Novel Aficionados and Doppelgängers: learning and evaluating distributional representations of individual entities

Andrea Bruera
Queen Mary University of London
School of Electronical Engineering and Computer Science
a.bruera@qmul.ac.uk

Aurélie Herbelot
University of Trento
Center for Mind and Brain Sciences

1 The issue with proper names

Learning and retrieving semantic representations for proper names is a task which, unlike other cognitive processes which are much more challenging for computers than for humans (e.g. Lake et al., 2015), seems difficult for both human beings (Semenza, 2009, Brédart, 2016, Brédart, 2017) and machine learning algorithms (Herbelot, 2015, Gupta et al., 2015, Aina et al., 2019, Almasian et al., 2019). Cognitive studies on the subject abound: it has been consistently found that proper names are both more difficult to acquire and retrieve from memory than common nouns and that, as a result of neurodegenerative diseases or vascular lesions, one category can be cognitively impaired independently of the other (Cohen, 1990, Martins and Farrajota, 2007). But the linguistic properties which make proper names more difficult than common nouns for computers are a relatively unexplored field in computational linguistics and NLP. Understanding this behaviour is important because it entails that proper names may require specific computational strategies to be processed in AI systems.

2 A novel referential task

Proper names are known to have different semantic properties from common nouns: the meaning of a proper name is exclusively the unique individual entity it refers to, whereas common nouns refer to classes of individuals (Kripke, 1972). The main hypothesis of our work is that this difference in semantic properties can be retrieved by distributional representations of meaning when tested over an appropriate referential task. That is, the distinction found in human cognition should be reflected in the distributional properties of proper names and common nouns.

To show that this is the case, we propose an original referential task, the *Doppelgänger test*, associated with a new dataset, the *Novel Aficionados* dataset, made of 59 novels. The Doppelgänger test evaluates whether each entity representation learned in one subcorpus (one half of a novel) can be correctly matched to its co-referring entity representation from another subcorpus (the second half of the same novel), choosing among all the other entity representations. The task is challenging in that the model must distinguish between very similar entities (people and entities engaged in shared activities in a common universe) using scarce data.

3 Results

Using the Doppelgänger test, we compare the distributional representations of the proper names referring to the novels' characters and those of similarly frequent common nouns mentioned in the novel. For robustness, we use several models (ELMO: Peters et al., 2018, BERT: Devlin et al., 2018, Word2Vec: Mikolov et al., 2013, and Nonce2Vec: Herbelot and Baroni, 2017). Distinct patterns emerge for the two linguistic categories across all models, mirroring human cognition.

As further analysis, we perform an RSA correlation study (Kriegeskorte et al., 2008), as well as introducing a challenging variation on the Doppelgänger test which requires linking entities across different corpora (the original novels and Wikipedia). These analyses highlight the disruptive effect of competing semantic representations, which disproportionately affect reference resolution for proper names, drawing a parallel with effects found in human semantic cognition (Abrams and Davis, 2017). Finally, we conduct a POS-based distributional analysis of the dataset, which shows that indeed characters and common nouns have a different distributional signature.

References

Lise Abrams and Danielle K Davis. 2017. Competitors or teammates: how proper names influence each other. *Current Directions in Psychological Science*, 26(1):87–93.

Laura Aina, Carina Silberer, Ionut Sorodoc, Matthijs Westera, and Gemma Boleda. 2019. What do entity-centric models learn? insights from Entity Linking in multi-party dialogue. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3772–3783.

Satya Almasian, Andreas Spitz, and Michael Gertz. 2019. Word embeddings for entity-annotated texts. In *European Conference on Information Retrieval*, pages 307–322. Springer.

Serge Brédart. 2016. Names and cognitive psychology. In *The Oxford Handbook of Names and Naming*.

Serge Brédart. 2017. The cognitive psychology and neuroscience of naming people. *Neuroscience & Biobehavioral Reviews*, 83:145–154.

Gillian Cohen. 1990. Why is it difficult to put names to faces? *British Journal of Psychology*, 81(3):287–297.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Abhijeet Gupta, Gemma Boleda, Marco Baroni, and Sebastian Padó. 2015. Distributional vectors encode referential attributes. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 12–21.

Aurélie Herbelot. 2015. Mr Darcy and Mr Toad, gentlemen: distributional names and their kinds. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 151–161.

Aurélie Herbelot and Marco Baroni. 2017. High-risk learning: acquiring new word vectors from tiny data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 304–309.

Nikolaus Kriegeskorte, Marieke Mur, and Peter A Bandettini. 2008. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2:4.

Saul A Kripke. 1972. Naming and necessity. In *Semantics of natural language*, pages 253–355. Springer.

Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. 2015. Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338.

Isabel Pavão Martins and Luisa Farrajota. 2007. Proper and common names: A double dissociation. *Neuropsychologia*, 45(8):1744–1756.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237.

Carlo Semenza. 2009. The neuropsychology of proper names. *Mind & Language*, 24(4):347–369.