Semantic Distance and Creativity in Linguistic Synaesthesia

Emmanuele Chersoni
The Hong Kong Polytechnic University
Chinese and Bilingual Studies
emmanuelechersoni@polyu.edu.hk

Francesca Strik Lievers
University of Genoa
Lingue e Culture Moderne
francesca.striklievers@unige.it

Chu-Ren Huang
The Hong Kong Polytechnic University
Chinese and Bilingual Studies
churen.huang@polyu.edu.hk

Abstract

In this work, we aim at quantitatively assessing the *creativity of linguistic combinations* in terms of semantic distance, using synaesthetic metaphors (e.g. *bitter voice*) as a case study. We created an evaluation dataset containing examples of synaesthesia that are actually occurring in corpora and automatically generated synaesthetic metaphors, together with a control set of non-synaesthetic adjective-noun combinations. Then, we tested on the dataset three quantitative models of linguistic creativity that have been proposed in the NLP and in the cognitive science literature, and we compared their performance in discriminating between creative and non-creative, directional and non-directional synaesthetic metaphors, and between synaesthetic metaphors and non-synaesthetic phrases.

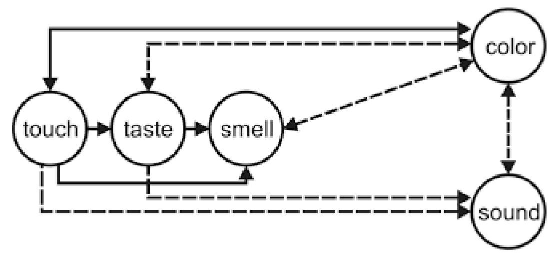

Figure 1: An example illustration of the directionality of synaesthetic transfers, graphically represented by the arrows that connect the senses (the image is taken from Werning et al., 2006), with the arrows indicating the directionality of the transfers. The details of intersensory connections may change depending on the specific language and on the specific study, but some common features are: a) transfers generally go from lower senses (touch, taste, smell) to the higher ones (sound and vision, which is sometimes identified with color, and sometimes divided between dimension and color); b) touch is the most common source and hearing the most common target; c) transfers between sound and vision are bidirectional.

1 Introduction

According to classical definitions, linguistic synaesthesia is a type of metaphor in which an experience related to a sensory modality (e.g. touch, hearing, etc.) is described through lexical means that are typically associated to a different sensory modality (Strik Lievers, 2015; Huang and Xiong, 2019). This figure is often discussed in studies on poetic and more generally literary texts (Ullmann, 1957; Shen and Cohen, 1998; Bretones-Callejas, 2001). On the one hand, synaesthesia has played an important role for literary poetics since the 19th century: see for instance the key role played by intersensory experiences in the works of symbolist poets such as Baudelaire and Rimbaud. On the other hand, research on synaesthesia as a clinical condition (Ramachandran and Hubbard, 2001; Simner and Hubbard, 2013) started a new trend of cognitively-inspired studies, putting the phenomenon in relation with the development of creative skills in the individuals.

More recent contributions focused instead on synaesthetic metaphors in ordinary language, on the basis of corpus-based analysis (see, for example, the studies of Marotta, 2012 on Italian; of Strik Lievers, 2015 on English and Italian; and of Jo, 2017; 2018 on Korean). A common point of agreement among most scholars is the observation that - both in literary and in ordinary language - synaesthetic transfers

Pacific Asia Conference on Language, Information and Computation (PACLIC 33), pages 370-378, Hakodate, Japan, September 13-15, 2019

are *directional* (among others, Ullmann, 1957, Shen and Cohen, 1998; for a critical discussion of this notion, see Strik Lievers, 2015; Winter, 2019a). That is, the synaesthetic transfers typically go from the "lower" senses (touch, smell and taste), which are the most common sources, to the "higher" senses (sight and sound), which are the most common targets (see Figure 1). For example, a synaesthetic metaphor like *sweet silence* is much more likely to occur than *silent sweetness* (Shen and Cohen, 1998).

In the present work, we adopt a different perspective on synaesthesia, since we are interested in the general notion of *linguistic creativity* and its quantitative assessment, as proposed in the recent cognitive science literature (Heinen and Johnson, 2017; Kenett, 2018a). [1] According to this view, the creativity of linguistic combinations can be seen as a function of *semantic distance*: the most creative combinations are those linking together concepts whose representations are far apart in the semantic memory space (Kenett, 2018a). Metaphors fit well this definition, as they typically link together concepts belonging to different conceptual domains (Lakoff and Johnson, 1980), and we believe this is the also the case for synaesthetic associations. [2] The first research question of our study is the following: *can models of semantic distance distinguish between synaesthetic metaphors and non-synaesthetic usages of sensory lexical items?* Secondly, are they able to detect *different degrees of creativity* in synaesthesia? And finally, *is semantic distance related to the directional tendency that has been observed by previous studies in synaesthesia*? In other words, can the rarity of some transfer types be explained by higher distances between concepts in the semantic memory?

Our paper is organized as follows. The computational models of creativity and semantic distance are briefly reviewed in Section 2, together with the literature on the creation of the sensory lexicon that we will use for querying synaesthetic metaphors in corpora. In Section 3, we present our procedure for generating a dataset including synaesthetic expressions with different degrees of creativity, and we describe the parameters of our experimental settings. In Section 4 we report the results of our experiments which are summarized and discussed in Section 5.

2 Related Work

2.1 Computational Models of Creativity

The traditional associative theory of creativity is probably due to Mednick (Mednick, 1962), according to which the notion involves the connection of remote, or weakly-related concepts. A common feature unifying this theory to modern research is the importance of the structure of human semantic memory in defining the distance between concepts (Kenett, 2018a). In the classic *Spreading Activation Model* (Collins and Loftus, 1975), the concepts in memory are organized in a network, and their proximity depends on their semantic similarity: concepts sharing many semantic properties will be connected by many links. Once a node in the network graph is activated, the activation spreads to all the direct neighbors, then decaying over time and space. In this model, semantic distance can be seen as the *length of the shortest path* connecting two concepts. This idea has been inspirational for some recent computational studies, which proposed to formalize the notion of semantic distance as the path length in a semantic network, and to conceive creative associations as new connections between distant nodes (Kenett, 2018b).

The most popular model for representing semantic distance in cognitive psychology is probably Latent Semantic Analysis (Landauer and Dumais, 1997). LSA models represent lexical items as vectors in a high-dimensional semantic space, on the basis of their distributional behavior in corpora. Vectors that are close correspond to semantically similar words, and the cosine between their angles is the most common similarity metric. LSA is a technique widely used also in the research field known as *Distributional Semantics* (see Lenci, 2008 for an overview), and such distributional models have been recently proposed, among the other things, to measure the semantic distance between the concepts involved in visual and linguistic metaphors (Bolognesi and Aina, 2019) and to account for the novelty and appropriateness of human noun-verb associations (Heinen and Johnson, 2017).

[1] For an overview of the different theories of linguistic creativity, see Veale (2012) and Jones (2016).

[2] Interestingly, conventional and creative metaphorical associations have been shown to activate different brain regions during sentence processing (Ahrens et al., 2007).

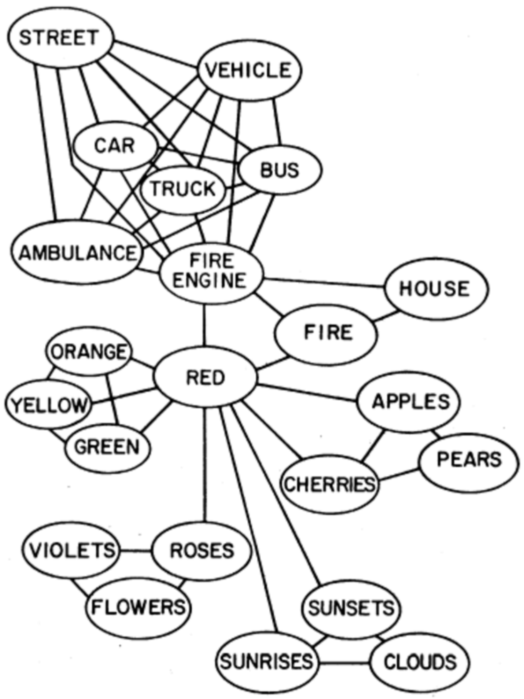

Figure 2: A schematic illustration from Collins and Loftus (1975) of the semantic memory structure. In the network graph, the shorter the line connecting the concepts, the higher their semantic relatedness.

A third metric that can be used to assess creativity, here adopted with this goal for the first time, is thematic fit (McRae and Matsuki, 2009; Lenci, 2011). Thematic fit can be described as the degree of compatibility, based on our event knowledge, between a verb and a given argument, but the notion can also be extended to adjective-noun combinations to measure the typicality of a given attribute for an entity (see the model of semantic anomaly for attributive adjective-noun pairs by Vecchi et al., 2011). In distributional models, the concept is often operationalized by means of *prototypes*: given a noun like *sound*, a representation of its typical attribute is built by averaging the vectors of the most typical co-occurring adjectives (e.g. *pleasant, nice, annoying, loud* etc.) and measuring the similarity between this prototype and a new, candidate attribute (e.g. *sweet*). Thus, for a synaesthetic combination like *sweet sound*, the degree of creativity will be an inverse function of the similarity between *sweet* and the typical adjectival modifiers of *sound*.

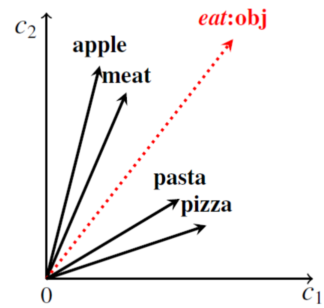

Figure 3: An illustration of a thematic fit model taken from Lenci, 2017 (oral presentation in Barcelona). In this case, the thematic fit for potential objects of *to eat* would be measured as their similarity with the prototype vector of the patient of the verb (in red).

2.2 Sensory Lexica for the Identification of Synaesthesic Metaphors

Two important requirements for building our evaluation dataset for synaesthesia are:

1. a methodology for automatically extracting synaesthetic metaphor candidates from corpora;

2. a list of words associated to different senses and annotated for part-of-speech, typically nouns and adjectives since noun phrases with an adjectival modifier are the most common form of synaesthesia [3] .

Adjective-noun combinations are the form of synaesthetic metaphor on which we will be focusing in the current study. As for the methodology, we adopt the dependency-based search proposed by Strik Lievers and Huang, 2016: given a parsed corpus, we look for all the adjective-noun phrases in which the adjective and the noun are typically associated with different sensory modalities.

Concerning the sensory lexicon, the first resource made available were probably the norms collected by Lynott and Connell (Lynott and Connell, 2009; Lynott and Connell, 2013), in which the association between words and sensory modality were generated on the basis of human ratings. An interesting

[3] For a systematic study on the distribution of lexical categories across different sensory modalities in the English sensory lexicon, see Strik Lievers and Winter, 2018.

feature of these datasets, respectively including 423 adjectives and 400 nouns, is that they also provide ratings reflecting to what extent each word tends to be associated with a single sensory modality. [4]

Another important resource is the sensory lexicon that has been built in a semi-automatic way by Tekiroglu et al., 2014, by using a list of words extracted from WordNet and expanded by means of NPMI association scores (Bouma, 2009). In terms of size, Sensicon is by far the biggest sensory dataset currently available, with associations for more that 22,000 words. However, being a large semi-automatically built resources, association scores have not been manually checked, and therefore it contains noisy data.

Finally, a wordlist annotated with sense associations was manually compiled by Strik Lievers, 2015, for a corpus-study on synaesthetic metaphors. Since the item selection was more controlled, we chose to use this wordlist as a reference for extracting a list of synaesthetic metaphors from corpora. However, for the purposes of our evalution we also wanted to generate a second list of *creative* synaesthetic metaphors that are unlikely to be found in corpora. For this task, we decided to rely on Lynott and Connell's dataset, which comes with ”monoesthesia” scores: the idea is that, the stronger the association of a word with a single sense, the less likely it will be that it enters into a synaesthetic association.

3 Experiments

3.1 Dataset Creation

The first step for us was to extract a first set of synaesthetic metaphor candidates from a parsed corpus: we chose the British National Corpus, since it is a balanced corpus containing a wide variety of textual genres (Leech, 1992). As we said, we only took into account adjective-noun phrases, and we used as a seed set the nouns and the adjectives manually annotated by Strik Lievers, 2015. The set includes 119 nouns (13 for smell, 22 for taste, 5 for touch, 59 for hearing, 20 for sight) and 190 adjectives (10 for smell, 28 for taste, 43 for touch, 30 for hearing, 79 for sight).

Source / Target	Sight	Sound	Taste	Smell	Touch
Sight	-	272	**16**	**13**	**23**
Sound	22	-	**3**	**1**	-
Taste	20	89	-	233	**1**
Smell	-	-	-	-	-
Touch	174	476	98	130	-

Table 1: Summary of the transfers types for the candidate synaesthetic metaphors from the BNC. For each cell, the sense in the row is the source, while the sense in the column is the target. In bold, the transfers that contradict the directionality principle of the classical sense hierarchies.

As a result of the extraction, we obtained 1571 occurrences of synaesthetic combinations from the BNC (after manually filtering out some noise, mainly due to metonymic expressions such as *black music*). The different types of transfer between senses are shown in Table 1: at a glance, it is clear that they mostly follow the principle of directionality of classical sense hierarchies, with some exceptions (in most cases, adjective-noun phrases with sight as the source sensory modality). As for the total number of synaesthetic *types*, we found 471 of them: we call this set SYN. That is, the SYN set includes types of **synaesthetic metaphors that occur in the British National Corpus**. Secondly, we generated two other sets of phrases, to be used for comparison with our newly-found synaesthetic metaphors:

- **control collocations**: for all words in the synaesthetic expressions of the SYN set, we aimed at extracting a common collocate. This means that, for nouns, we generated new adjective-noun phrases by combining nouns with their most typical adjectival modifiers (e.g. for *colour*: *bright colour*, *dark colour*, etc.). For adjectives, we did the same by combining them with the nouns that they typically modify (e.g. for *bitter*: *bitter disappointment*, *bitter taste*, etc.);
- **new synaesthetic metaphors**: in order to create new, creative synaesthetic metaphors, we adopted the view that the more creative lexical associations are those linking concepts that are very distant. Thus, in the case of synaesthesia, they are more likely to involve concepts

[4] We recently became aware that Lynott and colleagues have released a new and bigger sensory norms dataset on Psyarxiv in May 2019 (Lynott et al., 2019), including almost 40,000 words: it was unfortunately too late to use it for the present studies, but it will certainly be an important resource for future works.

that typically do not co-occur with more than one sense. By using the word-sense association scores of the datasets by Lynott and Connell, we randomly combined adjectives and nouns with very high degrees of monoestheticity (i.e. a score quantifying the tendency of being associated with a single sense only).

For the control collocations, the typical collocates have been extracted on the basis of Positive Local Mutual Information ($PLMI$) (Evert, 2004), measured from the co-occurrences of adjectives and nouns in the Wacky corpus (Baroni et al., 2009). This metric can be seen as a slightly modified version of the more common PMI (Church and Hanks, 1990), less biased towards rare events.

Given an adjective adj and a noun n, the $PLMI$ is computed as follows:

$$LMI(adj, n) = log\left(\frac{f_{adj,n} * C}{f_{adj} * f_n}\right) * f_{adj,n} \quad (1)$$

$$PLMI(adj, n) = max(LMI(adj, n), 0) \quad (2)$$

where f_{adj} and f_n are the respective frequencies of adj and n, $f_{adj,n}$ is the frequency of their joint co-occurrence, and C the number of observed word pairs in the corpus. In other words, the $PLMI$ measures the statistical association between adjectives and nouns by comparing their observed co-occurrence with the expected co-occurrence under the assumption of statistical independence between the two. Each adjective and noun of the Strik Lievers list has been combined with the top $PLMI$-scoring word for the other POS, to create examples of the usage of those words in ”standard”, non-figurative expressions. Notice that this methodology does not guarantee that collocates will be retrieved for all words appearing in the synaesthetic metaphors, as some of them might be rare in Wacky and/or might not have collocates with enough statistical association strength. Out of the 156 different words composing the expressions in SYN, we could retrieve collocates only for 132 of them to form the adjective-noun phrases of the CONTROL set.

Finally, for the set of the new synaesthetic metaphors, we adopted the following procedure: for each adjective or noun in the set SYN, we generated a new combination with a word of the other category included in the data by Lynott and Connell, provided that it has a monoestheticity score equal or superior to 6.[5] Then, we randomly sampled 471 of these expressions, in order to have a set of the same size of SYN. We refer to this new set as NEW_SYN.

3.2 Models and Experimental Settings

The three models of creativity that we will test in this study are *Latent Semantic Analysis*, *Thematic Fit* and *shortest path length*. All of them have been trained on the Wacky corpus (Baroni et al., 2009)

The models represent as targets all the words included in our datasets. As contexts for the target words, we use the 30,000 more frequent words in the Wacky corpus (only considering nouns, verbs and adjectives).

We trained LSA models in two different versions: an unweighted version, and a version where co-occurrences between targets and contexts have been weighted via Positive Pointwise Mutual Information ($PPMI$).

$$PMI(adj, n) = log\left(\frac{f_{adj,n} * C}{f_{adj} * f_n}\right) \quad (3)$$

$$PPMI(adj, n) = max(PMI(adj, n), 0) \quad (4)$$

We refer to these two versions of the model as LSA and LSA_PPMI. For both of them, we reduced the word-context matrix by setting the parameter of the SVD components to 300 (the only difference being that, with LSA_PPMI, the frequencies are weighted before the dimensionality reduction step).

The thematic fit models use dependency-based contexts: that is, each dimension of the semantic space is a combination of 1) one of the 30,000 words; and 2) a syntactic dependency relation linking the word with the target (for example, *ADJ_MOD:loud* might be a context for the target

[5]The threshold was empirically selected: with a higher threshold, the number of candidate words for creating new combinations was too small.

noise). We trained two different thematic fit models: one assigning a score to an adjective-noun phrase by measuring the similarity between the adjective and the prototype of the noun modifiers; and the other measuring the similarity between the noun and the prototype of the nouns modified by the adjective. We call these models $TFIT\_ADJ$ and $TFIT\_N$.

For building these prototypes, we averaged the vectors of the 10 most strongly $PLMI$-associated collocates for each relation (that is, the 10 adjectives with the highest $PLMI$ association score as modifiers of the noun, and the 10 nouns having the highest $PLMI$ association score with the adjective as a modifier). In case one of the top nouns or adjectives had been previously used for generating the CONTROL set, it was excluded from the list for generating the prototype.

As for the shortest path length model ($PATH$), we use the same $PPMI$ matrix of the weighted LSA model to generate an undirected graph. In this graph, the nodes correspond to the target words. Two words are linked by an edge if they co-occur and their $PPMI$ score is $>= 0$. The weight for each edge is equal to 1 divided by the $PPMI$ score for the two words, which means, edges connecting strongly associated words will have a lower traveling cost. In order to compute the scores for this model, for each adjective-noun phrase in our dataset we computed the shortest path length between the adjective and the noun in the graph, by using the classical Dijkstra algorithm (Dijkstra, 1959).

The output of each model will be a score of the semantic similarity between the two words in the adjective-noun phrase. As we explained in the introductory section, the more creative combinations are those linking together more distant concepts. Thus, the scores of the model have to be read as *inverse indexes of creativity*: the more two items are similar, the lower their semantic distance. If a model is doing well, we expect the scores to pattern in the following way: 1) the CONTROL items get the highest scores; 2) the NEW_SYN ones get the lowest scores and 3) SYN items score in the middle.

4 Results

A quick check of the Spearman correlations between the different metrics reveals that they are weakly correlated, the only exceptions being LSA and LSA_PPMI, with $\rho = 0.48$ (as expected, as they are just different versions of the same metric), and LSA_PPMI and $TFIT\_N$, with $\rho = 0.37$.

After computing the scores for all the four models, we ran statistical tests to see how good are the models in discriminating between the different experimental conditions. According to the Kruskal-Wallis test, all models can find a highly significant difference between conditions (for all of them, $p < 0.001$): the boxplots for $TFIT\_ADJ$ and LSA_PPMI are shown, as an example, in Figures 4 and 5.

We came then to the post-hoc tests: for the Wilcoxon rank sum test, again all models find a significant difference between CONTROL and NEW_SYN (for all of them, $p < 0.001$), between CONTROL and SYN ($p < 0.01$), and between SYN and NEW_SYN ($p < 0.01$). This is an interesting result: we expected the models to be able to easily discriminate between CONTROL and the other two conditions, but given the rarity of synaesthetic metaphors, it is surprising that they also manage to distinguish between those that were actually found in the BNC and the automatically generated ones. We should recall here that the latter ones were generated in order to be more "creative", by combining the words of the SYN set with words i) of a different sensory modality and ii) that are typically not associated with words of a different sensory modality. From this point of view, all models did a good job in recognizing different levels of creativity between the two sets of synaesthetic combinations.

Interestingly, there was a partial exception in our results: the simple $TFIT\_N$ model only found a marginally significant difference between NEW_SYN and SYN ($W = 70619$, $p < 0.05$), while the difference for the $TFIT\_ADJ$ was much larger ($W = 14368$, $p < 0.001$). Thus, the model based on the similarity between the actual and the prototypical modifier of the noun seems to have a higher discriminative power.

To address the last question of our exploratory study, i.e. whether semantic distance is related to directionality observed in the literature on synaesthetic metaphors, we assigned a class to each of the expressions in the SYN and in the NEW_SYN set, depending on the type of synaesthetic transfer: we

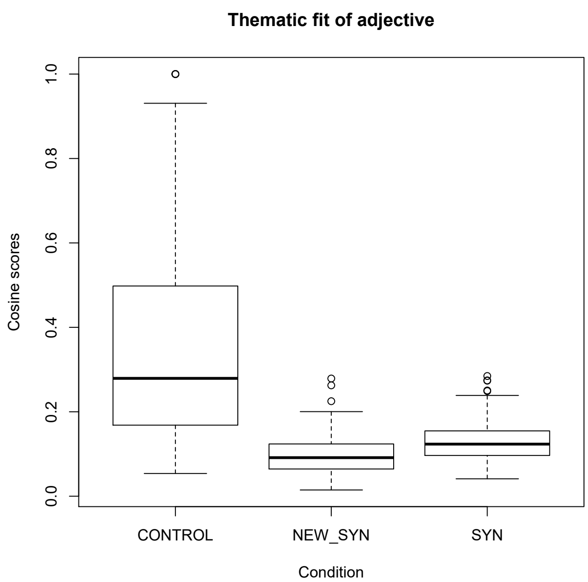

Figure 4: Cosine similarity scores assigned by $TFIT\_ADJ$ to the items in the three conditions.

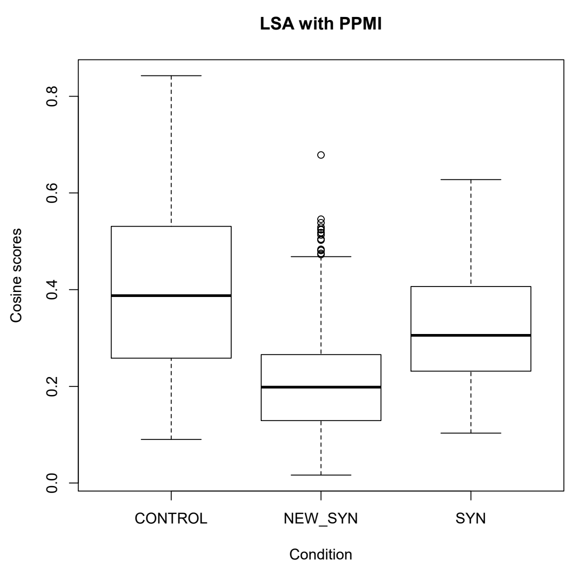

Figure 5: Cosine similarity scores assigned by LSA_PPMI to the items in the three conditions.

assigned a positive label to those expressions that are coherent with the directionality (e.g. pairs with a *taste* adjective and a *sight* noun) and a negative label to those that are not (e.g. pairs with a *sound* adjective and a *smell* noun). In total, we had 834 coherent and 108 non-coherent transfers. Again, we test our models in their ability of discriminating between the conditions, in order to check if semantic distance metrics confirm that transfers not respecting the common directionality are more creative. In this case, semantic distance scores assigned to the pairs with a negative label should be significantly higher (i.e. lower cosines for the distributional models, longer paths for the $PATH$ one).

It turns out that most models are not able to make the distinction: for the Wilcoxon rank sum test, $PATH$ ($W = 40450$, $p > 0.05$) and LSA ($W = 45644$, $p > 0.05$) failed to find a significant difference between directional and non-directional combinations. At a closer inspection, we found that a highly significant difference is found by LSA_PPMI ($W = 53937$, $p < 0.001$), but not in the expected direction: the cosines for the non-directional combinations are significantly higher, instead of being lower. On the other hand, thematic fit models struggle for the low coverage: out of the 108 items in the non-coherent set, only 40 take non-zero values for $TFIT\_ADJ$ and only 60 for $TFIT\_N$. By running the Wilcoxon test on the remaining phrases, both models fail to find a significant difference ($p > 0.1$). It should be noticed that we used a classical version of the thematic fit model, based on dependencies (Baroni and Lenci, 2010), which suffers by definition of more sparsity. If two vectors do not share any dependency-based context, their similarity will be zero, and from this point of view, the result could make sense, since the adjective-noun combinations of the non-coherent phrases are extremely unlikely. Actually, by taking into account also the phrases with a similarity score of zero, both $TFIT\_ADJ$ and $TFIT\_N$ assign significantly lower scores to the non-coherent combinations ($p < 0.01$ for both of them).

5 Conclusion

In this study, we tested three models of semantic distance to assess the creativity of linguistic combinations, taking over the task of distinguishing between synaesthetic metaphors and control expressions, and between the former and some automatically-generated, more creative synaesthetic combinations. We found that all models are able to find significant differences and to properly distinguish between the conditions.

Then, we also tested our models on the task of distinguishing between those combinations that are consistent with the directionality tendency (from the lower to the higher senses) observed in the studies on synaesthesia, and those that are not. We found that this task is much more difficult: thematic fit models might be the closest to identify this distinction, as their similarity assessment is based on the direction of the dependency relation between the adjective and the noun. Thus, they could implicitly incorporate some notion of the directionality of synaesthetic metaphors (i.e. what are the typical source and target domains). On the other hand, they suffer from data sparsity: most of the non-coherent phrases of our dataset got assigned a similarity score of zero, and we found a significant difference between the two conditions only by including these latter phrases.

To the best of our knowledge, this is the first study in computational modeling on the topic of linguistic synaesthesia, and the first trying to account for its combinatory patterns in terms of semantic distance. In a recent contribution, Jo (2018) pointed out that there was almost no connection between the corpus- and the cognitive science-oriented perspectives of research on the phenomenon. We believe that the notion of semantic distance, seen as a possible factor influencing the likelihood of sensory words combinations as observed in natural language corpora, could provide a link between these two trends of studies. On the one hand, the notion has been proposed by the modern research in cognitive science, but on the other hand it can be modeled in a straightforward way by means of corpus-based models of meaning.

Some promising models for our future tests include a thematic fit model based on dense spaces, in order to overcome the sparsity problem, or shortest path length models based on directed graphs, which should also be better at modeling the directionality of synaesthetic metaphors. Another possible direction is repeating the experiments with other languages for which large-scale modality exclusivity norms have been made available, such as Mandarin Chinese (Chen et al., 2019) and Italian (Morucci et al., 2019).

References

Kathleen Ahrens, Ho-Ling Liu, Chia-Ying Lee, Shu-Ping Gong, Shin-Yi Fang, and Yuan-Yu Hsu. 2007. Functional MRI of Conventional and Anomalous Metaphors in Mandarin Chinese. In *Brain and Language*, vol. 100, no. 2, pp. 163-171. Elsevier.

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. In *Language, Resources and Evaluation*, vol. 43, no. 3, pp. 209-226. Springer.

Marco Baroni and Alessandro Lenci. 2010. Distributional Memory: A General Framework for Corpus-Based Semantics. In *Computational Linguistics*, vol. 36, no. 4, pp. 673-721.

Marianna Bolognesi and Laura Aina. 2019. Similarity Is Closeness: Using Distributional Semantic Spaces to Model Similarity in Visual and Linguistic Metaphors. In *Corpus Linguistics and Linguistic Theory*, vol. 15, no. 1, pp. 101-137. De Gruyter.

Gerlof Bouma. 2009. Normalized Pointwise Mutual Information in Collocation Extraction. In *Proceedings of GSCL*.

Carmen Bretones-Callejas. 2001. Synaesthetic Metaphors in English. Technical Reports. University of California at Berkeley and International Computer Science Institute.

I-Hsuan Chen, Qingqing Zhao, Yunfei Long, Qin Lu, and Chu-Ren Huang. 2019. Mandarin Chinese Modality Exclusivity Norms. In *PLOS One*, vol. 14, no. 2.

Kenneth Ward Church and Patrick Hanks. 1990. Word Association Norms, Mutual Information, and Lexicography. In *Computational Linguistics*, vol. 16, no. 1. MIT Press.

Allan Collins and Elizabeth Loftus. 1975. A Spreading-Activation Theory of Semantic Processing. In *Psychological Review*, vol. 82, no. 6. American Psychological Association.

Edsger W Dijkstra. 1959. A Note on Two Problems in Connexion with Graphs. In *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271. Springer.

Stefan Evert. 2004. The Statistics of Word Co-occurrences: Word Pairs and Collocations. PhD Thesis, University of Stuttgart.

David JP Heinen and Dan R Johnson. 2017. Semantic Distance: An Automated Measure of Creativity That Is Novel and Appropriate. *Psychology of Aesthetics*

Creativity and the Arts. Amerian Psychological Association.

Chu-Ren Huang and Jiajuan Xiong. 2019. Linguistic Synaesthesia: An Introduction. *The Routledge Handbook of Chinese Applied Linguistics*. Routledge.

Charmhun Jo. 2017. A Corpus-Based Study on Synesthesia in Korean Ordinary Language. In *Proceedings of PACLIC*.

Charmhun Jo. 2018. Synaesthetic Metaphors in Korean Compound Words. In *Proceedings of the LREC Workshop on Linguistic and Neurocognitive Resources*.

Rodney H Jones. 2016. The Routledge Handbook of Language and Creativity. Routledge.

Yoed N Kenett. 2018. What Can Quantitative Measures of Semantic Distance Tell Us About Creativity? In *Current Opinions in Behavioral Sciences*, volume 27, pp. 11–16.

Yoed N Kenett. 2018. Investigating Creativity from a Semantic Network Perspective. In *Exploring Transdisciplinarity in Art and Science*, pp. 49-75. Springer.

George Lakoff and Mark Johnson. 1980. Metaphors We Live By. University of Chicago Press.

Thomas K Landauer and Susan Dumais. 1997. A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. In *Psychological Review*, vol. 104, no. 2. American Psychological Association.

Geoffrey Neil Leech. 1992. 100 Million Words of English: The British National Corpus (BNC).

Alessandro Lenci. 2008. Distributional Semantics in Linguistic and Cognitive Research. In *Italian Journal of Linguistics*, vol. 20, no. 1, pp. 1–31.

Alessandro Lenci. 2011. Composing and Updating Verb Argument Expectations: A Distributional Semantic Model. In *Proceedings of the ACL Workshop on Cognitive Modeling and Computational Linguistics*.

Dermot Lynott and Louise Connell. 2009. Modality Exclusivity Norms for 423 Object Properties. In *Behavior Research Methods*, vol. 41, no. 2, pp. 558–564. Springer.

Dermot Lynott and Louise Connell. 2013. Modality Exclusivity Norms for 400 Nouns: The Relationship Between Perceptual Experience and Surface Word Form. In *Behavior Research Methods*, vol. 45, no. 2, pp. 516–526. Springer.

Dermot Lynott, Louise Connell, Marc Brysbaert, James Brand, and James Carney. 2019. Lancaster Sensorimotor Norms. In *PsyArxiv*.

Giovanna Marotta. 2012. Sinestesie tra Vista, Udito e Dintorni. Un'Analisi Semantica Distribuzionale. In *Sinestesie e Monoestesie*, edited by Marina Catricalà. Milano: Franco Angeli.

Ken McRae and Kazunaga Matsuki. 2009. People Use Their Common Knowledge to Understand Language, and Do So as Quickly as Possible. In *Language and Linguistics Compass*, vol. 3, no. 6, pp. 1417–1429.

Sarnoff Mednick. 1962. The Associative Basis of the Creative Process. In *Psychological Review*, vol. 69: pp. 220-232. American Psychological Association.

Piermatteo Morucci, Roberto Bottini, and Davide Crepaldi. 2019. Augmented Modality Exclusivity Norms for Concrete and Abstract Italian Property Words. In *OSF Preprints*.

Vilayanur S Ramachandran and Edward Hubbard. 2001. Synaesthesia–A Window into Perception, Thought and Language. In *Journal of Consciousness Studies*, vol. 8, no. 2, pp. 3-34.

Yeshayahu Shen and Michael Cohen. 1998. How Come Silence Is Sweet but Sweetness Is Not Silent: A Cognitive Account of Directionality in Poetic Synaesthesia. In *Language and Literature*, volume 7, no. 2, pp. 123–140. Sage Publications Sage CA: Thousand Oaks, CA.

Francesca Strik Lievers. 2015. Synaesthesia: A Corpus-Based Study of Cross-Modal Directionality. In *Functions of Language*, volume 22, no. 1, pp. 69–95. John Benjamins.

Francesca Strik Lievers and Chu-Ren Huang. 2016. A Lexicon of Perception for the Identification of Synaesthetic Metaphors in Corpora. In *Proceedings of LREC*.

Francesca Strik Lievers and Bodo Winter. 2018. Sensory Language Across Lexical Categories In *Lingua*, pp. 45–61. Elsevier.

Julia Simner and Edward Hubbard. 2013. Oxford Handbook of Synaesthesia. Oxford University Press.

Serra Sinem Tekiroglu, Gözde Özbal, and Carlo Strapparava. 2014. Sensicon: An Automatically Constructed Sensorial Lexicon. In *Proceedings of EMNLP*.

Stephen Ullmann. 1957. The Principles of Semantics. Glasgow: Jackson.

Tony Veale. 2012. Exploding the Creativity Myth: The Computational Foundations of Linguistic Creativity. A&C Black.

Eva Maria Vecchi, Marco Baroni, and Roberto Zamparelli. 2011. (Linear) Maps of the Impossible: Capturing Semantic Anomalies in Distributional Space In *Proceedings of the ACL Workshop on Distributional Spaces and Compositionality*.

Markus Werning, Jens Fleischhauer, and Hakan Beseoglu. 2006. The Cognitive Accessibility of Synaesthetic Metaphors. In *Proceedings of CogSci*.

Bodo Winter. 2019. Sensory Linguistics. Amsterdam: Jonh Benjamins.

Bodo Winter. 2019. Synaesthetic Metaphors Are Neither Synaesthetic Nor Metaphorical. In *Perception Metaphors*. Amsterdam: Jonh Benjamins, pp. 105–126.