Inducing multi-sense word representations multilingually

Simon Šuster*°, Ivan Titov◊ and Gertjan van Noord*

*Dept. of Language Technology, University of Groningen
° CLiPS, University of Antwerp
◊Institute for Logic, Language and Computation, University of Amsterdam

Motivation

Ambiguity in L1 can correspond to smaller ambiguity in L2
[Snyder and Barzilay, 2010]

Disambiguate **polysemy** in L1 by looking at how words **translate**
[Diab, 2003, Brown et al., 1991]

- Translated words can be ”monosemous”
- L1 and L2 polysemies shouldn’t overlap
- Context around the translated word

Word embeddings

- Multi-sense [Neelakantan et al., 2014, Li and Jurafsky, 2015]
  - typically monolingual
- Multilingual
  - embeddings in the same semantic space [Gouws et al., 2014, Klementiev et al., 2012]
  - use target-language signal for better source-language embeddings [Hill et al., 2014, Faruqui and Dyer, 2014]

Can L2 signal improve multi-sense embeddings in L1?

Joint-learning scenario [Titov and Khoddam, 2015]

- **Encoding**: learn sense inventory and mapping
  - L2 in addition to L1 here
- **Decoding**: learn sense-specific word embeddings

rock_0
---
mud 0.897
grass 0.877
deep 0.874
sea 0.872
cloud 0.870
bush 0.858
canopy 0.856
reef 0.855
rough 0.851
vine 0.849
hollow 0.844
surrounding 0.841
boulder 0.840
leaf 0.839
spiral 0.839

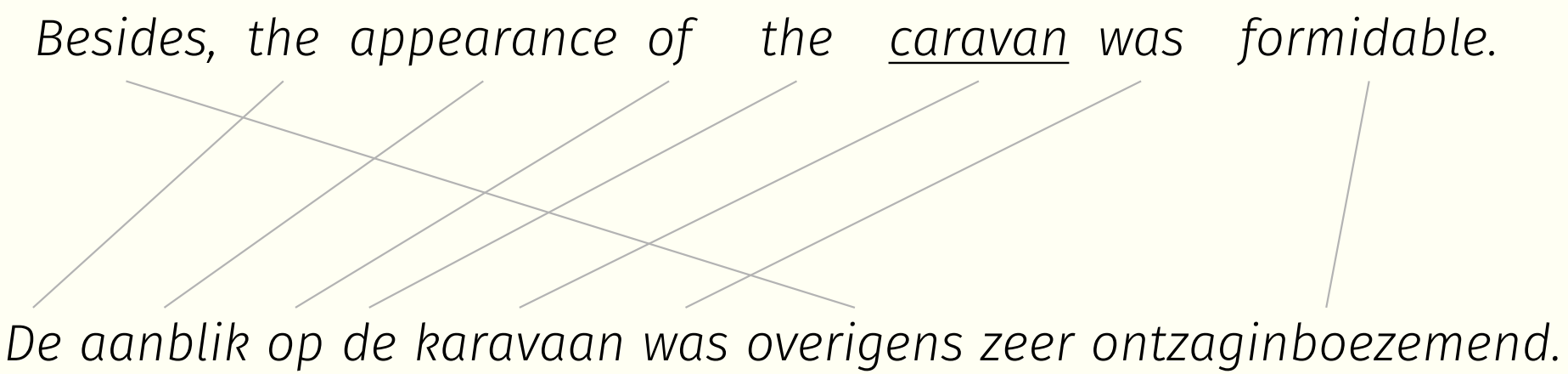
rock_1
---
band 0.919
pop 0.907
rapper 0.872
indie 0.870
punk 0.860
album 0.823
duo 0.820
supergroup 0.811
singer 0.784
metal 0.783
trio 0.781
songwriter 0.773
guitarist 0.764
Pop 0.759
metalcore 0.758

rock_2
---
disco 0.899
pop 0.891
roll 0.883
gospel 0.882
hip 0.867
psychedelic 0.862
hardcore 0.856
jazz 0.852
hop 0.847
contemporary 0.846
mainstream 0.842
grunge 0.841
techno 0.839
glam 0.837
progressive 0.836

Illustration

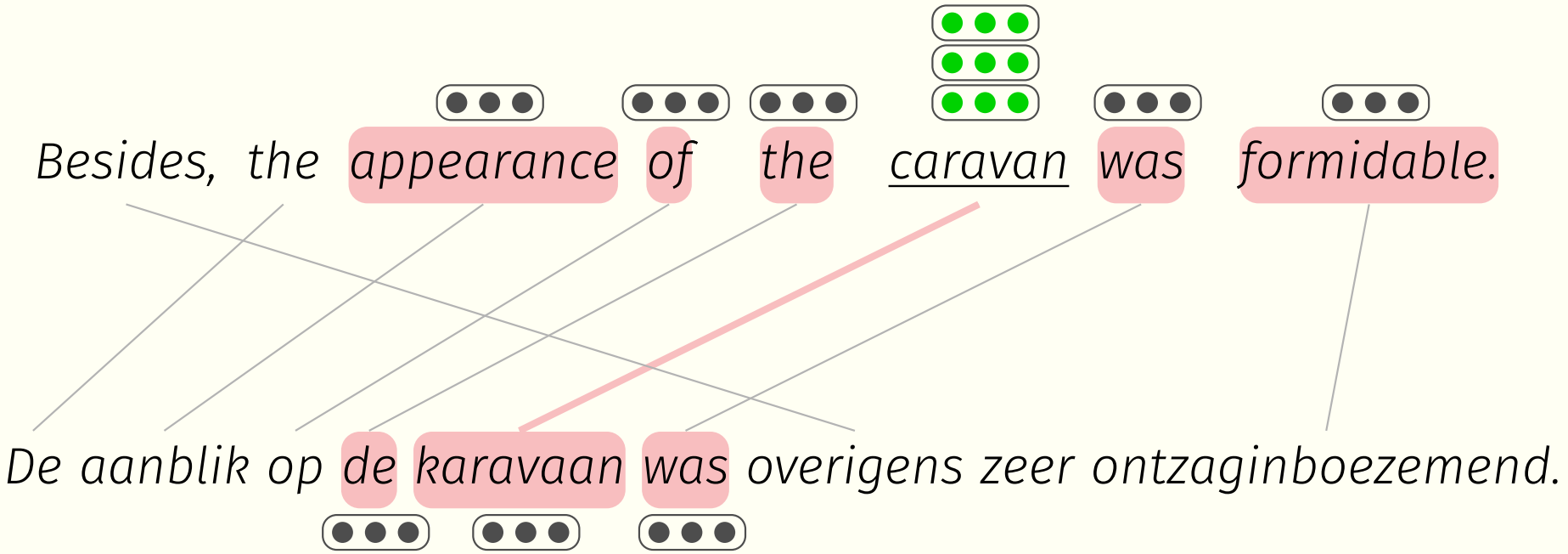
Besides, the appearance of the caravan was formidable.

Illustration



Source: `http://opus.lingfil.uu.se/Books/`

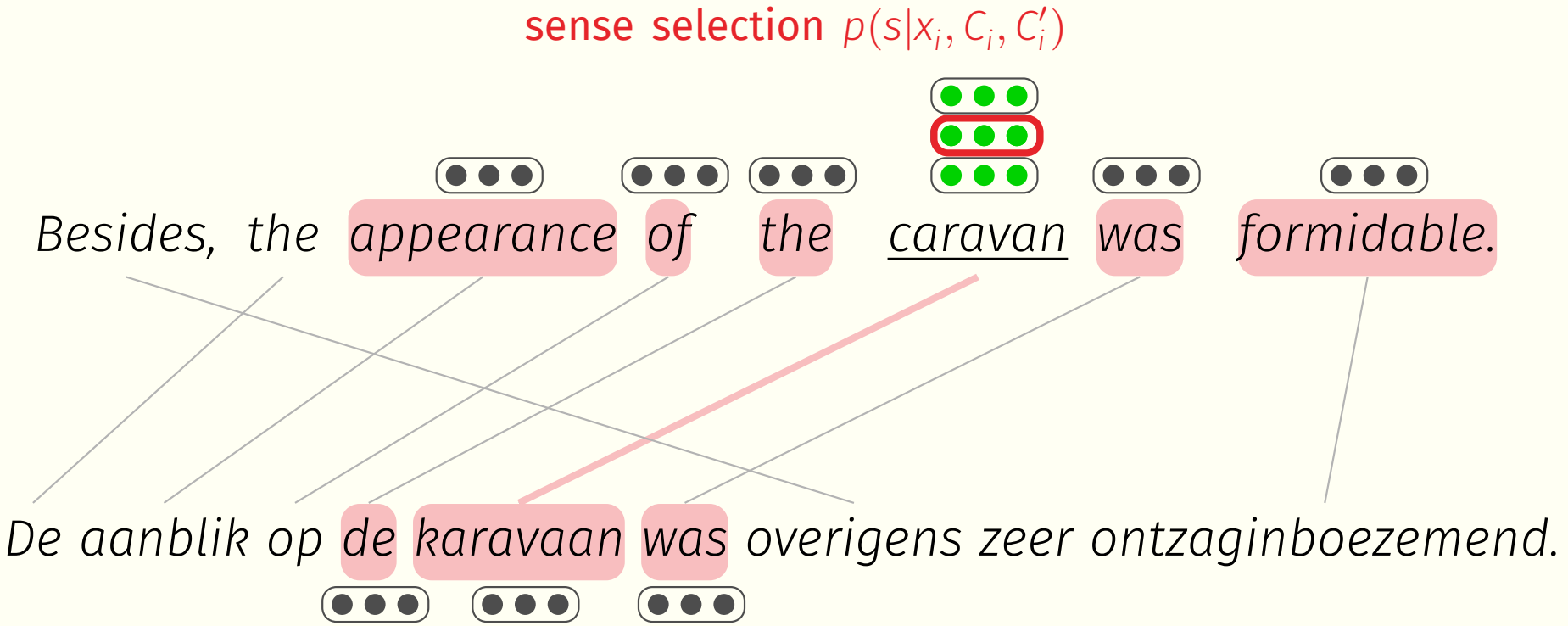
Illustration



— L1/L2 generic vector

— sense-specific vector

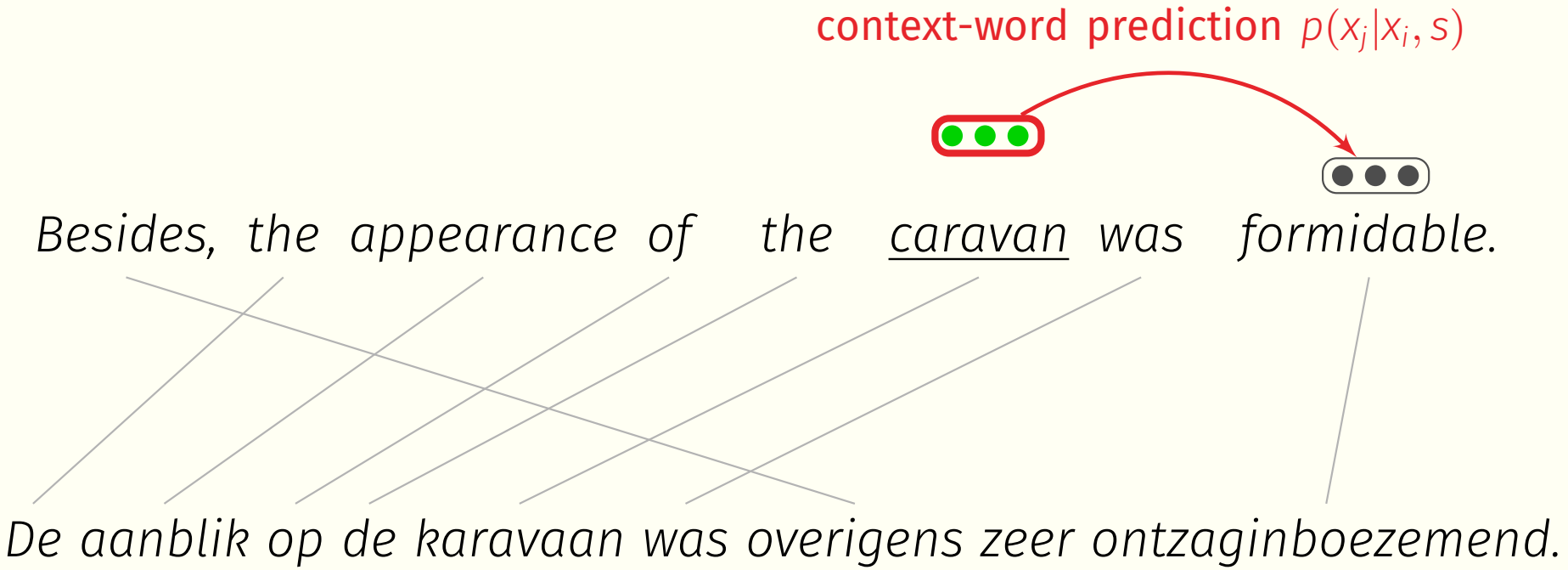
Illustration

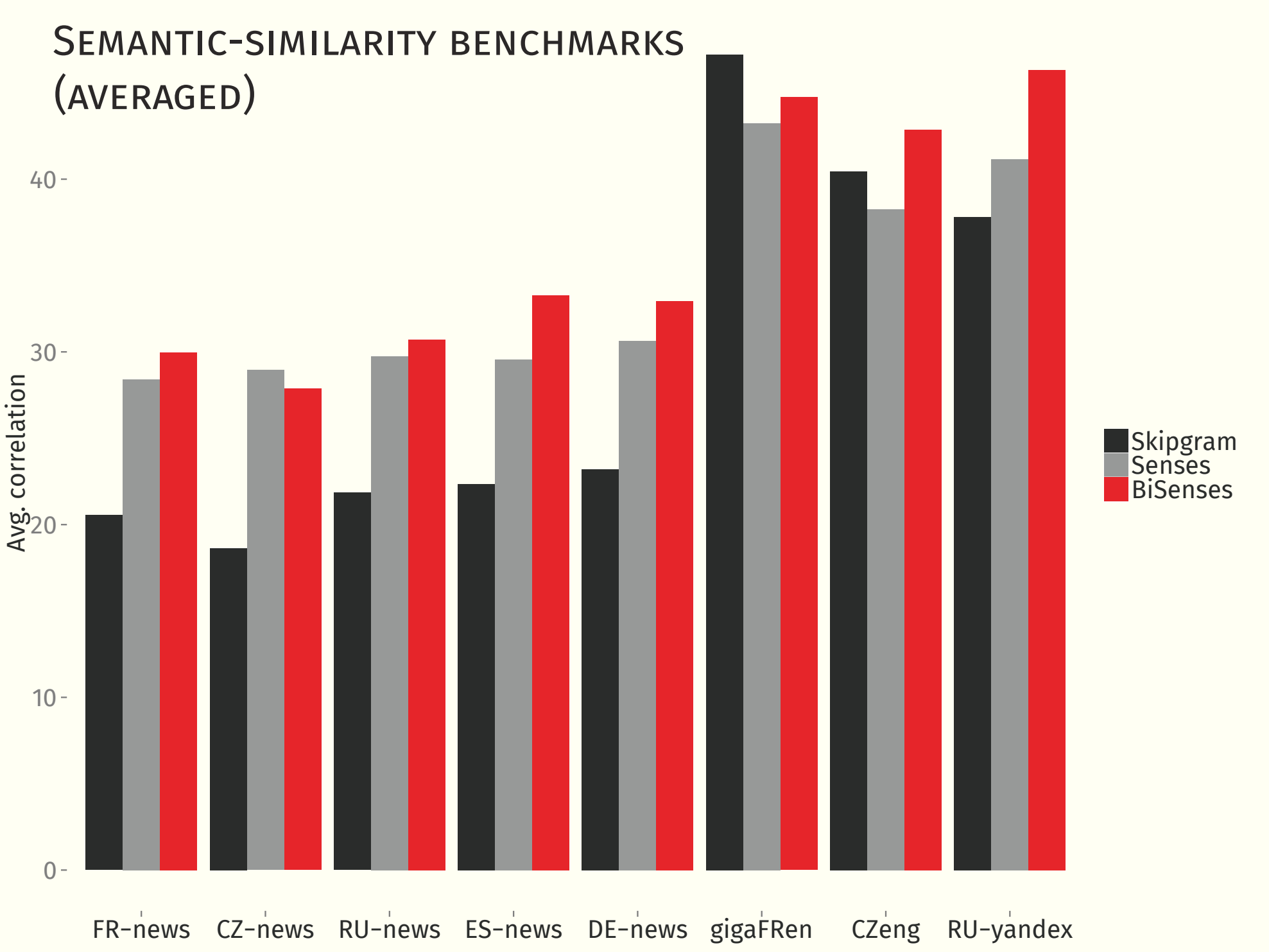


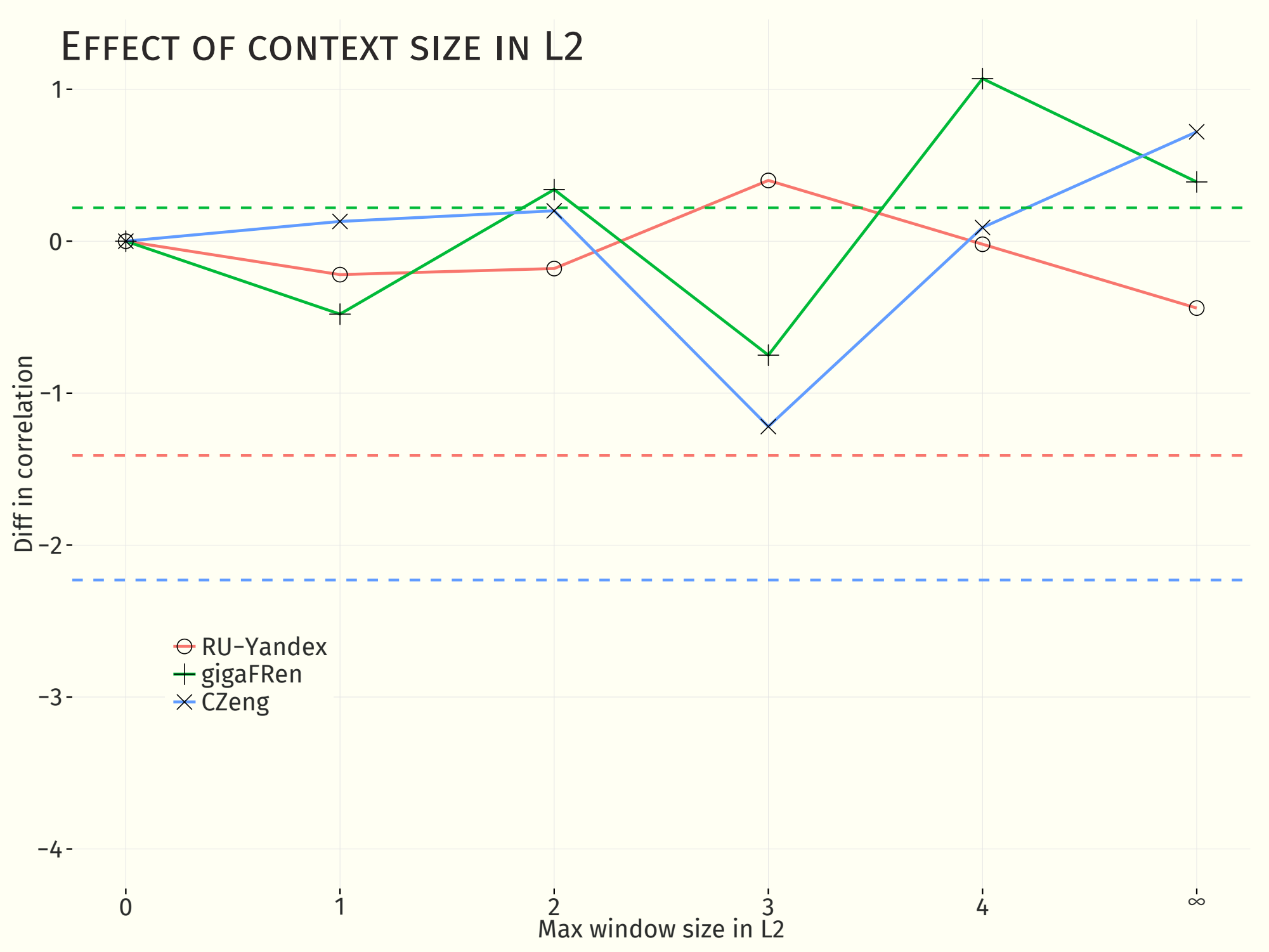
— L1/L2 generic vector

— sense-specific vector

Illustration

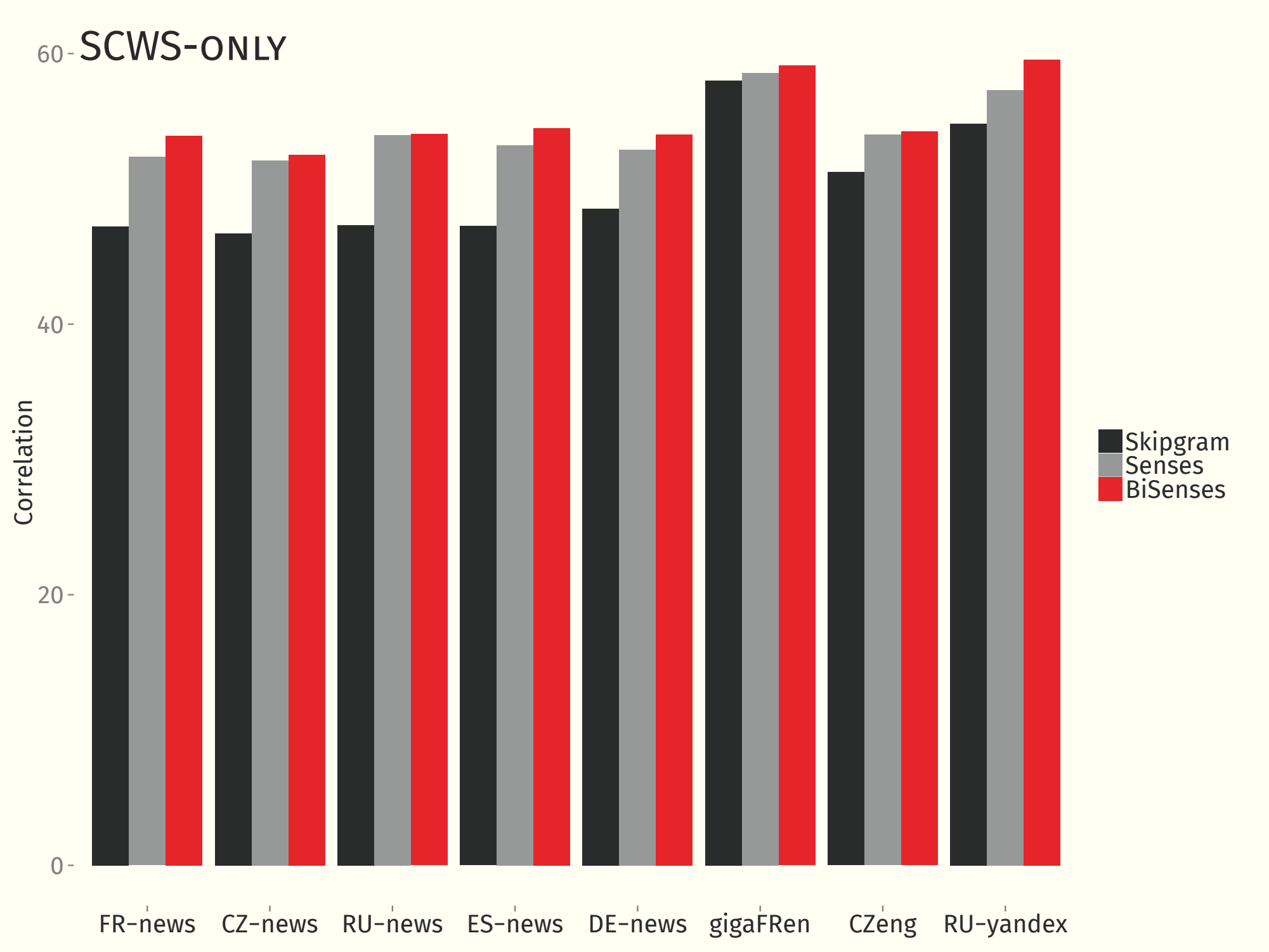



Semantic-similarity benchmarks (averaged)
Avg. correlation
0
10
20
30
40
FR-news
CZ-news
RU-news
ES-news
DE-news
gigaFRen
CZeng
RU-yandex
Skipgram
Senses
BiSenses


Effect of context size in L2
1
0
−1
−2
−3
−4
Diff in correlation
RU−Yandex
gigaFRen
CZeng
0
1
2
3
4
∞
Max window size in L2

This talk

- Jointly learning the sense predictor and the embeddings
- The role of bilingual training:
  - L2 signal improves L1 multi-sense embeddings intrinsically
  - uniform (sentence) alignment might be sufficient


SCWS-ONLY
60
40
20
0
Correlation
FR-news
CZ-news
RU-news
ES-news
DE-news
gigaFRen
CZeng
RU-yandex
Skipgram
Senses
BiSenses

Brown, P. F., Pietra, S. A. D., Pietra, V. J. D., and Mercer, R. L. (1991).
Word-sense disambiguation using statistical methods.
In *ACL*.

Diab, M. T. (2003).
Word sense disambiguation within a multilingual framework.
PhD thesis, University of Maryland at College Park.

Faruqui, M. and Dyer, C. (2014).
Improving vector space word representations using multilingual correlation.
In *EACL*.

Gouws, S., Bengio, Y., and Corrado, G. (2014).
BilBOWA: Fast Bilingual Distributed Representations without Word Alignments.
arXiv preprint arXiv:1410.2455.

Hill, F., Cho, K., Jean, S., Devin, C., and Bengio, Y. (2014).
Embedding word similarity with neural machine translation.
arXiv preprint arXiv:1412.6448.

Klementiev, A., Titov, I., and Bhattarai, B. (2012).
Inducing crosslingual distributed representations of words.
In *COLING*.

Li, J. and Jurafsky, D. (2015).
Do multi-sense embeddings improve natural language understanding?
In *EMNLP*.

Neelakantan, A., Shankar, J., Passos, A., and McCallum, A. (2014).
Efficient non-parametric estimation of multiple embeddings per word in vector space.
In *EMNLP*.

Snyder, B. and Barzilay, R. (2010).
Climbing the Tower of Babel: Unsupervised Multilingual Learning.
In *ICML*.

Titov, I. and Khoddam, E. (2015).

Unsupervised induction of semantic roles within a reconstruction-error minimization framework.
In *NAACL*.