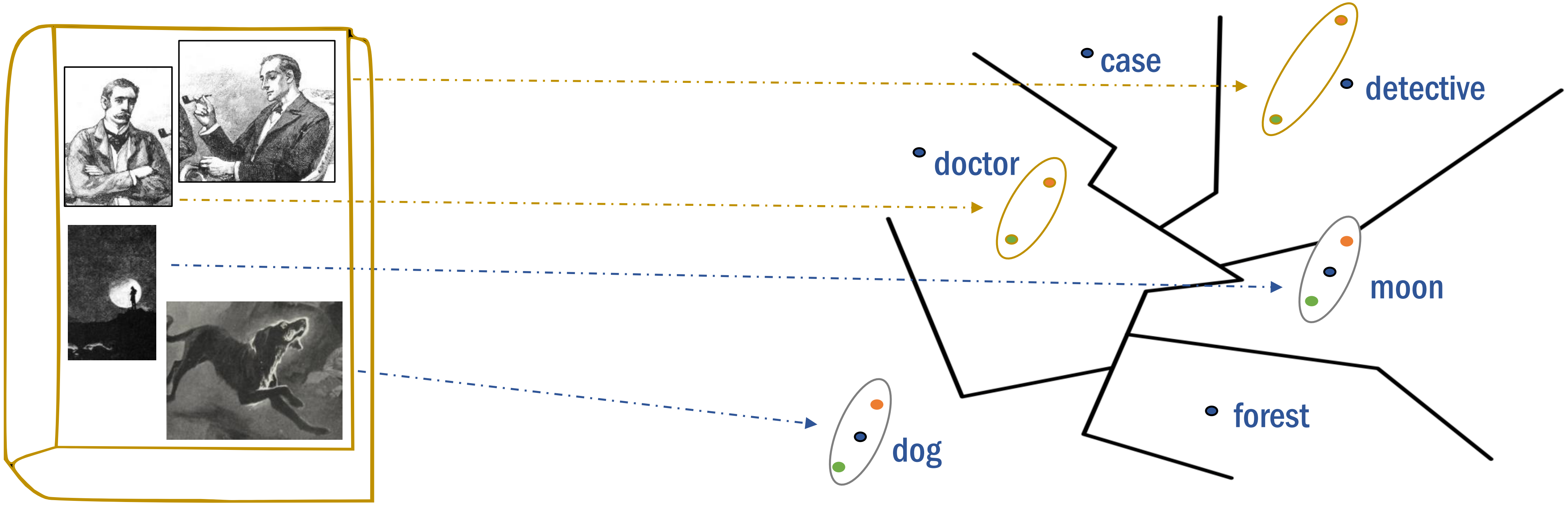
Novel Aficionados and Doppelgängers: learning and evaluating distributional representations of individual entities

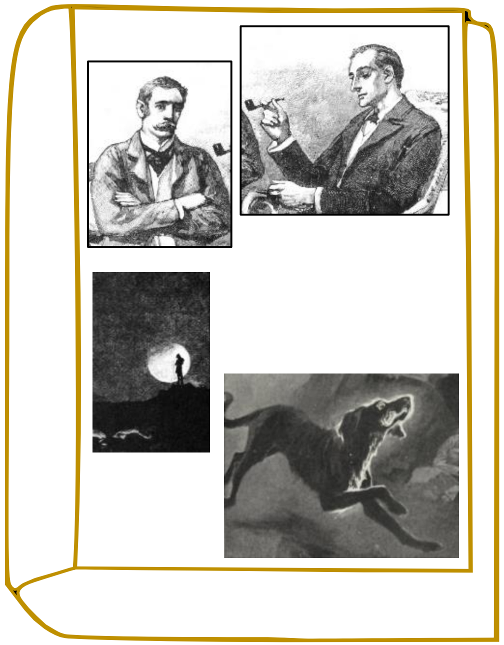
Andrea Bruera – *Queen Mary University of London, School of Electronic Engineering and Computer Science*
Aurélie Herbelot – *University of Trento, Center for Mind and Brain Sciences*



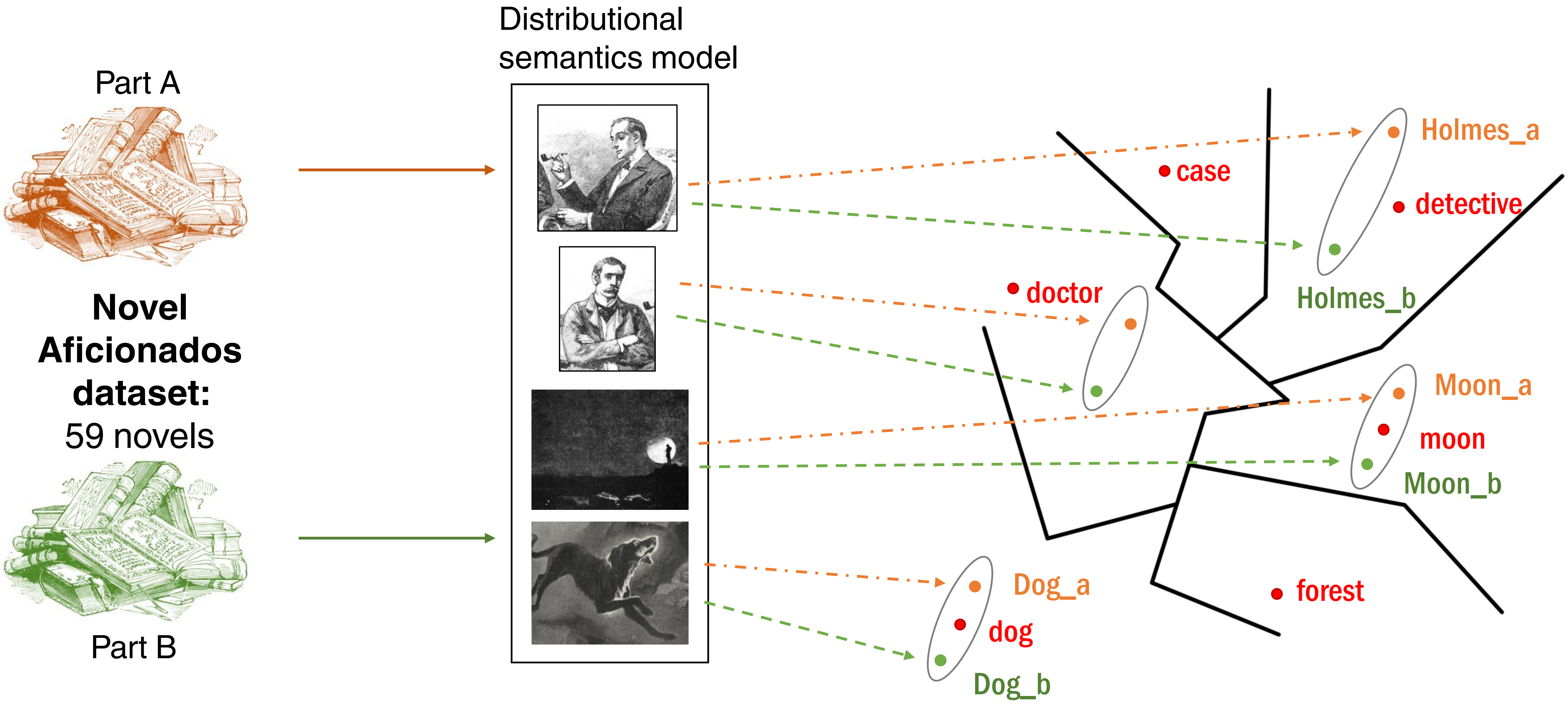
get in touch: a.bruera@qmul.ac.uk

Our approach to individual entities

- We define **individual entities** as entities to which are assigned **proper names** [1];
- We tackle the issue of their computational representation, by following three principles:
  - Contrarily to the mainstream approach (e.g. entity linking [2], knowledge-aware contextualized models [3]), we don't focus on the representation of frequent individual entities by using big corpora. Instead, we focus on a small-sized, cognitively interesting source of information about individual entities: **novels**;
  - We want to model the process by which **a reader forms a novel representation of previously unknown characters in a book**;
  - We take an unsupervised approach inspired by neuropsychological research on individual entities [4], where **individual entities (proper names) are usually contrasted to non-individual entities (common nouns).**



We propose the Doppelgänger test, a new referential task, as a way to compare the representations of proper names and common nouns in distributional spaces



Unsupervised evaluation for the Doppelgänger test

```
Result for the vector: mr_sherlock_holmes, coming from part
a of the book
Rank: 5 out of 7 characters
Cosine similarity to the co-referring vector: 0.855180033732
6522

1) john_ferrier_b - 0.9448971185149496

2) mr_drebber_b - 0.8777839614281693

3) mr_jefferson_hope_b - 0.8678383309937534

4) mr_joseph_stangerson_b - 0.8568128518089233

5) mr_sherlock_holmes_b - 0.8551800337326522

6) tobias_gregson_b - 0.8488438995759682

7) mr_lestrade_b - 0.7936464075060854
```

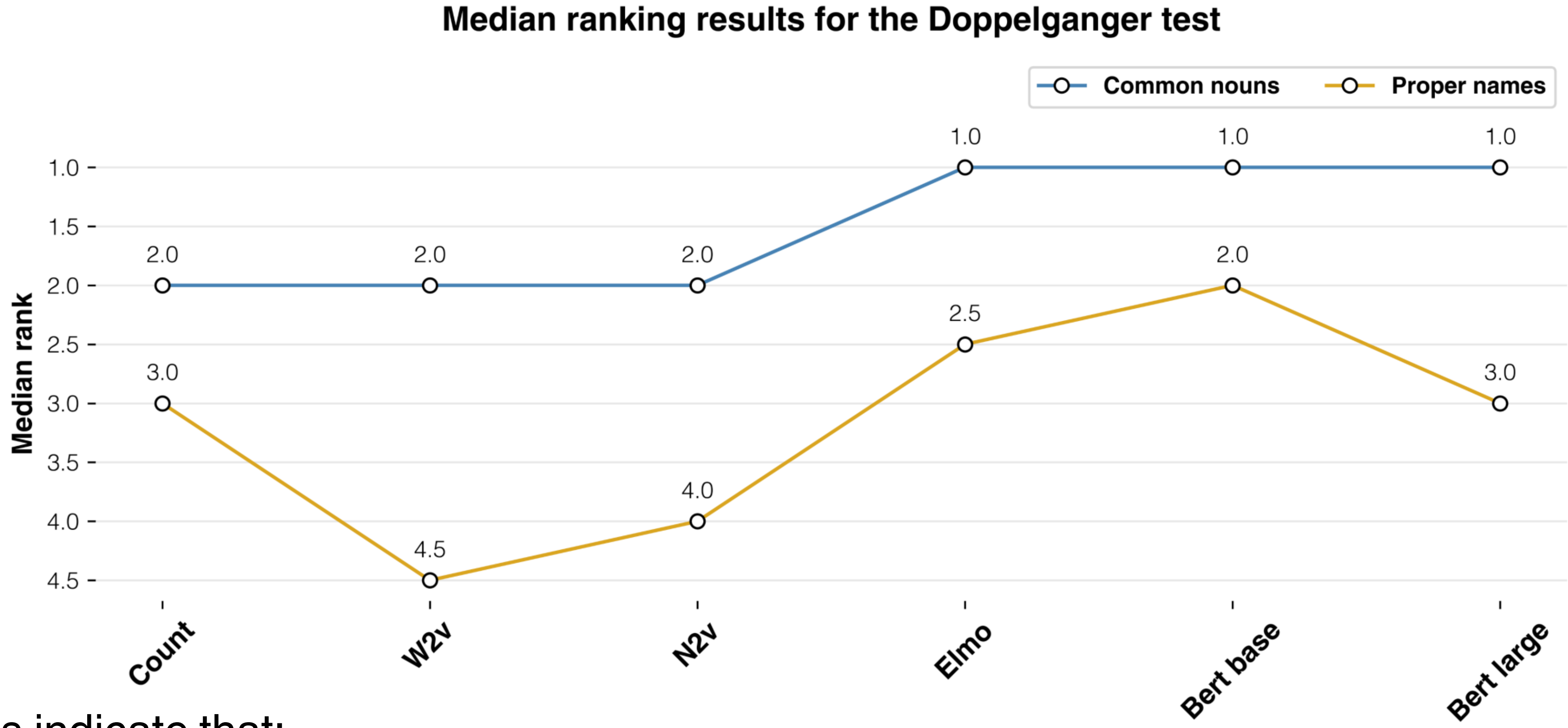
Individual entities

```
Result for the vector: hand, coming from part a of the book
Rank: 1 out of 8 characters
Cosine similarity to the co-referring vector: 0.942315760712
8779

1) hand_b - 0.9423157607128779

2) heart_b - 0.8774641156196594

3) life_b - 0.8587621962286366

4) word_b - 0.851731538772583

5) house_b - 0.7991881966590881

6) thing_b - 0.7865883708000183

7) night_b - 0.7779817084653744

8) matter_b - 0.7638704776763916
```

Non-individual entities

- We create two semantic spaces by employing two distinct halves of the novels;
- We **bootstrap** [5] from pre-trained models **new word vectors for the characters and the most common nouns**;
- We measure to what extent it is possible to **match each pair of vectors** referring to the same character/entity from the two spaces, by using a simple ranking test.

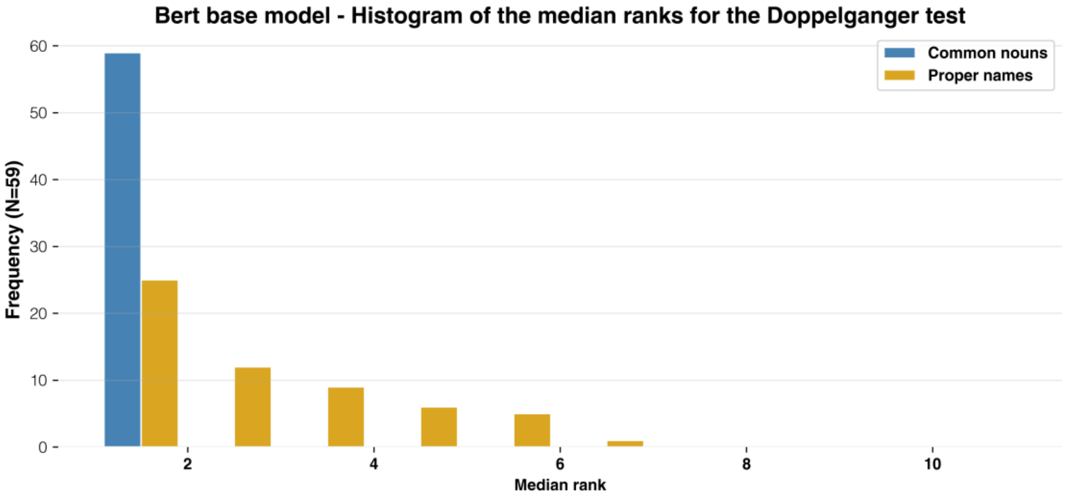
Individual entities and non-individual entities show different performance patterns #1

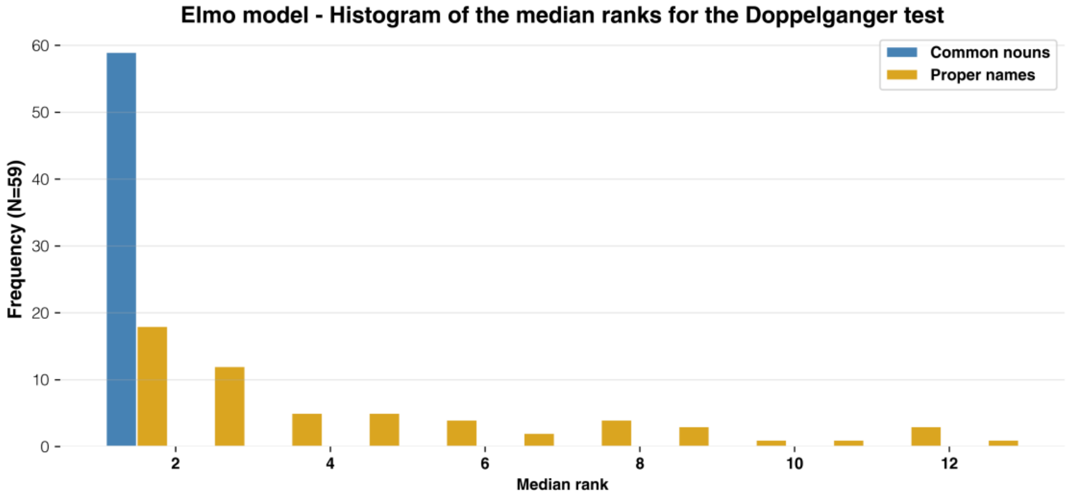


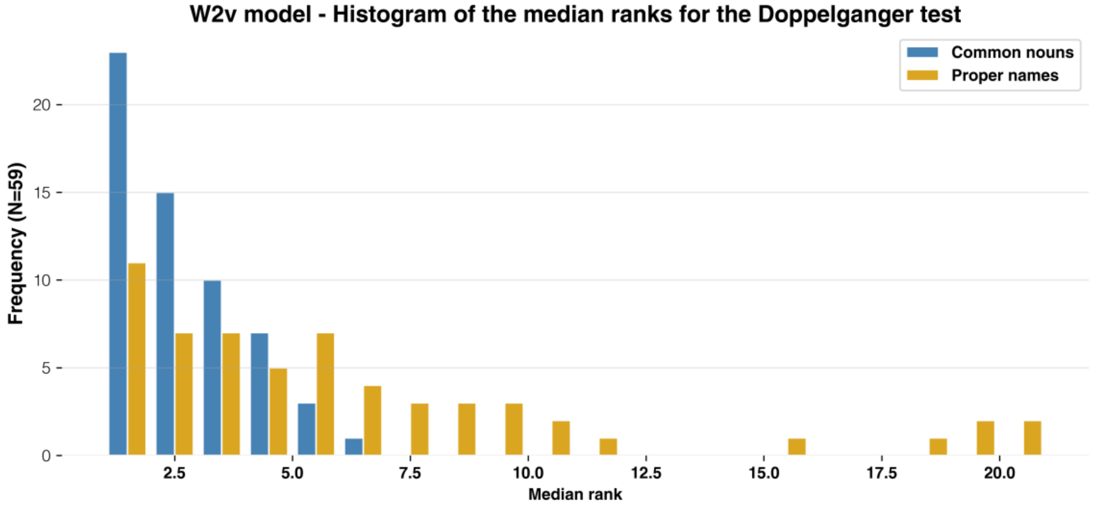
The results indicate that:

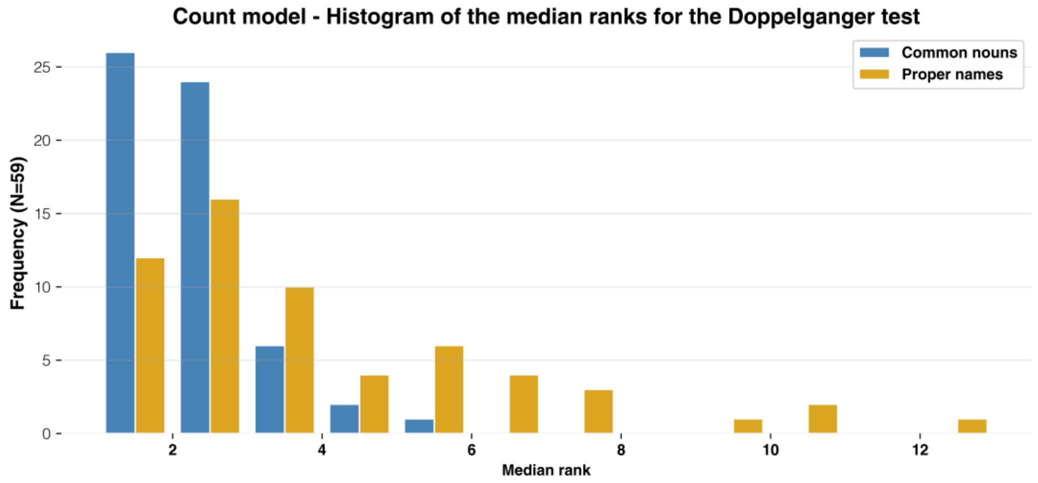
- the difference among the two categories of proper names and common nouns can be retrieved by their distributional properties;
- the representations of common nouns are much easier to match according to their reference than those of proper names - just as in human cognition.

Individual entities and non-individual entities show different performance patterns #2





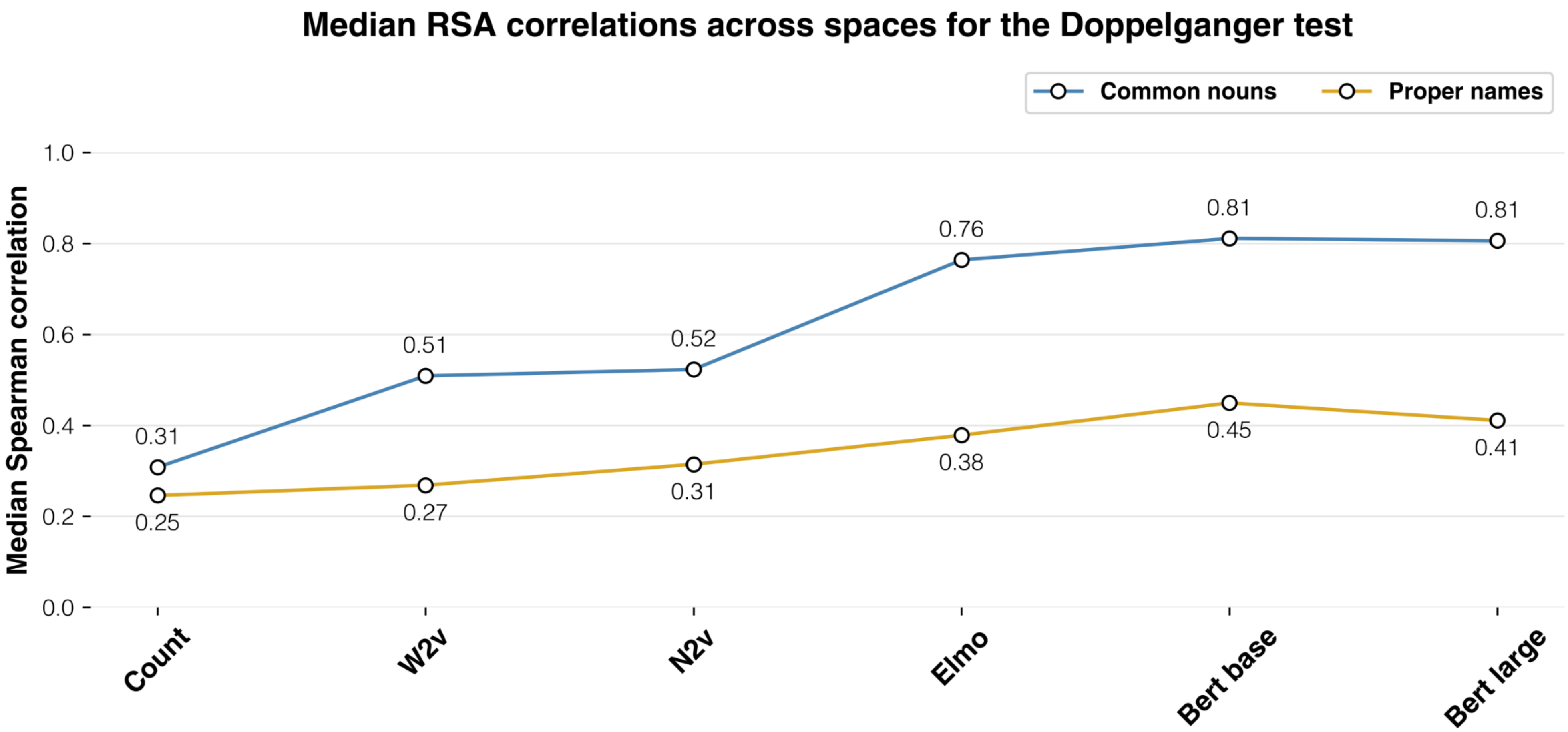




The histograms of the results again show clear differences for the two categories, with better results for common nouns.

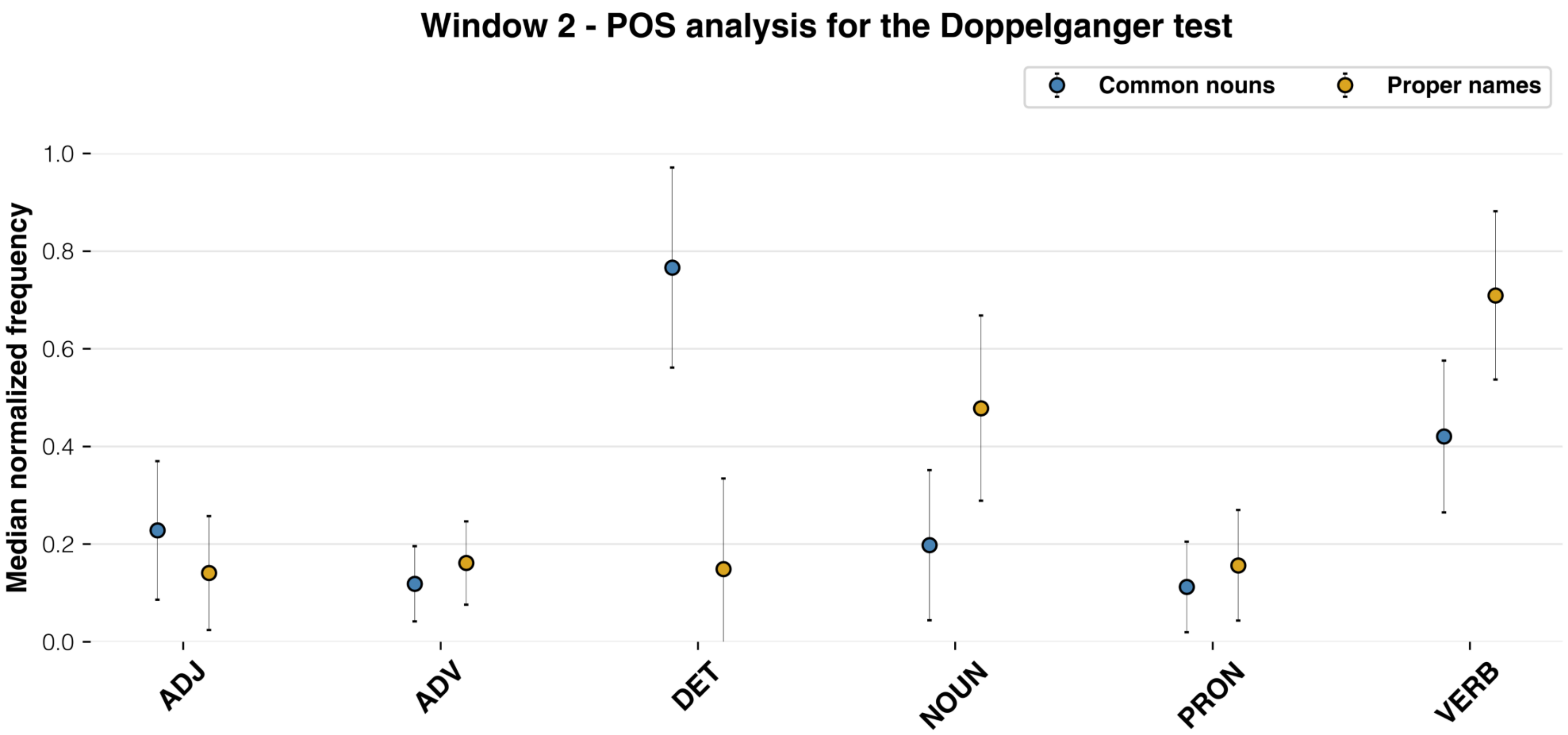
Why is this happening?

1. Representational similarity analysis of the two spaces reveals different stability in representational properties



The spaces made of the common nouns correlate among each other much more than those of the proper names. This hints at a reason why common nouns perform better than proper names at the Doppelgänger test: **their representational geometry is more stable across spaces** [6].

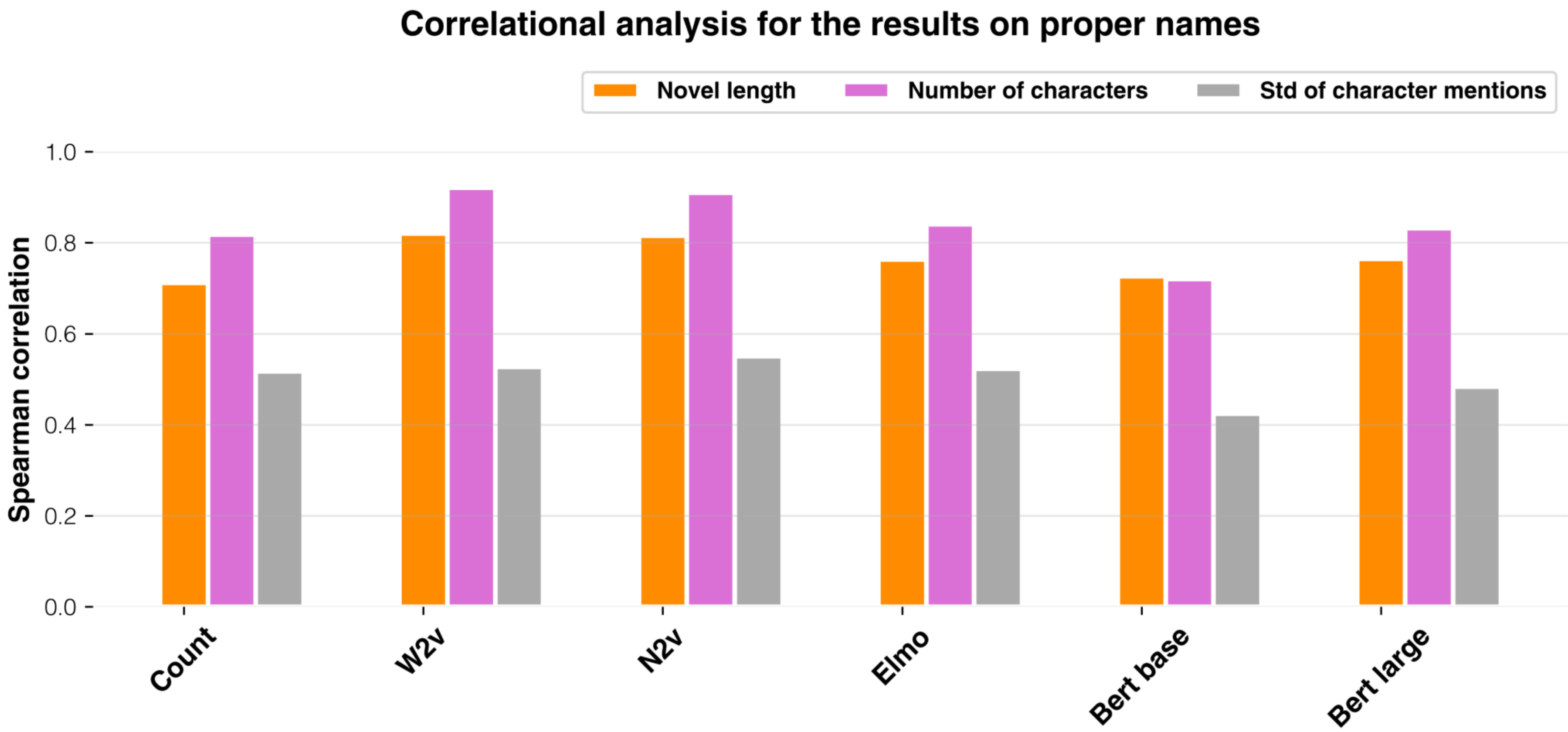
Why is this happening?
2. POS distributions are different



A POS distributional analysis of the two different categories reveals, once more, different distributional properties. This can arguably affect the representations of the two categories.

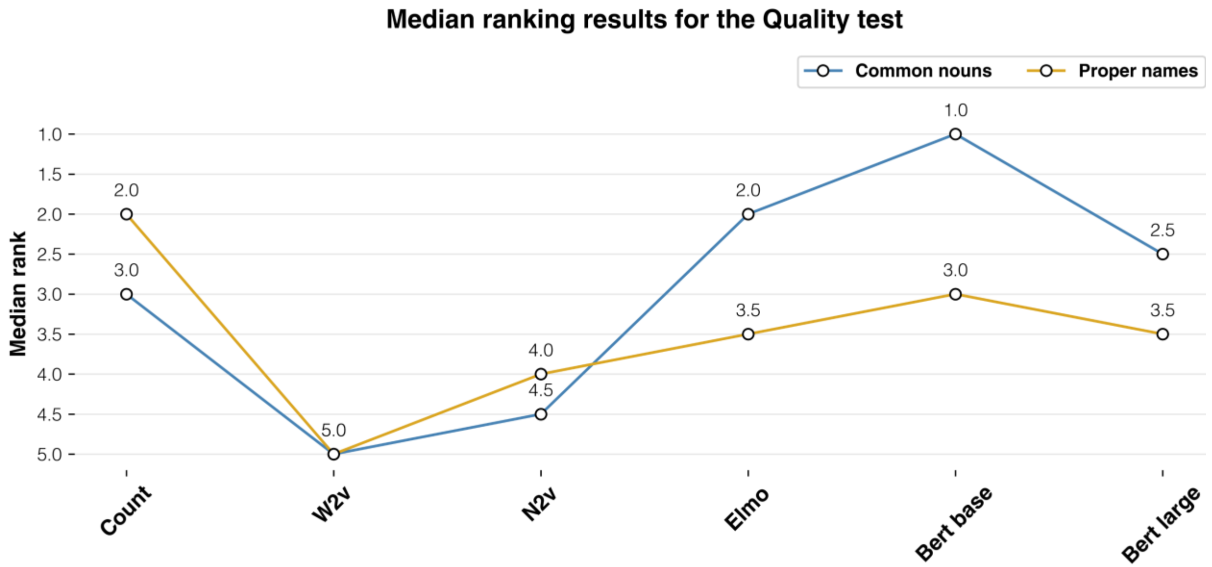
Why is this happening?

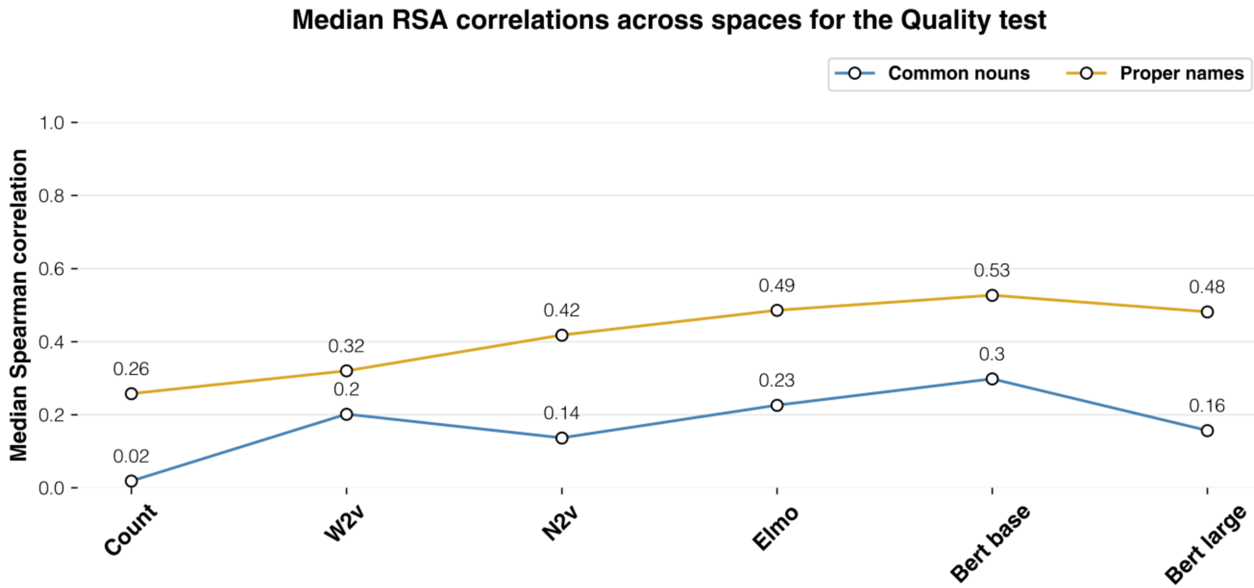
3. A correlational analysis shows that proper names get confused with one another



The variable which correlates the most with results is the number of characters per novel: **characters are susceptible to be confused with one another** (as is the case in human cognition, for instance with the tip-of-the-tongue phenomenon [7]).

A twist on the Doppelgänger test: the Quality test





- The Doppelgänger test setup can be adapted to other tasks - for instance, a **cross-document referential task**: what if we tried to match vectors obtained from the novels and vectors obtained from Wikipedia?
- We implement this research question in the **Quality test;**
- This task seems to be more difficult for both proper names and common nouns. Also, their representational geometry is consistenly less stable across spaces;
- In other words, there's loads to investigate... if you're curious, take a look to the **dataset** and take a **different approach** to the Doppelgänger test!

Summary

- We focus on **representations on individual entities in NLP** by taking inspiration from cognitive science, where they are usually contrasted to **non-individual entities**;
- We hypothesize that that the difference between **individual entities** (proper names) and **non-individual entities** (common nouns) **can be retrieved in distributional representations** when tested with an appropriate referential task;
- We present a **novel referential task, the Doppelgänger test**, to evaluate and contrast distributional representations of proper names and common nouns;
- We run the tests on an **original dataset, the Novel Aficionados dataset**, using a broad range of distributional semantics and word embeddings models, both contextualized and not;
- We show that proper names and common nouns show **distinct patterns of results at both tests**, with representations of proper names performing worse;
- We further investigate our results by means of **four analyses**: a variant of the Doppelgänger test, Representational Similarity Analysis, a part-of-speech distributional analysis and a correlational analysis;
- All analyses corroborate our results and give insights with respect to the reasons behind them, also allowing to **draw a parallel with human cognition.**

Thank you!

feel free to get in touch: a.bruera@qmul.ac.uk

*...and to **use the dataset**! it's freely available at*

https://github.com/andreabruera/novel_aficionados_dataset

References

1. Jeshion, R. (2009). The significance of names. *Mind & Language*, *24*(4), 370-403.
2. Ilievski, F., Vossen, P., & Van Erp, M. (2017, June). Hunger for Contextual Knowledge and a Road Map to Intelligent Entity Linking. In *International Conference on Language, Data and Knowledge* (pp. 143-149). Springer, Cham.
3. Peters, M. E., Neumann, M., Logan, R., Schwartz, R., Joshi, V., Singh, S., & Smith, N. A. (2019, November). Knowledge Enhanced Contextual Word Representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 43-54).
4. Brédart, S. (2016). Names and cognitive psychology. In *The Oxford Handbook of Names and Naming*.
5. Herbelot, A., & Baroni, M. (2017, September). High-risk learning: acquiring new word vectors from tiny data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 304-309).
6. Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in cognitive sciences*, *17*(8), 401-412.
7. Abrams, L., & Davis, D. K. (2017). Competitors or teammates: how proper names influence each other. *Current Directions in Psychological Science*, *26*(1), 87-93.

Models

- **BERT**: Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- **ELMO**: Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. arXiv preprint arXiv:1802.05365.
- **Word2Vec** (W2V): Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
- **Nonce2Vec** (N2V): Herbelot, A., & Baroni, M. (2017, September). High-risk learning: acquiring new word vectors from tiny data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 304-309).
- **Count -** no real reference for that, but a good introduction can be found in: Erk, K. (2012). Vector space models of word meaning and phrase meaning: A survey. Language and Linguistics Compass, 6(10), 635-653.