International Journal of Language and Literary Studies
Volume 4, Issue 3, 2022
Homepage : http://ijlls.org/index.php/ijlls

Identifying Light Verb Constructions in Indonesian: A Direct Translation Approach

Danang Satria Nugraha
Sanata Dharma University, Indonesia
d.s.nugrahausd.ac.id

DOI: http://doi.org/ 10.36892/ijlls.v4i3.1042

APA Citation: Nugraha, D. S..(2022). Identifying Light Verb Constructions in Indonesian: A Direct Translation Approach. *International Journal of Language and Literary Studies. 4*(3).298-311. http://doi.org/ 10.36892/ijlls.v4i3.1042

Received:
18/08/2022

Accepted:
00/00/2022

Keywords:
Light verb constructions, machine translation, direct translation approach, Indonesian.

Abstract

This study aimed to identify light verb constructions (LVCs) in Indonesian based on machine translation methods, namely binary translation or direct translation. Based on the method, the LVCs of the source language (SL) were selected and utilized to construct the LVCs counterpart of the target language (TL). The SL instrument has been developed using previous studies on English and Hungarian, namely the LVCs matrix formulated by Vincze (2011). The TL data were in the equivalent forms of LVCs in Indonesian. Data analysis was carried out based on morphosemantics and morphosyntax. Based on the study, the result was stated as follows, i.e. (i) there were translation work procedures that needed to be carried out to process LVCs from SL to TL regarding the type of translator, limited text selection, and final counterpart assignation; (ii) there were changes in the morphosemantic features of LVCs from SL to TL during the translation process; and (iii) there were grammatical exceptions that cannot be resolved through binary translation methods. This study also suggested a further analysis of LVCs identification through another method, for instance, information retrieval (IR) and information extraction (IE), in the light of computational linguistics.

1. INTRODUCTION

Language structure depends not only on the morphosyntactic features but also on morphosemantic characteristics. As in the light verb constructions (henceforth LVCs), one cannot analyse the phenomenon by utilizing a single facet of the approach in theoretical linguistics. Based on the assumption, this study aimed to identify LVCs in Indonesian based on the machine translation (henceforth MT) method, particularly binary translation or direct translation. As a branch of computational linguistics, MT was considered to be used for the initial identification of LVCs in Indonesian based on data set from the source language (henceforth SL). On the one hand, the initial identification was used as a preliminary study (cf. Kay, 2003; Mitkov, 2003; Shen, 2004; Simpson, 2001; Ivana & Sakai, 2007; Nugraha, 2021; Ong & Rahim, 2021; Hrenek, 2021). On the other hand, the initial identification can produce a projection of the grammatical universality of LVCs (cf. Spencer & Zwicky 2001; Haspelmath

& Sims 2010; Nugraha 2020; Snider, 2021; Tron et al., 2022; Jutunka & Attaviriyanupap, 2022; Eshaghi & Doustan, 2022). It is clear that this study is relevant to be carried out in terms of methodological background or substantial foundation.

Semantically, LVCs are verbs that have little semantic content on their own and form a predicate with some compulsory expression, which is occasionally a noun. According to Vincze (2011), LVCs consist of a nominal and a verbal component where the noun has taken in one of its literal senses. However, the verb usually loses its original sense to some extent. The term *nominal* refers to the common noun defined by Booij et al. (2000) as stated, "The *nomen appelativum* (nouns whose designation is a concept; not semantically definite as in proper nouns)" (p.746). Moreover, the term *verbal* refers to verbs, semantically defined as describing events, actions, and, in some languages, states (Booij et al., 2000). For instance, the sample of LVCs in Indonesian can be seen in the following (1) – (5).

(1) *membuat* *keputusan*
make decision – ACC
'to make a decision'

or

(2) *mengambil* *langkah*
take step – ACC
'to take a step'

or

(3) *memenuhi* *persyaratan*
meet requirement – ACC
'to meet a requirement'

or

(4) *memberikan* *nasihat*
give advice – ACC
'to give an advice'

or

(5) *memainkan* *peran*
play role – ACC
'to play a role'

Previously, several researchers have had researched the LVCs in various languages (cf. Vincze 2011; Fleischhauer et al. 2019; Vaidya et al. 2019; Fleischhauer and Neisani 2020; Nagy et al. 2020; Xu et al. 2022). Vincze (2011) and Nagy et al. (2020) examined the construction of LVCs in Hungarian and English. Vaidya (2019) examines LVCs in Indian Language. Fleischhauer and Neisani (2020) have analysed LVCs in Persian, having previously studied German (Fleischhauer et al., 2019). Xu et al. (2022) have analysed LVCs in Mandarin Chinese. So far, based on observation search on several journal publications and proceedings

and other forms of publication, there have not been many studies of LVCs on languages in the Southeast Asian Region, including Indonesian, a morphologically rich language spoken by around 3 million peoples (Sneddon et al. 2010; Nugraha & Baryadi 2019). Thus, it can be argued that the identification of LVCs in Indonesian based on the computational linguistic paradigm has not been comprehensively carried out. For this reason, this research was designed to apply this paradigm to identifying LVCs in Indonesian.

For a more structured presentation, the rest of this paper is laid out as follows. In section 2.1, the main theories are presented in a brief explanation of theoretical background, including the three research questions in an imperative mood of the sentence. In section 3, the brief methodology is outlined. In section 4, results and discussion were pointed out and described in the light of theory. In section 5, a conclusion and limitation are presented orderly. This paper is an expanded version of Nugraha (2022), containing additional features and a sample of data.

2. LITERATURE REVIEW

2.1. Light Verb Constructions

Verbs are constructions that can be identified as lexical and grammatical statues. As a lexical unit, verbs are usually used syntactically as predicates of a clause (Fleischhauer, 2021; Berenjian, 2021; Srinivas & Legendre, 2022; Kintz & Wrigth, 2022). The verbs also semantically represent the meaning of ACTION, PROCESS, and STATE (Coussé & Bouma, 2022; Purmohammad & Abutalebi, 2022; van Goethem & Koutsoukos, 2022). In addition, as a grammatical unit, verbs are seen as linguistic compositional units. The compositional form unit is based on combining at least two linguistic units as verb formers, for instance, the nouns and affixes. The verb in the grammatical context is morphologically the result of the word formation process. Besides derivational verbs, one of the other common types of grammatical verbs is light verb constructions (LVCs).

LVCs are grammatical verbs. It requires at least two linguistic units: verb (V) and noun (N). If analysed, the formation pattern is [LVCs: V + N]. Morphologically, the formation pattern can be assumed to occur in many languages worldwide. Because of these morphological features, LVCs are often classified as compound words. The morphological marker is also the main parameter in the initial identification of LVCs. In addition, LVCs also have syntactic features. Based on their function projection, LVCs tend to be the unit that fills the predicate function in a clause. Hypothetically, LVCs can become predicates in intransitive and transitive clauses. One should state that further analyses are needed to identify the transitivity of LVCs. Meanwhile, based on the semantics point of view, LVCs are understood as a construction of multiword expressions (MWEs) (Han, 2022; Villavicencio & Idiart, 2019; Ramisch, 2017; Kallens & Christiansen, 2022; Gries, 2022). In these constructions, the meaning of nouns dominates the meaning of verbs. The rule can take N > V, where N is the majority share of the meaning portion, and V is the minority share of the meaning portion. Hypothetically, LVCs are understood as linguistic units with the semantic meaning of nouns and the syntactic feature of verbs.

2.2. Direct Translation Approach

The fundamental theoretical background used in this study is the direct translation approach. The approach includes two folds. The first conjecture was rule-based machine translation and binary or direct translation approach. Ruled-based is understood as the most basic form of the machine translation system. It is the most basic because it does not involve any data extraction intervention on the initial data to be translated. The initial forms at the input level are processed according to the grammar algorithm that has been integrated into the translation machine. More specifically, according to Hutchins (2003), Forcada et al. (2011), Shiwen & Xiaojing (2014), and Hurskainen & Tiedemann (2017), the rule-based machine translation has been embedded

in the translator chosen by the researcher during the analysis. For public usage, the machine translator as Google Translate has been developed by using the rule-based. In this case, one may understand that the term rule-based refers to the specific rules embedded in the machine translator. The rule-based machine translation is realized through the binary or direct translation method as one of the optionality for research analysis (Olimqizi, 2022; Cui et al., 2022; Shang, Xia, & Yakovlev, 2002; Li, Wang, & Wang, 2021). Theoretically, the method is defined as the method where source language (SL) utilized as the primary source in the process of translation. There is no other consideration of language context that need to be include in the SL. The SL directly translate by using the selected machine translator to produce the equivalent form in the target language (TL).

The second supposition was the change of semantic meaning. The term "semantic meaning" refers to the linguistic senses of the language-specific units in a particular grammatical construction (Lieber 2004; Lieber and Štekauer, 2009). It is not the same as the utterance meaning from a pragmatical point of view. Semantic meaning is closely related to internal structure (Riemer, 2015; Jackendoff & Audring, 2020; Pavlick, 2022; Lieber & Plag, 2022). In machine translation, semantic meaning is defined as the meaning of the input and output of the text. On the one hand, the translation process will affect the meaning of the input text. In this case, the meaning of SL will change due to the translation process. The first level of meaning transformation is based on lexical. Since the form of SL data is in the form of compound words wherein the lexical meaning is not embedded, there is a possibility of error translation at the TL. Compound words are commonly identified as grammatical units instead of lexical ones. As well as LVCs in the TL, one cannot assume the exact equivalent translation in the TL.

On the other hand, the output meaning from translation is always changed by the nature of the translation process. Output meaning is defined as the meaning of equivalent form in TL. The other consideration is regarding the difference in structure. LVCs in English closely similar to their Indonesian counterparts because both languages use the serialization of word order, in this case, phrase order. Hence, the Hungarian and Indonesian are not closely similar since the first typologically employed the case marker in its construction. Based on that context, studying meaning in the LVCs across languages is a valuable phenomenon.

2.3. Research Questions

This research has been conducted to answer the following three research questions:

1) How do direct approaches translate LVCs from English and Hungarian as source languages to Indonesian as target languages?
2) What morphosemantic features of the LVCs from the source language changed during the translation?
3) What grammatical exceptions are obtained in the translation process of the LVCs?

3. METHOD

This study has been done in three stages as follows. The first stage is data collection. The instrument of study was obtained from English and Hungarian as the previous research context, namely the LVCs matrix formulated by Vincze (2011). The matrix of LVCs has been positioned as the source language (SL) of input material for the translation process in the main analysis phase. The other instrument was a set of machine-translator. The set composed by three machine-translation, i.e. (a) Google Translation (GT) (can be found at https://translate.google.com/), (b) Cambridge Dictionary Translation (CT) (can be found at https://dictionary.cambridge.org/translate/), and (c) Duolingo Dictionary Translation (DT) (can be found at https://www.duolingo.com/dictionary/id). The two instruments were utilized during the data collection stage, respectively.

The second stage is data analysis. By the machine-translation method, especially binary translation or direct translation, Indonesian LVCs were identified by looking for Indonesian equivalent words from the matrix of Vincze's English & Hungarian LVCs (2011). By the human-aided machine translation (HAMT) principles, the list of Indonesian LVCs was analyzed to determine the acceptability aspect of the equivalent form. To support the native speaker inspection, the determination of acceptance also depended on the corpora of the Indonesian, namely the Indonesian – Leipzig Corpora Collection (ILCC). ILCC_ind_mixed_2013 is an Indonesian mixed corpus based on material from 2013. It contains 74,329,815 sentences and 1,206,281,985 tokens. In particular, Figure 1 illustrates the analytical method in this study. In the input section, two matrix LVCs in English (ENG) and Hungarian (HUN) were prepared by utilizing the results of previous research conducted by Vinzce (2011). In the machine translation section, three types of online translators are selected based on their performance. In the output section, there is an Indonesian language slot as TL.

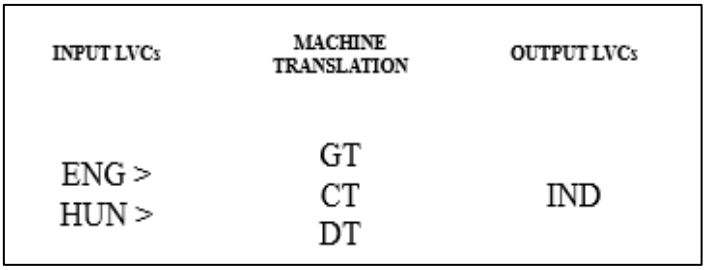

Fig. 1. Methodological Framework of Direct Translation Approach

The last stage is the results presentation. There were two styles of producing results: descriptive and table presentation. Since the primary data type of this study was in the qualitative form, the descriptive representation has been utilized to help explain the feature of LVCs. The part included the characteristics of the equivalent set of LVCs in Indonesian. The elements were in the sense of morphosyntax or morphosemantic ones. In addition, the table style has been formulated on the stage to show the details.

4. RESULTS AND DISCUSSION

4.1. Direct translation of LVCs

This section describes the translation results of LVCs from English and Hungarian as SL into Indonesian as TL. The descriptions presented are brief and only provide sample patterns. Not all LVCs of SL can be translated directly into Indonesian. Likewise, several translations or counterparts LVCs in Indonesian do not have grammatical meanings that can be understood based on grammatical knowledge. In addition to being based on grammatical intuition, the determination of the grammatical meaning is also justified in the ILCC corpus as the original documentation of the use of Indonesian. For instance, consider the examples bellow.

(6) English — Indonesian
a. *make a decision* — *membuat* *keputusan*
make — decision – ACC

'to make a decision'

b. *take a step* — *mengambil* *langkah*
take step – ACC

'to take a step'

(7) Hungarian — Indonesian
a. *üzletet köt* — *membuat* *kesepakatan*
make deal – ACC
'to make a deal'

b. *intézkedést tesz* — *mengambil* *tindakan*
take action – ACC
'to take an action'

Furthermore, the number of English LVCs found in the Indonesian equivalent is 101 constructions (see table 2 as the equivalent sample of LVCs). The number of Hungarian LVCs found in the Indonesian counterpart is 269 constructions (see table 3 as the equivalent sample of LVCs). This number tends to grow if different translation methods are used. Therefore, the amount is not final but a lasting result. This number is also the total number for which similar constructions have not been identified. Eventually, this means that no identification has been carried out to determine the intersection of the two counterparts. Regarding the similarity of the form of the counterpart in the TL, there will certainly be a decrease in the amount of LVCs counterpart.

No.	LVCs form in TL	Occurrences Rate Based on ILCC
1.	membuat keputusan	70.147
2.	mengambil langkah	49.422
3.	membuat kesepakatan	8.949
4.	mengambil tindakan	40.100

Table 1. Sample of Occurrences Rate of the Indonesian LVCs

Determination of acceptability based on the rate of occurrence (left neighbor cooccurrences) of constructions in the corpus of the language used. As a sample, consider the forms of LVCs in Table 1. In the table, there are four examples of LVCs with different levels of occurrence. LVCs *membuat keputusan* is most prevalent, and *membuat kesepakatan* is least occurring. Some types of LVCs translations also very rarely appear in Indonesian.

No	SL: LVCs of English	TL: Indonesian		
		GT	CT	DT
1	take place	Terjadi	berlangsung	dilakukan
2	make a decision	membuat sebuah keputusan	membuat keputusan	membuat sebuah keputusan
3	take part	ambil bagian	ambil bagian	ambil bagian
4	play a role	memainkan peran	memainkan peran	memainkan peran
5	take care	hati-hati di jalan	jaga diri	menjaga
6	take a decision	mengambil keputusan	mengambil keputusan	mengambil keputusan
7	make a remark	membuat komentar	membuat komentar	membuat komentar
8	take a look	Lihatlah	lihatlah	Lihatlah
9	give an order	memberi perintah	memberikan perintah	memberi perintah
10	make a mistake	membuat kesalahan	membuat kesalahan	membuat kesalahan

Table 2. Sample of the Indonesian equivalent

Based on the data in table 1, it can be stated that the construction of *membuat keputusan* is the type of productive LVC. Productive is limited in its understanding as Indonesian speakers use the verbal form most commonly used. Meanwhile, the construction of *mengambil kesepakatan* is a type of construction that tends to be not widely used in the comparison configuration in table 1. The striking comparison between the two constructions may be caused by extra lingual elements such as each construction's situation and speech context. However, these two constructs are LVCs which can be perfectly translated into Indonesian from English and Hungarian.

No	SL: LVCs of Hungarian	TL: Indonesian		
		GT	CT	DT
1	irányt ad	memberikan arah	memberikan arah	memberikan arah
2	szerzodést köt	masuk ke dalam kontrak	mengakhiri kontrak	masuk ke dalam kontrak

3	határozatot hoz	mengambil keputusan	mengadili	mengambil keputusan
4	döntést hoz	membuat sebuah keputusan	membuat keputusan	membuat sebuah keputusan
5	ajánlatot tesz	menawarkan	membuat penawaran	menawarkan
6	lehetoséget ad	memberikan kesempatan	memberi Anda kemungkinan	memberikan kesempatan
7	engedélyt ad	beri izin	memberikan izin	beri izin
8	tanácsot ad	memberikan nasihat	menasihati	memberikan nasihat
9	bérbe ad	Disewakan	sewa	disewakan
10	üzletet köt	membuat kesepakatan	membuat kesepakatan	membuat kesepakatan

Table 3. Sample of the Indonesian equivalent

On the one hand, binary translation can bring content and morphosemantic features from SL into TL. On the other hand, binary translation is not always optimal. Optimal is defined as carrying all the information in the deep layer or deep structure of LVCs. That is the main obstacle to translation based on the direct method that only relies on machines without any intervention from the language speakers. The limitation of the machine translator repository based on the binary model cannot transfer morphosemantic features from SL to TL, as discussed in subsections 4.2 and 4.3 below.

4.2. The Changes of Morphosemantic Features of the LVCs

There was evidence regarding the change of morphosemantic features, such as (i) translating LVCs into clauses and (ii) translating LVCs into (only) verbs that are not followed by nouns. These examples are listed in Table 4. There are LVCs constructions that are translated into clauses in Indonesian. The clause has a predicative or verbal element distributed with its grammatical subject. Meanwhile, other translations eliminate the noun elements from the source language LVCs.

No.	LVCs form	SL	TL Form	Grammatical Category
1.	take place	English	terjadi; berlangsung; dilakukan	Verb
2.	take into account	English	Memperhitungkan	Verb

3.	take a seat	English	silakan duduk	clause
4.	have a good time	English	selamat bersenang-senang	clause
5.	házasságot köt	Hungarian	Menikah	Verb
6.	megbeszélést folytat	Hungarian	Berdiskusi	Verb
7.	jutalmat kap	Hungarian	kamu mendapatkan hadiah	clause
8.	harcot vív	Hungarian	dia sedang bertarung	clause

Table 4. Sample of the change in morphosemantic features

In particular, the change in the LVCs category into clauses in Indonesian is evidence of the weakness of the binary translation system. The lexical identity or grammatical category of LVCs SL cannot be maintained, and its meaning changes. In the configuration of these changes, there needs to be intervention from language speakers who understand the three languages being processed: English, Hungarian, and Indonesian. Otherwise, a different type of machine translation should be applied to that purpose. According to previous findings, researchers like Huang et al. (2009) found that based on their large-scale experiments, almost all rules are binarizable. The resulting binarized rule set significantly improves the speed and accuracy of a state-of-the-art syntax-based machine translation system. In Wan et al. (2022) terminology, the change in morphosemantic features tends to be a *mistranslation error.*

4.3.The Grammatical Exceptions

This study found some unacceptable LVCs translations, as listed in table 3. Unacceptable means that either the construction did not appear in the corpus used in this study (ILCC) or undetected as a grammatically correct construction based on the underlying knowledge of Indonesian grammar. On the one hand, it means that the construction never appears in the daily use of Indonesian speakers; on the other hand, it also means that the constructions due to the machine translation process do not consider the context of language use. Translation based on the binary method removes the specific grammatical elements of the source language LVCs.

No.	Original LVCs form	SL	TL form	Occurrences rate based on ILCC
1.	bring into line	English	*membuat setuju	0
2.	take a measure	English	*mengambil ukuran	0
3.	give air	English	*memberi udara	0

4.	give a concert	English	?memberikan konser	0
5.	otthont ad	Hungarian	?memberi rumah	0
6.	háborút visel	Hungarian	*memakai perang	0
7.	szándéknyilatkozatot tesz	Hungarian	*membuat pernyataan niat	0

Table 5. Sample of the Ungrammatical Translation of LVCs

According to Sommers (2003), this grammatical exception is related to the fact that MT systems which are available on the World Wide Web, usually free, have introduced an essentially new and under-foreseen use for low-quality MT. Therefore, further analysis is needed in order to obtain valid translation results. The equivalence LVCs generated by binary method translation does not accurately record the source language's morphosemantic and morphosyntactic characteristics of the LVCs. Gimpel (2014) and Hadiwinoto (2017) suggested utilizing the linguistic syntax for either the source or target language concerning the combinatory model of phrases and dependency syntax and phrase-based and syntax-based translation.

5. CONCLUSION

To conclude this paper, two things are presented in this subsection: conclusions and limitations. The first is a description of the conclusion. Generally, it can be concluded that LVCs in Indonesian can be identified through a binary translation approach by utilizing the LVCs list of source languages (SL), namely English and Hungarian. However, there is a change in the morphosemantic characteristics of LVCs SL subjected to the translation process. In addition to the changes in these characteristics, there are also exceptions resulting from the translation process. The second is related to the limitations of the study. In particular, this study has not been able to describe in detail the pattern of changes in the translated LVCs semantic characteristics. In addition, the acceptability aspect of the counterpart construction/translation in Indonesian still needs to be tested on native speakers. For this reason, in further research, the construction of translation needs to be analysed experimentally based on its morpho-pragmatic aspects.

REFERENCES

Berenjian, S.R. (2021). An investigation into syntactic patterns carrying Persian light verb zadan. *International Journal of Language Studies*, 15(1).

Booij, G., Lehmann, C., Mugdan, J., Kesselheim, W., & Skopeteas, S. (2000). *Morphology: an international handbook on inflection and word-formation*. Berlin, New York: Walter de Gruyter.

Coussé, E., & Bouma, G. (2022). Semantic scope restrictions in complex verb constructions in Dutch. *Linguistics*, 60(1), 123-176. https://doi.org/10.1515/ling-2021-0172

Cui, J., Shinde, S., Sen, S., Saxena, P., & Yuan, P. (2022). Dynamic binary translation for sgx enclaves. *ACM Transactions on Privacy and Security*. https://doi.org/10.1145/3532862

Eshaghi, M., & Karimi Doustan, G. (2022). The study of Persian Light Verbs productivity. *Language Research*, 12(2), 1-28. https://dx.doi.org/10.22059/jolr.2021.315912.666682

Forcada, M. L., Ginestí-Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J. A., ... & Tyers, F. M. (2011). Apertium: a free/open-source platform for rule-based

machine translation. *Machine translation*, 25(2), 127-144. https://doi.org/10.1007/s10590-011-9090-0

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. In *Proceedings of the 13th International Conference on Computational Semantics-Long Papers*. 79–90. http://dx.doi.org/10.18653/v1/W19-0407

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

Fleischhauer, J. (2021). Light Verb Constructions and Their Families-A Corpus Study on German 'stehen unter'-LVCs. In *Proceedings of the 17th Workshop on Multiword Expressions (MWE 2021)* (pp. 63-69).

Gimpel, K. Noah, A. Smith. (2014). Phrase Dependency Machine Translation with Quasi-Synchronous Tree-to-Tree Features. *Computational Linguistics*, 40 (2): 349–401. https://doi.org/10.1162/COLI_a_00175.

Gries, S. T. (2022). Multi-word units (and tokenization more generally): a multi-dimensional and largely information-theoretic approach. *Lexis. Journal in English Lexicology*, (19). https://doi.org/10.4000/lexis.6231

Hadiwinoto, C. (2017). Syntax-Based Statistical Machine Translation. *Computational Linguistics*, 43 (4): 893–896. doi: https://doi.org/10.1162/COLI_r_00303.

Han, L. (2022). *An investigation into multi-word expressions in machine translation* (Doctoral dissertation, Dublin City University).

Haspelmath, M., Sims, A. (2010). *Understanding Morphology*. London: Routledge.

Huang, L., Zhang, H., Gildea, D., & Knight, K. (2009). Binarization of Synchronous Context-Free Grammars. *Computational Linguistic*, 35 (4): 559–595. doi: https://doi.org/10.1162/coli.2009.35.4.35406.

Hurskainen, A., & Tiedemann, J. (2017). Rule-based machine translation from english to finnish. In *Proceedings of the Second Conference on Machine Translation (WMT2017)*. The Association for Computational Linguistics. http://dx.doi.org/10.18653/v1/W17-4731

Hutchins, J. (2003). Machine translation: general overview. In: Mitkov, R. (ed). *The oxford handbook of computational linguistics*. New York: Oxford University Press.

Hrenek, É. (2021). Synonymous light verb constructions and synonymy groups. A study of verb variability in Hungarian. *Investigationes Linguisticae*, (45), 47-66. https://doi.org/10.14746/il.2021.45.4

Ivana, A., & Sakai, H. (2007). Honorification and light verbs in Japanese. *Journal of East Asian Linguistics*, 16, 171–191. https://doi.org/10.1007/s10831-007-9011-7

Jackendoff, R. & Audring, J. (2020). *The Texture of the Lexicon*. USA: Oxford University Press.

Jutunka, N., & Attaviriyanupap, K. (2022). Light Verb Constructions in German and Their Equivalents in Thai: A Contrastive Analysis through the Example of the Constitutional Texts. *Nusantara Science and Technology Proceedings*, 146-153. https://doi.org/10.11594/nstp.2022.1919

Kallens, P.C., & Christiansen, M.H. (2022). Models of language and multiword expressions. *Frontiers in Artificial Intelligence*, 5. https://doi.org/10.3389/frai.2022.781962

Kay, M. (2003). Introduction. In: Mitkov, R. (ed). *The oxford handbook of computational linguistics*. New York: Oxford University Press. xvii – xx.

Kintz, S., & Wright, H. H. (2022). Light verb production in healthy ageing and dementia. *International Journal of Language & Communication Disorders*. https://doi.org/10.1111/1460-6984.12721

Leipzig Corpora Collection: Indonesian mixed corpus based on material from 2013. Leipzig Corpora Collection. Dataset. https://corpora.uni-leipzig.de?corpusId=ind_mixed_2013.

Li, P., Wang, M., & Wang, J. (2021). Named entity translation method based on machine translation lexicon. *Neural Computing and Applications*, 33(9), 3977-3985. https://doi.org/10.1007/s00521-020-05509-y

Lieber, R. (2004). The semantics of verb formation. Morphology and Lexical *Semantics*, 76–97. https://doi.org/10.1017/CBO9780511486296.004

Lieber, R., & Štekauer, P. (2009). *The Oxford handbook of compounding*. Oxford: Oxford University Press.

Lieber, R., & Plag, I. (2022). The semantics of conversion nouns and-ing nominalizations: A quantitative and theoretical perspective. *Journal of Linguistics*, 58(2), 307-343. https://doi.org/10.1017/S0022226721000311

Mitkov, R. (2003). *The oxford handbook of computational linguistics*. New York: Oxford University Press.

Nagy T., I., Rácz, A., & Vincze, V. (2020). Detecting light verb constructions across languages. *Natural Language Engineering*, 26(3), 319–348. https://doi:10.1017/S1351324919000330.

Nugraha, D.S. (2020). The comparative analysis of syntactic features between Indonesian and English denominal verbs. *LiNGUA: Jurnal Ilmu Bahasa dan Sastra*, 15(1), 65–78. https://doi.org/10.18860/ling.v15i1.7680.

Nugraha, D.S. (2021). Makna-makna gramatikal konstruksi verba denominatif dalam bahasa Indonesia. *Bahasa Dan Seni: Jurnal Bahasa, Sastra, Seni, dan Pengajarannya*, 49(2), 224–239. https://doi.org/https://dx.doi.org/10.17977/um015v49i22021p224.

Nugraha, D.S. (2022). Identifying Indonesian light verb constructions: a computational linguistics approach. In: Krisztina, K., Etelka, G.T. (eds) 2022. *XVI. Alkalmazott nyelvészeti doktoranduszkonferencia*. Budapest: Az MTA Alkalmazott Nyelézseti Munkabizottsága és a Nyelvtudományi Kutatóközpont szervezésében.

Nugraha, D.S., & Baryadi, P. (2019). Perbandingan Fitur Morfologis antara Verba Denominatif dalam Bahasa Indonesia dan Bahasa Inggris. *Sirok Bastra*, 7(2), 107–117. https://doi.org/https://doi.org/10.37671/sb.v7i2.171.

Olimqizi, D.K. (2022). Comparison of Direct and Grammar Translation Methods. *Middle European Scientific Bulletin*, 22, 197-198.

Ong, C.S.B., & Rahim, H.A. (2021). Nativised structural patterns of make light verb construction in Malaysian English. *Concentric*, 47(1), 93-112. https://doi.org/10.1075/consl.00024.rah

Pavlick, E. (2022). Semantic structure in deep learning. *Annual Review of Linguistics*, 8, 447-471. https://doi.org/10.1146/annurev-linguistics-031120-122924

Purmohammad, M., Vorwerg, C., & Abutalebi, J. (2022). The processing of bilingual (switched) compound verbs: Competition of words from different categories for lexical selection. *Bilingualism: Language and Cognition*, 1-13. https://doi.org/10.1017/S1366728921001103

Ramisch, C. (2017). Putting the horses before the cart: identifying multiword expressions before translation. In *International Conference on Computational and Corpus-Based Phraseology* (pp. 69-84). Springer, Cham. https://dx.doi.org/10.1007/978-3-319-69805-2_6

Riemer, N. (2015). *The routledge handbook of semantics.* Abingdon: Routledge.

Shang, D., Xia, F., & Yakovlev, A. (2002). Asynchronous circuit synthesis via direct translation. In *2002 IEEE International Symposium on Circuits and Systems. Proceedings* (Cat. No. 02CH37353) (Vol. 3, pp. III-III). IEEE. https://doi.org/10.1109/ISCAS.2002.1010237

Shen, L. (2004). Aspect Agreement and Light Verbs in Chinese: A Comparison with Japanese. *Journal of East Asian Linguistics*, 13, 141–179. https://doi.org/10.1023/B:JEAL.0000019115.71381.48

Shiwen, Y., & Xiaojing, B. (2014). Rule-based machine translation. In *Routledge Encyclopedia of Translation Technology* (pp. 224-238). Routledge.

Simpson, A. (2001). Focus, Presupposition and Light Predicate Raising in East and Southeast Asia. *Journal of East Asian Linguistics*, 10, 89–128. https://doi.org/10.1023/A:1008360912957

Sneddon, J.N., Adelaar, K.A., Djenar, D., & Ewing, M. (2012). *Indonesian: A comprehensive grammar*. New York: Routledge.

Snider, T. (2021). Light Verbs in Biblical Hebrew. In *Linguistic Studies on Biblical Hebrew* (pp. 169-190). Brill.

Sommers, H. (2003). Machine translation: latest development. In: Mitkov, R. (ed) 2003. *The oxford handbook of computational linguistics.* New York: Oxford University Press.

Spencer, A., Zwicky, A.M. (2001). *The handbook of morphology*. Oxford: Blackwell Publishing.

Srinivas, S., & Legendre, G. (2022). Does D Select the CP in Light Verb Constructions? A Reply to Hankamer & Mikkelsen (2021). *Linguistic Inquiry*, 1-45.

Tron, A., Derevianko, O., Zhumbei, M., & Shpilchak, L. (2022). Light Verb Constructions as Means of Expressing Semelfactive/Multiplicative Meanings in JK Rowling's Discourse

(on the basis of novels 'Harry Potter and the Sorcerer's Stone'and 'Harry Potter and the Chamber of Secrets'). *Amazonia Investiga*, 11(50), 43-54. https://doi.org/10.34069/AI/2022.50.02.5

Vaidya, A., Rambow, O., & Palmer, M. (2019). Syntactic composition and selectional preferences in Hindi light verb constructions. *Linguistic Issues in Language Technology*, 17. https://doi.org/10.33011/lilt.v17i.1419

Van Goethem, K., & Koutsoukos, N. (2022). How typology shapes the constructional network: Denominal verb constructions in English, Dutch and German1. *Zeitschrift für Wortbildung/Journal of Word Formation*, 6(1), 7-57. http://dx.doi.org/10.3726/zwjw.2022.01.01

Villavicencio, A., & Idiart, M. (2019). *Discovering multiword expressions. Natural Language Engineering*, 25(6), 715-733. https://doi.org/10.1017/S1351324919000494

Vincze, V. (2011). *Semi-compositional Noun + Verb Constructions: Theoretical Questions and Computational Linguistics Analyses*. Hungary: University of Szeged.

Wan, Y., Yang, B., Wong, D.F., Chao, S.L., Yao, L., Zhang, H., & Chen, B. (2022). Challenges of Neural Machine Translation for Short Texts. *Computational Linguistics*. https://doi.org/10.1162/coli_a_00435.

Xu, H., Jiang, M., Lin, J., & Huang, C. R. (2022). Light verb variations and varieties of Mandarin Chinese: Comparable corpus driven approaches to grammatical variations. *Corpus Linguistics and Linguistic Theory*, 18(1), 145–173. https://doi.org/10.1515/cllt-2019-0049