Linguistic Duality

Rintaro Nakahodo
Department of Data Science
Musashino University
Tokyo, Japan
s1922069@stu.musashino-u.ac.jp

Virach Sornlertlamvanich
Department of Data Science
Musashino University
Tokyo, Japan
virach@musashino-u.ac.jp

Faculty of Engineering, Thammasat University
Bangkok, Thailand

Abstract—**Currently, the approaches from the results of word or document vectorization significantly achieve their success in large-scaled language processing, such as high-quality neural network machine translation trained on a large-scale parallel corpus. The data driven approach with minimum concern in language knowledge efficiently captures its syntactic information. However, there are some deficiency occurred in capturing the semantic information, such as in the similarity measure when performs on both sides of the language interpretation. We propose an add-on process to correct the similarity measure of the concepts defined in WordNet. Similarity measure is the fundamental procedure for language processing highly depending on the context integration and its representation method.**

Keywords—language duality, syntax, semantics, Word2Vec, WordNet, similarity measure

I. Introduction

In recent years, the processing power for natural language processing is being increased along with the progress in machine learning research. As a result, the language models can be generated with more context to solve linguistic problems in many tasks such as word disambiguation in machine translation, question-answering in chatbot, text classification in document recommendation. Word2Vec and WordNet are the current representative methods for word representation in terms of word context vectorization and hierarchical representation of sets of word synonym. They become major word representations for similarity measure. Word2Vec has a disadvantage due to the nature of bag-of-words computation. It does not take the order of word occurrence and the word polysemy into account in generating the word representation. On the other hand, the WordNet approach uses a set of synonyms (so called synset) to define the concept of a word. The synsets are then characterized in a form hierarchy by assigning an "is-a" relation between synsets. As a result, the synsets with positive and negative senses are assigned in the same hierarchical structure. It is indistinguishable when determining the concept similarity by simply measuring the distance between synsets. Therefore, the results of similarity measure in Word2Vec and WordNet are different. In this paper, we discuss the meaning of the similarity estimation between words from the perspective of linguistic duality in terms of these the two word representation approaches.

The remainder of this paper is organized as follows: Section 2 explains the related works, and Section 3 discusses the observation in linguistic duality. In Section 4, we discuss the pragmatic approach in word representation in terms of similarity measure between Word2Vec and WordNet. Section 5 discusses the problems of the current WordNet similarity measure and its solution is proposed. In Section 6, we extend our observation into the recent discussions on deep learning related to System 1 and System 2, and describe the future prospects of the discussions.

II. Related Works

Machine translation is an important task to challenge human ability in language usages. Research in the field of machine translation includes the work of Bonnie J. Dorr [1] who grounded that when one language is naturally translated into another language, the result has the property of being completely different from the original and diverging in translation. It can be divided into linguistically grounded classes underlying lexical semantic divergence, and techniques where lexical semantic divergence is resolved and developed a basis for evaluating the state of the system itself. Thus, it is necessary to study the semantics required for translation from a linguistic point of view in machine translation. In this paper, we elaborate the changes in the discussion of semantics and syntax, and then discuss what universal meaning is.

In a study on similarity measure, Yi Jiang, Shie-Jue Lee et al. [2] developed a similarity measure method between documents and evaluated it using several real-world data sets. It is reported that the proposed method performed better than others. How similarity should be measured is not a general one, but rather a limited similarity for a few cases. Up to present, the development of a method for measuring the similarity has been continuously done based on several grounded information for the word representation. This paper does not focus on the study of the development of such similarity methods per se, but rather on the study of cosine similarity and semantic measure of similarity currently used in deep learning approaches.

978-1-7281-6694-0/20/$31.00 ©2020 IEEE

III. WHAT IS LINGUISTIC DUALITY?

What we mean by 'linguistic duality' here is that, as of the present time, there are two major type of language modeling. The first is one that computes and estimates the word representation by calculating its syntactic information using the properties of its role in the syntactic structures. The second is one that is specialized in semantic interpretation. In this Section, we elaborate the essential difference between "syntactic information" and "semantic information" and its representative power to verify its nature.

A. Syntactic information

In syntactic parsing, we generate all possible trees for an input sentence. A syntactic tree is an intermediate representation showing the grammatical relation between the constituents of a sentence, conforming to the rules of a formal grammar. As a result, the constituents are labeled by a part-of-speech and its grammatical role. There are many types of approaches which can be grouped into a top-down and bottom-up parsing strategy. A top-down parsing strategy finds left-most derivations of an input sentence by searching for parse trees using a top-down expansion of the given formal grammar rules. A bottom-up parsing strategy starts with the input and attempt to rewrite it to the start symbol of the given formal grammar rules. Both strategy parsing keeps its all possible parsed trees due to the ambiguity, which occurs during the interpretation of the part-of-speech of a word and its grammatical role.

In general, parsing a natural language sentence has many typical difficulties, such as the complexity of many ad hoc variants, the ambiguity of word meaning. In addition, relying on syntactic context is not enough in some cases to determine a single appropriate syntactic tree for a given sentence. In a formal language, a context-free grammar can be used while for a natural language, a more linguistic approach is needed. To avoid the effects of the noisy input stream, shallow parsing that does not specify the internal structure of the sentence, nor does it specify the role of each component in the sentence, is applied.

It is obvious that both parsed trees in Figure 1 are syntactically acceptable. In Japanese, there are many sentences with difficulties in determining the word dependency, such as "美しい水車小屋の少女" (beautiful, watermill, of, a girl) which is syntactically acceptable but difficult to understand. In the case of such sentence, it is common to judge the syntactic parsing by the dominant pragmatical interpretation. In this example, the sentence has different meaning depending on whether the adjective "美しい" (beautiful) is attached to "水車小屋" (watermill) to mean "a beautiful watermill of a girl", or to "少女" (girl) to mean "a beautiful watermill girl".

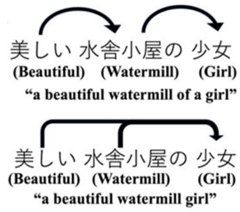

Figure 1. Ambiguity of dependency parsed trees in a Japanese sentence

B. Semantic information

Semantics, in contrasting to syntactic theory, is a study of combination of units of a language with consideration of their meanings. Theoretically, semantics has included the study of meaning and reference, truth conditions, argumentative structure, thematic roles, discourse analysis, and linking all of these to syntax.

In general, semantically there are ambiguities in natural language, as well as syntax. For example, according to Avram Noam Chomsky [3], there is a sentence "Colorless green ideas sleep furiously". Syntactically, words are put in a sequence conforming to the given grammar rules. In the following two sentences, given the same set of words, (1) is grammatically accepted, while (2) is a non-sentence.

(1) Colorless green ideas sleep furiously.

(2) *Furiously sleeping ideas green colorless.

However, sentence (1) is meaningless though it is grammatically correct.

Therefore, in this case, being a meaningful sentence, it must be determined semantically in addition to the syntactic soundness. The meanings are normally categorized into two main types

(1) The relationship that a symbol has to an object or situation.

(2) The relationship that a symbol has to other symbols, especially mental symbols called concepts.

The former is called the indicative meaning (reference) and the latter is called the intrinsic meaning (sense). In addition to these two types of meanings, discourse analysis makes it possible for semantics to show the hidden meanings. Traditionally, it reveals the truth conditions, term structures, subject roles that classify the semantic relations between predicates and terms. It also reveals the internal contradictions and accepted assumptions contained in a text. The power relations under what is regarded as various truths are constructed under the development of the discourse.

IV. DIFFERENCE BETWEEN WORD2VEC AND WORDNET

In this section, we discuss the characteristics of Word2Vec [4], where words are represented in a vector space, and WordNet [5, 6], where words are represented by a set of synonyms (so called synset) and labeled with a set of semantic relations, such as is-a, synonym-antonym, hypernym-hyponym, meronym-holonym, and troponymy. We consider the measure of similarity and its interpretation to compare between Word2Vec and WordNet based on experimental results.

A. Features of Word2Vec and WordNet

In Word2Vec, word is encoded as a vector in the word context space. The method is computed based on the distribution hypothesis. It acquires the distributed representations of words by using an inference-based method in which the weights are repeatedly updated over a small number of training samples using a neural network. There are two models proposed to generate the word vectors. In the CBOW model, the distributed representations of context (or surrounding words) are combined to predict the word in the middle. While in the Skip-gram model, the distributed representation of the input word is used to predict the context. As a result, Word2Vec calculates the word representation in the vector space. Therefore, when words are represented in a form of vector, the word vectors inherit the properties of vector in arithmetic calculation of cosine similarity, subtraction and addition. In the vector space, it is possible to show in a co-occurrence space and to measure the similarity of the word vectors. The cosine similarity can be obtained by Equation (1).

$$\cos(\vec{q},\vec{d}) = \frac{\vec{q}\cdot\vec{d}}{|\vec{q}||\vec{d}|} = \frac{\vec{q}}{|\vec{q}|}\cdot\frac{\vec{d}}{|\vec{d}|} = \frac{\sum_{i=1}^{|V|} q_i d_i}{\sqrt{\sum_{i=1}^{|V|} q_i^2}\cdot\sqrt{\sum_{i=1}^{|V|} d_i^2}} \quad (1)$$

The closeness between two vectors in the vector space does not infer the similarity in terms of their meaning. Instead, the similarity in terms of the context that the words may occur.

However, semantic relationships in pairing such as Man:Woman and King:Queen can be shared and paraphrased by the abstract set of all words that encodes the relationship of Male:Female. For example, The typical example of the arithmetic calculation of subtraction and addition is, king + woman - man = queen. The result of similar words are shown in Table I in the descendent order.

TABLE I. TOP 10 SIMILAR WORDS GENERATED BY USING WORD2VEC

king + woman – man	
Similar words	***Score of Word Similarity***
emperor	0.5768
Emperor	0.5687
queen	0.5471
shoot	0.5439
old	0.5384
Prince	0.5270
Queen	0.5243
Crown	0.5235
Hero	0.5232
fairy	0.5219

WordNet is a result of the effort to develop a word representation system according to the process model of human memory [7]. WordNet is a semantically rich English lexical database which is widely used as a lexical knowledge resource in many research and development areas, especially for solving semantic ambiguity problems. The database is grouped by part of speech into nouns, verbs, adjectives and adverbs. The lexicon is semantically organized in sets of synonyms, called 'synset' with a uniquely assigned ID, called 'synset ID'. Each synset represents the meaning of the word entry. Synsets are connected by several semantic relations, such as synonym-antonym, hypernym-hyponym, and meronym-holonym. It is successfully implemented in many applications, e.g., word sense disambiguation, information retrieval, text summarization, text categorization, etc. Inspired by the previous success cases, many languages have attempted to develop their own WordNets using WordNet as a model, for example, BalkaNet (Balkans languages), DanNet (Danish), EuroWordNet (European languages such as Spanish, Italian, German, French, English), Russnet (Russian), Hindi WordNet, Arabic WordNet, Chinese WordNet, Korean WordNet, Thai WordNet etc.

The main relation among words in WordNet is synonymy, as between the words shut and close or car and automobile. Synonyms--words that denote the same concept and are interchangeable in many contexts--are grouped into unordered sets (synsets). Each of WordNet’s 117 000 synsets is linked to other synsets by means of a small number of “conceptual relations.” Additionally, a synset contains a brief definition (“gloss”) and, in most cases, one or more short sentences illustrating the use of the synset members. Word forms with several distinct meanings are represented in as many distinct synsets [5, 6].

The calculation of word (synset) similarity in WordNet is based on its hierarchical is-a relation. One of the general paths based measures is proposed by Wu & Palmer [8]. The similarity measure takes the position of concepts c_1 and c_2 in the taxonomy relatively to the position of the most specific common concept $lso(c_1, c_2)$ into account. It assumes that the similarity between two concepts is the function of path length and depth in path-based measures. This function can be found in NLTK module *lcs* algorithm (longest common subsequence) [9] as shown in Equation (2),

$$sim_{w,p}(c_1,c_2) = \frac{2\cdot depth(lso(c_1,c_2))}{len(c_1,c_2)+2\cdot depth(lso(c_1,c_2)} \quad (2)$$

where, $lso(c_1, c_2)$ is the least common subsumer of c_1 and c_2, $len(c_1, c_2)$ is the length of the shortest path from synset c_i to synset c_j in WordNet, and $depth(c_i)$: the length of the path to synset c_i from the global root entity, and $depth(root)=1$.

B. Results of experiment

We perform word similarity tests between WordNet and Word2Vec. First, we perform two similarity estimation tests on adjectives and adverbs to test the differences between the two methods. Next, we investigate the relationship between word vectors in several types such as word inflection between “walk” and “walking”, which Word2Vec inherits the advantage of vector arithmetic calculation, comparing to WordNet which has a man-crafted synset for the word sense expression, and other semantic relations among the synsets. In this experiment, we adopt the

Word2Vec model generated from more than 4 GB of Japanese Wikipedia corpus [10]. We scale the score of similarity to the range between 0 to 1. To rank the similarity score, we confirm on the last four digits of the score.

Table II shows the list of similar words to "cute" ranked in descending order, resulting from the calculation of cosine similarity of the word vectors by Equation (1) comparing to the result of similarity of words from WordNet by Equation (2). The word "cute" is an adjective defined in the Longman Dictionary of Contemporary English LDOCE [12] to mean "very pretty or attractive," "especially American English sexually attractive," and "especially American English clever in a way that can seem rude". This word can be detected in both Word2Vec and WordNet. In WordNet, "cute" is classifies into two meanings which cute#1 has the meaning of cunning and attractive, and cute#2 has the meaning of precious and artful, as shown in Table II.

TABLE II. RESULTS OF THE CALCULATION OF SIMILARITY OF "CUTE" IN WORD2VEC AND WORDNET

Cute			
Word2Vec		***WordNet***	
Similar words	Score of Word Similarity	cute#1	cute#2
cafe	0.6756	cunning	precious
Style	0.6687	attractive	artful
Bee	0.660	-	-
girl	0.6564	-	-
Crea	0.6482	-	-
Nuts	0.6475	-	-
Honey	0.6425	-	-
girls	0.6393	-	-
Enjoy	0.6386	-	-
boy	0.6355	-	-

In the list of similar words resulted from Word2Vec by cosine similarity, we found that cafe, Style, Bee, girl, and such are the closest vectors by the order. They explicitly do not have similar meaning to the word "cute" at all. Instead, they are found in the similar context such as "cute cafe", "cute style", "cute bee", "cute girl" or in a limited contextual distance. They are associated to each other rather than having a similar meaning. On the other hand, WordNet gives a list of similar words to "cute" as "cunning", "attractive" and "precious", "artful". There are two lists because "cute" has two meanings (senses). Both lists neatly show the same or similar meaning to "cute". Therefore, it is meaningful to interpret the list of similar words from Word2Vec as a list of associated words, and the list of similar words from WordNet as a list of similar sense of words.

In similar case, Table III shows the list of similar words to "fast" ranked in descending order, resulting from the calculation of cosine similarity of the word vectors by Equation (1). The word "fast" is an adjective defined in the Longman Dictionary of Contemporary English LDOCE [11] to mean "moving or being able to move quickly", "doing something or happening in a short time" and "a clock that is fast shows a later time than the real time". This word can be detected in both Word2Vec and WordNet. In WordNet, "fast" as an adjective has ten meanings. As an example, for the meaning of "acting or moving or capable of acting or moving quickly" defined in fast#1, the similar words are listed as "accelerated", "alacritous", "blistering", for instance.

TABLE III. RESULTS FROM WORD2VEC AND WORDNET FOR THE SIMILAR AND RELATED WORDS TO "FAST"

Fast				
Word2Vec		***WordNet***		
Similar words	Score of Word Similarity	fast#1	…	fast#10
slow	0.8008	accelerated#1	…	…
high	0.7906	alacritous#1	…	…
off	0.7574	blistering#3	…	…
down	0.7494	double-quick#1	…	…
speed	0.7357	express#2	…	…
short	0.7354	fast-breaking#1	…	…
low	0.7335	fast-paced#1	…	…
out	0.7335	fleet#1	…	…
Non	0.7308	high-speed#1	…	…
up	0.7300	hurrying#1	…	…
…	…	…	…	…

In the list of similar words resulted from Word2Vec cosine similarity, we found that "slow", "high", "off", "down" and such are the closest vectors. They do not have similar meaning to the word "fast" at all. It also sounds strange to find the antonym "slow" as the closest vector in terms of cosine similarity. It seems to be very normal to find the antonyms as the closest vectors resulting from Word2Vec. This is because they occur in very similar contexts. In general, things can be referred to in a positive sense as well as in a negative sense. In terms of WordNet, viewing at the meaning of "acting or moving or capable of acting or moving quickly" or "fast#1", there are words such as "accelerated", "alacritous", "blistering" defined in the list of similar words. It can be concluded that the list of similar words from Word2Vec are interchangeable in the grammatically same context, while, the list of similar words from WordNet are interchangeable by maintaining the same meaning.

From the results of similarity measure for "cute" in Table II and for "fast" in Table III, we can observe that the distance in Wordnet is semantically defined more than the one in Word2Vec. The results from Word2Vec similarity measure, on the other hand, does not reflect the semantic closeness, but the similarity of the context of word occurrence. However, the results from WordNet similarity measure in Table IV show the problem of inclusion of antonyms because they are classified in the same class of "emotion-state" in the hierarchy.

Arithmetic calculation of vector is a substantial property of vector for deriving vector relationship which can encode the abstract set of all words that share a common property.

For example, “man” and “woman” share the relationship Male-Female. This means that “man” is the paraphrase of {woman, X} where X is the abstract set of all the words that encode the relationship Male-Female. Consequently, since “king” and “queen” share the same Male-Female relationship, then “king is the paraphrase of the set of words {queen, X}. Therefore, Equation (3) is satisfied by the definition of vector.

$$\begin{cases} U_{man} = U_{woman} + U_x \\ U_{king} = U_{queen} + U_x \end{cases} \Rightarrow U_{queen} = U_{king} - U_{man} + U_{woman} \quad (3)$$

This also occurs in the case of Country-Capital relationship such as Japan-Tokyo which can derives the relationship of Thailand-Bangkok, and part-of-speech relationship such as walk-walking which can derives the relationship of swim-swimming.

V. SOLUTION FOR CONCEPT SIMILARITY MEASURE

Currently, Wu-Palmer similarity measure is commonly used to obtain word similarity in WordNet. It is implemented in NLTK module by using *lcs* algorithm to measure the word similarity by calculating the path distance along the WordNet hierarchical is-a relation as shown in Equation (2). The result actually returns the concepts based on the path distance measure which reflects the closeness of the concepts in terms of hierarchically grouping rather than the semantic similarity. As a result, we can find the concept of “unhappiness” listed as the most similar concept to the concept of “happiness”. This case is very normal because both synonyms and antonyms belong to the same is-a class in WordNet.

TABLE IV. WORDNET SIMIRARITY OF “HAPPINESS”

Happiness	
WordNet	
	Similarity
hapiness#1	1
unhappiness#2	0.85
sadness#1	0.7142
surprise#1	0.6667
neutral#1	0.1538

In the experiment, we measure the similarity of the opposite concepts, “happiness” and “unhappiness”, and obtain 0.85 of similarity score. Although “unhappiness” is defined as the antonym of “happiness” in the lexicon, the similarity between the two concepts is the highest one in the result as shown in Table IV. To confirm the hierarchical expression in WordNet, we also investigate the results on four other words of sensitivity.

To solve the problem that the current WordNet does not provide such semantic similarity, we propose a new method of similarity measure for WordNet. Specifically, we propose an extension of NLTK’s similarity measure by removing antonyms from the result and utilizing the “similar to” relation to form the path of concept similarity in WordNet. As shown in Table IV, the current NLTK’s similarity measure calculates the similarity based on a mere is-a relation. In many cases, the antonyms are ranked at the top of the similarity list. In order to solve this problem, we add a process to remove the antonyms from the list by using the relation of “antonym” defined in WordNet.

For example, in the case of "happiness" and "unhappiness", both belong to the same hypernym of “emotion state”. Figure 2 shows a segment of the is-a structure retrieved from WordNet. We subtract the concept of “unhappiness” which is related as an antonym from the resulted list.

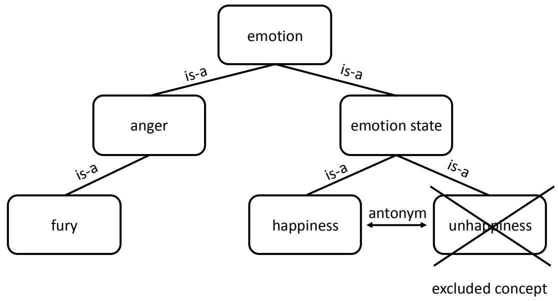

Figure 2. A segment of the is-a structure retrieved from WordNet

To form the path of concept similarity by the “similar to” relation in WordNet, according to the property of inheritance defined in the frame theory [13], we generate a virtual is-a relation for the similar concepts. The concepts, in the “similar to” relation, are expended to inherit the properties from their “attribute” relation.

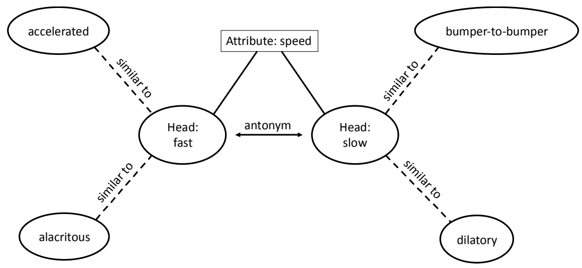

Figure 3. A segment of “attribute” and “similar to” relations retrieved from WordNet

Figure 3 shows a structure of a concept of “fast” with its relations of “attribute”, “similar to”, and “antonym” [13]. The similar concepts of “fast”, such as “accelerated”, are not the same node as “fast”, and their distance *len(fast, accelerated)* is 2 by assuming that “attribute” is a shared node between them, a virtual is-a relation per se. Therefore, according to Equation (2), the similarity between the concepts related by “similar to” is not 1.

VI. SYSTEM 1 AND SYSTEM 2

A. What is System 1and System 2?

According to Daniel Kahneman’s definition, System 1 is fast, automatic, frequent, emotional, stereotypic, and unconscious. In contrast, System 2 is slow, effortful, infrequent, logical, calculating, conscious [14].

System 1 tasks in machine learning and deep learning have been treated as a process of data modelling. Until now, most of the advanced research in developing a model in deep learning approaches are focused on the learning from static data sets, while perceptual tasks and ad hoc problem-solving abilities are difficult for deep learning to handle. Most of the tasks in System 1 can be automated in the bottom-up manner of data driven approaches. While, perceptual tasks and ad hoc problem-solving abilities are still difficult for deep learning to handle.

In recent years, with the shift in research direction and the emergence of new tools such as soft attention and deep reinforcement learning, there is a rise in new research to address each of the so-called System 2 tasks that humans consciously perform, such as reasoning, planning, understanding causality, and systematic generalization, in applications such as natural language processing. Yoshua Bengio [15] argues that such an extension of deep learning from a System 1 task to a System 2 task is important to achieve the traditional goal of deep learning, which is the discovery of high-level abstractions, and suggests that it should allow for out-of-distribution generalizations based on hypotheses of local (in time, space, and conceptual space) changes in the environment due to agent intervention. In this way, it goes without saying that the conventional research on deep learning and the current research on deep learning are located at different places.

B. Considering the world of similarity and meaning treated in the System 2

The results of the semantic similarity test between Word2Vec and WordNet show a significant difference in the sense of similarity expressed by both approaches. There are many recent studies using deep learning techniques without any concerns regarding to the different characteristics in Word2Vec and WordNet, especially when working with text data. A proper semantic interpretation is needed to result a reliable conclusion. Considering the part of vectorization, polysemy can easily confuse the process of vectorization in the space. Words which contain more than one meaning are encoded in one vector representation in the space. WordNet which has more precise in word representation in the form of synset can encode the meaning in different form.

Deep learning is a powerful neural network architecture to create a model from a huge training dataset. The result is more likely biased towards System 1. For example, when building artificial intelligence robots that interact with humans, do they act as System 2? For the current needs for robots, it may be more important to response timely to avoid obstacles in the moving path rather than taking time to understand the obstacles' thinking. However, System 2 is an aim for the artificial general intelligence research as reported in Yoshua Bengio's talk on "System 2 deep learning" [15].

VII. CONCLUSION

Languages are manipulated in terms of syntactic and semantic viewpoints. We discuss their ambiguities on both sides of analysis. Many approaches are proposed to include the possible contexts for better modeling and representation. To compensate for the shortcoming of WordNet similarity measure, we propose a method of discarding the antonyms from the similarity measure. The achievement in word or document vectorization has shown its potential to include the contexts for accurate language representation and modeling. However, there is still a gap of interpretation such as in the case of similarity measure based on the result of Word2Vec and WordNet. One solution has been proposed to solve the similarity measure in WordNet. A proper hybrid approach of the language duality is expected to leverage the advantages from both sides of language properties as discussed in the talk on "System 2 deep learning". As a future work, we are considering to propose a hybrid language modeling of WordNet in terms of semantic based representation and Word2Vec in terms of syntactic based representation to compensate the disadvantages of each other.

REFERENCES

[1] B. J. Dorr, "Machine Translation Divergences: A Formal Description and Proposed Solution", Computational Linguistics, Volume 20, Number 4, December 1994, pp. 597-633.

[2] Y. Jiang, S. J. Lee, A Similarity Measure for Text Classification and Clustering, IEEE Transactions on Knowledge and Data Engineering, 26(7), 2014, pp. 1575~1589

[3] N. Chomsky, "Syntactic Structures", Mouton & Co, 1957, pp. 15.

[4] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, "Distributed Representations of Words and Phrases and their Compositionality", NIPS,2013, pp. 3111-3119.

[5] G. A. Miller, "WordNet: A Lexical Database for English". Communications of the ACM Vol. 38, 1995, pp. 39-41.

[6] C. Fellbaum, "WordNet: An Electronic Lexical Database". Cambridge, MA: MIT Press, 1998, ed.

[7] A. K. Taylor, "Encyclopedia of Human Memory [3 volumes]", ABC-CLIO, 2013.

[8] Z. Wu and M. Palmer, "Verb semantics and lexical selection", Proceedings of 32nd annual Meeting of the Association for Computational Linguistics, June 27-30, 1994, Las Cruces, New Mexico.

[9] F. Baader, "Computing the Least Common Subsumer in the Description Logic EL w.r.t.Terminological Cycles with Descriptive Semantics", Lecture Notes in Computer Science, 2003, pp. 271-350.

[10] Japanese Wikimedia, 2020. Retrieved from https://dumps.wikimedia.org/jawiki/latest/.

[11] Peason Education, "Longman Dictionary of Contemporary English LDOCE6", 2014.

[12] S. C. Hayes, D. Barnes-Holmes, B. Roche, "Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition", 2001.

[13] V. Sheiman and T. Tokunaga "AdjScales: Visualizing Differences between Adjectives for Language Learners", IEICE Transactions on Information and Systems, Vol. E92–D, No. 8, August 2009, pp. 1542-1550.

[14] D. Kahneman, "Thinking, Fast and Slow", Flrrar, Status and Giroux, 2011.

[15] Y. Bengio, "From System 1 Deep Learning to System 2 Deep Learning", unpublished, NurlPS, 2019.