



## Recontextualizing *Pesantren* in International Scholarship: A Corpus-Based Semantic Profile Analysis

Ahmad Suhaili\*  
Sekolah Tinggi Ilmu Tarbiyah Al-Khairiyah  
Cilegon, Indonesia  
ahmadsuhaili@stitalkhairiyah.ac.id

Frans Sayogie  
Universitas Islam Negeri Syarif Hidayatullah,  
Jakarta, Indonesia  
frans.sayogie@uinjkt.ac.id

Muhammad Farkhan  
Universitas Islam Negeri Syarif Hidayatullah,  
Jakarta, Indonesia  
farkhan@uinjkt.ac.id

### **Purpose**

*The study aimed to show evolving trends and paradigms linked to Pesantren in Islamic studies in Indonesia by analyzing the corpus of PTKI's worldwide publications.*

### **Method**

*This study uses semantic prosody and semantic preference analysis in the Corpus-Assisted Discourse Studies (CADS) framework to examine the representation of pesantren in international scholarship using a specialized corpus of 6,774 peer-reviewed articles from Scopus-indexed journals of Indonesian Islamic Higher Education Institutions (2015-2025). Semantic preference identified the semantic sets linked with the node word 'pesantren' throughout two temporal periods, whereas semantic prosody examined its evaluative context. UCREL Semantic Analysis System (USAS) artificial semantic taggers label collocations to determine semantic preference.*

### **Result/Findings**

*Looking at concordance lines reveals semantic prosody. The study reveals that Period I (2015–2020) "pesantren" is associated with conventional Islamic education, institutional identity, and charismatic leadership. It may be linked to corporate management, higher education, and religious moderation debate in Period II (2021–2025). It also shows that the node word has prosody in both periods, although cultural preservation to institutional transition varies.*

### **Conclusion**

*International scholars are increasingly seeing pesantren in a positive light, not just as conventional schools but also as flexible, adaptable organizations that deal with modern problems like modernization, digitalization, and social growth.*

### **Keywords**

*Semantic preference, semantic prosody, pesantren, corpus linguistics, CADS*

\*) Corresponding Author

## Abstrak

### Tujuan

Studi ini bertujuan untuk menunjukkan tren dan paradigma yang berkembang terkait dengan pesantren dalam studi Islam di Indonesia dengan menganalisis korpus publikasi PTKI di seluruh dunia.

### Metode

Studi ini menggunakan analisis prosodi semantik dan preferensi semantik dalam kerangka Corpus-Assisted Discourse Studies (CADS) untuk meneliti representasi pesantren dalam kajian internasional menggunakan korpus khusus yang terdiri dari 6.774 artikel yang ditelaah sejawat dari jurnal-jurnal terindeks Scopus di Perguruan Tinggi Islam Indonesia (2015-2025). Preferensi semantik mengidentifikasi himpunan semantik yang terkait dengan kata kunci 'pesantren' di sepanjang dua periode waktu, sedangkan prosodi semantik meneliti konteks evaluatifnya. Penanda semantik buatan UCREL Semantic Analysis System (USAS) memberi label kolokasi untuk menentukan preferensi semantik.

### Hasil/Temuan

Melihat baris konkordansi mengungkapkan prosodi semantik. Studi ini mengungkapkan bahwa pada Periode I (2015–2020), "pesantren" dikaitkan dengan pendidikan Islam konvensional, identitas institusional, dan kepemimpinan karismatik. Pada Periode II (2021–2025), pesantren mungkin dikaitkan dengan manajemen perusahaan, pendidikan tinggi, dan debat moderasi agama. Studi ini juga menunjukkan bahwa kata kunci memiliki prosodi di kedua periode, meskipun terdapat variasi antara pelestarian budaya dan transisi institusional.

### Kesimpulan

Para sarjana internasional semakin memandang pesantren secara positif, bukan hanya sebagai sekolah konvensional tetapi juga sebagai organisasi yang fleksibel dan mudah beradaptasi yang menangani masalah modern seperti modernisasi, digitalisasi, dan pertumbuhan sosial.

### Kata Kunci

Preferensi semantik, prosodi semantik, pesantren, linguistik korpus, CADS.

## المخلص

### الهدف

تهدف هذه الدراسة إلى توضيح الاتجاهات والنماذج المتطورة المتعلقة بالمدارس الإسلامية الداخلية (بيسانترين) في الدراسات الإسلامية في إندونيسيا، وذلك من خلال تحليل مجموعة من المنشورات الصادرة عن مؤسسات التعليم العالي الإسلامية (PTKI) حول العالم.

### المنهجية

تستخدم هذه الدراسة التحليل الدلالي للنبرة والتفضيل الدلالي ضمن إطار دراسات الخطاب بمساعدة المدونات اللغوية (CADS) لدراسة تمثيل المدارس الإسلامية الداخلية (بيسانترين) في الدراسات الدولية. وباستخدام مدونة لغوية متخصصة تضم 6774 مقالة محكمة من مجلات ماهرة في سكوبس حول الجامعات الإسلامية الإندونيسية (2015-2025)، يحدد التفضيل الدلالي المجموعات الدلالية المرتبطة بالكلمة المفتاحية "بيسانترين" عبر فترتين زمنييتين، بينما يدرس التحليل الدلالي للنبرة سياقها التقييمي. وتستخدم الوسوم الدلالية التي أنشأها نظام التحليل الدلالي التابع لجامعة كاليفورنيا في إرفاين لتصنيف المتلازمات اللفظية وتحديد التفضيل الدلالي.

### النتائج/المخرجات

تكشف سطور التوافق عن النبرة الدلالية. تُظهر هذه الدراسة أنه في الفترة الأولى (2015-2020)، ارتبط مصطلح "المعاهد الدينية" بالتعليم الإسلامي التقليدي، والهوية المؤسسية، والقيادة الكاريزمية. أما في الفترة الثانية (2021-2025)، فقد يرتبط مصطلح "المعاهد الدينية" بالإدارة المؤسسية، والتعليم العالي، والنقاش الدائر حول الاعتدال الديني. كما تُظهر الدراسة أن الكلمات المفتاحية تحافظ على نبرتها في كلتا الفترتين، على الرغم من وجود تباين بين الحفاظ على التراث الثقافي والتحول المؤسسي.

### الخلاصة

ينظر الباحثون الدوليون بشكل متزايد إلى المعاهد الدينية نظرة إيجابية، ليس فقط كمدارس تقليدية، بل أيضاً كمنظمات مرنة وقابلة للتكيف تعالج قضايا معاصرة مثل التحديث والتحول الرقمي والنمو الاجتماعي.

### الكلمات المفتاحية

التفضيل الدلالي؛ النبرة الدلالية؛ المعاهد الدينية؛ علم اللغة الحاسوبي؛

## INTRODUCTION

In modern critical discourse and sociolinguistic research, social institutions are perceived not just as structural entities but also as discursive constructs whose meanings are influenced by repetitive and systematic language practices (Fairclough, 1995; Wodak & Meyer, 2009). In this setting, academic language does not operate neutrally; rather, it actively contributes to the creation of knowledge, legitimacy, and epistemic authority (Hyland, 2004). Examining an institution's representation in academic literature is essential for comprehending its status and significance within global discourse.

*Pesantren* are distinctive Indonesian Islamic educational institutions that have traditionally functioned as hubs for the dissemination of classical Islamic knowledge and the establishment of community-oriented religious authority (Bruinessen, 2018; Dhofier, 1980). Over time, *pesantren* have been more misinterpreted as conventional institutions and have experienced substantial alterations in their functions, aligning with the modernization of education, democratization, and the globalization of Islam (Azra, 2015; Hefner, 2010). This shift has resulted in *pesantren* becoming linked to modern concerns such as religious moderation, civic education, social development, and the interplay between Islam and the state.

The rising quantity of foreign papers generated by Indonesian Islamic Religious Colleges (PTKI) shows that Islamic boarding schools are becoming more prominent in worldwide academic discussions. Many studies show Indonesian scholars are presenting Islamic boarding schools as exemplars of moderate Islam and as advocates of Nusantara Islam in global discussions (Azra et al., 2010; Bruinessen, 2018). This portrayal is frequently inconsistent and often contingent upon the writers' conceptual frameworks and discursive objectives.

Prior research on Islamic boarding schools predominantly uses historical, anthropological, and sociological methodologies, concentrating on institutional frameworks, kiai leadership, intellectual networks, and pedagogical methods (Dhofier, 1980; Lukens-Bull, 2013; Zarkasyi, 2020). Although these methodologies have made significant contributions, they are predominantly small-scale qualitative studies and have not comprehensively analyzed how *pesantren* are represented through linguistic patterns in international academic literature. Islamic boarding schools are frequently regarded as social constructs rather than conceptual structures.

Recent bibliometric and corpus-based studies show that there are big gaps in *pesantren* studies in worldwide scholarship. (Suresman et al., 2025) looked at 86 Scopus-indexed publications from 2007 to 2025. They found that there has been a shift from traditional teaching methods to hybrid digital models, but most contributors are still in Southeast Asia, which limits global perspectives. *Pesantren* have changed into multidimensional institutions that promote democratic ideals, civil society growth, and inclusive education (Isbah, 2020; Muslim et al., 2024). However, the way they are discussed in international publications may not show how complex they are. New approaches in corpus linguistics have shown promise in revealing semantic structures and ideological biases in religious discourse (Bhatt et al., 2025; Elewa, 2025), especially in how Islamic concepts are negotiated between Arabic-Islamic terminology and Western educational frameworks. Despite these methodological advancements, corpus-based semantic analysis has seldom been utilized to investigate the linguistic construction of *pesantren* in international academic literature, indicating a substantial methodological deficiency in comprehending the discursive profile of *pesantren* within global scholarship.

In this context, research grounded in corpus linguistics presents substantial methodological advantages. The Corpus-Assisted Discourse Studies (CADS) method facilitates the identification of collocation patterns, semantic preferences, and semantic prosody that influence the meaning of a notion extensively and temporally (Baker, 2010; Partington et al., 2013). Multiple studies indicate that corpus-based analysis can reveal implicit representations and concealed ideologies in academic discourse that are not readily discernible

through manual reading (Baker, 2006; Gabrielatos & Baker, 2008; Stubbs, 2005). Thus far, this methodology has infrequently been utilized to examine the portrayal of *pesantren* in worldwide PTKI publications.

The lack of such studies reveals a distinct research gap: the absence of a thorough, big-data-driven analysis of the semantic character of *pesantren* within worldwide academic discourse. Without corpus analysis, comprehending the portrayal of Islamic boarding schools may be selective and subjective, neglecting to identify persistent and systematic discursive patterns (McEnery & Hardie, 2011).

This study shows evolving trends and paradigms linked to *Pesantren* in Islamic studies in Indonesia by analyzing the corpus of PTKI's worldwide publications. This study employs the CADS technique to provide a diachronic analysis of the semantic profile of Islamic boarding schools based on big data, reinforcing the methodological foundations of current Islamic studies.

## METHOD

This study employed the Corpus-Assisted Discourse Studies (CADS) methodology formulated by Partington et al. (2013). This study uses a method that combines a quantitative analysis of linguistic data with a qualitative examination of discourse. It aims to understand the meanings and language patterns related to how *pesantren* are portrayed in international research. This approach aligns with recent developments in Corpus-Assisted Critical Discourse Analysis (CACDA), which has been successfully used to analyze ideological ideas in different areas of discourse (Suhaili et al., 2024, 2025).

The present study is informed by a theoretical framework. This framework posits the significance of *pesantren*, a distinctive Indonesian Islamic educational system, within the context of international scholarship on Indonesian Islam (Azra, 2015; Bruinessen, 2018; Dhofier, 1980). Consequently, *pesantren* is designated as the primary keyword for analysis. This method allows researchers to systematically track how *pesantren* discourse has changed over time and to identify shifts in meaning across different periods.

The present investigation employs a curated corpus comprising 6,774 scholarly articles, all of which were published within the timeframe of 2015 to 2025. These articles originated from 37 journals indexed in Scopus and were affiliated with Indonesian Islamic Higher Education Institutions (PTKI). The data was sourced from the official portal of the Ministry of Religious Affairs (Kemenag, 20204).

Journals were chosen through purposive sampling, contingent upon their pertinence to Islamic studies and their publication frequency. The dataset was divided into two temporal segments: Period 1 (2015–2020), which included 5,073,336 tokens, and Period 2 (2021–2025), which contained 11,842,387 tokens. The ensuing analysis was conducted in three stages: (1) collocation analysis, which utilized the log-Dice method, applying a minimum frequency threshold of 5 and a context window of four to five words surrounding the target keywords; (2) semantic categorization, which employed the USAS system, organized into 21 primary domains and 232 subcategories; and (3) semantic profiling, which integrated semantic preference and prosody analysis to clarify the conceptual and evaluative aspects of *pesantren* within the framework of international discourse.

## RESULT/FINDING AND DISCUSSION

This study used the corpus assisted analysis approach to identify the 50 most significant collocates of the node word *pesantren* over two temporal phases using the logDice association measure from a 6,774-article corpus with 6,549 frequencies. In Period I, three main semantic clusters organize the conversation around *pesantren*: (1) a institutional-educational cluster with words like *pondok*, *boarding*, *traditional*, *educational*, *madrasah*, and *curriculum*; (2) a social actor cluster with words like *santri*, *kyai*, *kiai*, *alumni*, and *leaders*; and (3) an institutional identity cluster with proper nouns like *Gontor*, *Sidogiri*, *T*. Despite a semantic shift toward policy, governance, and socio-politics, the educational cluster remains important in Period II. This shows a progressive shift in *pesantren* discursive

sive posture from traditionalist-educational to institutional and civic. These findings suggest that educational-traditionalist discourse dominates both periods' discussions of *pesantren*. Stubbs's (1996) semantic prosody theory suggests that the words surrounding a central word affect its evaluative and associative meanings. These findings show how Indonesian public discourse has influenced the way *pesantren* is discussed.

These findings are particularly important for supporting and improving research on *pesantren* discourse. The strong connections in Period I to *kyai*, *salaf*, and tradition reinforce Dhofier's (1982) description of *pesantren* as an institution led by religious scholars and focused on traditional Islamic education. Van Bruinessen (1995) observed that Indonesian Islamic education negotiates between traditionalism and modernity, and introducing terminology like modern, development, and institutions demonstrates this discursive struggle. By Period II, discourse has partially resolved this tension, as *pesantren* increasingly aligns with national development agendas and regulatory frameworks, supporting Azra's (2003) claim that Indonesian Islamic educational institutions have indigenized modernity. In both periods, "female" and "male" are statistically significant collocates, adding a gender dimension to the discourse that past corpus-based studies of Islamic educational terminology did not. This suggests that the corpus data shows a discourse shift toward gender-inclusive narratives in *pesantren* institutions. This multi-layered diachronic pattern indicates that *pesantren* is a dynamic institution where tradition, modernity, governance, and gender are reinvented constantly.

### Collocation of *Pesantren* in Period I

The collocation patterns in Period 1 as available in Table 1 show that *pesantren* is primarily characterized as a traditional Islamic educational institution within the realm of international knowledge. The strongest collocates are words that show institutional identity, with "pondok" (logDice 10.05), "boarding" (logDice 9.52), and "Pondok" (logDice 9.45) being the most common. This makes *pesantren* fit firmly within the "Islamic boarding school" semantic frame.

This linguistic pattern shows that scholars always used the pondok *pesantren* construction to help people from other countries who didn't know much about Indonesian Islamic education understand the idea. The strong links to "traditional" (logDice 9.40), "educational" (logDice 9.08), and "education" (logDice 9.07) show that the main way people thought about *pesantren* was as an educational institution, with its traditional nature being what set it apart from modern schools.

The presence of certain *pesantren* names as strong collocates indicates the prevalence of the case study approach in contemporary academia. "Sukamanah" (logDice 8.70), "Sidogiri" (logDice 8.24), "Gontor" (logDice 8.05), and "Tebuireng" (logDice 7.88) are important collocates that show that these institutions were often used as examples in international scholarship. This trend indicates that *pesantren* rhetoric was created through comprehensive institutional profiles rather than extensive systemic investigation. The strong collocation of "madrasah" (logDice 8.77) shows that scholars are still interested in comparing or contrasting *pesantren* and madrasah systems. This is because there are still discussions regarding how to group Islamic schools in Indonesia. The collocation with "Java" (logDice 8.33) shows that most of the research on *pesantren* is done in Java, where Javanese institutions are the main places where international experts do their work. Words like "leaders" (logDice 8.54), "leader" (logDice 7.90), "kyai" (logDice 7.84), and "kiai" (logDice 7.78) show that leadership and human agency are important themes.

This pattern demonstrates the academic focus on charismatic authority and kiai leadership as fundamental characteristics of *pesantren* culture, aligning with traditional anthropological perspectives on Indonesian Islam that emphasized the examination of religious figures. The terms "santri" (logDice 8.03), "students" (logDice 8.62), "alumni" (logDice 8.01), and "teachers" (logDice 7.64) show that *pesantren* is a social arena where different people with different responsibilities interact. The words "female" (logDice 7.89) and "male" (logDice 7.32) show that gender is an important analytical category. This suggests

that researchers are paying more attention to gender dynamics in these traditionally male-dominated institutions, which may be because of the rise of female *pesantren* and gender-inclusive educational reforms. The conflict between tradition and modernity is seen in the contrasting collocates that place *pesantren* inside developmental narratives. "Traditional" (logDice 9.40), "tradition" (logDice 8.51), and "Salaf" (logDice 7.87) all stress the connection to classical Islamic education, while "modern" (logDice 7.67), "developed" (logDice 7.60), and "development" (logDice 7.61) place *pesantren* in the context of modernization discourses.

**Table 1.** The Collocate of *Pesantren* in Period I

No	Collocate	Freq	Coll. freq.	logDice	No	Collocate	Freq	Coll. freq.	logDice
1	pondok	98	141	10,0458	26	Tebuireng	21	27	7,87859
2	boarding	80	669	9,52184	27	Salaf	21	47	7,86876
3	Pondok	64	103	9,44925	28	kyai	22	258	7,83601
4	traditional	88	1364	9,40246	29	Lirboyo	20	27	7,8082
5	educational	67	1155	9,08167	30	kiai	21	240	7,77715
6	education	129	4951	9,07312	31	Nurul	19	74	7,71119
7	madrasah	43	335	8,76805	32	modern	29	1783	7,66569
8	Sukamanah	38	96	8,70055	33	institutions	29	1800	7,66046
9	Pesantren	45	708	8,67607	34	teachers	24	1053	7,63733
10	system	75	3313	8,62844	35	some	51	5544	7,62923
11	students	71	3036	8,6152	36	institution	25	1242	7,62879
12	leaders	52	1665	8,54499	37	development	37	3312	7,60931
13	tradition	56	2153	8,50529	38	developed	24	1170	7,59521
14	pesantren	49	2001	8,35673	39	studied	21	697	7,58094
15	Java	47	1883	8,33179	40	salaf	17	61	7,55706
16	schools	49	2108	8,32556	41	have	91	13231	7,53051
17	school	51	2605	8,24668	42	Jadid	16	37	7,48135
18	Sidogiri	27	30	8,23968	43	called	27	2123	7,46142
19	Gontor	24	73	8,04872	44	has	92	14577	7,43064
20	santri	27	484	8,03165	45	are	149	25695	7,41584
21	alumni	24	149	8,01226	46	taught	18	574	7,40879
22	curriculum	28	657	8,01213	47	Zainal	16	232	7,38851
23	established	31	1265	7,93114	48	learning	22	1537	7,34504
24	leader	29	1068	7,90488	49	male	17	592	7,31886
25	female	27	836	7,88877	50	particularly	19	1044	7,30358

This binary formulation is similar to bigger arguments in Indonesian Islamic studies regarding whether *pesantren* is a form of resistance to modernity or adaptive traditionalism. The fact that "curriculum" (logDice 8.01) is a major collocate shows that academics are interested in educational content and reform. This suggests that curriculum became a place where the conflict between tradition and modernity was most obvious and studied most often. Words like "system" (logDice 8.63), "schools" (logDice 8.33), "school" (logDice 8.25), and "institutions" (logDice 7.66) show institutional structures and systemic aspects. This pattern shows that while most research focused on individual *pesantren*, some scholars also tried to look at the whole *pesantren* system as an educational sector. The collocation with "established" (logDice 7.93) implies a focus on institutional histories and foundation narratives, whereas "learning" (logDice 7.35) and "taught" (logDice 7.41) denote an emphasis on pedagogical procedures. The mention of "Jadid" (logDice 7.48) in conjunction with "salaf" indicates scholarly focus on the reformist-traditionalist typology that has historically organized *pesantren* classification, il-

lustrating that international study has embraced Indonesian academic frameworks for categorizing *pesantren* varieties.

**Collocation of Pesantren in Period II**

Period 2 illustrates considerable topic development and variation in the language construction of *pesantren* throughout international scholarship (see Table 2). While basic institutional indicators endure—"Pondok" (logDice 9.74), "education" (logDice 9.23), "boarding" (logDice 8.95), and "traditional" (logDice 8.75) remain among the greatest collocates—novel associational patterns indicate changed scholarly ideas. The rise of "university" (logDice 8.55) and "universities" (logDice 7.93) as significant collocates signifies a pivotal transformation in the positioning of *pesantren* within Indonesia’s educational framework.

**Table 2** The Collocate of *Pesantren* in Period II

No	Collocate	Freq	Coll. freq.	logDice	No	schools	70	4375	7,71407
1	Pondok	179	318	9,73835	26	schools	70	4375	7,71407
2	education	325	11162	9,23194	27	Al-Furqon	42	46	7,70516
3	boarding	134	2319	8,9517	28	Indonesia	139	17140	7,58701
4	Assalaam	92	96	8,82551	29	Aceh	70	5531	7,56895
5	traditional	150	4866	8,75018	30	institution	51	2381	7,54798
6	Pesantren	96	1467	8,6163	31	within	77	7563	7,48215
7	university	87	1064	8,54869	32	leaders	53	3268	7,46685
8	Madura	75	288	8,48967	33	Salafi	42	1572	7,40488
9	Darul	72	451	8,39679	34	established	47	2571	7,39977
10	curriculum	84	1886	8,35015	35	environment	54	3983	7,3924
11	santri	75	1277	8,29476	36	community	99	13341	7,3496
12	Annuqayah	62	75	8,26071	37	madrasah	34	564	7,29122
13	Ulum	60	207	8,18494	38	culture	55	5579	7,21531
14	management	87	3562	8,13931	39	taught	34	967	7,21171
15	pesantren	95	4529	8,13445	40	alumni	31	304	7,21168
16	institutions	97	4968	8,10844	41	activities	66	8183	7,19773
17	system	117	7734	8,06835	42	role	66	8230	7,19313
18	BBI	50	55	7,95473	43	outside	37	1910	7,16318
19	three	96	6147	7,95288	44	values	73	10544	7,12871
20	modern	77	3697	7,94401	45	pondok	29	271	7,12242
21	universities	56	989	7,92739	46	teachers	39	2921	7,07625
22	Java	67	2707	7,88992	47	Islamic	210	45772	7,03914
23	educational	69	3548	7,80689	48	development	64	9420	7,03701
24	students	121	11335	7,79248	49	Kyai	28	473	7,02969
25	Nurul	44	228	7,733	50	developed	37	2843	7,01224

This trend shows that more and more research is looking at how *pesantren* are becoming part of or changing into higher education institutions. This is because several large *pesantren* have already set up their own universities. This shift toward higher education indicates that *pesantren* discourse has evolved beyond secondary-level Islamic boarding schools to include intricate, multi-tiered educational institutions.

The emergence of "management" (logDice 8.14) as a new strong collocate marks a shift in the way we look at *pesantren*, from cultural-anthropological to organizational-managerial. This administrative change shows that scholars are interested in governance structures, institutional sustainability, strategic planning, and professional management practices in the context of *pesantren*. The stronger connections to "institutions" (logDice 8.11) and "system" (logDice 8.07) show that *pesantren* is being studied more and more as part of Indonesia’s institutional infrastructure instead of as separate traditional communi-

ties. The fact that it is next to “curriculum” (logDice 8.35), which is stronger than in Period 1, suggests that there is still interest in educational content, but it is stronger than before. This could be because of discussions about how to include national curriculum standards, religious moderation content, or vocational training in traditional *pesantren* education.

Geographic diversification is one of the most obvious shifts in the collocation patterns of Period 2. New regional collocates like “Madura” (logDice 8.49), “Aceh” (logDice 7.57), and “Indonesia” (logDice 7.59) show that research has moved beyond Java-focused case studies to cover *pesantren* customs from all across the archipelago. The rise of Madura as a prominent collocate is especially interesting because Madurese *pesantren* have their own traditions that are generally stricter Salafi and have higher kiai authority than Javanese *pesantren*. “Assalaam” (logDice 8.83), “Annuqayah” (logDice 8.26), “Darul Ulum” (logDice 8.39-8.18), “Al-Furqon” (logDice 7.71), and “BBI” (logDice 7.95) are some of the new names that have become popular for *pesantren*. This geographic and institutional heterogeneity indicates that foreign scholarship has evolved from dependence on a limited number of canonical cases (Gontor, Tebuireng) to a more extensive examination of Indonesia’s diverse *pesantren* landscape.

The biggest change in themes in Period 2 is the addition of social, cultural, and community aspects. New semantic areas that weren’t in Period 1’s top collocates are “community” (logDice 7.35), “culture” (logDice 7.22), “values” (logDice 7.13), “role” (logDice 7.19), “activities” (logDice 7.20), and “environment” (logDice 7.39). This tendency shows that *pesantren* is being seen more and more as a multifunctional social institution that goes beyond formal education to include cultural output, moral formation, community organization, and environmental engagement. The emergence of “community” as a prominent collocate indicates scholarly focus on the integration of *pesantren* within local societies and their functions as community anchors, social service providers, and civil society participants. This sociological shift corresponds with international trends in religion studies that analyze religious institutions as venues for the cultivation of social capital and civic participation.

Although leadership-related collocates endure in Period 2, their relative attenuation indicates nuanced yet significant discursive transformations. “Leaders” (logDice 7.47) and “Kiai” (logDice 7.03) are still there, but their association strengths are lower than in Period 1. This suggests that charismatic authority is less important to how scholars frame *pesantren*. This reduced focus on individual authority figures may indicate various trends: heightened institutional complexity that allocates authority throughout organizational frameworks, generational shifts as founding kiai die, or academic inclinations favoring systemic rather than personality-driven analysis. The fact that “Salafi” (logDice 7.40) is replacing the general term “salaf” shows that people are paying more attention to specific ideological groups within *pesantren*. This may be because of arguments over transnational Islamic influences and concerns about radicalization. The collocation with “Islamic” (logDice 7.04)—which is not as strong as other collocates—suggests that people are starting to take the Islamic nature of *pesantren* for granted instead of making it clear. This shows that *pesantren* is becoming more common as a type of institution in international Islamic studies.

### Semantic Category for *Pesantren*

Semantic preference refers to how a word or phrase tends to be used with other words that belong to a specific area of meaning or semantic group. Semantic preference delineates the correlation between the collocates of *Pesantren* and other lexemes that possess analogous semantic attributes. Lindquist & Levin (2018) stated that semantic preference is the connection between a word and a group of words that belong to the same semantic category and are therefore meaningfully related. Table 3 displays the results of the categorization of the meaning field of the keyword *pesantren* from periods I and II based on the results of the USAS semantic annotation.

**Table 3** The collocates of the node word *Pesantren* Period I are categorized into their semantic sets

Category		Collocate
P1	Education in general	educational, education, students, schools, school, curriculum, alumni, teachers, taught, learning
Z99	Unmatched (istilah ocal)	pondok, Pondok, Sukamanah, Pesantren, pesantren, santri, Gontor, Tebuireng, Salaf, kyai, kiai, Nurul, Lirboyo, Sidogiri, salaf, Jadid, Zainal
S7.1+	Power, organizing	leaders, leader
S1.1.1	Social Actions, States And Processes	traditional, tradition
A6.2+	Comparing: Usual	traditional, tradition
T3++	Time: Old	traditional
M1	Moving, coming and going	boarding
H4	Residence	boarding
X4.2	Mental object: Means, method	system
Z2	Geographical names	Java
T2+	Time: Beginning	established
X6+	Decided	established
S2.1f	People: Female	female
N5	Quantities	some, have
A4.1	Generally kinds, groups, examples	institution
Other	Various minor categories	developed, studied, called, has, are, particularly, male

The semantic categories of notable collocate associated with "*pesantren*" in Period I reveal the predominance of category P1 (Education in general), which is represented in 8 of the top 50 collocates, including "educational," "education," "students," "schools," "school," "curriculum," and "teachers." This category dominance signifies that discussions regarding *pesantren* in international media from 2015 to 2020 continue to emphasize their primary role as educational institutions. *Pesantren* are characterized as educational institutions, highlighting their teaching function and formal academic framework.

The formation of category S1.1.1 (Social Actions, States, and Processes) in the terms "traditional" and "tradition" underscores the historical significance and cultural continuity of *pesantren* as indigenous Islamic educational institutions. *Pesantren* serve as an institution that embody the continuity of Islamic scholarly traditions and social practices historically entrenched in Indonesian society.

The Z99 (Unmatched) category is prevalent in numerous colloquial phrases, including *pondok*, *santri*, *kyai*, *kiai*, *salaf*, and particular *pesantren* names such as *Gontor*, *Tebuireng*, *Lirboyo*, and *Sidogiri*. This pattern signifies that *pesantren* terminology is excluded from the USAS semantic system, which relies on English corpus annotations. This illustrates the uniqueness of *pesantren* culture and its connection to local concepts that have not been completely assimilated into global academic discourse via translation or established semantic frameworks.

The A6.2+ (Comparative: Conventional) category and the T3++ (Temporal: Old, new, and young; age) category affirm the characterization of *pesantren* as enduring educational institutions dedicated to the preservation of ancient Islamic traditions. The boarding

categories, M1 (Moving, Coming, and Going) and H4 (Residence), highlight the residential aspect of Islamic boarding schools as total institutions, where education is perceived not merely as knowledge transfer but also as the cultivation of habitus through communal living in an immersive learning environment.

The creation of the Java category, Z2 (Geographical Names), signifies the spatial concentration of academic discourse in Islamic boarding schools on the island of Java. This trend shows a specific hegemony in knowledge creation that may mask the diversity of Islamic boarding school traditions in various regions of Indonesia. Simultaneously, the S7.1+ category (Power, Organizing), featured in the leaders and leadership categories, emphasizes the leadership aspect of Islamic boarding schools, specifically the *kiai*'s role as a charismatic authority within the traditional institutional framework of these schools.

Meanwhile, the results of the semantic annotation of the keywords of Islamic boarding schools period II along with the accompanying significant colloquialisms are shown Table 4 below.

**Table 4.** The collocates of the node word *Pesantren* Period II are categorised into their semantic sets

Category		Collocate
S7.1+	Power, organizing	management, leaders, leadership, institutions, institution, established
P1	Education in general	education, university, curriculum, students, educational, schools, taught, teachers, learning
Z99	Unmatched (istilah lokal)	Pondok, Assalaam, Pesantren, Madura, Darul, santri, Annuqayah, Ulum, pesantren, BBI, Nurul, Al-Furqon, madrasah, Salafi, pondok, Kyai
A2.1+	Change	development, developed
Z2	Geographical names	Indonesia, Aceh, Java, Madura
M1	Moving, coming and going	boarding
T3-	Time: New	modern, traditional
A6.2+	Comparing: Usual	traditional
W5	Green issues	environment
S5+c	Belonging to a group	community
A1.1.1	General actions / making	activities, role
S9	Religion and the supernatural	Islamic
Y2	Information technology	digital (implied in context)
Others	Various categories	within, outside, values, alumni, culture

The transition to Period II exhibited a significant semantic transformation. The S7.1+ (Power, Organizing) category witnessed a notable rise, manifesting in the vernacular phrases management, leaders, leadership, institutions, institution, and established. This discovery signifies a transition in analytical emphasis from cultural portrayals of Islamic boarding schools to investigations of organizational dynamics, management practices, and institutional fortification of these schools.

The A2.1+ (Change) category, characterized by terms like development and developed, signifies the enhancement of dialogue around the reform and modernization of Islamic boarding schools. The emergence of the colloquial term "university" in the P1/H1c (Education/Institutions) category signifies the elevation of Islamic boarding schools from traditional boarding establishments to recognized higher education institutions. The enduring use of the term curriculum reflects an ongoing emphasis on educational elements, but in a transformed context: a transition from traditional curriculum descriptions to ex-

aminations of curriculum innovation and the amalgamation of Islamic sciences with contemporary fields.

The most significant alteration is the growth of colloquial terminology that reflects the involvement of Islamic boarding schools with modern challenges. The digital colloquium in categories Y2 (Information Technology) and N1 (Numbers) illustrates the digitization of Islamic boarding school curriculum. The environmental colloquium categorized under W5 (Green Issues), W3 (Geographical Terms), and E1 (Emotional Actions, States, and Processes) signifies an increasing focus on environmental education and ecological consciousness inside Islamic boarding schools.

The elevated frequency of the within colloquium in category Z5 (Grammatical Bin) signifies a growing complexity in analyzing the internal and exterior relationships of Islamic boarding schools with the wider social system. The exceptionally high frequency of the Indonesian colloquium (139), together with the Aceh colloquium, illustrates the geographic diversification of research on Islamic boarding schools. Academic discourse has shifted from a focus on Islamic boarding schools in Java to include other locations throughout Indonesia, illustrating the democratization of knowledge production and acknowledgment of regional differences in Islamic boarding school traditions.

This movement signifies the reclassification of Islamic boarding schools within academic discourse, transitioning from subjects of anthropological-historical examination, perceived as vestiges of traditional society, to focal points of inquiry in educational organization and management studies. Pesantren are depicted as dynamic entities capable of innovation and adaptation to societal changes. This transition corresponds with the empirical evolution of Islamic boarding schools in Indonesia, which are progressively embracing modernity through incorporating digital technology, integrating contemporary sciences into the curriculum, the promotion of entrepreneurship, and involvement in the national development agenda. Modern academic discourse views Islamic boarding schools not as independent entities, but as integral components of the national education system and civil society. These institutions promote social cohesion, Islamic moderation, and sustainable development.

### Semantic Prosody of *Pesantren*

#### *Period I*

Semantic prosody in corpus linguistics denotes the habitual inclination of a lexical item to occur in situations imbued with specific evaluative connotations. Analysis of the concordance lines from Period I indicates that the keyword "*pesantren*" possesses a predominantly positive semantic prosody, marked by educational innovation, social empowerment, and cultural preservation. This semantic prosody is recognized in linguistic situations, as seen in the subsequent concordance lines:

1. *Pesantren contributes significantly to the religious education and character formation of Indonesian Muslim youth.*
2. *"The transformation of pesantren from traditional boarding schools into modern educational institutions demonstrates a high degree of institutional adaptability."*
3. *"Islamic education in pesantren offers alternative pedagogical approaches that integrate religious teachings with modern knowledge."*
4. *"The pesantren-based education system provides holistic development by combining spiritual, intellectual, and social competencies."*
5. *"The role of pesantren in preserving Islamic scholarly traditions while adapting to contemporary challenges is remarkable."*

Phrases like "contributes significantly," "demonstrates adaptability," "offers alternative approaches," and "remarkable" create a prosody that positions *pesantren* as useful, innovative, and strategically essential institutions within the context of modern Islamic education. This architecture reflects an intellectual recognition of the multifaceted nature of *pe-*

*santhren's* role, surpassing the cliché of a static, conventional institution.

This affirmative semantic prosody is bolstered by a collectional context that regularly links *pesantren* with normative values esteemed in modern educational and developmental discourse. The frequent appearance of terms like "empowerment," "leadership development," "community resilience," "cultural preservation," and "interfaith dialogue" in proximity to *pesantren* constitutes what Stubbs refers to as "attitudinal meaning" or "connotative coloring." This prosody portrays *pesantren* as innately progressive and socially responsive institutions. This is clearly demonstrated in the subsequent lines of the concordance:

6. *"Women's leadership development in pesantren institutions challenges traditional gender boundaries."*
7. *"Pesantren graduates demonstrate a strong foundation in religious knowledge as well as leadership capabilities across various sectors."*
8. *"The integration of classical Islamic scholarship with modern pedagogy in pesantren produces distinctive educational outcomes."*
9. *"Pesantren-based community development programs effectively address poverty and social inequality."*
10. *"The pesantren system fosters religious tolerance and cultural understanding among diverse student populations."*

## Period II

The concordance lines analysis indicates that the keyword "*pesantren*" in period II also has a predominantly positive semantic prosody, characterized by adaptability, relevance, and resilience. The key phrases "never disappeared," "still relevant," and the concessive construction "even though it is still traditional" form a prosody of sustainability and ongoing relevance that constructs *pesantren* not as an anachronistic institution but as an entity that can survive and adapt in the face of modernity.

11. *"Pesantren has existed long before the inception of formal schools. Today, hundreds of years into its existence, pesantren have never disappeared from Indonesia's educational realm"*
12. *"Therefore, even though it is still traditional, pesantrens are still considered relevant to Indonesia's modern education system. Pesantrens nowadays continue to receive"*
13. *"A study of Islamic education, pesantren occupies its role as a traditional education system that continuously develops from time to time"*
14. *"the spirit of education in many pesantrens is understood to reflect nothing less than the will of God"*
15. *"Pesantren consistently attempts to present its identity as a center of tafaquh fi al-din (deep understanding of religion)"*

This collectional context consistently associates *pesantren* with normative values that are highly regarded in contemporary educational discourse, thereby reinforcing this positive semantic prosody. Stubbs refers to the attitudinal meaning or connotative coloring that is formed by the repeated occurrence of phrases such as "religious moderation values," "tolerance values," "avoiding violent action," "moderate character building," and "cultural diversity" within the collectional radius of *pesantren*. The discursive significance of this moderation prosody lies in its function as a discursive strategy to establish *pesantren* in binary opposition to extremism and radicalism, rather than merely as a factual description. This is explicitly stated in the subsequent lines of the concordance:

16. *"Therefore, institutionally, the three pesantrens apply religious moderation values by avoiding violent action and the idea of religious extremism. Religious"*
17. *"to create an active tolerance in responding to the multi-identity situation. Therefore, institutionally, the three pesantrens apply religious moderation values"*
18. *"The three universities' pesantren system is not merely a place of religious education and training for students; the pesantren is designed and developed by a part of a cultured society"*
19. *"The togetherness of multicultural students within the scope of exclusive pesantren trains them to adapt several tolerance values"*
20. *"This study helps researchers and policy makers understand the dynamics of pesantren learning and education, particularly in relation to moderate character building"*

The semantic prosody of *pesantren* is not consistently positive; instead, it reveals complexity and discursive contention. In specific entries of the concordance, potentially adverse prosody is evident, exemplified by the phrase: "a prominent *pesantren* in Central Java has a close link with the Southeast Asian terrorist network." The presence of this negative prosody, albeit a minority within the corpus, signifies that academic discourse on *pesantren* is not uniform; it instead mirrors the empirical reality that *pesantren* institutions can express diverse ideological orientations, including affiliations with extremist networks. Additional settings that illustrate discursive tensions encompass:

21. *"This situation requires pesantren to responsively adapt to these challenges and to keep building moderate character through their educational process"*
22. *"For instance, the International Crisis Group (ICG) notes that a prominent pesantren in Central Java has a close link with the South East Asia terrorist network"*
23. *"some people often underestimate traditional pesantren and madrasah due to low market demand for their graduates"*
24. *"what we find in pesantren is far from this ideal standard. From the literature review, pesantren that try eco pesantren and childfriendly building principles have not met the ecologically sustainable development principles"*
25. *"In considering the values inculcated in pesantren related to commodification, it is worth looking closer to the concept of pious neoliberalism"*

The conflict between the prevailing positive-moderate prosody and the minority negative-extremist and critical-commodification prosody illustrates a wider discursive struggle about the portrayal of Islam and Islamic educational institutions (*Pesantren*) in global discourse.

Another aspect of the *pesantren's* semantic prosody is its focus on modernity and rejuvenation. The examination of the linguistic context reveals that *pesantren* frequently emerge in colloquial settings indicative of transformation and adaptation: "modernization of *pesantren* education," "renewal of Islamic teachings and methodologies," "*pesantren* have evolved and reconfigured," and "adapted through the integration of contemporary educational approaches." This modernizing prosody, however, possesses a unique attribute in that it is not framed as secularization or westernization, but rather as a process of selective appropriation that remains anchored in Islamic orthodoxy. The subsequent lines from the concordance exemplify this prosody:

This is evident in concordances such as "The modernization of education in Islamic boarding schools is designed to preserve the *pesantren's* heritage as Islamic orthodoxy" and the motto "al-muhafazah 'ala al-qadim al-salih wa al-akhdu bi al-jadid al-ashlah" (maintaining good traditions while adopting new, better ones). This semantic prosody builds modernity as continuity-in-transformation rather than a departure with tradition, a concept that is consistent with Islamic modernism theory, which encourages *tajdid* (renewal) while retaining *usul*.

25. "Given the globalization environment, Islamic orthodoxy in the pesantren needs turats modification towards modern character education. The modernization of pesantren education is"
26. "Modernity necessitates future-oriented work patterns; however, the modernization of education in pesantrens is designed to preserve the pesantren turats as Islamic orthodoxy"
27. "wal akhdu bi al-jadid al-ashlah (maintaining good tradition and employing the new-better one). Within 24-hour pesantren learning activities, this jargon is disseminated to the students"
28. "We have to renew Islamic teachings and methods, including in pesantren. Thus, it will not be outdated"
29. "post-conflict and post-disaster pesantren in Aceh have evolved and configured from monocultural and traditional pesantren into some patterns"

Commodification and the Muslim middle class bring forth more sophisticated semantic prosody. The use of terms like "middle-class *pesantren*," "affordable only for the upper-middle-class," "market demand," and "pious neoliberalism" in conjunction with *pesantren* generates an equivocal prosody. On the one hand, the phrases "commodification" and "low market demand for their graduates" have a subtle negative prosody, implying a condemnation of market logic's entrance into Islamic educational institutions. However, this prosody is descriptive-analytical rather than overtly critical, reflecting a sociological approach that seeks to understand the transformation of Islamic boarding schools within the context of Indonesian Muslim society's changing socioeconomic structure without making excessively normative judgments. The marketization of Islamic education is a complicated issue that cannot be reduced to a simple story of religious values being corrupted or adapting to modernity.

## CONCLUSION

It may be inferred that while *pesantren* is predominantly portrayed favorably in world-wide academia, the semantic profile indicates a complex and dynamic discursive construction influenced by temporal, geographical, and ideological elements. The change from anthropological-cultural framing to organizational-managerial framing shows a change in how Indonesian Islamic education fits into global knowledge economies. Additionally, it is important to acknowledge that the findings of this study are confined to 6,774 papers from Scopus-indexed journals of Indonesian Islamic Higher Education Institutions published between 2015 and 2025. This study's corpus represents a particular subset of international scholarship on *pesantren*. The researcher recommends that subsequent scholars explore *pesantren* representation within various linguistic communities (Indonesian, English, Arabic academic discourses), investigate specific types of *pesantren* (salafi vs. modern, urban vs. rural, male vs. female institutions), and perform comparative analyses with other Islamic educational institutions (madrasah, zawiyah, Islamic universities) to enhance the depth, cross-cultural perspective, and reliability of the findings in assessing the semantic preference and semantic prosody of *pesantren* discourse in global scholarship.

## REFERENCE

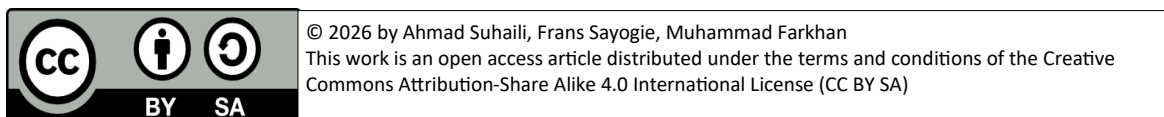
- Azra, A. (2015). The Significance of Southeast Asia (the Jawah World) for Global Islamic Studies: Historical and Comparative Perspectives.
- Azra, A., Afrianty, D., & Hefner, R. W. (2010). Chapter 8. Pesantren and Madrasa: Muslim Schools and National Ideals in Indonesia. In R. W. Hefner & M. Q. Zaman (Eds.), *Schooling Islam* (pp. 172–198). Princeton University Press. <https://doi.org/10.1515/9781400837458.172>
- Baker, P. (2006). *Using Corpora in Discourse Analysis*. Continuum. <https://>

- doi.org/10.5040/9781350933996
- Baker, P. (2010). *Sociolinguistics and corpus linguistics*. Edinburgh university press.
- Bhatt, I., Barnawi, O. Z., & Ahmad, R. (2025). Exploring a Sociolinguistics of Islam. *Applied Linguistics*, amaf043. <https://doi.org/10.1093/applin/amaf043>
- Bruinessen, V. M. (2018). Indonesian Muslims in a Globalising World: Westernisation, Arabisation and Indigenising Responses. <https://rsis.edu.sg/rsis-publication/rsis/wp311-indonesian-muslims-in-a-globalising-world-westernisation-arabisation-and-indigenising-responses/>
- Dhofier, Z. (1980). *The pesantren tradition: A study of the role of the kyai in the maintenance of the traditional ideology of Islam in Java* [The Australian National University]. <https://doi.org/10.25911/5D74E363A6DED>
- Elewa, A. (2025). A corpus-based analysis of gendered language in spoken religious discourse. *Applied Corpus Linguistics*, 5(3), 100137. <https://doi.org/10.1016/j.acorp.2025.100137>
- Gabrielatos, C., & Baker, P. (2008). *Fleeing, Sneaking, Flooding: A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005*. *Journal of English Linguistics*, 36(1), 5–38. <https://doi.org/10.1177/0075424207311247>
- Hefner, R. W. (2010). *Expressing Islam: Religious Life and Politics in Indonesia* \* Edited by GREG FEALY and SALLY WHITE. *Journal of Islamic Studies*, 21(3), 479–482. <https://doi.org/10.1093/jis/etq058>
- Hyland, K. (2004). *Genre and Second Language Writing*. University of Michigan Press/ELT. <https://doi.org/10.3998/mpub.23927>
- Isbah, M. F. (2020). Pesantren in the Changing Indonesian Context: History and Current Developments. *QIJIS (Qudus International Journal of Islamic Studies)*, 8(1), 65. <https://doi.org/10.21043/qijis.v8i1.5629>
- Lindquist, H., & Levin, M. (2018). *Corpus linguistics and the description of English (Second edition)*. Edinburgh University Press.
- Lukens-Bull, R. A. (2013). *Islamic Higher Education in Indonesia*. Palgrave Macmillan US. <https://doi.org/10.1057/9781137313416>
- McEnery, T., & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice (1st ed.)*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511981395>
- Muslim, A. B., Arroisi, J., Jan, J. M., & Ramazani, R. (2024). Indonesian pesantren-affiliated higher education: Universal Islamic values to develop locally engaged but globally oriented multilingual graduates. *Research in Post-Compulsory Education*, 29(1), 138–157. <https://doi.org/10.1080/13596748.2023.2285633>
- Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and Meanings in Discourse: Theory and practice in corpus-assisted discourse studies (CADS) (Vol. 55)*. John Benjamins Publishing Company. <https://doi.org/10.1075/scl.55>
- Stubbs, M. (2005). *Words and phrases: Corpus studies of lexical semantics (Reprint.)*. Blackwell.
- Suhaili, A., Sayogie, F., Farkhan, M., & Umami Arif, F. (2025). Racial Color Blindness: A Corpus-Assisted Discourse Analysis of Social Media Reactions to Disney's Little Mermaid Teaser. *GEMA Online® Journal of Language Studies*, 25(3), 622–641. <https://doi.org/10.17576/gema-2025-2503-03>
- Suhaili, A., Suganda, D., Darmayanti, N., & Yuliawati, S. (2024). *A Corpus-Assisted*

Critical Discourse Analysis of the Iranian People in the Corpus of Donald Trump's Social Media Texts. *Theory and Practice in Language Studies*, 14(1), 88–97. <https://doi.org/10.17507/tpls.1401.10>

Suresman, E., Faqihuddin, A., Jenuri, & Abdullah, M. (2025). From sorogan to digital learning: A systematic literature network analysis of pesantren learning models. *Cogent Education*, 12(1), 2580776. <https://doi.org/10.1080/2331186X.2025.2580776>

Zarkasyi, H. F. (2020). Imam Zarkasyi's Modernization of Pesantren in Indonesia (A Case Study of Darussalam Gontor). *QIJIS (Qudus International Journal of Islamic Studies)*, 8(1), 161. <https://doi.org/10.21043/qijis.v8i1.5760>



*Received (04-03-2026)*

*Accepted (01-04-2026)*

*Published (15-04-2026)*