International Journal of Education and Social Development
ISSN: 3078-9931 (Print), ISSN: 3078-2287 (Online) | Vol. 7, No. 3, 2026

A Review of the Relationship Between Sensory Encoding and Lexical Encoding from the Perspective of Working Memory Development

Longkun Duan

Ontario Institute for Studies in Education, University of Toronto, Toronto, Canada

Abstract: The acquisition and retention of linguistic knowledge are based on two main processes, namely sensory encoding and lexical encoding, although their developmental relation and the working memory architecture regulating this link are not yet well understood. This research analyses the multicomponent working memory model, in particular the phonological loop and visuospatial sketchpad, in the context of the interplay of modality-specific sensory information with lexical representations as a function of age. Drawing on longitudinal developmental studies, studies on the modality effect, studies on multisensory integration and L2 acquisition, the paper argues for systematic developmental change at the sensory-lexical interface in working memory: the initial dependence on auditory sensory encoding is gradually replaced by more flexible, multimodal lexical representations as the capacity of phonological working memory and the ability for strategic encoding develops. The episodic buffer is suggested as the fundamental convergence point between multi-sensory coding and long-term lexical knowledge with ramifications for vocabulary education and future study.

Keywords: Working memory; sensory encoding; lexical encoding; phonological loop; multimodal; language development; vocabulary acquisition.

1. Introduction

1.1. Research Background

One of the most enduring topics in cognitive science concerns the relation between perception and language. When learners encounter a new word, in speech, text, or the face of a speaker, they must first encode the perceptual information through modality-specific sensory channels, and then consolidate this signal into a lexical representation that is strong enough to facilitate later identification and use. The fundamental link in this translation from sensory coding to lexical form is working memory (WM), the limited-capacity cognitive mechanism that temporarily retains and manipulates information to assist ongoing mental activities. Understanding the course of this translation over the lifespan has crucial consequences for theories of language acquisition, reading development, and vocabulary education.

The dominant theoretical framework to understand this relationship is the multicomponent model of working memory (Baddeley and Hitch , 1974 , and subsequent refinements over the decades) . At the heart of the paradigm are two auxiliary systems for modality-specific information. These include the phonological loop which stores and rehearses speech-based representations and the visuospatial sketchpad which analyses visual and spatial information. Specifically, a fourth component, the episodic buffer, was hypothesised to accommodate the integration of information across subsidiary systems and with long term memory. Thus it provides a theoretically motivated site for the convergence of sensory and linguistic codes.

The development of working memory has been the subject of study for almost five decades and there is a vast parallel literature on lexical learning. However, there has been no comprehensive examination of the unique developmental trajectory of the sensory-lexical interface in working memory. Four theme categories are reviewed for evidence: working memory development across the lifespan, the phonological loop's role in vocabulary acquisition, the effect of modality on lexical encoding, and working memory multisensory integration. Taken together, these lines of data are consistent with a developmental model in which the sensory grounds of lexical encoding shift from auditory dominance in early childhood to flexible multimodal representation in mature learners.

1.2. Research Questions

The review answers three questions. RQ1. What is the developmental change in the link between sensory encoding and lexical encoding from infancy to adulthood in working memory? RQ2. How does the phonological loop and episodic buffer mediate the sensory-lexical interface? RQ3: What are the implications of developmental alterations in this interface for vocabulary acquisition and language teaching?

2. Theoretical Framework: The Multicomponent Working Memory Model

2.1. Core Architecture and Developmental Relevance

The multicomponent model of working memory has a central executive, an attentional control system of limited capacity and two domain-specific subsidiary systems. The phonological loop stores and rehearses phonological sequences via a passive phonological storage and an articulatory rehearsal mechanism, whereas the visuospatial sketchpad retains visual and spatial representations [1]. In 2000 Baddeley introduced the episodic buffer filling an important gap in the three-component model: the necessity for a limited capacity, multimodal store that combines information from the phonological loop, visuospatial sketchpad and long-term memory into coherent episodes [2].

The episodic buffer is largely accessed by conscious attention via the central executive and is a passive storage to which consciousness has access. It is the interface through which multisensory inputs are aligned with present lexical knowledge .

Support for the hypothesis that the phonological loop explains the learning of new words in early childhood but with growing involvement of present lexical knowledge in phonological WM as vocabulary increases is provided by a fifty year retrospective by Hitch, Allen and Baddeley [3]. Understanding how the sensory-lexical link develops over development requires understanding this shift from phonological WM as the main engine for word learning, to a system in which phonological and semantic long-term memory interact.

2.2. Working Memory Capacity Development

Cowan's fifty-year study of the development of working memory demonstrates that performance on WM tasks grows considerably across childhood and adolescence, but the reasons for this improvements are complicated and not reducible to a single capacity gain [4]. Developmental increases in WM are seen as increased storage capacity, more efficient rehearsal techniques, better filtering of distractors, and larger knowledge bases that promote chunking [4]. Importantly, Cowan considers the improvement of encoding technique to be a major aspect in WM progress. This implies that the coding of sensory inputs and the use of phonological recoding for visual material differ markedly with age in children.

This developmental trend has been observed in the verbal domain. Subvocal rehearsal strategies are not used for phonological working memory preservation in young children under about six years of age; rehearsal strategies arise around age seven and WM performance increases [1][4]. This milestone has direct consequences for the sensory-lexical interaction. When rehearsal strategies aren't available, visually displayed words are not phonologically recoded in the loop. Thus, visual sensory encoding and verbal encoding appear to be more autonomous in early children than in later learners.

3. The Phonological Loop as a Sensory-Lexical Bridge

3.1. Phonological Working Memory and Vocabulary Acquisition

The most extensively studied relationship between working memory and lexical development is that of phonological working memory in the acquisition of vocabulary. In longitudinal research Gathercole and Baddeley found that phonological WM capacity, measured by non-word repetition, was a robust predictor of vocabulary size in children aged four to eight, with the connection highest in younger children and reducing as vocabulary increases [5]. This made the phonological loop a language learning device, retaining the phonological shape of a new word long enough to acquire a permanent representation in long-term lexical memory [1][5].

The phonological loop is an asymmetrical link between sensory and lexical encoding. Heard words have immediate access to the phonological store; read words must be recoded into an articulatory form first. Thus, auditory inputs have a more direct route to lexical encoding than visual inputs – a prediction verified by patient evidence that phonological WM deficiencies primarily influence auditory vocabulary learning [1][2]. For preliterate children without accurate articulatory recoding, lexical learning is almost exclusively dependent on auditory sensory input from the phonological storage.

3.2. Phonological Short-Term Memory and Second Language Lexical Encoding

The role of the phonological loop as a sensory-lexical bridge has been most strongly supported in the domain of second language vocabulary acquisition, where learners must map wholly novel phonological forms onto an established lexical system. A meta-analysis of 79 samples by Linck and colleagues further demonstrated that phonological and executive WM had separate contributions to L2 learning, with phonological WM being more predictive of acquisition of new word forms [14]. In a more recent meta-analysis of 52 effect sizes, Kuribara found a small-to-medium effect of phonological short-term memory on auditory L2 vocabulary acquisition, with non-word repetition showing the strongest associations, consistent with the loop's role in encoding acoustic-sensory features into emergent lexical representations [6].

The age moderation found in Kuribara's meta-analysis is particularly important to the developmental angle of this review: the association between phonological short-term memory and L2 vocabulary was largest among younger learners and decreased with age [6]. This developmental gradient resembles the pattern that Gathercole and Baddeley found for first language vocabulary and suggests a general principle in development: early in the process of vocabulary development, whether in the first or second language, the capacity of the phonological loop to hold sensory-encoded phonological forms is the primary driver of lexical encoding and as vocabulary knowledge expands, semantic and contextual knowledge play a larger role in lexical encoding than phonological working memory in the form of the phonological loop.

4. Modality Effects and Multisensory Integration in Lexical Encoding

4.1. Sensory Modality Effects on Lexical Encoding and Retrieval

The modality of presentation of a word has a significant influence on its encoding in working memory and its lexical access. Studies of the modality effect have shown that the learning and retrieval of words presented in auditory-visual (AV) format is faster than the encoding and retrieval of words shown in auditory-only or visual-only format, and that the AV advantage is due to richer stimulus information as well as reduced attentional demands [7]. Visual-only encoding was significantly slower and less accurate, consistent with the phonological loop model's assumption that articulatory recoding is required prior to phonological storage [7]. The time course of lexical encoding and attentional demands routinely varies between modalities even when stimulus quality is controlled, demonstrating that the sensory channel impacts the architecture not only of the quality of lexical encoding.

Developmental research on the sensory format of lexical representations provides supplementary evidence for modality effects that change with age. Havy et al. showed that 14-month-old infants encode auditorily presented novel words in an auditory-specific structure, as they recognised the word in the auditory modality but did not demonstrate cross-modal transfer when tested with visual speech (silent talking face) [8]. By 18 months infants evidence growing cross-modal flexibility: auditorily learnt words are recognisable from visual speech, indicating that lexical representations have begun to incorporate visual as well as auditory sensory input [8][9]. Adults show full cross-modal flexibility in word recognition compatible with lexical models that integrate sensory information across modalities. This developmental sequence, from infant auditory-dominant modality-specific lexical encoding to adult flexible multimodal representation, reflects directly the developing capacity of the episodic buffer to integrate inputs from the phonological loop and visuospatial sketchpad with long-term lexical knowledge.

4.2. Multisensory Integration and Working Memory Development

There are major developmental changes in the ability to encode information from numerous sensory modalities simultaneously into working memory in childhood. Cowan and colleagues have shown a considerable developmental gain in the ability to encode visual and aural stimuli simultaneously over childhood [10]. Importantly, this enhancement was not due to sharing of attention: first-grade children performed equally well for successive and simultaneous audiovisual stimuli, suggesting that the enhancement is due to increased storage capacity or more effective binding, rather than to improved division of attention [10].

Louwerse et al. performed a systematic review and meta-analysis of multisensory processing and working memory development and found that multisensory verbal stimuli were more advantageous for verbal working memory capacity than unisensory auditory or visual stimuli alone and that the multisensory advantage was significantly modulated by age [11]. Meta-regression showed that the effect of multisensory processing on WM performance is age-dependent with younger children showing a greater proportional gain from multimodal encoding conditions. This age range is compatible with the episodic buffer integration account. As the buffer grows larger and its connections to long-term lexical memory get stronger, it becomes increasingly efficient and robust at integrating sensory codes from different modalities into unified lexical representations.

Recently, Arslan and colleagues have demonstrated evidence for a significantly higher attentional cost of cross-modal binding at retrieval than within-modality encoding in audiovisual working memory tasks, indicating the alpha oscillatory activity reflects the additional cognitive demands of cross-modal binding [12]. This attentional cost of multisensory lexical encoding has direct implications for understanding why young children, who have more limited central executive resources than older children and adults, disproportionately employ single-modality (typically auditory) sensory encoding as the basis for lexical representation in working memory.

5. Discussion

5.1. A Developmental Model of the Sensory-Lexical Interface

Together, the evidence in sections 3 and 4 supports a cohesive developmental model of the sensory-lexical interaction. In early infancy and childhood, lexical learning is largely mediated by auditory sensory encoding via the phonological loop, consistent with the primacy of speech and the limited capacity of the episodic buffer for multimodal integration in this developmental window. As phonological WM and rehearsal methods develop, visual inputs are progressively phonologically recoded and linked with auditory lexical representations. The increasing episodic buffer leads to more strong binding across modalities and with long-term lexical information and so generates the versatile multimodal representations that characterise adult word knowledge.

The developmental account provides principled explanations for several well-documented empirical findings: the primacy of auditory sensory input in infant word learning [8][9]; the strong predictive relationship between phonological working memory and vocabulary development in young children that attenuates as vocabulary grows [5][6]; the emergence of cross-modal word recognition flexibility between 14 and 18 months; and the developmental improvement in concurrent multimodal working memory encoding across the school years [10][11]. The model predicts that therapies focusing on phonological working memory should be most helpful for early vocabulary development, but interventions that enhance multimodal sensory integration should become increasingly significant as infants develop the capacity for multimodal lexical encoding.

5.2. Implications and Reinterpretation of Classical Findings

The developmental model we present here also implies a reinterpretation of classical WM effects. The phonological similarity effect (poor remembering of lists of phonologically similar words) has been historically attributed to interference inside the phonological loop [1]. However, it is a developmental matter that the magnitude of this effect varies with age: the phonological similarity effect is partially attenuated as semantic codes become available to WM in the form of links to long-term memory through the episodic buffer, reflecting the increasing importance of lexical-semantic rather than purely sensory-phonological encoding [13].

RQ1: The sensory-lexical link in working memory is consistently modulated by development, ranging from an auditory-dominant phonological loop-based lexical encoding in early childhood to a flexible multimodal lexical encoding supported by the emerging episodic buffer in adolescence and maturity.

Answer RQ2: The key sensory-lexical bridge in early development is the phonological loop , which provides temporary storage of new phonological forms. With development, the episodic buffer becomes more important, supporting multimodal sensory integration and the coupling of sensory codes to long-term lexical information.

6. Conclusion

The present review has shown that under the multicomponent working memory model the relation between sensory encoding and lexical encoding is not stable or unidimensional, but changes in a systematic and theoretically predictable fashion during the lifespan. Thus, the phonological loop is developmentally essential and most crucial for early vocabulary acquisition as a language learning device, momentarily holding sensory-encoded phonological patterns for lexical consolidation. The significance of the episodic buffer in integrating multimodal sensory information with long-term lexical information becomes increasingly important as the capacity of working memory and the skills involved in strategic encoding improve .

As for RQ3: The proposed developmental account has immediate implications for language education. The best ways for improving the quality of auditory sensory encoding for early childhood and early L2 acquisition are clear speech, low phonological competition and multimodal presentation that supports rather than divides attention. Such approaches should optimise lexical encoding via the phonological loop . Strategies for multimodal integration and episodic buffer binding (i.e. mixing auditory, visual, and semantic input) should exploit the growing ability for flexible multimodal lexical representation in older children and adult learners. Future research should investigate the neural correlates of the developmental trajectory of the episodic buffer and its association with individual differences in language acquisition.

References

[1] Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839. https://doi.org/10.1038/nrn1201.

[2] Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

[3] Hitch, G. J., Allen, R. J., & Baddeley, A. D. (2025). The multicomponent model of working memory fifty years on. *Quarterly Journal of Experimental Psychology*, *78*, 387–421. https://doi.org/10.1177/1747021824131425

[4] Cowan, N. (2022). Working memory development: A 50-year assessment. *Cognition*, *224*, 105075. https://doi.org/10.1016/j.cognition.2022.105075

[5] Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children. *Journal of Memory and Language*, *29*, 336–360. https://doi.org/10.1016/0749-596X(90)90022-Q

[6] Kuribara, K. (2026). Revisiting the effects of phonological short-term memory on L2 vocabulary acquisition: A meta-analysis. *ITL — International Journal of Applied Linguistics*, *176*, 1–32.

[7] Gordon-Salant, K. A., Sommers, J. B., & Zobel, J. M. (2020). Modality effects on lexical encoding and memory representations of spoken words. *Ear and Hearing*, *41*, 1243–1254. https://doi.org/10.1097/AUD.0000000000001546

[8] Havy, M. (2024). Toward an auditory dominance in early word learning. *First Language*, *44*, 203–226. https://doi.org/10.1177/01425694231224653

[9] Havy, M., & Floccia, J. (2018). The role of auditory and visual speech in word learning at 18 months. *Developmental Science*, *21*, e12480. https://doi.org/10.1111/desc.12480

[10] Cowan, N., Blume, J. S., & Saults, E. M. (2024). Attention-sharing in the development of audiovisual working memory. *Developmental Science*, *27*, e13552. https://doi.org/10.1111/desc.13552

[11] Louwerse, M. M., Stein, C., Briner, E., & Connell, J. (2025). From senses to memory during childhood: A meta-analysis of multisensory processing and WM development. *European Journal of Investigation in Health, Psychology and Education*, *15*, 157. https://doi.org/10.3390/ejihpe15030157

[12] Arslan, C., Schneider, D., Getzmann, S., Wascher, E., & Klatt, L.-I. (2025). Multisensory processing and attention in working memory: Audiovisual object storage. *Psychophysiology*, *62*, e70018. https://doi.org/10.1111/psyp.70018

[13] Camos, K., & Barrouillet, P. (2011). Working memory in development: A dual-component model. *Developmental Psychology*, *47*, 914–928. https://doi.org/10.1037/a0023736

[14] Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*, 861–883. https://doi.org/10.3758/s13423-013-0545-y