International Journal of Social Science, Education and Humanities
ISSN Print: 3104-4239
ISSN Online: 3104-4247
CD ISSN: 3104-4255
https://doi.org/10.63313/IJSSEH.9044

A Study on the Multimodal Synergetic Path of Junior High School English Vocabulary Memory under the Lexical Chunk Theory – Taking Unit 4 Eat Well as an Example

Ziyi Yin*. Ming Deng
China West Normal University, School of Foreign Language, Nangchong 637000, China
Email: 1610173139@qq.com

How to cite this paper: Yin, Z., & Deng, M. (2026). A study on the multimodal synergetic path of junior high school English vocabulary memory under the lexical chunk theory – Taking Unit 4 Eat Well as an example. International Journal of Social Science, Education and Humanities, 2(3), 63–73. ISSN Print: 3104-4239, ISSN Online: 3104-4247.
https://doi.org/10.63313/IJSSEH.9044
Published: 2026-06-29

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
CC BY 4.0

Abstract

Aiming at the problem of low memory efficiency in junior high school English vocabulary teaching, this study constructs a three-layer progressive vocabulary memory enhancement path of ' input-processing-output ' based on lexical chunk theory and multimodal semiotics theory. The study integrates the lexical chunk classification system and multimodal synergy theory, combining the cognitive characteristics of junior high school students, to design a multimodal corpus, a gamification strategy, and a high-frequency transfer mechanism. It then implements teaching practice with Unit 4 as a case. The results show that this model can effectively enhance the effect of vocabulary memory, facilitate the transformation from "mechanical memory " to "meaning construction," and provide an operational teaching paradigm for utilizing lexical chunks in the context required by the new curriculum standards.

Keywords

Lexical Chunk Theory; Multimodal Teaching; Junior High School English; Vocabulary Memory; Synergistic Path

1. Introduction

For years, junior high English vocabulary teaching has faced low memory efficiency and short retention, contradicting new curriculum standards. The 2022 Compulsory Education English Curriculum Standard defines vocabulary as all words and set phrases, stressing that learning goes beyond sound, form, and meaning to include understanding and using lexical chunks in context to master collocations, pragmatic functions, and cultural connotations. [1] Multimodal semiotics provides a solution to this issue, since meaning arises from the synergy of symbolic resources: language, image, sound, and action. [2] Lexical chunks, as prefabricated units, are learned

through multimodal integration of form-function-context; multimodal teaching activates sensory channels to build deep associations, aligning with human symbol cognition. [3] This paper intends to explore the connotations and origins of lexical chunk theory and multimodal theory, analyze their mutual adaptability and contemporary teenagers' memory characteristics, and then discuss their application paths in new textbooks so as to update teachers' methods and promote students' vocabulary development.

2. Research Significance

Junior high vocabulary learning suffers from low efficiency, short retention, and weak application. Drawing on lexical chunk theory and multimodal semiotics, and given teens' tech-dependence yet weakening autonomy, this study offers multidimensional solutions.

First, lexical chunks as prefabricated units enable holistic storage; multimodality integrates language, image, and sound. Their synergy reduces memory units, expands channels, and builds form–function–context links via multisensory interaction, boosting efficiency. Second, multimodal situational interaction and high-frequency chunk transfer, via gamification, turn chunks into embodied experiences. Multimodal coding plus contextual application extends memory traces and lowers cognitive load, converting short-term into long-term memory, solving the forget-after-learning issue. Third, cross-adaptation enables pragmatic chunk extraction in real contexts. Multimodal corpus input and task-based output let students perceive collocation rules and cultural connotations, shifting from mechanical memory to meaning construction, improving pragmatic competence. This integrated model effectively breaks through vocabulary difficulties.

3. Theoretical combing

3.1. Lexical chunk theory

3.1.1. Theoretical Development

Becker first proposed lexical chunks, stating language use relies on fixed/semi-fixed patterns, not single words, as basic communication units. [4]

A lexical chunk is a prefabricated, continuous or discontinuous multi-word unit, with structure and meaning, and is stored and retrieved as a whole in real-time communication. [5] Nattinger et al. emphasize the importance of lexical chunks at the pragmatic level, and divide lexical chunks into four categories according to their structural characteristics:

(1) Polywords, which refer to the fixed structure of words, such as calm down, as a result, well up;

(2) institutionalized expressions, which usually appear in fixed forms, including proverbs and social discourses, such as ' Well begun is half done’. Roses given, fragrance in hand;

(3) phrasal constraints (phrasal constraints), which are flexible and extensible, generated by embedding appropriate words or phrases into a given phrasal framework, such as far as... concerned, think highly / little of;

(4) Sentence builder refers to the lexical phrase that constructs the framework for the whole sentence, which is usually called sentence pattern, such as 'It is universally acknowledged that... '. [6] Lewis proposed the lexical approach, viewing language as grammaticalized lexis rather than lexicalized grammar. He argued that learners' knowledge is a collection of various chunks, not isolated words or grammatical rules. He emphasized effective communicative use and the importance of input/output, as words commonly occur in phrases and larger chunks. [7] Lexical chunks are language units with relatively stable grammatical structure, fixed collocational meaning, and specific pragmatic situations, skillfully integrating grammar, semantics, and context. Holistic retrieval avoids the tedious recombination of individual words and grammar rules, lowering language output difficulty and allowing focus on meaning and expression rather than on formal structure.

3.1.2. Definition and classification of lexical chunks

(1) Core definition of lexical chunks

Lewis subverts the traditional dichotomy, proposing that language consists of chunks, which combine to produce coherent text. These chunks are prefabricated, stored and extracted as wholes, and their nature is grammaticalized lexis rather than lexicalized grammar. They integrate grammatical, semantic, and pragmatic functions, forming the basic units of language communication. [8]

(2) Four classification systems of lexical chunks

The lexical chunk classification system reveals that ' the traditional distinction between ' fixed ' vocabulary and ' generative ' grammar is an invalid over-simplification '. Lewis classifies lexical chunks into four categories according to their structural features:

Type	Definition and features	Examples
Words & Polywords	Words which can stand alone…and polywords, arbitrary combinations, an assortment of mini-idioms	Book, by the way, nevertheless, all of a sudden
collocations	Certain words co-occur in natural text with greater than random frequency	Commit suicide, heavy rain, make a mistake, dramatically increase
Fixed Expressions	Fully fixed expressions must be acquired as whole	Good morning, happy new year, it's not my fault.
Semi-fixed expressions	Semi-fixed 'frames' with 'slots' which may be filled in a limited number of ways.	Could you pass…, please? What was really (adj.) was.

Tab 1. Categories of lexical chunks

Polywords are fixed and inseparable; collocations have a gradient, completely fixed (catch a cold) or relatively free (prices fell sharply); semi-fixed are the most generative, allowing finite substitutions, e.g., I haven't seen you for/since.

3.1.3. Cognitive basis of lexical chunk theory

Cognitive linguistics distinguishes two types of knowledge: declarative ('what') and procedural ('how'). Language learning aims to convert procedural into declarative through repeated practice, i.e., mastering lexical chunks in specific contexts, ultimately achieving automation. Automation, an ideal state for learners, means language is stored and used holistically without decomposition, saving time and boosting fluency and accuracy in input and output. Memory plays a vital role in vocabulary acquisition.

In language learning, knowledge initially goes into short-term memory, which retrieves related knowledge from long-term memory to process it. After repeated practice, the knowledge becomes permanently stored in long-term memory. However, short-term memory has a very limited capacity, only about 7 units (roughly 7 words), yet most English sentences exceed that length. An effective way to solve this problem is to agglomerate this information, thereby reducing the storage burden and ultimately increasing the total amount of storage. The number 1330109872461997 seems difficult to remember at first glance. If it is cut into pieces by 133 010 987 246 1997, it seems much easier. Short-term memory plays a crucial role. According to Gestalt psychology, people naturally grasp wholes before parts, and then parts within wholes—this order is irreversible. For example, we immediately recognize a circle even with a missing angle, or a house before its windows. However, traditional vocabulary teaching violates this cognitive law by isolating words from lexical chunks and contexts, cutting off internal semantic connections. The outcome is like blind men touching an elephant—students acquire only fragmented knowledge of individual words and fail to form any holistic impression. The superposition of these parts will not get an organic whole, which is equivalent to our usual problem that 1 + 1 is not equal to 2. [9]

That is to say, the application of lexical chunk theory to vocabulary memory is very beneficial to students ' vocabulary memory efficiency.

3.2. Multimodality theory and its adaptability to vocabulary teaching

3.2.1. The core connotation of multimodality theory

Based on social semiotics, multimodal theory advocates that human communication constructs meaning through multiple semiotic modes such as language, image, and sound. Its core connotation can be deconstructed into three dimensions:

First, from the perspective of social semiotics, modality is ' the meaning potential of material media shaped by society '. [10] In the digital age, multimedia transforms language into compound discourse, and multimodality becomes an inherent sociocultural attribute. [11] In class, blackboard, PPT, and body language form the symbol system of knowledge transfer.

Modality extends systemic functional linguistics, grounded in Halliday's three meta-functions. Visual grammar (Kress & van Leeuwen) realizes conceptual,

interactive, and compositional meanings—e.g., food images present, perspective engages, color highlights. Auditory grammar conveys social relations via sound hierarchy, as teacher intonation versus background noise implies interaction distance. Image-text relations show semantic complementarity; for instance, a recipe with slicing animation enriches meaning. These multimodal grammars collectively demonstrate how non-verbal symbols systematically construct meaning in communication.

Intermodal coordination and resemiotisation are key concepts in multimodal analysis. Coordination involves multiple modalities working through complementarity and reinforcement to generate holistic meaning, exemplified by science textbooks pairing formulas with animations. Resemiotisation refers to the reshaping of semiotic patterns due to social transformations, such as flipped classrooms reconfiguring spatial modes and thereby changing power dynamics.

3.2.2. Adaptation mechanism of multimodal theory and vocabulary teaching

The core characteristics of multimodal theory are deeply coupled with the needs of vocabulary teaching, and its adaptability is reflected in three dimensions:

Multimodal input strengthens vocabulary memory via PPT integrating text, image, and color. Concrete nouns like 'battle' use images; action verbs like 'stagger' use video. Multi-sensory coordination expands pathways, aligning with cognitive load theory—e.g., touching 'fragile' packaging links tactile, visual, and audio cues. [12]

Second, situational interaction drives vocabulary application. Based on interactive meaning theory, multimodal teaching creates immersive real contexts: body gestures simulate actions, and electronic hyperlink trees support independent chunk exploration. Modal synergy of audio-visual-speaking-writing stimulates active output, e.g., imitating dialogues after video, thereby enhancing pragmatic fluency. [13]

Technology empowers multimodal teaching by integrating image-text grammar to realize promptness (attention focus), intuitiveness (cognitive ease), and persuasiveness (engagement).

3.3. The intersection of lexical chunk theory and multimodal theory

3.3.1. The commonality of the theoretical core

The two theories resonate profoundly in language cognition. First, in contextual integration, lexical chunk theory posits that chunks, as grammaticalized lexis, are best acquired holistically within specific contexts, while multimodal theory asserts that meaning generation relies on the contextual interplay of various semiotic resources. Both oppose the conventional separation of vocabulary and grammatical rules, advocating a unified form-function-context dimension. Second, in cognitive economy, the prefabricated nature of chunks reduces memory load through clustering, and multimodal synergy expands retrieval channels via coordinated sensory inputs, jointly easing cognitive burden—an approach well-suited to junior

high students whose short-term memory holds only about 7±2 chunks. [14]

3.3.2. Complementarity of practice path

At the input level, multimodality provides concrete carriers—image, sound, action—for lexical chunks, e.g., "balanced diet" via visual food pyramid, audio nutritionist explanation, and kinesthetic classification gestures, thus satisfying both chunk theory's context embedding and multimodal theory's symbol integration. At processing, gamification turns chunks into multimodal interactive units; in a chunk puzzle, students combine text, picture, and sound to form a complete modal chain, demonstrating chunk theory's holistic extraction and multimodal intermodality. At output, through resemiotization in high-frequency scenarios, pragmatic functions are reconstructed, e.g., in restaurant theatre, students use body movements and menu props to interpret "order a dish," which verifies chunks' communicative value and multimodal theory's social interaction remodeling.

3.4. The characteristics of junior high school students ' cognitive memory

According to the CNNIC's December 2023 survey report on minors' internet use, among Chinese adolescents aged 6-18, there were 193 million internet users, with a penetration rate of 97.2%, 100% connectivity in primary and secondary schools, and 87% of them owning their own internet-capable devices. [15] Teenagers, as ' digital natives ' growing up in the Internet age, the accompanying information technology has developed their unique cognition, attitude, and learning habits. [16] When contemporary adolescents use Internet education applications, their memory activities show a distinct dual feature of ' technology empowerment ' and ' autonomy weakening ':

In junior high English, vocabulary acquisition is fundamental yet conventional methods often fail, leading to low efficiency, short retention, and weak application that hinder comprehensive literacy. Technology—apps with flashcards, spaced repetition algorithms, and multimedia content—improves retrieval speed and retention persistence through personalized, engaging interaction. However, it also induces cognitive offloading, multitasking distraction, and fragmented learning, which undermine autonomous deep processing and understanding—a classic double-edged sword of digital adolescence. Given this, lexical chunk and multimodal theories are well-suited: they address high sensory acceptance and low coding motivation. Multimodal input provides rich contextual cues, while chunk theory supplies prefabricated linguistic units for holistic retrieval, integrating fragmented information into structured chunks (e.g., dietary pyramid images) and reducing cognitive load. Their synergy enables efficient memorization via technological tools and activates deep processing through situational interaction, thus resolving the contradiction between technological dependence and memory weakening, and

ultimately transforming vocabulary instruction for the digital age.

4. Multimodal efficiency theory path design principles

Five principles guide the design: (1) theoretical integration of chunk and multimodal theories via form–function–context and sensory channels; (2) cognitive adaptation to junior high students' concrete thinking and tech sensitivity, transforming chunks into embodied experiences; (3) hierarchical progressiveness following input–processing–output for memory–understanding–application; (4) technological empowerment with AR/VR and corpus analysis for sensory regulation; (5) evaluation feedback combining reinforcement and constructivism to shift chunks from short-term to long-term use. Together, they shape corpora, gamification, and framework, offering a practical paradigm for contextual chunk use per curriculum standards.

5. Teaching Cases: The Path of Unit 4 'Eat Well ' Vocabulary Teaching

5.1. Input layer: Construction and application of multimodal corpus

At the input layer, grounded in multimodal theory and Piaget's concrete thinking, we extract high-frequency lexical chunks from PEP Grade 7 Unit 4, simultaneously collect visual pictures and gestures, auditory audio and rhythm, and tactile models, and structurally annotate them per Lewis' classification to build a form, function, and context multimodal database for vocabulary acquisition.

5.1.1. Extraction and classification of lexical chunks in textbooks

Based on the Unit4 vocabulary of the New People's Education Edition, the software is used to extract high-frequency lexical chunks to form a classification database:

Type	High Frequency Words
Words & Polywords	watermelon、cabbage、dumpling、coffee、salad、porridge、waiter、hamburger、taste、balanced、poor、menu；fish and chips, fast food, soft drinks, Gongbao chicken, Dongpo pork
Collocations	taste sweet, serve dinner, rice goes with meat；balanced diet, too much sugar, put on weight, heart disease; enough food, a common habit
Fixed expressions	What about..., after all, instead of, too...to..., away from
Semi-fixed expressions	What's your choice of...; You'd better (eat vegetables); Can I have...; It's + adj. + to do...（e.g., It's important to have a balanced diet）

Tab 2. High Frequency Words of Unit 4

5.1.2. Multimodal material development

(1) Visual modality: the situational construction of concrete symbols

Based on visual grammar theory, we construct a visual representation system for lexical chunks via an image-meaning mechanism. A food word cloud uses layered photos, forms, phonetics, and category labels to form a four-dimensional "form-sound-meaning-category" symbol group, with color contrast enhancing focus.

A restaurant mind map, centered on a menu, radiates related chunks (waiter, order, dish) with photos and arrows (e.g., menu→order→dish), embedding chunks into contexts and aligning with contextualized holistic acquisition.

(2) Auditory modality: Multidimensional coding of speech symbols

Based on the auditory grammar theory, we construct a prosodic and contextualized voice input system. First, order conversation audio records real customer-waiter interactions, highlighting continuous reading and stress to link phonetic form with pragmatic function. Second, rhythmic word songs incorporate target nouns into 4/4 beat children's melodies, with chunk repetition and context prompts; melody segmentation reinforces phonetic traces, as rhythmic information is easier to store.

(3) Kinesthetic modality: symbolic transformation of embodied cognition

Based on embodied cognition, tactile coding and motor memory of chunks are realized via action symbols. Gesture cards map actions (cutting, kneading) to chunks; imitation activates motor cortex, aligning with kinesthetic involvement. AR stickers trigger dynamic sequences (seeding to cooking) with voice feedback. This embodied design transforms abstract chunks into tangible experiences, solving memory difficulties and suiting junior high tech-driven learning.

5.2. Processing layer: Scene implementation of gamification memory strategy

Based on gamification and multimodal theories, three modal-fusion games are designed for Unit 4 lexical chunks. Through a task chain of symbol conversion, situation simulation, and competitive challenges, these games transform chunk memory from passive reception to active construction, leveraging digital natives' tech adaptability.

5.2.1. Symbol conversion class: Multimodal decoding training for lexical chunk cryptanalysis

Based on the mechanism of cross-channel information integration in cognitive psychology, we design a multimodal symbol conversion task for lexical chunk learning. It constructs a three-dimensional decoding model incorporating visual, auditory, and semantic cues, which requires students to extract target chunks under time constraints. Each password task contains 2-3 modal symbols, e.g., broken fries picture and chewing sound effect cue "fish and chips." For "cookie," students see fragment image, hear oven prompt, and read "Sweet treat" clue, then integrate them to output "cookie" and extend to "a box of cookies." This strengthens memory via cross-modal integration

5.2.2. Situational simulation: embodied semantic construction of VR role-playing in restaurants

Based on situated learning theory, we use Class VR to build an immersive virtual restaurant with menu, waiter, and customer roles. Students complete the

ordering–serving–checkout process in groups: customers use gestures to click items and initiate questions with semi-fixed expressions like "What about...?"; waiters respond appropriately. Mismatches (e.g., "hamburger goes with coffee") trigger red alerts. This embodied VR experience, aligned with multimodal re-symbolization, enables students to shift lexical chunks from formal memory to pragmatic competence through action simulation and language output.

5.2.3. Competitive challenge category: multi-modal collaborative breakthrough of lexical chunk energy station

Based on Vygotsky's collaborative learning principle, we design team-competitive lexical chunk training. The class is divided into 4-5 groups (e.g., Vegetable, Meat) to accumulate points via three multimodal levels. Level 1—listen to environmental sounds (e.g., coffee grinding) and respond with the correct chunk within 5 seconds for 10 points. Level 2—perform body language for a scene (e.g., stagger for fatigue) and other groups guess associated chunks like "put on weight" within 10 seconds. Level 3—modal relay: team members sequentially convert visual (salad image) to auditory ("balanced diet") to kinesthetic (simulating action), with bonus for the shortest time. This competition leverages emotional factors to enhance memory efficiency, and the relay reinforces multi-channel traces through attention-coding-output, aligning with cognitive load theory.

5.3. Output layer: multi-dimensional landing of high-frequency migration mechanism

Based on Wray's pragmatic theory of holistic chunk retrieval and Iedema's re-symbolization concept, we design a three-dimensional migration scenario (classroom–life–virtual space) for Unit 4 lexical chunks. A closed-loop mechanism of multimodal output, real-context application, and dynamic evaluation transforms chunks from memory storage to pragmatic automation. In class, multimodal flashcards use visual-auditory-language triggers: a visual card shows "fries" and "fish and" prompting "chips"; an audio clip "I'm thirsty" elicits "Have a soft drink." At home, students film family diet scenes with English narration (e.g., "Mom is serving porridge with beans"), and software annotates high-frequency chunks. This design leverages digital natives' tech adaptability, turning everyday contexts into chunk acquisition opportunities.

6. Conclusion

Based on lexical chunk theory and multimodal semiotics, this study constructs a three-layer progressive path of "input-processing-output" for junior high English vocabulary memory. The input layer integrates multimodal corpora, the processing layer employs gamified tasks, and the output layer facilitates real-context transfer. Teaching practice in Unit 4 of the New PEP Grade 7 textbook demonstrates that this model effectively promotes the transformation of chunk memory from mechanical

repetition to meaning construction through multimodal integration, gamification activation, and high-frequency application, directly meeting the new curriculum standard's call for contextualized chunk use. Theoretically, it confirms chunks as grammaticalized lexical units and underscores multimodal synergy in activating cognitive channels, offering an interdisciplinary perspective. The practical "multimodal chunk teaching template," empowered by AR/VR and gamification, overcomes traditional bottlenecks like lack of context and shallow memory, providing a feasible paradigm. Limitations include limited sample and basic tech use. Future work should explore chunk-modal matching, extend to reading/writing, and integrate AI for personalized recommendation, pursuing both theoretical and practical breakthroughs.

References

[1] Ministry of Education. English Curriculum Standards for Compulsory Education (2022 Edition).[M］. Beijing: Beijing Normal University Press. 2022.

[2] Zhanzi Li, Danyun Lu. Multimodal Semiotics: Theoretical Foundations, Research Approaches, and Development Prospects. [J]. Foreign Language Research,2012,(02):1-8.

[3] Zhuanglin Hu． From Literacy to Multiple Intelligences [C]. Beijing Forum (2007): The Harmony of Civilizations and Common Prosperity — The Pluralistic Development Model of Human Civilization: Papers and Abstracts of the Foreign Languages Sub-forum on “Language Identity and Evolution in the Conflict and Integration of Pluralistic Civilizations”.2007(4):84-101.

[4] Becker J. The phrasal lexicon [M]. Cambridge: Bolt Beranek & Newman，1975.

[5] Wary A. Formulaic language and the lexicon [M]. Cambridge： Cambridge University Press，2002.

[6] Nattinger. J R, Decarrico. J S. Lexical phrases and language teaching [M]. Oxford： Oxford University Press，1992.

[7] Lewis M. The lexical approach [M]. London: Language Teaching Publications, 1993.

[8] Lewis M. Implementing the Lexical Approach: Putting Theory into Practice [M]. London： Language Teaching Publications, 1993.

[9] Zihong Zhang. On the Application and Significance of the Lexical Chunk Approach [J]. Journal of China Youth University of Political Studies,2008,(06):134-138.

[10] Kress, G. & T. van Leeuwen. Multimodal Discourse: The Modes and Media of Contemporary Communication [M]. London: Arnold. 2001.

[11] Zhanzi Li, Danyun Lu. Multimodal Semiotics: Theoretical Foundations, Research Approaches, and Development Prospects. [J]. Foreign Language Research,2012,(02):1-8.

[12] Zou Bingjun, Zhou Ruizhen, Gan Luoying. A Multimodal Exploration of English Vocabulary Teaching. [J] Teaching and Administration,2013,(36):152-154.

[13] Zou Bingjun, Zhou Ruizhen, Gan Luoying. A Multimodal Exploration of English Vocabulary Teaching. [J] Teaching and Administration,2013,(36):152-154.

[14] Zihong Zhang. On the Application and Significance of the Lexical Chunk Approach [J]. Journal of China Youth University of Political Studies,2008,(06):134-138.

[15] China Internet Network Information Center. The Fifth National Survey Report on Internet Usage by Minors. [EB/OL].https://www.cnnic.cn/n4/2023/1225/c116-10908.html,2023-12-25.

[16] Xiaoqing Gu, Shili Lin, et al. Understanding and Responding: Digital Native Characteristics of Millennial Learners and Their Appeals for Learning Technology. [J].

Modern Distance Education Research,2012,(1):23-29.