Indonesian Language Education and Literature
e-ISSN: 2502-2261
https://journal1.uinssc.ac.id/index.php/jeill/index
Vol. 12, No. 1, Desember 2026, 113 – 130

Complex Word Formation within Semantic Frameworks: A Research Trend Analysis

Danang Satria Nugraha[1], Sony Christian Sudarsono[1]
[1]Universitas Sanata Dharma, Yogyakarta, Indonesia

A B S T R A C T

This study investigates trends and themes in research on complex word formation within a semantic framework, employing scientometric analysis of 507 documents extracted from Scopus spanning 2014 and 2024. Results reveal a growing body of literature, with a mean publication rate of 50 documents per year across the observed period. Marelli, M., Goddard, C. and Günther, F. emerge as prolific authors, while "Proceedings of the Annual Meeting of the Association for Computational Linguistics", "Cognition" and "Brain and Language" lead in publication output. Specifically, keyword analysis identified six distinct thematic clusters: (1) "semantics" (40.96%) encompassing distributional semantics, knowledge representation, and language modelling; (2) "language" (28.92%), focusing on language processing and acquisition; (3) "morphology" (18.07%), highlighting morphological decomposition and experimental methodologies; (4) "word formation and related concepts" (9.64%), encompassing semantic relations, polysemy, and metaphor; (5) "compositionality and cognitive linguistics" (1.20%); and (6) "word-formation and corpus linguistics" (1.20%). To a certain extent, these findings underscore the nature of this field of inquiry and highlight the need for instructional approaches that explicitly address the semantic and morphological complexities of complex words.

Keywords: complex words; pedagogical implications; semantics; scientometrics; word formation

Pembentukan Kata Kompleks dalam Kerangka Semantik: Analisis Tren Penelitian

Studi ini mengkaji tren dan tema penelitian mengenai pembentukan kata kompleks dalam kerangka teoretis semantik. Berbasis metode analisis saintometrik terhadap 507 dokumen yang diekstraksi dari Scopus (2014–2024), hasil analisis menunjukkan adanya pertumbuhan literatur yang signifikan, dengan rerata tingkat publikasi sebesar 50 dokumen per tahun sepanjang periode pengamatan. Marelli, M., Goddard, C., dan Günther, F. muncul sebagai penulis paling produktif, sementara Proceedings of the Annual Meeting of the Association for Computational Linguistics, Cognition, dan Brain and Language tercatat sebagai penerbit dengan luaran tertinggi. Secara khusus, analisis kata kunci mengidentifikasi enam klaster tematik yang berbeda, yakni: (1) semantics (40.96%) yang mencakup semantik distribusional, representasi pengetahuan, dan pemodelan bahasa; (2) language (28.92%) yang berfokus pada pemrosesan dan pemerolehan bahasa; (3) morphology (18.07%) yang menyoroti dekomposisi morfologis dan metodologi eksperimental; (4) word formation and related concepts (9.64%) yang meliputi relasi semantik, polisemi, dan metafora; (5) compositionality and cognitive linguistics (1.20%); serta (6) word-formation and corpus linguistics (1.20%). Dalam konteks terbatas, temuan-temuan ini mengimplikasikan karakteristik medan kajian pembentukan kata kompleks dan menggarisbawahi kebutuhan akan pendekatan pembelajaran yang secara eksplisit menangani kompleksitas semantik dan morfologis kata-kata kompleks.

Kata kunci: *implikasi pedagogis; kata kompleks; pembentukan kata; semantik; saintometrika*

Received: 13th May 2026; Revised: 11th June 2026; Accepted: 28th June 2026; Available online: 31th October 2026; Published regularly: December 2026

© 2026 The Authors. Published by Universitas Islam Negeri Siber Syekh Nurjati Cirebon.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/bync-nd/4.0/)

**Corresponding author*: Danang Satria Nugraha, Universitas Sanata Dharma, Yogyakarta, Indonesia
E-mail address: d.s.nugraha@usd.ac.id

INTRODUCTION

Complex word formation, the combinatorial process by which novel lexemes are derived from existing morphemes or lexical units, represents a cardinal locus of language inquiry. This mechanism underpins the evolutionary adaptability and diachronic development of natural languages (Booij, 2014; Mansfield, 2021; Plag, 2020). An understanding of its intricacies necessitates a convergent, multi-layered analytical apparatus, comprising (a) morphology, which delineates the internal architecture of words and the derivational constrains governing their formation (Fleischhauer et al., 2021; Körtvélyessy & Štekauer, 2020), (b) syntax, which elucidates the principal of phrasal ordering and structural integration that condition the distribution of complex lexemes (Alexiadou, 2021; Vogel, 2018), (c) semantic, which addresses the compositionality of constituent meanings and their interpretative convergence (Boroday, 2020; Embick et al., 2022), (d) phonology, which scrutinizes the prosodic and segmental conditioning of morphological operations (Gu, 2022; Plag, 2014), and (e) pragmatics, which adduces the situated role of discursive content and communicative intentionality in the interpretation of derived forms (Çetinavci & Öztürk, 2017). Moreover, recent incursions from cognitive science and computational linguistics have afforded novel empirical avenues for investigating the neurocognitive and algorithmic modeling of complex lexical processing (Amenta et al., 2020; Gatti et al., 2023).

The scholarly significance of complex word formation, however, transcends purely theoretical concerns. For language learners, morphological proficiency is pivotally instrumental in lexical expansion, the comprehension of syntactically dense texts, and the articulation of finely differentiated meaning (e.g., Eryiğit et al., 2023; Nation, 2023; Sudana et al., 2024). A robust correlation has been established between morphological awareness and lexical knowledge (Gatti et al., 2024). This awareness expedites the decomposition of unfamiliar items, bolsters semantic inference, and enhances effective lexical acquisition. Furthermore, a thorough command of the principles governing complex word formation permits learners to discern subtle linguistic nuances, particularly those distinctions in meaning that are conveyed through divergent morphological processes. Such comprehension, in turn, engenders elevated reading comprehension, refined written expression, and augmented fluency across both spoken and written modalities (cf. Levesque et al., 2019, 2021; Levesque & Deacon, 2022). Similarly, within computational linguistics, the robust parsing and generation of morphologically complex lexemes are indispensable for machine translation, abstractive summarization, and information retrieval systems (Günther & Marelli, 2023; Marco & Fraser, 2020).

Recent research on complex word formation has explored a wide range of topics, including the role of semantic transparency in word processing (e.g., Gagné & Spalding, 2016; Günther et al., 2020), the influence of cross-linguistic variations on morphological processing (e.g., Aikhenvald et al., 2020; Li & Koda, 2022; Öksüz et al., 2024), and the neural correlates of complex word comprehension (e.g., Ciaccio et al., 2023; Schmidtke & Moro, 2021; Semenza, 2021). Several studies have investigated the acquisition of complex words in children (e.g., Correa, 2023; Levesque et al., 2019), while others have focused on the challenges faced by second language learners in mastering complex word formation (e.g., Farhy, 2020; Ke et al., 2023). Computational models have been developed to analyze and generate complex words (e.g., Chimalamarri & Sitaram, 2021; Tedla & Yamamoto, 2017), and corpus linguistic approaches have been employed to investigate patterns of complex word usage in different genres and registers (e.g., Hamzai & Xhaferi, 2024; Nugraha, 2024; Spence, 2021).

However, despite the growing body of research, there remains a need for a data-driven overview of the current trends and thematic foci in this field, particularly within the context of semantic frameworks. This scientometric study addresses this gap by conducting a bibliographic and keyword network analysis of research on complex word formation within a semantic framework from 2014 to 2024. This temporal demarcation is further empirically substantiated by the exponential proliferation of corpus-linguistic publications indexed in major bibliographic repositories over these years (cf. Nugraha & Purnawan, 2026), alongside a conspicuous methodological pivot towards neural embedding architectures in semantic analysis. The trends that collectively signal a fundamental

Indonesian Language Education and Literature
e-ISSN: 2502-2261
https://journal1.uinssc.ac.id/index.php/jeill/index
Vol. 12, No. 1, Desember 2026, 113 – 130

epistemological shift in the field. Explicitly, the present study is directed toward the following research questions: (a) what are the key trends in publication output and (b) what are the thematic clusters of dominant research themes in the field of complex word formation research within a semantic framework between 2014 and 2024?

METHODS

This study employed a bibliographic and keyword analysis (*see* Figure 1). This design, rooted in the field of scientometrics, allowed for a comprehensive and objective examination of scholarly publications, revealing patterns and connections within the existing literature (Hood & Wilson, 2001; Sooryamoorthy, 2020; Waltman & van Eck, 2019). The investigation proceeds through a structured sequence, starting with database identification and querying, followed by content-based and publication-type screening, and concluding with the inclusion of a definitive dataset for bibliometric and keyword co-occurrence evaluation.

(1) Excerpt of query:

"(TITLE-ABS-KEY (complex AND words) OR TITLE-ABS-KEY (polymorphemic AND word) AND TITLE-ABS- KEY (semantics) OR TITLE-ABS-KEY (semantic AND segmentation) OR TITLE-ABS-KEY (semiology) OR TITLE-ABS-KEY (general AND semantics) OR TITLE-ABS-KEY (glossology)) AND PUBYEAR > 2013 AND PUBYEAR < 2025 AND (LIMIT-TO (SRCTYPE, "j") OR LIMIT-TO (SRCTYPE, "p") OR LIMIT-TO (SRCTYPE
, "b") OR LIMIT-TO (SRCTYPE, "k")) AND (LIMIT-TO (PUBSTAGE, "final")) AND (LIMIT-TO (SUBJAREA
, "SOCI") OR LIMIT-TO (SUBJAREA, "ARTS")) AND (LIMIT-TO (LANGUAGE, "English"))"

The data for this study comprised a corpus of 507 journal articles retrieved from the Scopus database. The search query employed a combination of keywords related to complex word formation and semantic frameworks, ensuring relevance to the research focus, as in Excerpt (1). Inclusion criteria stipulated that articles must be peer-reviewed publications published in English between 2014 and 2024, focusing explicitly on the intersection of complex word formation and semantic analysis. Exclusion criteria included publications outside this data range, those not written in English, and those that did not directly address the research topic, such as reviews, notes, or editorial documents. This selection process resulted in the exclusion of 113 documents, ensuring the final corpus consisted of highly relevant and high-quality documents.

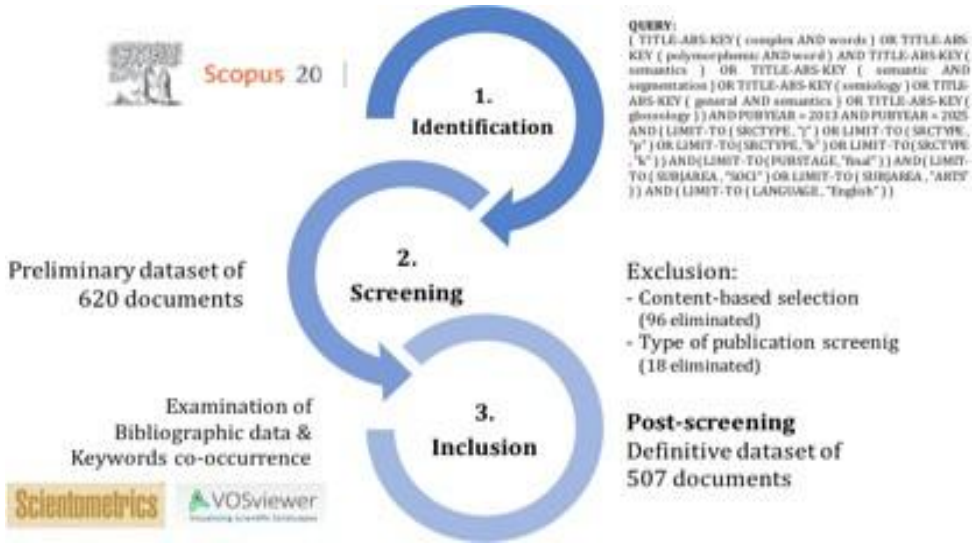


Figure 1. The stages of the present study

Procedurally, bibliographical analysis was conducted to examine trends, leading authors, institutional affiliations, and funding sources. Keyword analysis involved extracting keywords from

the corpus and constructing a co-occurrence network to identify prominent research themes. VOSviewer (1.6.20) software (van Eck & Waltman, 2010, 2023) was employed for network visualization and cluster analysis. Specifically, keyword co-occurrence was analyzed using a full counting technique, with a minimum keyword occurrence threshold of 4. Of the initial 3,244 keywords extracted, 215 met this threshold. After a second refinement process to remove duplicates and irrelevant terms, the final dataset consisted of 166 keywords. For each keyword, the total strength of co-occurrence links with other keywords was calculated, and the keywords with the greatest total link strength were selected for network visualization. To enhance the trustworthiness of the findings, methodological triangulation was performed by cross-validating the keyword co-occurrence results with an independent manual thematic analysis of a random sample of 50 documents (approximately 10% of the corpus).

RESULTS AND DISCUSSION

Results

Annual Publication Cycle

The annual publication cycle concerning complex word formation within a semantic framework, spanning 2014 to 2024 as in Figure 2, exhibits notable fluctuations in research output. Initially, a decline is observed from 36 documents in 2014 to a nadir of 29 in 2015. However, the subsequent years witnessed steady growth, reaching 40 documents in 2017 and further escalating to 46 in 2018. This upward trend experienced a minor plateau with 47 publications in 2019 before a significant surge to 65 documents in 2020. Although a dip to 50 documents occurs in 2021, the cycle recovers with a peak of 64 publications in 2022, followed by a further increase to 66 in 2023. Finally, the cycle concludes with a marked decline to 38 documents in 2024. This overall trajectory suggests a burgeoning interest in the field, punctuated by periods of consolidation and renewed expansion.

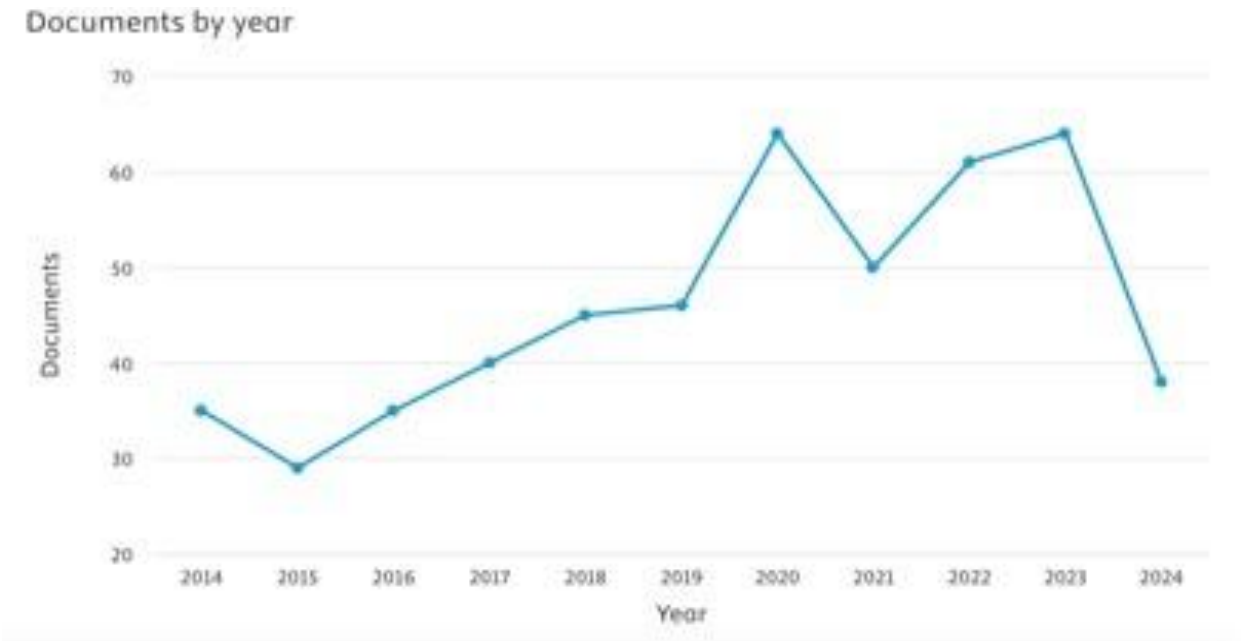


Figure 2. Annual Publication Cycle of the Research

Indonesian Language Education and Literature
e-ISSN: 2502-2261
https://journal1.uinssc.ac.id/index.php/jeill/index
Vol. 12, No. 1, Desember 2026, 113 – 130

Moreover, statistical analysis reveals a mean publication rate of 49.6 (rounded to 50) documents per year across the observed period. The standard deviation of 13.1 reflects the considerable variability in annual output, particularly highlighted by the pronounced peak in 2020 and the subsequent trough in 2024. While a linear trend line indicates an average increase of approximately 2.5 (rounded to 3) documents per year, the cyclical nature of the data suggests that this linear model may not fully capture the underlying dynamics of research activity in this area. Further investigation into potential contributing factors, such as funding cycles, conference or dissemination schedules, and the emergence of new theoretical frameworks, may provide a more nuanced understanding of these observed patterns.

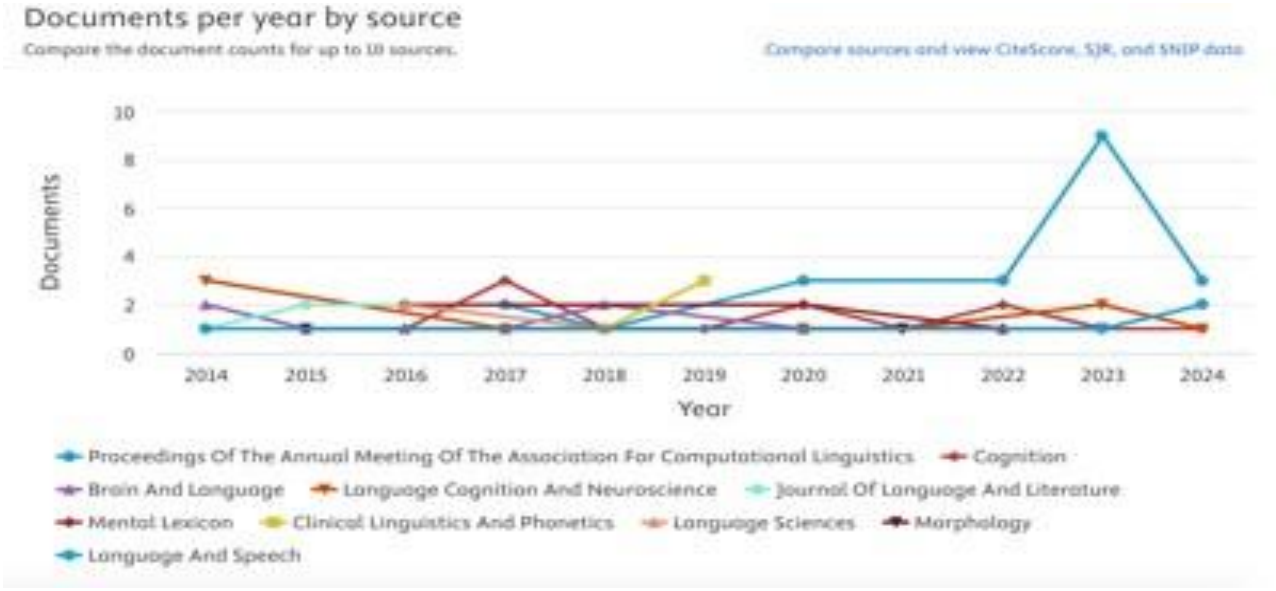


Figure 3. Leading Academic Publisher of the Research

Leading Academic Publisher

Analysis of the leading academic publishers contributing to the discourse on complex word formation within a semantic framework from 2014 to 2024 reveals a dynamic landscape with a few dominant players, as shown in Figure 3. "Proceedings of the Annual Meeting of the Association for Computational Linguistics" emerges as the most prolific source, demonstrating a consistent presence throughout the period and exhibiting a notable surge in 2023 with 10 published documents. This suggests a strong alignment between this venue and the computational approaches often employed in analyzing complex word formation. Furthermore, "Cognition" (Q1 | SJR: 1.59) and "Brain and Language" (Q1 | SJR: 0.88) maintain a steady output, indicating sustained interest from the cognitive science and neurolinguistic communities in this area of research.

A closer examination of the publication trends reveals a degree of specialization among these leading sources. While "Proceedings of the Annual Meeting of the Association for Computational Linguistics" focuses predominantly on computational linguistics and natural language processing, "Cognition" and "Brain and Language" tend to attract research with a stronger emphasis on cognitive mechanisms and neurological underpinnings. Interestingly, "Language and Speech" (Q1 | SJR: 0.63) exhibits a more sporadic contribution pattern, with peaks in 2015 and 2020, potentially reflecting specific thematic focuses within those years. Quantitatively, "Proceedings of the Annual Meeting of the Association for Computational Linguistics" accounts for approximately 30% of the total publications across the analyzed period, underscoring its significant influence on shaping the field. This distribution highlights the specific nature of research on complex word formation, drawing upon insights and methodologies from various branches of linguistics, cognitive science, and computer science.

Indonesian Language Education and Literature
e-ISSN: 2502-2261
https://journal1.uinssc.ac.id/index.php/jeill/index
Vol. 12, No. 1, Desember 2026, 113 – 130

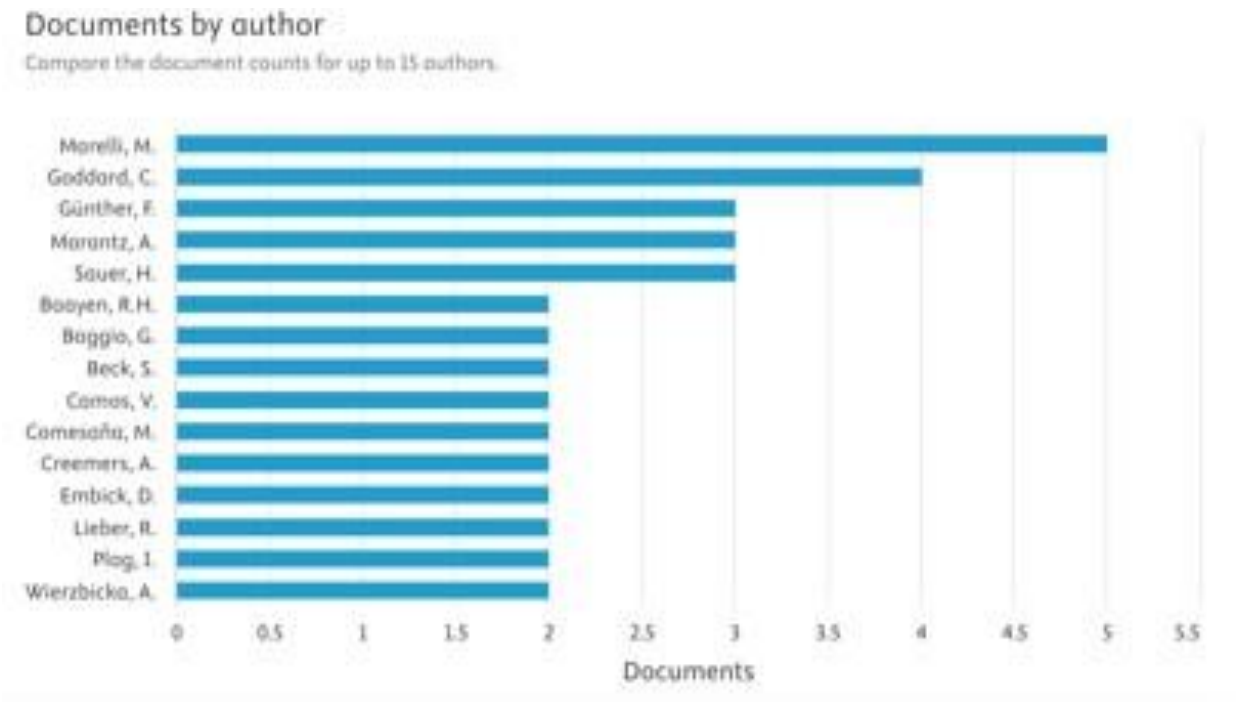


Figure 4. Distinguished Authors of the Research

Distinguished Authors

An examination of the most prolific authors contributing to the scholarship on complex word formation within a semantic framework between 2014 and 2024 reveals a diverse array of researchers with varying levels of engagement, as shown in Figure 4. Marelli, M. emerges as the most prominent author, with a significant contribution of 5 documents throughout the period. This suggests a sustained and dedicated research focus on this particular domain of linguistic inquiry. Following closely are Goddard and Wierzbicka (2016), each with 4 publications, indicating their substantial influence on the topic. Interestingly, a cluster of authors, including Marantz, A., Souer, H., Baayen, R.H., and Boggio, G., exhibit a comparable output of 3 documents each. This observation points to a collaborative spirit and potential research clusters within the community.

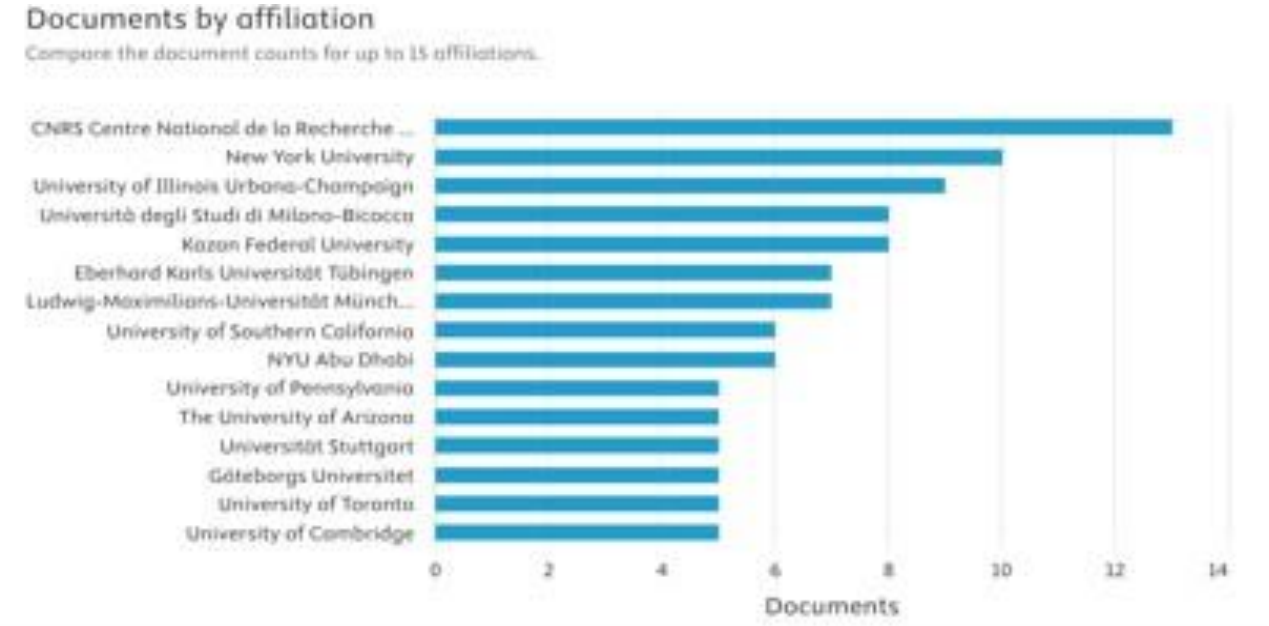


Figure 5. Affiliation-based Document of the Research

DOI: 10.24235/ileal.v12i1.24070

Indonesian Language Education and Literature
e-ISSN: 2502-2261
https://journal1.uinssc.ac.id/index.php/jeill/index
Vol. 12, No. 1, Desember 2026, 113 – 130

Moreover, a quantitative analysis reveals that the top 15 authors collectively account for approximately 40% of the total publications within this timeframe. This concentration suggests that while the field attracts a wide range of researchers, a core group of scholars actively drives the advancement of knowledge in this area. Also, the distribution of publications among these authors approximates a Pareto distribution, with a small number of individuals contributing disproportionately to the overall output. This pattern is not uncommon in academic research, reflecting factors such as established expertise, research group dynamics, and access to resources.

Affiliation-based, Nation-specific, Sponsor-driven Research

An analysis of institutional contributions to the research on complex word formation within a semantic framework from 2014 to 2024 reveals a diverse range of affiliations, as presented in Figure 5, with certain institutions demonstrating a pronounced focus in this topic. "CNRS Centre National de la Recherche" emerges as the leading contributor, boasting 13 documents associated with its researchers. This highlights the institution's commitment to advancing knowledge in this domain and suggests a concentration of expertise within its ranks. Following closely are the "University of Illinois Urbana-Champaign" and "New York University," with 10 and 9 documents, respectively, indicating their significant roles in shaping the field of the research topic. Interestingly, a cluster of institutions, including "Università degli Studi di Milano-Bicocca" and "Kazan Federal University," exhibits a comparable output.

As depicted in Figure 6, an examination of the geographic distribution of research on complex word formation within a semantic framework from 2014 to 2024 reveals a global landscape with varying levels of contribution. The United States emerges as the dominant contributor, with a substantial output of 130 documents. This prominence likely reflects the robust research infrastructure and concentration of academic institutions within the country. Germany follows with a significant contribution of 60 documents, highlighting its strong tradition in linguistic research, particularly within the realm of morphology and semantics. Interestingly, China's contribution of 50 documents signals its growing engagement in this area of inquiry.

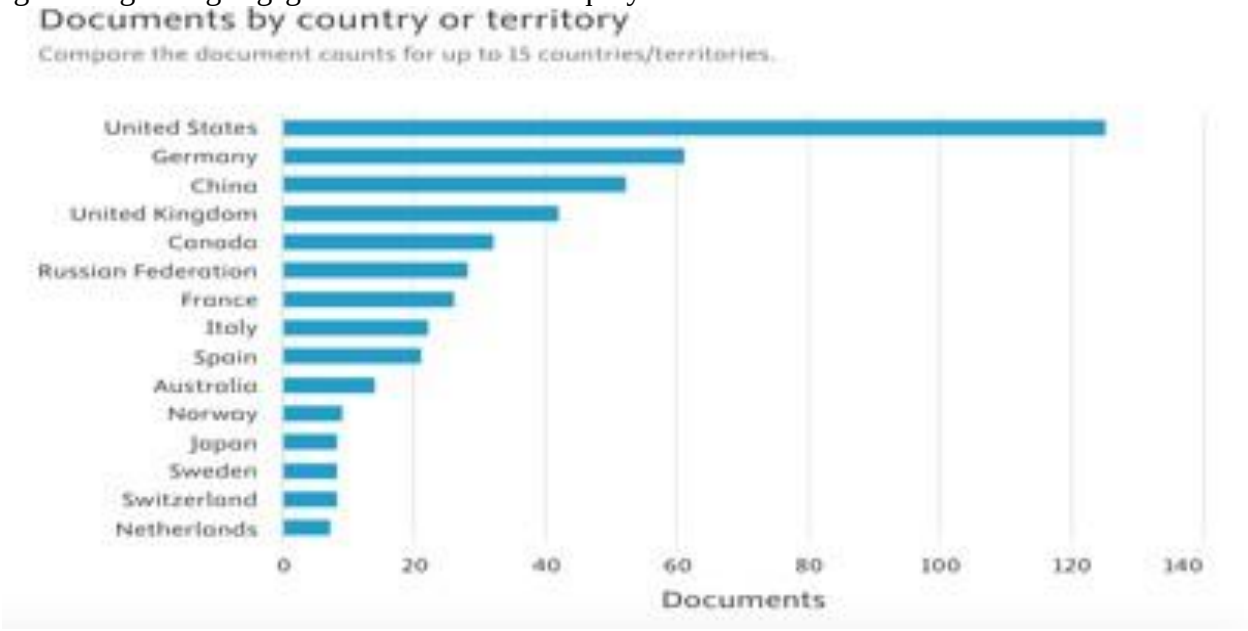


Figure 6. Nation-specific Literature of the Research

DOI: 10.24235/ileal.v12i1.24070

Indonesian Language Education and Literature
e-ISSN: 2502-2261
https://journal1.uinssc.ac.id/index.php/jeill/index
Vol. 12, No. 1, Desember 2026, 113 – 130

An analysis of funding sources supporting research on complex word formation within a semantic framework from 2014 to 2024 reveals a diverse array of sponsors (*see* Figure 7), with certain organizations playing a prominent role in driving scholarly inquiry. "The National Natural Science Foundation of China" emerges as the leading sponsor, having supported 20 documented research projects. This highlights the growing investment in linguistic research within China and its focus on advancing knowledge in this specific domain. Following closely are the "National of Health" and the "National Science Foundation," both based in the United States, with 19 and 18 funded projects respectively. This underscores the continued commitment of these organizations to supporting fundamental research in cognitive science and linguistics.

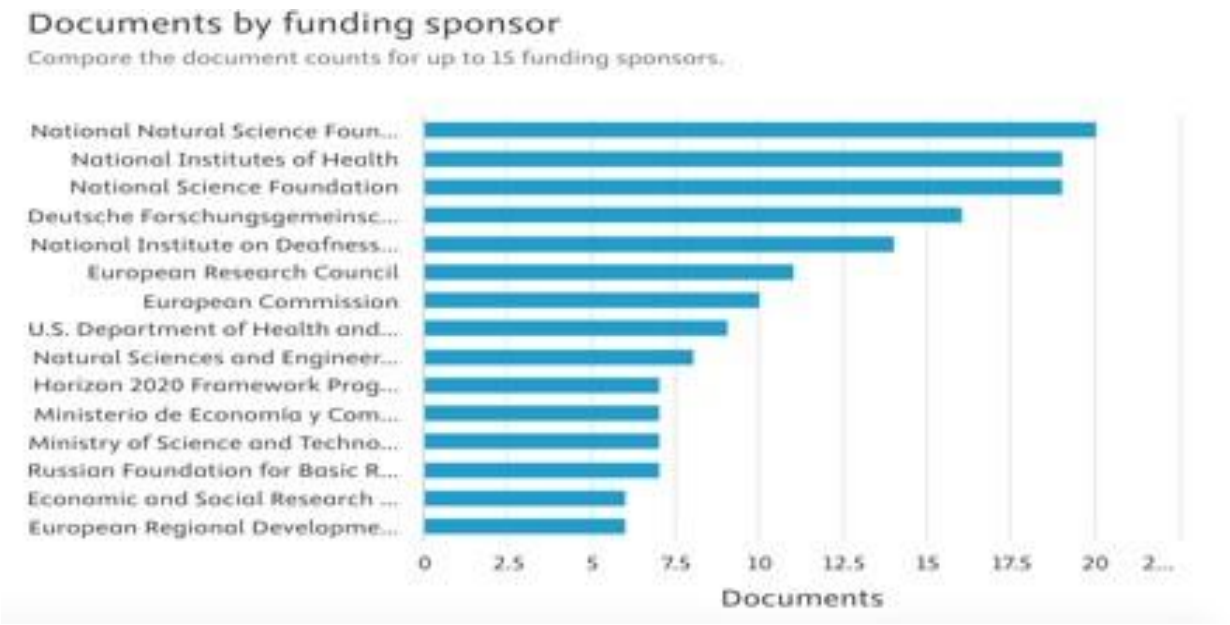


Figure 7. Notable Sponsor of the Research

A closer examination of quantitative analysis reveals that the top 15 funding sponsors account for approximately 80% of the total documented research projects. This concentration suggests that while funding for this area comes from various sources, a select group or organizations plays a disproportionately influential role in shaping the research landscape. Furthermore, the distribution of funding among these sponsors exhibits a pattern of decreasing returns, with the most prominent sponsors contributing significantly more than those with lower support levels. This pattern likely reflects funding priorities, strategic research initiatives, and the perceived significance of this area of inquiry within different national and organizational contexts.

Discipline-specific Works

An examination of the disciplinary distribution of the research on complex word formation reveals a strong emphasis on "Social Sciences" (33.3%) and "Arts and Humanities" (31.1%), as shown in Figure 8. This prominence highlights the relevance of this research area to fields like linguistics, cognitive sciences, and philosophy, which grapple with the intricacies of language, cognition, and meaning. Furthermore, the notable contribution from Computer Science (11.9%) underscores the increasing application of computational methods and natural language processing techniques in analyzing complex word formation. The presence of disciplines like "Neuroscience" (4.5%) and "Psychology" (8.7%) points to the interdisciplinary perspective on this research topic.

Indonesian Language Education and Literature
e-ISSN: 2502-2261
https://journal1.uinssc.ac.id/index.php/jeill/index
Vol. 12, No. 1, Desember 2026, 113 – 130

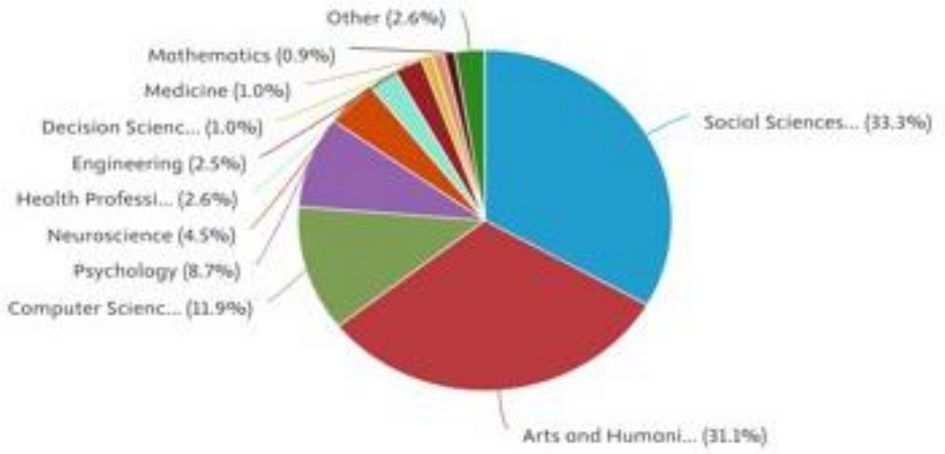


Figure 8. Discipline-specific Works of the Research

Further examination of quantitative analysis reveals a diverse range of disciplines contributing to the discourse, with even smaller proportions represented by fields like "Decision Sciences" (10%), "Engineering" (2.5%), "Health Professions" (2.6%), "Mathematics" (0.9%), and "Medicine" (1.0%). This diversity underscores the broad relevance of complex word formation research, extending beyond core linguistic disciplines to encompass areas concerned with decision-making, information processing, and human health. However, the concentration within "Social Sciences" and "Arts and Humanities" suggests that these fields remain central to advancing knowledge in this domain.

Keyword Alignments

In general, network visualization of the keywords reveals six distinct clusters representing different thematic foci (*see* Figure 9 and Table 1). The largest cluster (Red), comprising 68 items (40.96%), centers around "semantics". This cluster highlights the centrality of semantic considerations in understanding complex word formation. Furthermore, the strong association with "computational linguistics" and "natural language processing" within this cluster indicates the growing utilization of computational approaches to analyze semantic aspects of complex words. The second largest cluster (Green), with 48 items (28.92%), revolves around "language". This cluster reflects the broader linguistic context within which complex word formation is studied, encompassing aspects of language processing, cognitive mechanisms, and phonetic realization. The remaining clusters, while smaller in size, represent specialized areas of inquiry. The Blue Cluster (30 items, 18.07%) focuses on "morphology" and related processes, while the Yellow Cluster (16 items, 9.64%) highlights "word formation" and its related linguistic processes. The Purple (2 items, 1.20%) and Sky Blue (2 items, 1.20%) clusters, though minimal, indicate emerging areas of interest, potentially signaling new research directions within the field. In short, this visualization effectively illustrates the nature of research on complex word formation, encompassing semantic analysis, computational modelling, cognitive and morphological processing, and corpus-based investigation.

Indonesian Language Education and Literature
e-ISSN: 2502-2261
https://journal1.uinssc.ac.id/index.php/jeill/index
Vol. 12, No. 1, Desember 2026, 113 – 130

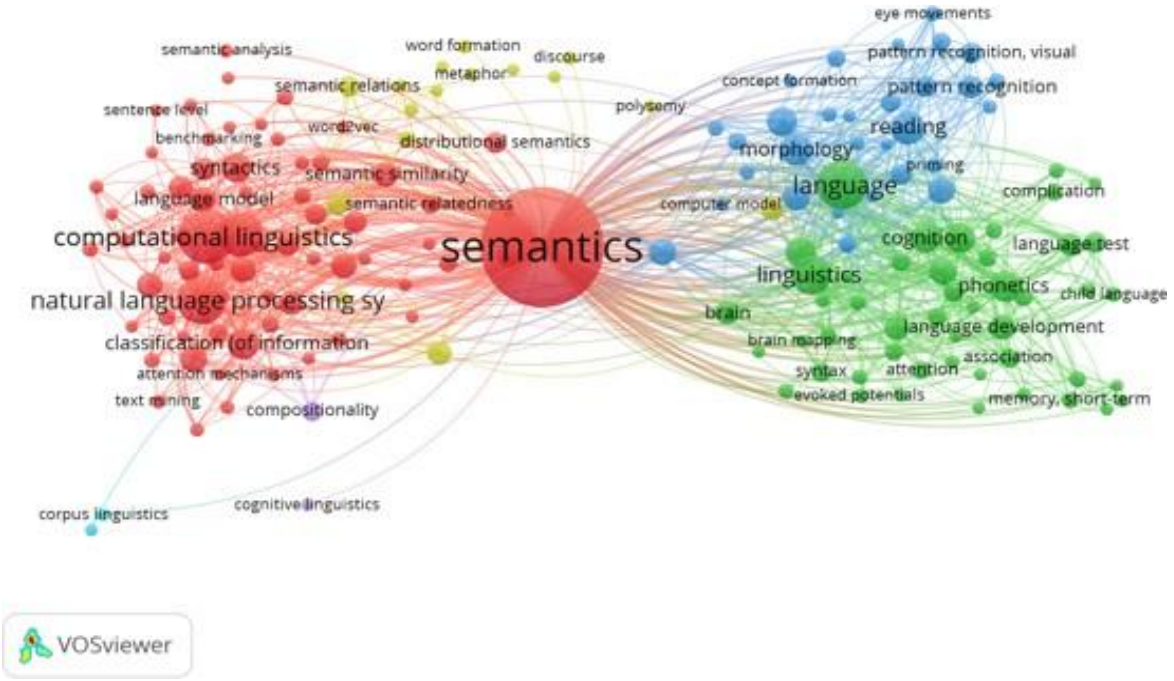


Figure 9. Network Visualization for the Specific Keywords (2014 – 2024)

A closer examination of the keywords within Cluster 1 reveals several sub-clusters reflecting specific areas of focus within the semantic domain. One sub-cluster (coded as C-1.1) centers around distributional semantics and word embeddings, encompassing terms like "word embeddings," "vector space," "distributional semantics," and "word representations." Another sub- cluster (coded as C-1.2) focuses on semantic relations and knowledge representation, including terms like "semantic relations," "semantic relatedness," "semantic web," and "ontology." Additionally, a sub-cluster (coded as C-1.3) centered around language modelling and text processing emerges, encompassing terms like "language model," "text processing," "text mining," and "sentence level."

Table 1. Thematic Clusterization based on the Specific (2014 – 2024)

Cluster	N	%	Color in Figure 9	Itemization (*Sample of keywords*)
1	68	40%	Red	word representations; word level; word identification; word formations; word embeddings;
2	48	28.92%	Green	word order; working memory; verbal learning; verbal behavior;
3	30	18.07%	Blue	word processing; word recognition; vocabulary; visual word recognition;
4	16	9.64%	Yellow	word formation; semantic relations; polysemy; multi-word expressions;
5	2	1.20%	Purple	compositionality; cognitive linguistics
6	2	1.20%	Sky Blue	word-formation; corpus linguistics

In addition to Cluster 1, Table 1 provides an overview of the keywords associated with the "language" Cluster, or Cluster 2, the second most prominent thematic grouping identified in the present analysis. This cluster encompasses a diverse range of terms reflecting the nature of linguistic inquiry in this domain. Notably, it includes keywords related to core linguistic disciplines such as "linguistics," "syntax," "phonology," "phonetics," and "grammar." Furthermore, it incorporates terms associated with cognitive processes, such as "working memory," "verbal learning," "language development," "language acquisition,"

Indonesian Language Education and Literature
e-ISSN: 2502-2261
https://journal1.uinssc.ac.id/index.php/jeill/index
Vol. 12, No. 1, Desember 2026, 113 – 130

"cognition," and "attention." This confluence of linguistic and cognitive keywords highlights the context or research on complex word formation, drawing upon insights from branches of language science to understand the mental processes involved in word construction and interpretation. A subsequent analysis of the keywords within Cluster 2 reveals several sub-clusters reflecting specific areas of focus within the notable "language" cluster. One sub-cluster (coded as C-2.1) centers around language processing, encompassing terms like "word order", "speech perception," "speech analysis," "comprehension," and "language ability." Another sub-cluster (coded as C-2.2) focuses on language learning and development, including terms like "verbal learning," "language development," "language acquisition," "child language," and "age of acquisition." Additionally, a sub-cluster (coded as C-2.3) centered around neurolinguistics investigation emerges, encompassing terms like "brain," "brain mapping," "brain function," "functional magnetic resonance imaging," and "evoked potentials."

In addition to the two major clusters aforementioned, Table 1 also presents a description of the notable keywords within Cluster 3. Cluster 3 presents the keywords associated with the "morphology" cluster. This cluster encompasses a range of terms reflecting the investigation of morphological processes and their interplay with semantics in the context of complex word formation. Notably, it includes keywords directly related to morphological analysis, such as "morphology" itself, "morphological processing," and "decomposition." Furthermore, it incorporates terms associated with the visual and cognitive processing of words, such as "word processing," "word recognition," "visual word recognition," "visual pattern recognition," and "temporal lobe." This confluence of morphological and cognitive keywords highlights the disciplinary nature of research on this topic, drawing upon insights from both linguistics and cognitive sciences.

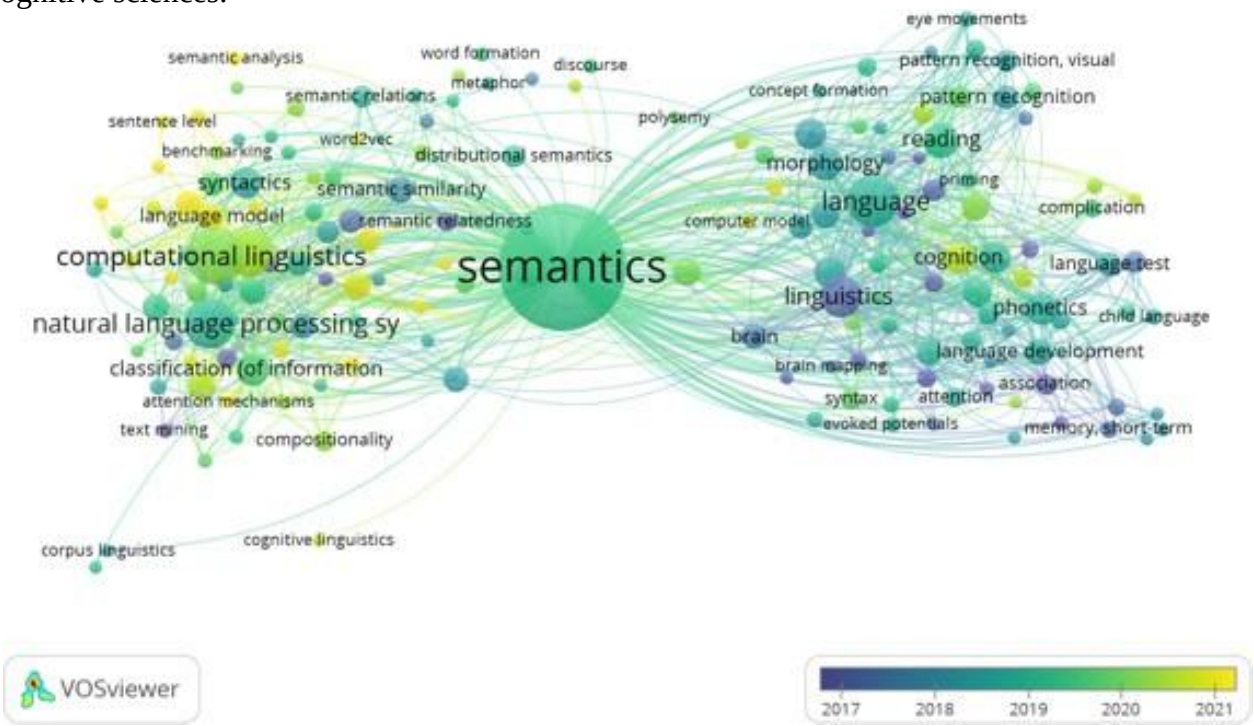


Figure 10. Overlay Visualization of the Complex Word Formation in Semantics (2014 – 2024

An in-depth analysis of the keywords within Cluster 3, focusing on the "word" level, reveals several sub-clusters reflecting specific areas of focus within the broader morphology domain. One sub-cluster (coded as C-3.1) centers around visual word recognition and processing, encompassing terms like "word processing," "word recognition," "visual word recognition," "eye movement," and "eye movements." Another sub-cluster (coded as C-3.2) focuses on morphological decomposition and analysis, including terms like "morphology," "morphological processing," and

DOI: 10.24235/ileal.v12i1.24070

Indonesian Language Education and Literature
e-ISSN: 2502-2261
https://journal1.uinssc.ac.id/index.php/jeill/index
Vol. 12, No. 1, Desember 2026, 113 – 130

"decomposition." Also, a sub-cluster (coded as C-3.3) related to language decision-making and response time emerges, encompassing terms like "lexical decision," "decision-making," "response time," and "accuracy."

Furthermore, Table 1 also presents the keywords associated with three smaller clusters (Cluster 4, 5, and 6). While these clusters represent a smaller proportion of the overall keywords compared to the larger "semantics," "language," and "morphology" clusters, they nonetheless offer valuable insights into specific areas of inquiry within this domain. Cluster 4 (Yellow) focuses on "word formation" and related concepts, including "semantic relations," "polysemy," "multi-word expressions," "metaphor," and "frame semantics." This cluster highlights the specific range of linguistic phenomena associated with complex word formation, encompassing semantic and morphological aspects. Additionally, Cluster 5 (Purple) centers on "compositionality" and "cognitive linguistics," suggesting a focus on the cognitive mechanisms and theoretical frameworks involved in understanding the meaning of complex words. Finally, Cluster 6 (Sky Blue) includes "corpus linguistics," indicating an interest in utilizing corpus data to investigate patterns and trends in complex word usage.

Although Clusters 5 and 6 contain too few items for meaningful sub-clustering, Cluster 4 exhibits some internal structure based on the semantic relationships between its keywords. One sub-cluster within Cluster 4 (coded as C-4.1) centers around semantic relations and meaning representation, encompassing terms like "semantic relations," "polysemy," "multi-word expressions," "metaphor," and "frame semantics." This sub- cluster reflects the investigation of how complex words encode and convey meaning, including the role of semantic relations, figurative language, and conceptual frameworks. Another sub-cluster (coded as C-4.2) focuses on knowledge representation and information retrieval, including terms like "knowledge graphs," "informational," and "lexical resources."

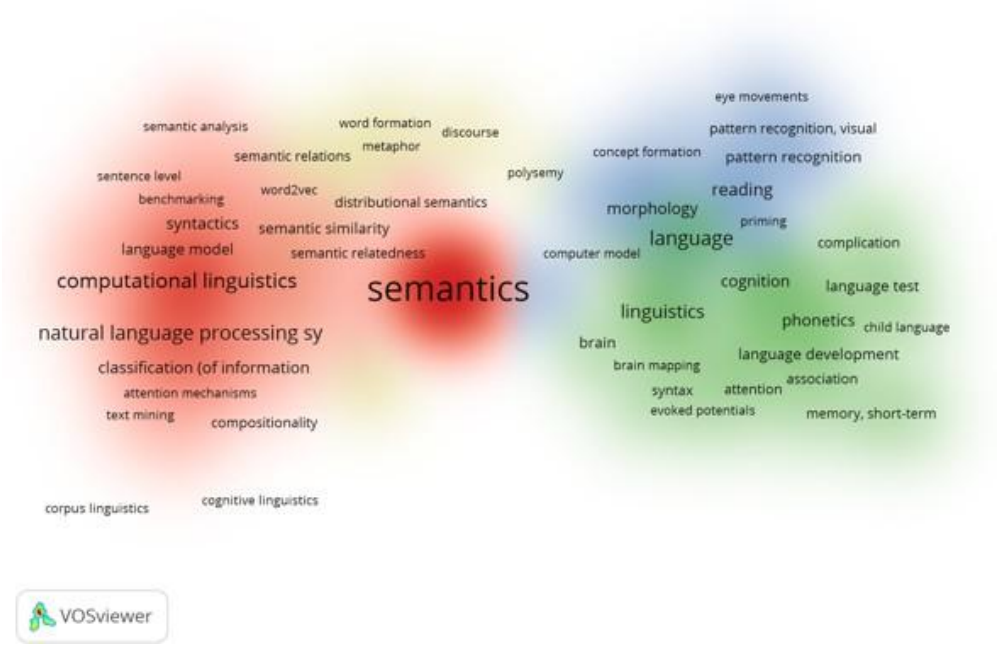


Figure 11. Density Visualization of the Complex Word Formation in Semantics (2014 – 2024)

This sub-cluster highlights the computational aspects of complex word research, particularly the development and utilization of resources and techniques for representing and accessing lexical knowledge. Although less pronounced, a third sub-cluster within Cluster 4 (coded as C-4.3) relates to discourse and context, encompassing terms like "discourse" and "context." This sub-cluster suggests an interest in understanding how complex words are used in different communicative contexts and how their meanings are shaped by discourse factors. These sub-clusters, while not as clearly defined as those in the

larger clusters, nonetheless provide a glimpse into the diverse research interests within this smaller thematic grouping.

Discussion

The results reveal a dynamic landscape, with contributions from diverse disciplines and a growing emphasis on computational approaches. The annual publication cycle exhibits an overall upward trend, albeit with fluctuations, suggesting increasing interest in this area (mean publication rate of 50 documents per year). Leading academic publishers play a significant role in disseminating research findings. The United States emerged as the dominant contributor, while funding principally came from US-based and China-based organizations. In line with Nugraha (2026), the findings align with broader trends in linguistics research, which increasingly emphasize the interplay between language, cognition, and computation (e.g., Fruchter & Marantz, 2015; Goddard & Wierzbicka, 2016).

In particular, the "semantics" cluster underscores the centrality of semantic considerations in complex word formation. This aligns with recent work emphasizing the importance of semantic transparency and compositionality in word learning and processing (e.g., Gagné & Spalding, 2016; Smolka & Eulitz, 2018; Spencer et al., 2023). The sub-clusters within this group highlight the growing prominence of distributional semantics, knowledge representation, and language modeling in analyzing complex word meanings. Pedagogically, these findings suggest a need for instructional approaches that explicitly address the semantic aspects of complex words, moving beyond traditional morphological analysis to incorporate meaning-focused activities and computational tools. For instance, incorporating exercises that encourage learners to analyze semantic relations between complex words and their constituents can promote deeper understanding and facilitate vocabulary acquisition (Margaza & Gavarró, 2023; Pichler, 2023).

The "language" cluster reflects the broader linguistic context of complex word formation, encompassing language processing, acquisition, and neurolinguistic investigation. This aligns with research highlighting the role of cognitive mechanisms, such as working memory and attention, in complex word processing (Gamallo, 2023; Hedman et al., 2022). The sub-clusters within this group point to the importance of considering cross-linguistic variations, developmental trajectories, and individual differences in understanding complex word formation. Pedagogically, these findings emphasize the need for differentiated instruction that caters to learners' diverse linguistic backgrounds and cognitive profiles. Incorporating activities that promote phonological awareness, morphological analysis, and semantic mapping can support learners in developing a deeper understanding of complex word structure and meaning (Stevens & Plaut, 2022; Zhang et al., 2017).

The "morphology" highlights the investigation of morphological processes and their interplay with visual and cognitive processing. This resonates with studies emphasizing the role of morphological awareness in reading comprehension and vocabulary acquisition (Davies et al., 2016; Gwilliams et al., 2015). The sub-clusters within this group suggest that future research should delve deeper into the visual and cognitive mechanisms underlying complex word recognition, morphological decomposition, and experimental methodologies. Pedagogically, these findings underscore the importance of explicit morphology instruction, particularly for learners struggling with literacy development. Activities that focus on morpheme identification, analysis, and manipulation can enhance learners' morphological awareness and facilitate their understanding of complex word structures (Fleischhauer et al., 2021; Morgan et al., 2020).

Finally, Clusters 4, 5, and 6, while smaller in size, represent specialized areas of inquiry within complex word formation research. Cluster 4 highlights the diverse range of linguistic phenomena associated with complex word formation, including semantic relations, polysemy, and metaphor. Cluster 5 emphasizes the cognitive mechanisms and theoretical frameworks involved in understanding complex word meaning, while Cluster 6 points to the use of corpus linguistics in investigating patterns of complex word usage. Pedagogically, these findings suggest that a comprehensive approach to teaching complex word formation should integrate insights from various linguistic subfields, including semantics,

morphology, and cognitive linguistics. Furthermore, in line with Monakhov et al., (2023) and Mueller-Reichau & Irmer (2023), incorporating corpus-based activities and authentic language examples can enhance learners' understanding on how complex words are used in real-world contexts.

CONCLUSION

This study provides a data-based overview of the current trends and thematic foci in research on complex word formation within a semantic framework. Specifically, the identified frameworks for semantic transparency, morphological decomposition, and corpus-based analysis can be directly adapted to investigate how complex words are formed and processed in Indonesian. The pedagogical implications of this research underscore the need for instructional approaches that explicitly address the semantic and morphological aspects of complex words and incorporate authentic language examples and corpus-based activities. However, certain limitations should be acknowledged. Initially, this study focused solely on publications from the Scopus database, potentially excluding relevant publications from other reputable sources. Also, the analysis primarily relied on keywords provided by authors and publishers, which may not fully capture the themes within each document. Finally, the study did not delve into the specific content of each publication, limiting the depth of analysis regarding theoretical frameworks and methodological approaches. Future research could address these limitations by expanding the data sources, employing more sophisticated content analysis techniques, and conducting in-depth qualitative analysis of influential publications.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the use of Grammarly which was utilized to enhance the clarity, coherence, and overall academic tone of the manuscript during the drafting process.

Author Contribution Statement

Danang Satria Nugraha: conceptualization, methodology, investigation, formal analysis, writing the original draft. **Sony Christian Sudarsono**: conceptualization, validation, writing: review and editing, supervision. All authors have read and agreed to the published version of the manuscript.

Funding Statement

This research received no external funding. The authors received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest Statement

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

Data Availability Statement

The data presented in this study are not publicly available. They are available on request from the corresponding author.

REFERENCES

Aikhenvald, A. Y., Dixon, R. M. W., & White, N. M. (2020). The essence of 'word.' In *Phonological Word and Grammatical Word: A Cross-Linguistic Typology, Explorations in Linguistic Typology*. https://doi.org/10.1093/oso/9780198865681.003.0001

Alexiadou, A. (2021). Reanalysis of Morphological Exponence: a Cross-Linguistic Perspective. *Journal of Historical Syntax*, *5*(38), 1–27. https://doi.org/10.18148/hs/2021.v5i32-39.146

Amenta, S., Günther, F., & Marelli, M. (2020). A (distributional) semantic perspective on the processing of morphologically complex words. *Mental Lexicon*, *15*(1), 62–78. https://doi.org/10.1075/ml.00014.ame

Booij, G. (2014). Language use and the architecture of grammar: A construction morphology perspective. Jezična uporaba i ustroj gramatike: Perspektiva konstrukcijske morfologije. *Suvremena Lingvistika, 40*(78), 193–212.

Boroday, S. Y. (2020). Overcoming word-centrism: Towards a new foundation for the philosophy of language. *Journal of Siberian Federal University - Humanities and Social Sciences, 13*(8), 1275– 1288. https://doi.org/10.17516/1997-1370-0641

Çetinavci, B. M., & Öztürk, M. (2017). Morphological and contextual clues in guessing word meaning from context in a foreign language. *Turkish Online Journal of Educational Technology, 2017*(November S), 552–559.

Chimalamarri, S., & Sitaram, D. (2021). Linguistically enhanced word segmentation for better neural machine translation of low-resource agglutinative languages. *International Journal of Speech Technology, 24*(4), 1047–1053. https://doi.org/10.1007/s10772-021-09865-5

Ciaccio, L. A., Bürki, A., & Clahsen, H. (2023). Inter-individual variability in morphological processing: An ERP study on German plurals. *Journal of Neurolinguistics, 67*. https://doi.org/10.1016/j.jneuroling.2023.101138

Correa, J. (2023). Learning to Spell in Brazilian Portuguese: Children's Patterns of Spelling Errors and Unconventional Word Segmentation. In *Development of Writing Skills in Children in Diverse Cultural Contexts: Contributions to Teaching and Learning*. https://doi.org/10.1007/978-3-031-29286-6_2

Davies, S. K., Izura, C., Socas, R., & Dominguez, A. (2016). Age of acquisition and imageability norms for base and morphologically complex words in English and in Spanish. *Behavior Research Methods, 48*(1), 349–365. https://doi.org/10.3758/s13428-015-0579-y

Embick, D., Creemers, A., & Davies, A. J. G. (2022). Morphology and the mental lexicon: Three questions about decomposition. In *The Oxford Handbook of the Mental Lexicon.* https://doi.org/10.1093/oxfordhb/9780198845003.013.17

Eryiğit, G., Bektaş, F., Ali, U., & Dereli, B. (2023). Gamification of complex morphology learning: the case of Turkish. *Computer Assisted Language Learning, 36*(8), 1421–1449. https://doi.org/10.1080/09588221.2021.1996396

Farhy, Y. (2020). Morphological generalization of Hebrew verb classes: An elicited production study in native and non-native speakers. *Mental Lexicon, 15*(2), 223–257. https://doi.org/10.1075/ml.19001.far

Fleischhauer, E., Bruns, G., & Grosche, M. (2021). Morphological decomposition supports word recognition in primary school children learning to read: evidence from masked priming of German-derived words. *Journal of Research in Reading, 44*(1), 90–109. https://doi.org/10.1111/1467-9817.12340

Fruchter, J., & Marantz, A. (2015). Decomposition, lookup, and recombination: MEG evidence for the full decomposition model of complex visual word recognition. *Brain and Language, 143*, 81–96. https://doi.org/10.1016/j.bandl.2015.03.001

Gagné, C. L., & Spalding, T. L. (2016). Written production of English compounds: effects of morphology and semantic transparency. *Morphology, 26*(2), 133–155. https://doi.org/10.1007/s11525-015-9265-0

Gamallo, P. (2023). Contextualized word senses: From attention to compositionality. *Linguistics Vanguard, 9*(1), 191–203. https://doi.org/10.1515/lingvan-2022-0125

Gatti, D., Marelli, M., Mazzoni, G., Vecchi, T., & Rinaldi, L. (2023). Hands-on false memories: a combined study with distributional semantics and mouse-tracking. *Psychological Research, 87*(4), 1129–1142. https://doi.org/10.1007/s00426-022-01710-x

Gatti, D., Rodio, F., Rinaldi, L., & Marelli, M. (2024). On humans' (explicit) intuitions about the meaning of novel words. *Cognition, 251*. https://doi.org/10.1016/j.cognition.2024.105882

Goddard, C., & Wierzbicka, A. (2016). "It's mine!". Re-thinking the conceptual semantics of 'possession' through NSM. *Language Sciences, 56*, 93–104.

https://doi.org/10.1016/j.langsci.2016.03.002

Günther, F., & Marelli, M. (2023). CAOSS and transcendence: Modeling role-dependent constituent meanings in compounds. *Morphology*, *33*(4), 409–432. https://doi.org/10.1007/s11525-021-09386-6

Günther, F., Petilli, M. A., & Marelli, M. (2020). Semantic transparency is not invisibility: A computational model of perceptually-grounded conceptual combination in word processing. *Journal of Memory and Language*, *112*. https://doi.org/10.1016/j.jml.2020.104104

Gwilliams, L. E., Monahan, P. J., & Samuel, A. G. (2015). Sensitivity to morphological composition in spoken word recognition: Evidence from grammatical and lexical identification tasks. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(6), 1663–1674. https://doi.org/10.1037/xlm0000130

Hamzai, J., & Xhaferi, B. (2024). Investigating the High Frequency and Morphological Productivity of Verb and Adjective Compounds in Legal English Corpus. *Palimpsest/ Палимпсест, 9*(17), 47–56. https://doi.org/10.46763/palim24917047h

Hedman, E., Hartelius, L., & Saldert, C. (2022). Word-finding difficulties in Parkinson's disease: Complex verbal fluency, executive functions and other influencing factors. *International Journal of Language and Communication Disorders*, *57*(3), 565–577. https://doi.org/10.1111/1460-6984.12707

Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, *52*(2), 291–314. https://doi.org/10.1023/A:1017919924342

Ke, S., Jin, R., & Koda, K. (2023). Adult L2 learners' morphological sensitivity in a morphosyllabic language. *Mental Lexicon*, *18*(3), 446–471. https://doi.org/10.1075/ml.22004.ke

Körtvélyessy, L., & Štekauer, P. (2020). Complex Words: Advances in Morphology. In *Complex Words: Advances in Morphology*. https://doi.org/10.1017/9781108780643

Levesque, K. C., Breadmore, H. L., & Deacon, S. H. (2021). How morphology impacts reading and spelling: advancing the role of morphology in models of literacy development. *Journal of Research in Reading*, *44*(1), 10–26. https://doi.org/10.1111/1467-9817.12313

Levesque, K. C., & Deacon, S. H. (2022). Clarifying links to literacy: How does morphological awareness support children's word reading development? *Applied Psycholinguistics*, *43*(4), 921–943. https://doi.org/10.1017/S0142716422000194

Levesque, K. C., Kieffer, M. J., & Deacon, S. H. (2019). Inferring Meaning from Meaningful Parts: The Contributions of Morphological Skills to the Development of Children's Reading Comprehension. *Reading Research Quarterly*, *54*(1), 63–80. https://doi.org/10.1002/rrq.219

Li, X., & Koda, K. (2022). Linguistic constraints on the cross-linguistic variations in L2 word recognition. *Reading and Writing*, *35*(6), 1401–1424. https://doi.org/10.1007/s11145-022-10266-6

Mansfield, J. (2021). The word as a unit of internal predictability. *Linguistics*, *59*(6), 1427–1472. https://doi.org/10.1515/ling-2020-0118

Marco, M. W.-D., & Fraser, A. (2020). Modeling word formation in English-German neural machine translation. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 4227–4232.

Margaza, P., & Gavarró, A. (2023). Word order with transitives and the interface hypothesis in L2 Greek and L2 Spanish. In *Advances in Second/Foreign Language Acquisition*. https://doi.org/10.1007/978-3-031- 38522-3_5

Monakhov, S. I., Turchanenko, V. V., & Cherdakov, D. N. (2023). Terminology use in school textbooks: corpus analysis | Школьный учебный текст в аспекте терминоупотребления: корпусный анализ. Research Result. *Theoretical and Applied Linguistics*, *9*(1), 27–49. https://doi.org/10.18413/2313- 8912-2023-9-1-0-3

Morgan, E. U., van der Meer, A., Vulchanova, M., Blasi, D. E., & Baggio, G. (2020). Meaning before grammar: A review of ERP experiments on the neurodevelopmental origins of semantic

processing. *Psychonomic Bulletin and Review, 27*(3), 441–464. https://doi.org/10.3758/s13423-019-01677-8

Nation, P. (2023). Word Parts and Language Learning. *JALT Journal, 45*(2), 235–245.

Nugraha, D. S. (2024). A Morphological Analysis of the Indonesian Suffixation: A Look at the Different Types of Affixes and Their Semantic Changes. *GEMA Online Journal of Language Studies, 24*(4), 109–132. https://doi.org/10.17576/gema-2024-2404-07

Nugraha, D. S. (2026). Corpus Linguistics and Morphology. In *International Encyclopedia of Language and Linguistics* (pp. 85–91). Elsevier. https://doi.org/10.1016/B978-0-323-95504-1.01540-4

Nugraha, D. S., & Purnawan, I. K. (2026). The evolution of compounding research in semantics: A decade of scientometric evidence (2014—2024). *GHANCARAN: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, *8*(1), 259-284. https://doi.org/10.19105/ghancaran.v8i1.23791

Öksüz, D., Brezina, V., Monaghan, P., & Rebuschat, P. (2024). Individual Word and Phrase Frequency Effects in Collocational Processing: Evidence from Typologically Different Languages, English and Turkish. *Journal of Experimental Psychology: Learning Memory and Cognition, 50*(8), 1287–1314. https://doi.org/10.1037/xlm0001324

Pichler, A. M. (2023). Diversity in (word) meaning: reducing the risk of bias in foreign language vocabulary teaching using prototype theory. *International Conference on Higher Education Advances*, 1081– 1088. https://doi.org/10.4995/HEAd23.2023.16148

Plag, I. (2014). Phonological and phonetic variability in complex words: An uncharted territory. *Italian Journal of Linguistics, 26*(2), 209–228.

Plag, I. (2020). Productivity and the Mental Lexicon. In *Language in Use: A Reader*.

Schmidtke, D., & Moro, A. L. (2021). Determinants of Word-Reading Development in English Learner University Students: A Longitudinal Eye Movement Study. *Reading Research Quarterly*, *56*(4), 819– 854. https://doi.org/10.1002/rrq.362

Semenza, C. (2021). Word formation in the brain: Data from aphasia and related disorders. In *Current Issues in Linguistic Theory* (Vol. 353). https://doi.org/10.1075/cilt.353.08sem

Smolka, E., & Eulitz, C. (2018). Psycholinguistic measures for German verb pairs: Semantic transparency, semantic relatedness, verb family size, and age of reading acquisition. *Behavior Research Methods*, *50*(4), 1540–1562. https://doi.org/10.3758/s13428-018-1052-5

Sooryamoorthy, R. (2020). *Scientometrics for the Humanities and Social Sciences*. Routledge. https://doi.org/10.4324/9781003110415

Spence, J. (2021). a Corpus Too Small: Uses of Text Data in a Hupa-English Bilingual Dictionary. *International Journal of Lexicography, 34*(4), 413–436. https://doi.org/10.1093/ijl/ecab006

Spencer, J., Kahraman, H., & Beyersmann, E. (2023). Positional Encoding of Morphemes in Visual Word Recognition. Journal of Experimental Psychology: *Learning Memory and Cognition, 50*(5), 712–739. https://doi.org/10.1037/xlm0001296

Stevens, P., & Plaut, D. C. (2022). From decomposition to distributed theories of morphological processing in reading. *Psychonomic Bulletin and Review, 29*(5), 1673–1702. https://doi.org/10.3758/s13423-022- 02086-0

Sudana, D., Hardini, T. I., & Zifana, M. (2024). Teaching Implication of Meaning Construction at Morphological Level of Affixation in a Complex Indonesian Verb Formation. *International Journal of Language Education, 8*(2). https://doi.org/10.26858/ijole.v8i2.64088

Tedla, Y., & Yamamoto, K. (2017). Analyzing word embeddings and improving POS tagger of Tigrinya. *Proceedings of the 2017 International Conference on Asian Language Processing*, IALP 2017, 2018-January, 115–118. https://doi.org/10.1109/IALP.2017.8300559

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

van Eck, N. J., & Waltman, L. (2023). *VOSviewer (1.6.20)*. Universiteit Leiden.

Vogel, I. (2018). The Morpho-Syntax-Phonology Interface in Complex Compounds. In *Studies in Natural Language and Linguistic Theory* (Vol. 94). https://doi.org/10.1007/978-3-319-90710-9_24

Waltman, L., & van Eck, N. J. (2019). Field Normalization of Scientometric Indicators (pp. 281–300). https://doi.org/10.1007/978-3-030-02511-3_11

Zhang, Q., Liang, L., Yao, P., Hu, S., & Chen, B. (2017). Parallel morpho-orthographic and morpho-semantic activation in processing second language morphologically complex words: Evidence from Chinese-English bilinguals. *International Journal of Bilingualism, 21*(3), 291–305. https://doi.org/10.1177/1367006915624249