A closer look at scalar diversity using contextualized semantic similarity

Matthijs Westera & Gemma Boleda, Universitat Pompeu Fabra

Sinn und Bedeutung 2019

1

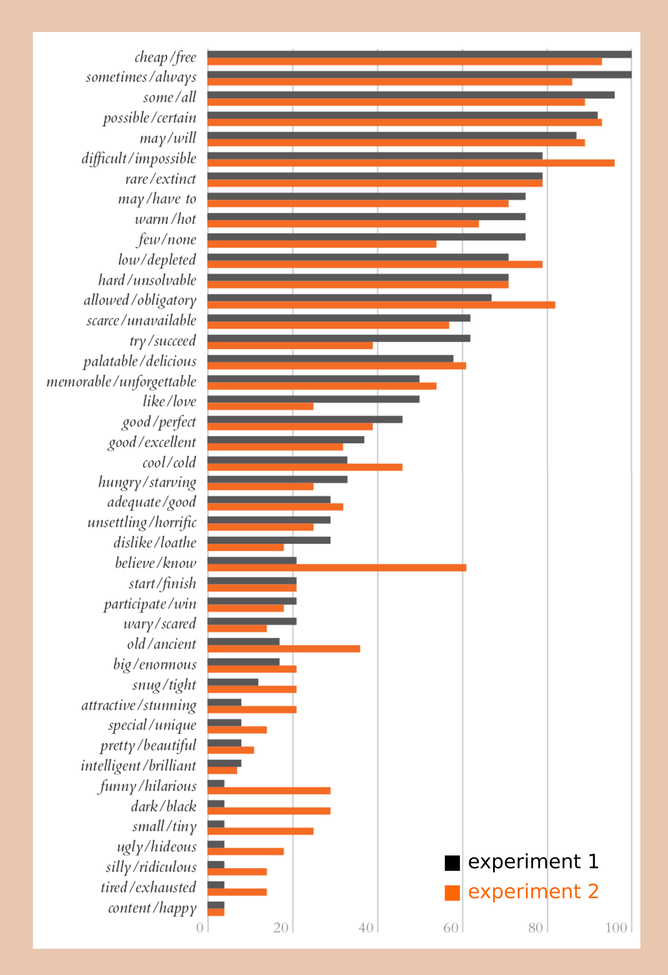
Previous results on **scalar diversity** are explained better by a notion of **semantic similarity** if it is **context-sensitive**.

2 Van Tiel et al. (2016)

▶ Testing pairs of adjectives, e.g., warm/hot:

John says: "*the {sand / soup / ...} is* ***warm***".
Would you conclude from this that, according to John, it is not **hot**? [Yes/No]

▶ They find **scalar diversity** (➡).



▶ And consider various explanatory factors:

Van Tiel et al. (2016)'s full mixed model:

Parameter	β	SE	Z	P	R^2
(Intercept)	−2.80	1.73	−1.62	0.104	–
Association strength	0.16	0.31	0.51	0.611	0.000
Grammatical class	−0.38	0.74	−0.52	0.606	0.001
Relative frequency	−0.15	0.21	−0.74	0.461	0.003
Semantic relatedness (LSA)	0.1	0.1	0.93	0.355	0.006
Semantic distance	0.65	0.27	2.36	0.018	0.027
Boundedness	−1.87	0.40	−4.72	0.000	0.108

3 McNally (2017)

▶ **How come semantic relatedness shows no effect?**

▶ LSA (see A⬇) is too coarse-grained, assigning similarity scores to words regardless of their context.

▶ But real semantic similarity is affected by what is likely relevant, given other words in the sentence:

(1) *"The sand is warm."* → not hot
Likely QUD: *"Is the sand safe to walk on?"*

(2) *"The soup is warm."* ↛ not hot
Likely QUD: *"Is it a warm or a cold soup?"*

(3) *"The salary is adequate."* ↛ not good
Likely QUD: *"Does it meet one's needs?"*

(4) *"The salary is good."*
Likely QUD: *"How does it compare to similar jobs?"*

▶ We need a notion of semantic similarity that reflects this.

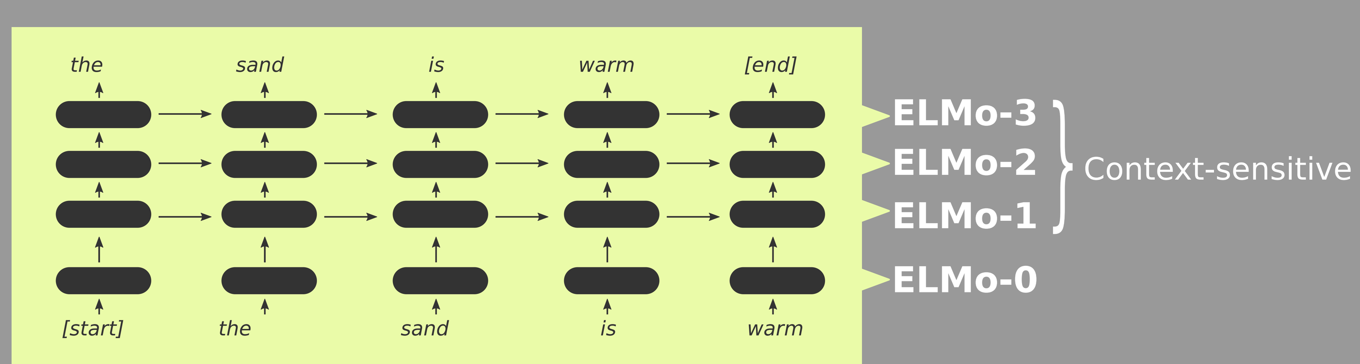
A LSA? ELMo?

▶ **LSA = Latent Semantic Analysis** [3];
ELMo = Embeddings from Language Models [4].

▶ Both represent words as high-dimensional vectors, where similarity = cosine of their angle.

▶ What is different is *how* the vectors are computed:
- **LSA: dimensionality reduction** on co-occurrence counts
- **ELMo: neural network** for predicting the next word.

▶ Specifically, **ELMo** is a 3-layer recurrent neural network:



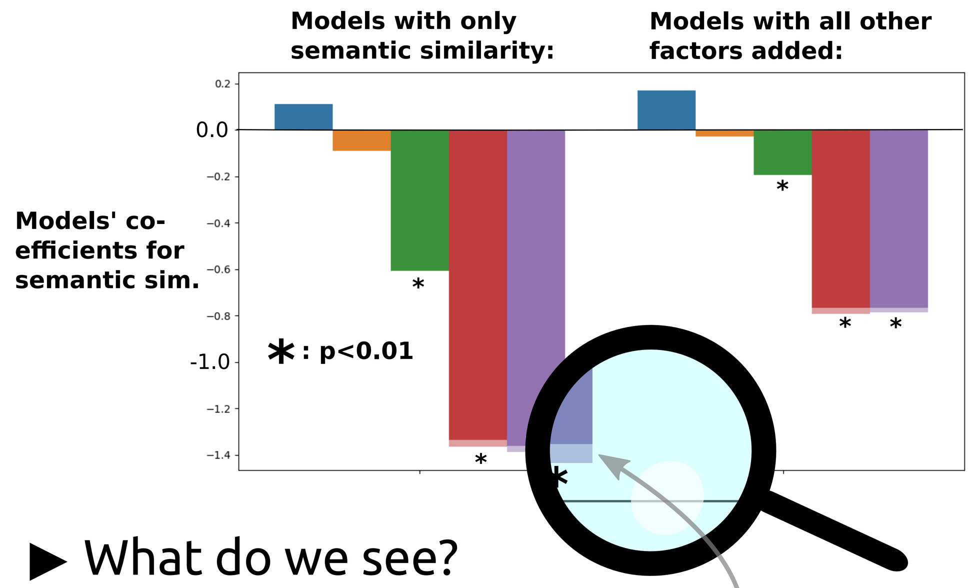
4 Method

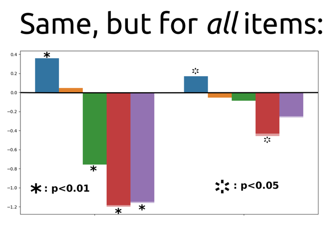
▶ We fit linear models on Van Tiel et al.'s data from exp.2, comparing notions of semantic similarity (see A⬋):
- context-insensitive: **LSA**, **ELMO-0**, vs.
- context-sensitive: **ELMo-1,2,3**.

▶ **Hypothesis:** The context-sensitive notions explain more.

Results

▶ Models for *open-class items* only (removes 4):





▶ What do we see?
- Larger, significant effects for ELMo-1,2,3.
- Even with other factors (and, less significantly, with closed-class items).

▶ But what's this?
- *Slightly* larger effect for ELMo-2,3 when sentences of the same pair (e.g., (1)/(2)) are averaged prior to model-fit.

6 Discussion

Some confirmation of McNally's proposal, but many questions:

▶ What makes ELMo-1,2,3 better than LSA?
- ELMo doesn't seem to need to distinguish sentences;
- But context-sensitivity somehow helps (compare ELMo-0).

▶ Why is the effect of ELMo-1,2,3 negative?
- Alternatives should be similar (lest they be unavailable) but not too similar (lest they won't be excluded); [1].
- Perhaps the task itself makes alternatives available anyhow, leaving only the latter, negative effect.

▶ Need a closer look at the data, e.g.:
- For (1)/(2), Van Tiel et al.'s data doesn't go in the direction suggested; but for (3)/(4) it does.

▶ How do Formal Semantics/Pragmatics relate to LSA, ELMo, and other vector-space models of meaning?

References

[1] Van Tiel, B., Van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar diversity. JoS 33.

[2] McNally, L. (2017). Scalar alternatives and scalar inference involving adjectives: A comment on Van Tiel, et al. 2016. *Essays in honor of Sandra Chung.*

[3] Landauer, T. K., P. W. Foltz & D. Laham. (1998), 'Introduction to latent semantic analysis'. *Discourse Processes* 25:259–84.

[4] Peters, M.E., Neumann, M., Iyyer, M., Gardner,M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *NAACL.*

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 715154). And from the Ramón y Cajal programme (grant RYC-2015-18907). We gratefully acknowledge the support of NVIDIA Corporation with the donation of GPUs used for this research. This paper reflects the authors' view only, and the EU is not responsible for any use that may be made of the information it contains.