



## Article

# Towards an Empirical Understanding of *membawa* ‘bring’: Corpus Insights into Indonesian Light Verb Constructions

Danang Satria Nugraha<sup>1</sup>, Veronika Vincze<sup>2</sup>

<sup>1</sup>Sanata Dharma University, Indonesia

<sup>1,2</sup>University of Szeged, Hungary

## SUBMISSION TRACK

Received: April 24, 2024  
 Final Revision: September 03, 2024  
 Accepted: September 07, 2024  
 Available Online: September 25, 2024

## KEYWORDS

Corpus analysis, Indonesian, light verb constructions, *membawa* ‘bring’, morphosemantics

## CORRESPONDENCE

E-mail: [d.s.nugraha@usd.ac.id](mailto:d.s.nugraha@usd.ac.id)

## ABSTRACT

This study analyzes the usage of the verb *membawa* (lit. bring) in Indonesian light verb constructions (LVCs). LVCs are ubiquitous in many agglutinative languages, including Indonesian. These constructions offer a unique way of combining a semantically light verb, which lacks inherent meaning, with another element to convey a specific concept. For instance, *membawa pesan* (lit. bring a message), *membawa kemenangan* (lit. bring a victory), *membawa perasaan* (lit. bring the feeling), *membawa perubahan* (lit. bring the change), amongst others. This study utilized corpus-based techniques to conduct a quantitative and qualitative analysis of 150 LVCs from the Indonesian-Leipzig Corpora Collection (ILCC) corpus. The findings of the study revealed three key points. Firstly, LVCs demonstrated a wide range of morphological and semantic usage, including primary and advanced areas, synonymic pair, and noun core proposition areas. Secondly, the morphosemantic relationships between core nouns in LVCs could be identified in the relations between base-abstract (0.15), base-concrete (0.37), affixed-abstract (0.36), and affixed-concrete (0.13). Thirdly, constructions with synonymic pairs dominated the total LVCs (0.75). However, all LVCs demonstrated the core proposition of the noun, which varied between ‘peoples’ (0.23), ‘materials’ (0.19), ‘objects’ (0.13), ‘qualities’ (0.17), ‘actions’ (0.11), and ‘processes’ (0.17). The findings suggest that LVCs in Indonesian are not merely syntactic constructions but rather reflect the intricate relationships between the verb and the noun, which can be analyzed in terms of their semantic and morphological properties. These findings can be applied to areas such as machine translation and computational linguistics, where a comprehensive understanding of LVCs is essential for accurate and effective language processing. Further corpus-based study could focus on analyzing the syntax and discourse functions of LVCs *membawa*.

## I. INTRODUCTION

Light verb constructions (henceforth: LVCs) are a linguistically intriguing phenomenon in many languages. LVCs are not restricted to inflectional languages, where verbs undergo morphological changes to indicate tense, mood, and aspect, but also exist in agglutinative languages, where morphemes with separate meanings are combined to create elaborate words (Hanks et al., 2006; Si, 2021; Sundquist, 2020; Vincze et al., 2013).

This remarkable flexibility of LVCs enables them to conform to the grammatical structures of diverse languages while providing speakers with an applicable device to convey a sizeable range of meanings. In most cases, the core of an LVC consists of a semantically bleached verb, often conveying a general meaning such as English *do* as in *do business* or *make* as in *make a claim*, combined with another element, usually a noun, that carries the specific meaning of the construction (Giparaitė,

2015, 2024; Ronan, 2014; Ronan & Schneider, 2015). The interplay between the light verb and its complement allows for subtle variations in meaning and offers many grammatical possibilities.

LVCs are ubiquitous in many agglutinative languages, including Indonesian (Nugraha, 2022a, 2022b). These constructions offer a unique way of combining a semantically light verb, which lacks inherent meaning, with another element to convey a specific concept. In Indonesian, *membawa* 'bring' is an exciting light verb due to its versatility. It can act as a grammatical scaffold, seamlessly combining various complements to create semantic nuances. The verb's meaning depends on the complement it partners with, allowing it to take on a range of interpretations. For instance, *membawa pesan* (lit. bring a message) ( $\Sigma$ -occurrence 6,723) expresses a 'straightforward action of carrying something'. However, *membawa kemenangan* (lit. bring a victory) ( $\Sigma$ -occurrence 2,669) takes on a more abstract meaning, implying the 'action to win'. *Membawa* can also be used with complements like *perasaan* as in *membawa perasaan* (lit. bring the feeling) ( $\Sigma$ -occurrence 243) to describe 'conveying an emotion' or with *perubahan* as in *membawa perubahan* (lit. bring the change) ( $\Sigma$ -occurrence 40,357) to 'signify introducing a new state'. This adaptability makes *membawa* a valuable tool in the Indonesian language, enabling the expression of a rich tapestry of ideas and actions.

Furthermore, the concept of LVCs has been the subject of numerous studies across various languages (Avdeeva, 2017; Eshaghi & Karimi-Doostan, 2023a; Kettnerová, 2023; Vilas, 2017; Wang & Wu, 2020). These studies are typically categorized into three groups: those focusing on isolating languages, those examining agglutinative languages, and those investigating fusional languages. In isolating and fusional languages, LVCs are often a critical area of investigation. LVCs leave behind isolative traces that can significantly impact the grammatical structure of a language (Buckingham, 2013; Dai, 2016; Vilas, 2015). These traces are not solely morphological but also carry vital semantic and syntactic information. In a language with free-word order, the construction of LVCs is a distinctive feature of a linguistic phenomenon that demonstrates universal characteristics (García Salido & Garcia, 2023; Wang et al., 2023; Yim, 2020). Consequently, LVCs are a critical aspect of understanding the

complexities of natural language processing and can provide valuable insights into the grammatical structures of various languages.

Agglutinative languages, on the other hand, characterized by the formation of words by combining morphemes (Haspelmath, 2009; Plungian, 2001), provide a fascinating area of study for LVCs. In particular, when examined in a fixed-word order context like SVO (subject-verb-object), LVCs in agglutinative languages exhibit morphological and semantic indications. These indications are fundamental to understanding LVCs. Previous research on Indonesian LVCs, notably the corpus-based studies by Nugraha (2023, 2024a, 2024b), has focused on the morphosemantic features of verbs like *membuat* 'make', *memberi* 'give', and *memenuhi* 'meet'. Despite these contributions, an investigation into other Indonesian LVCs remains relatively unexplored. This research gap, coupled with the unique characteristics of agglutinative languages, provides a compelling impetus for a thorough academic inquiry into *membawa* LVCs.

The present study, moreover, is based on two theoretical principles: the idiomatic combination of verbs and nouns and the morphosemantics of LVCs (Fleischhauer, 2021; Gibbs Jr., 1992; Hüning & Schlücker, 2015; Siefiring, 2004). In simple terms, idiomatic pairs refer to how language lexical elements can join uniquely to represent meaning (Ježek, 2023). This representation of meaning is no longer lexical but is at the grammatical and idiomatic level. The grammatical level is achieved through two morphemic units, each fulfilling a syntactic function (Butt, 2010). For instance, in LVCs, the verb part is usually the predicate function, and the noun is the object function. Therefore, LVCs are found in the core constituents of clauses, namely in their predicative elements. On the other hand, the idiomatic status of LVCs results from the meaning structure, which tends to be dominated by aspects of creating new meaning. LVCs comprising verb and noun elements usually have abandoned lexical meaning, especially for verb elements (Hellan, 2023). The verbs in LVCs have had their meanings merged and replaced by the meaning of the noun as the core element. At this level, it is as if the verb is only used as a unit that fulfils a predicative function because the semantic core comes from the paired noun (Nolan, 2014). This pattern of creating semantic meaning contrasts the tendency for proper verbs to dominate conventional constructions, such

as verbal phrases. If verbal phrases place verb elements as the core of compositional semantics, LVCs are the opposite. Therefore, LVCs that tend to be idiomatic are often classified as part of multi-word expressions (hereafter: MWEs).

LVCs, indeed, are the subject of morphological and semantic studies for their morphosemantic properties. Morphology provides theoretical and analytical tools to understand how morphemic elements combine to form new constructions and identify the core constituent types of LVCs (Di Salvo, 2023; Lu & Huang, 2023a). Semantics, on the other hand, provides the basis for analyzing the meaning of LVCs (Brugman, 2001; Eshaghi & Karimi-Doostan, 2023b). In analyzing LVCs more than morphological approaches are required, meaning analysis becomes crucial. Meaning analysis is carried out based on the semantic paradigm, and three main domains play an essential role in this study. These domains are noun semantics, synonymic pairs, and noun semantic prototypes. Noun semantics refers to the fundamental status of noun meaning in language, synonymous pairing refers to the presence or absence of similar meanings of non-LVC constructs for LVCs, and noun prototypes refer to the differentiation of types of nouns at the core of LVCs.

The corpus linguistic approach employed in this study underscores the significance of empirical analysis in understanding language phenomena. By examining a large dataset of Indonesian LVCs, the researchers were able to identify patterns and trends that would have been difficult to discern through introspection or intuition alone. This approach provides a solid foundation for understanding the complex interplay between the verb *membawa* and its accompanying nouns, revealing the multifaceted nature of these constructions and their role in conveying meaning within the Indonesian. Moreover, the findings offer insights for applications such as machine translation and computational linguistics, where a deep understanding of LVCs is crucial for accurate and effective language processing. Admittedly, this research has also potential contributions to the field of linguistics, i.e. (a) enhancing the understanding of morphological and semantic properties of LVCs in Indonesian and (b) providing insights into the underlying principles that govern language structure and typological variation. Specifically, the novelty of this study lies in its analysis by providing a more in-depth and

data-driven examination of LVCs *membawa*, i.e. (a) providing a detailed classification of LVCs, (b) identifying morphosemantic relationships, and (c) analyzing the dominance of synonymic pairs.

The current study, therefore, aims to conduct an analysis of the morphological and semantic aspects of LVCs formed with *membawa*. To achieve this, the research employs a corpus-based approach to examine a specific language data set of 150 instances of LVCs *membawa*. The two main research questions that the study seeks to answer are as follows. Firstly, it investigates the distribution of LVCs *membawa* in various grammatical structures and contexts to identify usage patterns and regularities. Secondly, the study explores the interaction between morphosemantic features of Indonesian LVCs in a fusional construction. The research question canters on how the morphological features of the verbs and nouns that comprise the construction impact the meaning of LVCs.

## II. METHODS

### Study Design

The current study adopted a corpus-based methodology to investigate the usage of the verb *membawa* (lit. bring) in Indonesian LVCs. A corpus of 150 instances of LVCs *membawa* was extracted as the main data. The analysis employed a combination of quantitative and qualitative techniques, allowing for both statistical analysis and in-depth examination of individual LVCs. This approach ensured a systematic investigation of the morphological and semantic properties of the verb *membawa* in Indonesian LVCs.

### Materials

This study scrutinized 150 instances of the LVCs *membawa* data obtained from the Indonesian–Leipzig Corpora Collection (ILCC), an Indonesian mixed corpus based on data from 2013 (label: `ind_mixed_2013`). There are 1,206,281,985 tokens and 74,329,815 sentences in it (website: [https://corpora.uni-leipzig.de/en?corpusId=ind\\_mixed\\_2013](https://corpora.uni-leipzig.de/en?corpusId=ind_mixed_2013)). The data was categorized into three groups, namely T1 for high frequency occurrence, ranks 1 to 50; T2 for medium frequency occurrence, ranks 50-100; and T3 for low frequency occurrence, ranks 101-150, based on the rate level. As an analytical instrument, we prepared four matrices consisting of two sets of indicators. The first set was a morphological analysis instrument consisting

of M-1 (for determining the status of primary morphemic forms in LVC constructions) and M2 (for determining the status of basic noun-core types in LVC constructions). The second set was a semantic analysis instrument consisting of S-1 (for determining the status of synonymous pairs of LVCs) and S-2 (for determining the propositional status of core nouns in LVCs).

### Procedure and Data Analysis

The initial phase of the research involved data collection. The data consisted of LVCs *membawa* construction, retrieved from the ILCC based on three parameters. These parameters were as follows: (a) LVCs with lexical meaning of verb-part were removed from the data, (b) nouns in the form of phrases were still considered as data that was suitable for analysis, and (c) each verb and noun pair had frequencies or tokens based on the corpus used. However, not all data collected was considered helpful for analysis, and data reduction was conducted to sort out valid data. The valid data was then categorized based on the frequency of occurrence. At the same time, a simple statistical description (see Table 1) was created to support the analysis phase.

The second stage of the research involved data analysis. The analysis was carried out on triangulated and clustered data based on the frequency of occurrence (see Table 2 for the high frequency, Table 3 for the medium frequency, and Table 6 for the low frequency)—the first analysis aimed to trace the distribution of LVCs in four different grammatical contexts. The first context was measured using the M-1 parameter (Blevins, 2016), which focused on tracing the distribution of LVCs *membawa*, the core noun's basic form and affixation form. The second context was measured using the M-2 parameter (Dal & Namer, 2015), which focused on tracing the distribution of LVCs on the primary classification spectrum of noun-core types. The third context was measured using the S-1 parameter (Hrenek, 2021), which focused on tracing the distribution of LVCs *membawa* and the construction of synonymous pairs. The fourth context was measured using the S-2 parameter (Aikhenvald, 2017), which focused on tracing the distribution of LVCs on the core proposition or semantic meaning of the core noun. These four types of analysis were used to describe the distribution of LVCs in Indonesian grammatical contexts.

In this research, the third stage involved presenting the results of our analysis. We presented the analysis results in two types: qualitative (Eatough & Tomkins, 2022) and quantitative (Gries, 2015; Sheng, 2023). Qualitatively, we provided a descriptive account of the distribution of LVCs *membawa*. Representative samples of LVCs complemented the descriptive presentation. The annotation framework is applied to portray a number of qualitative samples. UDPipe (website: <https://lindat.mff.cuni.cz/services/udpipe/>) was the primary tool employed for this technique. Quantitatively we also presented quantitative exposure as confirmation of the qualitative results carried out. Frequency analysis and statistical tests were employed to explore potential associations between distinct semantic interpretations and morphological attributes in LVCs *membawa*. We measured the extent of use (V(N)) and log potential productivity (P(N)) (Stupak & Baayen, 2022). Throughout the analysis process, we carried out extensive human-assisted checks to ensure correctness and consistency. Furthermore, we utilized an iterative refinement strategy that allowed us to continuously improve our analytic methods, coding schemes, and identification criteria in response to new patterns. This strategy made sure that our study had a solid data foundation and successfully addressed the relevant research objectives.

## III. RESULTS

### Distribution of LVCs *membawa*

This study used four different matrices to extensively investigate the prevalence of LVCs, including the verb *membawa* in Indonesian grammatical context (see Figure 1). M-1 and M-2 concentrated on morphological aspects, whereas S-1 and S-2 employed semantic indications to examine distribution patterns. Based on the M-1 indicator, the result suggests that the distribution of LVCs with *membawa* in Indonesian varies depending on whether the verb appears in its affixed or base form. LVCs with *membawa* in its affixed form appear a total of 72 times across the three tier phases (T-1, T-2, T-3), distributed across the three time periods (T-1: 19; T-2: 32; T-3: 21). This represents 48.00% of the total occurrences of LVCs with *membawa* observed in the data. LVCs with *membawa* in its base form appear a total of 78 times across the three time periods, distributed across the three time periods (T-1: 31; T-2: 18; T-3: 29). This

represents 52.00% of the total occurrences of LVCs with *membawa* observed in the data. Additionally, the data analyzed suggests that LVCs with the *membawa* + the base-form of noun were slightly more frequent than those with the affixed form (see Sample 1). However, it is essential to note that this is just one indicator out of the four employed in the study, and a more comprehensive understanding would require examining the data from the other indicators as well. The overall distribution of LVCs with *membawa* in its base form and the affixed form is relatively close, indicating that both forms are commonly used in Indonesian LVCs.

1) *Peristiwa Nagasaki dan Hiroshima membawa dampak bagi Indonesia* (ILCC, 2013).

Incident.NOUN Nagasaki.PROPN and.CCONJ Hiroshima.PROPN bring.VERB impact.NOUN for.ADP Indonesia.PROPN

‘The Nagasaki and Hiroshima events had an impact on Indonesia.’

2) *Nyepi membawa pesan kedamaian* (ILCC, 2013).

Nyepi.PROPN bring.VERB message.NOUN peace.NOUN

‘Nyepi brings a message of peace.’

Based on the M-2 indicator, the finding provided shows the distribution of LVCs with the verb *membawa* in Indonesian, according to the M-2 indicator. There were 76 LVCs with *membawa* classified according to the M-2 indicator. Of the

total, 21 LVCs were classified as Abstract in T-1, 27 in T-2, and 28 in T-3. The percentile for the abstract category is 50.67, which means that half of the LVCs with the M-2 indicator fell into the abstract category. On the other hand, 29 LVCs were classified as Concrete in T-1, 23 in T-2, and 22 in T-3, making a total of 74 LVCs. The percentile for the concrete category is 49.33, meaning the other half of the LVCs belonged to the concrete category. Based on the data, we can see that the total number of LVCs observed is almost the same for both Abstract and Concrete categories (76 vs. 74). There seems to be some variation in the distribution of LVCs across different periods (T-1, T-2, T-3). For example, there are more LVCs with *membawa* followed by a concrete complement in T-1 (29) compared to T-3 (22). The percentile values indicate that the proportion of Abstract and Concrete LVCs is relatively balanced across all periods, with Concrete LVCs having a slightly lower percentage overall (around 49%). While the total number of occurrences is close (76 for abstract vs. 74 for concrete), the percentile values reveal a slight bias towards abstract complements. This might indicate a tendency for *membawa* to be used more frequently with abstract concepts in the specific context of the corpus analyzed (see Sample 2). These categories likely refer to the type of complement following *membawa* in the LVCs. Abstract complements likely represent intangible

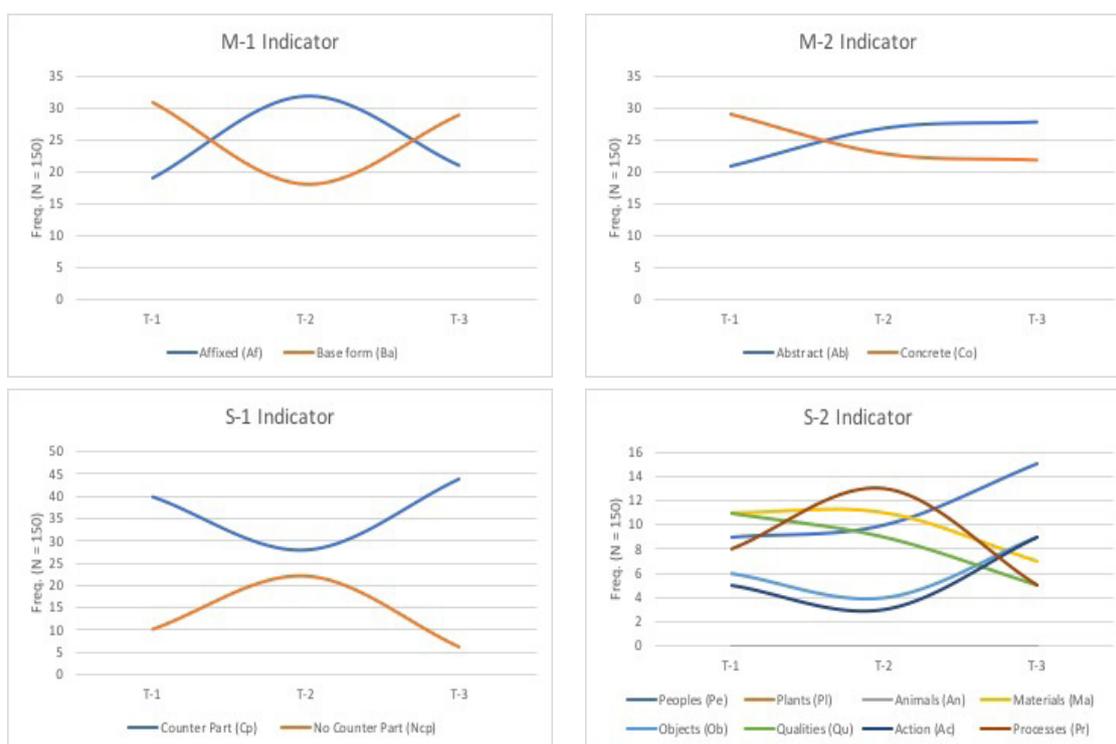


Figure 1. Four Circumstances were Employed to Assess the Distribution of LVCs with *membawa* in Indonesian Grammatical Context

concepts or ideas, while Concrete complements represent physical objects or entities.

Based on the S-1 indicator, the indicator refers to a specific semantic category being analyzed, and it is divided into two subcategories: Counter Part (Cp) and No Counter Part (Ncp). The former category refers to LVCs where *membawa* has a complete counterpart verb that can express the same meaning independently. The latter category refers to LVCs where *membawa* does not have a direct semantic equivalent and contributes essential meaning to the construction. The T-1, T-2, and T-3 categories reflect different contexts or scenarios in which the LVCs were analyzed. These refer to different grammatical constructions, registers, or genres where the LVCs were found. The Figure 1 shows the total number of LVCs observed for each subcategory across all three contexts. The percentile indicates the percentage of LVCs that fall into each subcategory within each context. For instance, 74.67% of LVCs observed in context T-1 belong to the Counter Part category, while the remaining 25.33% belong to the No Counter Part category. Also, the result indicates that most LVCs containing the verb *membawa* based on the S-1 indicator have a counterpart verb (Cp) and were found more frequently in the T-1 context than in T-2 and T-3. The Counter Part (Cp) category, which represents LVCs where *membawa* has a counterpart verb or verb phrase that complements it, is more prevalent across all three contexts compared to the No Counter Part (Ncp) category, which represents LVCs where *membawa* has no direct counterpart. The data suggests that the counterpart verb contributes significantly to the overall meaning of the construction and provides insights into the semantic flexibility and prevalence of different LVC constructions with *membawa* in Indonesian (see Sample 3). However, further information about the specific meaning S-1 and the context details of T-1, T-2, and T-3 is required to draw more conclusive results. Examples of LVCs with a counterpart verb include *membawa pergi* (lit. to bring away) or *membawa pulang* (lit. to bring back). In contrast, examples of LVCs with no direct counterpart include *membawa buku* (lit. to bring a book).

- 3) *Dan apakah akan membawa pengaruh yang cukup besar pada praktek pengadilan?* (ILCC, 2013).

And.CCONJ is.VERB will.AUX bring.VERB influence.NOUN which.PRON enough.ADV big.ADJ on.ADP practice.NOUN court.NOUN

'And will it have a significant impact on court practice?'

- 4) *Prestasi di dua event itu turut membawa nama Lopes <...>* (ILCC, 2013).

Performance.NOUN in.ADP two.NUM events.NOUN that.DET join.ADV bring.VERB name.NOUN Lopes.PROPN <...>

'Achievements in these two events also brought Lopes' name to adorn many products outside the racing arena.'

According to the S-2 indicator, People (Pe) are the most frequently used complement type with *membawa* in Indonesian, representing 22.67% (34 occurrences) of all complements (see Sample 4). On the other hand, Plant and Animal complements are rarely used with *membawa* based on the S-2 indicator, as these categories had zero occurrences. Materials are a relatively common type of complement for *membawa*, representing 19.33% (29 occurrences) of all complements, while objects represent a moderately frequent category with 12.67% (19 occurrences). Qualities are also a somewhat common type of complement for *membawa*, representing 16.67% (25 occurrences) of all complements, while actions are used occasionally with 11.33% (17 occurrences). Finally, processes are the second most frequent category of complements for *membawa* with 17.33% (26 occurrences) of all complements, according to the S-2 indicator. In details, based on the further analysis, S-2 Indicator categories can be grouped into three frequency ranges: High frequency (over 20%), which includes Peoples (Pe); Medium frequency (10-20%), which includes Materials (Ma), Qualities (Qu), and Processes (Pr); and Low frequency (below 10%), which includes Objects (Ob), Action (Ac), Plants (Pl), and Animals (An). The most frequent categories of complements with *membawa* according to S-2 Indicator are Processes (Pr) (17.33%), Qualities (Qu) (16.67%), and Materials (Ma) (19.33%). On the other hand, the least frequent categories are Plants (Pl), Animals (An), and Objects (Ob), with none found in T-1 and T-2, and a total of 0 occurrences across all text types. Despite Processes and Qualities being the most frequent, the distribution across categories is somewhat balanced, with all categories except Plants and Animals having at least some representation.

### Relation of Noun Semantic-head in the LVCs *membawa*

The analysis reveals four distinct paired occurrences of the relation (see Figure 2), which

Table 2. The First Tier of Occurrence

TIER 1					
LVCs	Freq.	Rank	LVCs	Freq.	Rank
<i>membawa dampak</i> 'have an impact'	78,721	1	<i>membawa kasus</i> 'bring the case'	2,521	26
<i>membawa perubahan</i> 'bring change'	40,357	2	<i>membawa suasana</i> 'brings atmosphere'	2,479	27
<i>membawa manfaat</i> 'bring benefits'	24,749	3	<i>membawa keberkahan</i> 'bring blessings'	2,275	28
<i>membawa berkah</i> 'bring blessings'	24,418	4	<i>membawa bukti</i> 'bring evidence'	2,269	29
<i>membawa nama</i> 'carry the name'	22,437	5	<i>membawa akibat</i> 'bring consequences'	2,122	30
<i>membawa pengaruh</i> 'bring influence'	16,695	6	<i>membawa kehancuran</i> 'bring destruction'	2,117	31
<i>membawa konsekuensi</i> 'bring consequences'	13,464	7	<i>membawa kemaslahatan</i> 'bring benefit'	2,115	32
<i>membawa misi</i> 'carrying a mission'	10,091	8	<i>membawa semangat</i> 'brings enthusiasm'	2,073	33
<i>membawa kebaikan</i> 'brings goodness'	9,258	9	<i>membawa kerugian</i> 'bring losses'	1,995	34
<i>membawa hasil</i> 'bring results'	9,229	10	<i>membawa rahmat</i> 'brings grace'	1,926	35
<i>membawa keberuntungan</i> 'bring luck'	8,335	11	<i>membawa kemakmuran</i> 'brings prosperity'	1,825	36
<i>membawa implikasi</i> 'carries implications'	8,040	12	<i>membawa damai</i> 'bring peace'	1,816	37
<i>membawa korban</i> 'carrying the victim'	7,456	13	<i>membawa bantuan</i> 'bring help'	1,808	38
<i>membawa pesan</i> 'carrying a message'	6,723	14	<i>membawa aspirasi</i> 'bring aspirations'	1,758	39
<i>membawa kemajuan</i> 'bring progress'	5,412	15	<i>membawa berita</i> 'bringing news'	1,637	40
<i>membawa beban</i> 'carry the load'	4,230	16	<i>membawa harapan</i> 'brings hope'	1,580	41
<i>membawa kebahagiaan</i> 'bring happiness'	4,103	17	<i>membawa hoki</i> 'bring luck'	1,533	42
<i>membawa kabar</i> 'bring news'	3,994	18	<i>membawa hikmah</i> 'brings wisdom'	1,479	43
<i>membawa efek</i> 'bring effect'	3,715	19	<i>membawa persoalan</i> 'bring problems'	1,467	44
<i>membawa petaka</i> 'bring disaster'	3,378	20	<i>membawa bencana</i> 'bring disaster'	1,377	45
<i>membawa masalah</i> 'bring trouble'	3,316	21	<i>membawa sial</i> 'bring bad luck'	1,369	46
<i>membawa kedamaian</i> 'bring peace'	2,900	22	<i>membawa konsekuensi</i> 'bring consequences'	1,359	47
<i>membawa malapetaka</i> 'bring disaster'	2,682	23	<i>membawa kesan</i> 'make an impression'	1,257	48
<i>membawa kemenangan</i> 'bring victory'	2,669	24	<i>membawa nuansa</i> 'bring nuance'	1,236	49
<i>membawa kesejahteraan</i> 'bring prosperity'	2,580	25	<i>membawa ketenangan</i> 'bring calm'	1,171	50
$\Sigma$	318,952		$\Sigma$	44,564	

are classified as follows: (a) base – abstract, (b) affixed – abstract, (c) base – concrete, and (d) affixed – concrete. These categories exhibit a distribution relationship, which can be quantified using two parameters. The first parameter measures the vocabulary size or number of word types ( $V(N)$ ) that appears in a sample of  $N$  tokens. The second parameter ( $P(N)$ ) captures the rate at which the vocabulary size increases as more tokens are sampled.

According to the first category of paired occurrence (Base & Abstract) as in sentence sample (5), Figure 2 shows the morphological productivity of a specific morphological category across three different text types of frequency (T-1, T-2, T-3). It compares two measures:  $\text{Log}(P(N))$  and  $\text{Log}(V(N))$ . Lower values of  $\text{Log}(P(N))$  indicate a lower probability of finding new word types as the sample grows, suggesting a limited pool of possible word types for the category. In this case, all three text types have negative values for  $\text{Log}(P(N))$ , meaning that the probability of encountering new word types is relatively low. Higher values of  $\text{Log}(V(N))$  indicate a more extraordinary richness in word types, suggesting a more productive morphological category. Here, the values are higher than those in  $\text{Log}(P(N))$ , indicating that the category has a decent number of word types present in the samples, even though the probability of finding new ones is low. The

morphological category is moderately productive based on the higher values in  $\text{Log}(V(N))$  compared to  $\text{Log}(P(N))$ . This suggests that the category has a reasonable number of distinct word types, but the likelihood of encountering entirely new ones is low as the sample size increases. There is some variation in productivity across text types. T-2 has the lowest values in both metrics, indicating potentially lower productivity in that specific text type compared to T-1 and T-3.

In analyzing the vocabulary size and probability of encountering new word types in three different text types, T-1 and T-3 were found to have a higher vocabulary size of around 0.90 compared to T-2, with a vocabulary size of around 0.78. This indicates that T-1 and T-3 have more word types in their respective samples. All three text types have negative values for the probability of encountering a new word type when the sample size increases. This indicates that the likelihood of finding a new word type decreases as the sample size grows. T-2 had the most negative value of -0.56, suggesting the lowest probability of finding new word types compared to T-1 and T-3, which have values of -0.44. The difference in vocabulary sizes might be minor than the values suggest due to the logarithmic scale. These findings highlight the differences in vocabulary size and probabilities of encountering new word types in different text types.

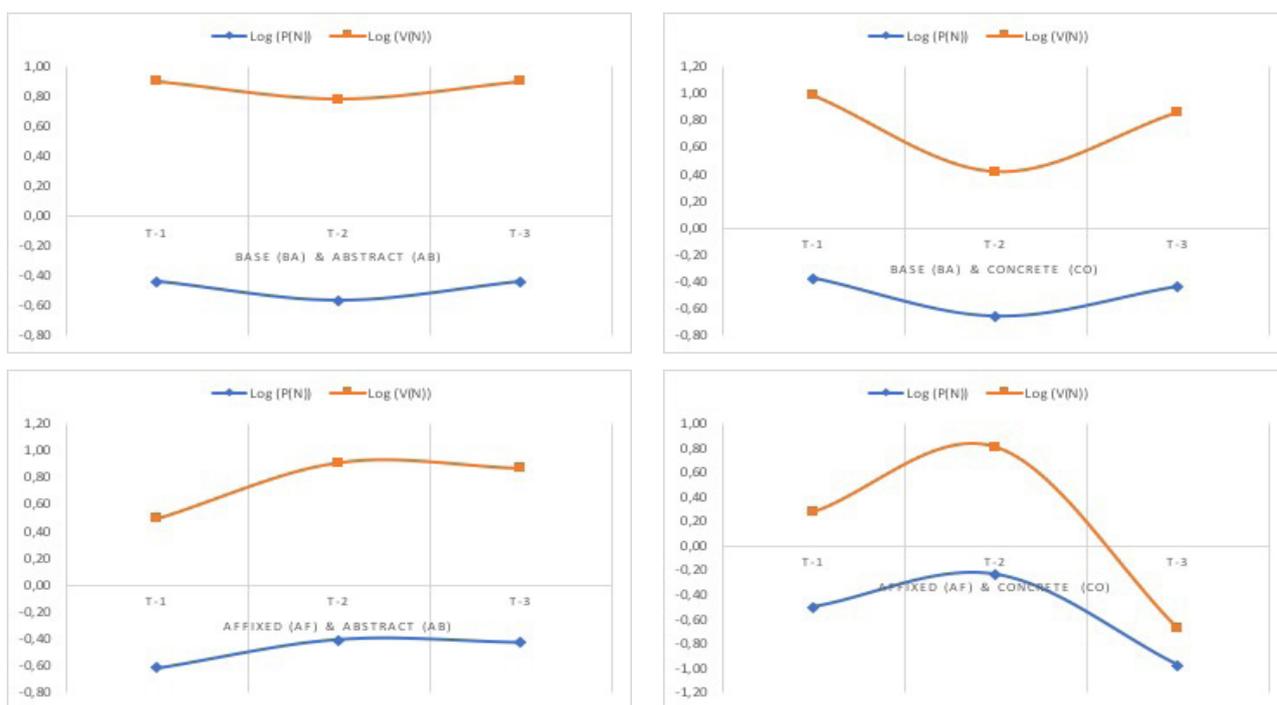


Figure 2. Four Distinct Paired-Occurrences of the Relation Measured by  $\text{Log}(P(N))$  and  $\text{Log}(V(N))$

5) *Sementara IM cukup ketakutan aksinya membawa petaka* (ILCC, 2013).

Temporary.SCONJ IM.PROPN enough.ADV afraid.  
ADJ action.NOUN bring.VERB disaster.NOUN

'Meanwhile, IM was quite afraid that his actions would bring disaster.'

According to the second category of paired occurrence (Affixed & Abstract) as in sentence sample (6), three text types were evaluated. T-1 had a vocabulary richness of 1.11 and a lower probability of encountering new word types with a log value of -0.62. T-2 had the richest vocabulary among the three, with a log value of 1.32 and a slightly higher probability of encountering new word types with a log value of -0.41 compared to T-1. T-3 had a similar vocabulary richness of 1.30 to T-2 and a similar probability of encountering new word types with a log value of -0.43. The analysis of three text types (T-1, T-2, T-3) revealed an increasing trend in vocabulary size (V(N)) with positive values of Log (V(N)). T-2 showed the highest vocabulary size, followed by T-3 and T-1. All three text types showed negative values for Log (P(N)), indicating a decreasing rate at which new word types are encountered with larger sample sizes. This suggests that the vocabulary richness is gradually approaching saturation and encountering entirely new word types becomes less likely. While all text types showed a similar trend of decreasing growth rate, the specific values for Log (P(N)) and Log (V(N)) differed, suggesting variations in vocabulary richness and growth patterns between the text types. T-1 had the lowest vocabulary size and the fastest decrease in encountering new word types, while T-2 had the highest vocabulary size and a slower decrease. T-2 and T-3 exhibit the richest vocabulary (highest Log (V(N))) for the morphological category under study. T-1 had the lowest probability (most negative Log (P(N))) of encountering new word types, suggesting a more limited variation in word forms compared to T-2 and T-3. While T-2 and T-3 had similar vocabulary richness, T-2 showed a slightly higher chance of encountering new word types than T-3.

The key observations from the analysis of three text types (T-1, T-2, and T-3) suggest that T-2 and T-3 exhibit the richest vocabulary (highest Log (V(N))) for the morphological category under study. All three text types show a positive value for Log (V(N)), indicating the presence of diverse word types. However, T-2 and T-3 have higher Log (V(N)) values than T-1, suggesting a

more incredible vocabulary richness in those text types. Additionally, T-1 has the lowest probability (most negative Log (P(N))) of encountering new word types, suggesting a more limited variation in word forms compared to T-2 and T-3. All three text types have negative Log (P(N)) values, but a negative value does not necessarily imply a low probability. Instead, it indicates a minor difference in probability compared to a reference value (usually  $P(N) = 1$ ). Comparing the values across text types, T-1 has the most negative value for Log (P(N)), indicating a relatively lower probability of encountering new word types than T-2 and T-3. This suggests that T-1 might have a less productive morphology, meaning new word types are formed less frequently in that context. In conclusion, the data suggests that T-2 and T-3 exhibit a richer vocabulary (higher Log (V(N))) and potentially a greater degree of morphological productivity (less negative Log (P(N))) compared to T-1.

6) *Di Jepang, Cina, dan bagian lain di Timur Jauh, kalung ikan ini membawa keberuntungan, kekayaan, dan kebahagiaan* (ILCC, 2013).

In.ADP Japan.PROPN China.PROPN and.CCONJ part.  
NOUN other.ADJ in.ADP Far.PROPN East.PROPN  
fish.NOUN necklace.NOUN this.DET bring.VERB  
luck.NOUN riches.NOUN and.CCONJ happiness.  
NOUN

'In Japan, China, and other parts of the Far East, this fish necklace brings good luck, wealth, and happiness.'

According to the third category paired occurrence (Base & Concrete) as in sentence sample (7), upon analyzing the data, it was observed that the vocabulary size in T-1 was the highest among the three text types, with a value of 1.36, indicating a more significant number of unique word types present in T-1 compared to T-2 and T-3. On the other hand, T-2 had the lowest vocabulary size of 1.08, suggesting a smaller number of unique word types. T-3, with a vocabulary size of 1.30, was closer to T-1 than T-2. Negative values were observed for Log (P(N)) across all three text types, indicating that the probability of encountering new word types was less than 1. This means it was less likely to find new words as the sample size increased. Among the three text types, T-2 had the most negative value (-0.66), suggesting the lowest probability of encountering new word types compared to T-1 and T-3. T-1 and T-3 had similar Log (P(N)) values (-0.38 and -0.44, respectively), indicating a relatively comparable likelihood of finding new word types when the sample size

Table 3. The Second Tier of Occurrence

TIER 2					
LVCs	Freq.	Rank	LVCs	Freq.	Rank
<i>membawa kesuksesan</i> 'bring success'	1,169	51	<i>membawa kelengkapan</i> 'bring equipment'	622	76
<i>membawa kenangan</i> 'brings memories'	1,147	52	<i>membawa perubahan-perubahan</i> 'bring about changes'	595	77
<i>membawa perbaikan</i> 'bring improvements'	1,139	53	<i>membawa kematian</i> 'brings death'	579	78
<i>membawa mudharat</i> 'bring harm'	1,061	54	<i>membawa kemanfaatan</i> 'bring benefits'	568	79
<i>membawa kesengsaraan</i> 'brings misery'	1,060	55	<i>membawa kesejukan</i> 'bring coolness'	560	80
<i>membawa pencerahan</i> 'bring enlightenment'	1,059	56	<i>membawa petunjuk</i> 'bring directions'	548	81
<i>membawa ajaran</i> 'bring teachings'	1,054	57	<i>membawa keberhasilan</i> 'brings success'	545	82
<i>membawa perdamaian</i> 'bring peace'	1,049	58	<i>membawa aura</i> 'carries an aura'	543	83
<i>membawa berkat</i> 'brings blessings'	1,010	59	<i>membawa makna</i> 'carry meaning'	536	84
<i>membawa persembahan</i> 'bring offerings'	963	60	<i>membawa pengalaman</i> 'bring experience'	521	85
<i>membawa penderitaan</i> 'brings suffering'	861	61	<i>membawa kesialan</i> 'bring bad luck'	519	86
<i>membawa sukacita</i> 'brings joy'	820	62	<i>membawa pembaca</i> 'bring the reader'	513	87
<i>membawa atribut</i> 'carries attributes'	803	63	<i>membawa nilai-nilai</i> 'bring values'	513	88
<i>membawa keselamatan</i> 'brings safety'	737	64	<i>membawa keceriaan</i> 'brings joy'	481	89
<i>membawa kebenaran</i> 'bring the truth'	727	65	<i>membawa persyaratan</i> 'bring requirements'	473	90
<i>membawa sentiment</i> 'brings sentiment'	725	66	<i>membawa penyakit</i> 'carry disease'	465	91
<i>membawa risiko</i> 'carries risks'	717	67	<i>membawa kemuliaan</i> 'bring glory'	460	92
<i>membawa diri</i> 'bring yourself'	717	68	<i>membawa kesegaran</i> 'bring freshness'	442	93
<i>membawa pesan-pesan</i> 'carrying messages'	706	69	<i>membawa maut</i> 'brings death'	436	94
<i>membawa kegembiraan</i> 'bring joy'	686	70	<i>membawa duka</i> 'bring sorrow'	435	95
<i>membawa citra</i> 'carries an image'	653	71	<i>membawa kerusakan</i> 'bring damage'	432	96
<i>membawa rezeki</i> 'bring fortune'	650	72	<i>membawa darah</i> 'carrying blood'	427	97
<i>membawa maksud</i> 'carry meaning'	649	73	<i>membawa imbas</i> 'bring impact'	420	98
<i>membawa maslahat</i> 'bring benefits'	638	74	<i>membawa rejeki</i> 'bring good fortune'	415	99
<i>membawa kejayaan</i> 'bring success'	635	75	<i>membawa keburukan</i> 'bring badness'	411	100
$\Sigma$	21,435		$\Sigma$	12,459	

increased in these two text types. It was also observed that the probability of encountering a new word type was relatively low across all text types, with T-2 having the lowest value (-0.66), suggesting the highest likelihood of encountering a new word type in that text type compared to the others. All Log (V(N)) values were positive, indicating a positive relationship between the sample size and the number of observed word types. This means that the vocabulary size also grew as the sample size increased. T-1 had the highest value of 1.36, suggesting the largest vocabulary size among the three text types.

In details, this analysis compares three text types based on their vocabulary size and probability of encountering new word types. The first text type, T-1, shows the largest vocabulary size and a relatively high probability of encountering new word types, suggesting it might be a richer source of diverse word forms. The second text type, T-2, has the lowest vocabulary size and the lowest probability of finding new word types, indicating it might contain fewer unique word forms and have a higher degree of repetition. The third text type, T-3, shows a vocabulary size closer to T-1 and a probability of new words similar to T-1, suggesting it might have a moderate level of lexical diversity and a somewhat lower chance of encountering completely new words compared to T-1.

7) *Kali ini, aku dan Brilian membawa berita yang bahagia* (ILLC, 2013).

Time.NOUN this.DET I.PS1 and.CCONJ Brilliant.  
PROPON bring.VERB news.NOUN which.PRON  
happy.ADJ

‘This time, Brilian and I brought happy news.’

According to the fourth category paired occurrence (Affixed & Concrete) as in sentence sample (8), based on the breakdown of data for each text type, it can be stated that T-1 has a low probability of encountering new noun types as the sample size increases, as indicated by a Log (P(N)) of -0.50. However, T-1 has a relatively high number of unique noun types observed in the sample, with a Log (V(N)) of 0.78. T-2, on the other hand, has a moderately low probability of encountering new noun types compared to T-1, with a Log (P(N)) of -0.24, but shows the highest number of unique noun types observed among the three text types, with a Log (V(N)) of 1.04. T-3, however, has the lowest probability of encountering new noun types among the three text types, with a Log (P(N)) of -0.98, and

the lowest number of unique noun types observed, with a Log (V(N)) of 0.30. These findings suggest that each text type has its unique characteristics and patterns that can be observed and analyzed.

8) *Penalti kembali membawa kemenangan buat Persib* (ILCC, 2013).

Penalty.NOUN return.ADV bring.VERB victory.  
NOUN for.ADP Persib.PROPON

‘Another penalty brought victory for Persib.’

The data suggests that T-2 has the highest overall morphological productivity for the measured category, with both a higher number of unique noun types observed (Log (V(N)) = 1.04) and a moderately higher probability of encountering new types (Log (P(N)) = -0.24) compared to the other text types. In contrast, T-3 has the lowest overall morphological productivity, with both a low number of unique noun types observed (Log (V(N)) = 0.30) and a very low probability of encountering new types (Log (P(N)) = -0.98). This suggests that the vocabulary used in T-3 might be limited and repetitive. T-1 has a relatively high number of unique noun types observed (Log (V(N)) = 0.78) but a low probability of encountering new types (Log (P(N)) = -0.50), suggesting that the vocabulary richness in T-1 might be due to the presence of frequent, common nouns. The data also suggests that the vocabulary size (number of word types) in T-2 is the largest, followed by T-1 and then T-3. This indicates that T-2 exhibits the most diverse use of the morphological category, while T-3 shows the least diversity. However, the probability of encountering new word types across all three text samples could be higher, especially in T-3. This indicates that the morphological category might be relatively closed, meaning that new word types are unlikely to be formed. It is essential to consider the logarithmic scale used here, as a slight change in the log value can represent a more considerable change in the actual probability. Overall, T-2 appears to have the highest morphological productivity based on the combination of a relatively high probability of encountering new word types (Log (P(N))) and the richest vocabulary (Log (V(N))). At the same time, T-1 shows a moderate vocabulary size but a low probability of encountering new words, suggesting a less productive morphology.

### Morphosemantics of Noun Semantic-head in the LVCs *membawa*

In this section of the analysis, we discuss the

distribution of noun-head meanings at the core of LVCs for the verb *membawa* ‘bring.’ This analysis highlights two semantic tendencies: the use of synonymous counterparts and the range of noun core propositions in LVCs (see Figure 3). The first tendency is the natural tendency for LVCs to use synonymous pairs. In simpler terms, this means that equivalent constructions are used. While LVCs are generally idiomatic and grammatically formed, synonymous pairs are lexical. The second tendency relates to the core proposition of nouns used in LVCs. Not all noun head core propositions are distributed equally. Certain propositions are more dominant than others. A quantitative analysis was used to describe both semantic tendencies.

The presented data of Table 4 shows the results of a two-sample t-test (assuming unequal variances) comparing the values of a variable named “Counter Part” (Cp) between two groups: those with a counterpart (Cp) and those without (Ncp). Both groups have the same variance (69.33), but the test assumes unequal variances for a more conservative approach. There are three observations in each group (Cp and Ncp), and the null hypothesis (H0) assumes no difference in the mean values between the groups, so it is set to 0.

The calculated t-statistic is 3.63, and the one-tailed p-value is 0.01, indicating we can reject the null hypothesis with a one-tailed test. This means we can conclude that the difference in means between the groups is statistically significant in the direction of the Cp group having a higher value (37.33) than the Ncp group (12.67). The two-tailed p-value (0.02) is also less than the alpha level, further supporting the conclusion that the difference in means is statistically significant, even considering both directions. However, the two-tailed test requires a more substantial level of evidence compared to the one-tailed test. The one-tailed critical t-value is 2.13 at an alpha level 0.05, and the two-tailed critical t-value is 2.78 at the same alpha level. The degree of freedom is 4, calculated as  $n_1 + n_2 - 2$  (number of observations in each group minus 2).

The data of Table 5 shows the results of a correlation test analyzing the semantic content occurrence of noun-heads. It is a correlation matrix where each row and column represent a semantic category (Peoples, Materials, Objects, etc.), and the values within the cells represent the correlation coefficient between the corresponding pairs of categories. Strong negative correlations are observed between Peoples (Pe) vs. Materials (Ma), Peoples (Pe) vs. Qualities (Qu), Materials (Ma) vs.

**Table 4. The result of T-Test (two-sample assuming unequal variances) for LVC’s Counter Part**

	Counter Part (Cp)	No Counter Part (Ncp)
Mean	37,33	12,67
Variance	69,33	69,33
Observations	3	3
Hypothesized Difference	Mean 0	
Df	4	
t Stat	3,63	
P(T<=t) one-tail	0,01	
t Critical one-tail	2,13	
P(T<=t) two-tail	0,02	
t Critical two-tail	2,78	

**Table 5. The Result of Correlation Test for Semantic Proposition of Noun-head**

	Peoples (Pe)	Materials (Ma)	Objects (Ob)	Qualities (Qu)	Action (Ac)	Processes (Pr)
Peoples (Pe)	1					
Materials (Ma)	-0,99	1				
Objects (Ob)	0,84	-0,92	1			
Qualities (Qu)	-0,98	0,94	-0,74	1		
Action (Ac)	0,88	-0,94	1,00	-0,79	1	
Processes (Pr)	-0,68	0,79	-0,97	0,54	-0,94	1

Table 6. The Third Tier of Occurrence

TIER 3					
LVCs	Freq.	Rank	LVCs	Freq.	Rank
<i>membawa kebanggaan</i> 'brings pride'	363	101	<i>membawa agenda</i> 'bring an agenda'	227	126
<i>membawa dosa</i> 'bring sin'	361	102	<i>membawa undangan</i> 'bring an invitation'	226	127
<i>membawa pembaharuan</i> 'bring renewal'	349	103	<i>membawa tanda</i> 'carrying a sign'	223	128
<i>membawa singgasana</i> 'carrying the throne'	345	104	<i>membawa cahaya</i> 'bring light'	209	129
<i>membawa konsep</i> 'bring the concept'	334	105	<i>membawa sengsara</i> 'brings misery'	203	130
<i>membawa arti</i> 'carries meaning'	326	106	<i>membawa inspirasi</i> 'brings inspiration'	201	131
<i>membawa spirit</i> 'bring spirit'	324	107	<i>membawa peruntungan</i> 'brings good luck'	199	132
<i>membawa kesedihan</i> 'brings sadness'	315	108	<i>membawa kehangatan</i> 'brings warmth'	193	133
<i>membawa tema</i> 'bring a theme'	310	109	<i>membawa faedah</i> 'bring benefits'	192	134
<i>membawa pembaharuan</i> 'bring renewal'	310	110	<i>membawa keteduhan</i> 'brings shade'	189	135
<i>membawa permasalahan</i> 'bring problems'	308	111	<i>membawa kemudahan</i> 'brings convenience'	187	136
<i>membawa mandat</i> 'carries a mandate'	301	112	<i>membawa celaka</i> 'bring harm'	184	137
<i>membawa amanah</i> 'bring trust'	297	113	<i>membawa kekecewaan</i> 'bring disappointment'	184	138
<i>membawa perpecahan</i> 'bring division'	287	114	<i>membawa aroma</i> 'brings fragrance'	184	139
<i>membawa keadilan</i> 'bring justice'	283	115	<i>membawa optimism</i> 'bring optimism'	181	140
<i>membawa energi</i> 'brings energy'	280	116	<i>membawa keunggulan</i> 'bring excellence'	180	141
<i>membawa cerita</i> 'carry the story'	272	117	<i>membawa bahaya</i> 'brings danger'	180	142
<i>membawa pengertian</i> 'brings understanding'	270	118	<i>membawa pengharapan</i> 'brings hope'	177	143
<i>membawa kesembuhan</i> 'bring healing'	264	119	<i>membawa nikmat</i> 'bring favors'	176	144
<i>membawa agama</i> 'bring religion'	261	120	<i>membawa pelita</i> 'carry a lamp'	176	145
<i>membawa imajinasi</i> 'bring imagination'	251	121	<i>membawa kemurkaan</i> 'bring wrath'	172	146
<i>membawa amanat</i> 'carrying a mandate'	245	122	<i>membawa masalah</i> 'bring trouble'	172	147
<i>membawa perasaan</i> 'brings feelings'	243	123	<i>membawa pahala</i> 'brings rewards'	171	148
<i>membawa tuah</i> 'bring good luck'	237	124	<i>membawa ketentraman</i> 'bring peace'	171	149
<i>membawa identitas</i> 'carry identity'	232	125	<i>membawa sensasi</i> 'bring sensation'	170	150
$\Sigma$	7,368		$\Sigma$	4,727	

Objects (Ob), Materials (Ma) vs. Qualities (Qu), Objects (Ob) vs. Processes (Pr), and Action (Ac) vs. Processes (Pr). These negative correlations suggest that when one category is frequent, the other tends to be infrequent in the same context. On the other hand, there are strong positive correlations between Objects (Ob) vs. Action (Ac) and Action (Ac) vs. Processes (Pr), indicating that these categories tend to co-occur frequently, meaning that if one category is frequent, the other is likely to be frequent as well. The value 1 appears on the diagonal of the matrix, which is expected as it represents the correlation between a category and itself. The strong negative correlations between specific categories (e.g., Peoples vs. Materials) suggest a degree of thematic coherence in the data. Texts frequently mentioning people are less likely to mention materials, and vice versa. The positive correlations between Actions, Objects, and Processes suggest that these categories often co-occur when describing activities or events.

The analysis also produced several correlation scores that indicate the strength and direction of the relationship between different categories. The analysis revealed that some categories exhibit a strong negative correlation, meaning they are mutually exclusive in the dataset. For example, Peoples and Materials have a score of -0.99, indicating that a noun head categorized as Peoples is unlikely to be categorized as Materials, and vice versa. Similarly, Objects and Materials have a score of -0.92, indicating that these categories are mutually exclusive. The analysis also found that Qualities have a solid negative correlation with Peoples and Materials, with scores of -0.98 and -0.94, respectively. This suggests that Qualities are rarely used as noun-heads referring to Peoples or Materials. Conversely, processes have moderate negative and positive correlations with Peoples and Materials, respectively, with scores of -0.68 and 0.79. This means that while there is a slight tendency for Processes to be less frequent with People and more frequent with Materials, the correlation is not as strong as with other categories. Finally, the analysis found a strong positive correlation between Objects and Action, with a score of 1.00. This indicates that Objects and Actions almost always co-occur as noun-heads in the corpus. This could be because many actions involve manipulating or interacting with objects. Overall, the remaining correlations between categories range from -0.79 to 0.94, indicating moderate positive or negative relationships.

#### IV. DISCUSSION

This section delves into the findings of this study that explored the distribution scope of LVCs *membawa* used in Indonesian. LVCs *membawa* display a wide range of morphological and semantic usage, including primary and advanced areas, synonymic pairs, and noun core proposition areas. On the one hand, the morphological distribution of LVCs is significant in two domains – the primary domain and the derivational domain. In the essential domain, the verb *membawa* is naturally distributed as LVCs with bare nouns unmarked by an affix. In the derivational domain, the verb *membawa* can be distributed as LVCs when adjacent to derivational nouns resulting from processes like affixation. This finding is further supported by the work of Nugraha, (2023a, 2024). In Nugraha's studies, the distribution of other light verbs, such as *memberi* 'give' and *membuat* 'make' is also mainly occurred in base area of morphological distribution. These observations are corroborated by extensive research documented in Mastrofini (2023), Mattissen (2023), Mehl (2019), and Pompei (2023).

On the other hand, LVCs also exhibit semantic distribution concerning the noun-head referent in the concrete and abstract domains. The concrete domain encompasses all real-world entities denoted by nouns with a minimum of two semantic specifications, i.e., [+PERCEPTIBLE] and [+PHYSICAL EXISTENCE]. Similarly, the abstract domain comprises all non-physical entities denoted by nouns with a minimum of two semantic specifications, i.e., [-PERCEPTIBLE] and [+INTANGIBLE]. This semantic distribution is an advanced area that demonstrates the ability to create word constructions that are both morphological and semantic. In other words, LVCs exhibit morphosemantic distribution in Indonesian grammatical contexts. Our argument draws parallels with the work of Fleischhauer et al. (2019), Karimi-Doostan & Bygi (2016), and Stojanovska-Ilievska (2021). The inclusion of semantic features in LVCs can be considered an essential aspect that cannot be denied. These features enable us to accurately describe the occurrence pattern of a given sample that has been tested. Semantic features carry great significance in analyzing and interpreting data, especially in cases where LVCs can significantly impact the analysis outcome. Moreover, semantic features in LVCs can help us identify and distinguish between different variants

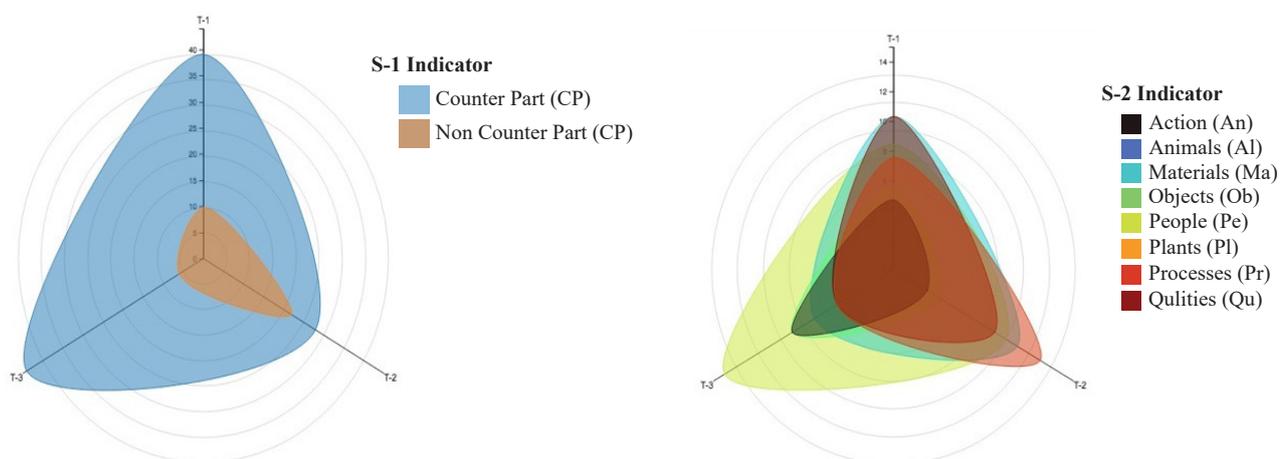


Figure 3. Semantic-Feature Spread Based on the S-1 and S-2 Indicators (left and right panels, respectively)

that may have a similar occurrence pattern, thereby enhancing the analysis's precision and accuracy.

Understanding the intricate connections between nouns in light verb constructions (LVCs) and their morphosemantic relationships, therefore, is a fascinating challenge. This complexity is particularly evident when exploring LVCs in the context of Indonesian grammar, where a nuanced correlation exists between base-abstract (0.15), base-concrete (0.37), affixed-abstract (0.36), and affixed-concrete (0.13) nouns (cf. Fleischhauer & Neisani, 2020; Ronan & Schneider, 2015). The first relationship highlights that nouns without affixes can still express abstract referents, entities outside of language that the five senses cannot directly identify (cf. Ivanová & Kyselová, 2016). On the other hand, the second relationship demonstrates that core nouns without affixes can still exhibit concrete referents, which refer to tangible entities outside of language that the five senses can perceive (cf. He & Wittenberg, 2020). Moreover, the third relationship, affixed-abstract, shows that a core noun with an affix can convey referent power towards an abstract extra-language entity. Conversely, the fourth relationship, affixed-concrete, demonstrates that a core noun with an affix can show referent power towards concrete external linguistic entities (cf. Pompei & Piunno, 2023). These four relationships are not just interesting linguistic phenomena, but they are crucial in understanding the meaning of LVCs. Nouns are the semantic core of LVCs, and therefore, identifying the meaning of the noun is a pivotal step in determining the constructional meaning of LVCs. Without a clear comprehension of the meaning of the noun, the meaning of LVCs remains elusive. The meaning of LVCs, rooted in nouns, accounts for this phenomenon. In other

words, if the noun paired with the light verb *membawa* has no meaning, then LVCs cannot be formed.

This study also comprehensively examines the distribution of Light Verb Constructions (LVCs), focusing on investigating the presence of synonymous pairs and the core propositions associated with nouns. In this study's context, synonymous pairs refer to LVCs that can be paired with other constructs with a similar meaning. The study found that while not all LVCs possess synonymous pairs, the data tested was dominated by LVCs that did. Specifically, constructions with synonymous pairs constituted most of the LVCs (0.75). This finding on synonymous pairs complements the existing study on topic, particularly the work of Baygi et al. (2018), Lu & Huang (2023), and Özge et al. (2022). However, the study also discovered that all investigated LVCs share a common feature: the core proposition of the associated noun. This core proposition varies depending on the noun, with people (0.23), materials (0.19), objects (0.13), qualities (0.17), actions (0.11), and processes (0.17) all being represented. As the domain of synonymous pairs expands, the types of noun core propositions become a source of variation (cf. Barking et al., 2022; Kovalevskaitė et al., 2020; Ronan, 2019). This last key finding, partially, echoes the argument expressed by Fleischhauer (2023). In short, this study provides a detailed and nuanced analysis of LVCs, offering valuable insights into their distribution and construction.

## V. CONCLUSION

The present study, in summary, described the distribution of LVCs *membawa* in the

Indonesian grammatical context. The construction spreads across at least four domains, including morphological and semantic areas. LVCs are distributed in the noun domain as basic or bare and affixed morphemes in the essential morphological area. In advanced morphological areas, LVCs are distributed in the scope of concrete and abstract nouns. LVCs also show two types of distribution in the semantic area: the distribution of synonymous pairs and the distribution of semantic core propositions. The distribution of synonymous pairs is more dominant than other types.

The current study, equally essential, has certain limitations, both methodologically and substantively. The methodological limitation is primarily due to using only one corpus as the data source for this study. This approach cannot compare the results with other corpora, which could have provided valuable insights into different distribution trends. To expand the scope of distribution studies, it is recommended to use two or more corpora as data sources. Including several corpora will also help deepen the study of light verb constructions (LVCs). Substantively, the study only discusses LVCs from a morphosemantic point of view, which can be considered a theoretical limitation. Adding basic theory, such as syntax, can add depth and breadth to the analysis of LVCs. In particular, the valence and transitivity analysis of LVCs is a critical domain that needs to be considered to enhance the study of LVCs.

Further research on LVCs *membawa* suggests that two domains require exploration. Firstly, syntactic analysis of LVCs *membawa* is necessary as these constructions may have a distinct valence structure from other verb constructions. This is because LVCs only functionally fill the predicative

component, possibly creating unique valence characteristics. Secondly, the presence or absence of arguments in LVCs requires investigation. In addition, analyzing LVCs within the context of their discursive function is essential. This approach allows for a concrete form of capturing the LVCs phenomenon in a textual context, leading to a more complete and holistic understanding.

### ETHICS STATEMENT

The authors confirm that they have read and adhered to the ethical requirements for publication in Jurnal Arbitrer. They further declare that the present work does not involve human subjects, animal experiments, or any data collected from social media platforms.

### CREDIT AUTHOR STATEMENT

All authors contributed to the following aspects of the project: Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing - Original Draft, Editing, Visualization, Project Administration. Additionally, all authors participated in Writing - Review & Editing, Validation, and Supervision.

### ACKNOWLEDGEMENTS

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### REFERENCES

- Aikhenvald, A. Y. (2017). A Typology of Noun Categorization Devices. In *The Cambridge handbook of linguistic typology* (pp. 361–404). Cambridge University Press. <https://doi.org/10.1017/9781316135716.012>
- Avdeeva, M. D. (2017). The problem of the origin and development of verbs of the lexico-semantic group 'giving light'. *Research Result. Theoretical and Applied Linguistics*, 3(4). <https://doi.org/10.18413/2313-8912-2017-3-4-3-9>
- Barking, M., Backus, A., & Mos, M. (2022). Similarity in language transfer - investigating transfer of Light Verb Constructions from Dutch to German. *Journal of Language Contact*, 15(1), 198–239. <https://doi.org/10.1163/19552629-15010005>
- Baygi, Z. R., Karimi-Doustan, G.-H., & Sharif, B. (2018). Explanation for alternating Light Verbs (LVs) in Persian complex predicates from a generative lexicon viewpoint. *Language Related Research*, 8(7), 429–452.
- Blevins, J. P. (2016). *Word and paradigm morphology*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199593545.001.0001>
- Brugman, C. (2001). Light verbs and polysemy. *Language sciences*, 23(4–5), 551–578. <https://doi.org/10.1016/S0388->

0001(00)00036-X

- Buckingham, L. (2013). Light verb constructions in Latin American Newspapers: Creative variants and coinages. *Spanish in Context*, 10(1), 114–135. <https://doi.org/10.1075/sic.10.1.05buc>
- Butt, M. (2010). The light verb jungle: Still hacking away. In *Complex Predicates: Cross-Linguistic Perspectives on Event Structure*. <https://doi.org/10.1017/CBO9780511712234.004>
- Dai, G. (2016). Light verbs in translated Chinese. In *New Frontiers in Translation Studies*. [https://doi.org/10.1007/978-981-10-0742-2\\_9](https://doi.org/10.1007/978-981-10-0742-2_9)
- Dal, G., & Namer, F. (2015). Frequency in morphology: For what usages? | La fréquence en morphologie: Pour quels usages? *Langages*, 197(1), 47–68. <https://doi.org/10.3917/lang.197.0047>
- Di Salvo, F. (2023). Light Verb Constructions in Latin. A study on (in) memoria and (in) animo habeo. In *Light Verb Constructions as Complex Verbs* (pp. 337–360). De Gruyter. <https://doi.org/10.1515/9783110747997-013>
- Eatough, V., & Tomkins, L. (2022). Qualitative methods. In *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science [HSK] 46/1* (pp. 163–182). De Gruyter. <https://doi.org/10.1515/9783110347524-008>
- Eshaghi, M., & Karimi-Doostan, G. (2023a). Persian Light Verbs as event determiners. In *Light Verb Constructions as Complex Verbs* (pp. 73–98). De Gruyter. <https://doi.org/10.1515/9783110747997-004>
- Eshaghi, M., & Karimi-Doostan, G. (2023b). Persian Light Verbs as event determiners. In *Light Verb Constructions as Complex Verbs: Features, Typology and Function*. <https://doi.org/10.1515/9783110747997-004>
- Fleischhauer, J. (2021). Light Verb Constructions and Their Families – A Corpus Study on German stehen unter-LVCs. *MWE 2021 - 17th Workshop on Multiword Expressions, Proceedings of the Workshop*, 63–69.
- Fleischhauer, J. (2023). Prospective aspect and current relevance: A case study of the German prospective Stehen vor NP light verb construction. *Journal of Germanic Linguistics*, 35(4), 371–408. <https://doi.org/10.1017/S1470542722000198>
- Fleischhauer, J., Gamerschlag, T., Kallmeyer, L., & Petitjean, S. (2019). Towards a compositional analysis of German Light Verb Constructions (LVCs) combining Lexicalized Tree Adjoining Grammar (LTAG) with frame semantics. *IWCS 2019 - Proceedings of the 13th International Conference on Computational Semantics - Long Papers*, 79–90.
- Fleischhauer, J., & Neisani, M. (2020). Adverbial and attributive modification of Persian separable Light Verb Constructions. *Journal of Linguistics*, 56(1), 45–85. <https://doi.org/10.1017/S0022226718000646>
- García Salido, M., & Garcia, M. (2023). On the unpredictability of Support Verbs. A distributional study of Spanish tomar. In *Light Verb Constructions as Complex Verbs* (pp. 239–256). De Gruyter. <https://doi.org/10.1515/9783110747997-009>
- Gibbs Jr., R. W. (1992). What do idioms really mean? *Journal of Memory and Language*, 31(4), 485–506. [https://doi.org/10.1016/0749-596X\(92\)90025-S](https://doi.org/10.1016/0749-596X(92)90025-S)
- Giparaitė, J. (2015). A Corpus-based Analysis of the Constructions Have/Take/Get a Bath and Have/Take/Get a Rest in British English. *Žmogus Ir Žodis*, 17(3), 37–53. <https://doi.org/10.15823/zz.2015.10>
- Giparaitė, J. (2024). A corpus-based analysis of Light Verb Constructions with MAKE and DO in British English. *Kalbotyra*, 76, 18–41. <https://doi.org/10.15388/Kalbotyra.2023.76.2>
- Gries, S. T. (2015). Quantitative designs and statistical techniques. In *The Cambridge handbook of English corpus linguistics*. <https://doi.org/10.1007/9781139764377.004>
- Hanks, P., Urbschat, A., & Gehweiler, E. (2006). German Light Verb constructions in corpora and dictionaries. *International Journal of Lexicography*, 19(4), 439–457. <https://doi.org/10.1093/ijl/ecl027>
- Haspelmath, M. (2009). An empirical test of the agglutination hypothesis. In: Scalise, S., Magni, E., Bisetto, A. (eds) *Universals of Language Today. Studies in Natural Language and Linguistic Theory*, vol 76. Springer, Dordrecht. [https://doi.org/10.1007/978-1-4020-8825-4\\_2](https://doi.org/10.1007/978-1-4020-8825-4_2)
- He, A. X., & Wittenberg, E. (2020). The acquisition of event nominals and Light Verb Constructions. *Language and Linguistics Compass*, 14(2). <https://doi.org/10.1111/lnc3.12363>
- Hellan, L. (2023). Unification and selection in Light Verb Constructions: A study of Norwegian. In *Light Verb Constructions as Complex Verbs: Features, Typology and Function*. <https://doi.org/10.1515/9783110747997-003>
- Hrenek, É. (2021). Synonymy in Light Verb Constructions of the type feledésbe + verb | Szinonímia a feledésbe + ige típusú funkcióiágés szerkezetek körében. *Magyar Nyelvor*, 145(3), 277–311. <https://doi.org/10.38143/NYR.2021.3.277>

- Hüning, M., & Schlücker, B. (2015). Multi-word expressions. *Word-Formation: An International Handbook of the Languages of Europe, 1*, 450–467. <https://doi.org/10.1515/9783110246254-026/HTML>
- Ivanová, M., & Kyselová, M. (2016). Eventuality semantics in Slovak light verb constructions. *Zeitschrift Fur Slawistik, 61*(4), 662–691. <https://doi.org/10.1515/slwa-2016-0048>
- Ježek, E. (2023). Semantic co-composition in Light Verb Constructions. In *Light Verb Constructions as Complex Verbs* (pp. 221–238). De Gruyter. <https://doi.org/10.1515/9783110747997-008>
- Karimi-Doostan, G., & Bygi, Z. R. (2016). A cognitive study of light verb polysemy: The case of *zadan*. *Language Related Research, 7*(3), 129–148.
- Kettnerová, V. (2023). Valency structure of complex predicates with light verbs: The case of Czech. In *Light Verb Constructions as Complex Verbs: Features, Typology and Function*. <https://doi.org/10.1515/9783110747997-002>
- Kovalevskaitė, J., Rimkutė, E., & Vilkaitė-Lozdienė, L. (2020). Light verb constructions in Lithuanian: identification and classification | Leksinės analitinės konstrukcijos: Nustatymas ir klasifikacija. *Studies About Languages, 36*, 5–16. <https://doi.org/10.5755/J01.SAL.0.36.22846>
- Lu, L., & Huang, C.-R. (2023a). A diachronic insight into the aspectual meaning in Light Verb Constructions. A case study in Mandarin Chinese. In *Light Verb Constructions as Complex Verbs* (pp. 305–336). De Gruyter. <https://doi.org/10.1515/9783110747997-012>
- Lu, L., & Huang, C.-R. (2023b). A diachronic insight into the aspectual meaning in Light Verb Constructions: A case study in Mandarin Chinese. In *Light Verb Constructions as Complex Verbs: Features, Typology and Function*. <https://doi.org/10.1515/9783110747997-012>
- Mastrofini, R. (2023). When lightness meets lexical aspect: A corpus-based account of English Light Verb Extensions. In *Light Verb Constructions as Complex Verbs: Features, Typology and Function*. <https://doi.org/10.1515/9783110747997-007>
- Mattissen, J. (2023). Light Verbs and ‘light nouns’ in polysynthetic languages. In *Light Verb Constructions as Complex Verbs* (pp. 275–304). De Gruyter. <https://doi.org/10.1515/9783110747997-011>
- Mehl, S. (2019). Light verb semantics in the international corpus of English: Onomasiological variation, identity evidence and degrees of lightness. *English Language and Linguistics, 23*(1), 55–80. <https://doi.org/10.1017/S1360674317000302>
- Nolan, B. (2014). Complex predicates and light verb constructions in Modern Irish. *Revista Espanola de Linguistica Aplicada, 27*(1), 140–167. <https://doi.org/10.1075/resla.27.1.06nol>
- Nugraha, D. S. (2022a). Identifying Indonesian Light Verb Constructions: a Computational linguistics approach. In Krisztina K. & G. T. Etelka (Eds.), *XVI. Alkalmazott nyelvészeti doktoranduszkonferencia. Az MTA Alkalmazott Nyelészeti Munkabizottsága és a Nyelvtudományi Kutatóközpont szervezésében*.
- Nugraha, D. S. (2022b). Identifying Light Verb Constructions in Indonesian: A Direct translation approach. *International Journal of Language and Literary Studies, 4*(3), 298–311. <https://doi.org/10.36892/ijlls.v4i3.1042>
- Nugraha, D. S. (2023). Morphosemantic features of *Membuat* ‘Make’ in The Light Verb Constructions of Indonesian. *LiNGUA: Jurnal Ilmu Bahasa Dan Sastra, 17*(2), 131–142. <https://doi.org/10.18860/ling.v17i2.17757>
- Nugraha, D. S. (2024a). A Corpus-Based study of *Memberi* ‘Give’ Light Verb Constructions. *International Journal of Society, Culture & Language*. <https://doi.org/10.22034/ijscsl.2024.2023112.3389>
- Nugraha, D. S. (2024b). Morphosemantic features of *Memenuhi* ‘Meet’ In The Light Verb Constructions of Indonesian. *Linguistik Indonesia, 42*(2), 461–480. <https://doi.org/10.26499/li.v42i2.638>
- Özge, D., Ünal, G., & Bayırlı, İ. K. (2022). Assigning meaning to light verbs in Turkish | Türkçede katkısız eylemlere anlam atanması. *Dilbilim Arastirmalari Dergisi, 33*(1), 1–27. <https://doi.org/10.18492/dad.917075>
- Pompei, A. (2023). How light is ‘give’ as a Light Verb? A case study on the actionality of Latin Light Verb Constructions (with some references to Romance languages). In *Light Verb Constructions as Complex Verbs* (pp. 149–200). De Gruyter. <https://doi.org/10.1515/9783110747997-006>
- Pompei, A., & Piunno, V. (2023). Light Verb Constructions in Romance languages: An attempt to explain systematic irregularity. In *Light Verb Constructions as Complex Verbs: Features, Typology and Function*. <https://doi.org/10.1515/9783110747997-005>
- Plungian, V. A. (2001). Agglutination vs. flexion. In *Language typology and language universals: An international handbook* (HSK), vol. I, ed. Martin Haspelmath et al., 669–678. Berlin: de Gruyter.
- Ronan, P. (2014). Light verb constructions in the history of English. In *Studies in Corpus Linguistics* (Vol. 63). <https://doi.org/10.1075/scsl.63.05ron>

- Ronan, P. (2019). Simple versus Light Verb Constructions in late Modern Irish English correspondence: A qualitative and quantitative analysis. *Studia Neophilologica*, 91(1), 31–48. <https://doi.org/10.1080/00393274.2019.1578182>
- Ronan, P., & Schneider, G. (2015). Determining light verb constructions in contemporary British and Irish English. *International Journal of Corpus Linguistics*, 20(3), 326–354. <https://doi.org/10.1075/ijcl.20.3.03ron>
- Sheng, D. (2023). Statistics in corpus linguistics: a practical guide. *Social Semiotics*, 33(4), 909–912. <https://doi.org/10.1080/10350330.2021.1969215>
- Si, F. (2021). Towards a cartography of light verbs. In *Linguistik Aktuell* (Vol. 267). <https://doi.org/10.1075/la.267.10si>
- Siefring, J. (2004). *Oxford Dictionary of idioms* (2nd ed.). Oxford University Press.
- Stojanovska-Ilievska, N. (2021). Observations on a subgroup of Light Verb Constructions with Dava in which the metaphorical extension of interpersonal communication is present. *Palimpsest*, 6(12), 17–26. <https://doi.org/10.46763/PALIM21612017si>
- Stupak, I. V., & Baayen, R. H. (2022). An inquiry into the semantic transparency and productivity of German particle verbs and derivational affixation. *Mental Lexicon*, 17(3), 422–457. <https://doi.org/10.1075/ml.22012.stu>
- Sundquist, J. D. (2020). Productivity, richness, and diversity of light verb constructions in the history of American english. *Journal of Historical Linguistics*, 10(3), 349–388. <https://doi.org/10.1075/jhl.19009.sun>
- Vilas, B. S. (2015). The role of determiners in Spanish light verb constructions. *Word*, 61(2), 178–200. <https://doi.org/10.1080/00437956.2015.1033176>
- Vilas, B. S. (2017). From the heavy to the light verb: An analysis of tomar ‘to take.’ *Lingvisticae Investigationes*, 40(2), 228–273. <https://doi.org/10.1075/li.00005.san>
- Vincze, V., István Nagy, T., & Zsibrita, J. (2013). Learning to detect english and Hungarian light verb constructions. *ACM Transactions on Speech and Language Processing*, 10(2). <https://doi.org/10.1145/2483691.2483695>
- Wang, C.-A. A., & Wu, H.-H. I. (2020). Light verbs in verbal reduplication. *Studia Linguistica*, 74(2), 337–359. <https://doi.org/10.1111/stul.12128>
- Wang, D., Jiang, G., & Zheng, Y. (2023). Walking out of the light verb jungle: Exploring the translation strategies of light verb constructions in Chinese–English consecutive interpreting. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1113973>
- Yim, C. (2020). Light verb ellipsis constructions in Korean. *Language and Linguistics*, 21(2), 351–374. <https://doi.org/10.1075/lali.00064.yim>