Temporal word embeddings in the study of metaphor change over time and across genres: a proof-of-concept study on English

Veronica Mangiaterra[1,*], Chiara Barattieri di San Pietro[1] and Valentina Bambini[1]

[1]Laboratory of Neurolinguistics and Experimental Pragmatics (NEPLab), Department of Humanities and Life Sciences, University School for Advanced Studies IUSS, Pavia, Italy

Abstract

Temporal word embeddings have been successfully employed in semantic change research to identify and trace shifts in the meaning of words. In a previous work, we developed an approach to study the diachrony of complex expressions, namely literary metaphors. Capitalizing on the evidence that measures of semantic similarity between the two terms of a metaphor approximate human judgments of the difficulty of the expression, we used time-locked measures of similarity to reconstruct the evolution of processing costs of literary metaphors over the past two centuries. In this work, we extend this approach previously used on Italian literary metaphors and we present a proof-of-concept study testing its crosslinguistic applicability on a set of 19th-century English literary metaphors. Our results show that the processing costs of metaphors changed as a function of textual genre but not of epoch: cosine similarity between the two terms of literary metaphors is higher in literary compared to nonliterary texts, and this difference is stable across epochs. Furthermore, we show that, depending on the metaphor structure, the difference between genres is affected by word-level variables, such as the frequency of the metaphor's vehicle and the stability of the meaning of both topic and vehicle. In a broader perspective, general considerations can be drawn about the history of literary and nonliterary English language and the semantic change of words.

Keywords

metaphors, distributional semantics, temporal embeddings

1. Introduction

Does the metaphor "The wind is a wrestler" convey the same feeling today, as it did in the 1888 when Gerard Manley Hopkins used it in the poem "That nature is a Heraclitean Fire and of the comfort of the Resurrection" [1]? The answer to this question is not trivial: human languages evolve constantly, alongside with the society in which they are used, so much so that the concepts associated with each word, as well as their semantic associations with other words, have changed to different degrees [2].

Studies on lexical semantic change have a long tradition [3, 4] but, with the increasing availability of historical language data and the development of new digital tools, they radically opened up to new approaches coming from computational linguistics and distributional semantics [5, 6, 7]. In the diachronic declination of the Distributional Hypothesis [8], it is said that changes in the contexts in which a word occurs over time may reveal a change in meaning [9]. Operatively, this means that by training vector space models on historical text corpora from different epochs, it is possible to create time-locked representations of words: if the meaning of a word changed over time, its vectorial representation at t_1 will be different from its vectorial representation at time t_2; conversely, if the two vectors of the same word at t_1 and t_2 are in close proximity, the meaning of the word has remained stable. Comparing words vectors diachronically, however, is not effortless and requires the temporal vector space models to be aligned. Alignment is a crucial step in diachronic distributional semantics and it has been tackled by different approaches [10, 11, 12]. Previous studies employing temporal embeddings have found that more frequent words change slower than less frequent words, and that polysemous words change faster than monosemous words [2], while synonyms tend to change meaning comparably [13]. However, temporal word embeddings have been mostly applied to the study of the semantic change of single words and only marginally to complex linguistic expressions leaving the field with a knowledge gap on the evolution of meaning of a widespread linguistic and textual phenomenon such as, for instance, metaphors.

Within the theoretical framework of Relevance Theory [14], metaphors are non-literal uses of language involving a conceptual adjustment described as context-driven broadening of lexically denoted meaning of words. In

CLiC-it 2024: Tenth Italian Conference on Computational Linguistics, Dec 04 — 06, 2024, Pisa, Italy

*Corresponding author.
veronica.mangiaterra@iusspavia.it (V. Mangiaterra); chiara.barattieridisanpietro@iusspavia.it (C. B. d. S. Pietro); valentina.bambini@iusspavia.it (V. Bambini)
0009-0001-7852-5259 (V. Mangiaterra); 0000-0003-4407-7037 (C. B. d. S. Pietro); 0000-0001-5770-228X (V. Bambini)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

terms of linguistic structure, metaphors normally involve two terms, the topic and the vehicle: for example, in the metaphor 'Sally is a chameleon', the topic *Sally* is described by the broadened vehicle *chameleon*, to indicate a person who changes attitude/behavior to fit their surroundings. While metaphors are broadly used in everyday communication, they are certainly a distinctive feature of literary texts, as long evidenced in stylistics [15]. Past studies on literary metaphors, however, report mixed results. The rating study by Katz et al. [16] found no difference between literary and everyday metaphors, while other studies showed that the former type is less familiar and more open-ended than the latter [17], but literary metaphors are rated as less difficult and more familiar when presented together with their original context [18]. Moreover, the processing of literay metaphors seems to be particularly effortful, given the multitude of possible meanings they evoke [19]. Therefore, open questions remain regarding how literary metaphors are processed. It must be also underlined that the literary metaphors used in previous studies were written tens or hundreds of years ago. Yet, the effect of this diachronic dimension on their processing costs, as well as its interplay with textual genre in which metaphors are embedded, remains an open question.

In addition to its diachronic application, the use of vector space models can help characterize metaphors thanks to the ability of these models to approximate human performance in psycholinguistic tasks. Measures derived from vector space models were shown to be able to approximate how humans process word meaning [20, 21, 22] and, more specifically to correlate with how humans perceive metaphorical expressions in terms of metaphoricity, difficulty, and other psycholinguistic dimensions [23, 24, 25]. In particular, semantic similarity, operationalized in vector space models as cosine similarity (CS) between topic and vehicle, has long been considered relevant for metaphor studies [26] and, more recently, for automatic metaphor identification [27].

In a previous study on Italian [28], we developed a novel method, employing the *Temporal Word Embeddings with a Compass* (TWEC) model [10] as training procedure, to capture the temporal dynamics of literary metaphors. This method combines the computational models' abilities to approximate human judgments and their diachronic applications, allowing to track the diachronic evolution of how literary metaphors are perceived by readers over the course of 200 years. In the present proof-of-concept study, we apply this approach to English, to test its crosslinguistic applicability and whether it can provide language-specific insights into the evolution of metaphors. We take the similarity between the topic and vehicle of a metaphor as a proxy for its difficulty and we analyze how it varies across time and textual genres. We also consider the role of word frequency (WF) and vector coherence (VC), two widely used measures in the study of semantic change [29, 30], as well as semantic neighborhood density (SND) in shaping the difficulty of the expression. WF and VC were considered to assess the effect of the semantic change of the single word on the evolution of whole metaphor understanding, while SND was considered to analyze the impact of a measure known to synchronically impacts metaphor understanding [31, 24] on its diachronic unfolding.

2. Methods

2.1. Dataset of metaphors

The study focuses on "classic" literary metaphors (i.e., metaphors found in 19th-century literary texts). In terms of metaphor structure, we focused on metaphors in the form of 'A is B' (e.g. "Stars are dancers") and 'A of B' (e.g., "Clouds of melancholy"), as they clearly display the two metaphorical elements (topic and vehicle) and allow to avoid possible confounding factors (length of expression, intervening words, etc.). Twenty- four (24) 'A is B' metaphors were taken from the dataset in Katz et al. [16] and 115 metaphors in the form 'A of B' were retrieved from a collection of literary texts of the 19th century. These latter were identified by PoS-tagging a corpus of literary texts from the 19th century (see below) with spaCy [32], and then extracting only the 'NOUN of NOUN' constructions. The resulting list was then further reduced by manually searching for words belonging to known sources of metaphors, such as atmospheric events (e.g., 'rain') or physical locations (e.g., 'river') [33], following the methodology in Bambini et al. (2014) [18].

2.2. Corpora and training

To test whether the processing costs of metaphors changed as a function of epoch, we collected corpora from the 19th century and from the 21st century. We also included different textual genres (literary vs. nonliterary) of the corpora, to examine whether the difficulty of the figurative expression is modulated by the stylistic features of different types of language. Following previous work [34], the corpora were built so as to be representative of the language to which speakers of the two epochs were exposed, and specifically by combining literary, nonfiction, and journalistic language for the 19th century, and literary and web language (which includes sections of newspapers, blogs, and other text types that can be found on the Internet) for the 21st century. Specifically, we trained four diachronic vector space models on four corpora:

- 19th-century literary corpus (32M tokens), consisting of a collection of literary texts (both narratives and poetry) retrieved from the Gutenberg project (gutenberg.org);
- a 19th-century nonliterary corpus (25M tokens), consisting of nonliterary texts, such as magazines or scientific essays, from the same online resource (gutenberg.org)
- a 21st-century literary corpus (16M tokens), collected from literary texts available on the web, employed without violating the "fair use" principle of copyright law;
- a 21st-century nonliterary corpus (46M tokens), collected from portions of the UMBC web- Base corpus [35].

To train aligned temporal vector space models, we followed the procedure by Di Carlo et al. [10]. The TWEC model is implemented on top of a Continuous Bag of Words (CBOW) architecture [36]. The TWEC model exploits the double representation learned by the CBOW model: the target matrix and the context matrix. First, a model, the so-called "compass", is trained on the whole corpus, creating time-independent word embeddings. The context matrix of the compass is then maintained fixed to train on each corpus a time- and genre-specific target matrix from which we derive the temporal word embeddings. The four sets of embeddings obtained for the four corpora will represent the meaning of words in each time slice for the two genres. To validate our models, following previous studies [2], we computed the synchronic (within time period) accuracy of each vector space model against the MEN dataset [37], which contains 3,000 pairs of words together with a semantic similarity score provided by humans. Finally, we tested whether our measure of metaphor difficulty (cosine similarity between topic and vehicle) correlated with the measure of difficulty in Katz et al. [16] dataset.

2.3. Measures of interest and analyses

For each metaphor, we collected four measures of interest, at the metaphor- and word-level.

- Cosine similarity (CS): the similarity between the two terms of the metaphor (topic and vehicle). It is computed as the cosine of the angle between the vectorial representations of the two words. CS is here considered as a proxy value of difficulty of the metaphors.
- Semantic neighborhood density (SND): a measure of the density of the semantic space around a word. Words with many closely related words have a higher semantic density while words whose neighbors are more distant and are sparsely distributed have a lower density. It is computed as the mean cosine similarity between the target word and its 500 closest neighbors (standard size from previous work, see [38]).
- Vector coherence (VC): a measure of the stability of a word's meaning, computed as the cosine similarity between the target word at t_1 the target word at t_2. Words with a high vector coherence are considered to have stable meaning through time, while a low vector coherence means that the word's meaning has changed.
- Word frequency (WF): computed as the logarithm of the frequency of the target word in the reference corpus.

Each measure was collected for all the temporal slices, extracted from the temporal vector space models (CS, SND, and VC) or corpora (WF). To analyze how the understanding of metaphors changed over time and if it was affected by genre and word-level variables, we fitted a set of Linear Mixed Models (LMMs) using the R package *lme4* [39]. The two metaphorical structures were treated separately, fitting distinct models for 'A is B' and 'A of B' metaphors.

The linear mixed model considers CS as dependent variable and the interaction between epoch and genre and word-level variables as predictors. In all models Items (metaphors) were added as random variables. The resulting formula was:

*lmer(cosine ∼ epoch * genre * (VC-topic + VC-vehicle + SND-topic + SND-vehicle + WF-topic + WF-vehicle) + (1|Item).*

Alpha level was set at .05.

3. Results

First, to test the validity of the meaning representation in the vector space models, we correlated the human scores of relatedness and the semantic similarity derived from our word embedding for each pair of words in the MEN dataset [37] (Table 1). These results show strong correlations, comparable to the results obtained by Hamilton et al. (2016) [2], indicating that the models accurately mimic humans' representation of meaning (i.e., they have a good synchronic accuracy).

19th Literary	19th Nonliterary	21st Literary	21st Nonliterary
.55	.58	.61	.59

Table 1
Results of correlation between models' semantic similarity scores and MEN dataset's semantic similarity scores. All the correlation have a $p < .001$.

Secondly, we tested whether cosine similarity between the two terms of a metaphor correlated with the measure

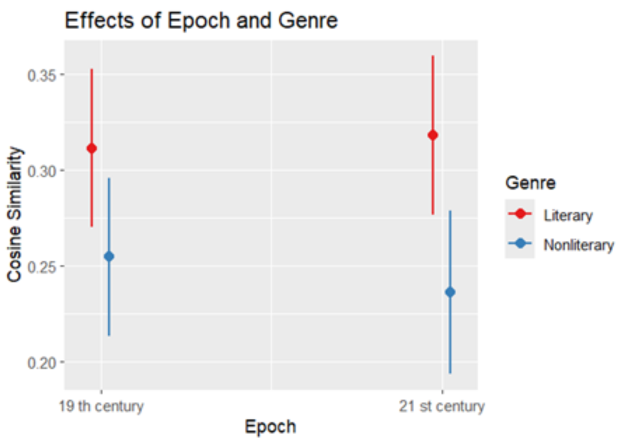

Figure 1: Effects of epoch and genre in defining the cosine similarity between the topic and vehicle of ‘A of B’ metaphors

of difficulty from the dataset by Katz et al. [16]. Results showed a moderate correlation ($r(26)$ = .49, p < .05): metaphors with higher semantic similarity between topic and vehicle were rated with lower values of difficulty by participants, coherently with previous studies.

Thirdly, we explored whether the change in the semantic similarity between the topics and the vehicles of literary metaphors is driven by the interaction between the Epoch, Genre and single-word variables. The results of our predictors of interest are reported below.

Concerning the ‘A of B’ metaphors’ mixed model, results showed a main effect of genre (β = 0.81, t = 2.44, p = .01) and a significant three-way interaction between epoch, genre and vector coherence, both of the topic (β = 0.34, t = 2.018, p = .04) and of the vehicle (β = -1.715, t = -4.954, p < .001). These results indicate that the cosine similarity of literary metaphors’ terms did not change over time, but it changed as a function of textual genres, resulting in greater difficulty (lower cosine similarity) in nonliterary texts than in literary (Figure 1). As shown by the three-way interaction between Epoch and Genre and the single-word variables in Figure 2, the effect of VC acted differently in the two time points and in the two genres. VC of the vehicle did not affect CS in literary and non- literary texts in the past; conversely, more stable vehicles significantly lowered CS in present literary texts and in- creased CS in present nonliterary texts. A similar trend can be observed for VC of the topic, where its stability did not affect CS in the past, regardless of the literary genres. Conversely, stability of the topic contributed to significantly increase CS in present literary texts, but less so in nonliterary texts.

For ‘A is B’, the model revealed a significant three-way interaction between epoch, genre, and the frequency of

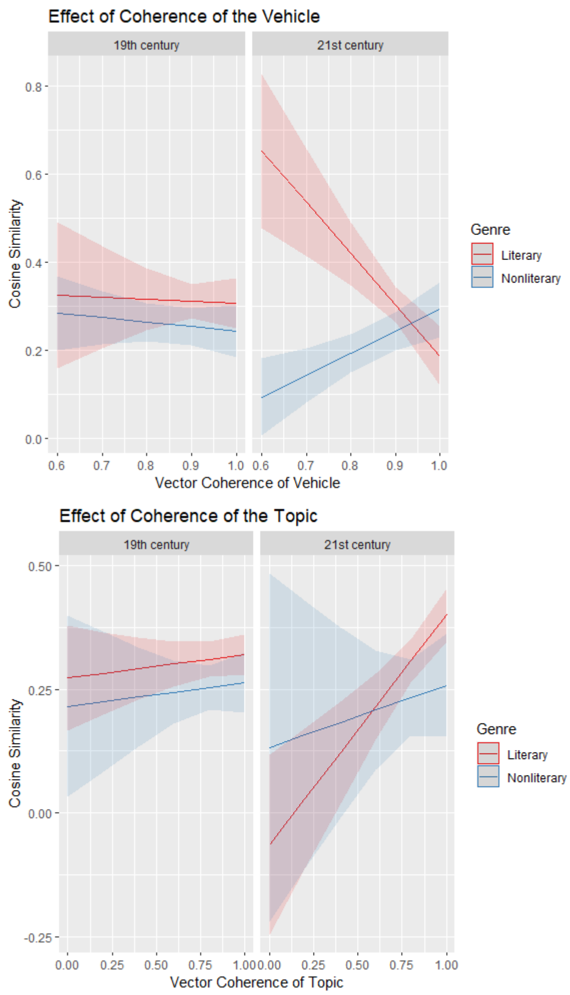

Figure 2: Effects of topic and vehicle VC in defining the cosine similarity between the topic and vehicle of ‘A of B’ metaphors

the vehicle (β = 0.06, t = 2.077, p = .04), but no main effects. The effect of WF of the vehicle showed different patterns in the two time points and in the two genres (Figure 3): while WF of the vehicle did not affect CS in literary texts both in the past and in the present, more frequent vehicles significantly increased CS in past nonliterary texts and lowered CS in present nonliterary texts.

4. Discussion

In this proof-of-concept study, we characterized the temporal dynamics of a set of English literary metaphors to understand whether their processing costs changed over time. We also explored if this change was affected by the genre of the texts, as well as by the semantic properties

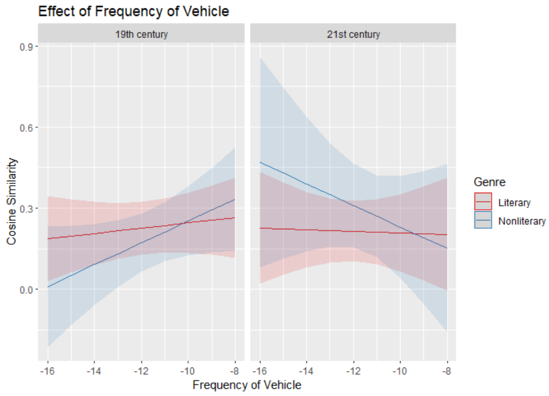

Figure 3: Effects of vehicle WF in defining the cosine similarity between the topic and vehicle of 'A is B' metaphors

of the constituting elements of the metaphors (topic and vehicle). By leveraging on the diachronic applications of distributional semantics and extending a method already applied to the study of Italian literary metaphors [28], we created a series of time-locked semantic representations of 139 English metaphors, from which we derived a measure of the cosine similarity between their terms (CS), taken as a proxy of their difficulty, together with semantic neighborhood density (SND), stability over time (VC), and, from four diachronic corpora, frequency (WF) of their topics and vehicles.

Results showed no effect of epoch for either 'A is B' or 'A of B' literary metaphors. Thus, no noticeable change in CS over time was revealed, suggesting that these metaphors come with similar processing costs for contemporary readers and for readers of the epoch in which the metaphors were created. The absence of an effect of epoch can be better understood by considering the historical evolution of the English language, and specifically its early standardization. As stated by Wyld [40], literary writing as early as the 18th century was considered 'English of our own age in all its essentials'. In line with this consideration, our results point to the stability of the main stylistic features of the English language in the last two centuries, including those related to metaphors.

While literary metaphors are not processed differently based on the epoch, the influence of textual genre is noticeable. This factor emerged both as a main effect and in different interaction patterns with single-word variables, varying according to the type of metaphor.

For 'A of B' metaphors, results revealed that the difficulty of these metaphors changed as a function of the genre. In particular, they are perceived as less difficult when found in literary contexts, compared to when encountered in nonliterary texts. Hence, the difficulty of these metaphors is sensitive to the style of the text in which metaphors are found: when read in a text that has a literary style and aesthetic intent, the metaphor is less striking than the same metaphor in a nonliterary text. Moreover, we found a strong effect of the stability of the meaning of the vehicle in interaction with epoch and genre. This suggests that 'A of B' metaphors with more unstable vehicles are perceived as less difficult than 'A of B' metaphors with vehicles whose meanings remained stable over time. We interpreted this result in light of Traugott's [41] theory of *metaphorization*, according to which the metaphorical use of a word can become one of its stable meanings. In the context of the present study, words that changed the most could have done so by incorporating meanings derived from their metaphorical uses. As a result, when these unstable and broadened vehicles are used, metaphors appear less difficult. The reader does not need to broaden the concept expressed by the vehicle to interpret the metaphor, because the metaphorical nuances have entered the standard meaning of the word. From a qualitative observation of the data, we can notice, for instance, that a metaphor such as "Wave of horror", where the vehicle *wave* incorporated the meaning of 'sudden increase in a particular phenomenon', is perceived as less metaphorical than "Clouds of doubt", whose vehicle *clouds* has maintained its original meaning.

For 'A is B' metaphors, instead, the statistical model highlighted an effect of the frequency of the vehicle in interaction with epoch and genre. In nonliterary texts, the perceived difficulty of 'A is B' metaphors differed as a function of the WF of their vehicle, to the point that metaphors showed opposite patterns in the past and in the present: in the past, the less frequent the vehicle, the more metaphorical the whole metaphorical expression; in the present, the less frequent the vehicle, the less metaphorical the metaphor. The pattern found in the 19th-century space model is in line with previous studies [42] that found that metaphors with less frequent vehicles are regarded as more metaphorical than those with highly frequent vehicles, indicating that the most metaphorical metaphors are those in which the vehicle communicates something new about the topic. Going back to Hopkins' metaphor "The wind is a wrestler", the vehicle *wrestler*, as a particularly low frequency word in the 19th century, was indeed communicating something new about the topic *wind*. As such, the metaphors might have been perceived as more difficult and "more metaphorical", leading to the creation of a new concept. The very same metaphor is nowadays perceived differently, because the frequency of the vehicle has changed: *wrestler* has become more frequent, and the whole expression has lost some of its metaphoricity for the 21st-century readers.

Overall, our results suggest that for the English language, metaphor processing costs are not affected by the temporal distance between the creation of metaphors, which occurred in the 19th century, and their processing by today's readers. Instead, the key factor modulating metaphor processing costs seems to be the textual genre in which they appear. This modulation, however, occur to a different extent depending on the syntactic structure of the metaphors and in interaction with single word measures. Indeed, we observe that in defining what drives the difficulty of metaphors, different patterns emerged for the 'A of B' and 'A is B' structures. While for the former, in addition to the main effect of genre, we found the effect of vector coherence in interaction with epoch and genre, for the latter the diachronic evolution of metaphor processing costs is related to the interaction of word frequency with epoch and genre.

While these differences might reflect genuine effects of the syntactic structure and how it impacts metaphorical predication [43, 44, 45], we must acknowledge that the numerosity of the two sets of items varies and this might obscure some of the effects in the less represented type (A is B). Future studies are needed to further explore the whole range of diachronic changes in processing related to structural differences.

In conclusion, this proof-of-concept study proposed an adaptation from Italian to English of a method employing temporal word embeddings to study the evolution of metaphors. Thanks to this approach, we could elucidate that the processing costs of English literary metaphors is stable over time (differently from Italian) but is dynamically affected by stylistic features of texts and by single-word measures. The proposed method seems to be sensitive to the specificities of the language under investigation, supporting its crosslinguistic applicability.

5. Ethic statement

The work aims to use computational tools for the study of literature, thus enhancing the literary heritage with innovative methods that can provide insights for scholars from a wide range of disciplines. We are aware, however, that the corpora used are not representative of the entire spectrum of varieties of English, but of educated, Western English. Hence, our results may not coincide with the general evolution of the language but provide a partial view of it.

6. Data availability

Temporal vector space models and metaphor datasets used in the study are available at https://osf.io/j8bd7/?view_only=4cd623d5622b4ed0bd1624c42aff0f40$.

7. Acknowledgment

This work received support from the European Research Council under the EU's Horizon Europe program, ERC Consolidator Grant "PROcessing MEtaphors: Neurochronometry, Acquisition and Decay, PROMENADE" [101045733]. The content of this article is the sole responsibility of the authors. The European Commission or its services cannot be held responsible for any use that may be made of the information it contains.

References

[1] W. H. Gardner, N. H. MacKenzie (Eds.), The Poems of Gerard Manley Hopkins, Oxford University Press, 1967.

[2] W. L. Hamilton, J. Leskovec, D. Jurafsky, Diachronic word embeddings reveal statistical laws of semantic change, arXiv preprint arXiv:1605.09096 (2016).

[3] E. C. Traugott, R. B. Dasher, Regularity in semantic change, Cambridge University Press, 2002.

[4] B. W. Fortson IV, An approach to semantic change, in: B. D. Joseph, R. D. Janda (Eds.), The Handbook of Historical Linguistic., Blackwell Publishing, Malden, MA, 2003, pp. 648–666.

[5] A. Kutuzov, L. Øvrelid, T. Szymanski, E. Velldal, Diachronic word embeddings and semantic shifts: a survey, arXiv preprint arXiv:1806.03537 (2018).

[6] N. Tahmasebi, L. Borin, A. Jatowt, Survey of computational approaches to lexical semantic change, arXiv preprint arXiv:1811.06278 (2018).

[7] X. Tang, A state-of-the-art of semantic change computation, Natural Language Engineering 24 (2018) 649–676. doi:10.1017/S1351324918000220.

[8] Z. Harris, Distributional hypothesis, Word World 10 (1954) 146–162.

[9] M. Hilpert, Germanic future constructions: A usage-based approach to language change, John Benjamins Publishing, Amsterdam, 2008.

[10] V. Di Carlo, F. Bianchi, M. Palmonari, Training temporal word embeddings with a compass, in: Proceedings of the AAAI conference on artificial intelligence, volume 33, 2019, pp. 6326–6334.

[11] K. Gulordava, M. Baroni, A distributional similarity approach to the detection of semantic change in the google books ngram corpus., in: Proceedings of the GEMS 2011 workshop on geometrical models of natural language semantics, 2011, pp. 67–71.

[12] V. Kulkarni, R. Al-Rfou, B. Perozzi, S. Skiena, Statistically significant detection of linguistic change, in: Proceedings of the 24th international conference on world wide web, 2015, pp. 625–635.

[13] Y. Xu, C. Kemp, A computational evaluation of two laws of semantic change., in: Proceedings of

the 37th Annual Meeting of the Cognitive Science Society, 2015, p. 2703–2708.
[14] D. Wilson, R. Carston, A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts, in: N. Burton-Roberts (Ed.), Pragmatics, Palgrave Macmillan, Basingstoke, 2007, p. 230–259.
[15] M. Fludernik, D. C. Freeman, M. H. Freeman, Metaphor and beyond: An introduction, Poetics today 20 (1999) 383–396.
[16] A. N. Katz, A. Paivio, M. Marschark, J. M. Clark, Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions, Metaphor and Symbol 3 (1988) 191–214. doi:10.1207/s15327868ms0304_1.
[17] E. Semino, G. Steen, Metaphor in literature, in: J. R. W. Gibbs (Ed.), The Cambridge handbook of metaphor and thought, Cambridge University Press, Cambridge, 2008, pp. 232–246.
[18] V. Bambini, D. Resta, M. Grimaldi, A dataset of metaphors from the italian literature: Exploring psycholinguistic variables and the role of context, PloS one 9 (2014) e105634. doi:10.1371/journal.pone.0105634.
[19] V. Bambini, P. Canal, D. Resta, M. Grimaldi, Time course and neurophysiological underpinnings of metaphor in literary context, Discourse Processes 56 (2019) 77–97. doi:10.1080/0163853X.2017.1401876.
[20] S. Bhatia, R. Richie, W. Zou, Distributed semantic representations for modeling human judgment, Current Opinion in Behavioral Sciences 29 (2019) 31–36. doi:10.1016/j.cobeha.2019.01.020.
[21] F. Günther, C. Dudschig, B. Kaup, Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies, Quarterly Journal of Experimental Psychology 69 (2016) 626–653. doi:10.1177/174569161986137.
[22] M. N. Jones, J. Willits, S. Dennis, M. Jones, Models of semantic memory, in: J. Busemeyer, Z. Wang, J. Townsend, A. Eidels (Eds.), Oxford handbook of mathematical and computational psychology, Oxford University Press, New York, NY, 2015, pp. 232–254.
[23] S. McGregor, K. Agres, K. Rataj, M. Purver, G. Wiggins, Re-representing metaphor: Modeling metaphor perception using dynamically contextual distributional semantics, Frontiers in psychology 10 (2019) 765. doi:10.3389/fpsyg.2019.00765.
[24] N. J. Reid, H. Al-Azary, A. N. Katz, Cognitive factors related to metaphor goodness in poetic and non-literary metaphor, Metaphor and Symbol 38 (2023) 130–148. doi:10.1080/10926488.2021.2011285.
[25] B. Winter, F. Strik-Lievers, Semantic distance predicts metaphoricity and creativity judgments in synesthetic metaphors, Metaphor and the Social World 13 (2023) 59–80. doi:10.1075/msw.00029.win.
[26] A. N. Katz, A. Paivio, M. Marschark, Poetic comparisons: Psychological dimensions of metaphoric processing, Journal of Psycholinguistic Research 14 (1985) 365–383. doi:https://doi.org/10.1007/BF01067881.
[27] E. Shutova, Design and evaluation of metaphor processing systems, Computational Linguistics 41 (2015) 579–623. doi:https://doi.org/10.1162/COLI_a_00233.
[28] V. Mangiaterra, C. Barattieri di San Pietro, V. Bambini, How literary metaphors change over two centuries (in prep).
[29] A. Englhardt, J. Willkomm, M. Schäler, K. Böhm, Improving semantic change analysis by combining word embeddings and word frequencies, International Journal on Digital Libraries 21 (2020) 247–264. doi:10.1007/s00799-019-00271-6.
[30] Q. Feltgen, B. Fagard, J.-P. Nadal, Frequency patterns of semantic change: corpus-based evidence of a near-critical dynamics in language change, Royal Society open science 4 (2017) 170830. doi:10.1098/rsos.170830.
[31] H. Al-Azary, A. N. Katz, On choosing the vehicles of metaphors 2.0: the interactive effects of semantic neighborhood density and body-object interaction on metaphor production, Frontiers in Psychology 14 (2023) 1216561. doi:10.3389/fpsyg.2023.1216561.
[32] M. Honnibal, I. Montani, spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing (2017). doi:https://doi.org/10.5281/zenodo.3358113.
[33] P. Hanks, Metaphoricity is gradable, Trends in Linguistics Studies and Monographs 171 (2006) 17. doi:https://doi.org/10.1515/9783110199895.17.
[34] V. Bambini, M. Trevisan, Esploracolfis: Un'interfaccia web per le ricerche sul corpus e lessico di frequenza dell'italiano scritto (colfis), Quaderni del Laboratorio di Linguistica 11 (2012) 1–16.
[35] L. Han, A. L. Kashyap, T. Finin, J. Mayfield, J. Weese, Umbc_ebiquity-core: Semantic textual similarity systems, in: Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity, 2013, pp. 44–52.
[36] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781 (2013).
[37] E. Bruni, G. Boleda, M. Baroni, N.-K. Tran, Distri-

butional semantics in technicolor, in: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2012, pp. 136–145.

[38] W. Kintsch, Metaphor comprehension: A computational theory, Psychonomic bulletin & review 7 (2000) 257–266. doi:10.3758/BF03212981.

[39] D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4, arXiv preprint arXiv:1406.5823 (2014).

[40] H. C. Wyld, A history of modern colloquial English, Basil Blackwell, 1936.

[41] E. C. Traugott, Semantic change, in: Oxford research encyclopedia of linguistics, Oxford University Press, 2017. doi:10.1093/acrefore/9780199384655.013.323.

[42] J. Littlemore, P. Pérez Sobrino, D. Houghton, J. Shi, B. Winter, What makes a good metaphor? a cross-cultural study of computer-generated metaphor appreciation, Metaphor and Symbol 33 (2018) 101–122. doi:10.1080/10926488.2018.1434944.

[43] V. Bambini, M. Ghio, A. Moro, P. B. Schumacher, Differentiating among pragmatic uses of words through timed sensicality judgments, Frontiers in psychology 4 (2013) 938. doi:https://doi.org/10.3389/fpsyg.2013.00938.

[44] R. Carston, X. Yan, Metaphor processing: Referring and predicating, Cognition 238 (2023) 105534. doi:https://doi.org/10.1016/j.cognition.2023.105534.

[45] E. Tonini, L. Bischetti, P. Del Sette, E. Tosi, S. Lecce, V. Bambini, The relationship between metaphor skills and theory of mind in middle childhood: Task and developmental effects, Cognition 238 (2023) 105504. doi:https://doi.org/10.1016/j.cognition.2023.105504.