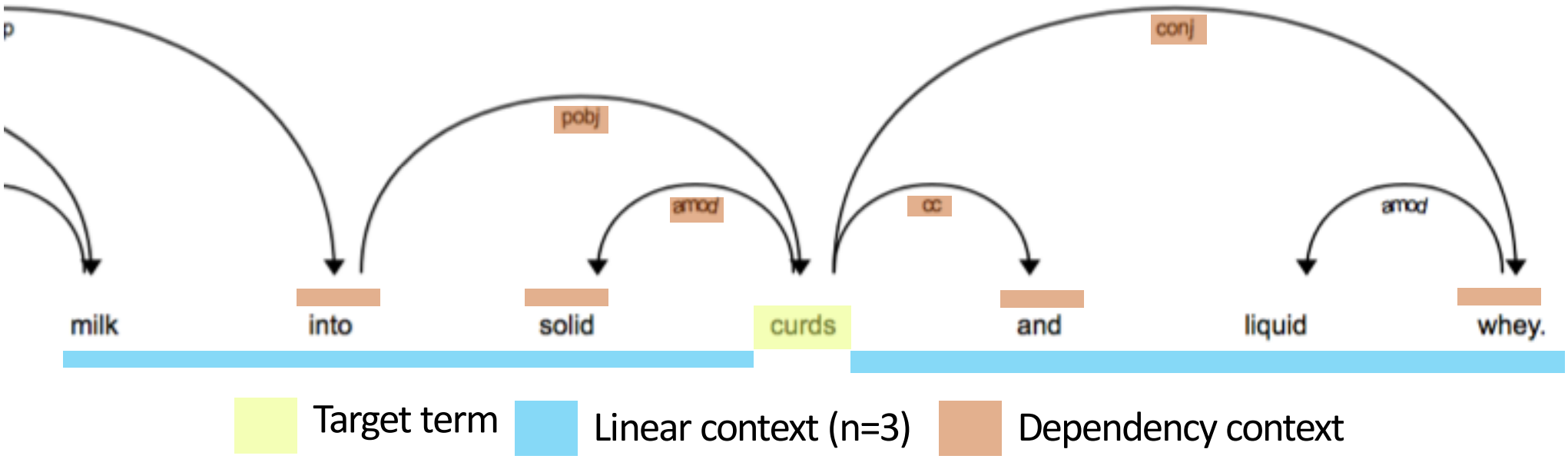
Dependency-Based Word Embeddings

- Levy and Goldberg (2014) challenged the practice of using linear term context windows when training word embeddings
- Embeddings trained with dependency context windows perform better at distinguishing similarity over relatedness

Dependency Parse of Sentence



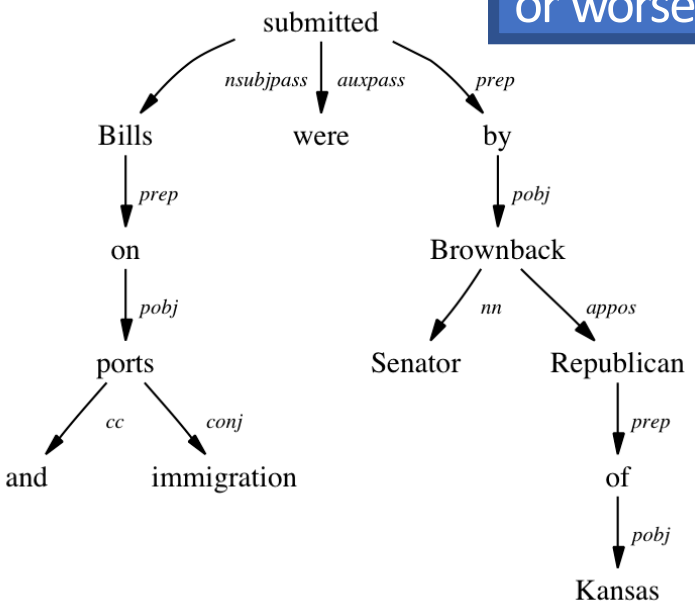
What's in a dependency context?



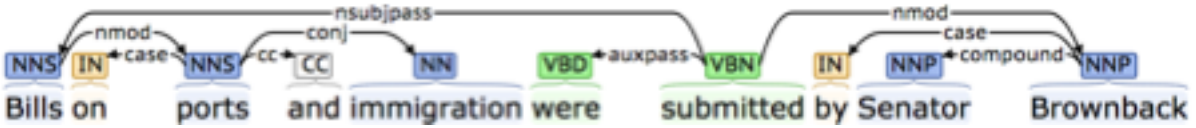
What's in a dependency context?

- Many decisions to be made, though…
- Stanford vs Universal dependencies?

Are the English-specific Stanford Dependencies better or worse than the general Universal Dependencies?



Stanford Dependencies



Universal Dependencies

(Levy and Goldberg (2014) only examined Stanford dependencies.)

What's in a dependency context?

- Many decisions to be made, though…
- Levels of enhancement:
  - Are the words in the context labeled?

Curds

Unlabeled Context:	**"Basic" Context:**
into	into**:pobj-reverse**
solid	solid**:amod**
and	and**:cc**
whey	whey**:conj**

What's in a dependency context?

- Many decisions to be made, though…
- Levels of enhancement:
  - Are the words in the context labeled?
  - Any post-processing of dependency labels?

Curds

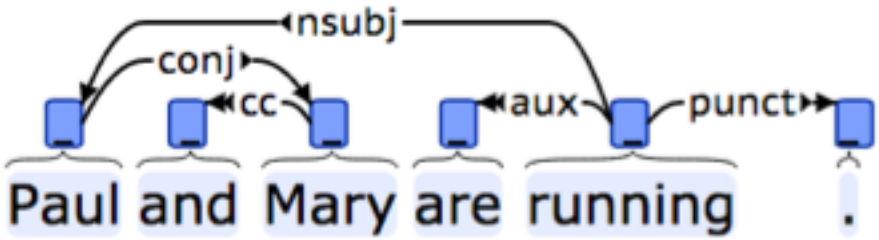
Simple Context:
into**:mod-reverse**
solid**:mod**
and**:coord**
whey**:coord**

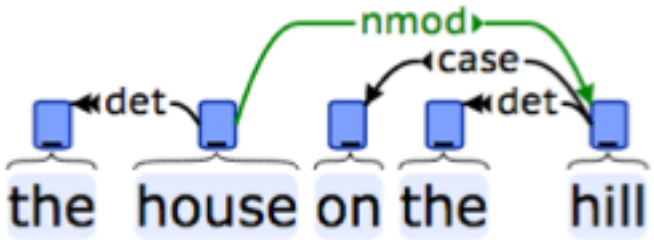
"Basic" Context:
into**:pobj-reverse**
solid**:amod**
and**:cc**
whey**:conj**

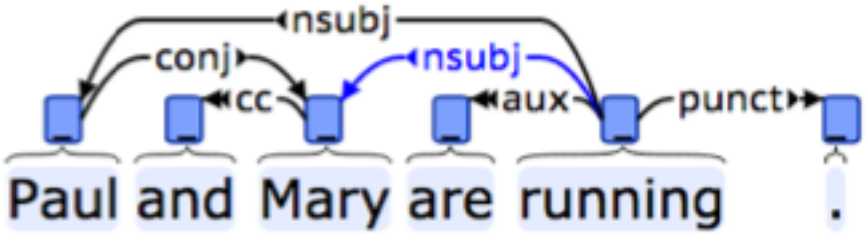
In simplified contexts, functionally similar dependency labels are collapsed.

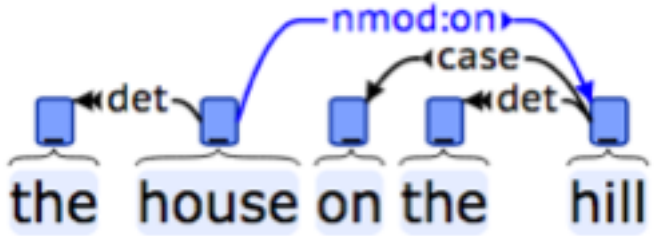
What's in a dependency context?

- Many decisions to be made, though…
- Levels of enhancement:
  - Are the words in the context labeled?
  - Any post-processing of dependency labels?





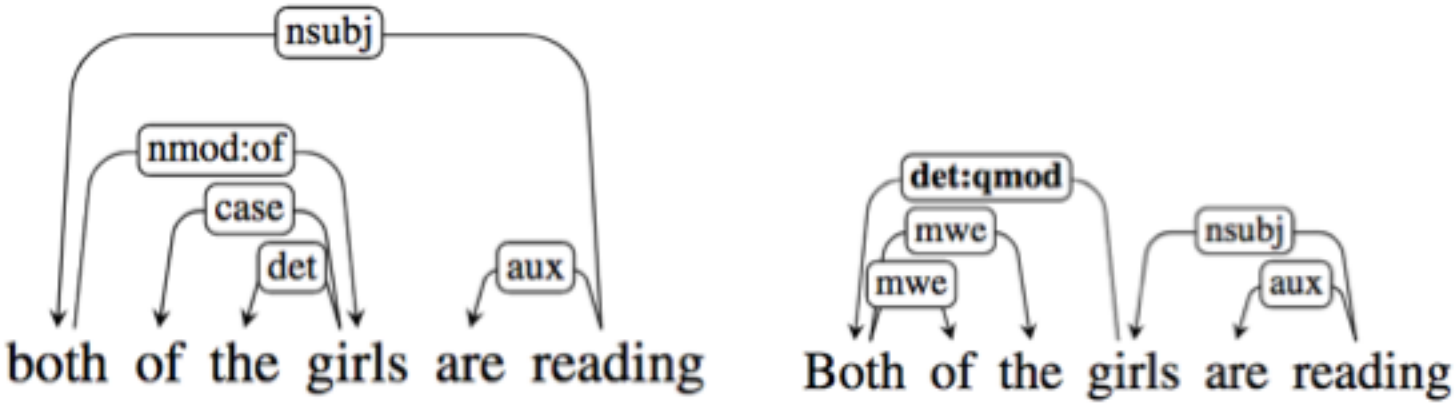




"Enhanced" dependencies: cross-lingual phenomena, e.g. conjoined subjects and case info.

What's in a dependency context?

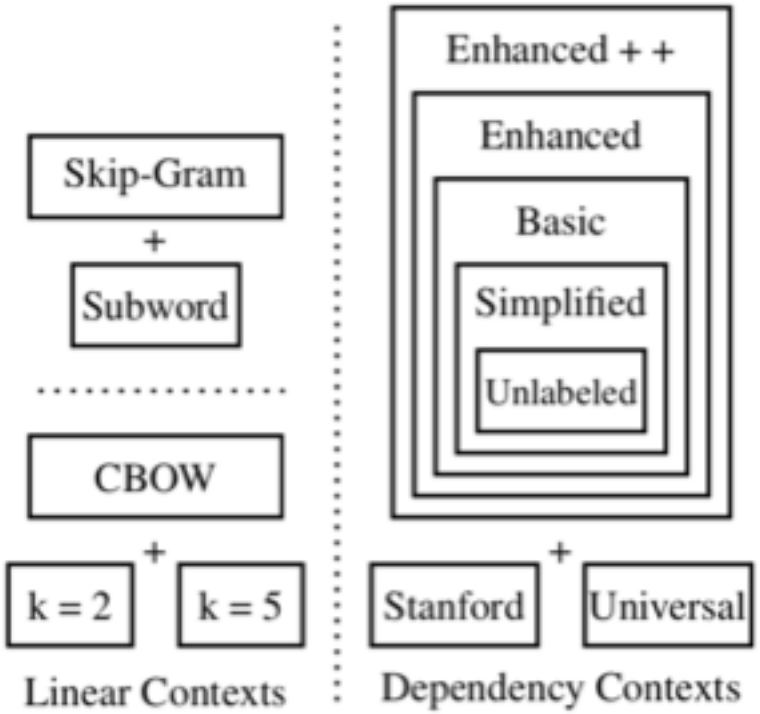
- Many decisions to be made, though…
- Levels of enhancement:
  - Are the words in the context labeled?
  - Any post-processing of dependency labels?



"Enhanced++" dependencies: English-tailored phenomena (e.g., partitive noun phrases).

Experiment

- Compare various types of dependency-based word embeddings to linear-context word embeddings



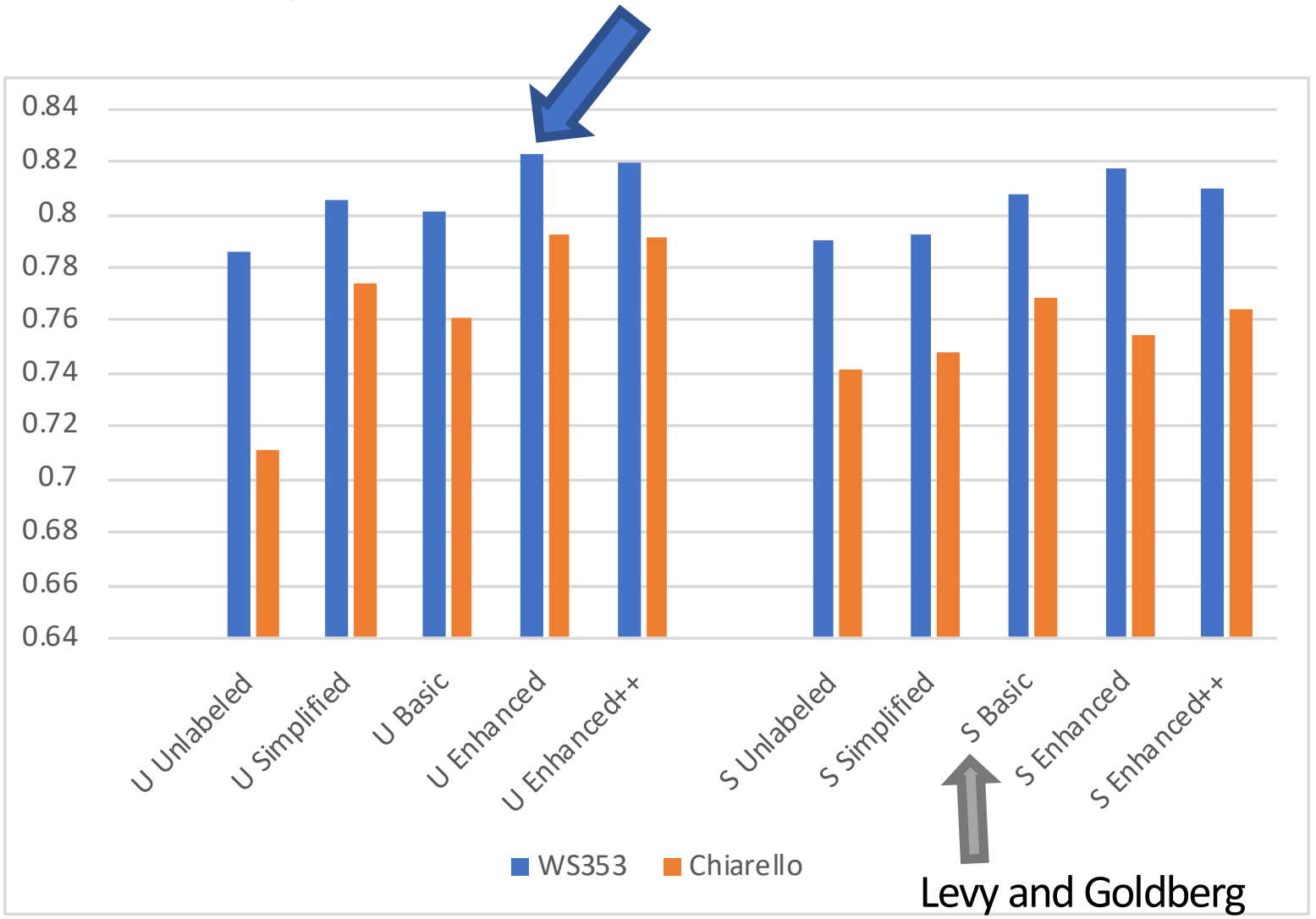
Experiment

- Compare various types of dependency-based word embeddings to linear-context word embeddings
- Use the CoreNLP dependency parser on a recent dump of Wikipedia (Nov 2017)
- Various evaluations:
  - How well do they rank similar words over related words?
  - What effect do they have on downstream tasks?

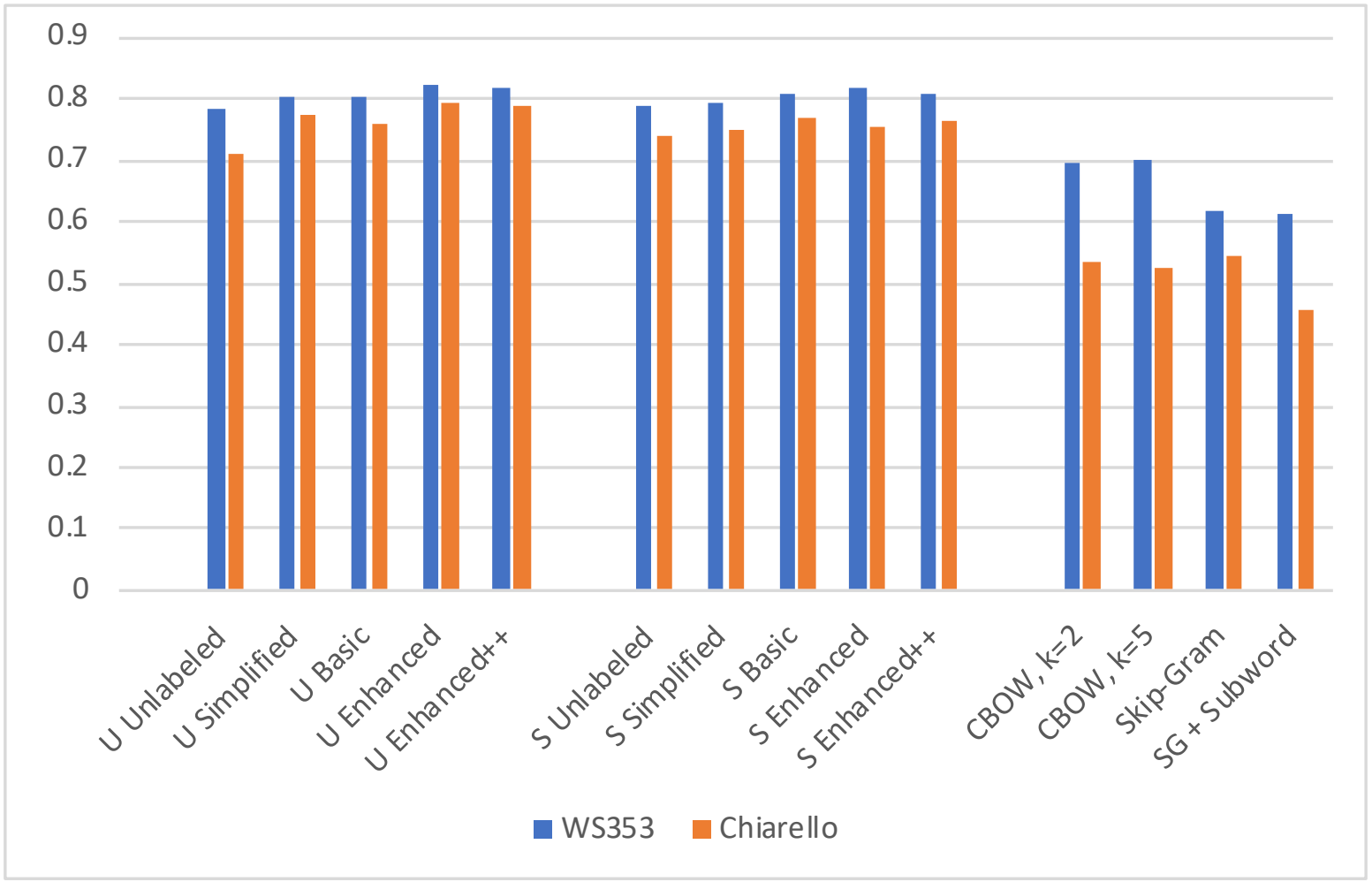
Similarity over Relatedness

- Given set of similar and related word pairs
  - Annotated WordSim-353 (Agirre et al., 2009)
  - Manually-curated (Chiarello et al., 1990)
- Measure area under precision-recall curve

Similarity over Relatedness



Similarity over Relatedness



Ranked Similarity

- We also examined the rankings of similar words, and found the rankings to be comparable.
- E.g., ceiling-roof > ceiling-floor > ceiling-cathedral

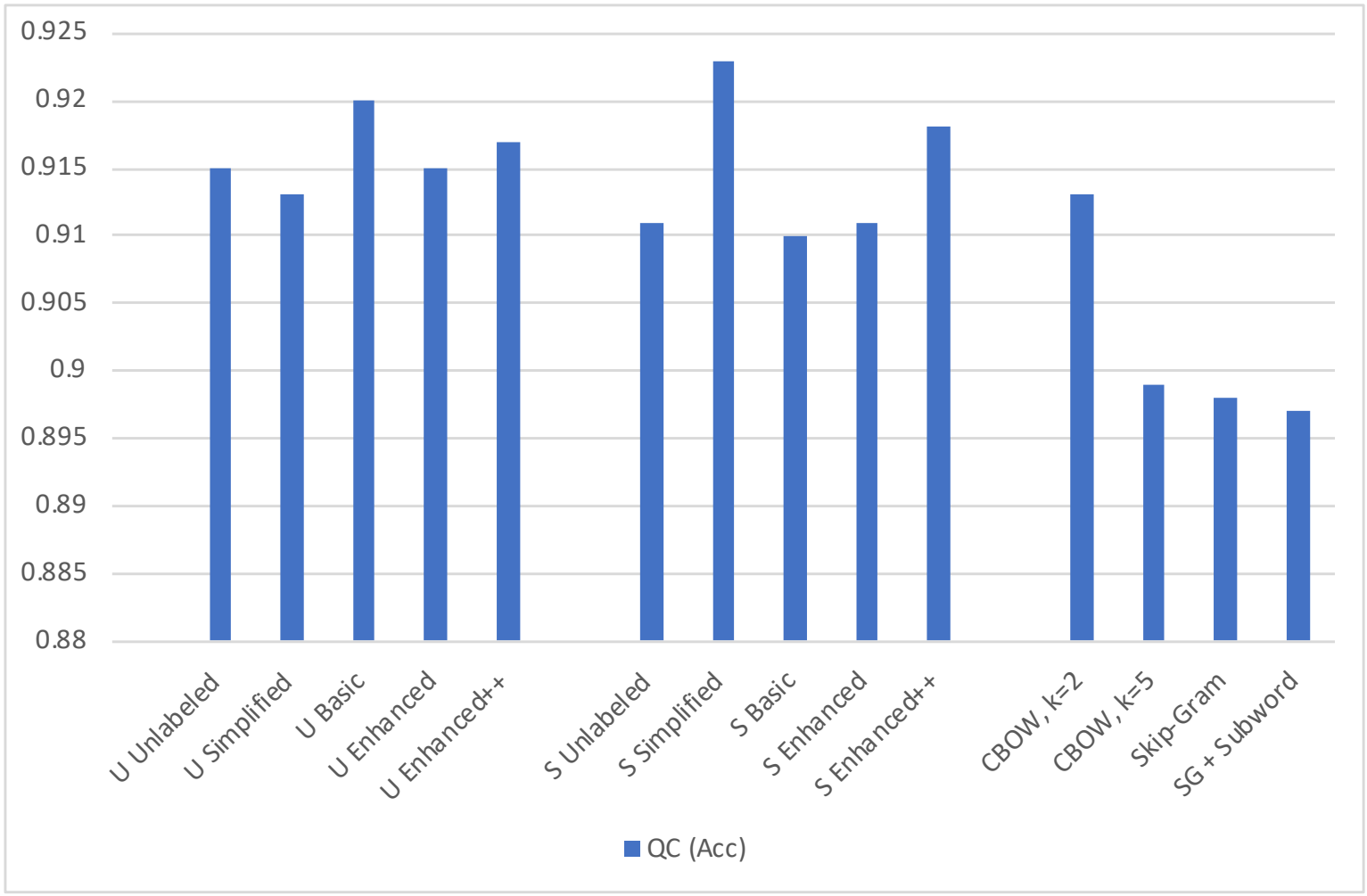
Downstream Tasks

- Fixed embeddings – no training
- Two tasks (Bi-directional LSTM implementations):
  - Question type classification (TREC QC dataset) [1]
  - Named entity recognition (CoNLL NER dataset) [2]

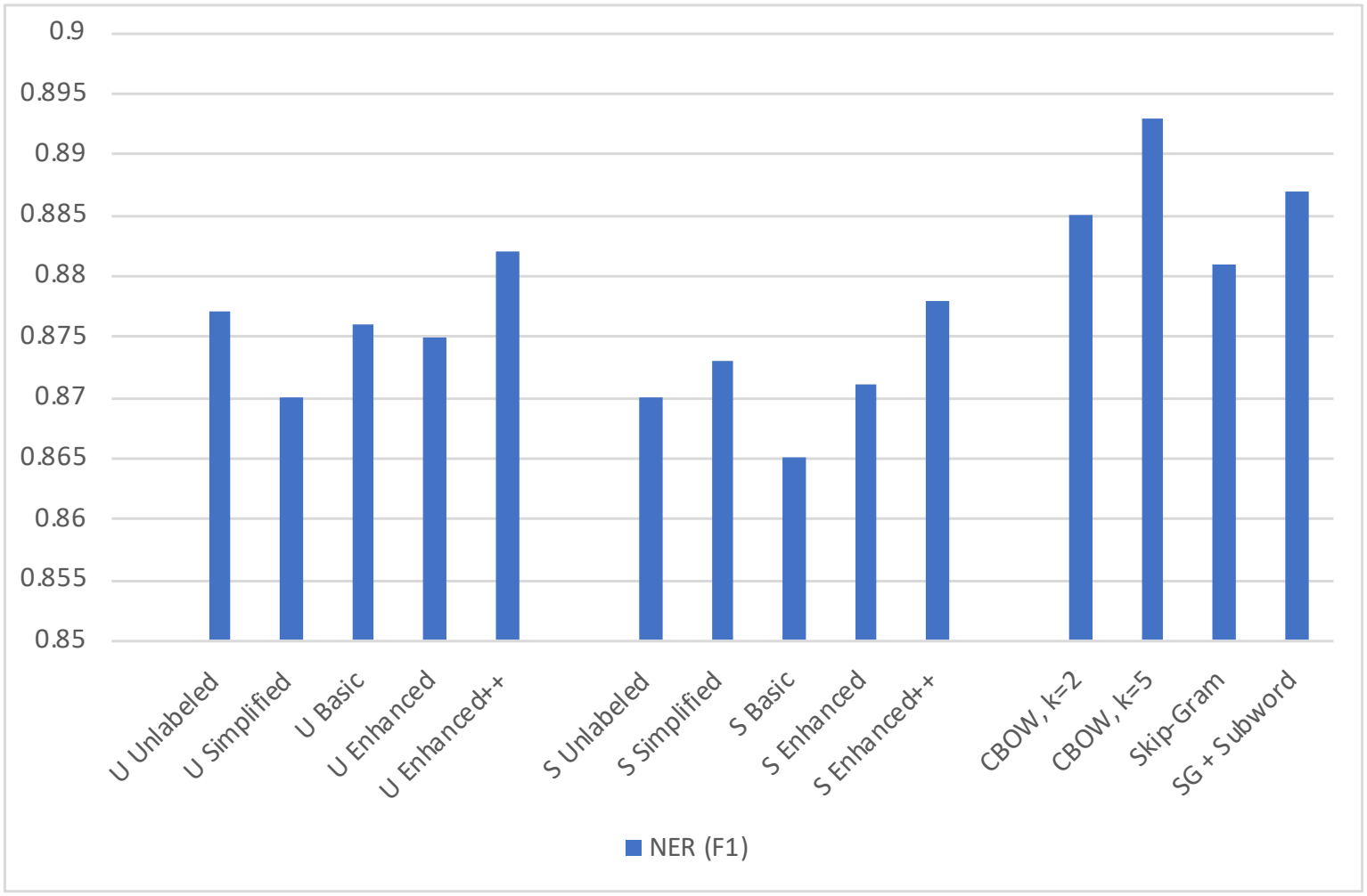
[1] https://github.com/zhegan27/sentence_classification

[2] https://github.com/Franck-Dernoncourt/NeuroNER

Downstream Tasks - QC



Downstream Tasks - NER



Conclusion

- Enhancing Universal dependencies can help distinguish similarity over relatedness (compared to previous work)
- Dependency-based word embeddings are valuable for some tasks
- Future work
  - More downstream tasks
  - Cross-lingual embeddings

- Questions?