International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 04 April 2026
DOI: 10.47191/ijcsrr/V9-i4-46, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

Surface and Underlying Valence of the Verb *"Spread"* in English–Indonesian Translation: Evidence from the OPUS Parallel Corpus

Janardana Dwipayana Ida Bagus[1*], Gede Primahadi Wijaya Rajeg[2], Ketut Artawa[3]
[1,2,3] Faculty of Humanities, Udayana University, Denpasar, Bali, Indonesia

ABSTRACT

Background: This study adopts a corpus-based approach to examine translation empirically through parallel corpora. Focusing on the verb *"spread"* in English–Indonesian data from the OPUS parallel corpus, it identifies empirical equivalents based on frequency and analyzes shifts in valence across languages.

Objective: This study aims to investigate the *surface valence* and *underlying valence* of the verb "spread" in an English–Indonesian parallel corpus. It further seeks to identify its empirical equivalents in the target language by analyzing shifts in syntactic structure and argument realization.

Methods: This study employs a corpus-based approach based on Catford's concept of empirical equivalence, focusing on actual translation usage rather than dictionary-based meanings. The data are drawn from the OPUS English–Indonesian bidirectional parallel corpus (Cooper, 2016), with a focus on medical discourse. A total of 86 instances of the verb *"spread"* are analyzed to identify recurring translation patterns. These patterns are used to determine empirical equivalents and to examine shifts in *surface valence* and *underlying valence*.

Results: The analysis of 86 instances shows that the verb *"spread"* is predominantly translated as *menyebar* (Category A), accounting for 80.23% of the data. Other realizations include *penyebaran* (C) at 10.47%, zero correspondence (0) at 5.81%, *menularkan* (D) at 2.33%, and *memengaruhi* (B) at 1.16%. These findings indicate that *menyebar* functions as the primary empirical equivalent, while the remaining categories reflect less frequent alternatives and instances of structural shift, including nominalization and omission.

Conclusion: This study shows that *menyebar* emerges as the primary translation of *spread* in the target language, reflecting its close alignment with the dominant dispersal sense. The findings also indicate that *spread*, in both the source and target languages, exhibits semantic ambivalence, as its interpretation varies across different contexts and valence realizations.

KEYWORDS: Empirical Equivalent, Valence, Frame Semantics, Spread.

INTRODUCTION

As the world continues to develop, discoveries in many areas of human life have become increasingly advanced, including in linguistics through the emergence of electronic corpora. This development, closely linked to the British National Corpus (Cooper, 2016), enables linguists to examine language empirically through real usage data, including multilingual datasets such as parallel corpora.

In translation, equivalence between the source language (SL) and target language (TL) does not always occur at the structural level due to typological and grammatical differences. According to Catford (1965), equivalence is best observed empirically in actual translation practice, where recurring patterns in a corpus indicate the most natural or preferred equivalents.

Translation often involves lexical and structural shifts, particularly when direct equivalents are unavailable, as seen in the verb *"spread,"* which may be rendered as *menyebar* or *menularkan* depending on context. From Catford's perspective, such variations reflect translation shifts arising from differences in linguistic systems and communicative needs.

These phenomena align with Newmark's (1988) concept of transference or borrowing, especially in scientific and technical domains. Similarly, Montalt and González Davies (2007) emphasize that medical translation often relies on cross-linguistic adaptation to maintain clarity and precision.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 04 April 2026
DOI: 10.47191/ijcsrr/V9-i4-46, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

Verbs play a central role in clause structure, as they determine valence and govern other sentence elements (Van Valin, 2001). In translation, differences between languages often lead to shifts in valence, affecting the number and type of arguments realized in the target language.

Previous studies have examined valence behavior in various verbs, highlighting their flexible argument structures (Rajeg & Rajeg, 2022; Rajeg & Arka, 2023). However, limited research has specifically investigated the verb *"spread"* in English–Indonesian translation, particularly within medical discourse where semantic precision is crucial.

This study addresses this gap by analyzing the valence structures of *"spread"* in English–Indonesian translation using data from the OPUS parallel corpus. By adopting a corpus-based approach, this research aims to identify empirical equivalents and examine shifts in valence across languages.

METHODS

This study employs a corpus-based, theory-driven approach (Tognini-Bonelli, 2001) within a deductive framework guided by Valence Theory (Van Valin, 2001) and Catford's (1965) concept of empirical equivalence. The data are drawn from the OPUS English–Indonesian bidirectional parallel corpus (Cooper, 2016), focusing on medical discourse, with 86 instances of the verb *"spread"* analyzed to identify recurring translation patterns.

The analysis applies theory-informed coding to examine structural patterns, semantic frames, and cross-linguistic correspondences. Clauses are used as the main unit of analysis, allowing the study to observe how the verb functions in terms of *surface valence* and *underlying valence* within a minimal grammatical structure.

The main instrument used in this study is the OPUS parallel corpus, supported by AntConc for extracting instances of the verb *"spread."* The search is conducted using the wildcard form *spread* (e.g., *spread*, *spreads*, *spreading*), and each occurrence is manually verified to ensure the accuracy of the retrieved data.

Microsoft Excel is used to organize and manage the dataset in a structured tabular format, facilitating systematic comparison of collocational patterns and translation outputs. This organization enhances transparency and supports the identification of translation shifts and argument variation (Silverman, 2013).

RESULTS

Table 1. Distribution of Prepositions Immediately Following the Verb "*Spread*"

Category	Frequency	Percentage
V + to	39	45.35%
V + non-preposition	21	23.42%
V + through/throughout	13	15.12%
V + from	5	5.81%
V + in	3	3.49%
V + by	2	2.33%
V + beyond	1	1.16%
V + across	1	1.16%
V + outside	1	1.16%

It is important to note that the data above represent the frequency of prepositions following the verb *"spread"* without explicitly categorizing their syntactic function as either arguments or adjuncts. In this study, the classification of *oblique arguments* and *oblique adjuncts* is determined based on the semantic roles of the constituents following the prepositions rather than the prepositions themselves. Accordingly, elements functioning as goals or endpoints (dispersal) and recipients (filling) are treated as oblique arguments, while source expressions, as well as those marked by *through* or *throughout*, are analyzed as adjuncts of place, with some instances also indicating adjuncts of time, particularly those introduced by *"in."*

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 04 April 2026
DOI: 10.47191/ijcsrr/V9-i4-46, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

Table 2. Distribution of Surface-Level Changes (Explicit Arguments) between SL and TL

Category	Percentage/Frequency
Valence Shift	6.98 %/ 6
Remain	93.02%/ 80

The findings indicate that *surface valence* shifts occur in 6 of 86 cases (6.98%), suggesting a high degree of structural stability in translation. These shifts are mainly linked to zero subject tendencies in Indonesian, where subjects are often omitted, leading to reduced argument realization. Furthermore, reductions are also driven by nominalization and language economy, reflecting the preference for more natural and concise constructions in Indonesian.

Table 3. Sense of *Spread* Frequency Percentage

Sense	Frequency	Percentage
Dispersal	47	54.65%
Filling	25	29.07%
Expansion	14	16.28%

These sense categories reflect how the underlying valence of *spread* is organized across different contextual domains. The **Dispersal** sense represents usages of *spread* in contexts involving large-scale distribution across macro-level spaces, ranging from local regions to countries. The **Filling** sense, on the other hand, captures instances where *spread* refers to micro-level diffusion, particularly among humans, animals, or other living entities, without explicit reference to geographical space. Meanwhile, the **Expansion** sense does not necessarily encode spatial or animate participants; instead, it often reflects more abstract or perceptual themes. In this dataset, Expansion also includes cases where *spread* refers to the transmission of information rather than physical propagation of viruses or pathogens. Importantly, all instances are still situated within a medical discourse context.

Additionally, each sense is characterized by a distinct set of Frame Elements (FEs): *dispersal* involves three FEs (Agent, Theme, Goal), *filling* involves three FEs (Agent, Theme, Recipient), while *expansion* involves a single FE (Theme).

DISCUSSION

Dispersal emerges as the most dominant sense in the dataset. However, a notable reduction in its valence structure is observed, with the three-argument configuration frequently reduced to two arguments, as the agent is often omitted in the data. Quantitatively, this pattern is strongly reflected in the distribution of surface valence in the source language (SL), where 70.9% of instances realize only two arguments (indicating agent omission), while only 11.6% retain the full three-argument structure (n = 86). This suggests that medical discourse in the OPUS corpus tends to foreground the trajectory of the pathogen and its goal or endpoint, rather than the initiating agent.

Similarly, in the *filling* sense, the agent is frequently unrealized, with discourse instead focusing on the pathogen and the recipient (typically an animate entity). The agent in these cases is likely recoverable through either endophoric or exophoric reference. However, since this study does not analyze broader discourse context and is limited to sentence-level alignments in the OPUS corpus, such recoverability remains inferential rather than explicitly traceable.

In the case of the *expansion* sense of *spread*, occurrences are frequently found in passive constructions. In these instances, the agent is systematically omitted as a consequence of passivization, which restructures the clause such that the theme becomes the syntactic and semantic focus. This represents a pivotal phenomenon in which the theme is promoted to the role of primary participant in the clause, while the agent is backgrounded or deleted.

A variation in the use of the preposition *in* is observed across the data. Conceptually, *in* may be associated with a *filling* interpretation, as it can encode the notion of an entity occupying or entering a container. However, the data reveal several instances that challenge this assumption. For example, in phrases such as *in the country*, the preposition does not signal a *filling* relation, but rather identifies a spatial domain, thus aligning more closely with a *dispersal* interpretation. In this case, the locative phrase functions as a place-denoting Goal rather than a recipient typically associated with animate or container-like entities. Furthermore, *in* is also used to mark temporal adjuncts, as seen in expressions such as *in February 2020*, where it encodes time rather than spatial or

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 04 April 2026
DOI: 10.47191/ijcsrr/V9-i4-46, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

participant roles. These variations suggest that the interpretation of *in* cannot be determined solely based on its form, but must be understood in relation to the broader semantic frame in which it occurs.

However, Indonesian demonstrates a more fine-grained distinction in prepositional marking, where *pada* is conventionally used for temporal adjuncts and *di* for locative relations. In contrast to the semantic flexibility of English *in*, this system reduces ambiguity by enforcing a clearer alignment between prepositional form and semantic function. This difference highlights a cross-linguistic asymmetry in how spatial and temporal relations are encoded, with Indonesian exhibiting a more explicit differentiation that may constrain potential frame interpretation.

Another notable insight concerns the role of the preposition *by* in relation to agent defocusing and the acceptability of the *expansion* sense. In passive constructions, *by* canonically marks the agent; however, its presence does not necessarily entail a strong degree of agentivity. In the data, the agent introduced by "*by"* (e.g., *Carnival festivals*) represents a non-prototypical, non-animate causal source, indicating a weakened degree of agentivity. Following David Dowty's Proto-Role Theory (1991), agentivity is a gradient property, and participants lacking volition and sentience occupy a lower position on the agentivity scale. This weakened agentivity facilitates agent defocusing, whereby the agent can be backgrounded or omitted without compromising the acceptability of the clause. As a result, the construction may surface as syntactically monovalent, with only the Theme overtly realized, despite an underlying multi-participant frame. This explains why the *expansion* interpretation remains acceptable: the event is construed as a process affecting the Theme, rather than as an agent-driven transfer, allowing the clause to maintain semantic coherence even in the absence of an explicitly realized agent.

CONCLUSION

This study provides empirical evidence that the meaning of *spread* in medical discourse is predominantly structured by a *dispersal* frame, characterized by systematic agent defocusing and reduced valence realization. The high frequency of two-argument constructions confirms that the propagation of the pathogen, rather than its initiator, constitutes the primary focus of representation. The analysis also challenges form-based assumptions by demonstrating that prepositions such as *in* and *by* do not consistently encode specific participant roles, but instead participate in flexible, context-dependent frame constructions. Cross-linguistic data further show that Indonesian resolves some of this ambiguity through more specialized prepositional marking. From a translational perspective, *menyebar* emerges as the most dominant equivalent of *spread*, reflecting its close alignment with the *dispersal* frame. In contrast, alternatives such as *menular* and *memengaruhi* are employed more selectively, depending on contextual factors—*menular* highlighting the mechanism or cause of transmission, and *memengaruhi* emphasizing broader impact rather than physical propagation. Taken together, these findings underscore that the interpretation and translation of *spread* are governed by underlying semantic frames and discourse priorities, rather than by surface lexical or syntactic correspondences.

Additionally, *spread*, in both the source and target languages, can be considered semantically ambivalent based on the empirical evidence observed in the data. This verb may be broadly associated with verbs of change of location as proposed by Beth Levin (1993), while simultaneously exhibiting flexibility in its argument realization. In particular, *spread* is attested in monovalent constructions, divalent constructions where the agent is only recoverable at the discourse level, and fully trivalent constructions involving agent, theme, and goal or recipient. This variability further supports the view that its interpretation is not fixed, but dynamically shaped by contextual and structural factors.

REFERENCES

1. Cooper I. *Bidirectional corpora and translation studies.* London: Routledge; 2016.
2. Catford J.C. *A linguistic theory of translation: An essay in applied linguistics.* London: Oxford University Press; 1965.
3. Newmark P. *A textbook of translation.* New York: Prentice Hall; 1988.
4. Montalt V., González Davies M. *Medical translation step by step: Learning by drafting.* Manchester: St. Jerome Publishing; 2007.
5. Van Valin RD Jr. *An introduction to syntax.* Cambridge: Cambridge University Press; 2001.
6. Rajeg GPW, Arka IW. Usage-based perspective on argument realisation: A corpus study of Indonesian BUY verbs in applicative construction with -kan. *NUSA: Linguistic Studies of Languages in and around Indonesia.* Tokyo: Tokyo University of Foreign Studies; 2023.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 09 Issue 04 April 2026
DOI: 10.47191/ijcsrr/V9-i4-46, Impact Factor: 8.048
IJCSRR @ 2026
www.ijcsrr.org

7. Rajeg GPW, Rajeg IM. A corpus linguistic study of constructional equivalence for the Indonesian translation of rob and steal based on the Open Subtitles Parallel Corpus. *PAROLE: Journal of Linguistics and Education.* 2022.
8. Tognini-Bonelli E. *Corpus linguistics at work.* Amsterdam: John Benjamins; 2001.
9. Silverman D. *Doing qualitative research.* 4th ed. London: SAGE Publications; 2013.
10. Fillmore C.J. Frame semantics and FrameNet. In: *The Oxford handbook of linguistic analysis.* Oxford: Oxford University Press; 2010.
11. Dowty D. Thematic proto-roles and argument selection. *Language.* 1991;67(3):547–619.
12. AntConc. Version 4.2.0 [computer software]. Tokyo: Waseda University; 2022. Available from: https://www.laurenceanthony.net/software/antconc/
13. OPUS. OPUS: Open Parallel Corpus [Internet]. Uppsala: Uppsala University; 2012– [cited 2026 Apr 20]. Available from: https://opus.nlpl.eu/
14. Levin B. *English verb classes and alternations: A preliminary investigation.* Chicago: University of Chicago Press; 1993.

Cite this Article: Ida Bagus, J.D., Wijaya Rajeg, G.P., Artawa, K. (2026). Surface and Underlying Valence of the Verb "Spread" in English–Indonesian Translation: Evidence from the OPUS Parallel Corpus. International Journal of Current Science Research and Review, 9(4), pp. 2143-2147. DOI: https://doi.org/10.47191/ijcsrr/V9-i4-46