Siirt Üniversitesi Eğitim Fakültesi Dergisi (SÜEFD)

Cognitive, Pedagogical, and Technological Perspectives of EFL Listening: The Factor of Vocabulary Development

Melisa KANDEMIR[1], Muhammed Fatih GÖKMEN[2]

1 Siirt Üniversitesi, 0009-0008-3981-3742, melisakandemir06@gmail.com

2 Siirt Üniversitesi, 0000-0002-0868-7623, mehmedfatih1907@gmail.com

ABSTRACT

This article reviews research on second/foreign language (L2) listening as a cognitively and affectively demanding skill that remains underemphasized in instruction. It examines how working memory, cognitive load, anxiety, motivation, and self-efficacy interact with bottom-up and top-down processes to shape listening comprehension. Key listening modalities—extensive and intensive, interactive and passive, academic and real-world, and bimodal/captioned listening—are discussed in terms of their cognitive demands and pedagogical value. The review also explores input-related factors, such as authentic versus simplified materials, multimodal input, modified speech, input flood, and input salience, with particular attention to incidental vocabulary learning. Major instructional approaches, including task-based listening, strategy instruction, integrated skills work, and project- and experiential learning, are evaluated for their contributions to metacognitive awareness, autonomy, and engagement. Finally, the role of digital tools, AI-based applications, and captioned viewing is synthesised within key frameworks, including the Input Hypothesis, Cognitive Load Theory, Dual Coding, Multimedia Learning, and Sociocultural Theory.

Keywords: *Listening comprehension; vocabulary acquisition; captioned video; cognitive load; technology-enhanced learning*

ARTICLE INFORMATION

Article History:
Receipt: *15.11.2025*
Corrected version received: *19.12.2025*
Accepted:*25.12.2025*
Published online:*30.12.2025*
Article Type: *Review article*

Yabancı Dil olarak İngilizce Dinlemenin Bilişsel, Pedagojik ve Teknolojik Perspektifleri: Kelime Gelişiminin Etkisi

Melisa KANDEMIR[1], Muhammed Fatih GÖKMEN[2]

1 Siirt Üniversitesi, 0009-0008-3981-3742, melisakandemir06@gmail.com

2 Siirt Üniversitesi, 0000-0002-0868-7623, mehmedfatih1907@gmail.com

ÖZ

Bu makale, dil öğretiminde yeterince vurgulanmayan, bilişsel ve duygusal açıdan zorlu bir beceri olan ikinci/yabancı dilde dinleme konusundaki araştırmaları incelemektedir. Bu yayında çalışma belleği, bilişsel yük, kaygı, motivasyon ve öz yeterliliğin, dinlemeyi şekillendiren alt düzey (bottom-up) ve üst düzey (top-down) süreçleriyle nasıl etkileşime girdiği incelenmektedir. Kapsamlı ve yoğun dinleme, etkileşimli ve pasif dinleme, akademik ve gerçek yaşam bağlamlı dinleme ve iki kipli/altyazılı dinleme gibi başlıca dinleme yöntemleri bilişsel gereksinimleri ve pedagojik değerleri açısından tartışılmaktadır. Bu inceleme ayrıca, özgün ve sadeleştirilmiş materyaller, çok kipli girdi, indirgenmiş metin, girdi seli ve girdi belirginliği gibi girdi ile ilgili faktörleri, özellikle rastlantısal kelime öğrenimine odaklanarak ele almaktadır. Görev tabanlı dinleme, strateji öğretimi, bütünleşik beceri çalışması, proje ve deneyimsel öğrenme gibi başlıca öğretim yaklaşımları, dinlemede metabilişsel farkındalık, öğrenen özerkliği ve katılım üzerindeki katkıları açısından değerlendirilmektedir. Son olarak, dijital araçların, yapay zeka tabanlı uygulamaların ve altyazılı izlemenin rolü, Giriş Hipotezi, Bilişsel Yük Teorisi, İkili Kodlama, Çoklu Ortam Öğrenmesi ve Sosyokültürel Teori gibi temel çerçeveler içinde sentezlenmektedir.

Anahtar Kelimeler: *Dinleme becerileri; kelime dağarcığı geliştirme; altyazılı video; bilişsel yük; teknoloji destekli öğrenme*

MAKALE BİLGİ

Makale Tarihçesi*:*
Alındı*: 15.11.2025*
Düzeltilmiş hali alındı*: 19.12.2025*
Kabul edildi*: 25.12.2025*
Çevrimiçi yayınlandı*: 30.12.2025*
Makale Türü*: Derleme Makalesi*

Extended Summary

This literature review offers an extensive analysis of second language (L2) listening as a cognitively, affectively, and socially intricate skill that is fundamental yet frequently underestimated in second language acquisition (SLA). The skill is characterised as a transient and demanding process that requires learners to decode, interpret, and integrate fleeting auditory information in real time. Based on Krashen's Input Hypothesis (1985), the review stresses that listening is a key way to receive comprehensible input, but traditional teaching methods have often placed greater emphasis on testing comprehension than on building the cognitive processes that accompany it. Empirical work indicates that successful comprehension depends on a range of cognitive and emotional factors, including working memory, processing speed, anxiety, motivation, and self-efficacy, alongside bottom-up decoding and top-down inferencing. Sweller's Cognitive Load Theory (CLT) (1988) elucidates how task complexity, rapid speech, and inadequately designed materials can overwhelm learners' finite cognitive resources, highlighting the necessity for meticulously scaffolded and multimodal instructional support.

A significant part of the review focuses on the cognitive and emotional aspects of L2 listening. Rather than a passive reception of sound, the process is depicted as dynamic and interactive, requiring integration of linguistic signals with prior knowledge and contextual cues. Working memory plays a central role because it supports the temporary storage and manipulation of incoming speech. Comprehension can be greatly affected by a high cognitive load from many new words, unfamiliar accents, or rapid delivery. Affective factors also have a significant impact: anxiety about listening makes it harder to pay attention and prompts people to avoid it, which aligns with Krashen's Affective Filter Hypothesis. The review emphasises that explicit metacognitive instruction, which teaches students how to plan, monitor, and evaluate their listening, improves both understanding and emotional resilience, encouraging independence and self-control.

The review also brings together research on different types of listening and ways to teach. Extensive listening is an approach based on the principles of comprehensible input that helps students build fluency by letting them listen to enjoyable, authentic, or graded materials without being tested all the time. Intensive listening, on the other hand, focuses on closely examining short texts to improve certain skills, such as breaking them into parts, recognising sounds, and being aware of grammar. Evidence indicates that a balanced model integrating extensive and intensive methodologies produces the most substantial learning outcomes. There are also differences between interactive listening, which is part of communicative tasks that require negotiation and feedback, and passive listening, which occurs in monologic settings like lectures or movies. Academic listening requires organised note-taking, inference, and proficiency in discipline-specific terminology, whereas real-world listening involves spontaneous, context-dependent communication. Dual Coding Theory and Multimedia Learning Theory examine bimodal listening, in which audio is paired with text, such as captions. These theories show how combining audio and video can help with understanding, vocabulary learning, and confidence, especially for learners with lower skills. However, relying too heavily on captions may make it harder for learners to process sound independently.

The review then turns to input manipulation and comprehension enhancement. Authentic materials expose learners to natural language, pragmatics, and discourse features, while simplified materials facilitate comprehension by decreasing complexity. Studies indicate that both play crucial and synergistic roles when properly sequenced. When designed according to cognitive principles, multimodal input that combines audio, images, and text can help people learn new words and make sense of what they read. In the beginning, it helps to change the way you speak by slowing down and speaking more clearly. However, to improve aural skills in the real world, you need to gradually switch to real speech. Input flood and input salience techniques are shown to be very useful for helping people

learn new words by accident, especially when the target words are common, easy to see, and make sense in the context.

There are many different ways to teach, such as task-based listening, strategy instruction, integrated skills teaching, and project-based or experiential learning. Task-based listening embeds comprehension in intentional, meaning-focused activities that boost engagement and facilitate deeper cognitive processing. Teaching students to use cognitive, metacognitive, and socio-affective strategies is an effective way to improve both their performance and their metacognitive awareness. Integrated skills approaches emphasize the inherent interconnection of listening with speaking, reading, and writing, facilitating cross-modal reinforcement of vocabulary, grammar, and discourse competence. Project-based and experiential methods, such as podcasts, interviews, and presentations, enhance listening within genuine contexts, fostering motivation, autonomy, and sustained attention.

Technological advancements are integral to contemporary auditory comprehension. Digital tools like language-learning apps, streaming services, and interactive softwares create personalised, multimodal, and often gamified environments that make learning more fun and help people better understand. Podcasting, vlogging, and interactive media support both receptive and productive skills, fostering authenticity and multimodal literacy. With new developments in AI, like speech-to-text, automatic captioning, adaptive learning algorithms, and listening analytics, it is now possible to give personalised feedback and help people improve specific skills. The review emphasises the significance of critical digital literacy and highlights issues pertaining to privacy, algorithmic bias, and the potential overdependence on automated systems, despite these tools providing unparalleled opportunities for personalised learning.

A thorough investigation of captioned listening provides substantial evidence that captions enhance comprehension, segmentation, and vocabulary acquisition. Full captions help with overall understanding, keyword captions draw attention to important vocabulary, and glossed captions give clear linguistic support for target items. Research consistently demonstrates that L2 captions are more effective than L1 subtitles for language acquisition, as they facilitate processing within the target language, whereas L1 subtitles may enhance comprehension at the cost of auditory processing. Captions serve as an input enhancement method, directing learners' focus to forms without changing meaning, thereby enhancing phonological awareness, grammatical recognition, and lexical development. To encourage independence, it is best to move from full captions to limited or no captions in steps.

The review shows how closely receptive processing and learning new words are connected. Spoken input helps people learn new words both on purpose and by accident. The best ways to teach this are to combine rich exposure with tasks that focus on specific forms. High lexical coverage (95–98%) is demonstrated to be essential for comprehension because of the ephemeral quality of oral input. Repetition, encounter frequency, and semantic richness are recognised as essential elements in vocabulary retention. Well-structured activities, including pre-listening vocabulary preparation, recognition tasks during listening, and post-listening summarisation and discussion, enhance vocabulary acquisition and fortify both receptive and productive knowledge.

Several core theoretical frameworks underlie these discussions: Krashen's Input Hypothesis, Cognitive Load Theory, Dual Coding Theory, Multimedia Learning Theory, and Sociocultural Theory. Together, they explain how input, cognition, multimodality, and social interaction affect aural skills' development. The review concludes that effective listening instruction necessitates a cohesive, research-based methodology that acknowledges cognitive constraints, utilises technological and multimodal resources, cultivates strategic proficiency, and encourages significant, socially facilitated engagement with spoken language.

1. Introduction

Listening comprehension is a crucial yet underrepresented skill in second language acquisition (SLA). Unlike reading and writing, listening is transient, requiring learners to process auditory information quickly amidst contextual and linguistic uncertainties. This makes it particularly challenging for learners. Aural comprehension functions as a key communication channel and source of language input, as posited by Krashen's Input Hypothesis, which emphasises the importance of exposure to comprehensible input just above the learner's current level. However, listening instruction often lacks quality, overly emphasising comprehension tasks rather than the cognitive processes essential for decoding spoken language.

Modern research regards this skill as an active interpretive process that requires both bottom-up and top-down processing. Bottom-up processes involve decoding phonological and lexical elements, while top-down strategies leverage prior knowledge and contextual clues to derive meaning. Additionally, various cognitive and affective factors, including working memory, processing speed, anxiety, motivation, and self-efficacy, complicate aural comprehension. Working memory is particularly crucial in second-language listening, as it enables the retention and processing of auditory information. Sweller's Cognitive Load Theory (1988) highlights that instructional design should accommodate the finite nature of cognitive resources by managing intrinsic, extraneous, and germane loads. Aural tasks that overload cognitive resources can hinder comprehension and retention; therefore, effective listening instruction should incorporate strategies such as pre-listening scaffolding and multimodal support to reduce cognitive demands.

Emotional and psychological factors significantly affect aural processing performance, with listening-specific anxiety reducing attention and triggering avoidance behaviours, as highlighted by Krashen’s Affective Filter Hypothesis (1982). High anxiety for that specific skill correlates with decreased motivation, lower self-esteem, and weaker performance. To combat these issues, metacognitive strategy education has proven effective. It enhances learners' awareness of their cognitive processes, leading to improved listening comprehension, learner autonomy, and emotional resilience. Research indicates that explicit metacognitive training fosters better management of challenging information and deeper cognitive processing.

The effectiveness of aural processing styles on learning is influenced by individual traits and includes various types, such as extensive, intensive, interactive, and passive listening. Extensive listening enhances fluency through enjoyable exposure, while intensive listening focuses on detailed analysis of texts. Interactive listening fosters collaboration, and passive listening is often found in lectures. Recent emphasis on bimodal listening, which combines audio with visual texts like captions, is supported by theories such as Dual Coding and Multimedia Learning. This approach benefits comprehension and motivation, especially for lower proficiency learners. However, over-reliance on captions may impede independent auditory skills, suggesting a gradual shift to minimal or no captions is often recommended.

The design and quality of input significantly impact learners' comprehension and acquisition. Research highlights the benefits of authentic versus simplified input, modified speech, and instructional methods like input flood and salience. Authentic resources expose learners to real language use, while simplified materials aid lower-level learners. Input enhancement strategies, such as highlighting and glossed captions, support vocabulary learning. Furthermore, technological advancements in language instruction, including apps, captioned videos, and AI tools, promote personalised learning experiences. Underlying theories like Krashen's Input Hypothesis (1982), Sweller's Cognitive Load Theory (1988), and Vygotsky's Sociocultural Theory inform these pedagogical practices, emphasising the importance of meaningful input, manageable tasks, and social interaction in language acquisition.

In the overall study, this review examines the key cognitive, emotional, pedagogical, and technological factors that influence second/foreign language aural comprehension. It examines how variables such as working memory, cognitive load, anxiety, motivation, and self-efficacy shape aural processing performance, and explores the effects of different listening modalities, including extensive vs. intensive, interactive vs. passive, academic vs. real-world, and captioned/bimodal listening. The study also considers how input features such as authenticity, multimodality, speech modification, and input salience interact with major instructional approaches like task-based listening, strategy instruction, integrated skills work, and project-based learning, with emphasis on metacognitive strategy use. Furthermore, it analyses the roles of digital tools and caption types in enhancing listening and vocabulary acquisition, and investigates vocabulary learning processes, including incidental vs. intentional learning, lexical coverage, and repetition. Grounded in prominent theories including the Input Hypothesis, Cognitive Load Theory, Dual Coding, Multimedia Learning, and Sociocultural Theory, the study aims to provide a holistic understanding of the factors shaping L2 listening and to offer evidence-based implications for research and classroom practice.

1.1. Cognitive and affective dimensions of listening

Listening is essential in second language acquisition (SLA) but is often overlooked and cognitively challenging. It involves coordinating cognitive and emotional mechanisms to understand spoken language. Recent studies emphasise that aural comprehension in a second language (L2) relies on both bottom-up and top-down processing, working memory capacity, cognitive load, and affective factors like anxiety. Processing models have evolved from basic frameworks to more interactive models that integrate prior knowledge and contextual cues. Cognitive Load Theory suggests that task complexity and poor input design can impede comprehension. Learners' emotional states also play a critical role, with anxiety impacting attention and processing, as supported by Krashen’s Affective Filter Hypothesis, which indicates that higher anxiety can hinder language acquisition and comprehension.

Metacognitive methods are essential for enhancing learners' understanding and emotional resilience. Metacognitive awareness allows students to manage their listening skills effectively, leading to greater confidence and independence, as shown in studies by Cross (2009) and Goh and Taib (2006). This literature review integrates cognitive processing models, working memory, cognitive load, and affective factors to provide a comprehensive view of the cognitive and emotional aspects involved in L2 aural processing. It highlights the need for instructional approaches that not only develop language proficiency but also support students' mental and emotional well-being for effective real-world listening.

1.2. Types and approaches of listening in second language acquisition

Listening is a complex skill crucial to second language acquisition (SLA) and is characterised by cognitive, linguistic, and environmental processes. It is recognised as an active interpretative process that varies with the type and mode of information. Literature categorising that skill into dichotomies such as extensive versus intensive, interactive versus passive, academic versus real-world, and bimodal listening, each highlighting different cognitive, educational, and sociocultural aspects of auditory language processing. The subsequent sections of this chapter will explore these dimensions in detail through the headings Extensive vs. Intensive Listening, Interactive vs. Passive Listening, Academic vs. Real World Listening, and Bimodal Listening, and will conclude with a final section that synthesises the main insights.

Extensive and intensive listening represent key dichotomies in our pedagogy. Extensive listening promotes repeated exposure and supports vocabulary recognition and fluency development (Gökmen, 2021). Extensive listening (EL) draws on the extensive reading movement, emphasising enjoyable,

comprehensible content with minimal instructor intervention or comprehension checks. The theoretical underpinning aligns with Krashen's Input Hypothesis, which suggests that language acquisition occurs when learners receive input slightly beyond their current proficiency level (i+1). In contrast, intensive listening focuses on the detailed analysis of short, carefully selected texts to enhance specific skills, such as phonemic discrimination and grammatical recognition. This method is grounded in skill-learning theory, which asserts that deliberate practice leads to automaticity, with tasks such as transcription and dictation aimed at refining learners' phonological and lexical awareness. Research indicates that extensive listening boosts listening fluency and vocabulary growth, while intensive listening hones attention to linguistic nuances. An educational consensus is emerging in favour of a hybrid model that combines the fluency benefits of extensive listening with the skill development offered by intensive listening tasks.

Interactive listening, which involves reciprocal dialogue and feedback, is contrasted with passive listening, where comprehension occurs without engagement (Brown, 2007). Both Communicative Language Teaching (CLT) and Task-Based Language Teaching (TBLT) promote interactive listening, enhancing comprehension and student engagement (Pica et al., 1989). This approach aligns with Vygotsky's Zone of Proximal Development, emphasising meaning-making through collaboration with more competent speakers, ultimately fostering pragmatic competence and strategic aural comprehension skills such as clarification requests (Rost, 2011). Passive listening, while limited in interaction, offers substantial advantages for improving aural fluency and comprehension. It provides learners with exposure to authentic speech and various genres, particularly in monologic contexts such as lectures, news broadcasts, and films. Field (2008) suggests that passive listening aids in the encounter with genuine speech features, such as reduced forms, connected speech, and discourse markers, particularly when paired with pre- and post-listening activities.

The distinction between academic and real-life processing of spoken input tasks is significant. Academic listening involves formal communication styles typically found in university settings, such as lectures and debates. It emphasises abstract reasoning, organisation of ideas, and the use of academic vocabulary. Essential skills for academic include note-taking, making inferences, and summarising, which require advanced cognitive and metacognitive abilities. Listening in the real world is casual and context-dependent, in contrast to structured forms. It encompasses daily conversations in various settings. Lynch (2001) emphasises that effective listening involves practical knowledge, including understanding meanings and cultural references, as well as addressing conversational issues. Students should learn to manage unclear or incomplete information effectively. A curriculum that integrates both academic and practical listening enhances learners' transferable skills. Gilmore (2007) endorses the use of authentic materials for practical language training, yet standardised assessments prioritise academic listening, influencing teaching focuses (Field, 2008). To address this discrepancy, activities that simulate both contexts are essential to improve students' adaptability across various discourse settings.

Bimodal listening, which combines auditory input with synchronised text, is an effective educational tool supported by Paivio's Dual Coding Theory and Mayer's Cognitive Theory of Multimedia Learning. Research indicates that captions enhance understanding and memory by aiding lexical segmentation, pronunciation identification, and meaning inference. They particularly benefit students struggling with rapid speech. Studies show that captioned videos significantly improve comprehension and retention, while also increasing student motivation and confidence. Zarei (2009) cautions that excessive reliance on captions can hinder the development of essential auditory processing skills. Consequently, educational approaches often implement a gradual shift from complete captions to keyword captions or listening without transcripts. Vanderplank (2016) emphasises that the deliberate use of bimodal input is

crucial for improving comprehension and fostering independence in the advancement of that specific receptive skills.

The classification of listening in second language acquisition into various types, extensive versus intensive, interactive versus passive, academic versus real-world, and bimodal, highlights the complex nature of auditory processing. Each type engages different cognitive and emotional elements, presenting unique pedagogical opportunities. Research advocates for an integrated approach that combines multiple aural comprehension modalities to enhance transferable listening proficiency, enabling educators to improve the precision and fluidity of learners' L2 understanding.

1.3. Input factors and listening input enhancement in second language acquisition

Listening comprehension is vital for second language acquisition (SLA) and is influenced by cognitive processing and linguistic input. Researchers underscore the role of input in language learning and advocate for the careful design and enhancement of specific materials. This literature review explores four key areas of input manipulation: authentic versus simplified materials, multimodal input, modified speech, and input flood with salience, all of which significantly impact learner outcomes in aural comprehension and vocabulary acquisition.

The debate continues over the use of real versus simplified auditory input in second-language learning. Authentic materials, such as interviews and movies, reflect natural language use by native speakers and highlight the complexities of real-world speech, incorporating reduced forms, idiomatic phrases, and other pragmatic elements to engage students. To simplify understanding, graded materials adjust language by altering words and sentence structures, often slowing down speech and prioritising high-frequency vocabulary (Yano, Long, & Ross, 1994). Proponents of authentic materials argue they enhance real-world communication skills and pragmatics (Gilmore, 2007; Rost, 2011). However, their complexity may overwhelm lower proficiency students, hindering comprehension and motivation. Simplified input aligns with Krashen’s Input Hypothesis (1985), which posits that learning occurs when input is slightly beyond the learner’s current proficiency (i+1). Research indicates that simplified texts and authentic input serve complementary roles in language learning. Nation and Newton (2009) recommend starting with simplified materials and gradually moving to authentic ones as learners develop their processing skills. Widdowson (1979) cautions against oversimplification that misrepresents real language use, highlighting the importance of a structured pedagogical approach.

By personalizing the difficult tasks and assessments, AI-driven technologies bring tailored education and more effective assessment of individual progress (Alqahtania et al., 2023).

As technology advances in language education, multimodal input combining sounds, images, and text has emerged as crucial in listening education. Technology can help by personalizing the difficult tasks and assessments, AI-driven technologies bring tailored education and more effective assessment of individual progress (Alqahtania et al., 2023). These approaches are supported by the Cognitive Theory of Multimedia Learning and the Dual Coding Theory, which assert that multisensory information enhances processing and retention. Multimodal input provides contextual cues and visual support, aiding comprehension in inferring, lexical segmentation, and mapping form to meaning. Research indicates that English as a Foreign Language (EFL) learners benefit from watching captioned movies, as they gain more vocabulary and comprehension than those who rely solely on audio. Captions are essential in bimodal listening, aiding both top-down and bottom-up processing. Winke et al. (2010) highlight that captions improve lexical recognition, facilitate fluent speech processing, and reduce anxiety among novice learners. Conversely, Mayer (2009) cautions that poorly designed multimodal input can overload working memory, leading to split attention effects that impair comprehension. Customised multimodal

input, progressing from full to keyword captions and finally to caption-free video, enhances auditory decoding skills supported by visual and textual aids (Vanderplank, 2016).

Modified speech, or foreigner speak, involves adjustments in speech rate, pronunciation, and sentence complexity to enhance comprehension for L2 learners. Key characteristics include longer delivery, precise articulation, extended pauses, and reduced grammatical complexity. The rationale for this speech modification is based on Comprehensible Input Theory and the Interaction Hypothesis, which suggest that such modifications improve understanding and promote interaction. Additionally, Schmidt's Noticing Hypothesis posits that modified speech highlights linguistic forms, making them more apparent to learners. Empirical research demonstrates that moderating speech rate can improve understanding for novice learners. However, over-modification may disrupt natural prosody and intonation, hindering real-world listening applications. Field (2008) noted that overly restricting input can lead learners to rely on artificial speech rather than authentic language. Consequently, while modified speech is important in early education, it should gradually be replaced with more genuine input to facilitate adaptation to authentic contexts. Nation and Newton (2009) recommend structured exposure to natural speech characteristics, such as reduced forms and varying accents, to enhance learning experiences.

Empirical research shows that a moderate reduction in speech rate improves understanding for novices, but excessive modifications may hinder natural prosody, leading to reliance on non-authentic speech patterns. As modified speech is essential in initial education, it should gradually be replaced with authentic input to help learners adapt to real life contexts. Input flood, which entails providing abundant instances of specific language forms, enhances vocabulary acquisition when the target items are relevant and perceptible. Salience in listening can be increased through prosodic stress, visual aids, or contextual redundancy. Acknowledgement of input is necessary for vocabulary gain, and studies confirm that input flood effectively aids incidental learning. Combining input flood with enhancement techniques further draws attention to target forms, while factors such as prior knowledge and working memory significantly affect outcomes. Vandergrift and Goh posit a strategy-oriented model that encourages focused monitoring of key lexical items during listening.

Improving auditory input is crucial for student learning and vocabulary acquisition. Authentic materials enhance fluency, while simplified materials provide essential support. Multimodal input, which combines sound, sight, and text, fosters better processing and engagement. Initially, modified speech aids comprehension, and frequent, meaningful input enhances vocabulary learning. Effective instructional design must consider learners' skill levels, task goals, and cognitive load to create supportive environments that promote long-term retention.

1.4. Instructional approaches to listening in second language acquisition

Listening is now recognised as an active, purposeful activity crucial for communicative competence in second language acquisition (SLA), necessitating effective instruction that transcends traditional drills. Recent pedagogical frameworks emphasise diverse approaches to aural skills, including task-based, strategy, integrated skills education, and experiential learning. This literature review consolidates theoretical justifications, empirical evidence, and educational implications of these methodologies, with a focus on learner engagement, strategic proficiency, and practical relevance.

Task-Based Language Teaching (TBLT) focuses on meaning-oriented activities that foster authentic communication in second-language acquisition. It structures listening activities as objective-driven tasks, enabling learners to extract and analyse auditory information for specific purposes. Different types of task based activities include information gap tasks, where students match or sequence events, and problem solving tasks that encourage discussion. These projects enhance understanding and engagement, often integrating listening, speaking, or writing skills. Nunan (2004) and Vandergrift

(2007) demonstrate that task based increases student engagement, aids in real time information processing, and enhances both inferential and decoding skills. This approach aligns with constructivist learning principles, emphasizing active meaning making, student independence, and contextual language use. Effective task based listening requires careful planning, clear goal setting, and reflective learning. Willis and Willis (2007) advocate for assignments that align with students' skill levels while progressively increasing cognitive and linguistic challenges.

Listening methods are essential for improving learners' comprehension and independence. According to O'Malley and Chamot's taxonomy (1990), aural comprehension strategies are classified into three categories: metacognitive (planning, monitoring, measuring performance), cognitive (inferencing, note taking, summarising), and socio-affective (managing anxiety, motivation, social relationships). Research supports the value of explicit strategy instruction; Vandergrift and Tafaghodtari (2010) found that metacognitive teaching significantly enhanced students' listening scores and self-regulation. Goh and Taib (2006) demonstrated that young learners can also benefit from metacognitive approaches involving prediction, verification, and reflection. Cognitive techniques such as elaboration, translation, and resourcing improve input processing when learners use them flexibly, as noted by Graham (2006). Socio-affective methods are important for alleviating listening anxiety, a recognised obstacle to performance, according to Kim (2000) and Gonen (2009). Vandergrift & Goh (2012) proposed an integrated model that fosters metacognitive awareness in learners, enabling them to select and adapt strategies based on task needs. This approach not only enhances understanding but also encourages learner autonomy and ongoing progress in spoken language compehension skills.

In real-world communication, aural comprehension is typically not an isolated skill. Integrated skills instruction highlights the importance of combining listening with other language skills such as speaking, reading, or writing. This approach aligns with communicative competency frameworks, fostering cross-modal processing and contextual comprehension, ultimately reinforcing input through output. The most prevalent combination is listening and speaking, particularly in interactive settings such as interviews, role-plays, and group discussions. Research indicates that improved aural comprehension enhances speaking clarity and response time. Moreover, combining these skills facilitates advancements in vocabulary, grammar, and the structural organisation of conversations. Listening and reading integration enhances learning in educational and workplace settings by enabling students to engage with lectures while accessing supplementary materials. This method fosters schema activation, inference making, and critical thinking, aligning with high-stakes tests such as TOEFL and IELTS. Grabe's (2009) meta-analysis indicates that integrated training promotes interconnected language skills, deep cognitive engagement, and skill transfer across different modalities. However, it is essential to manage this integration in the classroom to avoid overshadowing vital skills such as speaking or writing.

Project-based Learning (PBL) and experiential learning are increasingly used in Second Language Acquisition (SLA) to foster engagement and collaborative learning through activities such as presentations, interviews, and podcasts. These approaches extend listening beyond traditional assignments to real-world contexts, supporting revisionist and experiential learning theories. They encourage students to engage critically with input, enhancing comprehension and purposeful listening. Research highlights that project based listening boosts motivation, autonomy, and retention. Watanabe and Swain (2007) found that collaborative tasks enhance attention and processing. Coyle, Hood, and Marsh (2010) emphasize their importance in Content and Language Integrated Learning (CLIL) as students use language for subject specific goals. The effectiveness lies in genuineness and learner autonomy, though it requires more time, scaffolding, and guidance. Key to successful implementation are explicit objectives, organized tasks, and reflective exercises linking auditory information to linguistic and cognitive growth.

Instructional methodologies for listening in second language acquisition have evolved to include interactive frameworks such as task-based ones, which emphasises practical communication; strategy instruction, which fosters self-regulated learners through metacognitive training; integrated skills instruction, which connects listening with broader communicative contexts; and project-based listening, which encourages engagement and autonomy. Together, these approaches facilitate a collaborative pedagogy that enhances learners' active engagement in meaning-making, thereby improving aural skills and communicative competence.

1.5. Technology and listening development in second language acquisition

The evolution of digital and intelligent technologies has significantly transformed second-language listening instruction, moving from traditional methods to interactive, learner-centred experiences. As second language acquisition merges with educational technology, the literature explores how digital environments enhance listening skills, enrich input, and foster learner autonomy. This review focuses on three main areas: digital tools, podcasting and interactive media, and AI-driven applications, analysing their theoretical foundations, pedagogical uses, and empirical effectiveness.

Digital tools such as learning apps (Duolingo, ELSA Speak), video streaming services (YouTube, Netflix), and audio software (EnglishCentral, FluentU) are increasingly used in second-language training. These resources offer multimodal input, adaptive feedback, and self-paced learning, creating an environment that traditional classrooms struggle to replicate. Their effectiveness is rooted in constructivist and interactionist theories, emphasising meaningful interaction with the content. Video platforms, for instance, enable simultaneous audio and caption viewing, supporting Paivio's Dual Coding Theory and Mayer's Multimedia Learning Theory, which highlight that dual information channels enhance comprehension and retention. Research indicates that tools such as captioned videos enhance understanding and vocabulary acquisition (Montero Perez et al., 2014), and that digital listening environments increase learner engagement and motivation (Cross, 2011). Mobile applications such as Duolingo provide gamified auditory exercises, promoting practice and interest with minimal risks (Loewen et al., 2020). However, challenges remain, including screen fatigue, variable quality of user-generated content, and limited interactive features in some apps. Consequently, experts emphasise the need for pedagogical mediation, wherein teachers guide students in strategically and critically utilising digital resources (Hubbard, 2009).

The advent of user-generated content and Web 2.0 technology has allowed educators and students to integrate podcasting, vlogging, and interactive digital media into listening instruction. These tools emphasise authentic input, learner autonomy, and content creation, aligning with connectivist and sociocultural learning theories. Podcasting, as both a receptive and productive task, facilitates learner engagement with authentic information in various formats, enhancing sustained attention, aural fluency, and cultural competency. Additionally, student-produced podcasts encourage deeper processing, criticality , and verbal expression. Video blogging and interactive media platforms like Flipgrid, TikTok, and Edpuzzle enhance education by promoting active engagement through visual cues and peer interaction. These tools allow for the integration of comprehension questions to transform passive viewing into active listening. Platforms such as VoiceThread facilitate asynchronous communication, blending auditory reception with verbal expression. While they offer benefits such as authenticity and improved multimodal literacy, challenges include technical issues, complex assessments, and the need for instructional support to navigate interactive content effectively.

Recent advancements in artificial intelligence (AI) and machine learning have created new opportunities for developing aural skills. AI-powered tools such as speech-to-text engines and adaptive learning systems offer personalised and interactive experiences tailored to individual learners. Tools like Google's

Live Transcribe and Otter.ai enhance real-time captioning and transcription, significantly improving form-meaning mapping and comprehension for students, particularly those who face challenges in these areas. Research indicates that voice recognition technologies facilitate better understanding and note-taking among students (Bui & McDaniel, 2015; Huang et al., 2021). Listening analytics in platforms like EnglishCentral and Yabla provide feedback on students' listening habits by tracking actions such as replaying, pausing, and correct responses. This data enables teachers to customise lessons and allows students to monitor their progress. Additionally, AI-driven adaptive learning tools like Speakly and LingQ adjust input difficulty and feedback based on learner performance, reflecting principles similar to Dynamic Assessment (DA) frameworks, which focus on providing feedback within the learner's developmental zone.

Experts caution against over-reliance on AI in education, highlighting potential issues such as data privacy, algorithmic bias, and dehumanisation. CALL academics stress that AI should complement, not replace, human teaching and advocate for concurrent digital literacy training (Godwin Jones, 2021).

Technology has transformed the teaching of listening skills in second-language environments through digital tools such as apps and video platforms that create adaptable, gamified learning experiences. Innovative methods such as podcasting and interactive media foster engaging activities, while AI tools offer adaptive feedback and data-driven training. The effectiveness of these technologies depends on their strategic integration into teaching frameworks that promote learner autonomy and teacher support. Future research should explore how emerging technologies, including voice AI, eye tracking, and virtual reality, can further enhance the L2 listening experience.

1.6. Captioned viewing and listening in second language acquisition

Adding captioned audio and video to second language learning environments enhances aural skills and vocabulary acquisition. Captioning aids comprehension by providing written text alongside audiovisual content, facilitating word learning, form meaning mapping, and improved listening fluency. This part consolidates evidence and theory regarding the effects of various caption types, the significance of L1 and L2 subtitles, and captions' role in vocabulary learning and input improvement.

Captions vary in format and information depth, primarily categorised as full captions, keyword captions, and glossed captions. Full captions, which provide literal audio transcriptions, are well researched and enhance learners' understanding by aiding decoding and segmentation. Studies by Montero Perez et al. (2014) and Winke, Gass, and Sydorenko (2010) indicate that complete captions significantly improve aural comprehension and support incidental language acquisition by visually reinforcing auditory signals. Keyword captions reduce cognitive tension by limiting text, helping learners focus on key vocabulary, especially benefiting those with lower skills. Park (2014) found that keyword captions assist in attention management without overwhelming learners. Glossed captions provide more detailed support, improving vocabulary retention and understanding, as shown by Peters et al. (2016). However, they can increase cognitive load, with effectiveness contingent on learners' reading speed and memory capacity (Zarei, 2009). Each form of caption serves unique pedagogical purposes: full captions enhance overall understanding, keyword captions improve focus, and glossed captions support vocabulary acquisition. A progressive scaffolding approach is recommended, initially using whole or glossed captions and gradually moving to keyword or no captions, to balance support and challenge (Vanderplank, 2016).

The distinction between captions (L2) and subtitles (L1) significantly impacts language learning outcomes. Captions facilitate the simultaneous processing of visual and auditory information in one language, whereas subtitles require translation between L2 audio and L1 text, potentially disconnecting from the auditory input. Research shows that L2 captions generally enhance language development more

effectively than L1 subtitles. Specifically, L2 captions improve comprehension and vocabulary acquisition, as highlighted by Markham and Peter (2003). Meanwhile, L1 subtitles may help comprehension but do not contribute to language acquisition and can hinder phonological processing and lexical retrieval by reducing reliance on auditory signals. Several studies suggest that L1 subtitles can be helpful for beginners lacking sufficient L2 comprehension. A proposed solution is to use multilingual subtitles that display L1 and L2 text simultaneously, which, despite being cognitively taxing, may aid in form-meaning integration and comparison noticing.

Captions serve as an effective tool for vocabulary acquisition, exposing learners to words in context without direct instruction. They enhance visibility and awareness of new language elements, which is crucial for language learning. Research has demonstrated that captioned videos significantly aid in vocabulary retention and comprehension. For instance, a study by Montero Perez et al. (2013) found that students exposed to captioned TV shows recalled more target vocabulary than those who were not. Similarly, Peters and Webb (2018) found that consistent interaction with captioned audio-visual content notably improved learners' understanding and usage of vocabulary. The effectiveness of captions in vocabulary acquisition is determined by several factors, such as frequency of exposure, word difficulty, input modality, and learner proficiency. Nation and Webb (2011) emphasise the importance of repeated, spaced exposure to lexical items across multiple viewing sessions to enhance learning retention. Captions aid in the learning of multiword units, including collocations and phrasal verbs, which are often difficult to master through isolated word lists. The research conducted by Boers and Lindstromberg (2009) underscores the educational value of integrating lexical chunks within meaningful audio-visual contexts.

Captions are a form of input enhancement that highlight the significance of linguistic elements, drawing learners' attention without altering the meaning of the content (Sharwood Smith, 1993). They clarify the connection between spoken and written language, enabling learners to recognise grammatical gaps, create hypotheses, and link phonetics with spelling. In the context of instruction, captioning supports Processing Instruction (VanPatten, 2004) and the Multimodal Information Hypothesis (Plass & Jones, 2005), both advocating for structured information to aid in form-meaning mapping and reduce cognitive load. Research indicates that captions enhance the learning of morphosyntactic elements such as verb tenses, articles, and word order. Winke et al. (2010) found that learners exposed to captioned movies demonstrated a better understanding of grammatical structures and improved in speaking and writing activities. Additionally, captioning promotes phonological development by helping learners identify word boundaries, stress patterns, and intonation contours. Vandergrift (2011) states that bimodal input helps learners bridge auditory information with its written form, thereby enhancing phonological encoding. Captions are a vital educational tool that enhances understanding and focused learning, particularly within a structured listening curriculum that incorporates pre, during, and post viewing activities.

Captioned viewing and listening significantly enhance second language acquisition. Different caption types (full, keyword, glossed) serve various educational purposes, such as comprehension and vocabulary enhancement. Subtitles in the second language (L2) are generally more effective than first language (L1) subtitles, which may aid understanding but limit auditory processing. Captioning fosters incidental vocabulary acquisition, functioning as an effective input enhancement method that supports grammatical and phonological development. For optimal results, captions should be integrated into structured, task-oriented activities to direct learner focus and promote active engagement with audio-visual materials.

1.7. Listening and vocabulary learning in second language acquisition

Listening plays a vital role in second language acquisition (SLA), as it provides essential language input that supports vocabulary development. Unlike reading, which allows for visual confirmation and revision, aural processing requires immediate comprehension of transient information, rendering the acquisition of vocabulary both cognitively demanding and intricately linked to learning. In examining how listening contributes to vocabulary learning, this section will also address key dimensions such as incidental versus intentional vocabulary acquisition, the lexical coverage needed for effective comprehension, the role of frequency and repetition in strengthening vocabulary retention, and how specific tasks can be used to reinforce vocabulary learning, culminating in a concluding discussion that synthesizes these perspectives.

Auditory processing can facilitate vocabulary acquisition both incidentally and purposefully. Incidental learning often occurs through exposure to real audio materials, aligning with Krashen's Input Hypothesis, which emphasizes understanding. However, it may result in incomplete knowledge of word usage. In contrast, purposeful acquisition involves strategies like pre-teaching and focused attention on form-meaning relationships, which studies suggest improve memory retention, particularly for productive vocabulary. Current research supports a hybrid methodology that combines intentional tactics with incidental listening. Peters and Webb (2018) found that learners using captioned films with vocabulary-focused exercises demonstrated improved vocabulary recall.

Lexical coverage is the percentage of words in auditory texts that learners need to understand for comprehension; at least 95% is necessary for basic understanding, and 98% for optimal comprehension. Unlike reading, listening provides less context and fewer opportunities to revisit content, making high lexical coverage crucial. Insufficient coverage can lead students to focus too much on specific words, hindering overall comprehension and vocabulary acquisition. Vocabulary acquisition necessitates that learners recognise and extract new words from continuous speech, which requires a balance between clarity and difficulty in the input. Graded or scaffolded resources, such as simplified podcasts or captioned videos, can improve lexical coverage for learners at different levels of competence.

Frequency and repetition are crucial for vocabulary retention in second language learning, with learners needing to encounter a word 6 to 10 times for comprehension (Nation, 2013). Aural input poses challenges due to its transient nature, but tools such as captioned videos and interactive transcripts can effectively provide repetition benefits similar to those of written forms (Webb & Rodgers, 2009). Research by Peters (2012) and Vidal (2011) indicates that repeated exposure to words across different contexts enhances memory retention. Semantic richness contributes to deeper vocabulary by illustrating the diverse contexts in which words are used. Resources that present language across various syntactic and pragmatic contexts are more effective at enhancing both receptive and productive vocabulary.

Listening tasks play a critical role in enhancing vocabulary acquisition and retention. They encompass various activities such as comprehension checks, gap fills, dictogloss, and focused vocabulary tasks, enabling students to concentrate on information retrieval and connect new inputs to their prior knowledge. Pre-listening activities prepare learners by introducing essential vocabulary to facilitate understanding. During listening tasks, such as transcribing and multiple choice quizzes, allow students to track new words in context. Post-listening tasks, including summarisation and discussions, aid in reinforcing memory and deepening comprehension. Research indicates that tasks emphasising vocabulary, particularly those involving delayed recall and form-focused feedback, are highly effective. Including both explicit and incidental learning components is vital to enhancing the breadth and depth of vocabulary knowledge.

Listening is crucial for vocabulary acquisition in second-language learning, offering both incidental and deliberate learning opportunities. Factors influencing vocabulary learning include lexical coverage, repetition, and semantic context. Utilising aural exercises and scaffolded materials, such as captions, enhances retention and understanding. A beneficial approach is to integrate incidental exposure with deliberate attention to improve vocabulary through listening.

1.8. Theoretical frameworks relevant to listening in second language acquisition

Listening is essential for effective second language acquisition (SLA), involving cognitive, emotional, and social processes for real-time comprehension. Various theoretical frameworks aid in understanding and enhancing aural skills. This includes Krashen's Input Hypothesis, Cognitive Load Theory, Dual Coding Theory, Multimedia Learning Theory, and Sociocultural Theory, which inform instructional materials, learner behaviour analysis, and pedagogical strategies in listening education.

Stephen Krashen's Input Hypothesis is a key theory in Second Language Acquisition, asserting that learners achieve linguistic proficiency most effectively with input that is understandable yet slightly above their current competence level (i+1). It highlights the importance of "comprehensible input," particularly through engaging and meaning-focused activities. Krashen distinguishes between acquisition, an unconscious process akin to first language learning, and learning, a conscious understanding of rules. The theory advocates extensive listening practices with authentic audio materials, promoting richer language exposure than traditional drills. Critics of the Input Hypothesis argue it overlooks the importance of output, interaction, and form-focused training, yet Krashen's emphasis on input has impacted various teaching methods, such as extended listening and captioned viewing.

Cognitive Load Theory (CLT) examines learning through the lens of cognitive psychology, emphasising the constraints of working memory. Introduced by Sweller in 1988, "cognitive load" refers to the mental effort required for learning. CLT categorises cognitive load into three types: intrinsic load (material difficulty), extraneous load (presentation issues), and germane load (mental resources allocated for learning and schema development). In aural comprehension, Communicative Language Teaching (CLT) plays a crucial role by helping learners process rapid auditory input. Challenges such as high working memory demands, fast speech, or poor audio quality can impair understanding. To alleviate cognitive strain, auditory materials should be designed thoughtfully, using strategies like segmenting audio, providing vocabulary pre-listening, or integrating visual aids. CLT also supports the use of multimodal resources, such as subtitles and visual organisers, which enhance comprehension while respecting cognitive limits, thereby furthering multimedia and dual-coding approaches. Teng (2024) demonstrates that learners' working memory capacity significantly moderates vocabulary gains from captioned viewing, supporting Cognitive Load Theory by showing that multimodal input is most effective when cognitive demands remain manageable.

Allan Paivio's Dual Coding Theory (DCT) posits that information is processed in two interrelated systems: a verbal system for linguistic data and a nonverbal system for imagery. This theory suggests that when learners engage both auditory and visual modalities, these systems work together to improve recall and comprehension. In the context of language acquisition, especially through auditory exercises using audiovisual materials, DCT illustrates how integrating spoken language with relevant visuals or subtitles enhances understanding. For example, associating the word "guitar" with an image of the instrument allows for a more effective recall than hearing the word alone. Furthermore, Paivio's framework offers important implications for auditory education that employs visual aids. Tools such as captions, gestures, and contextual images reinforce the auditory input, making abstract language more understandable. Additionally, students facing auditory processing challenges benefit significantly from

this dual-channel information approach, contributing to a more inclusive and effective listening experience.

Mayer's Cognitive Theory of Multimedia Learning (CTML) posits that learning is most effective when individuals utilise both auditory and visual channels, addressing cognitive limitations while engaging in deep cognitive processing. CTML is based on three key assumptions: dual channel processing, the limited capacity of working memory, and the necessity for active learning through the selection, organisation, and integration of information. Mayer has outlined criteria for effective multimedia training that are particularly relevant to listening activities. The document outlines key principles for optimising video-based activities in language acquisition, including the Coherence, Redundancy, Signalling, Segmenting, and Modality Principles. These principles advocate for minimising distractions, combining spoken and visual elements, emphasising essential information, and structuring content in manageable segments. By incorporating keyword captions and visual cues aligned with storytelling, educators can enhance understanding and retention, ensuring that aural tasks are effective in technology-enhanced language learning environments.

Sociocultural Theory (SCT), rooted in Lev Vygotsky's research, emphasises the social aspects of learning, particularly the roles of interaction, mediation, and scaffolding in cognitive development. Vygotsky introduced the Zone of Proximal Development (ZPD), indicating the difference between what learners can achieve independently and with support. In SCT, listening is seen as an active, collaborative process enhanced through engagement in diverse environments, where students interact with skilled speakers for feedback and meaning clarification. Donato's concept of collaborative scaffolding highlights peer assistance in tasks beyond individual capabilities, especially in pair or group listening activities that involve prediction, clarification, and discussion to improve auditory processing. SCT highlights the importance of learning tools and artefacts such as audio texts, subtitles, and discussion platforms. These technologies shape learners' cognitive environment, affecting their processing and understanding of spoken language. Listening instruction, as per SCT, emphasizes interaction, collaboration, and the social creation of meaning.

The enhancement of aural skills in second language acquisition can be understood through various theoretical frameworks. Krashen's Input Hypothesis highlights the need for intelligible information, while Cognitive Load Theory and Multimedia Learning Theory address cognitive limitations and effective design principles for listening activities. Dual Coding Theory emphasizes the benefits of audiovisual input, and Sociocultural Theory points out the importance of social interaction and scaffolding in cognitive development. Together, these frameworks form a robust foundation for effective instruction that supports learners' cognitive, social, and linguistic growth.

2. Conclusion

This thorough examination of second-language listening underscores its complexity as both a cognitive and an emotional process, essential to language acquisition. Often overlooked, listening is a demanding skill that requires quick interpretation of fleeting auditory information, challenging working memory and cognitive resources. Several theoretical frameworks, such as Krashen's Input Hypothesis, Sweller's Cognitive Load Theory, Paivio's Dual Coding Theory, Mayer's Multimedia Learning Theory, and Vygotsky's Sociocultural Theory, illustrate the complex interplay between input, cognition, multimodality, and social mediation in the development of second language listening.

Being shaped by bottom up and top down processing, influenced by cognitive factors such as processing capacity and personal characteristics like anxiety and motivation. Metacognitive strategy training enhances learners' ability to manage their aural comprehension. The distinction among various listening types underlines the importance of diverse inputs in education. Key characteristics affecting

understanding include authenticity, simplicity, multimodality, and modified speech. Techniques such as input flooding and salience enhancement support incidental learning, particularly when combined with scaffolding and focused feedback. Additionally, captioned viewing with different caption types aids comprehension and vocabulary development.

Technological innovation has significantly transformed listening pedagogy through digital platforms, podcasts, and AI, creating multimodal and gamified learning experiences. These advancements promote learner autonomy and motivation while providing adaptive feedback, contingent on intentional instructional integration. Modern methodologies emphasise active participation, practical usage, and the development of strategic skills, moving beyond traditional comprehension exercises to include task-based listening and experiential projects.

Improving aural comprehension capacity in second-language acquisition requires a comprehensive educational approach that integrates cognitive psychology, sociocultural theory, and multimedia learning. Effective instruction must address cognitive limitations, leverage multimodal resources and technology, and encourage learner autonomy through strategic training, facilitating meaningful interaction with spoken language in various contexts.

References

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189–208.

Chang, A. C. S., & Millett, S. (2014). The effect of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal, 68*(1), 31–40.

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta: Journal des Traducteurs, 49*(1), 67–77.

Ellis, R. (2003). *Task based language learning and teaching*. Oxford University Press.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367–413.

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal, 57*(4), 325–334.

Field, J. (2008). *Listening in the language classroom*. Cambridge University Press.

Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice*. Cambridge University Press.

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*(1), 55–75.

Gökmen, M. F. (2021). The role of extensive listening in raising students' metacognitive awareness of listening skill. ELT Research Journal, 10(2), 162–189.

Guillory, H. G. (1999). The effects of keyword captions to authentic French video on learner comprehension. *CALICO Journal, 15*(1–3), 89–108.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge University Press.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

Krashen, S. D. (1985). *The Input Hypothesis: Issues and implications*. Longman.

Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task induced involvement. *Applied Linguistics, 22*(1), 1–26.

Long, M. H. (1985). Input and second language acquisition theory. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 377–393). Newbury House.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2013). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning, 27*(1), 79–98.

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology, 18*(1), 118–141.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59–82.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Nation, I. S. P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. Routledge.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Heinle Cengage Learning.

Peters, E. (2012). Learning vocabulary through listening to oral and written stories: Do modality and story complexity matter? *Studies in Second Language Acquisition, 34*(1), 113–138.

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 40*(3), 551–577.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Teng, M. F. (2024). Incidental vocabulary learning from captioned viewing: The roles of prior vocabulary knowledge and working memory. System, 120, 103170.

Toscu, S. (2025). A Systematic Review on AI Technologies in English as a Foreign Language Education. Sakarya University Journal of Education, 15(2), 156-172.

Van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609–624.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219–258.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46–65.

Webb, S., & Rodgers, M. P. H. (2009). The vocabulary demands of television programs. *Language Learning, 59*(2), 335–366.

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology, 14*(1), 65–86.

Zarei, A. A. (2009). The effect of bimodal, standard, and reversed subtitling on L2 vocabulary recognition and recall. *Journal of English Language Teaching and Learning, 53*(212), 65–85.

Zarei, A. A., & Rashvand, F. (2011). The effect of interlingual, intralingual, and reversed subtitling on L2 vocabulary recognition and recall. *The Journal of Teaching Language Skills, 3*(1), 95–117.

Makale Bilgi Formu	
Yazarın Katkıları	Makale iki yazarlıdır.
Çıkar Çatışması Bildirimi	Yazar tarafından potansiyel çıkar çatışması bildirilmemiştir.
Destek/Destekleyen Kuruluşlar	Bu araştırma için herhangi bir kamu kuruluşundan, özel veya kar amacı gütmeyen sektörlerden hibe alınmamıştır.
Etik Onay ve Katılımcı Rızası	"Cognitive, Pedagogical, and Technological Perspectives of EFL Listening: The Factor of Vocabulary Development" başlıklı çalışma etik kurul onayı gerektirmemektedir. Yazım sürecinde bilimsel, etik ve alıntı kurallarına uyulduğu, toplanan veriler üzerinde herhangi bir tahrifat yapılmadığı yazar tarafından beyan edilmiştir.