Nomen est omen: fictional characters’ names encode polarity, gender and age

Fabiënne Reedijk, Stefano Scola, Niccolò Minetti, Niveditha Subramaniam and Giovanni Cassani (Tilburg University)
g.cassani@tilburguniversity.edu

We report preliminary results from an ongoing investigation of whether authors choose the names of fictional characters to indicate their gender, age, and polarity, following up on research about the sound-symbolic connotations of real first names [1,2] and fictional characters [3,4]. In order to explore the importance of shared language experience in driving the choice and interpretation of character names, we consider *real first names* (e.g., Julia), names which are words with existing meanings (*aptronyms,* e.g., Spark), and *made-up names* (e.g., Arobynn). 63 character names (19 *real first names*, 24 *made-up names*, 20 *aptronyms*) are extracted from a crawled corpus of fantasy fan fiction stories consisting of approximately 17M tokens. Another 120 names (40 per category) come from a curated corpus of children and young adult literature, consisting of 2.6M tokens. Names were selected to ensure they were sufficiently frequent in the story they featured in to learn reliable semantic representations from textual co-occurrences.

We leverage recent models from computational psycholinguistics that extract distributed meaning representations from word form alone by exploiting statistical regularities in how word forms relate to lexical meanings [5,6,7]. We first featurize English words into letter n-grams and features inspired by studies on sound-symbolism. We then derive co-occurrence-based word embeddings from our corpora [8] for general words as well as character names, ensuring embeddings from different stories are aligned and exist in the same embedding space for comparability by using Compass-Aligned Distributional Embeddings (CADE, [9]). Finally, we learn a mapping function between n-gram and semantic representations considering the general vocabulary, and use it to generate form-based semantic embeddings for character names that can be directly compared to the co-occurrence-based embeddings.

We evaluate to what extent names reflect semantic properties by feeding the vectors for character names in the different feature spaces (letter n-grams, phonological features, form-based and co-occurrence-based embeddings) into a Linear Discriminant Analysis (LDA) classifier, to probe how discriminable names are in each feature space.

Classification experiments show that names tend to be discriminable in form and semantic feature spaces, with accuracies between 0.59 and 0.7 and Wilk’s lambda between 0.04 and 0.33, depending on the input features (ngrams, phonological features, word embeddings) and the target attribute (age, polarity, gender). Interestingly, names for ambivalent characters (as coded following cognitive literary theory) tend to be less discriminable in form space than plain evil or good ones, suggesting that sound symbolic devices may be used in subtle ways to convey expectations. Form-based word embeddings, on the contrary, tend to be less discriminable, suggesting that the statistical relations between form and meaning in the general vocabulary are not reliable to infer the semantic connotations of the names we considered. Further inspection revealed that form-based semantic vectors tend to be poorly discriminative in general, clustering around the centroid of the embedding space, in contrast with evidence from [5]. This pattern likely originates in the semantic representations used: whereas [5] used sparse word embeddings learned on the TASA corpus (10M tokens) using Naïve Discriminative Learning, we relied on dense representations obtained using CADE to ensure the alignment across stories. We plan on experimenting with different semantic representations in the future.

In conclusion, our work highlights that, beyond describing characters sharing an attribute in similar ways across different stories (as captured using aligned word-embeddings [9]), authors name them in ways which already convey attributes such as gender, polarity and age [2,3,4,5]. We are now collecting behavioral intuitions about character names to analyze whether readers are sensitive to the patterns in names and whether association patterns in form and meaning predict human intuitions about characters’ attributes based on their names only, and do so differently depending on whether a name is routinely used to name people, is made-up or leverages a word with established meaning.

Technical details of the computational model

In order to derive word embeddings for fictional characters that could be compared to each other in spite of the fact that characters appear in different stories, we leveraged CADE, a model which has already been fruitfully used for narrative understanding [10]. CADE leverages the Skip-Gram with Negative Sampling (SGNS) model from *word2vec* [8], which uses two matrices to learn word embeddings, a target matrix and a context matrix. CADE exploits this aspect by first training a general embedding using the whole corpus, ignoring the different stories. This embedding space is the *compass*, i.e. a general representation to which the embedding spaces derived from each story are aligned. The context matrix of the compass is extracted and used to initialize (and freeze) the second matrix of a story-specific SGNS embedding space. This approach ensures that all story-specific embedding spaces share the same context matrix, making the story-specific embeddings aligned and directly comparable.

In detail, we trained two CADE models, one for the fan fiction corpus (window size=5, min count=5 in each story, dimensionality=300) and one for the children and young adult corpus (window size=5, min count=5 in each book, dimensionality=50). Hyper-parameters were selected based on a grid-search and an intrinsic evaluation carried out using the MEN dataset for semantic relatedness as a benchmark [11]. Parameter optimization is carried out with a learning rate of 0.025 and 10 negative samples. We use 5 iterations to train the compass embeddings and 5 iterations to train the slice specific embeddings. We initialize all the other hyper-parameters using the default settings provided by CADE.

In order to learn form-based semantic vectors, we leverage Linear Discriminative Learning (LDL, [5,6]) and an extension of Orthography-Semantic-Consistency (OSC, [7]). LDL learns a mapping function from form vectors to semantic vectors using multivariate multiple regression. The encoding of word form exploits letter tri-grams [5,6]. The mapping function is obtained by multiplying the semantic matrix, the compass matrix from CADE, by the pseudo-inverse of the form matrix and minimizes the reconstruction error when predicting semantic vectors from form vectors. This mapping function is finally applied to the form vectors of our character names and yields form-based semantic vectors which depend on names alone.

OSC, on the contrary, leverages local similarity in form and semantic space. So far, OSC has only been used to estimate (pseudo)words' semantic neighborhood density, but we extend it to generate semantic vectors based on word form alone. First, the 5 nearest neighbors of character names in form space are retrieved (in case of ties, all words at the same distance are considered). Then, the semantic embeddings of these neighbors are fetched from the semantic space and averaged (weighted by the inverse of the distance of the word to the target in form space) to yield a semantic vector for the target names which combines the semantics of similar words in form space.

As we mentioned in above, we use LDA to assess whether different feature representations capture semantic attributes. For semantic embeddings and n-gram vectors, we feed the raw vectors as input. For theory-driven phonological vectors, however, we manually coded names following previous work which highlighted sound-symbolic features (since we do not have an accepted pronunciation for *made-up names*, we limited our analysis to features that we could code based on the orthographic form). In detail, we used the following features when predicting polarity: sonorants, voiced stops, voiceless stops, /f/, /g/, /tʃ/, /s/, /d/ and /ɹ/. These phonological features were chosen because the literature suggests that they are representative of valence [12] and round and spiky shapes [13,14,15,16], which were used as a proxy for polarity (good characters tend to be portrayed using round shapes, in contrast to evil characters which are associated to angular shapes [17]). In order to predict age, we leveraged the following features: /n/, /t/, /g/, /k-/, voiceless fricatives, voiced fricatives, voiceless stops, voiced stops, and sonorant consonants. These features were based on sounds related to size iconicity [18,19,20] given that young characters tend to be smaller than old characters. Finally, when predicting gender, we relied on the same sound-size iconicity features as the prediction of age, because smallness is related to femininity [19]. Additionally, the amount of syllables was used because in English women's names tend to be longer [21].

References:

[1] Sidhu, D. M., & Pexman, P. M. (2015). What's in a Name? Sound Symbolism and Gender in First Names. *PLoS One, 10*(5), e0126809. doi:10.1371/journal.pone.0126809

[2] Sidhu, D. M., Deschamps, K., Bourdage, J. S., & Pexman, P. M. (2019). Does the name say it all? Investigating phoneme-personality sound symbolism in first names. *J of Exp Psych: Gen, 148*(9), 1595-1614. doi:10.1037/xge0000662

[3] Elsen, H. (2017). The Two Meanings of Sound Symbolism. *Open Ling, 3*(1). doi:10.1515/opli-2017-0024

[4] Smith, R. (2006). Fitting Sense to Sound: Linguistic Aesthetics and Phonosemantics in the Work of J.R.R. Tolkien. *Tolkien Studies, 3*(1), 1-20. doi:10.1353/tks.2006.0032

[5] Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The Discriminative Lexicon: A Unified Computational Model for the Lexicon and Lexical Processing in Comprehension and Production Grounded Not in (De)Composition but in Linear Discriminative Learning. *Complexity, 2019*, 1-39. doi:10.1155/2019/4895891

[6] Cassani, G., Chuang, Y. Y., & Baayen, R. H. (2020). On the Semantics of Nonwords and Their Lexical Category. *J of Exp Psych: Learn, Mem, and Cogn, 46*(4), 621-637. doi:10.1037/xlm0000747

[7] Marelli, M., Amenta, S., & Crepaldi, D. (2015). Semantic transparency in free stems: The effect of Orthography-Semantics Consistency on word recognition. *QJEP, 68*(8), 1571-1583. doi:10.1080/17470218.2014.959709

[8] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th NeurIPS International Conference - Volume 2*, Lake Tahoe, Nevada.

[9] Bianchi, F., Carlo, V., Nicoli, P., & Palmonari, M. (2020). Compass-aligned Distributional Embeddings for Studying Semantic Differences across Corpora. *ArXiv, abs/2004.06519*

[10] Volpetti, C., Vani, K., & Antonucci, A. (2020). Temporal word embeddings for narrative understanding. In *Proc of the 2020 12th Intern Conf on Mach Learn and Comp* (pp. 68-72).

[11] Bruni, E., Tran, N. K., & Baroni, M. (2014). Multimodal distributional semantics. *JAIR*, *49*, 1-47.

[12] McCormick, K., Kim, J., List, S., & Nygaard, L. C. (2015). Sound to Meaning Mappings in the Bouba-Kiki Effect. In *CogSci* (Vol. 2015, pp. 1565-1570).

[13] Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proc Nat Acad Sci*, *113*(39), 10818-10823.

[14] Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *J of Exp Psych: Learn, Mem, and Cogn*, *38*(5), 1152.

[15] Sidhu, D. M., Westbury, C., Hollis, G., & Pexman, P. M. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psych Bull & Rev*, 1-9.

[16] Adelman, J. S., Estes, Z., & Cossu, M. (2018). Emotional sound symbolism: Languages rapidly signal valence via phonemes. *Cognition*, *175*, 122-130.

[17] Solarski, C. (2013). “Gamasutra - The Aesthetics of Game Art and Game Design”. Gamasutra.com. Available at this link.

[18] Taylor, I. K., & Taylor, M. M. (1965). Another look at phonetic symbolism. *Psych Bull*, *64*(6), 413.

[19] Klink, R. R. (2000). Creating brand names with meaning: The use of sound symbolism. *Marketing Letters*, *11*(1), 5-20.

[20] Shih, S. S., Ackerman, J., Hermalin, N., Inkelas, S., & Kavitskaya, D. (2018). Pokémonikers: A study of sound symbolism and Pokémon names. *Proc Ling Soc of America*, *3*(1), 42-1.

[21] Cutler, A., McQueen, J., & Robinson, K. (1990). Elizabeth and John: Sound patterns of men's and women's names. *J of Ling*, *26*(2), 471-482.