INTERNATIONAL JOURNAL OF TRENDS IN EMERGING RESEARCH AND DEVELOPMENT
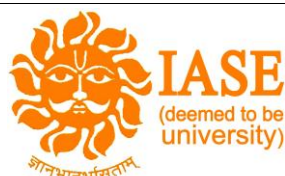
Language and Literature as Knowledge Carriers

Dr. Devangana Pareek

Assistant Professor, Department of Art & Humanities, Government College Gharsana, Rajasthan, India

Corresponding Author: Dr. Devangana Pareek

Abstract

The integration of digital technologies into education has dramatically reshaped pedagogical practices worldwide. In India, this shift carries unique implications for the promotion and revitalization of the Indian Knowledge System (IKS), which encompasses traditional sciences, philosophies, and cultural practices. This paper critically explores the role of digital tools in advancing IKS, addressing the effectiveness of digital versus traditional pedagogies, ethical and legal considerations, privacy in AI-based education, and stakeholder dynamics in innovation ecosystems. Drawing from empirical studies and theoretical frameworks, the research highlights the opportunities digital tools present-such as personalized learning, community-based collaboration, and immersive simulations-while also examining infrastructural, ethical, and socio-cultural challenges. The paper advocates for a balanced, inclusive, and ethically grounded digital approach to preserve and promote IKS within the evolving Indian educational landscape.

Keywords: Digital Education, Educational Technology, Personalized Learning, Indigenous Knowledge, Digital Pedagogy, AI in Education

Introduction

Language and literature, as the primary vehicles of human expression and intellectual endeavour, have long been recognized as central carriers of knowledge across cultures and epochs. The intricate relationship between language-the system of signs and meanings that enables communication-and literature-the creative and systematic deployment of language for aesthetic, cognitive, and social purposes-constitutes the foundation of knowledge transmission, preservation, and innovation. In contemporary academic discourse, the convergence of language and literature with computational and artificial intelligence (AI) methodologies has catalyzed novel forms of knowledge encoding, analysis, and dissemination (Rahimi, 2015; Tovar, 2023) [3, 6]. Simultaneously, advances in large language models (LLMs), natural language processing (NLP), and vision-language models (VLMs) have redefined the epistemic roles of language and literature as structured, dynamic, and scalable knowledge carriers (Du et al., 2025; Tao et al., 2025) [2, 5].

This research paper explores the multifaceted roles of language and literature as knowledge carriers in both traditional and computational contexts. Drawing on recent empirical studies in corpus linguistics, computational linguistics, AI-driven literature review, vision-language modelling, and knowledge transfer in LLMs, the paper examines how language and literature encode, structure, and mediate knowledge. The paper further analyzes the technological advancements that augment or transform these knowledge carrying functions, considering their implications for research, pedagogy, and knowledge management. The discussion integrates findings from state-of-the-art datasets, automated literature review systems, and LLM knowledge fusion frameworks, providing a

comprehensive evaluation of how language and literature, both as natural and artificial constructs, function as primary knowledge carriers in the 21st century.

Theoretical Foundations: Language, Literature, and Knowledge
Language as a Structured Knowledge Carrier
At its core, language is more than a means of communication; it is a systematized repository and transmitter of knowledge. Language encodes conceptual structures, social practices, cultural values, and scientific information, making it indispensable for the intergenerational transmission of knowledge. As Biber, Conrad, and Cortes (2004) contend, language is replete with recurring lexical patterns-such as lexical bundles and phraseological units-that facilitate the efficient packaging and retrieval of information (Rahimi, 2015) [3]. These units, often unconsciously employed by native speakers, exemplify language's capacity to condense and contextualize knowledge within specific genres and discourses.

In computational terms, language's role as a knowledge carrier is formalized in the construction of corpora, the development of NLP tools, and the design of LLMs. Through the systematic analysis of large textual datasets, researchers can extract patterns, structures, and semantic relationships that embody domain-specific knowledge (Rahimi, 2015) [3]. For instance, the extraction and classification of lexical bundles from an 8-millionword corpus of computational linguistics academic literature reveal how language naturally organizes knowledge through frequent, formulaic expressions (Rahimi, 2015) [3]. These bundles function as cognitive scaffolds, supporting both the production and comprehension of specialized knowledge.

Literature as a Repository and Mediator of Knowledge
Literature, encompassing both creative and scientific writing, extends language's knowledge-carrying capacity by embedding information within structured, meaningful, and often aesthetically rich narratives and expositions. Scientific literature, in particular, serves as the cornerstone of academic knowledge, facilitating the accumulation, synthesis, and critical evaluation of research findings (Tovar, 2023) [6]. The process of conducting literature reviews, for example, is itself a knowledge-generative activity, as it requires the identification, extraction, and synthesis of relevant information from a vast array of texts (Ali et al., 2023; Tovar, 2023) [2, 6].

The emergence of AI-driven literature review tools has further highlighted literature's role as a dynamic knowledge carrier. Automated systems leveraging LLMs and NLP techniques can now search, organize, extract, and synthesize information from scholarly texts, thus streamlining the process of knowledge discovery and transmission (Ali et al., 2023; Tovar, 2023) [2, 6]. These developments underscore literature's dual function: as a repository of accumulated knowledge and as a mediated interface for knowledge interaction and production.

The Interplay Between Language, Literature, and Knowledge Technologies
The intersection of language and literature with knowledge technologies-such as AI, LLMs, and VLMs-has profoundly impacted the ways in which knowledge is represented, accessed, and synthesized. Through the formalization of language in computational models and the algorithmic processing of literature, knowledge is rendered both scalable and adaptable to new contexts (Du et al., 2025; Tao et al., 2025) [5]. This interplay foregrounds the need to understand language and literature not only as passive repositories but as active, evolving carriers of knowledge that shape and are shaped by technological innovation.

Empirical Perspectives: Corpus Linguistics and Lexical Bundles as Knowledge Structures
Lexical Bundles and the Organization of Academic Knowledge
Corpus linguistics, as an empirical approach to language study, provides powerful methodologies for uncovering how language structures knowledge in academic discourse. Lexical bundles-recurrent sequences of words that co-occur more frequently than expected by chance-serve as salient indicators of how specialized knowledge is packaged, transmitted, and internalized within academic communities (Rahimi, 2015) [3]. The systematic extraction and classification of lexical bundles from large corpora, such as the computational linguistics academic literature corpus, reveal the functional and structural regularities that underpin disciplinary knowledge (Rahimi, 2015) [3].

Rahimi's (2015) [3] analysis of an 8-million-word corpus identified 591 distinct bundles, ranging from two-word to five-word sequences, which accounted for a significant proportion of the corpus's lexical density. These bundles were categorized according to their frequency, structure, and function. Structurally, bundles were found to cluster into types such as noun phrases with of-phrases ("the end of the"), prepositional phrases ("in terms of"), anticipatory it + verb phrases ("it is important to"), and passive verb + prepositional phrase fragments ("is shown in figure") (Rahimi, 2015) [3]. Functionally, bundles supported research-oriented (e.g., quantification, procedure, description), text-oriented (e.g., structuring, framing, transition), and participant-oriented (e.g., stance, engagement) roles (Rahimi, 2015) [3].

This structural-functional mapping of lexical bundles demonstrates the implicit ways in which language encodes and organizes knowledge within academic texts. Bundles such as "as a result of," "in the case of," and "the extent to which" provide both referential precision and rhetorical cohesion, enabling authors to articulate complex arguments, frame findings, and engage readers (Rahimi, 2015; Biber et al., 2004) [3]. The prevalence of such bundles in scientific writing attests to their epistemic significance: they are not merely linguistic artifacts but are integral to the construction and negotiation of disciplinary knowledge.

Pedagogical and Cognitive Implications
The identification and pedagogical deployment of lexical bundles have far-reaching implications for language teaching, academic writing, and cognitive modeling. As Rahimi (2015) [3] and Salazar (2011) [4] argue, teaching learners to recognize and employ key lexical bundles can enhance their fluency, coherence, and native-like proficiency in academic genres. Such instruction moves

beyond rote memorization of isolated vocabulary, focusing instead on the acquisition of formulaic language patterns that encapsulate domain-specific knowledge.

From a cognitive perspective, the use of lexical bundles reflects the brain's tendency to chunk information into manageable units, facilitating both storage and retrieval. This chunking mechanism mirrors knowledge structures in memory, whereby complex ideas are encoded as interconnected phrases or schemata. Consequently, the study of lexical bundles offers insights into how language mediates cognitive processes of knowledge acquisition, retention, and application.

Language, Literature, and Computational Intelligence: The Evolution of Knowledge Representation
Large Language Models and the Formalization of Knowledge

The advent of large language models (LLMs) marks a paradigm shift in the representation and manipulation of knowledge through language. LLMs, such as GPT-3.5 and LLaMA, are trained on vast corpora of text, enabling them to learn intricate patterns, semantic relationships, and contextual dependencies (Du *et al*., 2025; Ali *et al*., 2023) [2, 1]. These models function as both repositories and generators of knowledge, capable of answering questions, summarizing texts, and synthesizing information across domains.

A critical challenge in LLM research is cross-capability transfer-the ability to transfer knowledge and skills from one model or task to another (Du *et al*., 2025) [2]. Traditional approaches, such as knowledge distillation and parameter merging, have limitations in scalability, adaptability, and knowledge retention. Du *et al*. (2025) [2] introduce the GraftLLM framework, which encodes source model capabilities into compact, transferable Skill Packs, thereby enabling efficient knowledge fusion, continual learning, and forget-free adaptation. By modularizing and compressing task-specific knowledge, GraftLLM exemplifies how language, as instantiated in LLMs, can serve as a scalable and composable knowledge carrier (Du *et al*., 2025) [2].

The implications of this development are profound: language, as modeled by LLMs, is no longer a static container of information but an active agent of knowledge transfer, integration, and innovation. The SkillPack paradigm aligns with the cognitive and social functions of language, facilitating the selective activation, recombination, and adaptation of knowledge units in response to evolving tasks and contexts.

Automated Literature Review: Integrating Language, Literature, and AI

The process of conducting literature reviews epitomizes the knowledge-carrying functions of language and literature in academia. Traditionally, literature reviews require the identification, extraction, organization, and synthesis of information from a dispersed and ever-growing body of texts (Tovar, 2023; Ali *et al*., 2023) [6, 1]. This process is labor-intensive and susceptible to information overload, necessitating efficient tools for knowledge management.

Recent advances in AI and NLP have led to the development of automated literature review suites that leverage LLMs, semantic search, and text embeddings to streamline the review process (Tovar, 2023; Ali *et al*., 2023)

[6, 1]. Tovar's (2023) [6] AI Literature Review Suite, for example, integrates modules for knowledge gathering (searching and downloading PDFs), knowledge extraction (semantic search and querying), and knowledge synthesis (summarization and literature clustering). By automating these stages, the suite transforms literature from a static repository into an interactive, modular, and dynamic knowledge environment.

Ali *et al*. (2023) [1] further compare multiple NLP and LLM-based approaches to automated literature review, demonstrating that retrieval-augmented generation using LLMs (e.g., GPT-3.5-turbo) outperforms frequency-based and transformer-based methods in both accuracy (as measured by ROUGE scores) and user experience. The integration of AI-driven tools with literature review processes exemplifies the convergence of language, literature, and computational intelligence as collaborative knowledge carriers.

Vision-Language Models and Multimodal Knowledge Representation

While traditional language models operate on textual data, vision-language models (VLMs) extend the knowledge-carrying capacity of language to multimodal contexts, integrating visual and linguistic information. In domains such as remote sensing, digital humanities, and scientific imaging, VLMs enable the joint analysis and interpretation of images and texts, facilitating richer and more nuanced knowledge representation (Tao *et al*., 2025) [5].

Tao *et al*. (2025) [5] introduce the FarmSeg-VL dataset, a large-scale image-text benchmark for farmland segmentation, which incorporates language-based descriptions of farmland alongside high-resolution remote sensing imagery. By annotating images with semantically rich captions that capture spatiotemporal characteristics, phenological features, and environmental context, the dataset enables VLMs to model the complex relationships between visual and linguistic knowledge. Experiments show that models trained on image text pairs significantly outperform those trained solely on visual labels, underscoring the epistemic value of language as a structured knowledge carrier in multimodal settings (Tao *et al*., 2025) [5].

This multimodal approach reflects the broader trend toward integrated knowledge systems, where language and literature are no longer confined to text but are embedded within and interconnected with other modalities. The result is an expansion of the scope and granularity of knowledge that can be encoded, retrieved, and synthesized.

Knowledge Encoding, Transfer, and Synthesis: Mechanisms and Challenges
Encoding Knowledge in Language and Literature

The encoding of knowledge in language and literature involves multiple layers of structure, from lexical and syntactic patterns to discourse and genre conventions. Lexical bundles, as discussed earlier, provide a micro-level mechanism for encoding frequently cited concepts, procedures, and relationships (Rahimi, 2015) [3]. At the macro level, genres such as research articles, reviews, and narrative fiction encapsulate domain knowledge through standardized structures (e.g., IMRaD format in scientific

writing), rhetorical strategies, and intertextual references.

In computational frameworks, encoding knowledge in language involves the design of datasets, annotation schemas, and model architectures that capture relevant features and relationships (Tao *et al*., 2025; Du *et al*., 2025) [5, 2]. The development of dedicated image-text datasets, such as FarmSeg-VL, requires the careful selection and formalization of

descriptive elements that are both semantically rich and machine-interpretable (Tao *et al*., 2025) [5]. Similarly, the modularization of LLM knowledge into SkillPacks necessitates the identification and compression of task-specific parameter deltas, preserving task knowledge while optimizing storage and transferability (Du *et al*., 2025) [2].

Knowledge Transfer: From Human Learners to Artificial Models

The transfer of knowledge-whether between individuals, across generations, or from human experts to artificial models-depends on the effective encoding and retrieval of information in language and literature. In educational settings, explicit instruction in lexical bundles, discourse markers, and genre conventions facilitates the transfer of academic literacy and domain expertise (Rahimi, 2015; Salazar, 2011) [3, 4]. In computational contexts, knowledge transfer involves the adaptation of pretrained models to new tasks or domains, often through techniques such as fine-tuning, distillation, or parameter grafting (Du *et al*., 2025) [2].

Du *et al*. (2025) [2] identify key challenges in cross-capability transfer for LLMs, including the risk of catastrophic forgetting, parameter conflicts, and the limitations of full-parameter fine-tuning. The GraftLLM approach addresses these challenges by modularizing and compressing knowledge into SkillPacks, enabling both efficient transfer and continual learning without loss of general capabilities. This approach mirrors cognitive theories of knowledge transfer, which emphasize the importance of modular, context-sensitive knowledge units that can be selectively activated and recombined.

Knowledge Synthesis: Toward Collective and Automated Intelligence

Knowledge synthesis-the integration of information from multiple sources to generate new insights or comprehensive overviews-is a central function of language and literature. Literature reviews, meta-analyses, and systematic reviews are prominent examples of knowledge synthesis in academic research (Tovar, 2023; Ali *et al*., 2023) [6, 1]. The automation of knowledge synthesis through AI tools augments human capabilities, enabling researchers to process larger volumes of information and to identify patterns, trends, and gaps more efficiently.

Tovar's (2023) [6] AI Literature Review Suite and Ali *et al*.'s (2023) [1] comparative study of NLPbased literature review systems demonstrate the feasibility and effectiveness of automated knowledge synthesis. By leveraging LLMs, semantic search, and clustering algorithms, these systems can extract, summarize, and organize information from diverse sources, providing structured outputs such as literature tables, thematic clusters, and synthesized narratives. These developments suggest a future in which language and literature, mediated by AI, serve as both sources and agents of collective intelligence.

Case Studies: Language and Literature as Knowledge Carriers in Practice
Case Study 1: Lexical Bundles in Computational Linguistics Literature

Rahimi's (2015) [3] corpus-based study of lexical bundles in computational linguistics academic literature offers a concrete illustration of how language structures and transmits knowledge within a scientific community. By analyzing an 8-million-word corpus, Rahimi identified high-frequency bundles that function as linguistic shortcuts for expressing complex ideas, procedural steps, and evaluative judgments. The study's categorization of bundles into structural and functional types highlights the regularities and variations that characterize disciplinary discourse.

The findings reveal that research-oriented bundles (e.g., "is shown in figure," "can be used in"), text-oriented bundles (e.g., "on the other hand," "as a result of"), and participant oriented bundles (e.g., "we're going to talk about," "it can be seen that") collectively support the efficient encoding, retrieval, and synthesis of domain knowledge. The pedagogical implications extend to academic writing instruction, where explicit teaching of these bundles can improve learners' fluency and cognitive access to disciplinary knowledge (Rahimi, 2015) [3].

Case Study 2: Farm Seg-VL and Vision-Language Modeling in Remote Sensing

Tao *et al*.'s (2025) [5] construction of the Farm Seg-VL dataset exemplifies the integration of language and vision as knowledge carriers in scientific research. The dataset pairs high resolution remote sensing images of farmland with detailed, semi-automatically generated captions that describe spatiotemporal features, crop phenology, spatial distribution, and environmental context. The inclusion of language-based descriptions enables VLMs to model the dynamic and heterogeneous nature of farmland, overcoming the limitations of label-driven deep learning approaches.

Performance analyses indicate that models trained on image-text pairs achieve higher segmentation accuracy and greater generalizability across seasons and regions, demonstrating the added value of language as a structured knowledge carrier (Tao *et al*., 2025) [5]. The study underscores the importance of multimodal datasets and annotation strategies in advancing knowledge representation and discovery in complex domains.

Case Study 3: AI-Driven Literature Review and Knowledge Synthesis

The proliferation of scholarly publications has necessitated the development of automated literature review tools that can handle the scale and complexity of contemporary research landscapes. Tovar's (2023) [6] AI Literature Review Suite and Ali *et al*.'s (2023) [1] comparative evaluation of NLP and LLM-based systems illustrate the practical benefits and challenges of automating knowledge synthesis. By integrating modules for document retrieval, content extraction, semantic search, and summarization, these tools enable researchers to rapidly synthesize information from large corpora of scientific texts.

Empirical results show that LLM-based approaches, such as GPT-3.5-turbo, achieve higher accuracy in generating coherent and relevant literature reviews compared to older NLP methods (Ali *et al*., 2023) [1]. The capacity to cluster, summarize, and synthesize literature at scale redefines the epistemic functions of language and literature, positioning them as active interfaces for collective knowledge generation.

Case Study 4: Knowledge Grafting in Large Language Models

Du *et al*.'s (2025) [2] GraftLLM framework addresses the challenge of cross-capability transfer in heterogeneous LLMs by encoding task-specific knowledge into modular SkillPacks. This approach facilitates efficient knowledge transfer, fusion, and continual learning, enabling models to acquire new abilities without forgetting previous knowledge. The modularization of knowledge mirrors the cognitive organization of language and supports scalable, composable intelligence.

Experimental results demonstrate that GraftLLM outperforms existing techniques in knowledge transfer, fusion, and forget-free learning, providing a scalable and efficient solution for integrating knowledge across diverse models and tasks (Du *et al*., 2025) [2]. The SkillPack paradigm illustrates how language, as instantiated in LLMs, can be engineered to function as a flexible and adaptive knowledge carrier. Challenges, Limitations, and Future Directions

Challenges in Encoding and Transferring Knowledge

Despite significant advances, several challenges persist in encoding, transferring, and synthesizing knowledge in language and literature. In human contexts, the implicitness, ambiguity, and contextuality of language can hinder precise knowledge transfer, particularly across languages, cultures, and disciplinary boundaries. Non-native speakers, for example, often struggle to reproduce the native-like use of lexical bundles, affecting the clarity and fluency of their academic writing (Rahimi, 2015; Salazar, 2011) [3, 4]. In computational contexts, the alignment of semantic representations across tasks and domains remains a complex problem, with risks of information loss, bias, and overfitting (Du *et al*., 2025) [2].

Technological solutions, such as modularization, compression, and multimodal integration, offer promising avenues for addressing these challenges. However, the rapid evolution of AI and NLP tools necessitates ongoing research into the interpretability, transparency, and ethical implications of automated knowledge systems.

Limitations of Current Approaches

Current approaches to automated literature review, knowledge synthesis, and multimodal modeling are limited by the quality and coverage of datasets, the robustness of annotation schemas, and the scalability of model architectures. For instance, while LLM-based literature review tools outperform older methods in accuracy and coherence, they may still struggle with nuanced domain-specific synthesis, hallucination, or failure to capture emerging trends (Ali *et al*., 2023; Tovar, 2023) [1, 6]. Similarly, vision-language models depend on the availability of high-quality, semantically rich image-text datasets, which may be lacking in less-studied domains (Tao *et al*., 2025) [25].

The modularization of LLM knowledge into SkillPacks, while effective in many scenarios, may encounter difficulties in capturing complex, interdependent knowledge structures or in adapting to radically new tasks. These limitations underscore the need for continued methodological innovation and interdisciplinary collaboration.

Future Directions

The future of language and literature as knowledge carriers lies in the development of more integrated, adaptive, and transparent systems that harness the strengths of both human and artificial intelligence. Research priorities include:

1. **Expanding and Diversifying Datasets:** The creation of larger, more diverse, and more representative corpora-encompassing multiple languages, genres, and modalities-will enhance the capacity of language and literature to encode and transmit knowledge.

2. **Advancing Multimodal and Multilingual Modeling:** Integrating textual, visual, and auditory modalities, as well as supporting cross-linguistic transfer, will enable richer and more flexible knowledge representation.

3. **Improving Interpretability and Transparency:** Developing tools and frameworks for explaining how language models encode, transfer, and synthesize knowledge will enhance trust and usability in academic and practical settings.

4. **Ethical and Social Considerations:** Addressing issues of bias, fairness, and inclusivity in automated knowledge systems is essential to ensuring that language and literature remain equitable and accessible knowledge carriers.

5. **Human-AI Collaboration:** Designing interfaces and workflows that facilitate effective collaboration between human experts and AI systems will maximize the epistemic potential of language and literature.

Conclusion

Language and literature have long served as the primary carriers of human knowledge, encoding, structuring, and mediating information across generations, disciplines, and cultures. The advent of computational linguistics, NLP, LLMs, and VLMs has transformed and expanded these knowledge-carrying functions, enabling new forms of encoding, transfer, and synthesis that transcend traditional boundaries. Empirical studies on lexical bundles, multimodal datasets, automated literature review, and knowledge grafting in LLMs illustrate the evolving landscape of knowledge representation and management.

Despite ongoing challenges and limitations, the integration of language and literature with AI and computational methodologies offers unprecedented opportunities for collective intelligence, innovation, and discovery. As we continue to develop and refine these systems, it is imperative to preserve the richness, diversity, and contextuality of language and literature, ensuring that they remain flexible, transparent, and inclusive carriers of knowledge in an increasingly complex world.

References

1. Ali NF, Mohtasim MM, Mosharrof S, Krishna TG. Automated literature review using NLP techniques and LLM-based retrieval-augmented generation [Internet]. 2023 [cited 2025 Oct 1]. Available from: http://arxiv.org/pdf/2411.18583v1
2. Du G, Zhou X, Li J, Li Z, Shi Z, Lin W, *et al*. Knowledge grafting of large language models [Internet]. 2025.
3. Rahimi A. Lexical bundles in computational linguistics academic literature [Internet]. 2015.
4. Salazar D. Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching. Amsterdam: John Benjamins Publishing Company; c2011.
5. Tao C, Zhong D, Mu W, Du Z, Wu H. A large-scale image-text dataset benchmark for farmland segmentation [Internet]. 2025.
6. Tovar DA. AI literature review suite [Internet]. 2023.