Computational Linguistics in the Netherlands Journal 10 (2020) 5–17 Submitted 10/2020; Published 12/2020

Comparing Frame Membership, WordNet-Based Similarity and Distributional Similarity

Esra' M. Abdelzaher* ESRA.ABDELZAHER@GMAIL.COM

* *University of Debrecen, Doctoral School of Linguistics/ University of Debrecen, Institute of English and American Studies, Debrecen, Hungary*

Abstract

Frame co-membership is a relation between lexical units occurring within the same lexico-syntactic environment and representing the same cognitive structure (i.e., frame). This informal relation is valuable to the construction of predicate-modeled language resources, automatic induction of lexical units and semantic role labeling. However, it requires extensive human effort, which slows the progress of FrameNet (FN) and undermines the construction of similar databases. The current study first addresses the challenge of converting frame membership into a numerical similarity relation. This conversion should facilitate the comparison between frame membership, WordNet-based similarity and distributional similarity. The study then identifies the most statistically compatible measures with frame membership. The proposed measure of degree of frame co-membership (DFCM) is entirely based on the FN database. It embraces the unique features of Frame Semantics and does not account for any frame-external data. Accordingly, it preserves the individual approach of the theory and the distinctive criteria for word grouping. Although DFCM does not reflect the lexical or numerical relations between words in WordNet (WN) or distributional semantics, it is compatible with similarity scores obtained from the WN database and through distributional tools. The results may have considerable implications for the enrichment of FrameNet's lexicon without jeopardizing the precision of the database or maintaining the sole dependence on the manual effort of lexicographers.

1. Introduction

The FrameNet (FN)[1] database contains 13676 lexical units (i.e., a word paired with one of its senses) and 1224 frames, which are "schematic representation of a situation" in which several participants (i.e., Frame Elements) are involved. According to FN, the frame of ADDICTION[2], for instance, schematically describes the "physiological or psychological compulsive dependence" of an Addict on an Addictant. The core participants in this situation include Addict, Addictant, Compeller and Degree. FN places *compulsive*.a and *alcoholism*.n as co-members in the same frame, despite their different parts of speech and semantic fields. Frame co-membership is subject to the judgment of lexicographers and the support of annotated examples. Lexical units would be co-members if they evoke the same frame and share the same lexico-syntactic environment (Ruppenhofer et al. 2006).

The manual annotation of the FN database guarantees the precision of the included valence and semantic information. At the same time, it slows the process of developing the database or constructing FN-like resources (Baker 2012). FN provides software programs with valuable information regarding lexical units, the situations in which they are used (frames), the participants involved in every scenario (frame elements), other relevant situations and associated lexical units to solve several tasks in Natural Language Processing (Tóth 2014). However, the restricted coverage of annotated data and the limited scope of polysemous verbs in FN are significant challenges to frame-based tasks in NLP (Moor et al. 2013). Besides, the unique annotation scheme used in FN is hardly compatible

1. Project status information is retrieved from framenet.icsi.berkeley.edu on June 10, 2020
2. Retrieved from framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Addiction, on June 10, 2020

with other language resources, such as WordNet (WN)[3] (Chow and Wong 2006), and it does not account for any statistical information (De Bleecker 2005). Moreover, FN allows the inclusion of formally dissimilar lexical units in the same frame, which may describe opposite events, and permits the inclusion of related lexical units in different frames (Virk et al. 2016).

These challenges undermine the possibility of relating frame co-membership to other measures of similarity. The present study compares frame co-membership to WN-based and distributional similarity. It relies on frame membership to assign similarity values to lexical units in FN and, accordingly, facilitate the comparison between the three measures. The measure is designed to reflect the co-membership relation in FN without the use of any frame-external knowledge, whether exported from WN or distributional tools. Therefore, the numerical values obtained from this measure can be interpretable within the linguistic context of FN.

Therefore, the main objectives of the current study are (a) assigning numerical values to co-lexical units in FN and (b) measuring the statistical correlation between the proposed FN-based similarity scores and the similarity scores retrieved from distributional similarity and WN-based measures. Accordingly, the study addresses the following questions.

(a) How can frame membership be reflected quantitatively without reliance on extra-frame knowledge?

(b) Does the introduced measure for frame co-membership correlate with WN-based measures or distributional similarity?

The rest of this paper consists of five more sections. The diversity of theoretical bases and practical methods adopted in FN, WN and distributional similarity is outlined briefly in the next section (section 2). Section 3 reviews related work, while section 4 explains the data and methods used in this paper. Section 5 displays the results and, finally, section 6 presents the concluding remarks.

2. Measures of Similarity and Relatedness

On the one hand, corpus evidence and the judgment of expert lexicographers are FN's tools for grouping or splitting lexical units in the same frame or in different frames. FN does not use formal criteria, such as part of speech (POS) or ontological hierarchies, in identifying the similarity of words. FN annotators have been trying to formulate clear criteria for identifying frame co-members. To elaborate, lexical units must occur with the same number and type of arguments, and the relation between the lexical units and their arguments must be the same in order to place them in the same frame (Ruppenhofer et al. 2006). The need for expert human annotators is inevitable to the application of these similarity criteria. Besides, the null instantiation of arguments (i.e., linguistically absent but cognitively perceived) and the frame specificity of arguments, which assigns the distinctive frame elements of Abuser, Aggressor and Assailant to the frames of ABUSING, VIOLENCE and ATTACK respectively, necessitate the decision of a lexicographer. Complicating matters, *buy* and *sell*, for instance, are not co-lexical units in FN because they adopt different perspectives (COMMERCE BUY and COMMERCE SELL) of the experiential knowledge of COMMERECE GOODS TRANSFER.

On the other hand, WN depends on the POS and the relations of lexical semantics such as synonymy, hyponymy, troponymy and antonymy in detecting the similarity between words. WN places synonymous word senses in a synset (i.e., synonymy set). Synsets are hierarchically related to more general (hypernym) and specific (hyponym or troponym) synsets. WN creates semi-formal ontological relations between word senses (Miller 1998). These clear criteria of relating word senses in WN motivated several scholars to develop numerical similarity measures based on the WN database. These measures rely on the glosses, which help in the disambiguation of word senses, and the vertical path between synsets and the lowest common synset, which subsumes two similar synsets.

3. Online version of the project is accessible through wordnetweb.princeton.edu/perl/webwn

Unlike FN and WN, the distributional similarity is not theoretically grounded. Distributional similarity relates words to each other based on their co-occurrences with the same words because it hypothesizes that similar words occur in similar contexts (Firth 1957). Whereas scholars in computational linguistics appreciate the role of distributional semantics in meaning representation, cognitive linguists vigorously deny its role. Lenci (2008) discusses the possibility of summarizing the lexical properties of a word in its statistical association patterns and the existence of a causal relationship between the contexts of a word and the semantic representation of meaning as the main reasons for questioning the adequacy of the distributional representation of meaning. He argued for the strong distributional hypothesis, which regards distributional similarity as a way to explain the semantic content of a word at the cognitive level. Accordingly, it is not just a statistical way of representing words. Unlike Lenci (2008), Ruppenhofer et al. (2006) undervalue the importance of statistical co-occurrence to FN's lexicographic information. After all, distributional models are widely used in computational linguistic applications because they require minimal human supervision and make use of large raw corpora.

The different theoretical and practical approaches followed in FN, WN and distributional semantics lead to a glaring discrepancy between similar words determined by the three approaches. For instance, *feeling*, *impression* and *belief* are placed within the same synset in WN, whereas they evoke the frames of FEELING, REGARD and OPINION in FN. Similarly, *emotion*.n, *experience*.v and *full*.a are co-lexical units in the frame of FEELING, although they belong to different synsets in WN. Distributional similarity, however, judges *feel*, *very*, *feels*, *felt*, *feelings* and *extremely* as the most similar words to *feeling* according to SpaceXplorer, a tool for distributional similarity detection based on a corpus of Wikipedia snapshot (Tóth 2015). Despite the differences, the rich and accurate valence information in FN, comprehensive lexical-semantic coverage of WN and the automatically-driven and statistically-evident relations of distributional similarity encouraged scholars to combine the different types of knowledge.

3. Review of the Literature

Previous attempts at integration mainly addressed the drawbacks of FN, WN or both. Some studies depended on the similar linguistic features in the two resources, while others tried to link the frame knowledge to the statistical representation of meaning.

Tonelli and Pighin (2009) made use of the broad lexical coverage of WN to expand FN's lexicon through linking FN's lexical units to WN's synsets. They relied on the shared linguistic features between FN's definitions of lexical units and WN's glosses of synsets. They detected the overlaps between the definitions and glosses and checked common lemmas between synsets and frames in order to assign WN synsets to frames. They introduced MapNet, which aligns LUs to synsets, and enriched FN with 4,265 lexical units imported from WN. They reported 0.78 precision of the new words added to FN, based on the evaluation of a 200-word sample. The results suggest a promising degree of similarity between the frame knowledge and synset information.

Similarly, Ferrández et al. (2010) aligned the lexical units in FN to their relevant synsets in WN. The ultimate goal was to build a joint hierarchy that overcomes the shortcomings in the two resources. However, they relied more on proposing numerical measures than on tracing similar linguistic features. The alignment correlated the synset that includes the target lemma with the lexical units that belong to the same lemma. They depended on calculating the neighboring similarity scores between word senses of the same lemma in WN and FN and measuring the distance between them. They assigned numerical scores to WN's relations, including synonymy and hyponymy, and FN's frame-to-frame relations. Although the experiment was promising in adding the synonyms and hyponyms of the synset to the lexical units in the frame, the words' POS imposed a considerable challenge to this calculation-based alignment. Nouns, verbs and adjectives can be embraced in a single frame, whereas WN creates a net for each POS.

Laparra et al. (2010) somewhat overcame this challenge in their attempt to map WN's synsets to FN's lexical units. They introduced WordFrameNet, which increased the coverage of FN with the synset assigned to a lexical unit, and linked unrelated synsets (especially those belonging to different parts of speech) in WN through frame co-membership relations. It thus turned the challenge of having different parts of speech in a frame into a new semantic relation between synsets in WN. The challenge they could not address was the fine-granularity of senses in WN, which led to linking the same lexical unit to several WN synsets.

Instead of using WN's lexicon to expand FN's, Pennacchiotti et al. (2008) depended on distributional tools. They attempted to represent the frame knowledge distributionally through a vector space model. They used word-based and document-based spaces to represent words through vectors and measured the similarity between pairs of co-lexical units and unrelated ones. They concluded that distributional semantics is, to a large extent, valuable for modeling the notion of a frame as reflected by co-lexical units, and that word-based and document-based spaces capture similar properties. Although the results were promising, they did not report the features that were successfully modeled and those that were missing, and they did not interpret the results.

Kleinbauer and Trost (2018) pursued a similar goal and compared the distributional and frame semantic properties of words. They resorted to the distributional representation of words through vectors. Then, they calculated the distance between the nearest neighbors of words and measured the Euclidean distance and the angles between pairs of word vectors. Although some features were distributionally and semantically conventional, co-lexical units were, in several instances, farther from each other than unrelated words belonging to different frames. Given the fuzziness of the distributional similarity relation, this finding was not further interpreted.

4. Data and Methods

The current study uses the database of FN 1.7 to select a sample of lexical units to which the suggested frame-based measure is implemented. The WN database is not directly employed, but eight WN-based measures of similarity are adopted to calculate the similarity scores between FN's lexical units. BNC and enTenTen are used to measure the distributional similarity between the selected lexical units from FN.

4.1 Frame Membership

According to FN, a lexical unit is "a pairing of a lemma and a frame," and a lemma may carry "one or more senses." Fillmore et al. (2004) explained that membership in the same frame is one of the ways FN uses to link lexical units. However, co-membership in a frame does not indicate equal representation of the frame by each lexical unit. Fillmore et al. (2003a) acknowledged that some frame members are "central" to the frame, and the exploration of these "central members" in the corpus is a prerequisite to verifying the understanding of the semantic and syntactic features of frame members. *Tie*.v, for instance, is chosen by Fillmore et al. (2003b) as one of the central members of the ATTACHING frame, although the initial list of candidates included more than 20 lexical units, including *attach*.v. The manual exploration of *tie* as a lemma, in corpus examples, revealed that it has different senses belonging to several frames such as KNOT CREATION. The centrality is, therefore, not semantically associated with the polysemy or monosemy of a word. In addition, the centrality of *tie*.v to ATTACHING did not correspond to the centrality of *tie*'s senses in dictionaries. That is to say, the centrality is not lexicographically-motivated. Furthermore, FN does not consider quantitative information in the creation of the database and, accordingly, the centrality of lexical units to frames is not judged numerically. Although the centrality of some lexical units to a frame has not been discussed by FN creators, it is presupposed before and during the construction of frames ((Fillmore et al. 2003a); (Fillmore et al. 2003b)).

The current study attempts to address a similar point quantitatively and without reliance on any extra-frame knowledge. It hypothesizes that there is a degree of frame membership (DFM) for each lexical unit (LU) in a frame (F). DFM depends on the number of frames evoked by the lemma (L) containing the LU. The highest DFM equals 1 and occurs only if the lemma is monosemous, i.e., the LU has the membership of a single frame. Polysemous lemmas distribute the membership of the LUs over two or more frames. Polysemy, consequently, decreases the degree of belongingness to each frame.

$$DFM(LU) \subset L = \frac{1}{Fn} \tag{1}$$

The suggested measure of frame membership assigns diversified numerical values for LUs in FN. Monosemous lemmas assign their lexical units the highest degree of frame membership. For instance, 1 was assigned to *able*.a and *capable*.a as prototypical members of the CAPABILITY frame because their lemmas do not include any other LUs and, accordingly, are not members in any other frame.

$$DFM(able) \subset able.a = \frac{1}{1} \tag{2}$$

On the contrary, *state*.n evokes five frames and this, therefore, reduces its DFM. Worth mentioning, this measure is POS-sensitive, i.e., it considers only a single POS at a time. Therefore, DFM for *state*.n is 0.2, whereas it is 1 for *state*.v.

$$DFM(state) \subset state.n = \frac{1}{5} \tag{3}$$

$$DFM(state) \subset state.v = \frac{1}{1} \tag{4}$$

In FN, co-LUs also vary in their association with each other. Although all co-LUs must display similar lexico-grammatical behavior to be included in the same frame, some co-LUs are more similar to each other than to other LUs in the same frame. *August*.n and *April*.n are more similar to each other than to their co-members *age*.n and *year*.n, in CALENDERIC UNIT. Baker et al. (2003) refer to this "identical" behavior of a sub-set of co-LUs as "by analogy" in the annotated database. Worth mentioning, this relation is labeled "BTDT" in the current version of FN, and it is limited to 19 LUs (all evoke a single frame).

The present study maintains the assumption of having different degrees of similarities among co-LUs, but it adopts another (quantitative) perspective. First, it argues that these variable similarity degrees are applicable to any pair of co-LUs. Second, it relies on the proposed DFM of two LUs and the number of the shared frames (SFs) their separate lemmas have in common to measure the degree of frame co-membership (DFCM). The strength of association between two monosemous co-LUs or polysemous co-LUs sharing all of their frames will be the same. Unlike the degree of membership, the strength of association between two LUs is not affected by the polysemy of the lemmas. It is affected by the number of frames they have in common.

$$DFCM(LU1, LU2) = (DFM(LU1) + DFM(LU2)) \times (SF)) \div 2 \tag{5}$$

The most substantial co-membership relation between LU1 and LU2 equals 1 and occurs only when the lemmas of the two LUs evoke the same frame(s). For instance, *cook*.v and *fry*.v evoke the same three frames of APPLY HEAT, COOKING CREATION and ABSORB HEAT while *gossip*.v and *converse*.v evoke the same single frame of CHATTING. The two pairs are assigned the highest degree of frame co-membership (i.e., 1). The strength of co-membership descends when the LUs have fewer frames in common.

Dark.a and *warm*.a, for instance, place their lexical units in 5 and 7 frames, respectively, and this reduces their DFM to 0.2 for *dark*.a and 0.14 for *warm*.a. The two lemmas have a single frame in common. Therefore, the DFCM between *warm*.a and *dark*.a descends to 0.17.

$$DFCM(dark.a, warm.a) = ((0.2) + (0.14) \times 1) \div 2 \tag{6}$$

where 0<= DFCM <= 1, the 0 value is retrieved only when the lexical units have no frames in common (i.e., are not co-LUs). The measure is applicable to any pair of LUs whether or not they belong to the same POS; however, the study implemented it only on pairs of the same POS to avoid calculation errors when WN-based measures are applied to the same pairs.

4.2 WN-Based Similarity Measures

The obtained numerical values reflecting DFCM are compared to WN-based measures of similarity. For adjectives, only two WN-based measures are applicable because other measures rely on the hierarchical relations absent from the adjectives net. Banerjee and Pedersen (2002) LESK measures the similarity between any pair of words (having the same or different parts of speech) based on the overlap between their glosses. LESK similarity scores can be 0 if the glosses do not overlap at all, and it has no maximum score because it detects overlaps over the glosses of the synsets containing the two words and their extended synsets. Hirst and St.Onge (1995) HSO considers the lexical chains, whether hierarchical or horizontal, between two concepts to measure their similarity. Therefore, it applies to any pair of words as long as they belong to the same POS. It returns values that vary from 0 if the distance between the two concepts exceeds five chains to 16 when the two words belong to the same synset.

Other WN-based measures stipulate that word pairs are either nouns or verbs. Wu and Palmer (1994) WUP is implementable only within the nouns and verbs nets because it considers the depth of the lowest common subsumer (LCS) node between the two concepts in addition to calculating the concepts' depths. Similarly, Leacock and Chodorow (1998) LCH and Lin (1997) Lin decide upon the similarity of the words based on the shortest hierarchical path between the two concepts and their LCS. Jiang and Conrath (1997) JCN and Resnik (1995) Res are also applicable only on the nets of nouns and verbs because they make use of LCS. However, they consider the information content of the two concepts and of their LCS, not the length of the path between them. The similarity scores were retrieved through Shima's (2013) web-based software "WordNet Similarity for Java."[4]

4.3 Corpora and Distributional Similarity

The scores retrieved from FN-based and WN-based measures are finally compared to the distributional similarity scores obtained from enTenTen2015 corpus and the British National Corpus (BNC). Rychlý and Kilgarriff (2007) algorithm is executed to calculate the similarity scores between LUs as it enables the unification of the POS, which guarantees a degree of specification and satisfies the POS criterion adopted by the previous measures. Kilgarriff and Duvuru (2011) explain that the lemmatization and POS-tagging of a corpus are prerequisites for the implementation of the algorithm. The first step to generate a distributional thesaurus of a word is creating its word sketch. Word sketches are one of the unique features in Sketch Engine, which summarize the grammatical and collocational behavior of a word. They record words that co-occur with the target word in a grammatical relation (e.g., a subject of, an object of, adjectival modifier). Similar to several collocational measures, word sketches order the collocates of a word according to their statistical significance. As a demonstration, the word sketch of *challenge*.n, in the BNC, identifies *pose*, *relish* and *mount* as the typical verbs co-occurring with *challenge*.n in an "object of" grammatical relation while *face*, *confront* and *lay* are the typical verbs with which *challenge*.n occurs in the subject position.

Worth mentioning, Atkins et al. (2003b) describe the word sketch representation as "more sophisticated and informative" than the traditional KWIC representation of corpus examples. Also, Atkins et al. (2003a) highlight the potential contribution of word sketches to the identification of the FEs and senses of LUs in FN if compared to the KWIC approach. Later, Baker (2012) reported on

4. The software is accessible through: ws4jdemo.appspot.com/

the ongoing developments of new tools for FN based on the algorithm of word sketches. The project aims at saving the time and effort of lexicographers to create new frames, determine their FEs and decide on their evocative LUs.

Due to the advantages of the word sketch representations of a target word at the grammatical and collocational levels, Rychlý and Kilgarriff (2007) formulated an algorithm to measure the distributional similarity between two words based on the similarities in their word sketches. The word sketch adds a grammatical feature to the collocational patterns of a word, and so does the distributional algorithm (i.e., it includes grammatical relations as part of the context). It captures the categorical similarity between *beer*.n and *wine*.n, for instance, based on their statistically significant co-occurrence in the object position of *drink*.v. The grammatical component added to the traditional second-order co-occurrence distinguishes this algorithm from other distributional measures of similarity and motivates scholars to adopt it in language learning (Baisa and Suchomel 2014), facilitating the creation of dictionary definitions (Stará and Kovár 2016), and in constructing semantic fields cross-linguistically (Zakharov et al. 2020). The algorithm is accessible through Sketch Engine, which also hosts the two corpora used in the present study.

5. Results and Discussion

Measure	Adjectives	Nouns	Verbs
Correlations between the suggested DFCM and WN-based similarity			
DFCM and Lesk	-0.0509	0.1054	-0.0092
DFCM and LCH	N/A	0.3047*	0.0911
DFCM and Res	N/A	0.4842*	0.1018
DFCM and WUP	N/A	0.4291*	0.1686
DFCM and Path	N/A	0.2572	0.096
DFCM and Jcn	N/A	0.2627	0.173
DFCM and HSO	0.0885	0.2656	0.0532
DFCM and Lin	N/A	0.4052*	0.1111
Correlations between the suggested DFCM and distributional similarity			
DFCM and BNC	-0.369*	0.058	-0.0661
DFCM and eTenTen	-0.2426	0.3147*	0.2213

Table 1: Correlations between DFCM, WN-based similarity and distributional similarity scores

The study compares the similarity scores retrieved from FN, WN and the distribution algorithm to each other using Pearson's correlation test to identify the most compatible measures among the studied constructs. The correlation matrix was calculated between all measures. The concluded correlations, which are based on 280 observations, are considered statistically significant if $p < 0.05$.

The correlation scores for each POS are provided in table 1. As tabulated, the DFCM among the three groups (i.e., nouns, verbs and adjectives) correlated differently with the other measures of similarity. All WN-based measures apply to nouns and verbs, but only the similarity scores calculated by LCH, Res, WUP and Lin correlated significantly with DFCM among nouns. None of the WN-based measures correlated with DFCM for verbs, however. Most of the WN-based measures are not applicable (N/A) to adjectives. Moreover, adjectives did not significantly correlate with HSO or LESK, which are the two applicable measures.

As for the distributional similarity scores, enTenTen provided more compatible scores with the DFCM than the BNC, although it is the main corpus of FN. Moreover, enTenTen-based distributional similarity correlated significantly with the DFCM of nouns. The glaring discrepancy between the size of the BNC (112,345,722 tokens) and that of enTenTen (15,411,682,875 tokens) affects the frequency of word occurrences and, accordingly, influences the richness of word sketches and the generation of distributional thesauri, which are based on the two corpora. To elaborate, the word sketch of *captivating*.a, based on the BNC, captures a very limited number of collocates (39 words), participating with the target word in only five grammatical relations. This is due to the infrequent use of *captivating*.a (only 44 times) in the BNC. On the contrary, *captivating*.a occurs 34,524 times in enTenTen, which gives sufficient attestations to describe the collocational and grammatical patterns in a word sketch. The word sketch for *captivating*.a includes more than 200 collocates co-occurring with it in 9 grammatical relations. This provides the thesaurus with rich input to create the list of distributionally similar words. The poor coverage of adjectives, in particular, in the BNC maybe a substantial reason for the significant anti-correlation with the DFCM between adjectival pairs.

5.1 DFCM among Nouns

For nouns, the correlations between DFCM and all WN-based and distributional measures were positive, and they varied from moderate correlations with Res, WUP, Lin, enTenTen and LCH to the weakest correlation with the BNC. DFCM's correlations with the similarity scores provided by Res, WUP, Lin, enTenTen and LCH were statistically significant at p <0.05. Although the BNC is the corpus mainly used in building and annotating the dataset of FN, the correlation between distributional similarity scores retrieved from the BNC and the DFCM scores was not statistically significant.

The unification of POS of the analyzed pairs of LUs increased the possibility of having synonymy or hyponymy relations among co-LUs. It also guaranteed the existence of a node subsuming the two LUs. Several cases in which the highest DFCM corresponded to high scores of WN-based similarity involved one of WN's relations. Synonymy and sister-terms were the most dominant relations traced between FN's co-LUs. In the absence of a WN relation between co-LUs, relatively high DFCM corresponded to high WN-based similarity scores, too. When the DFCM decreased because of the distributed DFM across several frames, the similarity scores also dropped down, and even the LCS between words in the studied pairs was too distant to establish a strong similarity relation. Table 2 compares the DFCM to the WN-based similarity in a sample of the studied noun pairs.

Co-LUs	DFCM	Res	WUP	LCH	Lin	WN relation
Abuse.n- maltreat.n	1	9.2008	1	3.6889	1	Synonymy
Alliance.n- coalition.n	1	9.2008	1	3.6889	1	Synonymy
April.n- August.n	1	6.2403	0.8889	2.5903	0.7133	Sister-terms
Bonnet.n- cap.n	1	7.7404	0.8696	2.59	0.7761	Sister-terms
Go.n- push.n	0.75	2.6044	0.7059	1.8971	0.0763	None
Family.n- crowd.n	0.666	3.5267	0.7143	2.0794	0.4621	None
Chaos.n- order.n	0.555	3.1688	0.7143	2.0794	0.3651	None
Paper.n- operation.n	0.333	0.7794	0.5	1.743	0.1331	None

Table 2: Examples of DFCM values and statistically significant WN-based similarity scores among nouns

Also, Pearson correlations between Res, WUP, Lin, and LCH scores for co-LUs on the one hand, and unrelated LUs, on the other hand, were moderately negative (Figure 1).

Figure 1a visualizes WUP similarity scores among 41 co-LUs (represented by the blue line) and 41 unrelated LUs (represented by the red line). Whereas the y-axis represents the range of WUP similarity scores (0<WUP score<=1), the x-axis represents the 41 pairs of LUs. As illustrated, co-LUs have remarkably higher similarity scores than unrelated LUs. WUP assigns the maximum score to several co-LUs and high similarity values to all co-LUs. The lowest similarity score WUP

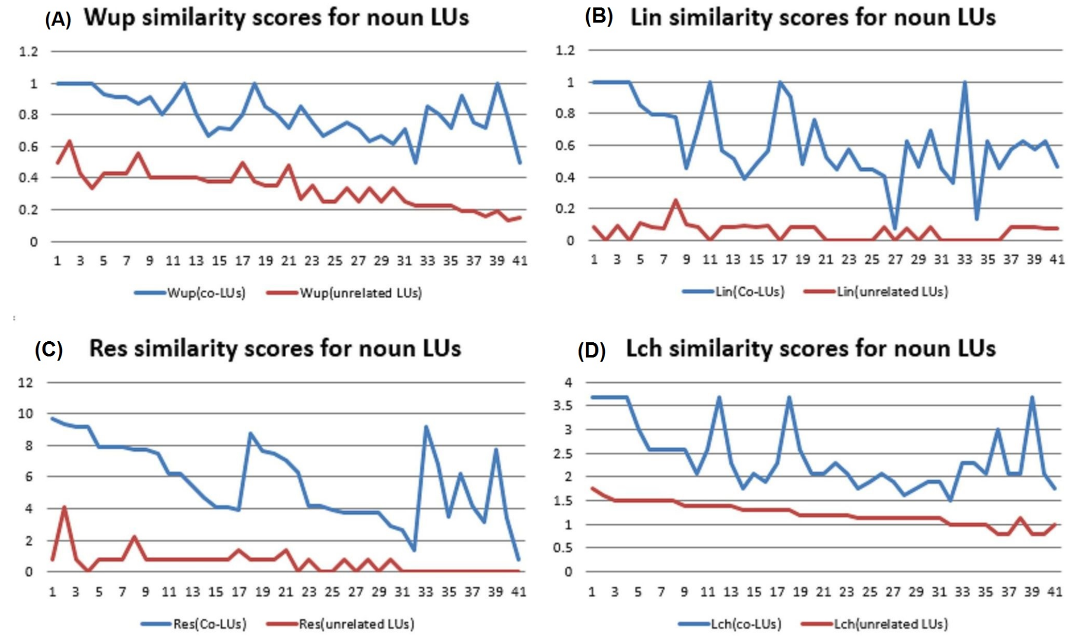

Figure 1: WN-based similarity between noun LUs

assigns to noun co-LUs is 0.5. The majority of unrelated nouns have a similarity range from 0.4 to 0.1, with only a few exceptions, according to WUP.

Lin, however, retrieves a wider range of similarity values to noun co-LUs. Figure 1b clarifies that many co-LUs have the maximum similarity score (i.e., 1) whereas other noun co-LUs are assigned relatively low similarity values (e.g., 0.4, 0.3, 0.1 and 0.07). The decrease in the similarity scores with unrelated nouns is salient, although the ranges of the similarity values between noun co-members and unrelated noun LUs intersect at some points.

Figure 1c displays Res similarity scores, which, in theory, range from 0 to the number of the tokens in a corpus. Based on the WN database, the highest similarity value between the studied noun pairs is 9.6, while the lowest one is 0, as reflected in the x-axis. Again, the quantitative gap between co-LUs and unrelated LUs (with regard to nouns) is obvious. The same is applicable to Figure 1d, which shows the variation between the similarity scores LCH calculates for related and unrelated noun pairs.

The statistical anti-correlation between these values mirrors the linguistic-based distinction between co-members and unrelated LUs in FN. Although there is no fixed WN relation that can be detected among all co-LUs, the absence of any WN relation among most unrelated LUs is evident. The similarity scores among unrelated LUs declined considerably according to Res, WUP, LCH and Lin computations. Furthermore, the negative correlation between Res similarity scores among co-members and those among unrelated LUs was statistically significant. This further recommends Res as the most compatible measure with DFCM for nouns.

Distributional similarity among nouns in enTenTen also correlated significantly with the DFCM. DFCM does not accurately reflect the polysemous and monosemous nature of words. A polysemous word may evoke a single frame in the FN database because of its incompleteness. Accordingly, the retrieved distributional similarity corresponding to high DFCM were, in several cases, considerably low. It descended to 0 for most of the nouns that are not related in FN. However, the fuzziness of the distributional similarity relation was detected when co-LUs were assigned zero similarity scores and unrelated LUs were assigned relatively high distributional similarity. BNC was not beneficial in obtaining similarity scores because it retrieved 0 values for numerous nouns that were related according to the DFCM and other WN-based similarity scores. Table 3 displays the distributional similarity between noun pairs as retrieved from enTenTen and BNC.

Pair of LUs	DFCM	D.S in enTenTen	D.S in BNC
Addict.n- habit.n	0.75	0.071	0
Attempt.n- effort.n	1	0.415	0.271
Bracelet.n- chain.n	0.666	0.117	0.05
Shop.n- practice.n	0.4167	0.267	0.16
Accident.n- backfire.n	0	0	0
Behavior.n- deceit.n	0	0	0
Company.n- age.n	0	0.319	0.155
Order.n- disarray.n	0.555	0	0

Table 3: Examples of DFCM and distributional similarity among nouns

This correlation supports the potential role of enTenTen fostered by, at least, the Res measure in the automatic expansion of FN's lexicon when it comes to nouns. The implementation of the POS-sensitive distributional algorithm on enTenTen can retrieve candidate members for a frame. Calculating the similarity between the retrieved candidate and the target LU using Res would further filter the results to preserve the most potential candidates only. These steps should save the lexicographic effort partially and speed up the enrichment of the lexicon without decreasing the accuracy of the database.

5.2 DFCM Among Adjectives and Verbs

For adjectives, DFCM displayed weak negative correlations with LESK and the similarity scores obtained from both BNC and enTenTen. The correlation was statistically significant only with the BNC similarity scores. This undermines the potential role of BNC in the automatic distributional-based expansion of FN's lexicon. The context in BNC was not rich enough to drive distributional similarity scores for some words and, accordingly, they were assigned 0 similarity scores. Furthermore, many synonymous words did not appear among the 5000 most distributionally similar words. On the contrary, the distributional algorithm, when applied to BNC, assigned non-zero similarity values to unrelated adjectives in FN. Table 4 lists a sample of FN's adjectives and their similarity scores in BNC.

Pair of LUs	DFCM	D.S in BNC
Able.a- capable.a	1	0.217
Amazing.a- astounding.a	1	0
Captivating.a- interesting.a	0.75	0
Following.a- early.a	0.666	0
Good.a- terrific.a	0.8	0
Able.a- accurate.a	0	0.06
Dark.a- fresh.a	0	0.216
Fresh.a- pale.a	0	0.17

Table 4: Examples of DFCM for adjectives and the corresponding distributional similarity in BNC

In the case of verbs, all correlations between DFCM scores and WN-based measures were not statistically significant, and neither were the correlations between DFCM and distributional similarity scores obtained from BNC and enTenTen.

Few studies used Sketch Engine's algorithms in the qualitative exploration of FN for lexicographic purposes. Atkins et al. (2003b) compare the word sketch of *argue*.v its entries in FN. They argue that the word sketch is obviously more suggestive of the target word's valence patterns and senses than the KWIC approach. The separation of the senses of *argue*.v in FN can be deduced from its word sketch in the BNC. In addition, the manual processing of word sketch results requires considerably less time and effort from the lexicographer than processing all or most of the concordance lines.

In the same vein, Abdelzaher and Tóth (2020) use Sketch Engine's distributional algorithm to compare the distributional similar words to *crime*.n to its related words in FN qualitatively. The new words that are distributionally similar to *crime*.n expose new senses, of the word, which are absent from the database of FN and from dictionaries. The multifaceted approach which integrates FN's database and distributional similarity motivates the proposal of four criteria (i.e., top-level frame, FEs, distributionally similar words and lexical preference) to differentiate between the five senses of *crime*.n systematically. The approach is also recommended to be part of the manual process of constructing frames because it does not compromise the precision of the database and, at the same time, saves the lexicographer a lot of time and effort wasted in the manual exploration of concordances.

6. Conclusion

To conclude, numerical values can be assigned to co-LUs in FN based on frame membership. The proposed DFCM is totally based on the FN database and it bears the consequences of this. It embraces the unique features of Frame Semantics and does not account for any frame-external data. It, accordingly, preserves the individual approach of the theory and the distinctive criteria for word grouping. Although the proposed DFM is by no means equivalent to the intuitive judgement of FN annotators, it reflects the information stored in FN and it is as dynamic as the FN database is. It can accommodate the continuous changes in the relation between lemmas, LUs and frames. At the theoretical level, there is no lexical or conventional relation among LUs with the highest DFCM, which is typical of the frame semantic approach. In addition, opposite words and words belonging to the same POS can enjoy higher DFCM than the derivational forms of the same base (e.g., DFCM for *happy*.a and *happily*.n is 0 because they have no shared frames). At the practical level, FN's database is under development, and sometimes a single sense is recorded for polysemous words. This drawback results in assigning the same DFM to monosemous and (poorly covered) polysemous words. That is to say, the DFCM score reflects both the uniqueness and the drawbacks of FN.

Statistical correlations yield promising results only for nouns and, therefore, the suggested approach can be more suitable to speeding up the process of frame constructions than to expanding the FN database automatically. The use of Rychlý's and Kilgarriff's (2007) distributional algorithm can provide the lexicographer with a list of potential co-LUs and, more importantly, present word sketches to justify their similarity at the grammatical and collocational levels. The possible qualitative interpretation of the distributional results offered by word sketches maintains the qualitative precision of FN's database. Moreover, this correlation recommends that future attempts to develop FN may benefit from a POS-sensitive algorithm and classification of data. Future research can also make use of Word2Vec distributional measure although it does not enjoy the same interpretative power offered by the word-sketch based distributional measure.

References

Abdelzaher, E. and Á. Tóth (2020), Defining crime: A multifaceted approach based on lexicographic relevance and distributional semantics, *Argumentum* **16**, pp. 44–63.

Atkins, S., C. Fillmore, and C. Johnson (2003a), Lexicographic relevance: Selecting information from corpus evidence, *International Journal of Lexicography* **16** (3), pp. 251–280.

Atkins, S., M. Rundell, and H. Sato (2003b), The contribution of framenet to practical lexicography, *International Journal of Lexicography* **16** (3), pp. 333–357.

Baisa, V. and V. Suchomel (2014), Skell: Web interface for english language learning, *Proceedings of the Recent Advances in Slavonic Languages Natural Language Processing*, Brno, The Czech Republic.

Baker, C. (2012), Framenet, current collaborations and future goals, *Language Resources and Evaluation* **46** (2), pp. 269–286.

Baker, C., C. Fillmore, and B. Cronin (2003), The structure of the framenet database, *International Journal of Lexicography* **16** (3), pp. 281–296.

Banerjee, S. and T. Pedersen (2002), An adapted lesk algorithm for word sense disambiguation using wordnet, *International conference on intelligent text processing and computational linguistics*, Mexico City, Mexico, p. 136–145.

Chow, I.C. and T. Wong (2006), Axiomatizing relational network for knowledge engineering - exploring wordnet and framenet, *Proceedings of the 2006 IEEE International Conference on Information Reuse and Integration*, Hawaii, USA.

De Bleecker, I. (2005), Towards an optimal lexicalization in a natural-sounding portable natural language generator for dialog systems, *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, Michigan, USA.

Ferrández, O., M. Ellsworth, R. Munoz, and C.F. Baker (2010), Aligning framenet and wordnet based on semantic neighborhoods, *Language Resources and Evaluation* **10**, pp. 310–314.

Fillmore, C., Johnson C., and Petruck M. (2003a), Background to framenet, *International Journal of Lexicography* **16** (3), pp. 235–250.

Fillmore, C.J., C.F. Baker, and H. Sato (2004), Framenet as a" net", *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, Lisbon, Portugal.

Fillmore, C.J., M.R. Petruck, J. Ruppenhofer, and A. Wright (2003b), Framenet in action: The case of attaching, *International journal of lexicography* **3** (3), pp. 297–332.

Firth, J.R. (1957), A synopsis of linguistic theory, *Studies in linguistic analysis*, Blackwell, Oxford, pp. 1–32.

Hirst, G. and D. St.Onge (1995), Lexical chains as representations of context for the detection and correction of malapropisms, *in* Fellbaum, C., editor, *WordNet: An Electronic Lexical Database*, MIT Press, Massachusetts, p. 305–332.

Jiang, J. and D. Conrath (1997), Semantic similarity based on corpus statistics and lexical taxonomy, *Proceedings of the 10th Research on Computational Linguistics International Conference*, Taipei, Taiwan.

Kilgarriff, A. and G. Duvuru (2011), Large web corpora for indian languages, *Proceedings of the International Conference on Information Systems for Indian Languages*, Patiala, India.

Kleinbauer, T. and T.A. Trost (2018), Comparing distributional and frame semantic properties of words, *Proceedings of 14th Conference on Natural Language Processing - KONVENS 2018*, Viena, Austria.

Laparra, E., G. Rigau, and M. Cuadros (2010), Exploring the integration of wordnet and framenet, *Proceedings of the 5th Global WordNet Conference*, Mumbai, India.

Leacock, C. and M. Chodorow (1998), Combining local context and wordnet similarity for word sense identification, *in* Fellbaum, C., editor, *WordNet: An Electronic Lexical Database*, MIT Press, Massachusetts, pp. 265–283.

Lenci, A. (2008), Distributional semantics in linguistic and cognitive research, *Italian journal of linguistics* **20** (1), pp. 1–31.

Lin, D. (1997), Automatic retrieval and clustering of similar words, *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics*, Montreal, Canada.

Miller, G. (1998), Nouns in wordnet. wordnet: An electronic lexical database, *in* Fellbaum, C., editor, *WordNet: An Electronic Lexical Database*, MIT Press, Massachusetts, p. 23–46.

Moor, T., M. Roth, and A. Frank (2013), Predicate-specific annotations for implicit role binding: Corpus annotation, data analysis and evaluation experiments, *Proceedings of the 10th International Conference on Computational Semantics*, Potsdam, Germany.

Pennacchiotti, M., D. De Cao, P. Marocco, and R. Basili (2008), Towards a vector space model for framenet-like resources, *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, Marrakech, Morocco.

Resnik, P. (1995), Using information content to evaluate semantic similarity in a taxonomy, *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montreal, Canada.

Ruppenhofer, J., M. Ellsworth, M. Schwarzer-Petruck, C.R. Johnson, and J. Scheffczyk (2006), Framenet ii: Extended theory and practice. https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf.

Rychlý, P. and A. Kilgarriff (2007), An efficient algorithm for building a distributional thesaurus (and other sketch engine developments, *Proceedings of the 45th annual meeting of the ACL*, Prague, Czech Republic.

Stará, M. and V. Kovár (2016), Options for automatic creation of dictionary definitions from corpora, *Proceedings of the Recent Advances in Slavonic Languages Natural Language Processing.*

Tonelli, S. and D. Pighin (2009), New features for framenet: Wordnet mapping, *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, Boulder, Colorado.

Tóth, Á. (2014), The distributional compatibility relation, *Argumentum* **10**, pp. 588–599.

Tóth, Á. (2015), Exploring distributional semantics with the spacexplorer tool, *BAS British and American Studies* **21**, pp. 217–224.

Virk, S.M., P. Muller, and J. Conrath (2016), A supervised approach for enriching the relational structure of frame semantics in framenet, *Proceedings of the 26th International Conference on Computational Linguistics*, Osaka, Japan, p. 3542–3552.

Wu, Z. and M. Palmer (1994), Verbs semantics and lexical selection, *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, Las Cruces, New Mexico, USA.

Zakharov, V., S. Pivovarova, E. Gvozdyova, and N. Semenova (2020), Corpus methods and semantic fields: the concept of empire in english, russian and czech, *Proceedings of the III International Conference on Language Engineering and Applied Linguistics*, Saint Petersburg, Russia.