

## Research Article

# A Novel Procedure for Measuring Semantic Synergy

**Yair Neuman, Yiftach Neuman, and Yochai Cohen**

*Ben-Gurion University of the Negev, 84105 Beer-Sheva, Israel*

Correspondence should be addressed to Yair Neuman; [yneuman@bgu.ac.il](mailto:yneuman@bgu.ac.il)

Received 8 September 2016; Revised 3 November 2016; Accepted 5 December 2016; Published 15 January 2017

Academic Editor: Sergio Gómez

Copyright © 2017 Yair Neuman et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One interesting characteristic of some complex systems is the formation of macro level constructions perceived as having features that cannot be reduced to their micro level constituents. This characteristic is considered to be the expression of synergy where the joint action of the constituents produces unique features that are irreducible to the constituents isolated behavior or their simple composition. The synergy, characterizing complex systems, has been well acknowledged but difficult to conceptualize and quantify in the context of computing the emerging meaning of various linguistic and conceptual constructs. In this paper, we propose a novel measure/procedure for quantifying semantic synergy. This measure draws on a general idea of synergy as has been proposed in biology. We validate this measure by providing evidence for its ability to predict the semantic transparency of linguistic compounds (Experiment 1) and the abstractness rating of nouns (Experiment 2).

## 1. Introduction

One interesting characteristic of some complex systems is the formation of macro level constructions perceived as having features that cannot be reduced to their micro level constituents. For example, we perceive water as “wet” while “wetness” does not characterize either the hydrogen or the oxygen molecules from which water is composed. The same emerging behavior is evident in natural language, where the meaning of word compounds, such as Hotdog, cannot be trivially reduced to the meaning of their constituents or/and their compositionality. This characteristic may be considered to be the expression of *synergy* where the joint action of the constituents produces unique features that are irreducible to the constituents isolated behavior or their simple composition.

Various measures of synergy have been developed in the natural sciences (e.g., [1–3]) (for a comprehensive review, see [4]). Most of them exclusively rely on the concept of mutual information, as epitomized, for instance, by the information decomposition frameworks developed by [3]. Nevertheless, there is no one agreed measure of synergy, and it is nontrivial

to apply the various measures of synergy, which rely on mutual information, to the “semantic” context. In other words, the synergy, characterizing complex systems, has been well acknowledged but difficult to conceptualize and quantify in the context of computing the emerging *meaning* of various linguistic and conceptual constructs. We are not familiar with any paper in which the synergy of semantic constructs, here described as “semantic synergy,” has been scientifically measured.

In this paper, we propose a novel measure/procedure for quantifying semantic synergy. This measure draws on a general idea of synergy as has been proposed in biology. We validate this measure by providing evidence for its ability to predict the semantic transparency of linguistic compounds (Experiment 1) and the abstractness rating of nouns (Experiment 2). For clarity, we first explain and illustrate the meaning of semantic synergy and then introduce the measure and test it. As the measure is of relevance for people from social sciences and humanities, who do not necessarily have the background of information theory, efforts have been made to provide a clear exposition of the main idea through several worked-out examples.

## 2. Semantic Synergy and Word Compounds

For introducing the idea of semantic synergy, we use the example of linguistic compounds. Linguistic compounds are formed when two or more words, or more accurately “lexemes,” are joined to produce a new word. Compound words are not marginal linguistic elements of the lexicon, and it is argued that compounding is one of the first processes that accompanied the emergence of language [5, p. 1].

The extent in which the constituents contribute to the meaning of the compound is discussed under the title of *semantic transparency* [6–9]. For instance, Seafood is a highly transparent compound as both Sea and Food contribute to its meaning. After all, Seafood is food that we get from the sea. In contrast, the word Dog seems to contribute a little to understanding the meaning of the word compound Hotdog unless one understands the visual similarity between the sausage denoted by “Hotdog” and a dachshund, an understanding which is grounded in a concrete cultural knowledge. The example of Hotdog is a clear illustration of semantic synergy as the meaning of this compound cannot be reduced either to Hot or to Dog or even to their simple composition. An intelligent alien trying to understand the meaning of Hotdog by inquiring the definitions of Hot and Dog in the Oxford English Dictionary would probably experience a failure. An anecdote may further illustrate this point. An old friend of the first author visited the UK in the early sixties and for the first time in his life saw a stand selling sausages under the sign of “Hotdogs.” He was shocked by the idea that, despite their great fondness for dogs, the British eat sausages made out of dog meat.

The semantic synergy evident in the case of Hotdog is also evident in cases in which the semantic transparency of the compound is much higher. The word compound Guidebook is rated among the top compounds in terms of its semantic transparency [10]. However, even in this relatively simple case, there is no trivial computational process to infer its meaning from the meaning of its constituents, as indicated by its relatively old age of acquisition which is approximately five. In this context, it must be clarified that the terms “semantic transparency” and “semantic synergy,” despite being conceptually related, are not conceptually confounded. Semantic transparency is a token instantiating the general mechanism of semantic synergy. However, semantic transparency cannot be equated with semantic synergy. The extent in which the meaning of a compound can be derived from its constituents is not the same as the extent in which the information encapsulated in the compound provides us with an added value beyond the information provided by the constituents. As we show in Experiment 1, the correlation between semantic transparency and our proposed measure of semantic synergy is significant but far from perfect, as should have been the case if the two concepts would have been identical.

There has been intensive work in understanding the cognitive processes underlying the comprehension of linguistic compounds. In addition, there have been some notable attempts to build computational models of “semantic

composition” (e.g., [11]). However, understanding the cognitive processes underlying the comprehension of compounds and developing workable computational models of these processes are far from being solved issues and totally beyond the scope of the current paper. We use the contexts of word compounds and words’ level of abstractness only for validating our measure of semantic synergy. Hence, our main aim is the development and validation of the new measure.

In this paper, we first use the case of word compounds in order to test our measure of semantic synergy. While linguistic and computational models of compounds have shown the contribution of the constituents to the semantic transparency of the compound (e.g., [6, 12]), it is clear from psychological studies [7] that the meaning of the compound is not simply constructed from its parts. Therefore, the semantic transparency of word compounds may serve as a test case for validating our measure of semantic synergy. However, and we repeatedly emphasize this point, the paper has no pretensions to model the cognitive processing of compounds or to build computational models for understanding compounds, although it probably has some relevance for both of these challenges.

## 3. Meaning and Distributional Semantics

Previously, we have mentioned the fact that synergy measures heavily rely on the idea of mutual information. Mutual information-based measures of synergy are not required to deal with “meaning” as they focus on the reduction of uncertainty only. However, any attempt to develop a measure of *semantic* synergy should explain the specific sense of meaning on which it relies and the way this sense is embedded in the measurement process. Here, we adopt the idea of “distributional semantics” ([13, 14]) in order to clarify how meaning can be represented. “Distributional semantics is based on the hypothesis that words that occur in similar contexts tend to have similar meanings (Harris, 1954; Firth, 1957). This hypothesis leads naturally to vector space models, in which words are represented by context vectors (Turney & Pantel, 2010)” [14, p. 1]. In this context, we may consider the meaning of Hotdog, for instance, by using the words collocated with it in a large corpus of the English language. For instance, searching the Corpus of Contemporary American English (abbreviated as COCA) [15] for the words collocated up to 4 positions to the right/left of Hotdog, we identify collocations such as Bun, Ketchup, and Mustard. We may organize these words as a *basis* for a “context vector” in which the values are the collocations’ frequencies or probabilities. This vector may be used for representing the meaning of Hotdog for various practical applications and indeed this simple idea of distributional semantics has been proven to be extremely powerful in the context of natural language processing and computational cognition. Given that the meaning of Hotdog and its constituents (i.e., Hot and Dog) may be represented by using their context vectors, how can we measure the synergy of a word compound?

TABLE 1: The context vectors of Seafood, Sea, and Food.

	Restaurant	Crab	Shrimp	Service	Wine	Water	Salt
Seafood	.70	.05	.12	.02	.04	.05	.02
Sea	.09	0	0	.22	.17	.49	.03
Food	.01	.01	.02	.02	0	.30	.64

#### 4. Measuring Semantic Synergy

In the context of genes interaction, [1] proposed that the synergy between  $G1$  and  $G2$  with respect to  $C$  may be formalized as follows:

$$\text{Syn}(G1, G2; C) = I(G1, G2; C) - [I(G1; C) + I(G2; C)], \quad (1)$$

where “ $I$ ” signifies the mutual information between the constituents. When applied to the context of compounds, the idea behind this formulation is appealing in its simplicity; synergy is what remained after we subtract the sum of each constituent’s unique information about the compound from the constituents’ joined contribution. Analogically, we suggest that the semantic synergy of words  $W1$  and  $W2$ , with respect to a word compound  $W1W2$ , may be conceptualized as the information gained when using the unique context vectors of  $W1$  and  $W2$  for approximating the context vector of  $W1W2$  minus the information gained when approximating  $W1W2$  by the simple addition of the information gained by  $W1$ ’s approximation of  $W1W2$  and the information gained when using  $W2$  for approximating  $W1W2$ . This idea may be better clarified by a toy example in which we apply the *Kullback-Leibler Divergence* measure.

Let us assume that we would like to measure the semantic synergy of the word compound Seafood. We search COCA for the collocations of Seafood, Sea, and Food and group these collocations into a bag of words that serves as the basis of our vector. This bag of words may include the following words: Crab, Salt, Service, Shrimp Restaurant, Water, and Wine.

Next, we use the above bag of words to form a shared *basis* for the context vectors of Seafood, Sea, and Food. We load our vectors with values that indicate the probably of each word to appear in the context of the target word (i.e., Seafood, Sea, and Food). See Table 1.

At this point, we may want to approximate the distribution of Seafood (i.e., its meaning) by using the distributions of Sea and Food. According to the distributional hypothesis, the meaning of Seafood can be represented by using its context vector. Therefore, approximating the distribution of Seafood by using the distribution of Sea and Food may indicate how much information we have gained when revising our beliefs from the prior distributions (i.e., senses) of Sea and food, to the a posteriori distribution of Seafood. If the meaning of Seafood is a synergetic product of its constituents, then approximating the meaning of Seafood, as represented by its context vector, through the senses of its constituents, as represented by their context vectors, should involve some gain in information as the information represented in Seafood cannot be fully approximated through

the information encapsulated in Sea and Food. The higher the gain, the higher the synergy. Here, we may use the Kullback-Leibler Divergence which is an *asymmetric* measure of the difference between two probability distributions  $P$  and  $Q$ :

$$D_{KL}(P \parallel Q) = \sum P(i) \log \frac{P(i)}{Q(i)}. \quad (2)$$

We may use  $D_{KL}$  to approximate the distribution of the word compound Seafood by using each of its constituent’s distributions. Here is an elaborated example showing how we may approximate the distribution of Seafood through Sea and Food.

Table 1 shows us that the probability of Restaurant to be collocated with Seafood is 0.70 and that its probability to be collocated with Sea is 0.09. Therefore, we first calculate

$$0.70 * \log \frac{0.70}{0.09}. \quad (3)$$

Next, we see that the probability of the word Crab to be collocated with Seafood is 0.05 and with Sea 0 (which is converted to 0.001). Therefore, we calculate

$$0.05 * \log \frac{0.05}{0.001} \quad (4)$$

and so on. We sum these expressions as follows:  $D_{KL}(\text{Seafood} \parallel \text{Sea}) = \sum 0.70 * \log(0.70/0.09), \dots, 0.02 * \log(0.02/0.30)$ .

Similarly, we calculate the divergence measure for Food and Seafood.  $D_{KL}(\text{Seafood} \parallel \text{Sea}) = 2.85$ , and  $D_{KL}(\text{Seafood} \parallel \text{Food}) = 4.70$ . These results suggest that it is easier to approximate the context vector of Seafood by using Sea rather than by using Food.

### 5. The Experiments

#### 5.1. Experiment 1

**5.1.1. Methods.** In the sections below, we introduce our dataset and the procedure we have applied for measuring semantic synergy.

**5.1.2. Dataset.** For the first experiment, we have used the dataset provided by [10]. This dataset includes 629 linguistic compounds and their rating by human judges according to several relevant measures such as familiarity (FAM), age of acquisition (AoA), semantic transparency (TRANS), and imageability (IMG). Semantic transparency concerns the extent in which the meaning of the compound can be inferred from its constituents. Imageability concerns the extent in which the compound may be imagined. As TRANS may be considered to be the expression of semantic synergy, it is the focus of the first experiment.

**5.1.3. Procedure.** We symbolize each word compound as  $W1W2$ , the left word as  $W1$ , and the right word as  $W2$ . For each word, we searched the COCA for collocations that exhibit up to 4 positions to the right/left of our target word.

Out of this list, we filter up to the 300 most frequent words in the forms of lemmas that (1) belong only to the parts of speech of Noun, Verb, and Adjective and have (2) mutual information  $\geq 3$  with our target word ( $W1$ ,  $W2$ , and  $W1W2$ ). The second constraint aims to filter out highly frequent and noisy collocations. Each list is stored as a bag of words and the bags of words of  $W1$ ,  $W2$ , and  $W1W2$  are united into a single list of unique words (i.e., lemmas) that can have a maximal cardinality of 900. From this list, we remove  $W1$ ,  $W2$ , and  $W1W2$ . The list is organized as an alphabetical list of unique words that form the basis of the context vector. This basis is used to describe  $W1$ ,  $W2$ , and  $W1W2$ . Next, we construct the specific vectors for  $W1$ ,  $W2$ , and  $W1W2$  by loading the basis with the frequencies of each lemma and converting the frequencies into probabilities. We define three sets of words and four vectors:

- (1)  $W1W2$ : the vector comprised of the probabilities of the lemmas collocated with  $W1W2$
- (2)  $set\_W1$ :  $(W1 \cap W1W2) - (W2 \cap W1W2)$

This means that, first, we identify the words at the intersection of  $W1$  and  $W1W2$  and sum their frequencies. From this list of words, we “remove” the words that also exist at  $W2 \cap W1W2$ . The unique words remaining after this process has been completed are identified in the basis and loaded with values to form a vector which is titled *Unique\_W1*.

- (3)  $set\_W2$ :  $(W2 \cap W1W2) - (W1 \cap W1W2)$

It is calculated the same as  $set\_W1$  but this time we identify words that exist at  $W2 \cap W1W2$  but do not exist for  $W1 \cap W1W2$ . Again, the final vector, titled *Unique\_W2*, is comprised of the same 900 words that form the basis, but we have values different from 0 only for words that exist at  $W2 \cap W1W2$  but do not exist in  $W1 \cap W1W2$ .

- (4)  $set\_Joint = W1 \cap W2 \cap W1W2$

This time, we identify the words that exist at the intersection of the three words (i.e.,  $W1 \cap W2 \cap W1W2$ ) and the vector, titled *Joint*, includes the 900 words but only the words included in  $set\_Joint$  have values different from 0 which is the sum of their frequencies at  $W1$ ,  $W2$ , and  $W1W2$  as converted to probabilities.

Each vector is converted into a vector of probabilities where a value of 0 is transformed to 0.001. In the next step, we produce three measures that are based on a “softer” version of the Kullback-Leibler Divergence. These are measures of information *gain*, as they give us an indication about the information we have gained when revising our beliefs from the prior distributions of the context vectors of the constituents, or their shared semantic space, to the context vector of the compound. The general structure of the measures is very simple:

$$\sum_{i=1}^N \frac{p(P_i)}{p(Q_i)}, \quad (5)$$

where  $P$  is always  $W1W2$  and  $Q$  is either *Unique\_W1*, *Unique\_W2*, or *Joint*.

TABLE 2: Pearson correlations between the measures of the linguistic compounds.

	FAM	AoA	TRANS	IMG
FAM		-.672	<b>.288</b>	.495
AoA			-.397	-.642
TRANS				<b>.396</b>

- (1) The first measure is titled *Gain\_W1*, where  $P = W1W2$  and  $Q = W1$ . In this case,  $W1$  is the unique vector of  $W1$  as we calculated before.
- (2) The second measure is titled *Gain\_W2*, where  $P = W1W2$  and  $Q = W2$ . In this case,  $W2$  is the unique vector of  $W2$  as we calculated before.
- (3) The third measure is titled *Gain\_Joint*, where  $P = W1W2$  and  $Q = Joint$ .

The semantic synergy measure of a word compound is calculated as follows:

- (5)  $SemSyn = (Gain\_Joint * |set\_Joint|) - (GainW1 * |set\_W1| + GainW2 * |set\_W2|)$ .

Semantic synergy is thus defined as

- (1) the information gained when trying to approximate the distribution of the compound using the prior distribution of the elements shared by  $W1$ ,  $W2$ , and  $W1W2$  and multiplied by the cardinality of  $set\_Joint$ , minus
- (2) the sum of the information gain from the *unique* prior distribution of  $W1$  multiplied by the cardinality of  $set\_W1$  plus the gain from the *unique* prior distribution of  $W2$  multiplied by the cardinality of  $set\_W2$ .

We have not used the Kullback-Leibler Divergence as its component of log function condenses differences we would like to use in order to identify the synergetic effect. When experimenting with a version of the above function using  $D_{KL}$ , the results were slightly inferior. In addition, we multiplied the information gain by the cardinality of each relevant set, as the number of words that exist at the intersecting semantic fields of the words was found to be associated with the semantic transparency of the compound, a finding which is in line with the literature indicating the role of the constituents in determining the transparency of the compound [12].

**5.1.4. Analysis and Results.** Table 2 presents the Pearson correlations between the various measures of the linguistic compounds ( $N = 616$ ). Only results significant at  $p < .001$  are reported and the correlations with TRANS are emphasized.

**Hypothesis 1.** We hypothesized that if our measure of semantic synergy (i.e., *SemSyn*) is valid, then a negative correlation should be expected between *SemSyn* and TRANS (Hypothesis 1 a) as the more synergetic the word compound is, the less transparent it is semantically, the more the time it takes



TABLE 3: Pearson correlations between *SemSyn* and the measures of the linguistic compounds.

	FAM	AoA	TRANS	IMG
<i>SemSyn</i>	-.397	.439	-.400	-.483

to learn it (higher age of acquisition) (Hypothesis 1 b), the less familiar it is (Hypothesis 1 c), and the less imaginable it is (Hypothesis 1 d). Table 3 presents the correlations of our synergy measure with the above measures.

All results are statistically significant at  $p < .001$ . One may wonder whether the correlation between TRANS and *SemSyn* is influenced by the frequency of *W1* and *W2* in our corpus. To address this question, we measured the correlation again, this time by controlling for the frequency of *W1* and *W2*. The resulting Pearson correlation was only slightly lower than the one gained before ( $r = -.392$ ) indicating that the frequency of the constituents does not have a major impact on the correlation between TRANS and *SemSyn*.

We can see that our first hypothesis with its various variants has been supported. It is interesting to notice that, within the measures of the compounds, the highest correlation TRANS has is with AoA (-.397) and IMG (.394). This means that the higher the semantic transparency of the compound is, the earlier it is learned by children and the more imaginable it is. In this context, it is interesting to see in Table 3 that the semantic synergy measure was correlated with TRANS, to the *same degree* that TRANS was correlated with AoA and IMG. This result indicates that the new measure of semantic synergy may predict semantic transparency to the same level as the compound's age of acquisition and its degree of imageability.

Another way of testing the major research hypothesis (i.e., Hypothesis 1 a) is by comparing the synergy scores of compounds rated high or low on the semantic transparency measure. In this case, we apply the *extreme groups research design*. We have identified the top 25 percent of the compounds that scored the highest on TRANS (H) and compared them to the 25 percent of the compound that scored the lowest (L). Using one-way ANOVA, the difference between the two groups was found to be statistically significant ( $F(1, 299) = 85.82, p < .001$ ) with Partial Eta Squared = 0.223, which according to Cohen's norms in the behavioral sciences is considered to be a *large* effect size. As expected, compounds that were less semantically transparent (L) scored on average higher on the *SemSyn* measure (-250 versus -541, resp.).

*Hypothesis 2.* The fact that *SemSyn* was found to be linearly correlated with TRANS is not an indication of its ability to successfully classify compounds as transparent or not. Using *SemSyn* in a classification task, by applying Machine Learning procedures, may further support its validity. Therefore, we have also tested the validity of the semantic synergy measure in a classification task, where *SemSyn* was used as the only feature for classifying the compounds as L or H on semantic transparency.

TABLE 4: Results of the classification procedures (rounded percentages).

	Precision	Recall	Accuracy
CRT	72	87	76
K-NN	69	78	72

\* All results are in round percentages and those involving significance tests are significant at  $p < .001$ .

We have hypothesized that *SemSyn* will provide us with a significant increase in prediction of which compound has been rated low on transparency. This significant increase in prediction is judged by comparing the probability that a compound is rated low on transparency given that the classifier predicted it as such (i.e., the precision of the classifier). The precision is compared to the prediction we may gain using the base rate of L cases in our dataset. In our dataset, 50% of the compounds are tagged as L and therefore any measure of precision which is higher than 50% may be considered significant. We used two Machine Learning classification procedures:

- (1) The Classification and Regression Tree (CRT) model with a tenfold cross validation procedure
- (2) The *k*-nearest neighbors algorithm (K-NN) with an Euclidean metric for distance computation and tenfold cross validation procedure

The results are presented in Table 4 where following the norms of Machine Learning and natural language processing, the precision, recall, and accuracy of the classifier are used to evaluate its performance.

Precision is the true percentage of L words out of the words identified as L cases. Recall is the percentage of cases of the classifier identified as L out of all L cases in our dataset. Accuracy is the overall measure of the classifier's correct identification of L cases (i.e., compounds that were correctly identified by the classifier as characterized by low semantic transparency) and correct identification of non-L cases, out of the total cases. The measures of precision and recall refer to the success in predicting compounds with low transparency. We can see that both classifiers produced a significant improvement in prediction over the base rate (50%) with an average improvement of 20%. These results further support the validity of our measure.

In the first experiment, we have validated our measure of semantic synergy in the context of word compounds. It is highly important to emphasize the fact that we have used the context of word compounds to validate our measure and have no intentions whatsoever to compete with algorithms that aim to predict various aspects of word compounds such as semantics transparency. We have mentioned some of these attempts (e.g., [11]) but in this paper have a totally different aim. Nevertheless, it is interesting to examine the predictive value of our semantic synergy measure with regard to the predictive power of measures of semantic distance. It is reasonable to hypothesize that the semantic transparency of a compound can be predicted based on

the semantic distance between the compound and its constituents. Whether the semantic synergy measure has any contribution beyond the predictive value of these semantic distances is an open question. To answer this question, we used a vector space model of semantics [14] by specifically using the *term-to-context* matrix developed by [16]. We prefer this matrix over term-to-document LSA models as it preserves the common sensical meaning of similarity in terms of term-to-lexical-context association and since it has been successfully used in various studies of cognitive and social computing. Using the term-to-document matrix, we have measured the semantic distance between each word compound and its first constituent and titled this new measure *w1\_compound*. Similarly, we have measured the semantic distance between the compound and the second word and titled the new measure *w2\_compound*. To recall, the Pearson correlation between our synergy measure and the semantic transparency of the compound was  $-0.40$ . The Pearson correlation between *w1\_compound* and TRANS was  $r = .325$  ( $p < .001$ ) and between *w2\_compound* and TRANS was  $r = .299$  ( $p < .001$ ). SemSyn was negatively correlated with *w1\_compound* ( $r = -.496$ ) and *w2\_compound* ( $r = -.547$ ,  $p < .001$ ), meaning that the higher that semantic synergy of the constituents, the lower the distance between their vectors and the vector of the word compound, as trivially expected. A better way of measuring the relative contribution of *w1\_compound*, *w2\_compound*, and SemSyn for predicting the semantic transparency of the compound is by repeating the CRT analysis and classifying the compounds into low transparency compounds and high transparency compounds. Using the CRT classifier with tenfold cross validation and *w1\_compound* and *w2\_compound* as features gave us better precision than the one we have gained by using only SemSyn (76% versus 72%, resp.), but a lower recall (69% versus 87%, resp.) and slightly lower accuracy (74% versus 76%, resp.). When entering *w1\_compound*, *w2\_compound*, and SemSyn into the model, we have gained the highest precision (82%) and accuracy (77%) but a slightly lower recall (71%). The normalized importance of the features shows that SemSyn had the highest normalized importance in the model (100%) followed by *w1\_compound* (89%) and *w2\_compound* (40%). Again, designing the best algorithm for predicting semantic transparency is totally beyond the scope of the current paper. However, the supplementary analysis we have conducted proves that our semantic synergy measure may contribute to the prediction of semantic transparency beyond the contribution of the semantic distance of the constituents, as measured by a powerful model.

Another complementary analysis to the main research question may involve the role of Lexeme Meaning Dominance [10]. Lexeme Meaning Dominance (LMD) measures the relative dominance of the first/second lexeme in determining the meaning of the entire compound. In our dataset, LMD is measured on a scale ranging from 0 (i.e., the meaning of the entire compound is in the first lexeme) to 10 (i.e., the meaning of the entire compound is determined by the second lexeme). In [10], compounds that scored 4 or lower on the LMD measured were titled as “Headed” and those that scored 6 or higher were titled as “Tailed.” In our dataset, 139

compounds have been identified as “Headed” (45%) and the rest, 167, have been identified as “Tailed” (55%), indicating the dominance of the second lexeme.

To recall, we have calculated the information gained in trying to approximate the vector of the compound through the vector of the first lexeme (i.e., Gain\_W1) and the second lexeme (i.e., Gain\_W2). The higher the Gain score is, the more difficult it is to approximate the meaning of the compound using the vector of the constituent. LMD can be interpreted by using these measures, as the relative dominance of the first and second lexeme should be expressed in terms of the information gained when trying to approximate the meaning of the compound by using the vector of each constituent. Therefore, we may hypothesize that significant differences will be found when comparing Headed and Tailed compounds on the gain measures.

Using MANOVA with LMD as the factor (i.e., Headed versus Tailed compounds) and the two gain measures as the independent variables, a statistically significant difference was found between the groups ( $F(2, 303) = 9.82$ ,  $p < .001$ ). However, the difference was statistically significant only for the Gain\_W2 measure ( $p < .001$ ) where Tailed compounds scored lower ( $M = 1674$  versus  $M = 2038$ , resp.). This means that Tailed compounds, where the second lexeme determines the whole meaning of the compound, are compounds in which the context vector of the second lexeme is much “closer” to the meaning of the context vector of the whole compound. Interestingly, there was no symmetry with regard to the information gain of the first lexeme, indicating that lexical semantic dominance is not symmetrically and trivially associated with the information provided by the first and the second lexeme.

**5.2. Experiment 2.** In the first experiment, we have used word compounds in order to validate our measure. We hypothesized that if the semantic synergy score of the compound is valid, then a significant negative correlation should be found between the semantic synergy score and the semantic transparency score. The results provide some empirical support for the validity of our measure. However, validating the measure cannot be exhausted by a single case as semantic synergy does not have to be necessarily expressed in semantic compounds. Another test case involves the abstractness/concreteness rating of words as explained in the next section.

As suggested by [17], the meaning of words may be attributed to two dimensions: experiential and distributional. The experiential dimension concerns the perceptual aspect of the meaning. For instance, the meaning of a Cherry is to a large extent determined by the perceptual aspects of cherries, being red, small, round, sweet, and so forth. As can be immediately comprehended, the meaning of concrete words seems to lean more heavily on their perceptual experiential dimension or the perceptual dimension of words collocated with them. Identifying the collocations of Cherry in COCA and grouping them in mind, we may easily identify that they are organized around the themes of Nature (e.g., Tree) and Food (e.g., Pie).

The distributional dimension of meaning concerns the way in which the meaning of a word is derived from its connections with other words. For instance, the word Democracy denotes a concept that has no reference to a perceptible entity. The meaning of Democracy is exclusively determined by its connections with other words that determine its meaning as a specific form of government. As we can see, the distributional dimension of meaning is deeply connected with the abstractness level of a word. The meaning of words that are more abstract relies more heavily on the distributional dimension. In this context, we may hypothesize that the concreteness/abstractness rating of a word should be correlated with its score of semantic synergy as the meaning on an abstract word cannot be trivially reduced to any perceptual entity or to the meaning of other words through which it is defined; while the meaning of Democracy may be a synergetic product of other words defining its semantic network, the meaning of a Cherry is probably less synergetic as it relies on a simple sum of its perceptual constituents (e.g., color, shape, size, and taste) or the meaning of words to which it is linked in the semantic network (e.g., apple, strawberry). Therefore, we have hypothesized that the abstractness of a word will be positively correlated with our measure of semantic synergy.

**5.2.1. Dataset.** We have used a dataset that includes the concreteness ratings of 37,058 English words, obtained from over 4,000 participants [18]. Each participant was asked to rate the word using a 5-point rating scale ranging from abstract to concrete. We have identified the nouns in the dataset and selected the top 150 concrete words and the top 150 abstract words.

**5.2.2. Procedure.** Let us denote each of the target words we have analyzed ( $N = 300$ ) as Abs. We have first identified the collocations of each target word according to the same procedure used in Experiment 1. Next, we selected the two top rated collocations (nouns only), denoted as  $W1$  and  $W2$ , and recursively identified their own collocations. From this point on, the procedure was exactly the same as the one used in Experiment 1, but with Abs as analogously standing for  $W1W2$ . The idea was that, according to the distributional dimension of meaning, updating our beliefs from the prior context vectors of words that form the semantic context of the target, to the a posteriori context vector of the target, requires more efforts. Hence, abstractness will be positively correlated with semantic synergy.

**5.2.3. Analysis and Results.** There were  $N = 296$  words in our dataset, 149 of them denoting the most concrete nouns in the database (51%) and the rest the most abstract words (49%). The Pearson correlation between abstractness and *SemSyn* was found to be statistically significant ( $r = .537$ ,  $p < .001$ ), where the sign of *SemSyn* has been turned to negative in order to provide a simple interpretation of the results where the abstraction ranking starts with the concrete to abstract. Linear Regression was calculated to predict abstraction level

TABLE 5: Results of the classification procedures (rounded percentages).

	Precision	Recall	Accuracy
CRT	72	85	76
K-NN	70	76	71

of a word based on its synergy score. A statistically significant regression equation was found ( $F(1, 292) = 118.23$ ,  $p < .001$ ), with an  $R^2$  of 0.288. In an attempt to examine the predictive power of *SemSyn*, we have applied the two Machine Learning classification procedures. The results are presented in Table 5.

Given that the base rate of abstract words in our dataset is 49%, the average precision is 71% which means 22% improvement in prediction. These results are indicative of the predictive power of synergy and hence provide another layer of empirical support for its validity.

## 6. Conclusions

The famous Gestalt slogan “the whole is different from the sum of its parts” is a clear indication of the way emerging structures bear the fingerprint of synergetic processes. It must be noted that this synergy is difficult, at least theoretically, to capture through measures relying on mutual information. Synergy involves a shift in scales of analysis while the measure of mutual information does not directly address this change in scaling. This theoretical point may be explained through the physics of computation and Landauer’s principle [19]. Reference [19] argued that whenever a system erases some information then this process is irreversible and is accompanied by a minimal price of entropy which is released to the environment. This price is evident in a process of computation where a certain output is produced from certain inputs, such as in the case of the formation of abstract semantic categories from particular instances [20]. For instance, the computation of Hotdog from Hot and Dog requires that some information about the constituents will be lost when the new structure of Hotdog is computed. In other words, the formation of the compound requires that some information is lost and some information is gained when shifting between different scales of analysis in a process of irreversible natural computation. In this context, a mutual information-based measure of synergy cannot be used for directly representing the *gain* and *loss* accompanying the emergence of semantic structures.

In this paper, we have made a first step in developing a measure of semantic synergy. This measure, which takes into account the information gain (and loss) accompanying the shift between context vectors, preserves the idea of information gain/loss that accompanies a synergetic process as well as the semantic representations on which the vectors operate in the context of semantic synergy.

This measure of semantic synergy has various applications beyond the specific and limited contexts in which it has been validated in this specific study. For instance, one may be interested in studying the way in which the meaning of

certain concepts (e.g., God, Love, and Darwinism) has been changed through history. One possible way of addressing this challenge is by tracing the changes undergone by the semantic fields of these concepts as instantiated by their context vectors. However, quantifying this change is far from trivial. One possible approach derived from our measure of semantic synergy is to trace the way in which the semantic synergy of a concept has been changed. Analyzing the trajectory of this measure may be used for identifying “tipping points” in the evolution of the concept and historical landmarks where its meaning has been transformed. This idea and others are probably worth developing by researchers to whom the concept of semantic synergy may be of value.

## Competing Interests

The authors declare that they have no competing interests.

## References

- [1] D. Anastassiou, “Computational analysis of the synergy among multiple interacting genes,” *Molecular Systems Biology*, vol. 3, no. 1, article 83, 2007.
- [2] V. Griffith, *Quantifying synergistic information [Ph.D. dissertation]*, California Institute of Technology, 2014.
- [3] P. L. Williams and R. D. Beer, “Nonnegative decomposition of multivariate information,” <https://arxiv.org/abs/1004.2515>.
- [4] N. Timme, W. Alford, B. Flecker, and J. M. Beggs, “Synergy, redundancy, and multivariate information measures: an experimentalist’s perspective,” *Journal of Computational Neuroscience*, vol. 36, no. 2, pp. 119–140, 2014.
- [5] C. Semenza and C. Luzzatti, “Combining words in the brain: the processing of compound words. Introduction to the special issue,” *Cognitive Neuropsychology*, vol. 31, no. 1-2, pp. 1–7, 2014.
- [6] M. J. Bell and M. Schäfer, “Modelling semantic transparency,” *Morphology*, vol. 26, no. 2, pp. 157–199, 2016.
- [7] S. Frisson, E. Niswander-Klement, and A. Pollatsek, “The role of semantic transparency in the processing of English compound words,” *British Journal of Psychology*, vol. 99, no. 1, pp. 87–107, 2008.
- [8] G. Libben, “Semantic transparency in the processing of compounds: consequences for representation, processing, and impairment,” *Brain and Language*, vol. 61, no. 1, pp. 30–44, 1998.
- [9] G. Libben, M. Gibson, Y. B. Yoon, and D. Sandra, “Compound fracture: the role of semantic transparency and morphological headedness,” *Brain and Language*, vol. 84, no. 1, pp. 50–64, 2003.
- [10] B. J. Juhasz, Y.-H. Lai, and M. L. Woodcock, “A database of 629 English compound words: ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience,” *Behavior Research Methods*, vol. 47, no. 4, pp. 1004–1019, 2015.
- [11] P. D. Turney, “Semantic composition and decomposition: from recognition to generation,” <https://arxiv.org/abs/1405.7908>.
- [12] S. Reddy, D. McCarthy, and S. Manandhar, “An empirical study on compositionality in compound nouns,” in *Proceedings of the 5th International Conference on Natural Language Processing*, pp. 210–218, Chiang Mai, Thailand, November 2011.
- [13] S. Clark, *Vector Space Models of Lexical Meaning. Handbook of Contemporary Semantics*, Wiley-Blackwell, New York, NY, USA, 2nd edition, 2012.
- [14] P. D. Turney and P. Pantel, “From frequency to meaning: vector space models of semantics,” *Journal of Artificial Intelligence Research*, vol. 37, pp. 141–188, 2010.
- [15] M. Davies, “The 385+ million word Corpus of Contemporary American English (1990–2008+): design, architecture, and linguistic insights,” *International Journal of Corpus Linguistics*, vol. 14, no. 2, pp. 159–190, 2009.
- [16] P. D. Turney, Y. Neuman, D. Assaf, and Y. Cohen, “Literal and metaphorical sense identification through concrete and abstract context,” in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP ’11)*, pp. 680–690, Edinburgh, Scotland, July 2011.
- [17] M. Andrews, G. Vigliocco, and D. Vinson, “Integrating experiential and distributional data to learn semantic representations,” *Psychological Review*, vol. 116, no. 3, pp. 463–498, 2009.
- [18] M. Brysbaert, A. B. Warriner, and V. Kuperman, “Concreteness ratings for 40 thousand generally known English word lemmas,” *Behavior Research Methods*, vol. 46, no. 3, pp. 904–911, 2014.
- [19] C. H. Bennett and R. Landauer, “The fundamental physical limits of computation,” *Scientific American*, vol. 253, no. 1, pp. 48–56, 1985.
- [20] B. Tamir and Y. Neuman, *The Physics of Categorization*, Complexity, 2015.



